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

Doc. 2.1
(21.II.2011)
v1.0.0.3

3rd GRUAN Implementation-Coordination Meeting (ICM-3)
Queenstown, New Zealand
28 February – 4 March 2011

Session 2

The GRUAN Manual

(Submitted by Greg Bodeker)

Summary and Purpose of Document

The purpose of this manual is to establish the operational philosophy under which GRUAN will operate and to inform current and future GRUAN sites of the expected modus operandi for GRUAN. It defines the requirements for GRUAN site operations, including requirements on expected accuracy, longterm stability, and uncertainty measures. The description of this document as a ‘manual’ is consistent with the WMO nomenclature i.e. it is a document that provides higher level directives and where underlying ‘guides’ provide more detailed and specific information. Therefore, rather than prescribing the methods, techniques and processes that should be employed in GRUAN, it provides higher level principles that are intended to direct the development of the methods, techniques and processes needed to achieve the stated goals of GRUAN. Where possible, the document does provide more in-depth detail on specific methodologies appropriate for incorporation into existing WMO literature.

THE GRUAN MANUAL

Version 1.0.0.3

PREFACE

The purpose of this manual is to establish the operational philosophy under which GRUAN will operate and to inform current and future GRUAN sites of the expected *modus operandi* for GRUAN. It defines the requirements for GRUAN site operations, including requirements on expected accuracy, long-term stability, and uncertainty measures. The description of this document as a ‘manual’ is consistent with the WMO nomenclature i.e. it is a document that provides higher level directives and where underlying ‘guides’ provide more detailed and specific information. Therefore, rather than prescribing the methods, techniques and processes that should be employed in GRUAN, it provides higher level principles that are intended to direct the development of the methods, techniques and processes needed to achieve the stated goals of GRUAN. Where possible, the document does provide more in-depth detail on specific methodologies appropriate for incorporation into existing WMO literature.

Table of Contents

Executive Summary

The development and current operation of the GRUAN network is managed through a number of distinct but often overlapping documents, including GCOS-112, GCOS-121, GCOS-134, web-based material, reports from GRUAN Task Teams and papers published in the international peer reviewed literature. The purpose of this manual is not to supersede that documentation, but rather to provide a vehicle for communicating and documenting messages important to the operation of GRUAN but which may not find a natural home elsewhere in the GRUAN body of documentation. As such, this document is neither complete nor comprehensive in its coverage of GRUAN. The high level messages emerging from this GRUAN manual are summarized in this executive summary.

Goals: 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.
- Provide a calibrated reference standard for global satellite-based measurements of atmospheric essential climate variables.
- Fully characterize the properties of the atmospheric column.
- Ensure that potential gaps in satellite programmes do not invalidate the long-term climate record.

Partner networks: GRUAN will not operate in isolation but will connect with a number of existing networks, some of which are already making measurements pertinent to GRUAN. Duplication with these networks should be avoided. Wherever possible, QA/QC techniques developed within those networks should be adopted within GRUAN.

Managing change: GRUAN will not be a static network. Change, resulting from the availability of new, improved instruments, the generation of new knowledge about calibration procedures, and the adoption of more exact standards, will be inevitable. Such changes need to be managed carefully to avoid introducing discontinuities in long-term measurement time series.

Measurement uncertainty: A focal point for GRUAN, and one which differentiates it from many other networks, is the emphasis on deriving robust values for the uncertainty on each measurement. This involves a process of describing and analyzing all sources of uncertainty in any measurement, quantifying and synthesizing the contribution of each source of uncertainty to the total measurement uncertainty, and verifying that the derived net uncertainty is a faithful representation of the true uncertainty.

Key requirements for GRUAN sites: Three essential climate variables have been identified as priority 1 measurements for GRUAN, viz.: temperature, water vapour, and pressure. The goal is to measure these through:

- 1 weekly production radiosonde flight using the best technology currently available at the site.
- 1 monthly radiosonde capable of measuring water vapour in the upper troposphere and lower stratosphere and all other priority 1 variables to the best level possible with current technology, launched together with the weekly radiosonde.
- Regular 00 and 12 LST (as a preference over UTC) launches of a production radiosonde with best technology currently available.
- Dual launches of sondes with highest quality humidity sensing capability in the upper troposphere and lower stratosphere.
- Periodic intercomparisons of a large range of sonde types.

Only the first two criteria are deemed an initial requirement. Equally important as implementing this measurement schedule is establishing and documenting the methods used to quantify the uncertainty on each measurement. No measurement programme from any site should be adopted into GRUAN until a detailed, traceable account of the measurement uncertainty has been established. Therefore, in additional to implementing the required measurement programmes, sites should also be encouraged to develop detailed documentation around these measurement programmes which can then also be used to trace ongoing improvements in measurement precision and accuracy, and in the derivation of measurement uncertainties.

1. INTRODUCTION

1.1. A brief summary of GRUAN

The reliable detection of the vertical structure of changes in climate variables in the atmosphere requires very high quality atmospheric observations with well characterised measurement uncertainties. While the GCOS Upper Air Network (GUAN) provides upper air measurements over large regions of the globe, these are primarily for operational weather forecasting and as a result seldom include systems to guarantee data quality such that the data are suitable for long-term trend detection. For example GUAN does not include, as part of its design, regular intercomparisons between measurements at different sites to ensure homogeneity in data quality and traceability. In addition GUAN does not provide global coverage; in particular ocean regions are poorly sampled.

The need for a reference upper-air network to better meet the needs of the international climate research community has long been recognized. This was formalized between 2005 and 2007 when a reference upper-air network consisting of eventually 30-40 sites worldwide was planned. This is the GCOS Reference Upper-Air Network (GRUAN; GCOS-112, GCOS-134). In contrast to GUAN, which is based on weather observing stations, GRUAN is specifically designed for climate research. GRUAN will provide reference observations of upper-air essential climate variables (ECVs), through a combination of in situ measurements made from balloon-borne instruments and from ground-based remote sensing observations. Furthermore, unlike other GOS networks, management decisions in GRUAN are driven by the requirements of long-term climate trend detection. Nonetheless, there are aspects of GRUAN operations that clearly link to GUAN. As such GRUAN has a somewhat split personality with a dual-purpose nature. On one hand the GRUAN network is a research network constantly striving to improve measurement techniques, quantify and reduce measurement uncertainties, and improve precision and accuracy. On the other hand the measurements need to be made in a stable way over decadal time scales to achieve data homogeneity both in time and between measurement stations. In this sense GRUAN will operate more like a long-term monitoring network for the detection of climate change. These two aspects of GRUAN operations are not mutually exclusive, but do need to be carefully balanced. This dual-purpose nature of GRUAN has been accommodated in this manual.

GRUAN's goals are to:

i) Provide vertical profiles of reference measurements suitable for reliably detecting changes in global and regional climate on decadal time scales. The uniformity and coherence of standard operating procedures at GRUAN stations and the resultant homogeneity of GRUAN data products will provide a global reference standard for GUAN stations. In this way improved detection of changes in the climate of the troposphere and stratosphere will be achieved.

ii) Provide a calibrated reference standard 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 detection of trends in climate in the upper troposphere and stratosphere.

iii) Fully characterize the properties of the atmospheric column. This is necessary for process understanding and for radiative transfer modelling.

iv) Ensure that potential gaps in satellite programmes do not invalidate the long-term climate record.

In achieving these four goals, GRUAN will address the current deficiencies of the GUAN network. In the context of the WMO networks, GRUAN will effectively be the climate reference backbone of the existing GUAN. 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 are not repeated here.

1.2. GRUAN Governance

A schematic outline of the GRUAN governance structure is given in Figure 1. GRUAN measurement sites are guided directly by the GRUAN Lead Centre, currently hosted by the Lindenberg Meteorological Observatory, Germany. The Lead Centre is responsible for implementation of GRUAN, for managing various systems that apply to GRUAN as a whole, and for collecting and integrating best practices across the network. The GRUAN Lead Centre is designated by WMO who also sponsors the GCOS Steering Committee. The GCOS steering committee in turn guides the GCOS/WCRP Atmospheric Observation Panel for Climate (AOPC). The AOPC in turn guides the Working Group for Atmospheric Reference Observations (WG-ARO) which guides the development of GRUAN, is responsible for GRUAN site selection (see Section 5.2), develops guidelines for observations and data and ultimately guides the GRUAN Lead Centre. The GCOS Secretariat provides additional support to the GCOS Steering Committee, the AOPC, the WG-ARO and the GRUAN Lead Centre. The GRUAN Lead Centre acts as the interface between GRUAN and the community of users of GRUAN products. For example, data transfer to end-users is not made from GRUAN measurement sites but is first shared within the GRUAN community, subjected to the QA/QC procedures developed within GRUAN, and then submitted by the GRUAN Lead Centre to the GRUAN data repository (NCDC – see Section 8.5).

GRUAN aims to me more than the sum of its Lead Centre and measurement sites. To achieve this GRUAN might benefit from having access to resources that can be used to address issues that are relevant across the network as a whole. An example would be the development of a system for

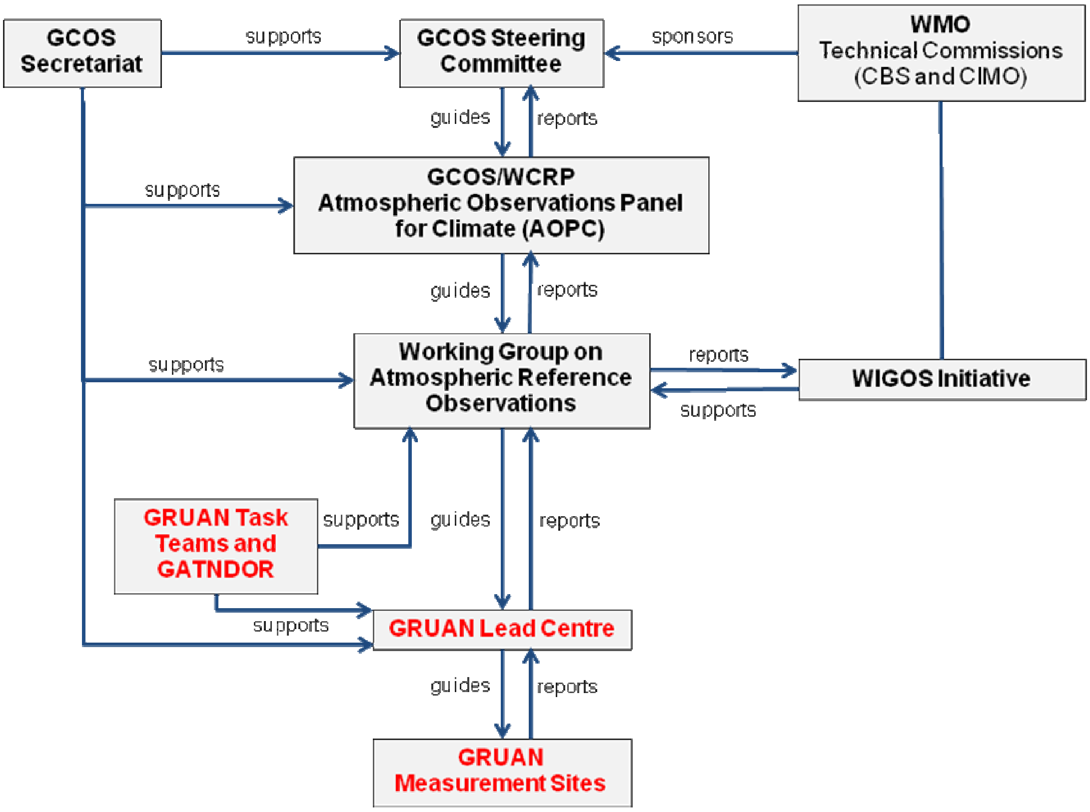

Figure 1: Schematic outline of the structure of GRUAN. GRUAN elements are shown in red while external support structures are shown in black. GATNDOR=GRUAN Analysis Team for Network Design and Operations Research.

reporting bugs in data by the GRUAN data user community. To this end a trust fund, established in a similar manner to the GCOS Cooperation Mechanism, might provide the necessary facility to receive voluntary contributions for implementing activities across the GRUAN network for the communal benefit of the network.

1.2.1. Internal GRUAN structure

Internally, GRUAN comprises the Lead Centre, GRUAN measurement sites, GRUAN task teams, and an ad hoc GRUAN Analysis Team for Network Design and Operations Research (GATNDOR). The Lead Centre is responsible for implementing GRUAN, system management and for collecting and integrating best practices. GRUAN measurements are made at the Lead Centre and at GRUAN measurement sites which should replicate the measurement practices recommended by the Lead Centre. The GRUAN Lead Centre may also conduct targeting training programmes for instrument operators at various GRUAN sites to encourage uniformity of instrument operation between sites.

GRUAN task teams support the WG-ARO and the Lead Centre in implementing GRUAN by considering specific issues in support of network design and decision making, and entraining operational and other relevant expertise. The task teams 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 should 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 should also be used to support the Lead Centre in guiding individual stations through instrumental and operational changes without impacting long-term measurement time series. Possible avenues for expansion of task teams are discussed in Section 1.3.1. Following ICM-1, the GATNDOR team was established to undertake scientific investigations (in additional to the more operational investigations undertaken by the task teams) in support of GRUAN decision making and to report at subsequent ICM meetings. The team undertakes focused, short-term research to address specific topics identified by the GRUAN science and management community. GATNDOR activities are coordinated with the GRUAN task teams and with national GCOS programmes when appropriate. To best serve the needs of climate monitoring and research, it is essential that GRUAN be informed by a good understanding of the evolving science issues that drive the measurements and accuracy of the GRUAN data. Therefore, the establishment of an internal or external science advisory panel is being considered.

1.3. Links to partner networks

GRUAN will not operate in isolation of existing networks and GRUAN is not intended to replace in any way any existing networks. In fact many GRUAN initial and candidate sites already belong to existing networks such as GUAN, GAW, NDACC, BSRN and SHADOZ. One of the essential characteristics of a successful GRUAN is 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 collocation database to enable cross-calibration 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 GRUAN is required on a continuous basis.

There are 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, expert teams from existing networks should be approached to support GRUAN operations and to avoid duplication of effort by utilizing existing scientific knowledge. It is especially important to note that establishing GRUAN is not just an exercise in adding another acronym label to existing measurement sites. While in the 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 will require a mode of operation, and the establishment of measurement programmes, currently not available anywhere. The purpose of this section is to provide, as early as possible in this document, a context for GRUAN in the broader community of climate monitoring networks.

A number of networks currently in operation make measurements which fall within the scope of GRUAN. Of particular interest are those stations that make upper air measurements that are not part of the typical meteorological measurements of temperature, pressure and humidity. Many of these networks 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 cross-calibrated and can be quantitatively linked).

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

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

The NDACC comprises more than 70 high-quality, remote-sensing research stations 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. Because GRUAN and NDACC share a number of common science goals, it has been debated whether GRUAN is necessary and whether NDACC could achieve the goals of GRUAN. There are a number of key differences between NDACC and GRUAN that require GRUAN to operate as a new and independent network, including:

- 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 factors driving changes in climate. This is evidenced by the fact that NDACC does not include measurements of incoming and outgoing longwave and shortwave radiation, nor measurements of various cloud parameters such as cloud amount/frequency, base height, layer heights and thicknesses, cloud top height, cloud top pressure, cloud top temperature, and cloud particle size. While GRUAN does include measurements of a few trace gases (ozone, methane, and carbon dioxide) it excludes the wide range of trace gases measured within NDACC.
- 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 humidity.
- NDACC operates as a federation of independent measurement sites. While NDACC does have in place stringent standards which must be met for measurement programmes to become part of the network, unlike GRUAN, the NDACC network is not controlled by a Lead Centre that aims to implement standard operating procedures across the network as a whole.
- One of the primary goals of GRUAN is to detect long-term climate trends above the Earth's surface. NDACC does aim to make long-term measurements of changes in chemical composition in the upper troposphere and stratosphere but this is not the primary focus of the network.

- Because NDACC does not operate under the control of GCOS, it does not have the institutional mandate to act as the reference standard for the GUAN which is a key purpose of GRUAN.

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:

- The NDACC has established 'working groups' that are primarily centred on specific instruments used within the NDACC. GRUAN task teams currently include a mix of teams focussing on specific measurements systems (radiosondes and precipitable water) and on network wide operational issues. Some consideration should be given to later expanding the ‘Ancillary Measurements’ Task Team to include specific measurement systems, and then to have 'crosscutting' teams that focus on issues common to the network as a whole. This could be achieved through assigning ‘instrument mentors’ as recommended in GCOS-112. Task teams focussing on specific measurement systems or on specific ECVs would better link to advisory groups within partner networks e.g. the Scientific Advisory Groups within GAW (see Section 1.3.4). SCOPE-CM (see Section 1.4) intends to establish one or two centres to lead the generation and provision of fundamental climate data records for each ECV and so establishing task teams within GRUAN focussed on specific ECVs or groups of ECVs would mirror the structure within SCOPE-CM and thereby facilitate interactions with the satellite-based measurement community (one of the key clients of GRUAN).
- Measurements of vertical ozone and water vapour profiles made within the 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.
- Measurement redundancy in the NDACC network 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).

In light of the commonalities between the GRUAN and NDACC networks, consideration should be given to including an NDACC representative on the GRUAN steering committee to ensure close cooperation and coordination between these two networks.

1.3.2. BSRN (Baseline Station Radiation Network)

The BSRN provides a worldwide network to continuously measure radiative fluxes at the Earth's surface. The network comprises about 40 stations between 80°N and 90°S many of which began operation in 1992 and each year more stations are added to the network. These stations provide data for the calibration of the GEWEX Surface Radiation Budget (SRB) Project and other satellite-based measurements of radiative fluxes. BSRN data are also used to validate radiative flux models. BSRN data are archived at the Alfred Wegener Institute (AWI) in Bremerhaven, Germany. In 2004, BSRN was designated as the global surface radiation network for the GCOS. The BSRN stations also contribute to GAW (see Section 1.3.4).

The primary goal of BSRN is to monitor the background shortwave and longwave radiative components and their changes with the best methods currently available. Therefore the measurements of longwave and shortwave incoming and outgoing 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) would be highly advantageous to GRUAN. The BSRN includes a working group on measurement uncertainties (currently led by Bruce Forgan of the Australian Bureau of Meteorology) that should provide guidance for establishing the radiation measurement uncertainties within GRUAN.

1.3.3. WOUDC (World Ozone and UV Data Centre)

The WOUDC is one of the World Data Centres which are part of the GAW (see Section 1.3.4) 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 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.
- 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 multi-decadal trend detection.

1.3.4. GAW (Global Atmospheric 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. As such, GAW improves understanding of the interactions between the atmosphere, the oceans and the biosphere. As with the NDACC, the primary focus of GAW is on changes in atmospheric composition. GAW has strong linkages to GCOS and as such is likely to have skills and resources that could be used to support GRUAN.

1.3.5. SHADOZ (Southern Hemisphere Additional Ozonesondes)

The SHADOZ project was initiated to remedy the lack of consistent tropical ozonesonde observations by augmenting ozonesonde launches at operational ozone observing stations (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 target research questions such as quantifying the wave-one pattern in equatorial vertically resolved ozone.

1.3.6. 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 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 spectral aerosol optical depth (AOD), inversion products, 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 the level 2.0 data that are primarily 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 implement additional quality checks.

1.3.7. Atmospheric Radiation Measurement (ARM) Programme

The goal of the 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.

Of particular interest to GRUAN, ARM has a dedicated Data Quality (DQ) Office which was established in July 2000 to coordinate and implement efforts to ensure the quality of the data collected by ARM field instrumentation. The DQ Office has the responsibility for ensuring that quality assurance results are communicated to data users so that they may make informed decisions when using the data, and to ARM's Site Operations 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. 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.

1.3.8. Meteorological agencies

Meteorological agencies producing global reanalyses (e.g. NCEP/NCAR, ECMWF, JMA and NASA) are likely to be users of the high quality data produced by GRUAN. 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). The additional value provided by the GRUAN measurements in such data assimilations should be quantified since this would provide additional scientific justification for GRUAN operations. Once a sufficiently large database of GRUAN measurements has been accumulated, such a study could be undertaken through collaboration between the GATNDOR group within GRUAN and perhaps the SPARC data assimilation activity. 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. Care will need to be taken when comparing satellite-based measurements against the GRUAN reference e.g. by smoothing the GRUAN vertical profile measurements to match the intrinsic resolution of the satellite-based measurements (Rodgers and Connor, 2003).

1.4. Link to satellite-based measurement programmes

GRUAN will provide data sets, not currently available, that will be useful to the satellite measurement community for calibrating and validating satellite-based sensors, and for removing offsets and drifts between satellite-based data sets when creating merged data products. Because the GRUAN measurements are likely to serve a wide range of end-users within the satellite measurement community, this manual recommends that a task team/working group be established within GRUAN to liaise with key clients within the satellite community, and with other data providers, to ensure that GRUAN data products are tailored, where possible, to best meet the needs of this community. Once GRUAN datasets are available, pilot studies on enhanced datasets using these reference measurements need to be undertaken.

1.4.1. Calibration and validation of satellite-based sensors

To be useful for climate monitoring, satellite radiances require calibration against a ground truth to unambiguously remove non-climatic influences (Ohring et al., 2005). GRUAN and the GSICS (Global Space-Based Intercalibration System) are complementary in meeting this need. The data products derived from the satellite-based radiance measurements also require validation and this is usually achieved through comparison of the derived data products with independent ground-based measurements. Vömel et al. (2007a) demonstrate how reference-quality in situ water vapour measurements can be used to validate current satellite-based observations.

New satellite missions have higher resolution and better station-keeping, resulting in better control of diurnal sampling. Global Positioning System Radio Occultation (GPS RO) measurements are also highly promising, at least for upper-tropospheric and lower-stratospheric temperature. Even though they represent a significant step forwards, these more recent satellite observing systems will not be adequate for climate purposes unless they can be suitably validated. To this end GRUAN will also provide shorter-term quality assured measurements for the validation of satellite-based retrievals.

The need for inter-station homogeneity within GRUAN has special significance for validation of satellite-based measurements. If satellite-based measurements agree well with ground-based measurements made at one GRUAN station but disagree with measurements made at another, this will significantly weaken the utility of GRUAN measurements for satellite instrument validation.

The issue of measurement scheduling within GRUAN to accommodate satellite validation activities is discussed further in Section 7.1.

1.4.2. Creating global homogeneous trace gas data bases

While satellite-based measurements have the advantage of providing global or near-global geographical coverage, the quality and usefulness of the measurements is compromised by an inability to conduct regular calibrations, limited vertical resolution, difficulties in continuity due to drifting orbits (which, for species showing strong diurnal variation can alias into apparent trends), and limited instrument lifetimes which require data series from multiple instruments to be spliced together to form long-term data records. Discontinuities between satellite-based measurements of climate variables, while not important for weather forecasting purposes, can be ruinous for detecting long-term changes in climate. The reference measurements that GRUAN will produce can be used to remove offsets and drifts between these separate satellite-based measurement series i.e. GRUAN will provide a 'gold-standard' that will serve as a common baseline when splicing satellite-based measurement time series. Specifically, differences between a given satellite-based data set and the GRUAN gold standard can be analyzed using the algorithms detailed in Alexandersson et al. (1997) and Khaliq et al. (2007) to 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 gold standard. Similar approaches using the global ground-based Dosbon and Brewer spectrophotometer networks to create long-term global total column ozone records from multiple satellite-based measurements have been developed (Bodeker et al., 2001).

By contributing to the creation of global homogeneous trace gas data bases, 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 can 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'.

2. REFERENCE MEASUREMENTS

2.1. The concept of a reference measurement

As denoted by its title, GRUAN will provide reference quality measurements for a range of upper-air climate variables. Reference quality atmospheric observations are based on key concepts in metrology, in particular traceability. Metrological traceability is the process whereby a measurement result, i.e. a measurement and its uncertainty, can be related to a reference through a documented, unbroken chain of calibrations, each of which contributes to the measurement uncertainty.

A reference measurement does not refer to a measurement that is perfect, nor to a measurement that will never change. Rather it refers to 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. Reference measurement accommodate the unavoidable sources of uncertainty in the compilation of the net measurement error while excluding those source 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 systematic sources of uncertainty must be avoided.

A reference measurement may not necessarily be the outcome of a measurement by a single instrument but may be an average of measurements from one instrument or an average of results from multiple instruments. 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 is expressed as measurement uncertainty and is a property of the measurement that combines instrumental as well as operational 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. An important point is that within GRUAN this uncertainty will be vertically resolved and each measurement in a profile will be treated as a single measurement result requiring both the measurement and its uncertainty. To provide the best estimate 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 calibrating or validating other systems, have been defined in Immler et al. (2010).

A reference measurement typically results from a measurement procedure that provides sufficient confidence in its results by relating to well-founded physical or chemical principles, or a measurement standard that is calibrated to a recognized standard, in general a standard provided by a National Metrological Institute (NMI). 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 systematic errors have been identified and corrected, and that the uncertainty of these 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.

Another important requirement is that the methods by which the measurements are obtained and the data products derived must 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 GRAUN data for decades to come. They should 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.

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

2.2. Managing Change

GRUAN recognizes that change is inevitable – changes in instrumentation, changes in operating procedures, changes in data processing algorithms and changes in operators. Such changes introduce sources of operational uncertainty into GRUAN data products. Rather than designing a system that is resistant to change, GRUAN appreciates that without change, improvement is impossible. Therefore, the goal is to manage change in a way that does not compromise the integrity of the long-term climate records being measured. To this end the GRUAN network must develop detailed guidelines for managing change. One of the core tasks for GATNDOR is to develop that guidance.

The first focus in managing change is to ensure that when transitioning from older to newer instrumentation, that a sample of coincident measurements, sufficient to quantify any biases between the two systems, is obtained before the older system is retired. For example 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). The length of time for which the older and newer systems should be run in parallel, and the frequency with which coincident measurements should be made, will depend on the instruments used and on an in-depth understanding of the measurement technique. Such decisions should be informed by robust scientific investigations (e.g. by GATNDOR). Until the results of such research are available, sites should err on the side of caution and undertake a super-saturation approach to overlap so that sub-sampling can be undertaken later to determine a minimum safe level of overlap required to preserve the record. Where it may not always be feasible to operate older and newer instruments side-by-side for extended periods of time e.g. with balloon-borne instruments, alternating between the newer and older instruments is particularly useful in diagnosing and correcting systematic inter-instrument differences; regression analysis techniques including a basis function that is set to 1 for one measurement system and to 0 for another can be used to extract biases between the two systems. These biases can be derived as functions of other state variables such as air pressure, temperature, time of day, solar zenith angle etc.

As new and more in-depth knowledge of various measurement systems is gained, reprocessing of historical data will be necessary. Data reduction processes and data archiving within GRUAN need to be designed with this in mind i.e. that the original raw data (which must always be archived) can be easily and regularly reprocessed, as required, to form a single homogeneous time series that is then provided to end-users. Each change in instrumentation, operating procedure or data processing algorithm is likely to require reprocessing of historical raw data. Protocols need to be established to indicate when reprocessing of the full measurement record is justified. Every reprocessing generating a new homogeneous time series over the complete measurement period should be reflected in a change in the data version and such updates need to be communicated to users who have accessed earlier versions of the data (see Section 8.5). For this reason it is also important that all older versions of any data set are always made available through the GRUAN archives.

A discussion of specific sources of changes is presented below but in general this requires dealing with breakpoints in the measurement time series. It is far more preferable that these changes are identified *a priori* through the available meta-data that identifies such changes. However, it is also possible to identify breakpoints in measurement time series based on the statistical behaviour of the data themselves. Significant resources and techniques have already been developed within the surface climate community around this issue (see e.g. http://www.homogenization.org).

These techniques must be grounded in quantitative understanding of the causes of offsets and drifts between two different measurement systems i.e. the reliance should not be on the implementation of signal processing techniques that identify and correct for offsets and drifts in time series. This quantitative understanding in turn should emerge from the meta-data associated with each measurement and from in-depth knowledge of each measurement system.

2.2.1. Changes in instrumentation

Changes in instrumentation are both inevitable and desirable if they lead to more precise measurements of the true atmospheric state. Instrument changes will also often be driven by the necessities of production engineering (when instrument components become unavailable or too expensive) and decisions will have to be made as to what level of component change requires additional change testing. 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 should be a pre-requisite for the adoption of any instrument within the GRUAN network. Outcomes from such intercomparisons would form an important component of the meta-data archived at GRUAN. GRUAN need not necessarily organise these intercomparisons themselves. WMO and partner networks (e.g. NDACC) often run instrument intercomparison campaigns and GRUAN should participate in these 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.

Following a scientifically robust replacement strategy that maximises the maintenance of long-term climate records will be important for ensuring the integrity of the GRUAN data products in the face of change. GATNDOR has been tasked with developing such scientifically robust strategies. Specifically a goal within the 'Management of Change' research topic of the GATNDOR team is to provide scientific bases to develop operational practices to better manage instrument changes at GRUAN sites and to accurately merge disparate data segments to create a homogeneous time series (led by June Wang). Consideration will need to be given to the desired strategy when more than one station 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.

Measurement redundancy (see 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. The same advantages could be achieved through the use of a travelling standard instrument. For *in situ* balloon-borne instruments consistent ground-check routines between new and old instruments will minimize changes in procedural uncertainty contributions.

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 will 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. HMEI has suggested that a workshop specifically for manufacturers and open to all HMEI members would be helpful.

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

2.2.2. Changes in operating procedures

Even if instruments themselves do not change, changes in the operating procedures for an instrument may also introduce breakpoints 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. reprocessing of historical data to homogenize the time series and redistribution of the data with an updated version number will almost certainly be required. The expectation is that standard operating procedures for all instrument types within GRUAN will be archived at the Lead Centre and changes in standard operating procedures at individual stations will be managed through the Lead Centre.

2.2.3. 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. At the very least 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 one location and one location only (either the Lead Centre or some other GRUAN site with particular expertise in that measurement), changes in data processing algorithms should 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 even if those changes are well document and follow the prescriptions listed above. This more central, 'top-down' approach to data processing is different to 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 stations, but also between stations. In support of maintaining consistency in the use of data processing algorithms within GRUAN, the Lead Centre should be tasked with maintaining an archive of data processing algorithms which then also comprise an important part of the meta-data archive for GRUAN.

Tension may arise where a site may wish to implement a non-standard (at least non-standard for GRUAN) data processing algorithms 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.

2.2.4. Change 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. Meta-data should include codes (not names to protect the privacy of operators) to denote where different operators have been responsible for measurements.

3 MEASUREMENT UNCERTAINTY

3.1 Estimating measurement uncertainty

No measurement can be made perfectly and estimating measurement uncertainty is a central tenet in GRUAN's operations. A common GRUAN definition of measurement uncertainty and a common procedure to establish measurement uncertainties is required to homogenize uncertainty estimates across the network. It is also needed to make the steps leading to the determination of measurement uncertainty traceable. This common definition should, ideally, be adopted by instrument providers as well.

Achieving a useful 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 quality 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 and will elevate the GRUAN measurements above what is currently available.

The availability of sufficiently detailed meta-data is vital to quantifying random and systematic errors in measurements. The more detailed the meta-data, the 'deeper' the measurement uncertainty can be traced. The approach that should be followed is that where some calibration, reference standard, application of an operating procedure, or use of a data processing algorithm introduces a source of uncertainty into a measurement, complete details about that uncertainty source must be available through the meta-data tagged to that measurement. Such sources of meta-data 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 all sources of measurement uncertainty are made transparently available to end-users of GRUAN measurements.

A particular challenge for GRUAN in estimating 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, viz. balloon-borne sondes, are used for measurements of single profiles. The well calibrated instruments with quantified measurement errors 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.

Because one of GRUAN's primary goals is to detect long-term climate trends in the upper atmosphere, the primary consideration might be to work towards reducing the random error in measurements i.e. to emphasize reproducibility. However, 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, reducing systematic errors to achieve the best possible accuracy also needs to be a priority. 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 mantra for dealing with measurement uncertainty should be:

i) *Describe/Analyze* all sources of measurement uncertainty.
ii) *Quantify/Synthesize* the contribution of each source of uncertainty to the total measurement uncertainty.
iii) *Verify* that the derived net uncertainty is a faithful representation of the true uncertainty.

3.1.1. Describe/Analyze sources of measurement uncertainty

The first step in the process of deriving an uncertainty associated with any measurement is to first fully explore and describe each source of uncertainty in the form of systematic and random errors. Contributions to the net measurement uncertainty are likely to include sensor calibration, sensor integration, sensor 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 in some way on another sensor that performs less well, this source of uncertainty needs to be accounted for. For example, if a very precise and accurate temperature measurement is made but the vertical coordinate for that measurement is a less precise pressure measurement, in the presence of large $\partial T/\partial p$, the uncertainty in pressure can introduce significant uncertainty in the temperature measurement. Therefore uncertainty in the geo-location and time coordinates associated with each measurement should also be considered when identifying and describing sources of measurement uncertainty. A full list of sources of measurement uncertainty will be defined in the GRUAN common definition of measurement uncertainty terms. Every GRUAN station should measure, collect, and provide all information necessary to establish an uncertainty budget for every measurement.

3.1.2. Quantify/Synthesize sources of uncertainty

The second step is, where possible, to quantify and correct for any systematic biases. Uncertainty in such bias corrections, which must also be diagnosed, documented and quantified, then contributes to the random error on the measurement. Once all systematic biases have been corrected for, and assuming all remaining random errors are normally distributed about the mean, the resultant net error on the measurement can be reported as a single value i.e. the first standard deviation of the distribution (1σ errors). Where systematic biases cannot be determined, or perhaps can be determined but cannot be corrected for, or when remaining random errors are not normally distributed about the mean, a different approach will be required for quantifying the net uncertainty on the measurement. In such cases because the net error is no longer represented by a Gaussian distribution, it cannot be reported as a single value. Techniques to fully describe the shape of the error distribution 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).

3.1.3. Verify measurement uncertainties

The uncertainty budget for every GRUAN measurement should be verified at regular intervals using redundant observations from complementary instruments (see Section 6.2). 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 uncertainties. If measurements agree within their uncertainties, the error estimates on the measurements are more likely to be correct. Formal methods have been developed to achieve this (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 error bars, then either a systematic effect in either or both measurement sets has been overlooked, or the uncertainties have been under-estimated. On the other hand, if much less that 32% of measurement differences are smaller than the 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 error 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. In the case of remote sensing methods, the calibrated data are in physical units of radiance and/or frequency, which are then analyzed to provide estimates of the underlying climate variable of interest. Validation of data products, which is equivalent to verifying measurement uncertainties, is therefore a two-step process whereby the accuracy of both the instrument calibration and the analysis algorithm, are validated.

3.2 Reporting measurement uncertainty

An overarching principle for the operation of GRUAN is that no measurement should be provided without also providing an estimate of the measurement uncertainty. Where all sources of systematic error in the measurement have been identified and corrected for, the measurement uncertainty can be quoted as the standard deviation of the random error. As discussed above, where systematic biases remain in the measurement, or where the net random error in the measurement does not follow a Gaussian distribution, alternative methods for reporting the measurement uncertainty should be considered. This may be in the form of establishing 1σ upper and lower bounds on the measurement uncertainty to denote that the uncertainty is asymmetric – generally reported as X_{-l}^{+u} where X is the measurement, u is the 1σ uncertainty in the positive direction and l is the 1σ uncertainty in the negative direction. 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).

3.3 Reducing measurement uncertainty

Changes in instrumentation or standard operating procedures may lead to reductions in measurement uncertainty. In such circumstances it is important that the same detail of uncertainty analysis is conducted for the new instrument/operating procedure as has been done for the instrument/operating procedure to be replaced.

In some circumstances, e.g. in the presences of high natural variability, reducing measurement uncertainty has little impact on derived trends since the primary source of the variability in the trend estimate might be the noise on the signal being analyzed. It is therefore important that scientific analyses guide where reducing measurement uncertainties is most likely to lead to reductions in uncertainties in trend estimates.

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. 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 should be developed at the GRUAN Lead Centre and then disseminated to all sites making that particular measurement.

3.5 Validating measurement uncertainty

Once the uncertainty on a measurement has been calculated, the question then becomes: how well does this measure of uncertainty represent the degree of confidence we should have in this measurement? Two approaches are available for validating the derived uncertainty on any measurement, viz.:

3.5.1. Comparison of redundant measurements

A traditional way of validating measurement uncertainty is to measure the quantity of interest through two (or more) techniques, based on physically different measurement principles. Because the different techniques are subject to unique 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*.

3.5.2. Laboratory analysis of the measurement system

The ability to simulate a specific measurement in the laboratory can permit an in-depth investigation of the various sources of uncertainty in the measurement. For example, the environmental simulation facility at the Research Centre Juelich (Smit et al., 2007) has provided information to validate measurement uncertainty in ozonesondes.

4 ESSENTIAL CLIMATE VARIABLES MEASURED IN GRUAN

Since GRUAN's goal is not only to provide long-term high quality climate records, but also the ancillary data required to interpret those records, a number of parameters in addition to the fundamental atmospheric state variables of temperature, pressure, humidity and wind will need to be measured. High quality measurements of atmospheric state variables, trace gas concentrations, the atmospheric radiation environment, and cloud and aerosol properties will be required. Many of these parameters have been identified by GCOS as Essential Climate Variables (ECVs; GCOS-92). A subset of ECVs has been selected as the most scientifically important and most tractable for GRUAN (see Appendix 1 of GCOS-112). As scientific research into the underlying causes of observed changes in upper-air climate advances, and as the capabilities of GRUAN sites expand, this list is likely to grow.

4.1 Justification and context for Essential Climate Variables

The complete list of ECVs targeted by GRUAN is listed in Appendix 1 of GCOS-112. The purpose of this section is to provide additional scientific justification and context, and more general guidelines for the measurement requirements for those ECVs listed as priority 1 for GRUAN, viz. temperature, pressure, and water vapour. Similar material for the priority 2, 3 and 4 variables is provided in Appendix B. As such this section provides clear expectations for the measurement of priority 1 ECVs for GRUAN sites. However, this manual recognizes the heterogeneity of the network and its state of development. Therefore, the requirements imposed on current and putative GRUAN sites, as detailed in GCOS-112, may not be immediately achievable. In such cases the 'Site assessment, certification and expansion' Task Team (see Appendix A) will provide possible incremental approaches to achieving the target attributes for each measurement. Because the desired operations parameters for each of the ECVs are based on the scientific requirements of the data and not on current instrument performance, they may not be currently achievable. Therefore, as stated in GCOS-112, these GRUAN requirements should be interpreted as eventual measurement goals of any given network site. Setting these parameters ambitiously high may discourage potential sites from joining GRUAN since they may not be able to immediately achieve these standards. On the other hand setting the parameters low is likely to result in stagnation since once achieved there will be little incentive to advance. For this reason the tables below are different to classical WMO/CBS requirement tables and should be interpreted in a different manner to WMO/CBS requirement tables. The values in Appendix 1 of GCOS-112 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. In no case should an inability to achieve these targets result in the exclusion of a site or a measurement programme from the GRUAN network. This manual recognizes that GRUAN is less about meeting prescribed measurement standards and more about establishing an approach that continually strives to improve measurement precision and accuracy, extend the range of coverage, and achieve higher sampling.

The measurement ranges prescribed in Appendix 1 of GCOS-112 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. However, measurement precision 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. Measurement frequencies are not specified because they may vary over time. Measurement accuracy refers to the systematic error of a measurement (the difference between the measured or derived value, and the true value). 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 error over time, which is a critical aspect of the reference network. To ensure that realistic climate trends can be derived from the dataset, the effect of any intervention to the measurement system on measurement error, such as a change in instrument, should be smaller or quantified to a much greater degree than the value given for long-term stability. Long-term stability is a measure of the acceptable systematic changes to the measurements on multi-decadal timescales. The requirements stated in Appendix 1 of GCOS-112 are largely consistent with the GCOS ECV observation requirements, as detailed in the WMO/CEOS database.

4.2 Priority 1 ECVs

4.2.1. Temperature

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. Other drivers of atmospheric temperature changes, e.g. changes in solar output, would not have the same vertical profile fingerprint. Remaining discrepancies between temperature trends derived from satellite-based measurements and from radiosondes weaken the attribution of changes in temperatures to changes in GHGs. High quality temperature measurements within GRUAN will contribute to the resolution of these discrepancies.

Because radiosondes will remain the primary workhorse within GUAN for the measurement of temperature, pressure and humidity, it is imperative that GRUAN sites establish state-of-the-art radiosonde measurement programmes that continually strive to improve the quality of radiosonde measurements. Other measurement techniques can and should be developed to extend the height range of the temperature profile measurements and to improve the precision and accuracy of 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 to 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/NCAR and ECMWF.
- Provide input to numerical weather prediction models if and when submitted shortly after the measurement. Upper-air measurements of temperature and relative humidity are two of the basic measurements used in the initialization of numerical weather prediction models for operational weather forecasting.

Satellite-based measurements of this ECV will be provided by MSU (Microwave Sounding Unit) instruments and by GPS radio occultation (RO) measurements. However, these measurements are unlikely to extend deep into the troposphere and so GUAN radiosonde measurements are likely to remain the primary data set for trend detection in this region. Recent research has shown that the RO technique has the potential to provide high resolution profiles of atmospheric refractive index in the middle to lower troposphere, which combine the effects of temperature and water vapour in this region. Requirements for precision, accuracy and long-term stability need to be guided by the requirements of end-users and in particular the requirements for detecting trends in temperature time series which include natural, unforced climate variability. This becomes a signal-to-noise ratio problem and climate models should 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 4th assessment report).

It is particularly important that trends in the tropical cold point tropopause temperatures are accurately detected since this controls the flux of water vapour into the stratosphere (Gettelman et al., 2002) 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). 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) would be particularly advantageous.

Measurement range: Ideally temperature measurements should cover the range 170 – 350 K to span the range of measurements encountered between the Earth's surface and the upper stratosphere. Currently available technology can meet this requirement.

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). Ideally GRUAN measurements of the vertical temperature profile should extend from the surface to ~50 km. Vertical temperature profiles are most routinely measured using radiosondes which seldom reach above ~35 km altitude (noting that radiosondes flown to provide input to NWP models aim only to reach ~25 km). However, if used to provide a reference standard for temperature over the lower portion of satellite-based measurements of the vertical temperature profile, and then if combined seamlessly with those satellite-based measurements, the goal of achieving coverage from the surface to the stratopause (and even higher) would be achieved. Ideally temperature profiles from the surface to the upper stratosphere/lower mesosphere, measured by a single instrument, should be the GRUAN goal since these would provide the most robust signal of climate change. Use of GRUAN radiosonde temperature profiles as a standard for other GUAN stations would increase the geographical coverage in the troposphere.

Vertical resolution: Given that it is primarily balloon-borne instruments that provide high resolution profiles of the vertical temperature profile in the atmosphere, a resolution of 100 m or better below 30 km altitude and a resolution of ~500 m above 30 km altitude is appropriate.

Precision: ≤0.2 K in measurement repeatability.

Accuracy: Uncertainties of ≤0.1 K in the troposphere and ≤0.2 K in the stratosphere. This is significantly more stringent than the 0.5 K in the troposphere and 1 K in the stratosphere prescribed in WMO-No. 8 and is currently unrealistic since the perhaps most accurate temperature sonde, the 'Accurate Temperature Measuring Radiosonde' (Schmidlin, 1991), claims an uncertainty of 0.3 K throughout most of the upper troposphere and the stratosphere. This suggests that while GRUAN should proceed with the best technology available, emphasis also needs to placed on the development of new technology to achieve higher accuracy. It might be that higher accuracy is achievable from nighttime soundings where the radiation correction, the dominant source of uncertainty in the stratosphere (Immler et al., 2010), is significantly reduced. Accuracy can also be improved by reducing systematic biases in the measurements. For temperature this may be partially accomplished by using a three-thermistor set with different radiative properties (e.g. white, black and silver) to quantify the uncertainty in the radiation correction which is the largest source of measurement bias towards the top of the flight.

Long-term stability: 0.05 K. The signal of change over the satellite era is in the order of 0.1–0.2K/ decade requiring long-term stability to be an order of magnitude smaller to avoid ambiguity.

4.2.2. Water vapour

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 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.

For weather forecasting, boundary layer and lower tropospheric humidity measurements (or total column water vapour, which is dominated by the lower troposphere) are of primary interest. However, changes in water vapour in the UT/LS exert a greater radiative forcing than changes elsewhere (Solomon et al., 2010). Unfortunately standard radiosonde humidity sensors have very poor response at the low temperatures, pressures, and water vapour concentrations of the UT/LS (Wang et al., 2003). 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).
ii) Changes in convection. Convective transport of ice particles into the UT/LS can provide a path with bypasses the limitation imposed by the cold-point tropopause temperature.
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 is relatively easy to measure, the water vapour content of the upper atmosphere is difficult to measure accurately; the current generation of operationally-deployed balloon-borne instruments, and the satellite data record to date do 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 (see section B9).

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 these data in long-term trend analyses (Rosenlof et al., 2001). High precision measurements of water vapour profiles 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 Fluorescent Advanced Stratospheric Hygrometer for Balloon (FLASH-B) Lyman-alpha instrument, the Snow White chilled mirror hygrometer, or the Vaisala RS92 (Suortti et al., 2008) or RS-90 FN (Leiterer et al. 1997), may be used for reference measurements in their respective, valid altitude range. Other proven reference instruments may be introduced, with careful attention to data continuity concerns.

Many sites are currently developing the capability to observe and analyze data from ground-based GPS receivers. These data provide continuous high-quality estimates of column water vapour which can be used to partially validate the vertical humidity profile measurements; total precipitable water calculated from the radiosonde measured temperature and humidity profiles should compare well with that measurement by the GPS receiver.

Measurement range: 0.1 – 90000 ppm. The large range in values that needs to be covered by these measurements 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.

Vertical range: 0 to ~40 km.

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

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

Accuracy: 2% in mixing ratio throughout the profile. 1% for total column. This is more stringent than the 5% standard prescribed in WMO-No. 8.

Long-term stability: 1% (0.3%/decade) in mixing ratio and for the total column.

4.2.3. Pressure

Scientific justification: Accurate measurements of pressure from the surface to the upper stratosphere are necessary for relating measurements made in different vertical coordinates e.g. radiosonde (pressure) and lidar (geometric height) measurements, or model output which is often provided with geopotential height as the vertical coordinate. Uncertainty in calculated geopotential heights will result from uncertainties in temperature, pressure and water vapour measurements. The extent to which calculated geopotential heights/geometric heights agree with GPS derived altitudes can provide an indirect validation of the accuracy of the temperature, pressure and water vapour measurements. If pressure measurements drift in the presence of a steep vertical gradient in some target trace gas, this will alias into an apparent trend in that trace gas. It is therefore essential that pressure profile measurements maintain long-term stability.

Measurement range: 1 – 1100 hPa

Vertical range: 0 – 50 km

Vertical resolution: 0.1 hPa

Precision: 0.01 hPa

Accuracy: 0.1 hPa. This is more stringent than the 1 hPa to 2 hPa in the troposphere and 2% in the stratosphere requirements listed in WMO-No. 8.

Long-term stability: 0.1 hPa

4.3 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 Appendix 1 of GCOS-112. An approach to expanding site measurement capabilities to eventually cover as many of the specified variables as possible, whilst recognising that not all variables may be observed at all stations, is required.

5 GRUAN SITES

5.1 Site certification and assessment

GRUAN site selection is likely to happen through two possible routes, viz.:

1) Sites being approached by GRUAN and invited to become GRUAN stations. This would be true for most of the candidate sites listed in GCOS-121.
2) Sites being proposed externally e.g. through the National Weather Service of the host country.

In either case clear protocols for achieving site certification, and ongoing site assessment, need to be developed so that there is no ambiguity around site selection. The process must be transparent and applied equally to all candidate sites. This is especially important for sites proposed externally, or for sites seeking to have the GRUAN label, and where those sites may not be prepared to work towards achieving the standards set by GRUAN.

Once a site has been identified for possible inclusion in GRUAN, through either of the routes listed above, the following sequence of events is proposed as the protocol for achieving site certification:

1) Communication of GRUAN requirements to the candidate site by the Lead Centre. The Lead Centre will provide documentation outlining in detail the standards required for the operation of a GRUAN site. This should include the GCOS documents relevant to GRUAN, this manual, and a number of guides providing more detail around required standard operating procedures. In particular the minimum measurement requirements detailed in Section 5.2 and, equally importantly, the manner in which those measurements must be made will be the focus for the requirements of a candidate site. The GRUAN Lead Centre will also provide documentation around data submission protocols and the procedures that must be followed when data are submitted to the internal GRUAN archives (see Section 8.1).

2) Communication of the current status from the candidate site to the GRUAN Lead Centre. The candidate site should respond by providing the Lead Centre with documentation detailing:
    i) 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 ongoing operation of the site.
    ii) A description of the current measurement programmes at the site that will provide data to GRUAN and of the technical expertise available at the site to maintain these measurement programmes at the required standard.
    iii) A description of which databases these measurements have previously been submitted to and are currently being submitted to.
    iv) Detailed standard operating procedures for each of the measurement programmes that will be providing data to GRUAN, including a description of data storage policies.
    v) A description of how measurements to date at the site have been processed to come as close as possible to achieving GRUAN standards. 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 suitable for long-term trend detection. Since the historical database of measurements will be taken up into GRUAN, it is particularly important that the historical data can meet the stated GRUAN requirements for long-term homogeneity.

   - vi) A description of how systematic and random uncertainties in the measurements are currently being derived and how these measurement uncertainties are being reported.
   - vii) Any other meta-data describing key aspects of the measurement programmes to date.
   - viii) A list of the scientific experts employed at the site who would likely participate in the analyses of the data collected within GRUAN.

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 GRUAN Lead Centre will then write a short recommendation. This, together with the documentation from the candidate site, will then be submitted to the 'Site assessment, expansion and certification' Task Team (see Appendix A). This Task Team will make the final decision as to whether the candidate site will be certified as a GRUAN site. Important aspects on which this decision should be based would include:
   - i) Adherence to GRUAN protocols and requirements: More generally sites must have an operational philosophy of continually striving to improve measurement accuracy.
   - ii) Data quality (complete uncertainty analysis): Sites must be accountable for every measurement made. Specifically the calibration methods applied to each measurement, what sources of measurement uncertainty were accounted for, and what sources of measurement uncertainty were not accounted for.
   - iii) Operational standards: If necessary, sites must be prepared to forgo locally established operating procedures and adhere to the standard operating procedures imposed by the Lead Centre.
   - iv) Meta-data completeness: Sites must have procedures in place to ensure that detailed meta-data for all measurement systems are regularly submitted to the Lead Centre for inclusion in GRUAN data archives.
   - v) Traceability: Every measurement must be traceable to fundamental standards and calibrations through well documented routes.
   - vi) Management of change: Sites must be prepared to work under the guidelines outlined in Section 2.2.
   - vii) Commitment to long-term measurements: Since GRUAN is a climate monitoring network, sites must be prepared to commit to multi-decade measurement programmes of the essential climate variables. It is also essential that there be full host institution commitment to GRUAN-related activities at any particular site and that this commitment is not dependent on a single Principal Investigator.
   - viii) While a demonstrated track record in long-term monitoring would be advantageous, this is not essential. If a site with the instrumentation required to meet the GRUAN monitoring requirements exists, then it should not be overlooked simply because it has not been observing for decades.

   Since few, if any, planned GRUAN sites are likely to be immediately able to measure all required ECVs to the required levels of precision, accuracy and stability, achieving GRUAN status is likely to be an incremental process. Therefore, in developing the network and associated protocols, some degree of leeway in this regard is needed.

5) If the site is selected as a GRUAN site, a formal Memorandum of Understanding (MoU) between the GRUAN Lead Centre and the GRUAN site will be signed. This MoU would include a statement from the GRUAN site that the site agrees to operate under the protocols established within GRUAN, agrees to implement the standard operating procedures prescribed by GRUAN through a series of guides, and agrees to submit data to the GRUAN archives as detailed in the data submission protocols. In return the Lead Centre would agree to assist the

site with all operations related to GRUAN and to act as the liaison between the site and the international community of GRUAN data users.

In addition to the initial process of site certification, GRUAN sites should also undergo periodic assessments as being part of the network. This should include periodic site visits by members of the 'Site assessment, expansion and certification' Task Team and should include formal reports submitted to and archived by the GRUAN Lead Centre. It is important for external perceptions of GRUAN integrity that these audits are conducted by a GRUAN task team and not based on e.g. annual station reports. If conducted regularly, a series of such site assessment reports would clearly document the progress being made by sites towards achieving GRUAN standards. Should an existing GRUAN site show significantly reduced observational capability over more than a year, as evaluated by the criteria listed above, the task team should investigate the circumstances at that site, and, if needed, suspend its membership in the network.

5.2 Site selection

The process by which new sites will be selected/accepted into GRUAN is currently being addressed by the 'Site assessment, expansion and certification' Task Team and has not yet been finalized. This section defines the more general principles under which GRUAN site selection should be considered. Foremost is the measurement and operational capabilities of any putative GRUAN site i.e. the availability of necessary instruments, infrastructural support, and ability to adhere to the site requirements listed in Section 5.1. This may depend in part on the membership of that site in other measurement networks (e.g. NDACC, GAW, BSRN). In such cases, this should be seen as an advantage since it reduces the start up costs for establishing the GRUAN site and it quantitatively links the GRUAN measurements to the measurements being made in those other networks.

GCOS-121 suggested that an interim starting point for radiosonde observations at GRUAN sites should be made at tiered levels, ideally consisting of:

- 1 weekly production radiosonde measurement of temperature, pressure and humidity with the best technology currently available. High quality surface measurements of these same variables are also required to provide pre-launch calibration of the instruments onboard the sonde. While weekly sampling under-estimates monthly standard deviations in temperature by up to 90% smaller and 100% larger than true values, differences between detectable trends for weekly sampling compared to 12 hourly sampling are smaller (Seidel and Free, 2006).
- 1 monthly radiosonde capable of measuring water vapour in the UT/LS and all other priority 1 variables (see Section 4.1) to the best level possible with current technology, launched together with the weekly radiosonde. Given that high frequency natural variability in the lower stratosphere is small, sites should launch these radiosondes in those conditions most likely to lead to a successful launch and measurement throughout the column, but particularly in the upper reaches of the ascent. Typically this may be under cloud free conditions at night but site staff will be best placed to make this call.
- Regular 00 and 12 LST (as a preference over UTC) launches of a production radiosonde with the best technology currently available. Local operational constraints may lead to other launch schedules at some stations, which should not preclude these stations from being designated as GRUAN stations. Where feasible, occasional soundings at both 00/12 LST and UTC could be used to establish a temperature difference climatology, including uncertainties, which could thereafter be used to relate measurements made at one standard time to measurements made at another. It should be noted that 00/12 UTC observations are no longer as important for NWP since 4D data assimilation is now more common.

- 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).
- Periodic intercomparisons of a large range of sonde types.

This interim starting point for required GRUAN site capabilities will be expanded as more quantitatively defensible assessments (e.g. following research by GATNDOR) become available. Only the first two criteria were considered in GCOS-121 to be absolute requirements. However, it should be noted that weekly measurements made using the best radiosonde technology currently available may be prohibitively expensive. Since the focus is not only on making very precise and accurate measurements, a compromise would be to make 1 weekly radiosonde measurement of temperature, pressure and humidity using the usual radiosonde used at the station but to fly this together with a second sonde, either from another manufacturer (to test network homogeneity) or from the same manufacturer (to test repeatability). Both approaches would assist in validating measurement uncertainty which is equally important for these measurements (see Section 6.2)..

Geographical coverage of GRUAN sites is also an important consideration. GATNDOR have been tasked to assess the scientific desirability of station locations from a variety of perspectives. Because GRUAN will act as the reference standard for the current 167 GUAN sites located world-wide performing primarily radiosonde observations, it is important that each GUAN site is, eventually, located sufficiently close to a GRUAN site to allow meaningful intercomparisons. 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..

It is not necessary that GRUAN provides globally complete and spatially homogeneous coverage - rather GRUAN should provide a reference anchor for other ground- and satellite-based networks which would then provide the required global coverage. However, it would be advantageous if GRUAN could sample all major climatic regimes and environment types to ensure that different temperature and radiation environments are reliably calibrated. Expansion of the network should concentrate on climatic zones and regions that are under-sampled in the initial network configuration. Geographical coverage of GRUAN sites should also be tailored to meet the specific needs of end-users e.g. the satellite-based measurement community is likely to want validation data in key regions of the atmosphere.

Candidate GRUAN sites will have to be able to demonstrate reasonable expectations of funding to maintain operations over many decades. The GRUAN executive should have in place procedures for supporting long-term funding applications to local funding agencies for sites seeking to join GRUAN. At present most national funding agencies are challenged by requirements for funding over multi-decade timescales. This may therefore requite higher level (GCOS) education of national funding agencies for maintenance of the global climate observing system. Having sites transiently joining and then leaving the GRUAN network could compromise the goal of ensuring data homogeneity across the network e.g. if trends in ECVs differ at two different stations which measured the ECV over different time periods it is not clear whether the differences arise from the geographical separation or from different time periods being sampled.

It may be the case that while a single station might not be able to provide the full range of ECV measurements required by GRUAN, a group of two or more stations, located sufficiently close together might have the combined capability of providing the full range of measurements. Such a collection of stations may then act as a single GRUAN site. A key question here is what is meant by 'sufficiently close'? This is a research question currently being addressed by the GATNDOR team.

Sites may be selected and invited to join GRUAN, subject to the requirements listed above. However, it is also possible that some countries may propose the inclusion of specific sites in GRUAN such as during the 15th session of AOPC, when Japan Meteorological Agency (JMA) offered to contribute the Tateno site. A formal mechanism therefore needs to be established to deal with such offers should they arise. This needs to balance the needs of all stakeholders but recognise that at this stage a willingness to participate is highly desirable.

6 INSTRUMENTATION

6.1 Instrument selection

The choice of what instruments should be deployed within GRUAN will not be a one-off decision. Periodic review of instrumentation likely to be of use within GRUAN needs to be undertaken since instrument technology is constantly evolving. It also needs to 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 introducing a discontinuity in the measurement time series. The emphasis is therefore not on prescribing an instrument, but rather on prescribing the capabilities of an instrument and allowing individual sites to select an instrument that achieve those capabilities. That said, the fewer the number of instrument types deployed within GRUAN, the more likely network homogeneity will be achieved.

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

- 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 sufficiently long for the expected multi-decade deployment within GRUAN?
- 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 accuracy, precision and stability standards (see Section 4.1) required by GRUAN?
- Information content: Are the temporal and spatial resolution, dynamic range, and other characteristics of the measurements made by the instrument consistent with GRUAN requirements?
- Manufacturer support: Is the manufacturer committed to a process of improving the performance of the instrument based on findings made by the GRUAN user community? 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? A case in point regarding this last question – Immler et al. (2010) were unable to adequately assess the radiation correction made in three different radiosondes because the correction algorithm applied by the radiosonde software would not be disclosed by the manufacturer. For a consistent uncertainty analysis it is imperative that the algorithms used for corrections within the data processing software are made publicly available by the instrument manufacturers. Unwillingness for the manufacturer to do so, should count against the selection of that instrument for use within GRUAN.
- 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 errors in measurements. A project within GATNDOR has been tasked with quantifying the value of redundant measurements and assessing 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. A case study underway within GATNDOR is using vertical profile measurements of temperature and water vapour at the GRUAN sites at Beltsville, Cabauw, Lindenberg, Potenza (all ARM sites) to quantify the error reduction resulting from increasing redundancy of measurements. This requires an assessment of the uncertainty of the temperature and water vapour vertical profiles retrieved using each of the considered techniques and then the investigation of possible sensors' synergies to reduce the uncertainty. The investigation will be carried out focusing on the most common instruments at the considered GRUAN sites: for temperature, radiosonde soundings and microwave profilers; for moisture, radiosonde soundings, Raman lidars, microwave profilers, and GPS receivers. 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
- Quality control/assurance with a focus on instrument performance in different meteorological conditions.

As for much of the other research underway to support the operational framework for GRUAN, this is work in progress and the true value of having multiple measurements of the same climate variables at GRUAN stations will become clear in time.

One important factor for GRUAN is that independent 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 (see Section 3.1.3).

6.3 Surface measurements

While GRUAN is, by definition, an upper-air network, the availability of coincident surface measurements is likely to be advantageous to GRUAN for a number of purposes, including:

- Providing ground-truthing for vertical profile measurements. For example, comparisons between ozonesonde measurements of ozone at the surface against a high precision standard can be used to identify uncertainties in the ozonesonde measurement.
- Some remote sensing instruments that derive vertical profile data from e.g. optimal estimation techniques can benefit from having a surface measurement to constrain 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 the Total Carbon Column Observing Network (TCCON).

While there are no formal requirements for GRUAN stations to include surface measurements, the guideline 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

A discussed in Section 4.2.1, radiosondes will remain the primary workhorse within GUAN for the measurement of vertical profiles of temperature, pressure and humidity. The fact that these instruments are not recovered has important implications for GRUAN operations, viz.:

- The instruments must be low cost, and because they are low cost, the sensors on sondes are unlikely to be the best commercially available. Therefore, certain compromises in system measurement accuracy have to be accepted by users, taking into account that radiosonde manufacturers are producing systems that need to operate over an extremely wide range of meteorological conditions.
- 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.

Because GRUAN will make only weekly high quality measurements of temperature, pressure and humidity (see Section 5.2) rather than the 12 hourly profile measurements required at GUAN stations, more expensive (and hopefully more accurate) sensors can be used.

6.4.2. Remote sensing instruments

Material to come in here from Task Teams 2 and 5

6.5 Instrument co-location

As discussed in Section 5.2, 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. Therefore, appropriate collocation requirements for variables and instrumentation should be established to ensure the representativeness of measurements. These considerations should be site and parameter-specific.

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, will be 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 2.2) and avoid introducing discontinuities into the time series, it is equally important that calibration procedures do not compromise the goal of achieving homogeneity across the GRUAN network as a whole so that a measurement of some parameter at one site is directly comparable to a measurement of the same parameter at a different site. A guiding principal that will achieve this goal is that the same when two identical instruments are deployed at two different sites, they should also use the same calibration procedures, preferably tied to the same absolute standards, and should also employ identical data processing algorithms. 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 Sections 2.2.3 and 8.1), the same approach cannot be used for calibration procedures. To this end achieving inter-site homogeneity will be improved by developing travelling calibration standards which can be taken to different GRUAN stations to be used in on-site calibration or intercomparisons. A current example of this would be Dobson Spectrophotometer #83 which is used in the NDACC and WOUDC networks to achieve homogeneity across the global Dobson network (see Sections 1.3.1 and 1.3.3). Such travelling standards for ground-checks for radiosondes (temperature and humidity sensor checks) would be particularly valuable.

Traceability to recognized measurement standards (e.g. SI standards) that can be reproduced globally and over long periods of time will be the key component enabling 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 back 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 common fundamental standards are available across the GRUAN network this will support the goal of achieving coherence across the network.

GRUAN stations should maintain a “GRUAN site working standard” for each basis 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 needs to be put in place to address the compatibility of those systems that may not be traceable to SI standards with the rest of the network.

Use of traceable calibration standards 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 facilitate the network to incorporating new scientific insights and new technological developments, while maintaining the integrity of the long-term climate record. To achieve traceability, meta-data on all aspects relating to a measurement and its associated uncertainty will need to be collected. Each station will need to maintain accurate meta-data records and provide these to the GRUAN archives. Copies of calibration certificates should be submitted to the GRUAN meta-database.

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 meta-data, 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 those newly-discovered issues.

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 likely be site specific. A standard recommendation for the use of redundant instrumentation and remote sensing instrumentation should be developed 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 maintenance

GRUAN sites are equipped with sophisticated, state-of-the-art instrumentation and should comply with strict requirements of station 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 should 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 should not be conducted more frequently than is necessary. Maintenance schedules should be developed for all instruments. All maintenance actions on instruments need to be documented as part of the meta-data associated with the measurements made by that instrument.

7 METHODS OF OBSERVATION

7.1 Measurement scheduling

The development of measurement scheduling protocols is undertaken by the 'Measurement schedules and instrument-type requirements' Task Team (see Appendix A). The highest priority is that measurement schedules are established to achieve the four primary goals of GRUAN (see Section 1.1). Specifically it should be noted that measurement scheduling should be designed not only for the purposes of long-term trend detection but to fulfil all goals of GRUAN. The required measurement frequency will differ depending on the parameter being measured. Measurements need to be sufficiently frequent to capture important scales of temporal variability, both for trend analysis and for process understanding. In cases where oversampling would allow averaging of measurements to reduce the net random error, and where this is technically feasible, measurement schedules should be set so as to achieve this. Where measurement redundancy (see Section 6.1) allows measurements of the same variable to be made with more than one instrument, sampling intervals and data averaging schemes need to be applied similarly to both instruments to allow the resultant values to be comparable.'

Measurement frequency may also vary regionally and seasonally. In places and seasons where the parameter is being measured is more variable, measurements should be made more frequently so that the effects of that variability can be accounted for in trend analyses. The degree of autocorrelation in the measured time series is also likely to affect measurement frequency requirements. Measurement scheduling requirements should be informed by quantitative studies that are regionally and seasonally specific and that perhaps sample model output to understand how measurement scheduling may affect the ability to detect long-term trends. It may be that trend detection is limited by natural variability rather than by the precision of the measurement, in which case more resources should be invested in increasing measurement frequency rather than increasing measurement precision. In some cases this may reduce to a cost-benefit analysis where the cost to detect a putative trend of X%/decade (perhaps based on projections from climate models or chemistry-climate models) over N years is minimized. A cheaper instrument making a less precise but more frequent measurement might be selected over a more expensive instrument making a more precise less frequent measurement, since the greater frequency leads to detection of the expected trend either in fewer years or at a lower cost.

Measurement frequency should also be set to permit a statistical separation of the different drivers of changes in the observed variable. Statistical studies should inform the process of establishing measurement schedules. Where possible, and where is does not compromise achieving the highest priority, measurement schedules should be adapted to meet the needs of other end-users e.g. 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. If, however, the variable being measured showed a strong diurnal cycle, shifting the measurement time away from the norm would introduce an anomaly which might then later compromise the interpretation of those measurements. Clear protocols therefore need to be established to ensure that meeting the needs of secondary users of GRUAN products does not compromise the quality of the data provided to the primary users.

For some measurements, scheduling with respect to UTC or Local Solar Time (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), while for ensuring coincidence with GUAN stations, or to be used as input to initializing NWP models, having all measurements made at the same UTC might be more appropriate. As detailed in Section 5.2), the current intention is that radiosonde flights will be made at the same LST within GRUAN, however, this decision has not been finalized.

7.2 Operation and maintenance, quality standards

The more traditional approach of setting a quality standard and then assessing whether each measurement meets that standard is less applicable in GRUAN where the emphasis is more on describing, quantifying and verifying measurement uncertainty estimates and then communicating the quality of the measurement through that uncertainty estimate. That said, standards of operation and maintenance for each instrument used in GRUAN should be developed to ensure that minimum quality standards are achieved. This will be necessary to minimize sources of error 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 and maintenance protocols should be such that collection of detailed meta-data is mandatory as these meta-data will be vital to establishing measurement uncertainties.

Because GRUAN is not being established as a network of completely new stations, and because many of the initial stations within GRUAN have been in operation in some cases for decades, sites collecting data from different instruments will almost certainly currently use different averaging and data processing algorithms, different instrument pre-checks, different instrument post data checks, etc.. GRUAN will not consist of a set of identical sites supported by a single funding agency. A process for achieving convergence on agreed on operations and maintenance procedures that will be applied across the network therefore needs to be developed. Furthermore, many of the initial sites report to numerous networks and their governance and stated aims differ substantially. It is therefore essential to have in place protocols and agreements, such as a Manual of Operations, including common quality assurance procedures that allow the required flexibility, whilst maintaining the fundamental quality of the observations necessary to meet GRUAN aims.

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 2.

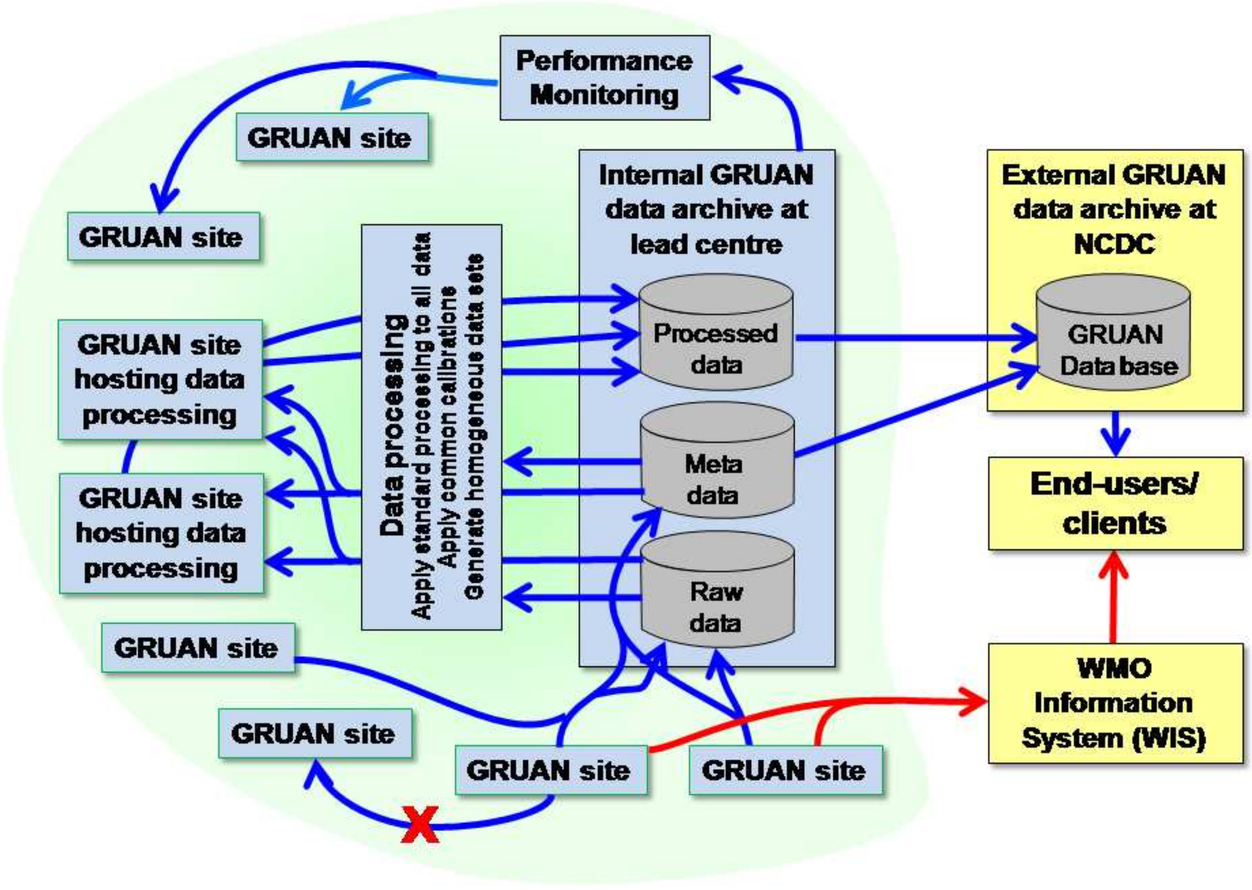

Figure 2: A schematic representation of the flow of data in GRUAN. Blue arrows show the standard flow of data. The red arrows show the flow of near-real time data. Data provided to end-users via red routes are not ‘GRUAN data’. Different data exchange protocols should operate for exchange of data within GRUAN (shaded green region) and from the GRUAN external data archive to end-users.

Raw measurement data and meta-data, referred to as Level 1 (L1) data, are ingested from all GRUAN sites into the internal GRUAN data archive hosted at the Lead Centre (see Section 8.4). L1 data will typically be the ‘rawest’ form of data available e.g. measured voltages before any processing has been applied. Direct exchange of L1 data between sites should be discouraged since this would circumvent the data versioning protocols, network wide application of calibration techniques, and other pre-processing of raw data that would be implemented at the Lead Centre or at a centralized GRUAN data processing site (see below). The only likely pre-processing of L1 data at the measurement site would be the conversion to a common format (e.g. NetCDF). It is also expected that L1 data would be archived at the measurement sites.

Where GRUAN sites have agreed to the near-real time release of their data, these data will be made immediately available via the WIS. This will almost certainly require some local site-based processing of the L1 data to create data suitable for submission to the WIS.

Processing of the L1 data held in the GRUAN internal data archive to produce a GRUAN data product, referred to as Level 2 (L2) data, will occur either at the Lead Centre or at a GRUAN station 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 L2 data, including its meta-data and documentation, are provided to the user community through the external GRUAN data archive 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

GRUAN data should be made freely and publicly available. Specifically GRUAN data dissemination and use should comply with WMO Resolution 40 (Cg-XII). However, because some GRUAN stations are likely to be providing data to other networks which may have policies in place to protect the rights of the data providers to their own data, some flexibility may need to be shown regarding timeframes for making the data publicly available. GRUAN meta-data should include all information related to acknowledgements and/or co-authorship on publications making use of the data. Two different 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 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 8.5). A different policy should be implemented to control the dissemination of GRUAN data at this level.

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, common measurement technology, generally well understood, and 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, would recommend contact to site scientist for correct interpretation of data, and would require considerable efforts to maintain continuous measurements and high quality of data.

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

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.

The primary goals of GRUAN (see Section 1.1) are not consistent with near real-time dissemination of measurements made at GRUAN sites. Generating high precision, high quality measurements with well characterized uncertainties takes a significant investment of time and effort. In GRUAN the balance is tipped strongly in favour of the provision of high quality measurements rather than the provision of near real-time measurements. However, it is recognized that GRUAN measurements are likely to be very useful to a number of users requiring data in near real-time 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 real-time data to end-users via the WIS. These, however, should not be termed ‘GRUAN’ data since they would not have been subjected to the stringent QA/QC procedures that are core to GRUAN’s operation. Rather they are what might be termed ‘pre-GRUAN’ data. In this context, greater emphasis should be placed on the submission of real-time data required for real-time applications such as NWP model initialization e.g. those listed in Annex 1 of Resolution 40 (Cg-XII). These are also more likely to be 'standard data' as described above. The WIS requirements, e.g. on meta-data, and the possibility to transmit near-real time data via the Global Telecommunication System (GTS) should be explored. Where sites do not currently have the infrastructure or expertise in making such submissions, WMO should be approached for assistance in the form of hardware and/or training. There may be advantages to submitting data in near real-time 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 meta-data. Near real-time release of standard GRUAN data will also facilitate the quality control link between GRUAN and GUAN.

8.3 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 meta-data 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 meta-data in each data file, should be an open format/standard that requires no licensing, 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.

8.4 Data submission

If sites elect to submit near real-time data to end-users, this should be done directly through the WIS. 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 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 this time. A facility for imposing time limits on making the data available to the end-user community for different stations should be implemented as this does not contravene WMO Resolution 40 (Cg-XII). In this way stations 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 meta-data 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 (see Section 1.2).

8.5 Data dissemination

Dissemination of GRUAN data products to end-users/customers should occur through an official GRUAN data Centre 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 stations.

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 GRAUN data should 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.

Such a facility might exist independently of the GRUAN NCDC archives.

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. The difficulty arises in that data submitted through a non-GRUAN network may be subject to different data processing, different QA/QC procedures, and different calibrations. This would result in two different versions of ostensibly the same data being publicly available. Such a situation should be avoided since it would undermine the confidence that users would have in GRUAN data products.

Users of GRUAN data need to know the version of any dataset they are using and whether newer versions might be available. Having the names of data files include the data version number would be helpful in this regard. A facility to periodically check for updates of GRUAN data files found on a client computer with the database at NCDC would be very advantageous.

8.6 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 should be 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.

8.7 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 meta-data submitted etc.. Existing algorithms, potentially supplemented by future algorithms to be developed, will need to be used operationally to identify systematic errors, anomalies or instrumental issues. Results of such tests should 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, communicating quality assurance results to GRUAN site operators and engineers will facilitate improved instrument performance and thereby minimize the amount of unacceptable data collected.

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. 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 specializing in processing of that particular data set, should be tasked with data version control and ensuring that the necessary meta-data 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.

10 QUALITY ASSURANCE

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. Where total measurement uncertainties lie 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 in 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.

Visual inspection of all data by science/instrument experts will be required for all instruments to minimize issues that slip through automated routines. The Lead Centre should coordinate this effort, which should be distributed across different GRUAN sites. As outlined in Section 3.1.3, 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.

The use of GRUAN data in data assimilation 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 within the known internal variability of the system.

ACRONYMS

ARM: Atmospheric Radiation Measurement programme
ACRF: ARM Program Climate Research Facility
AOD: Aerosol Optical Depth
AOPC: Atmospheric Observation Panel for Climate
CBS: WMO Commission for Basic Systems
CIMO: WMO Commision for Instruments and Methods of Observation
GATNDOR: GRUAN Analysis Team for Network Design and Operations Research
GCOS: Global Climate Observing System
GHG: Well-mixed greenhouse gas (CO_2, CH_4, N_2O, CFCs, HFCs, PFCs, SF_6, etc.)
GOS: Global Observing System
GRUAN: GCOS reference upper air network
GTS: Global Telecommunication System
GUAN: GCOS upper air network
ICM: Implementation - Coordination Meeting (GRUAN)
ISCCP: International Satellite Cloud Climatology Project
NCDC: National Climate Data Centre
NWP: Numerical Weather Prediction
PDF: Probability Distribution Function
RMS: Root Mean Square
UT/LS: Upper troposphere/lower stratosphere
WIGOS: WMO Integrated Global Observing System
WMO CBS: World Meteorological Organisation Commission for Basic Systems
WWW: World Weather Watch

Appendix A – Task Teams

Task Team 1
Radiosondes

Chairs
Franz Immler
Masatomo Fujiwara

Members
Joe Facundo
Sasha Kats
Bruce Sumner
Rolf Philipona
Larry Miloshevich
Tim Oakley
Frank Schmidlin

Task Team 2
GPS Precipitable Water

Chairs
June Wang
Kalev Rannat

Members
Seth Gutman
John Braun
Galina Dick
Yoshinori Shoji
Siebren De Haan

Task Team 3
Measurement schedules and instrument-type requirements

Chairs
Tom Gardiner
Dave Whiteman
Howard Diamond

Members
Besty Weatherhead
Reinout Boers

Task Team 4
Site assessment, expansion and certification

Chairs
Russ Vose
Steve Williams

Members
Dian Seidel
Mike Kurylo
Anna Kuhn
Geir Braathen

Task Team 5
Ancillary measurements

Chairs
Tony Reale
Thierry Leblanc

Members
Seth Gutman
John Braun
Galina Dick
Yoshinori Shoji
Siebren De Haan

Task Team 6
GRUAN sites

Chairs
Belay Demoz
Dale Hurst

Members
Martin de Graaf
Paul Johnston
Rigel Kivi
Gelsomina Pappalardo
Rolf Philipona
Hakaru Mizuno
Holger Vömel
Russ Vose
Jimmy Voyles

Appendix B – Expanded details on Essential Climate Variables

B.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.

B.2. Wind direction (priority 2)

No supplementary comments yet.

B.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.

B.4. Methane (priority 2)

No supplementary comments yet.

B.5. Net radiation (priority 2)

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

B.6. Incoming short-wave 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 sub is not obscured by cloud, reflections off clouds can enhance surface short-wave radiation significantly. The prescribed precision and accuracy values of 3 and 5 W/m^2 respectively, match the requirements for the BSRN network.

B.7. Outgoing short-wave radiation (priority 2)

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

B.8. Incoming long-wave radiation (priority 2)

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

B.9. Outgoing long-wave radiation (priority 2)

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

B.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.

B.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 scientifically useful if the aerosol optical depth is large, when the aerosol optical depth is small, measurements of other aerosol parameters become less valuable.

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

Size-fractionated measurements are required.

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

Size-fractionated measurements are required.

B.14. Aerosol light scattering (priority 2)

Size-fractionated and spectral measurements are required.

B.15. Aerosol light absorption (priority 2)

Size-fractionated and spectral measurements are required.

B.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).

B.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 20 m/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). It should be noted that the trends reported in Chernykh have been questioned by Seidel and Durre (2003).

B.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).

B.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.

B.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.

B.21. Cloud top pressure (priority 3)

No supplementary comments yet.

B.22. Cloud top temperature (priority 3)

No supplementary comments yet.

B.23. Cloud particle size (priority 4)

No supplementary comments yet.

B.24. Cloud optical depth (priority 4)

No supplementary comments yet.

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

No supplementary comments yet.

REFERENCES

Alexandersson H. and A. Moberg, Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends, *Int. J. Climatol.*, 17, 25-34, 1997.

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.

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

Bodeker G.E., J.C. Scott, K. Kreher, and R.L. McKenzie, Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978-1998, *J. Geophys. Res.*, 106 (D19), 23029-23042, 2001.

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.

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.

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

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.

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. *Atmospheric Measurement Techniques*, 3, 1217–1231, 2010.

Khaliq M.N. and T.B.M.J. Ouarda, On the critical values of the standard normal homogeneity test (SNHT), *Int. J. Climatol.*, 27, 681-687, 2007.

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.

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.

Rodgers C.D. and B.J. Connor, Intercomparison of remote sounding instruments. *J. Geophys. Res.*, 108 (D3), 4116, doi:10.1029/2002JD002299, 2003.

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.

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.

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 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.

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.

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.

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.

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.

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.