GAIA-CLIM Report / Deliverable D3.4

Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring:

D3.4 Measurement mismatch studies and their impact on data comparisons

A Horizon 2020 project - Grant agreement: 640276

Date: 28 February 2017

Lead Beneficiary: University of Bergamo, Italy

Nature: R

Status: Final

Dissemination level: PU

European Commission

BIRA-IASB

NPL National Physical Laboratory

Consiglio Nazionale delle Ricerche

FINNISH METEOROLOGICAL INSTITUTE

Work Package	WP3 – Comparison error budget closure – Quantifying metrology related uncertainties of data comparisons
Deliverable	D3.4
Title	Measurement mismatch studies and their impact on data comparisons
Nature	R
Dissemination	PU
Lead Beneficiary	University of Bergamo, Italy
Date	28 February 2017
Status	Final
Editors	Alessandro Fassò (UNIBG), Tijl Verhoelst (BIRA-IASB) and Jean-Christopher Lambert (BIRA-IASB)
Contributors	Alessandro Fassò, Ilia Negri, Francesco Finazzi (UNIBG), Tijl Verhoelst, Jean-Christopher Lambert (BIRA-IASB), Timo H. Virtanen, Rigel Kivi (FMI), Tom Gardiner (NPL), Fabio Madonna, Lucia Mona, Nikolaus Papagiannopuolos (CNR) and Peter Thorne (NUIM)
Reviewers	Peter Thorne (NUIM), Corinne Voces (NUIM), Richard Davy (NERSC)
Contacts	Alessandro.Fasso at unibg.it, Tijl.Verhoelst at aeronomie.be
URL	http://www.gaia-clim.eu/

This document has been produced in the context of the GAIA-CLIM project. The research leading to these results has received funding from the European Union's Horizon 2020 programme under grant agreement n° 640276.

Table of content

Executive summary

GAIA-CLIM, a H2020 project funded by the European Commission, aims to support Europe's Earth Observation programme Copernicus by assessing and improving the fitness-for-purpose of sub-orbital (ground- and balloon-based) reference measurements in the validation of observational data sets from satellites. In particular, the project aims at improved traceability and uncertainty characterization, of the individual sub-orbital measurement systems and of the comparison with satellite data.

A key issue in the geophysical validation of satellite data sets with respect to sub-orbital reference measurements is the interpretation of their differences in terms of known, quantified, uncertainties. This aspect includes not only the measurement uncertainties associated with the individual measurements, but also the additional uncertainties that appear when comparing different perceptions of the inhomogeneous and variable atmosphere, that is, when comparing data sets characterized with different sampling and smoothing properties, both in space and time. Those "comparison uncertainties" are the main topic of investigation for GAIA-CLIM Work Package 3.

Deliverable 3.2 (D3.2) describes the concept of co-location mismatch, and how the resulting uncertainties can be decomposed and quantified through a careful metrological analysis of the measurements and their comparison. It also provides an overview of available methods to quantify these uncertainties so that they can be taken into account when interpreting the results of a data comparison, for example like in data validation and in data assimilation.

The present GAIA-CLIM deliverable 3.4 (D3.4) elaborates further on the quantification of co-location mismatch detailed in D3.2 and applies statistically-based and model-based approaches to several case studies carried out on the Essential Climate Variables (ECVs) targeted within GAIA-CLIM:

- Sections 3 and 4 address two meteorological profile variables: temperature and humidity. In particular, Section 3 discuss IASI, RAOB and GRUAN comparisons which are analyzed using advanced statistical methods developed by UNIBG and CNR. Moreover, Section 4 summarises the sonde intercomparisons as performed by NPL.
- Section 5 addresses ozone. Here BIRA-IASB presents the OSSSMOSE results on ozone profile comparisons, considering ozonesondes, micro-wave radiometers and LIDAR.
- Sections 6 and 7 address aerosols. In particular, in Section 6 it is reported the comparisons of AERONET vs AATSR, which is developed by FMI. Eventually, in Section 7, the comparison of EARLINET vs CALIPSO is developed by UNIBG and CNR.

1 Introduction

High-quality observational datasets from satellites constitute a key component of the European Commission’s Copernicus programme, which aims at providing users (mainly policy makers and public authorities) with reliable and up-to-date information related to environmental and security issues. The climate change and atmosphere monitoring services in particular rely heavily on observations from current and future satellite instruments measuring both key meteorological variables such as temperature and humidity, and atmospheric composition, including greenhouse gases and health-endangering pollutants.

For these services to be reliable and effective, it is a prerequisite that the underlying datasets are fit-for-purpose, i.e. that their quality is assured and that they meet user requirements. Quality assurance in the context of satellite remote sensing has been defined in the context of the CEOS- and GEO-endorsed Quality Assurance for Earth Observation (QA4EO) framework, as the need *for fully traceable Quality Indicators*. In practice, the extent to which the satellite measurements agree with ground-based reference measurements is an essential such quality indicator. Clearly, this agreement needs to be assessed in the context of the reported uncertainties, both those on the satellite and on the reference measurements. As such, traceability of the data production and of the associated uncertainties is another crucial quality indicator. The H2020 project GAIA-CLIM aims specifically at improving the traceability and uncertainty characterization of the sub-orbital reference measurements used to assess the quality of the satellite data sets. Moreover, it also addresses the uncertainty budget of these crucial satellite-to-reference comparisons from which several quality indicators are derived, but which require careful consideration of the additional errors due to co-location mismatch, i.e. the unavoidable differences in measurement times, locations and spatiotemporal smoothing.

1.1 Context and aims of this document

Two key scientific objectives of GAIA-CLIM concern the uncertainties on sub-orbital reference measurements. In addition, uncertainties related to the intercomparison of measurements from different instruments and obtained at different locations and measurement times. In particular

- objective S3, mainly targeted within WP2, is to provide *reference quality measurement uncertainties that are traceable to recognized measurement standards*, and
- objective S4, mainly targeted within WP3, is to understand and quantify *the metrology of a data comparison, including the additional uncertainties that arise from spatiotemporal mismatches between both observing systems*.

As such, these objectives target the uncertainty budget of a comparison between a satellite measurement and a sub-orbital reference measurement, which, in the ideal case of perfect co-location, can be represented mathematically as:

$$|m_1 - m_2| \le k\sqrt{u_1^2 + u_2^2} \tag{1.1}$$

where the left-hand side represents the observed difference between both measurements, $\boldsymbol{k}$ is a coverage factor, and $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$ represent the measurement uncertainties (e.g. Immler et al., 2010). However, in case of spatiotemporal mismatch, i.e. non perfect co-location, an additional uncertainty term, $\boldsymbol{\sigma^2}$, must be included:

$$|m_1 - m_2| \le k\sqrt{\sigma^2 + u_1^2 + u_2^2} \qquad (1.2)$$

The present document, Deliverable 3.4, represents key output from WP3, and targets mainly objective S4, i.e. $\boldsymbol{\sigma}^2$ in Eq. (1.2). In particular, it reports on several case studies, which cover temperature, humidity, ozone and aerosols.

Figure 1.1 visualizes the relation between the current deliverable, the different tasks within WP3, and the other future deliverables.

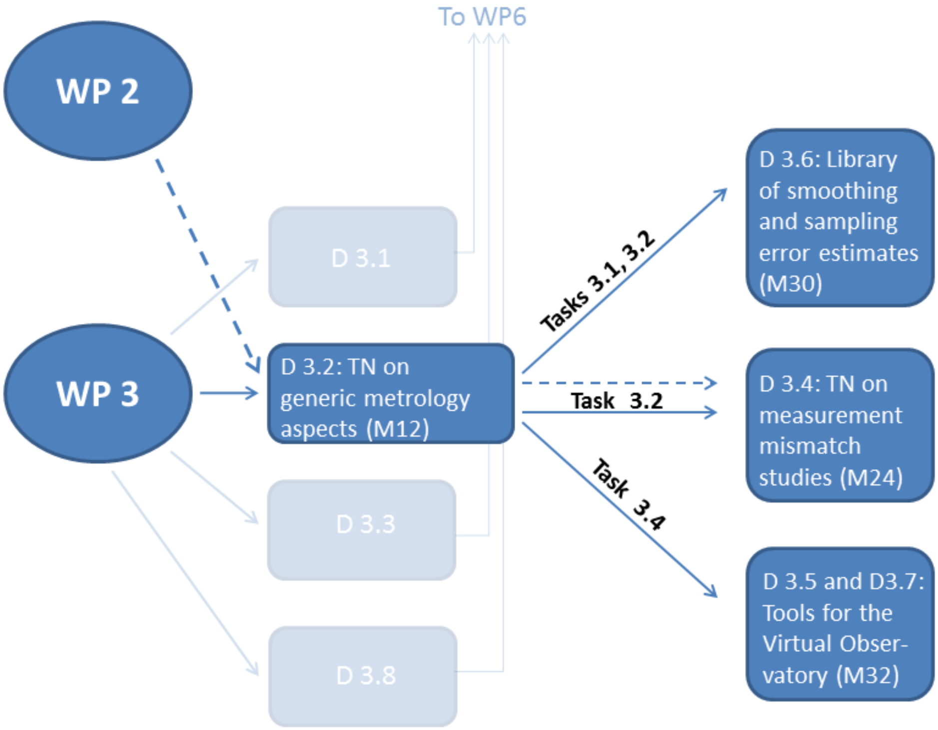


Figure 1.1: Inter-linkages between D3.2 and other deliverables from WP3. Because WP2 also deals with some metrology aspects and provides the measurement uncertainties required for uncertainty budget closure aimed for in Task 3.2 and deliverable D3.4, it is included in this graph.

Since the ultimate aim is to close the uncertainty budget of a comparison between satellite and sub-orbital reference measurements (cfr. Eq. 1.2), the availability of reliable and traceable measurement uncertainties (S3, WP2) is essential to the work performed within WP3. Consequently, the present document will also briefly cover those metrology aspects that are primarily dealt with by WP2, in this way further clarifying which metrological aspects are (or should be) part of the reported measurement uncertainties, and which are to be taken into account in addition to the measurement uncertainties when performing intercomparisons.

Note that the current document deals with Level-2 data only, i.e. (retrieved) columns and profiles, as this is the focus of WP3 within GAIA-CLIM. For Level-1 radiance data, similar considerations may apply, but this is beyond the scope of the work planned here. Level-3 data, i.e. gridded averages of Level-2 data, add another layer of metrological issues, in particular regarding representativeness of the averages due to the particular sampling patterns of the sounders and networks, but also this topic is largely beyond the scope of the current document.

1.2 Metrology of a data comparison and associated errors

In most data comparison endeavors, such as ground-based validation exercises, a compromise must be made between on the one hand abundance of comparisons pairs, and on the other hand additional comparison errors, not related to the actual measurements uncertainties but due to non-perfect co-location in space and time. This non-perfect co-location is a consequence of both a difference in spatio-temporal sampling, i.e. a satellite pixel centre never coincides exactly with the ground station location, and a difference in resolution, i.e. in the way each instrument has a smoothed perception of the inhomogeneous and variable atmosphere. This is visualized conceptually in Figure 1.2.

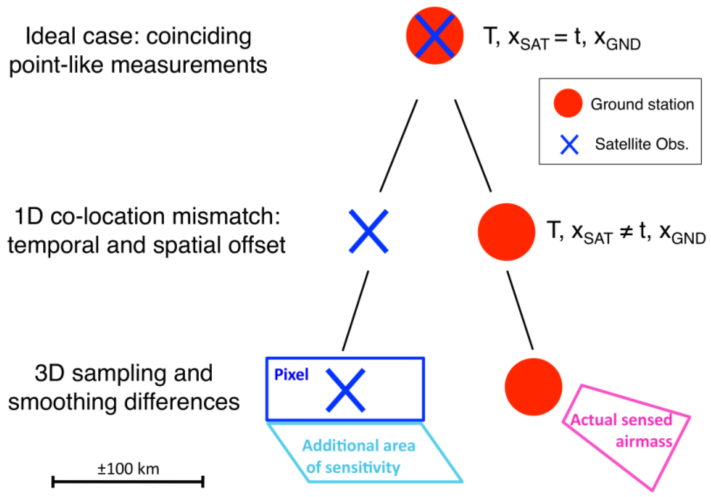


Figure 1.2: Conceptual visualization of the metrology of a satellite-to-ground measurement comparison. Ideally, both measurements are point-like in space and time, and coincide perfectly. In practice, the sampling pattern of the satellite sounder and the fixed locations of the ground network induce sampling difference errors. Furthermore, differences in resolution, or, more broadly, in area of actual measurement sensitivity, induce additional smoothing difference errors. Figure reproduced from Verhoelst et al. (2015).

A formal representation of the uncertainty budget of a comparison was already explored by Rodgers (1990, 2000) and by Rodgers and Connor (2003), and further elaborated by von Clarmann (2006). Lambert et al. (2012) includes the multi-dimensional perspective, dealing with horizontal smoothing errors and errors due to less than perfect co-incidence. While the above-mentioned papers deal with uncertainties in terms of covariance matrices, and thus assume the errors to have a Normal distribution, it is instructive to focus first on the decomposition of individual differences (i.e. per comparison pair) in terms of the different error sources, before deriving an uncertainty budget. A pair of co-located measurements, $\boldsymbol{x}_{\mathbf{SAT}}$ and $\boldsymbol{x}_{\mathbf{GND}}$, can be related to each other as:

$$x_{SAT} = x_{GND} + \varepsilon_{total}$$

and the error

$$\varepsilon_{total} = x_{SAT} - x_{GND} \tag{1.3}$$

may be written as follows:

$$\varepsilon_{total} = \varepsilon_{SAT,rand} - \varepsilon_{GND,rand} + \varepsilon_{SAT,sys} - \varepsilon_{GND,sys} + \varepsilon_{S4D} + \varepsilon_{dx/d4D}, \tag{1.4}$$

where

- $\varepsilon_{SAT,rand}$ and $\varepsilon_{GND,rand}$ represent the random errors related to the measurement uncertainty of both sensors,
- $\varepsilon_{SAT,sys}$ and $\varepsilon_{GND,sys}$ represent the systematic errors related to the measurement uncertainty of both sensors,
- ε_{S4D} represents the so-called smoothing difference error, which contains horizontal, vertical and temporal components, and
- $\varepsilon_{dx/d4D}$ represents the so-called sampling difference error, which also contains horizontal, vertical and temporal components.

The sampling difference error is the error that would occur even if the measurements were point-like, but not perfectly coinciding. Note that this is not the same as the sampling error resulting from an incomplete sampling of a signal (e.g. von Clarmann, 2006). The smoothing difference error is the error that would occur even if the measurements have coinciding nominal locations (e.g. station location and pixel centre coincide), but different resolutions. Notice that in Eq. (1.4), due to definition of ε_{total} in Eq. (1.3) some errors have negative signs, of course in computing the total uncertainty these components do not cancel. For example in case of error uncorrelation the total uncertainty is given by the sum of all the uncertainties corresponding to the right hand side of Eq. (1.4).

From the error budget described in Eq. (1.4), it is in principle possible to derive an uncertainty budget in terms of variances (column measurements) or covariance matrices (profile measurements), e.g. to calculate $\boldsymbol{\sigma}^2$ in Eq. (1.2), but that would implicitly assume a symmetric distribution of the errors, and the absence of correlations between the different terms. As shown in Verhoelst et al. (2015), these assumptions are not always valid, and it is therefore advisable to investigate errors, and their probability density functions (PDFs), instead of uncertainties, whenever possible.

1.3 Definitions

The nomenclature followed throughout this document is based as far as possible on the international conventions published by the *Bureau International des Poids et Mesures* (BIPM) in the form of two key documents: the *Vocabulaire International de Métrologie* (VIM), and the Guide to the expression of Uncertainty in a Measurement (GUM), see also Sect. 1.4. Further definitions were taken from a list of authoritative documents and compiled into a reference table in the framework of the CEOS WG on Calibration and Validation (WGCV) with support from the EC FP7 project QA4ECV (http://qa4ecv.eu). These conventions are applied strictly in the current document, and they are provided as Annex A for reference. Frequently used terms and specific concepts not defined in the QA4ECV table are listed below. Note that these are consistent with the summary on terminology compiled within GAIA-CLIM as the “Guide to Uncertainty in Measurement and its Nomenclature”.

- **Metrology**: Definition 2.2 in the VIM: “science of measurement and its application (NOTE: Metrology includes all theoretical and practical aspects of measurement, whatever the measurement uncertainty and field of application).”
- **Measurement error**: Definition 2.16 in the VIM: “Measured quantity value minus a reference quantity value”.

- **Systematic measurement error**: Definition 2.17 in the VIM: “component of measurement error that in replicate measurements remains constant or varies in a predictable manner”

- **Random measurement error**: Definition 2.19: “component of measurement error that in replicate measurements varies in an unpredictable manner”

- **Measurement uncertainty**: Definition 2.26 in the VIM: “non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used”

- **Measurement bias**: Definition 2.18 in the VIM: “estimate of a systematic measurement error”

- **Uncertainty budget**: Definition 2.33 in the VIM: “statement of a measurement uncertainty, of the components of that measurement uncertainty, and of their calculation and combination”

- **Error budget**: Undefined in the VIM, but easily derived from the definition of the uncertainty budget: “statement of a measurement error, of the components of that measurement error, and of their calculation and combination”

- **Smoothing error**: the difference between the measurement and the truth at the nominal measurement location due to the smoothing properties of the instrument

- **Smoothing difference error**: Not to be confused with the smoothing error, the smoothing difference error represents the difference between two measurements, due to the differences in smoothing of the truth. E.g. for measurements with very similar smoothing properties, the smoothing difference error may be much smaller than the individual smoothing errors. See also Section 1.2.

- **Sampling error:** the difference between the measurement and the truth due to an incomplete sampling of the signal

- **Sampling difference error**: Not to be confused with the sampling error, the sampling difference error represents the difference between two measurements due to differences in sampling of the truth. See also Section 1.2.

- **Co-location mismatch**: Generic term implying the mismatch between two co-located measurements in spatiotemporal smoothing and sampling. It causes smoothing and sampling difference errors in the horizontal, vertical and temporal domains.

1.4 Related documents and projects

Within GAIA-CLIM

- The **Gap Assessment and Impacts Document** (GAID), in particular the gaps identified by WP2 and WP3, identified as G2.xx and G3.xx
- Deliverable D3.1, the **initial input from WP3 to the GAID**, which includes a literature review
- The GAIA-CLIM Guidance Note **Guide to Uncertainty in Measurement & its Nomenclature**
- Deliverable D3.2: **Generic metrology aspects of an atmospheric composition measurement and of data comparisons**
- Deliverable D3.6 (upcoming): **Library of smoothing/sampling error estimates for key atmospheric composition measurement systems, and smoothing/sampling error estimates for key data comparisons**

Nomenclature and metrology principles

- VIM, 3rd edition: ***International Vocabulary of Metrology – Basic and General Concepts and Associated Terms*** (VIM 3rd edition) JCGM 200:2012
- GUM: **Evaluation of measurement data – Guide to the expression of uncertainty in measurement** , JCGM, JCGM 100:2008, 2008, http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf
- **Data Modeling for Metrology and Testing in Measurement Science**, Pavese, F. & Forbes, A. B. (Eds.), Springer Science + Business Media, 2009
- **Measurement Uncertainty Analysis Principles and Methods**, NASA Measurement Quality Assurance Handbook – ANNEX 3, 2010
- **Annex A of the current document: the QA4ECV terms and definitions**

Related projects

- QA4ECV (www.qa4ecv.eu): Quality Assurance for Essential Climate Variables, aims at "*developing an internationally acceptable Quality Assurance (QA) framework that provides understandable and traceable quality information for satellite data used in currently evolving climate and air quality services.*" It serves in particular as preparation for the Copernicus Climate Change Service.
- FIDUCEO (www.fiduceo.eu): Fidelity and uncertainty in climate data records from Earth Observations aims at building "*nine new climate datasets from Earth Observation using a rigorous treatment of uncertainty, informed from the discipline of metrology*."
- MetEOC/MetEOC-2 (http://www.emceoc.org) aims at improving the metrology in Earth Observation (EO), and includes WPs on satellite calibration test sites, climate indicators, SI traceability of biophysical parameters, solar irradiance, and ECV measurements.
- GeoMON (website no longer active), was an EC FP6 project focusing on atmospheric composition and with multiple aims, ranging from better data production to validation and integration in models. It dealt extensively with the characterization of the horizontal smoothing properties of key atmospheric composition measuring instruments.
- NORS (nors.aeronomie.be), the "demonstration Network Of ground-based Remote Sensing observations in support of the Copernicus Atmospheric Service" was an EU FP7 project aimed at demonstrating the value of ground – based remote sensing data from the Network for the Detection of Atmospheric Composition Change (NDACC) for quality assessment and improvement of the Copernicus Atmospheric Service products.

2 Overview of approaches to quantify co-location mismatch

In Section 3, temperature and humidity profiles are considered. In particular an empirical approach is used to understand both the vertical smoothing and the spatio-temporal mismatch of IASI in the comparison with RAOB radiosonde profiles. Vertical smoothing is assessed by a data harmonization approach. In particular, RAOB profiles are vertically smoothed to match satellite data, using a flexible weighting function based on the Generalized Extreme Value pdf (GEV, Kotz and Nadarjah, 2000), which is independent of the averaging kernel and is fitted optimizing the distance w.r.t. IASI. To do this RAOB data have been interpolated using Hermite cubic splines and the quality of this interpolation has been assessed using reference GRUAN data. The spatio-temporal mismatch assessment is based on the harmonized RAOB and is based on an isotonic regression approach which gives the uncertainty for each combination of air distance, time delay and altitude.

In Section 4, temperature and humidity profiles are considered. In particular, the work on temporal mismatch uncertainties in radiosonde profiles aims to provide a direct means of estimating the temporal mismatch uncertainty in both temperature and humidity as a function of altitude, season and time of day, for selected sites using ERA Interim model data. The approach is validated by comparing the ERA Interim results with GRUAN-processed radiosonde data from those sites where long-term high frequency data is available. The method provides an estimate of mismatch uncertainty for any given site using global reanalysis data, and would enable an appropriate sampling strategy to be put in place for an intercomparison measurement application in order to meet a defined comparison uncertainty requirement.

In Section 5, ozone profiles are considered. The OSSSMOSE system (see Verhoelst et al., 2015, for a more in-depth description) is used to estimate the errors and uncertainties related with the particular sampling and smoothing properties of several ozone profiling instruments and their intercomparison. OSSSMOSE (Observing System of Systems Simulator for Multi-missiOn Synergies Exploration) uses (1) the metadata of actual measurement data sets, (2) parametrizations of the spatio-temporal sampling and smoothing properties of the measurement systems, and (3) high-resolution gridded representations of the atmosphere (such as reanalyses) to simulate measurements and their intercomparisons. As such, the system allows an estimate of co-location mismatch errors, which is based on a realistic representation of the real atmosphere and its variability. This method is independent from the measured data (e.g. those that are being validated) as only the metadata are used. The main advantages of this approach are (1) the possibility to estimate both smoothing and sampling (difference) uncertainties, (2) access to both random and systematic errors, and (3) the independence from actual measured values in the products to be validated, allowing a full consistency test on both the data and their reported uncertainties. The challenging part is the need for a reliable gridded representation of the atmosphere, at sufficient spatio-temporal resolution to resolve the smoothing and sampling mismatch.

In Section 6, columnar aerosol is considered. In particular, the total atmospheric column aerosol optical depth (AOD) from satellite (AATSR) and AERONET is compared. To do this, the spatial standard deviation of satellite AOD around the AERONET site is used to understand if the co-location mismatch is dominated by the spatial effects. The increase of Person's correlation coefficient among the ground and satellite AOD applying different threshold to spatial variability, permits interpretation of the spatial standard deviation of satellite AOD as a measure of co-location mismatch. However, this is not a direct measure of the co-location mismatch uncertainty because the variation of the AATSR AOD values around a site is very likely dominated by ADV/ASV retrieval errors.

In Section 7, aerosol profiles are considered. In particular, the comparison is related to CALIOP/CALIPSO and EARLINET aerosol backscatter profiles. In order to understand the co-location mismatch uncertainties, the co-location mismatch has been investigated as a function of the observational site, the CALIOP horizontal smoothing and altitude. To do this the horizontal smoothing level has been changed from 5 to 205 km along the satellite trajectory and the optimal horizontal smoothing has been computed for each site and atmospheric region. Optimal smoothing here, means the horizontal smoothing that minimizes co-location uncertainty. Differences among the various smoothing levels obtained permit interpretation of the co-location uncertainty in terms of local factors

3 Temperature and humidity profiles: comparison of IASI and RAOB

In this chapter, considering temperature and humidity in a region of continental Europe, the comparison of satellite and radiosonde data is organized as follows: in Section 3.4, IASI vertical smoothing is assessed by an independent method, this step allows harmonisation of RAOB measurements approximately to the same vertical smoothing as IASI products for further comparisons. In this first step, since RAOB products have a limited vertical resolution their information content is assessed by comparing RAOB and GRUAN products where available. As a result of this assessment, two additional uncertainties are defined: a sparseness uncertainty and a processing uncertainty.

In Section 3.5, using the harmonized data, the joint impact on co-location uncertainty profile of air distance and delay between the two products is considered and nighttime vs daytime results are compared.

The datasets considered here are provided by NOAA through the NPROVS system, discussed in some details in Section 3.3. In fact, NPROVS routinely (daily) compiles datasets of collocated radiosonde (and other products) collocated with satellites and sounding product suites. Hence, using this source of data makes the Virtual Observatory potentially permanently operative, see Section 3.6.

3.1 IASI description

The infrared atmospheric sounding interferometer (IASI) is a Fourier transform spectrometer based on the Michelson interferometer, associated with an integrated imaging system (Blumstein et al., 2004). As part of the payload of the MetOp series of polar-orbiting meteorological satellites, there are currently two IASI instruments in operation: on MetOp-A (launched 19 October 2006) and on Met-Op B (launched 17 September 2012) with the third due for launch in 2017.

IASI is a nadir-viewing instrument recording infrared emission spectra from 645 to 2760 cm^{-1}. It has 8461 spectral samples that are aligned in 3 bands within the spectral range. Correspondingly, the spectral resolution at which the measurements are made is 0.5 cm^{-1} after apodisation. Although primarily intended to provide information in near real-time on atmospheric temperature and water vapour to support weather forecasting, the concentrations of various trace gases can also be retrieved from the spectra.

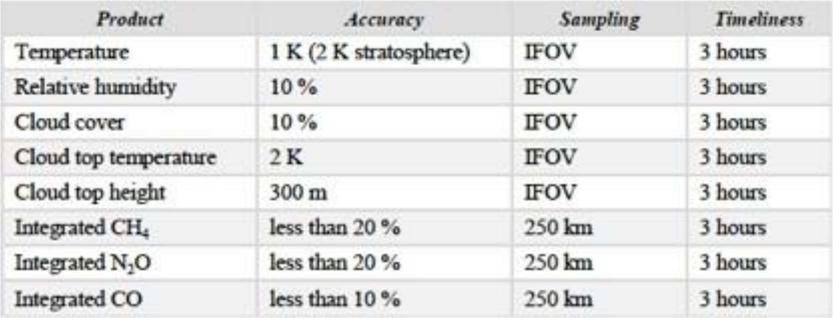
IASI belongs to the thermal infrared (TIR) class of spaceborne instruments, which are devoted to tropospheric remote sensing. On the operational side, it is intended as a replacement for the HIRS instruments, whereas on the scientific side, it continues the mission of instruments dedicated to atmospheric composition, which are also nadir viewing, Fourier Transform instruments (e.g. Atmospheric Chemistry Experiment).

The IASI spectral range has been chosen such that the instrument can record data from the following ranges:

- carbon dioxide strong absorption around 15 µm,
- ozone absorption around 9.6 µm,
- water vapour strong absorption between 6.1 and 8.2 µm,
- methane absorption up to the edge of TIR.

Horizontal sampling

IASI is an across-track scanning system with scan range of ±48° 20´, symmetrically with respect to the nadir direction. A nominal scan line covers 30 scan positions towards the Earth and two calibration views. One calibration view is into deep space, the other observes the internal black body. The scan starts on the left side with respect to the flight direction of the spacecraft. The elementary (or effective) field of view (EFOV) is the useful field of view at each scan position. Each EFOV consists of a 2 x 2 matrix of so-called instantaneous fields of view (IFOV). Each IFOV has a diameter of 14.65 millirad (mrad), which corresponds to a ground resolution of 12 km at nadir and a satellite altitude of 819 km. The 2 x 2 matrix is centered on the viewing direction. The instrument points spread function (PSF) is defined as the horizontal sensitivity within an IFOV. The IFOV diameter (D = 14.65 mrad) is defined so that the integral of the PSF over this circular area is larger than 95 %. The non-uniformity within the inner 80% of the IFOV (D=11.72 mrad) is not larger than ± 5 %. **Table 3.1** below summarises the requirements for the IASI Level 2 products with respect to product accuracy, sampling and timeliness (EUMETSAT, 2014).



Product	*Accuracy*	*Sampling*	*Timeliness*
Temperature	1 K (2 K stratosphere)	IFOV	3 hours
Relative humidity	10 %	IFOV	3 hours
Cloud cover	10 %	IFOV	3 hours
Cloud top temperature	2 K	IFOV	3 hours
Cloud top height	300 m	IFOV	3 hours
Integrated CH$_4$	less than 20 %	250 km	3 hours
Integrated N$_2$O	less than 20 %	250 km	3 hours
Integrated CO	less than 10 %	250 km	3 hours

Table 3.1. Requirements for IASI level 3 products.

Temporal sampling

The IASI instrument produces around 1,300,000 spectra every day. It takes around 8 seconds for IASI to acquire data from one complete across track scan and the onboard calibration. The former consists of 120 interferograms, each one corresponding to one pixel. At nadir, the instrument samples data at intervals of 25 km along and across track.

Vertical smoothing

The IASI sounding products represent thermodynamic states of deep atmospheric layers at variable depths, due to the integrating nature of the radiation measurements at the top of the atmosphere. The maximum number of independent pieces of information determined in the temperature profile is 14, with a maximum of 10 pieces for the moisture profile—but these numbers vary with the atmospheric situation. In summary, the true vertical resolution of the retrieved profiles is lower than the vertical grid defined in the products. Profiles retrieved from such radiance measurements are smoothed versions, where the smoothing functions are given by the averaging kernels.

An example of a set of averaging kernels for temperature and humidity is shown in Figure 3.1. Two things can be seen: the vertical extent over which a particular kernel averages, and the amplitude, which shows how sharply a kernel peaks at a particular height. Higher amplitudes indicate more information about the corresponding layer. For example, an amplitude of one would indicate perfect measurements at a distinct level; however, this is purely hypothetical and does not exist. Nevertheless, the retrieved profiles are represented on a fine vertical grid for the reason that the averaging kernels vary with atmospheric situation.

Consequently, the vertical resolution and the central altitudes of the resolved layers vary also. The actual variation is not known *a priori*, so the retrieval is performed on a fixed, fine pressure grid and the smoothing is represented by the *a posteriori* error covariance matrix, which is part of the product and represented on the same pressure grid. The off-diagonal elements of the covariance matrix describe the inter-relationship between the state-vector elements and provide information about the actual vertical resolution.

In particular, the IASI dataset used within GAIA-CLIM WP3, is provided by NOAA/NESDIS. The atmospheric profiles produced by NOAA/NESDIS (Reale et al., 2010) are provided at 100 pressure levels from 1100 to 0.0161 hPa. A statistical method—PWLR (Piece-Wise Linear Regression) is applied using the input measurements available: IASI and AMSU/MHS, or IASI only. It provides estimates of the temperature, water-vapour and ozone profiles as well as surface skin temperature in all-sky conditions—clear and cloudy scenes. The PWLR retrievals serve as “first guess” to the optimal estimation method. The latter essentially involves comparing the measured spectra with an a priori spectrum. Subsequently, the a priori spectrum is perturbed as many times as needed to adjust the simulated spectrum such that it resembles the measured one as closely as possible (e.g. the amount of the measured species allowing the retrieval is perturbed, for example the amount of CO2 is variated to allow the retrieval of temperature profiles). It must be noted that a variety of errors must be taken into consideration while perturbing the a priori, such as the error on the a priori, the instrumental error and the expected error.

3.2 RAOB description

A radiosonde is a battery-powered telemetry instrument package carried into the atmosphere usually by a weather balloon that measures various atmospheric parameters and transmits them by radio to a ground receiver. The radiosonde contains instruments capable of making direct in-situ measurements of air temperature, humidity and pressure with height, typically to altitudes of approximately 30-35 km, depending on several factors (including location and season).

Conventional radiosonde observations (RAOBs) have been used historically as a commonly de facto accepted reference dataset in satellite measurement and derived product validation. The upper-air data that are collected and transmitted during the flight of a radiosonde include the air pressure, air temperature and humidity measured continuously by the instruments aboard the radiosonde package.

Worldwide there are more than 800 radiosonde launch sites. Most countries share data with the rest of the world through international agreements. Nearly all routine radiosonde launches occur 45 minutes before the official observation time of 0000 UTC and 1200 UTC, more rarely at 6 or 18 UTC.

The radiosonde transmits temperature and relative humidity data at each pressure level. The altitudes of these levels are calculated using GPS. In older models, this is done using the hypsometric equation that relates the vertical height of a layer from the mean layer temperature, the humidity of the layer and the air pressure at top and bottom of the layer. Significant levels where the vertical profiles of the temperature or the dewpoint undergo a change are determined from the sounding. The height of the troposphere and stability indices are calculated.

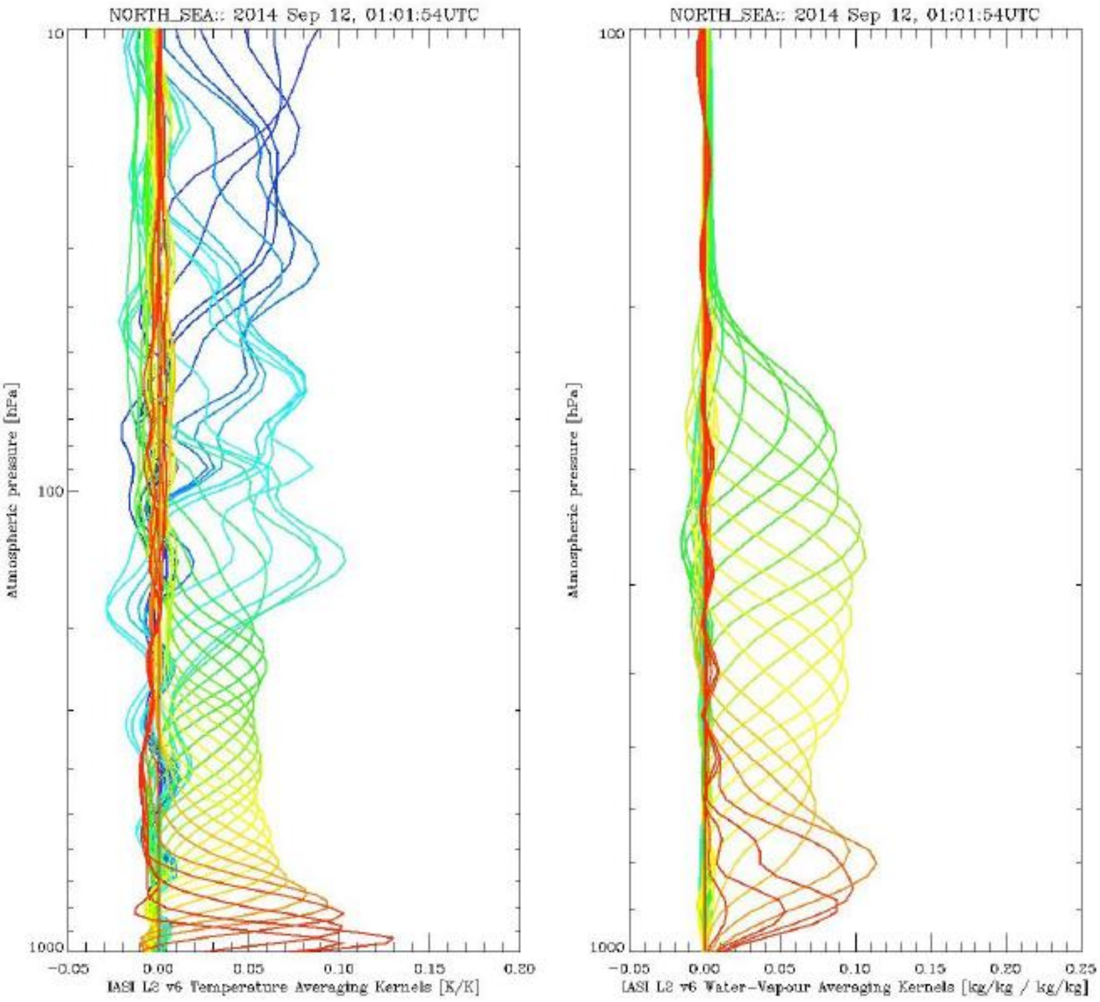


Figure 3.1. Averaging kernels for temperature (left) and water-vapour (right) profiles for a mid-latitude ocean sample taken on 12/09/2014. Sampling

RAOB observations are processed, tabulated and encoded for transmission over communication networks. While the radiosonde transmits an essentially continuous stream of temperature and humidity information back to the station (each 5-10 m of altitude, measured each 1-2 s). The RAOB data include temperature, humidity, and wind profiles at 22 mandatory levels (including surface) and various significant levels:

- **MANDATORY LEVELS:** By international convention, the following 22 specific pressure levels must be reported in the RAOB message: the surface, 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, 1 hPa. The information for some of these levels is plotted routinely on constant pressure charts to show the spatial variability of these levels and it is used as input in the numerical weather prediction models. Note that many sonde sites do not routinely measure to the highest levels owing to balloon burst.

- **SIGNIFICANT LEVELS:** In addition to the mandatory levels, as part of the RAOB message are included pressure levels at ***significant*** or abrupt changes and extrema in the vertical temperature and/or dewpoint temperature profiles. In a radiosonde observation, a significant level is a level other than a mandatory level, for which values of pressure, temperature, and humidity are reported either because they are sufficiently important or unusual to warrant the attention of the forecaster, or these values are required for the reasonably accurate reproduction of the radiosonde observation. There are definite rules governing the selection of significant levels, set forth in the Manual of Radiosonde Observations, WBAN Circular P, 7th ed. rev., 1957. As a result, significant levels change among different profiles and, on average, we have 28 significant levels per profile in the dataset considered. By assuming that the temperature and dewpoint profiles change linearly with height between significant levels (*i.e.,* constant environmental lapse rate), a reasonably accurate reproduction of the RAOB sounding can be made from the sequence of RAOB message information at significant levels and supplemented by the mandatory levels.

3.3 Datasets

The analysis of RAOB-IASI co-locations is based on the dataset provided by NOAA through the co-location software known as NPROVS system (http://www.star.nesdis.noaa.gov/smcd/opdb/nprovs/) and two auxiliary datasets, namely ECMWF and GRUAN. The NPROVS dataset includes K=3900 co-locations, from unbalanced design over 21 RAOB stations across the central European area (C-EU) in Figure 3.2 over the period January 2015 – February 2016. The three continental GRUAN stations at Cabauw, Lindenberg and Payerne are included (Section 3.3.2).

Each co-location record includes RAOB and IASI measures for temperature and water vapor mixing ratio (WVMR). Considering IASI data, NPROVS dataset provided also IASI retrieval uncertainties, but individual averaging kernels discussed above were not available in this dataset.

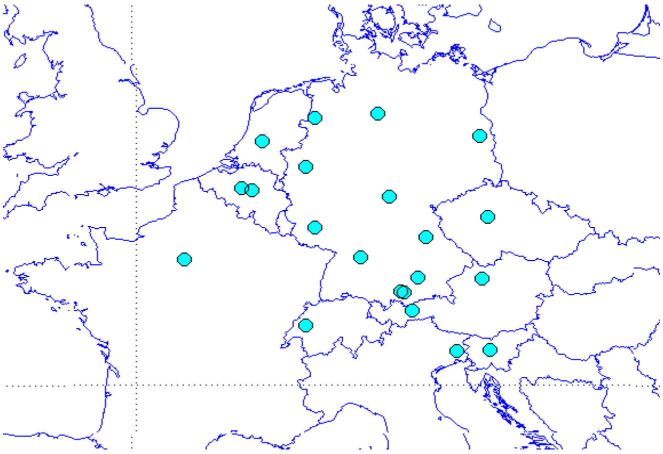


Figure 3.2. Central European RAOB network (C-EU) including three continental GRUAN stations at Cabauw, Lindenberg and Payerne.

3.3.1 ECMWF

ERA-Interim is a global atmospheric reanalysis from 1979, continuously updated in real time. The data assimilation system used to produce ERA-Interim is based on a 2006 release of the Integrated Forecast

System (IFS - Cy31r2). The system includes a 4-dimensional variational analysis (4D-Var) with a 12-hour analysis window (Dee et al., 2011).

Each single analysed observation is described as instantaneous, though it represents an average over the model time step (30 minutes for ERA-Interim). Depending on the parameter, forecast data in ERA-Interim is either instantaneous or accumulated from the beginning of the forecast (twice daily forecasts starting at 00 and 12 UTC). Parameters such as precipitation and radiation are accumulated. The accumulated parameters in ERA-Interim are listed in Tables 9, 13-15 of the ERA-Interim archive document (http://www.ecmwf.int /en/elibrary/ 8174-era-interim-archive-version-20).

Analyses are available every 6 hours (0, 6, 12, 18 UTC) and the twice daily forecasts (from 00 and 12 UTC) provide output for surface and pressure level parameters, at forecast steps every 3 hours to 24 hours, then with decreasing frequency to 10 days.

Product description

The ERA-Interim dataset contains atmospheric and surface parameters: 6-hourly atmospheric fields on model levels, pressure levels, potential temperature and potential vorticity; 3-hourly surface fields and daily vertical integrals; monthly averages of daily means and Synoptic monthly averages at 0 UTC, 6 UTC, 12 UTC, 18 UTC.

The temporal coverage is from 1 January 1979 to present and spatial coverage is global.
The spectral resolution is T255 (T255 spherical-harmonic representation of the basic dynamic fields), while tentative horizontal resolution is ~80 km (reduced Gaussian grid N128); ~83km/0.75 deg when interpolated to a regular lat/long grid.

There are 60 vertical levels from the surface up to 0.1 hPa. The list of ERA-Interim fields (parameters) available for downloading can be accessed in the ECMWF dataserver. Note that while ERA-Interim "runs" in near real time, data is published with a few months delay. Note also that several quality issues are known with ERA-Interim data and ERA-Interim daily retrieval efficiency documentation gives indications for best practices in retrieving ERA-Interim data.

For the GAIA-CLIM WP3 objectives, ERA-INTERIM forecast daily data have been co-located with all the RAOB stations in a box of latitude (35°, 70°) and longitude (-10°, 35°) . The investigated dataset covers the period 01-Jan-2015 – 29-Feb-2016. For a detailed documentation of the ERA-Interim Archive see Berrisford et al. (2011) .

3.3.2 GRUAN

RAOB represents the comprehensive contribution of the radiosounding measurement to the global observing system. However, RAOB data cannot provide any reference-quality (Immler et al. 2010, hereafter I2010) in situ and ground-based remote sensing observations of upper-air essential climate variables (ECVs; Seidel et al. 2009; Bojinski et al. 2014).

For this reason, the GCOS (Global Climate Observing System) Reference Upper-Air Network (GRUAN) data processing for the Vaisala RS92 radiosonde was developed to meet the criteria for reference measurements (Dirksen et al., 2014). These criteria stipulate the collection of metadata, the use of well-documented correction algorithms, and the estimation of the measurement uncertainty. GRUAN radiosounding profiles are provided at high vertical/temporal resolution (1-2 s equivalent to 5-10 m in altitude). An important and novel aspect of the GRUAN processing is that the uncertainty estimates are vertically resolved.

An additional GRUAN requirement for performing reference measurements with the RS92 is that the manufacturer-prescribed procedure for the radiosonde's preparation, i.e. heated reconditioning of the sensors and recalibration during ground check, is followed. In the GRUAN processing however, the recalibration of the humidity sensors that is applied during ground check is removed. For the dominant error source, solar radiation, laboratory experiments were performed to investigate and model its effect on the RS92's temperature and humidity measurements. GRUAN uncertainty estimates are 0.15 K for night-time temperature measurements and approximately 0.6 K at 25 km during daytime but vary vertically and on a per ascent basis. The other uncertainty estimates are up to 6% relative humidity for humidity, 10–50 m for geopotential height, 0.6 hPa for pressure, 0.4–1 m s−1 for wind speed, and 1° for wind direction. Daytime temperature profiles for GRUAN and Vaisala processing are comparable and consistent within the estimated uncertainty. GRUAN daytime humidity profiles are up to 15% moister than Vaisala processed profiles, of which two-thirds is due to the radiation dry bias correction and one-third is due to an additional calibration correction. Redundant measurements with frost point hygrometers (CFH and NOAA FPH) show that GRUAN-processed RS92 humidity profiles and frost point data agree within 15% in the troposphere. No systematic biases occur, apart from a 5% dry bias for GRUAN data around −40 °C at night. GRUAN radiosonde daytime dry biases with respect to LBLRTM and IASI have been of the order of 2.5 % in the upper troposphere, though based on a relatively small sample size of comparisons (Calbet et al., 2016).

GRUAN data processing is applied currently on the RS92 radiosondes. In the future also measurements by other radiosonde types will be used to apply GRUAN processing. An example of CFH/RS92/RS41 comparisons is shown in Figure 3.3.

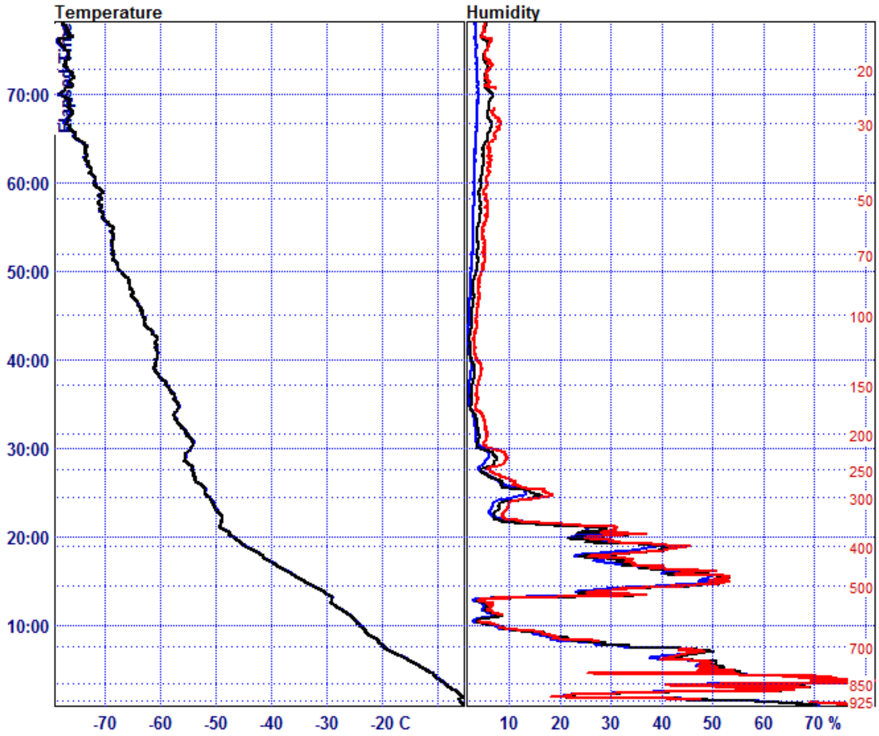


Figure 3.3. Right panel: relative humidity measurements by three different sensors in the same payload: CFH (red), RS41 (black) and RS92 (blue). The comparison flight took place at the GRUAN station Sodankylä on November 23, 2015. Left panel: corresponding temperature profile measured by RS41 radiosonde.

For the purposes of GAIA-CLIM WP3, GRUAN reference measurements will be used to quantify the co-location error with a high vertical resolution and to support the co-location error quantification for the comprehensive measurements provided by RAOB. To do this, the data of the continental GRUAN stations at Cabauw, Lindenberg and Payerne will be used.

3.4 Vertical smoothing

RAOB and IASI profiles are not immediately comparable because they are not available at the same pressure levels and have different vertical smoothing. In particular, IASI observations come at fixed pressure level grid, while RAOB pressure level grid changes across launches. Moreover, IASI observations, due to kernel averaging, have an important vertical smoothing, which is essentially absent in radiosonde. In fact, the IASI profiles are available at fixed pressure levels, but their effective vertical resolution is much coarser because of smoothing resulting from the use of averaging kernels. RAOB radiosounding profiles are reported at 22 mandatory plus approximately 28 significant pressure levels which differ among different profiles. Although there is no smoothing effect, the reporting at only these few atmospheric vertical levels does not allow to catch all the gradients of the temperature and water vapor field.

For these reasons, the approach reported below has the aim to reconcile the difference between the two measurement techniques and, then, to assess the co-location uncertainty after considering sparseness and vertical smoothing uncertainties.

3.4.1 Vertical harmonization

For a given co-location k, let $\boldsymbol{x}_G$ be the RAOB measure vector related to the pressure levels $\boldsymbol{p}_G = (p_{G,1}, \dots, p_{G,N_k})$, while let $\boldsymbol{x}_S$ be the IASI measure vector related to the pressure levels $\boldsymbol{p}_S = (p_{S,1}, \dots, p_{S,M})$. Note that the number of pressure levels for RAOB depends on co-location k while M is fixed. Moreover, in general RAOB and IASI measurements are never available at the same pressure level, that is $p_{G,l} \neq p_{S,h}, l = 1, \dots, N_k, h = 1, \dots, M$. The solution we adopt is to transform discrete RAOB into continuous profiles in order to calculate the differences at the fixed IASI levels $\boldsymbol{p}_S$.

Therefore, the choice of the profile to be transformed into a continuous profile is crucial, as implications are different. In our case, the choice is driven by the smoothing properties of RAOB and IASI measurements. As discussed in Section 3.1, RAOB measurements have practically no vertical smoothing while IASI measurements have a relevant vertical. In order to understand the effect of the IASI vertical smoothing on the difference between RAOB and IASI, the vertical smoothing must be allowed to vary. A natural solution is to vary the smoothness of RAOB as the original measures are not smoothed. This requires to compute integrals of a continuous profile and for this reason only RAOB profiles are transformed into continuous profiles.

3.4.2 Sparseness and processing uncertainty

In this Section the properties of RAOB interpolation are discussed. In fact, the limited number of mandatory and significant levels entails a loss of information w.r.t. to high resolution vertical profiles such as GRUAN profiles. Moreover, comparing reference measures with RAOB at observation levels allows to assess Vaisala processing uncertainty.

To see this, RAOB profiles $\boldsymbol{x}_G$ observed at discrete levels are transformed into continuous profiles $\boldsymbol{x}_G^S$ using spline interpolation. Within the spline family, we adopt interpolating cubic Hermite splines since they avoid unnatural oscillations of the profile and ensure that the profile $\boldsymbol{x}_G^S$ satisfies the following property (Catmull and Rom, 1974).

Internality Property. *Consider RAOB profile* $\boldsymbol{x}_G$ *observed at pressure levels* $p_{G,l}, l = 1, \dots, N_k$ *as defined in Section 3.4.1 and its cubic Hermite spline denoted by* $\boldsymbol{x}_G^S$*. For pressure level* p *such that* $p_{G,l} \le p \le p_{G,l+1}$*, we have that*

$$x_{min}(p) \le \boldsymbol{x}_G^S(p) \le x_{max}(p)$$

where $x_{min}(p) = min\left(x_G(p_l), x_G(p_{l+1})\right)$ *and* $x_{max}(p) = max(x_G(p_l), x_G(p_{l+1}))$ *for* $l = 1, \dots, N_k$.

RAOB vertical profiles are first interpolated using Hermite splines. In order to assess the accuracy of this approach, the interpolated RAOB are validated using the GRUAN dataset of Section 3.3.2, which covers the same period and a subset of RAOB stations, namely Cabauw, Lindenberg and Payerne. For this reason, validation is based only on 306 launches for temperature and 439 launches for WVMR.

The spline uncertainty is assessed by considering the difference between GRUAN measures and RAOB splines. Figures 3.4 and 3.5 show the uncertainty with respect to pressure. Minima of the graph are the mandatory pressure levels, where there is always a RAOB measure. Note that these minima are not zero as GRUAN retrievals are obtained using GRUAN processing discussed in Section 3.3.2, which is different from Vaisala processing, used in the RAOB dataset. Hence these minima may be used as an estimate of Vaisala-GRUAN processing uncertainty. In particular for temperature, this uncertainty is nearly constant, about 0.12 K up to 300 hPa, and increases up to 0.39 K above. For humidity, the Vaisala-GRUAN processing uncertainty has an approximately linear decrease from 0.1 g/kg at 925 hPa to about 0.01 g/kg at 300 hPa.

Once this Vaisala processing uncertainty is eliminated the remaining part is related to lack of information of RAOB products related to vertical sparseness. For temperature, in the range 1000-300 hPa this is smaller than 0.48 K but increases to about 1 K in upper reaches of the measured profile. For humidity this is smaller than 0.22 g/kg up to 500 hPa and smaller than 0.05 g/kg above.

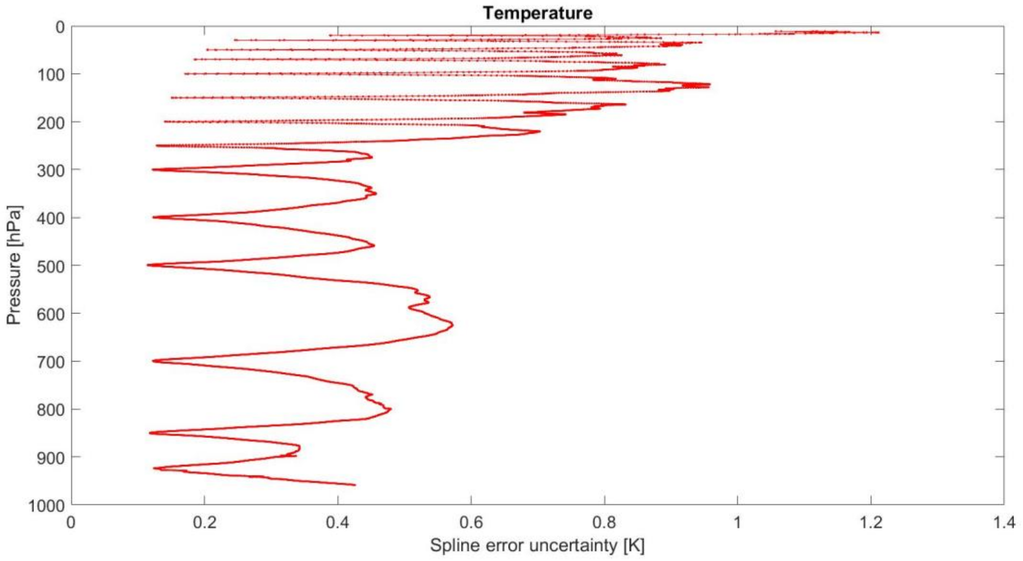


Figure 3.4. Temperature: Spline interpolation uncertainty assessed through GRUAN-RAOB comparison. Local minima of interpolation uncertainty occur at mandatory levels: 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20 hPa.

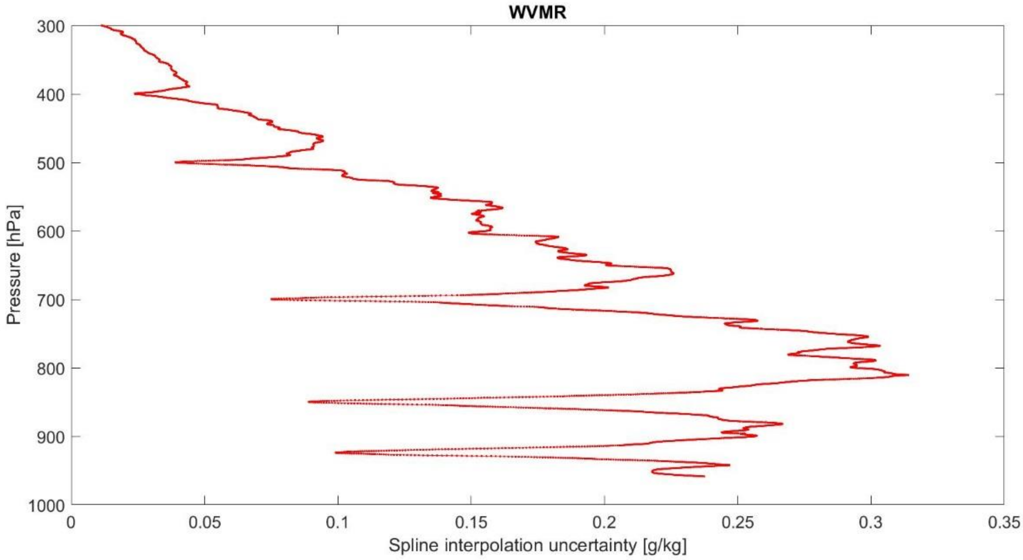


Figure 3.5. WVMR: Spline interpolation uncertainty assessed through GRUAN-RAOB comparison. Local minima of interpolation uncertainty occur at mandatory levels: 925, 850, 700, 500, 400 hPa.

3.4.3 Vertical smoothing uncertainty

In the comparison of co-located RAOB-IASI profiles, we consider that the different vertical smoothness of the two data products contributes to the total co-location uncertainty. In order to assess the vertical smoothing uncertainty, we first set up an optimization problem which aims at deriving the IASI vertical smoothness at each pressure level $p_{S,h}$.

To do this, we introduce the smoothed RAOB measure at pressure level $p_{S,h}$, which is given by

$$\tilde{x}^S_{G,h}(\boldsymbol{\theta}) = \frac{\int \boldsymbol{x}^S_G(p) w(p;\boldsymbol{\theta}) dp}{\int w(p;\boldsymbol{\theta}) dp}$$

where $w(p;\boldsymbol{\theta})$ is a weight function with parameter vector $\boldsymbol{\theta}$. The formula above is adjusted to avoid biases at the borders of the vertical domain.

We tested weight functions with different shapes and, in particular, the flat (rectangluar) function, the cosine function, the Gaussian function and the generalized extreme value (GEV) density function. The first three functions are symmetric and centered at the nominal pressure level $p_{S,h}$. On the other hand, the GEV function is characterized by three parameters, so that $\boldsymbol{\theta} = (shape, scale, position)$. Notice that GEV may show both left and right skewness and its peak may have an offset from the nominal pressure level.

For each weight function and each pressure level $p_{S,h}$, the parameter vector $\boldsymbol{\theta}$ is obtained by minimizing the distance between IASI $x_{S,h}$ and smoothed RAOB $\tilde{x}^S_{G,h}$. To do this, consider K different RAOB-IASI co-location profiles observed across space and time and let $E()$ be the average operator over these K co-locations. Hence the optimal smoothing is given by the following minimum co-location mean square error problem:

$$\boldsymbol{\theta}_h = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, E\left[\left(x_{S,h} - \tilde{x}^S_{G,h}(\boldsymbol{\theta})\right)^2\right]$$

and the quantity

$$\tilde{x}^S_{G,h} = \tilde{x}^S_{G,h}(\boldsymbol{\theta}_h)$$

is the optimally smoothed RAOB. Additionally, the vertical smoothing uncertainty at pressure level $p_{S,h}$ is given by

$$u^2_{v,h} = E\left[\left(\tilde{x}^S_{G,h} - x^S_{G,h}\right)^2\right]$$

and, in practice, can be computed by

$$u^2_{v,h} = u^2_h - u^2_{c,h}$$

where u^2_h is the total co-location uncertainty at altitude h, which includes differences in horizontal and temporal co-location, as well as differences in vertical smoothing, that is:

$$u^2_h = E\left[\left(x_{S,h} - x^S_{G,h}\right)^2\right].$$

Moreover, $u^2_{c,h}$ is the adjusted co-location uncertainty, corrected for differences in vertical smoothing, that is:

$$u^2_{c,h} = E\left[\left(x_{S,h} - \tilde{x}^S_{G,h}\right)^2\right].$$

3.4.4 Analysis and results

The original NPROVS dataset is first filtered in order to have RAOB profiles that cover the vertical range [958.6-10.0] hPa, for temperature, and the vertical range [958.6-300.0] hPa for WVMR. In order to give good spline interpolation, RAOB profiles are additionally selected to have at least 20 standard and significant level measures for temperature and 14 for moisture. This filtering results in the availability of 1596 co-locations for temperature and 2648 co-locations for WVMR.

Smoothing optimization

RAOB and IASI vertical profiles are first interpolated using Hermite splines. Subsequently, the optimization procedure described above is implemented in order to estimate $\boldsymbol{\theta}_h$ at each IASI pressure level and for each weight function. Figures 3.6 and 3.7 depict the optimization results for temperature and WVMR, respectively. For both temperature and humidity, it can be seen that the GEV function attains the lowest uncertainty for most of the pressure levels. Therefore, all the subsequent analysis are based on the GEV function. Figures 3.8 and 3.9 show the shape of the GEV functions for each IASI pressure level while Figures 3.10 and 3.11 show the offset and the width of the GEV functions. In particular, the offset is the difference between the IASI nominal level and the location of the GEV function peak, while the width is given in terms of Full Width at Half Maximum (FWHM).

It is important to note that the above results were obtained considering all the (filtered) co-locations, regardless of their date, location, spatial mismatch and temporal mismatch. This implies that the results hold for any past and future co-location observed in the Central European area (C-EU). If this result has the value of generality, it may be considered sub-optimal, in the sense that seasonal and daily variations of atmosphere are not considered. In fact, a dynamic procedure covering for daily atmospheric variations will be used for the Virtual Observatory which is discussed in Section 3.6.

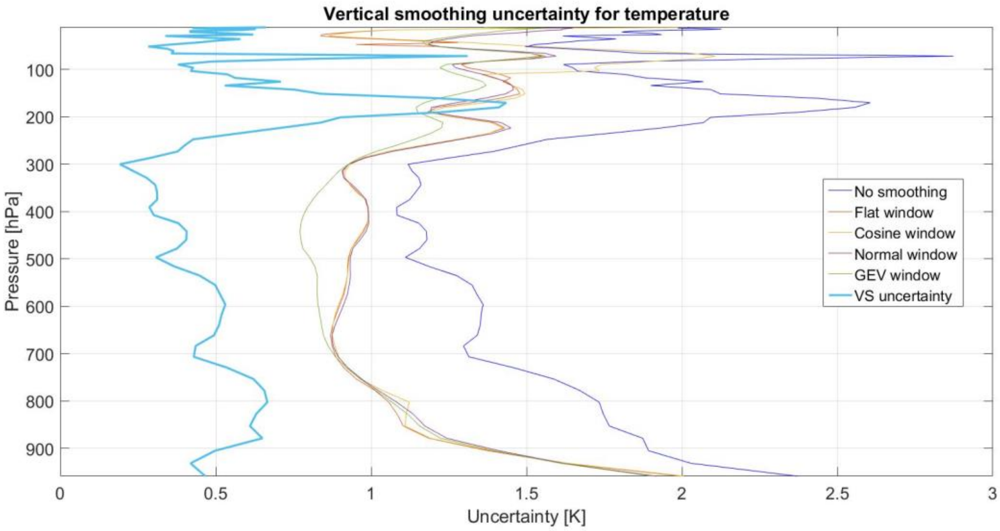


Figure 3.6. Uncertainty decomposition for temperature. The dark blue curve is the total co-location uncertainty u_h while the red, yellow, purple and green curves are the adjusted co-location uncertainty $u_{c,h}$ for the flat, cosine, normal and GEV weight functions, respectively. The thick light blue curve is the vertical smoothing uncertainty $u_{v,h}$ based on the GEV function, which is the best weight function.

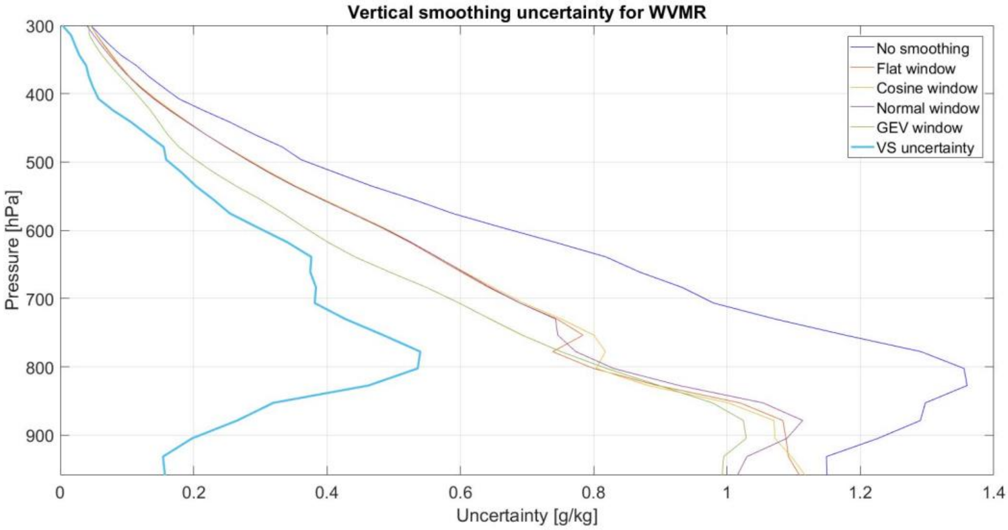


Figure 3.7. Uncertainty decomposition for WVMR. The dark blue curve is the total co-location uncertainty u_h while the red, yellow, purple and green curves are the adjusted co-location uncertainty $u_{c,h}$ for the flat, cosine, normal and GEV weight functions, respectively. The thick light blue curve is the vertical smoothing uncertainty $u_{v,h}$ based on the GEV function, which is the best weight function.

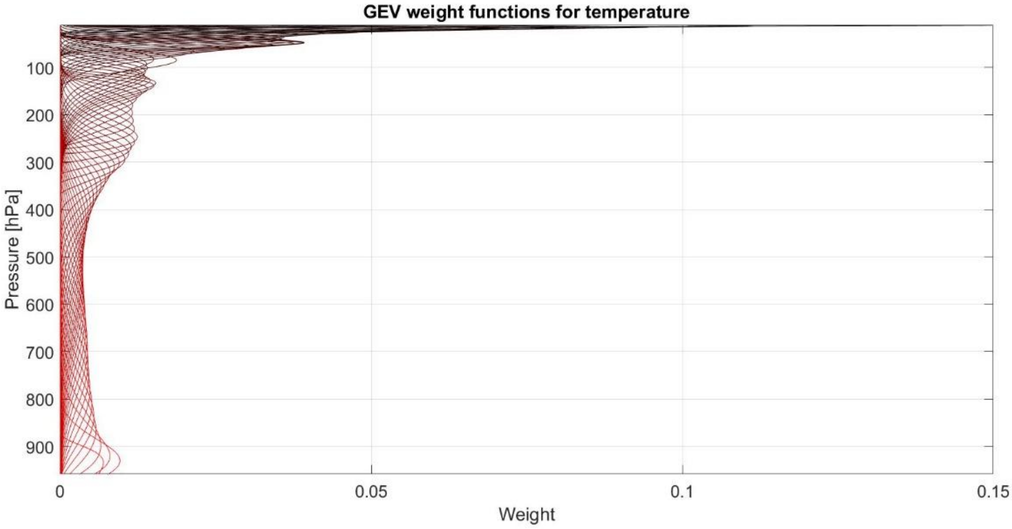


Figure 3.8. GEV weight functions for temperature.

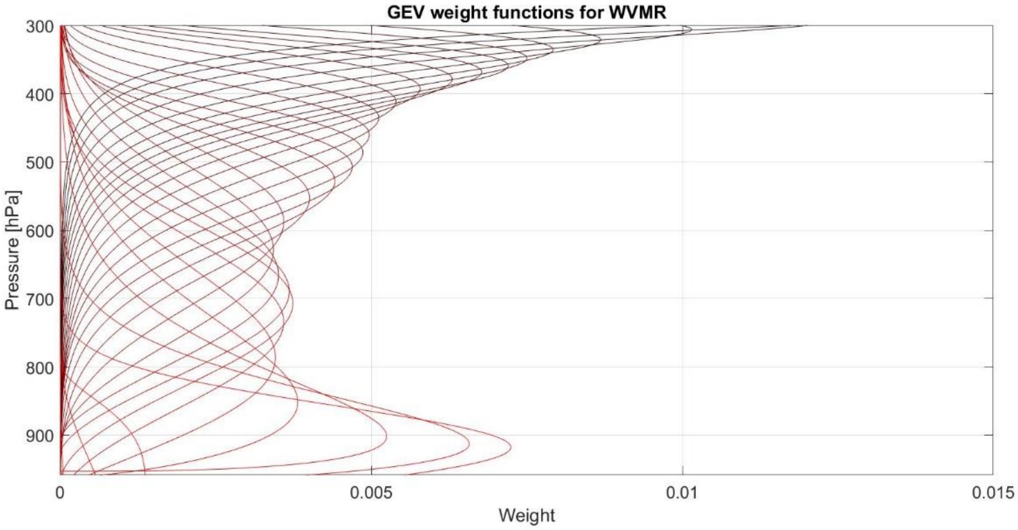


Figure 3.9. GEV weight functions for WVMR.

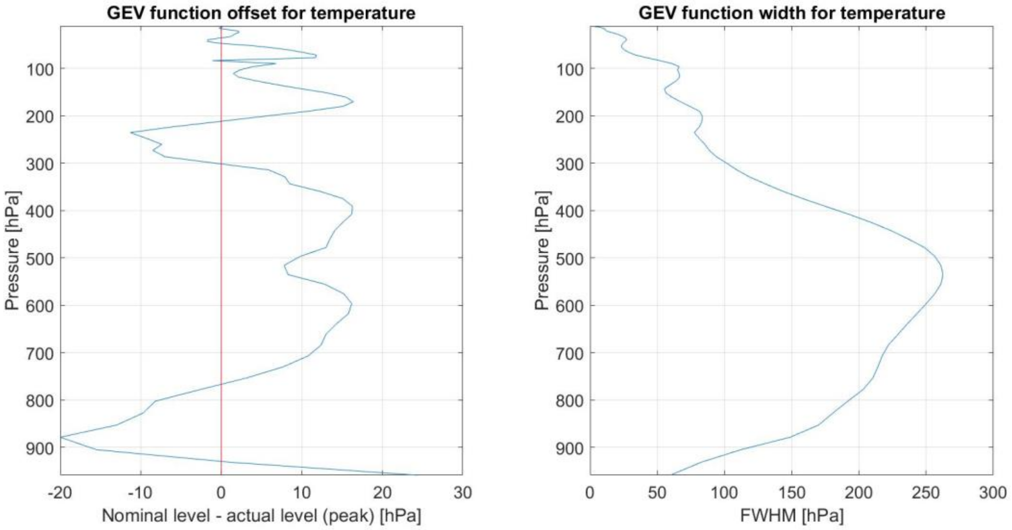


Figure 3.10. GEV function offset and width for temperature.

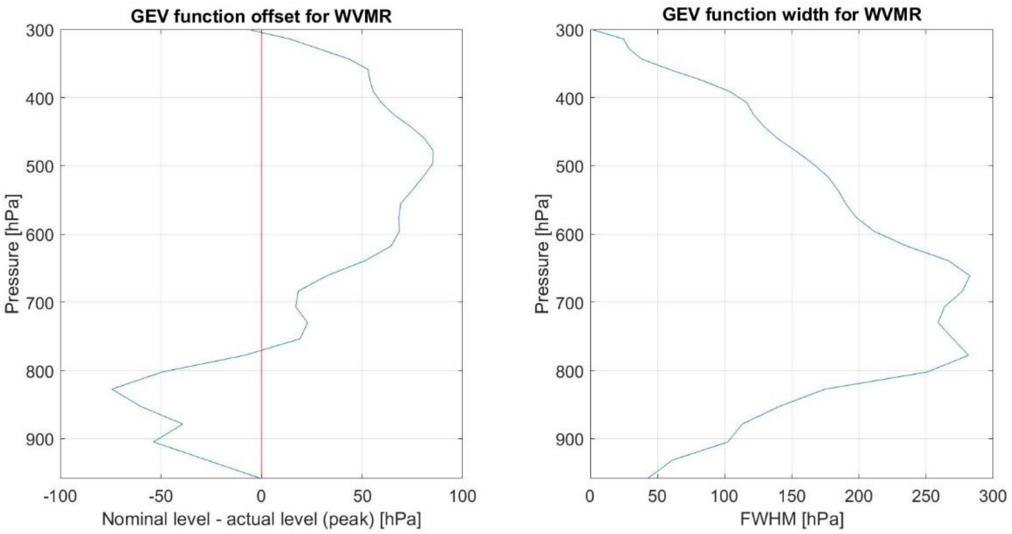


Figure 3.11. GEV function offset and width for WVMR.

Vertical smoothing and co-location uncertainty

The optimization results of the previous paragraph are used to evaluate both the vertical smoothing uncertainty $u_{v,h}^2$ and the adjusted co-location uncertainty $u_{c,h}^2$. In fact, the latter is a byproduct of the optimization procedure since $u_{c,h}^2$ describes the irreducible error, namely the error budget component that cannot be reduced by changing $\boldsymbol{\theta}$. On the other hand, the vertical smoothing uncertainty is obtained by first evaluating the total uncertainty at each pressure level and by computing $u_{v,h}^2 = u_h^2 - u_{c,h}^2$. This uncertainty decomposition is displayed in Figures 3.6 and 3.7 and is based on the GEV weight function. In particular, both total co-location uncertainty, u_h, and vertical smoothing uncertainty, $u_{v,h}$, show the same periodic behavior with local minima at mandatory levels as in Figures 3.4 and 3.5, due to sparseness uncertainty discussed in Section 3.4.2. Interestingly, this effect is largely reduced in the co-location uncertainty $u_{c,h}$. This is due to the fact that also the sparseness error is smoothed by our procedure. Note that for humidity, the sparseness effect is a smaller fraction of total co-location uncertainty due to the larger atmospheric variability of humidity, giving a smoother behavior of total co-location and vertical smoothing uncertainties.

Considering temperature, we see that vertical smoothing uncertainty, in comparison to total and co-location uncertainties is roughly constant, about 0.5 K, or slowly decreasing up to 300 hPa; above, it increases more than co-location uncertainty which is more stable. On the other side, considering humidity, total and vertical smoothing uncertainties have a maximum near 800 hPa and above they smoothly decrease. In particular, vertical smoothing uncertainty decreases to zero faster than the other two involved uncertainties.

Note that the above results do not make use of IASI averaging kernels, which were not available in NPROVS as discussed in Section 3.3. Although, in principle, it may be interesting to compare the results obtained by our smoothing and the application of original IASI kernels to RAOB data, we consider valuable the introduction of a methodology, which is able to perform optimal vertical harmonization in absence of direct information on vertical smoothing of one of the two products being compared.

3.5 Spatio-temporal mismatch

In this section, we consider the effect of horizontal and temporal mismatch to the co-location uncertainty of the previous section. In particular, the vertically harmonized data of the previous section are considered with a focus on the co-location mean square error

$$u_{c,h}^2 = E\left[\left(x_{S,h} - \tilde{x}_{G,h}^S\right)^2\right]$$

where, as before, $x_{S,h}$ is IASI temperature or humidity and $\tilde{x}_{G,h}^S$ is the corresponding optimally smoothed RAOB measurement.

Using the C-EU data of Section 3.3, the objective here is to understand the variation of the above co-location uncertainty as a function of the horizontal distance between IASI and RAOB measurements at each IASI pressure level $p_{S,h}$, that is

$$u_{c,h}^2 = \sigma^2(\partial a, \partial t)$$

Similarly to Sofieva et al. (2008) $\partial a = air.distance$ is the air mass distance, that is the distance between IASI and RAOB adjusted for advection due to wind. Namely $\partial a = |\partial s + w\partial t|$ where ∂s is the displacement vector among RAOB and IASI measurements at height h, w is the wind vector and ∂t is the time delay between measurements.

Since, from physical considerations, we may assume that $\sigma^2(\partial a, \partial t)$ is a nondecreasing function in each of its coordinates, we estimate σ^2 by fitting a bidimensional isotonic regression model to squared co-location errors

$$\epsilon^2 = (x_{S,h} - \tilde{x}_{G,h}^S)^2.$$

The bidimensional isotonic regression technique is a constrained least square method, which gives an estimate of σ^2 such that $E(\sigma^2 - \epsilon^2)^2$ is a minimum, conditional on marginal monotonic constraints, that is $\sigma^2(\partial a + d, \partial t) \geq \sigma^2(\partial a, \partial t)$ and $\sigma^2(\partial a, \partial t + d) \geq \sigma^2(\partial a, \partial t)$ for any $d > 0$. The optimization is carried out by the quadratic programming technique proposed by Meyer (2013).

3.5.1 Analysis and results

The isotonic model has been fitted separately to daytime and nighttime co-locations. Since atmosphere is more stable under nighttime conditions, we start with nighttime uncertainty.

Nighttime temperature co-location uncertainty

The temperature co-location uncertainty $\sigma(\partial a, \partial t)$, or co-location root mean square error (RMSE), is reported in Figures 3.12-3.14 at three different pressure levels for nighttime data. These figures show the variation of co-location uncertainty in the *air.distance*×*delay* domain, respectively in the lower troposphere, upper troposphere and lower stratosphere. The vertical profile of uncertainty σ is reported in 3.15, which gives minimum, average and maximum uncertainty with respect to *air.distance* ×*delay*. As a result, overall in C-EU, co-location uncertainty is smaller than 1.5 K, except near ground level where it reaches 1.8 K. But large uncertainties (3.0-3.5K) are observed when both *air.distance* and *delay* are large. For example, at 11 hPa, Figure 3.14, shows an increase up to 3.6 K at large distances.

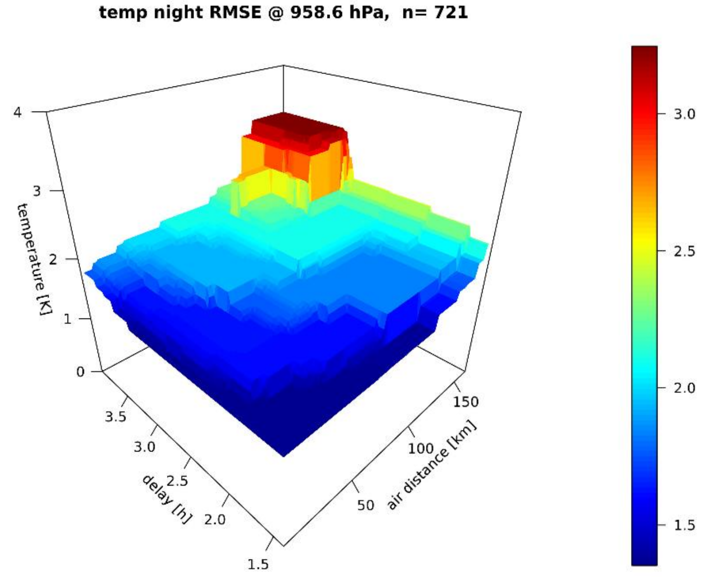


Figure 3.12. Temperature co-location uncertainty, estimated by isotonic regression, nighttime data, at 958.6 hPa.

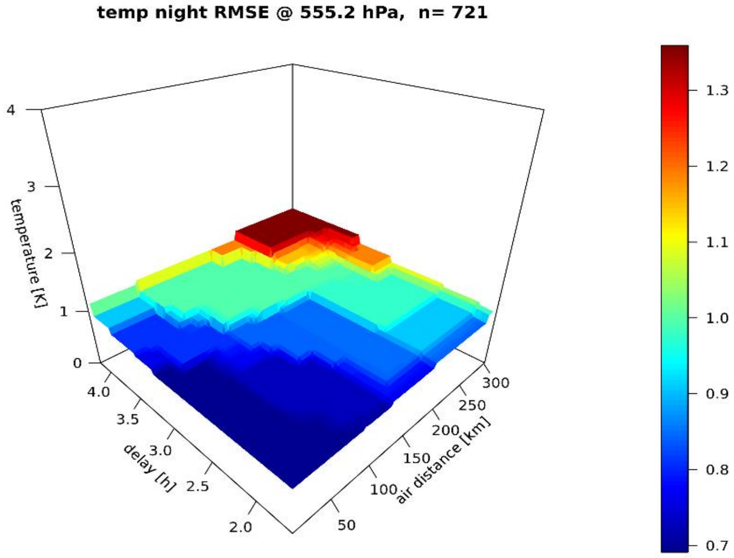


Figure 3.13. Temperature co-location uncertainty, estimated by isotonic regression, nighttime data, at 555.2 hPa.

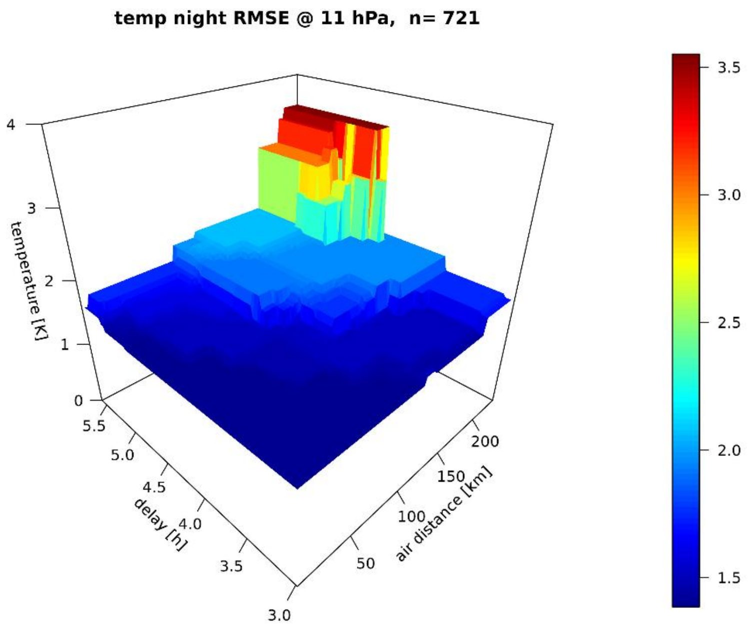


Figure 3.14. Temperature co-location uncertainty, estimated by isotonic regression, nighttime data, at 11 hPa.

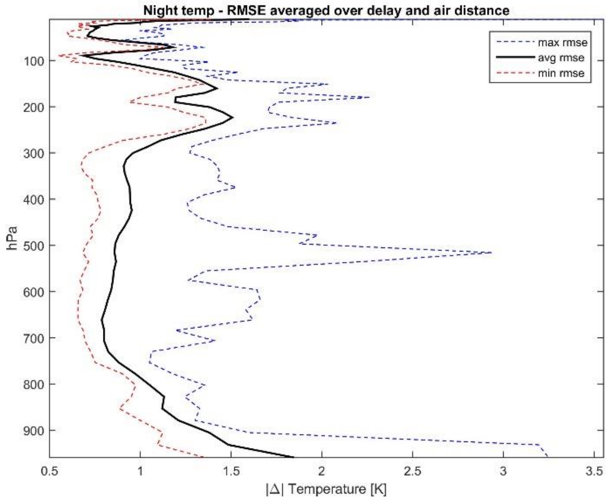


Figure 3.15. Temperature. Minimum, average and maximum uncertainty in central European area.

Nighttime humidity co-location uncertainty

The co-location uncertainty of water vapor mixing ratio (WVMR), estimated using the isotonic regression, is reported in Figure 3.16-3.18 at three different pressure levels for nighttime data. These figures show the variation of co-location uncertainty in the *air.distance*×*delay* domain, from lower to upper troposphere. The vertical profile of uncertainty σ is reported in Figure 3.19, which gives minimum, average and maximum uncertainty with respect to *air.distance* ×*delay*. Overall, this shows that the co-location uncertainty can be considered small in C-EU at all pressure levels, say smaller than 1 g/kg. But large uncertainties (1.5-3.0 g/kg) are observed when both *air.distance* and *delay* are large, in particular around 550 and 720 hPa.

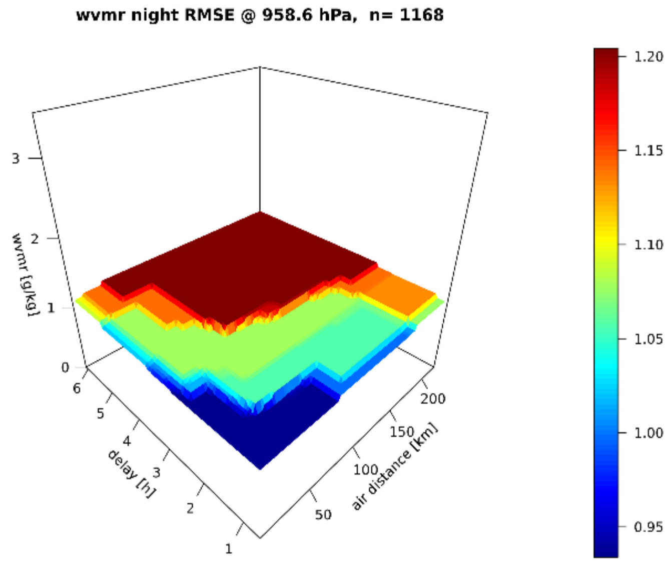


Figure 3.16. Humidity. Co-location uncertainty, estimated by isotonic regression, nighttime data at 958.6 hPa.

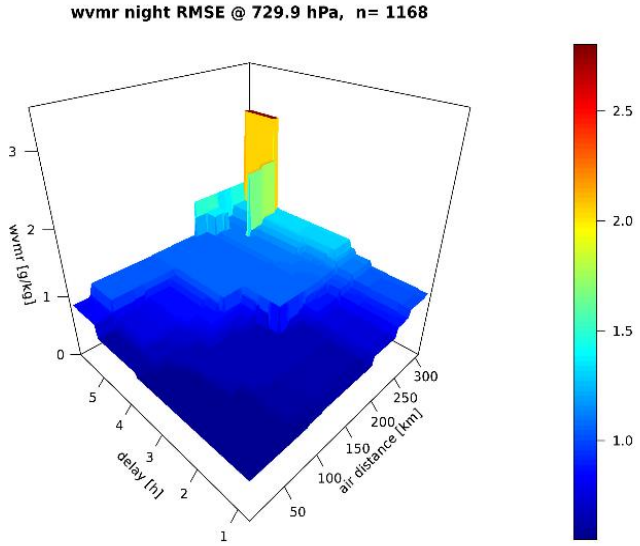


Figure 3.17. Humidity. Co-location uncertainty, estimated by isotonic regression, nighttime data at 729.9 hPa.

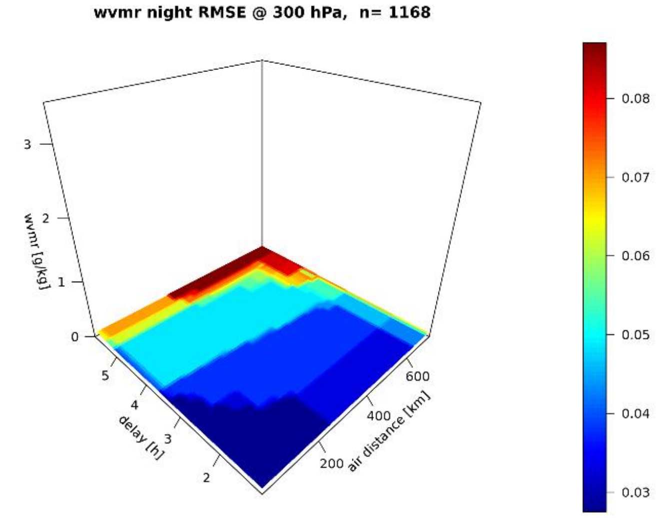


Figure 3.18. Humidity. Co-location uncertainty, estimated by isotonic regression, nighttime data at 300 hPa.

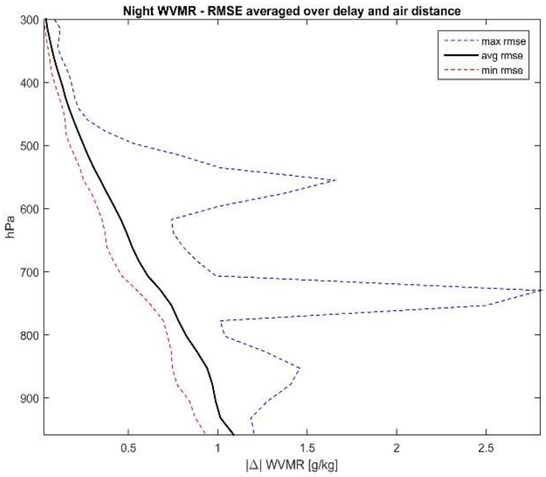


Figure 3.19. Humidity. Minimum, average and maximum uncertainty in central European area.

Nighttime vs daytime uncertainty

The isotonic regression approach has been applied successfully also to daytime data giving the estimated uncertainty surfaces at various pressure levels for temperature and humidity. We omit here the surfaces corresponding to nighttime analysis. Instead, in order to highlight principal similarities and differences, Figures 3.20-3.21 report the differences day-night of average uncertainty and the differences day-night of maximum uncertainty for temperature and WVMR, respectively, in the study area (C-EU). It is seen that average uncertainty is almost the same during day or night, while the difference between day and night maximum uncertainty, that is uncertainty at large spatio-temporal distance, reaches 3.5K and 1.2 g/kg for temperature and humidity respectively. In particular for temperature, the difference of maxima is always smaller than1.2 K in absolute value except between 50 and 30 hPa.

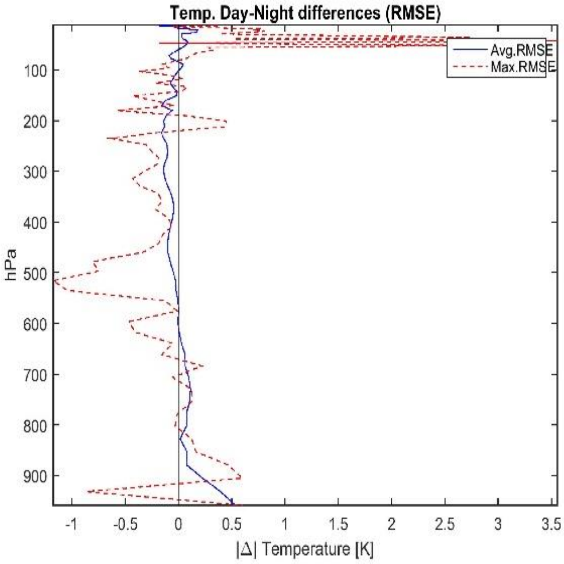


Figure 3.20. Temperature day-night difference. Blue solid line: difference between C-EU averages. Red dashed line: difference between C-EU maxima.

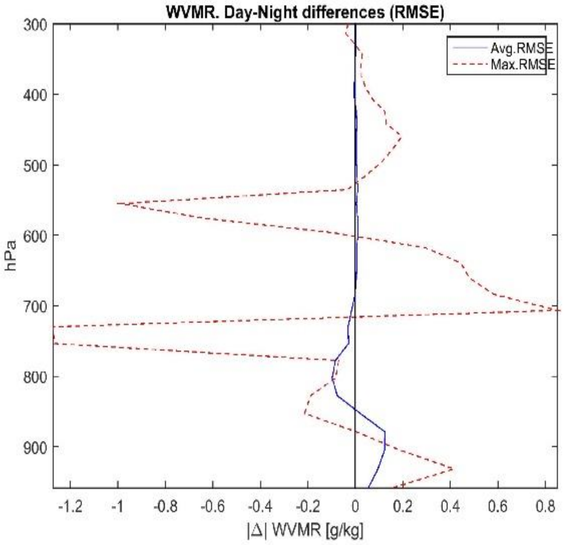


Figure 3.21. Humidity day-night difference. Blue solid line: difference between C-EU averages. Red dashed line: difference between C-EU maxima

3.6 Conclusions and inputs to Virtual Observatory

The comparison of temperature and humidity profiles of IASI and RAOB instruments has been considered in Central Europe (C-EU) in order to understand both the vertical smoothing and the spatio-temporal mismatch of IASI in the comparison with RAOB radiosonde profiles. To do this the information content of RAOB data has been assessed w.r.t. to vertical sampling using GRUAN as a high vertical resolution reference. This resulted in sparseness uncertainty and processing uncertainties which are useful to understand the behavior of total co-location uncertainty.

The assessment of vertical smoothing uncertainty has been based on radiosounding harmonization w.r.t. the strongly vertically smoothed IASI product. The proposed method is based on flexible GEV weight function and has the capability to estimate the vertical smoothing even in the absence of detailed information about averaging kernels used for the satellite retrieval. In fact, in this co-location exercise based on NOAA's NPROVS dataset, the averaging kernels were not available.

Note that, although the algorithm proposed in Section 3.4 considers a time invariant vertical smoothing, in the version under development for the Virtual Observatory of WP5, a dynamic procedure which covers seasonal and daily atmospheric variability will be used. As a consequence, the results of this D3.4 are expected to give an upper limit for the adjusted co-location uncertainty which will be obtained using a time varying harmonization algorithm.

Temperature: Vaisala-GRUAN processing uncertainty is nearly constant and about 0.12 K up to 300 hPa. It increases up to 0.39 K in the lower stratosphere, possibly due to solar radiation. Once Vaisala-GRUAN processing uncertainty is eliminated, vertical sparseness uncertainty of RAOB profiles is smaller than 0.48 K below 300 hPa but increases to about 1 K in lower stratosphere. Vertical smoothing uncertainty, in comparison to total and co-location uncertainties is roughly constant, about 0.5 K, slowly decreasing up to 300 hPa; above this level, it increases more than co-location uncertainty which is more stable. The spatio-temporal mismatch uncertainty is based on the co-location uncertainty adjusted for vertical smoothing and considered as a function of air distance and co-location delay. As a result, overall in C-EU, co-location uncertainty is smaller than 1.5 K, except near ground level, where it reaches 1.8 K. But large uncertainties (3.0-3.3K) are observed when both air distance and delay are large.

Humidity: Vaisala-GRUAN processing uncertainty has an approximately linear decrease from 0.1 g/kg at 925 hPa to about 0.01 g/kg at 300 hPa. Once this Vaisala-GRUAN processing uncertainty is eliminated vertical sparseness uncertainty is smaller than 0.22 g/kg up to 500 hPa and smaller than 0.05 g/kg above. On the other side, considering humidity, total and vertical smoothing uncertainties have a maximum near 800 hPa and above they smoothly decrease. In particular vertical smoothing uncertainty decreases to zero faster than the other two involved uncertainties. Overall, this shows that the co-location uncertainty can be considered small in C-EU at all pressure levels, say smaller than 1 g/kg. But large uncertainties (1.5-3.0 g/kg) are observed when both air distance and delay are large, in particular around 550 and 720 hPa.

Inputs to Virtual Observatory

The "Virtual Observatory" (VO) under development by WP5 will enable end users to the access and use of satellite to non-satellite data comparisons, to explore, interrogate, extract and analyse co-locations between satellite data and high-quality reference and baseline network data. The Virtual Observatory shall be built to showcase potential methods. Nevertheless, the VO design will allow to operate it subsequently as an operational service, though within GAIA-CLIM shall only serve as a proof-of-concept facility.

From the work described above three main outputs will be offered to the end users through the VO: sparseness and processing uncertainties of RAOB products; vertical smoothing uncertainties and harmonized RAOB profiles; and joint spatio-temporal mismatch uncertainty for RAOB-IASI co-locations. These WP3 outputs may be pre-computed and used as inputs to be uploaded off-line to VO.

Regarding Hermite splines, the possibility to implement a "spline engine" in the VO will be considered in D3.5 and evaluated with respect to the alternative to upload to VO pre-computed high resolution spline outputs.

Sparseness and processing: This kind of uncertainty analysis is available only for GRUAN sites. The outputs will be in the form of uncertainty profiles to be attached off-line to the relevant RAOB profiles and made available to the end users of the VO.

Vertical smoothing: The outputs of the above vertical smoothing uncertainty analysis are in the form of harmonized profiles, which can be off-line uploaded to VO. Similarly, the adjusted co-location uncertainties should be attached off-line to the IASI-RAOB co-locations.

Spatio-temporal mismatch: The co-location uncertainty has been estimated for temperature and humidity as a function of air distance and delay, at pressure levels available for IASI.
Hence, 3D contours plot of the uncertainty will be made available through the VO, either as numerical look-up tables or graphical representations. As a result, this tool allows the user to retrieve the co-location uncertainty for IASI temperature and humidity profiles. Moreover, given an uncertainty threshold u_0, which may rise from a specific end-user application, it will be easy to check if a specified co-location (x_G, x_S), at spatio-temporal distance $(\partial a, \partial t)$, satisfies the uncertainty requisite in the form $\sigma(\partial a, \partial t) < u_0$. More interestingly, this approach allows a co-location system to select all co-locations satisfying an uncertainty requirement u_0 fixed by the user. Notice that, this can be done either for a fixed uncertainty threshold u_0 or for an uncertainty threshold profile $u_0(p)$.

4 Temperature and humidity profiles: radiosonde temporal mismatch

The goal of this activity is to enable estimation of the co-location uncertainty in non-simultaneous temperature measurements due to atmospheric variability, both in terms of the systematic bias between measurements and the random noise. The availability of such information provides an estimate of temperature mismatch uncertainty for any given site using global reanalysis data, and would enable an appropriate sampling strategy to be put in place for a given intercomparison measurement application in order to meet a defined comparison uncertainty requirement.

Previous work (Butterfield and Gardiner, 2015) established that 4 launches per day (or 6 hourly updates) provided a reasonable measure of the variability during the day. It also provided estimates from actual sonde datasets, as a function of time of day, altitude and season – but only available for the very limited number of sites that provide long-term data at the required high level of sonde launch frequency. The objective of the work within GAIA-CLIM is to extend this analysis, through the use of ECMWF re-analysis model data, to provide a globally-applicable co-location uncertainty assessment tool. This work has initially focused on temperature measurements, and is being extended to humidity.

4.1. Methodology

Radiosonde temperature readings and modelled temperatures are processed in the same way by amalgamating into altitude bins 500m high, labelled as the centre of each bin, i.e. 0 to 500m labelled as 250m. As the ERA-Interim altitude data is represented as local altitude it is first corrected to height above sea level so it can be directly compared with the radiosonde data. The temperatures in each altitude bin are averaged to provide a mean temperature, *T*, for that specific altitude. The rate of change in temperature between single launches 6 hours apart, at each altitude, were calculated according to Eqn (4.1). The mean rate of change in temperature between each launch separation and altitude, $\frac{dT}{dt_n}$, for a defined group of launches were then calculated according to Eqn (4.2).

$$\frac{dT}{dt_n} = \frac{T_n - T_{n-1}}{t_n - t_{n-1}} \qquad (4.1)$$

Where t_n is the launch time for the sonde.

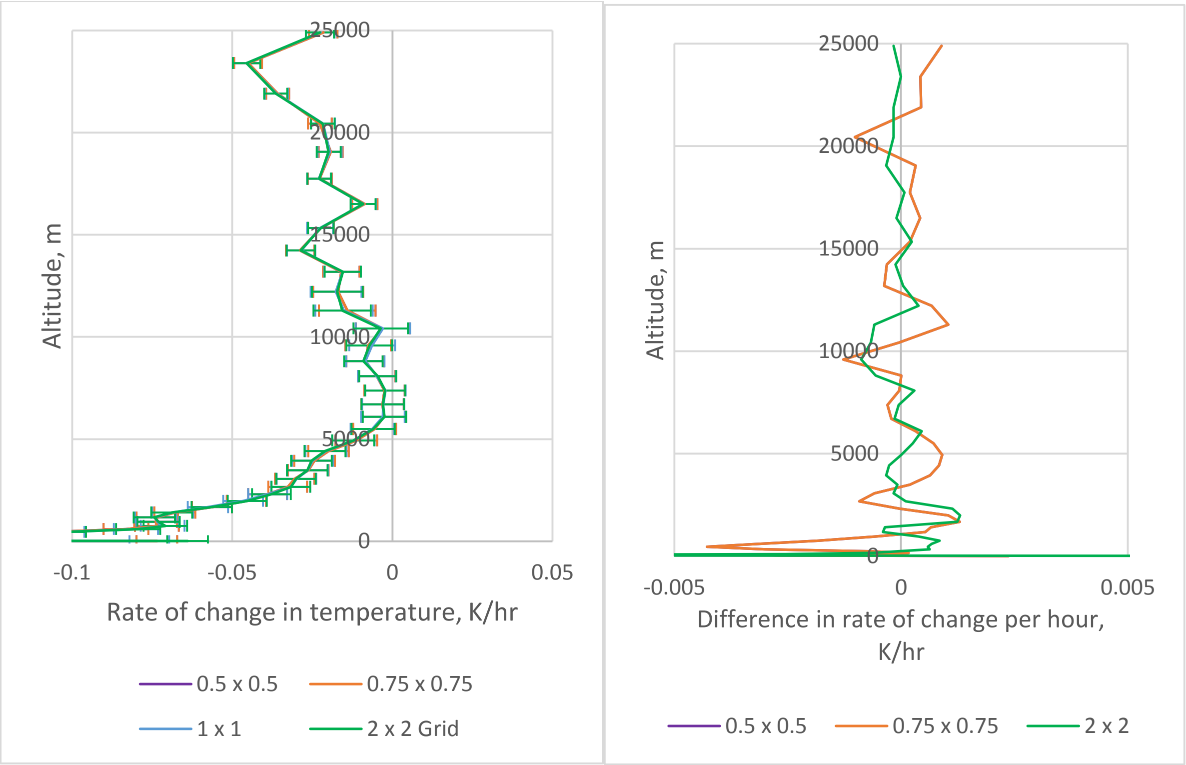
$$\overline{\frac{dT}{dt_n}} = \frac{\sum \frac{dT}{dt_n}}{i} \qquad (4.2)$$

Where *i* = the number of launch pairs.

The result of this calculation provides the mean rate of change of temperature as a function of altitude for launches grouped by time of day and season across the datasets being investigated, with the standard error on the mean providing a measure of the random uncertainty resulting from the variability observed for that particular grouping.

4.2. Comparison between radiosonde and ERA-Interim reanalysis

The spatial resolution of the ERA-Interim data set is approximately 80 km on 60 vertical levels from the surface up to 0.1 hPa. As the ERA-Interim model is available at different spatial grid resolutions, a comparison was made between temperature data produced at grid resolutions of: 0.5^{o} x 0.5^{o}, 0.75^{o} x 0.75^{o}, 1.0^{o} x 1.0^{o} and 2.0^{o} x 2.0^{o}, corresponding to the Lindenberg radiosonde launch site for the period 2009 to 2012. Figure 4.1 shows an example of this comparison. For altitudes above 500m the differences in the mean rate of change in temperature for each grid size was below 0.005 K per hour when compared to the 1^{o} x 1^{o} result. Based on this it was decided that a grid size of 1^{o} would be used for all future data processing.



Note: +ve result means that result is larger than the 1° x 1° result.

Figure 4.1. (a) Lindenberg, mean rate of change in temperature between 01LT and 07LT (b) Difference between each grid size and the 1o x 1o result for the 01LT – 07LT.

Comparisons between ERA-Interim (1° x 1° grid) and radiosonde temperature stability data were then made for 3 long-term radiosonde data sets: Lindenberg 1999 – 2008 (Vaisala RS90), Lindenberg 2009 – 2014 (Vaisala RS92-SGP), Southern Great Plains 2006 – 2014 (Vaisala RS92) and 1 short-term data set: Manus Island 24th September 2011 to 31st March 2012 (Vaisala RS92-SGP).

As only 2 GRUAN sites regularly launch radiosondes on a 6 hour frequency, the global extent of the investigation was extended by analysing the temperature stability of the ERA-Interim data at 12 GRUAN site locations as detailed in Table 4.1 below, over the same period as when GRUAN data is available.

Location	**Latitude**	**Longitude**	**Elevation, m above sea level**	**Period**
Ny Alesund	78.92 N	11.92 E	5	01/01/11 to 31/12/14
Barrow	71.32 N	156.61 W	8	01/01/09 to 31/12/14
Sodankyla	67.37 N	26.63 E	179	10/09/10 to 31/12/14
Lindenberg	52.21 N	14.12 E	98	01/01/09 to 31/12/14

Lauder	45.05 S	169.68 E	370	11/12/12 to 31/12/14
Potenza	40.60 N	15.72 E	720	22/02/11 to 31/12/14
Boulder	39.95 N	105.20 W	1743	15/06/11 to 31/12/14
SGP	36.60 N	97.49 W	320	01/01/06 to 31/12/14
Tateno	36.06 N	140.13 E	31	01/01/10 to 31/12/14
La Reunion	21.00 S	55.00 E	2200	01/01/10 to 31/12/14
Darwin	12.43 S	130.89 E	35	01/01/10 to 31/12/14
Manus	2.06 S	147.42 E	6	01/01/11 to 31/12/14

Table 4.1. Location of GRUAN sites selected for ERA-Interim Analysis.

The standard error of the mean can be taken as an estimation of the uncertainty of the mean value as it quantifies both the scatter in the data and the sample size. If the data points are normally distributed the standard error of the mean will reduce as the sample size increases, at the rate of the square root of the sample size. To test if this is really the case for the distribution of rates of change in temperature for RS-92 radiosonde and ERA-Interim data sets, the data sets were randomly sampled over variable sample sizes. The standard error was calculated from the resampled standard deviations and compared to the theoretical standard error for the different sample sizes. Figure 4.2 shows an example of the comparison of the resampled standard error compared to the theoretical standard error – in this case for the two ECMWF sites with the lowest and highest standard deviation in mean rate of change in temperature at an altitude of 15 km. The excellent agreement between the resampled and theoretical results confirmed that the data behaves as a set of stationary variables.

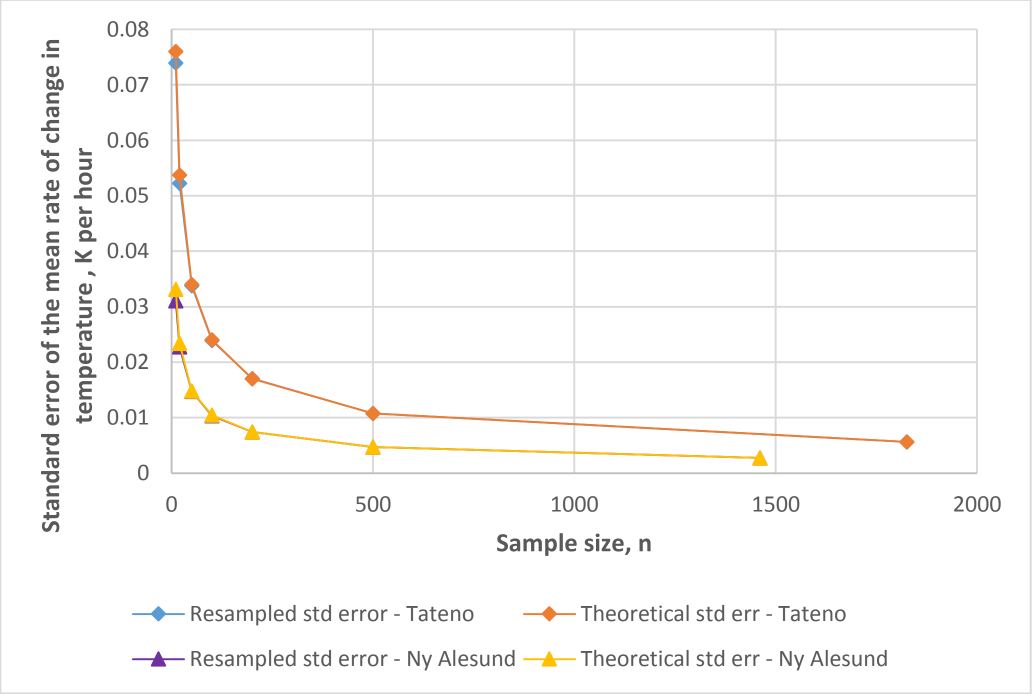


Figure 4.2. Resampled standard error and theoreical standard error of the mean rate of change in temperature for Tateno and Ny Alesund ECMWF sites at an altitude of 15,000m.

The variability in temperature stability data at 6 altitudes (3250m, 5250m, 8250m, 12250m, 15250 and 20250m) was then determined for both ERA-Interim and radiosonde data sets by repeatedly sampling the data with different sample sizes to determine how many data points would be needed to achieve a particular level of mismatch uncertainty, and therefore what would be needed for a valid radiosonde / model / satellite overpass intercomparison with a given uncertainty specification. Figure 4.3 provides examples of these analyses, showing the standard deviation of rate in change in temperature at an altitude of 5.25 km for (a) the radiosonde and ERA-Interim data for the 3 long-term data sets, and (b) the ERA-Interim data for the 12 selected GRUAN sites.

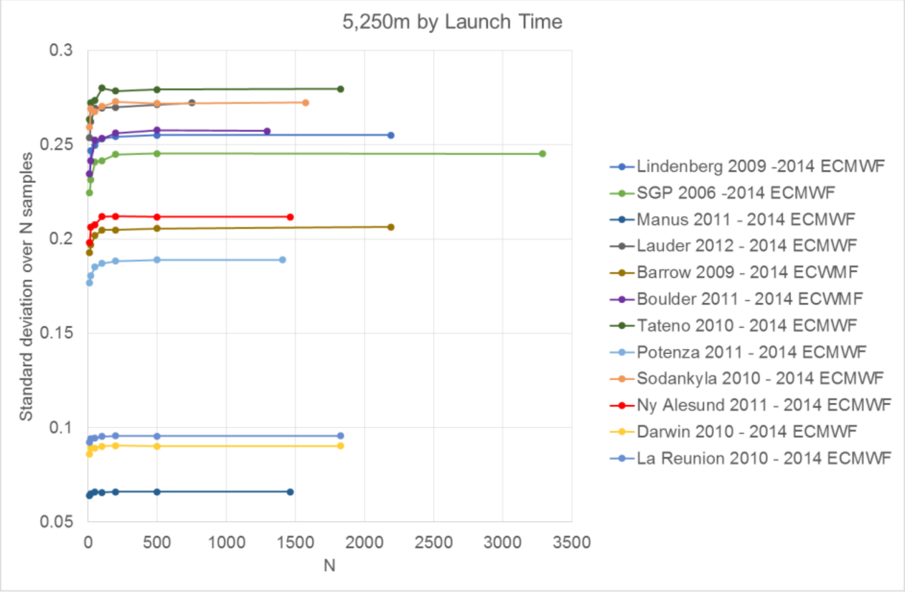


Figure 4.3. (a) Standard deviation of rate in change in temperature of radiosonde and ERA-Interim data for the 3 long-term data sets. (b) Standard deviation of rate of change in temperature of ERA-Interim data for 12 GRUAN sites.

The results shown in Figure 4.3a above show clear differences in the random component of the difference between the radiosonde and ERA-Interim data. Differences are also seen in the systematic value of the rate of change in temperature. Such differences are to be expected as the derived mismatch values are a combination of the underlying atmospheric variability and the uncertainty due to the particular measurement (or model) technique. However, if the differences between the two methods lie within the expected uncertainty bounds of the methods, then this gives us confidence that they are in agreement, and providing a reliable estimate of the underlying atmospheric variability.

A detailed study of the uncertainties in a GRUAN-processed RS-92 radiosonde profile is presented in Dirksen et al, 2014. There will also be uncertainties in the ERA-Interim results, both from the meterological model itself, and from the spatial representativeness of the model results when compared to the small spatial sample from the sonde results. These uncertainties are harder to quantify. However, if there is agreement between the two methods at, or close to, the level of the sonde uncertainties, then this confirms the comparability of the methods.

As indicated above there are two elements to the uncertainty analysis – the random (uncorrelated) component given by the standard deviation of a large number of samples (e.g. Figure 4.3) that will result in an uncertainty that reduces with the number of measurements made (see Figure 4.2), and a systematic (correlated) component that is the difference in the underlying rate of change in temperature. Dirksen et al, 2014 separate the sonde measurement uncertainty sources into correlated and uncorrelated contributions. However, in their work correlated refers to correlation within one sonde ascent, and this study addresses the difference between sequential sonde launches, so only those uncertainty sources that remain correlated between launches would contribute to the systematic element. As a result the majority of the uncertainty described by Dirksen et al would contribute to the random component in this analysis.

Table 4.2 shows the random component of uncertainty for the sonde and ERA-Interim results for Linbenberg (combined across seasons and launch times), and compares the difference between them to the predicted uncertainty in the difference between two sonde measurements based on Dirksen et al. These results are given separately for the mid-troposphere (Trop) and upper-troposphere/lower stratosphere (UTLS), as the uncertainty behavior is different in these regions.

	Trop	UTLS
Sonde	0.283 K/hr	0.202 K/hr
ERA-Interim	0.252 K/hr	0.148 K/hr

Sonde uncertainty	0.035 K/hr	0.058 K/hr

Table 4.2. Random components of uncertainty in temperature differences.

The random component is consistently higher for the sonde measurements compared to the ERA-Interim model, and the differences are of the order of the predicted sonde measurement uncertainties. This result, together with the expected differences due to the higher spatial resolution and therefore spatial variability of the sonde measurements, confirms the general agreement between the sonde and ERA-Interim results and indicates that the differences in the random component of the uncertainty are consistent with the expected uncertainties in the data sets.

The magnitude of the systematic component of the temperature difference at Lindenberg is typically ~0.021 K/hr for the sonde results and ~0.014 K/hr for the ERA-Interim data, with the difference between them consistently less than 0.01 K/hr (with similar results in Trop and UTLS regions). This level of agreement is within that expected for the likely sources of correlated uncertainties in the methods.

4.3. Humidity data analysis

Following the successful development of the mismatch analysis tool for temperature the scope of the GAIA-CLIM activity has been extended to cover humidity. A similar set of analyses is therefore being undertaken for radiosonde and reanalysis humidity data. Note that, in order to directly compare the measured and modelled results it is necessary to convert the specific humidity provided in the ERA-Interim data to the relative humidity over water, as reported by the radiosondes. Figure 4.4 shows an example of the comparison between measured and modelled humidity differences, in this case for the Spring data in the GRUAN-processed Lindenberg radiosonde dataset. The red lines shows the rate of change in humidity (%RH per hour) over the indicated time period as a function of altitude (up to 25 km) for the sonde data, while the black lines shows the equivalent assessment for the ERA-Interim data.

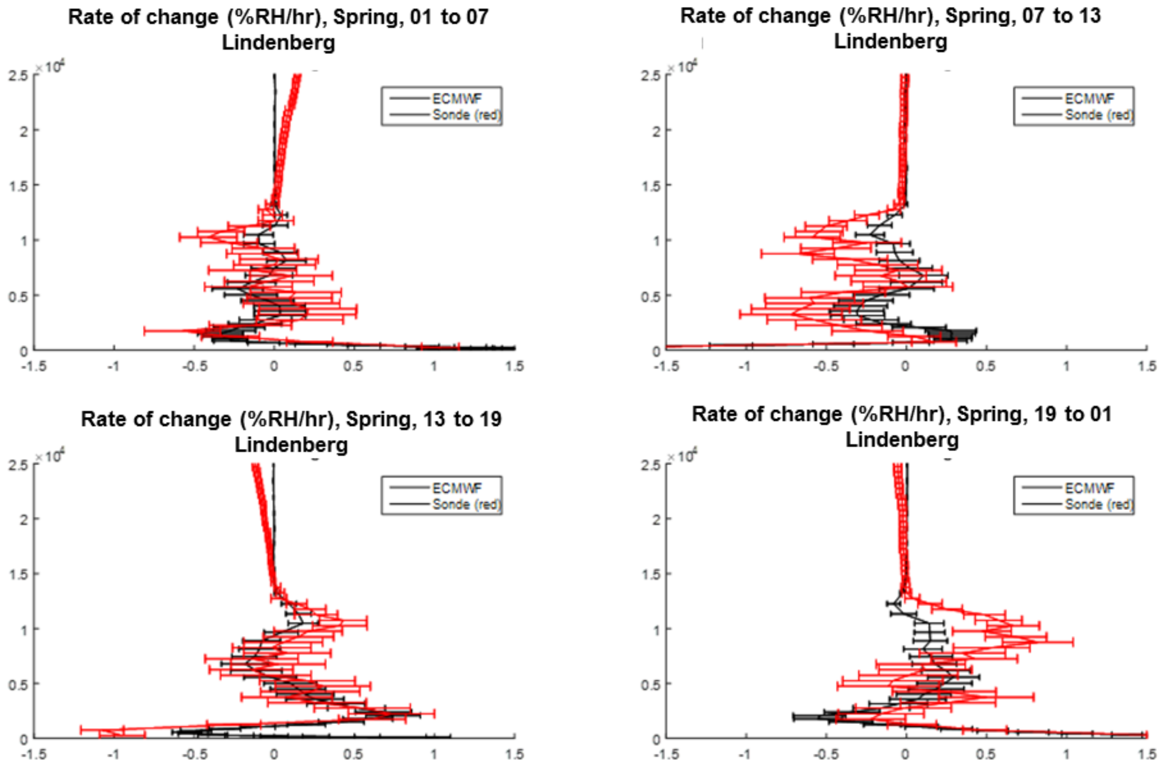


Figure 4.4. Comparison in the rate of change in relative humidity (% per hour) between the measured (red) and modelled (black) profiles at Lindenberg, shown for different times of day in Spring. (top left panel: 01:00 to 07:00, top right panel 07:00 to 13:00, bottom left panel: 13:00 to 19:00, bottom right panel: 19:00 to 01:00). Error bars show the standard uncertainty on the average rate of change.

A similar exercise with the SGP data set shows the improvement that is gained by using the GRUAN-processed humidity data rather than the standard sonde humidity data product. Figure 4.5 shows two examples of the rate of change profiles (upper: Spring, 01:00 to 07:00; lower: Spring 13:00 to 19:00) for the standard data product (left panels) and the GRUAN-processed product (right panels).

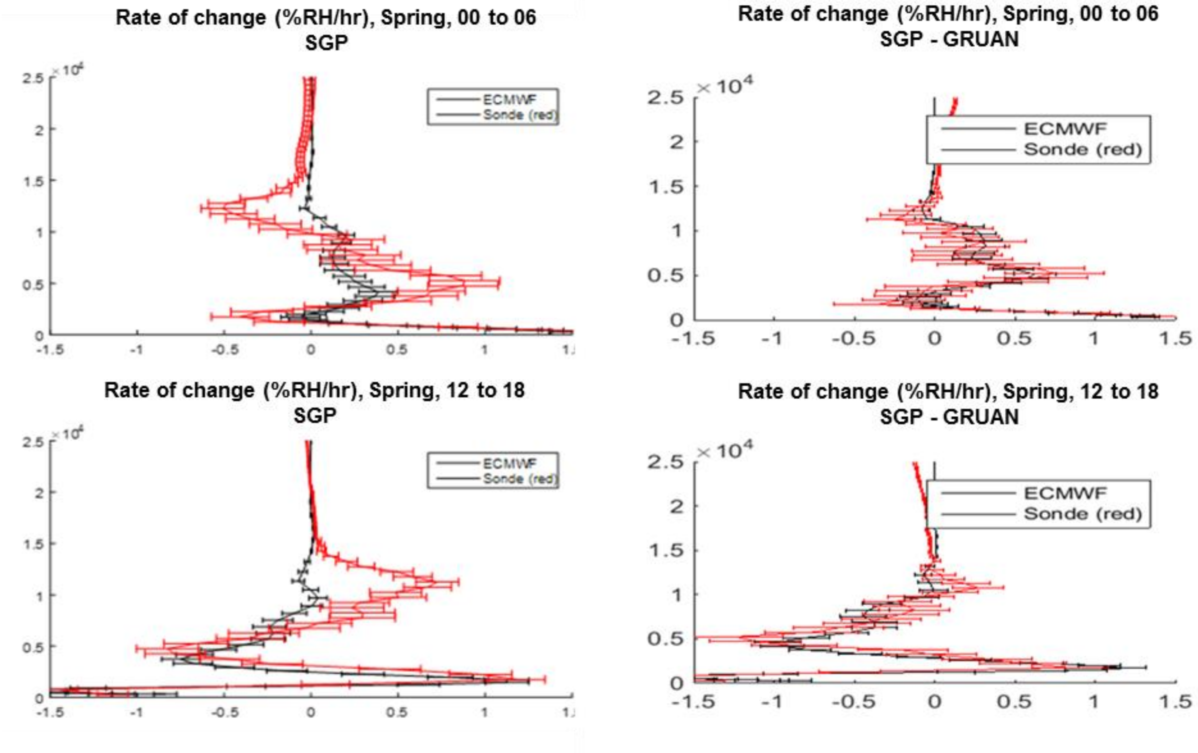


Figure 4.5. Comparison in the rate of change in relative humidity (% per hour) between the measured (red) and modelled (black) profiles at SGP for standard sonde product (left panels) and GRUAN sonde product (right panels), shown for different times of day in Spring. Error bars show the standard uncertainty on the average rate of change.

The level of agreement between the GRUAN-processed sonde and ERA-Interim results in figures 4.4 and 4.5 indicates that the overall method for deriving mismatch uncertainties can be applied to humidity as well as temperature, and a similar uncertainty analysis on the humidity differences is now underway to confirm this.

4.4. Conclusions

The work on temporal mismatch uncertainties in radiosonde profiles provides a direct means of estimating the temporal mismatch uncertainty in temperature as a function of altitude, season and time of day, for the selected sites using ERA Interim model data. This approach has been validated by comparing the ERA Interim results with GRUAN-processed radiosonde data from those sites where long-term high frequency data is available. Analysis of the results have shown that the differences between the radiosonde and ERA-Interim temperature data can be explained by the expected uncertainties in the different data sources. Similar studies are underway for the humidity data, and this has already demonstrated the improved agreement between measured and modelled differences when the GRUAN-processed humidity data is used in the analysis. A paper summarising the overall results of this study is in preparation. This work has been based on the available reference data with traceable uncertainties i.e. the GRUAN RS92 sonde dataset. As additional reference datasets and uncertainties become available they can be integrated into this analysis. This will be particularly relevant with higher temporal resolution data

5 Ozone profiles

Vertical ozone profiles are measured from the ground with a variety of techniques, covering both in-situ measurements (balloon-borne ozonesondes) and active (LIDAR) and passive (MWR, FTIR) remote sensing. Each technique has its own distinct properties in terms of vertical coverage and resolution, temporal coverage and resolution, measurement uncertainty etc. While a large body of literature exists on the vertical properties of these measurements, we present here new results on the impact of the horizontal and temporal smoothing properties, in particular in the context of using these data for the validation of satellite measurements. More specifically, we look here at the impact of (1) balloon drift on ozonesonde measurements, (2) the low elevation angle of the line-of-sight used in MWR measurements, and (3) the significant integration times required for precise O_3 profile measurements with LIDAR systems. The method used is based on a sophisticated simulation of measurements and observing systems, including their full 4-D spatio-temporal sampling and smoothing characteristics, as implemented in the OSSSMOSE system developed at BIRA-IASB.

Integration in the Virtual Observatory

The Virtual Observatory developed in WP5 of GAIA-CLIM will offer access to O_3 profile data from ozonesondes, LIDAR instruments, and FTIR instruments contributing to the NDACC. Besides a focus on the “reference” quality of the measurement in terms of traceability and uncertainty quantification, part of the aim is also to provide the user with estimates of the additional uncertainties due to spatio-temporal features such as the balloon drift and temporal integration described here. Since performing OSSSMOSE simulations within the VO environment is beyond the scope of the current project, the results described below for representative case studies will be extended to all NDACC-contributing stations and converted into look-up tables that can easily be used within the VO. More specifically, for each station a look up table will be constructed that contains altitude-resolved mean (i.e. “climatological”, as derived from multiple years of data) spatio-temporal smoothing and sampling errors and their spread. Further details and illustrations of the use of these look up tables will be provided in D3.5.

5.1 Ozonesondes: balloon drift

Ozonesondes and the corresponding long-term data records play a fundamental role in the monitoring of the vertical ozone profile throughout the troposphere and into the lower stratosphere, addressing scientific needs regarding stratospheric ozone recovery, tropospheric ozone pollution, and ozone as a climate forcing agent. Moreover, they provide crucial in-situ reference data for the ground-based validation of satellite ozone profilers (e.g. Keppens et al., 2015, and Hubert et al., 2016). While great effort is being invested in the homogenization of the ozonesonde network and in controlling and minimizing the measurement uncertainties, e.g. within the O3S-DQA, the impact of (horizontal) balloon drift on the representativeness of the measurements has hitherto remained uninvestigated.

In this section, we address this gap with a quantification of the errors and uncertainties due to balloon drift using an Observing System Simulation Experiment (OSSE) performed with the OSSSMOSE system and using the latest generation of global ozone reanalyses: MERRA-2 and the MACC (now CAMS) IFS-MOZART reanalysis. We quantify how drift-related errors depend on altitude, season, particular atmospheric regimes, etc. In many cases, these errors exceed the (random) measurement uncertainties. Finally, we put these results in context of ozonesondes as a source of reference data for satellite validation, and how balloon drift can/must be taken into account in order to optimize their fitness-for-purpose for the upcoming Copernicus validation work.

Balloon drift statistics as derived from radiosonde networks

Balloon drift has hitherto not been studied on a global scale for ozonesondes. However, since the weather balloons used for ozonesonde launches are similar to those used in the extensive radiosonde networks, a first estimate of drift statistics can be obtained from published statistical analyses on those radiosondes. Indeed, while the payload for an ozonesonde is significantly heavier than that of a regular radiosonde, this is compensated by a larger balloon in order to ensure similar ascent rates (R. Van Malderen, Belgian Met Office, KMI, private communication). Assuming a strong coupling between horizontal wind components and the balloon movement, this should also lead to similar balloon drift trajectories. Seidel et al. (2011) present such an analysis of radiosonde drift trajectories and their figure summarizing the median drift distance (at the highest level, i.e. just before balloon burst) during northern winter months is shown in Figure 5.1. As is clear from this graph, drift distances vary considerably across the globe, from less than 25km in northern Brazil up to 200km and more near Japan in winter.

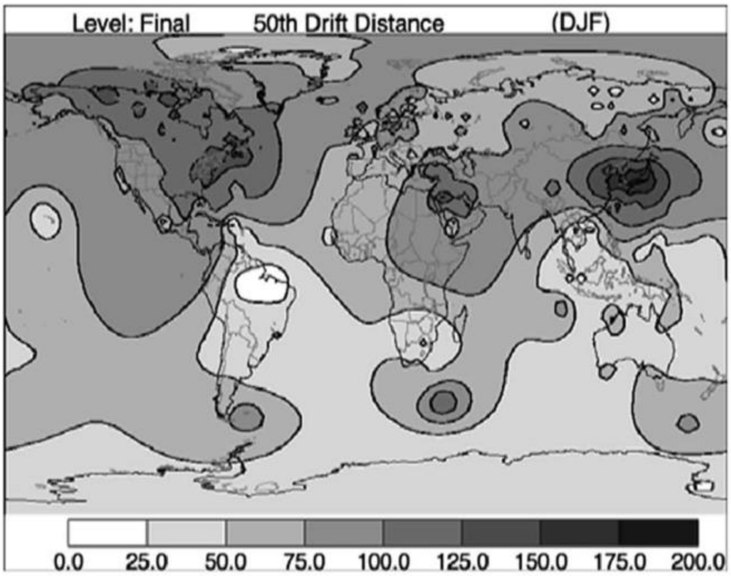


Figure 5.1. Median drift distance at the highest level reached by the balloons of the global radiosonde network, during northern winter months (December, January, and February), as derived by Seidel et al. (2011).

To what extent this balloon drift leads to an ozone concentration measurement that differs significantly from that directly above the station depends not only on the drift distance, but also on the local variability of the ozone field. To gauge this variability across the globe, we looked at the variability in global reanalysis fields at a scale of a few hundred kilometres. Such a map, corresponding to the variability in the UT/LS ozone column for northern winter months, is shown in Figure 5.2.

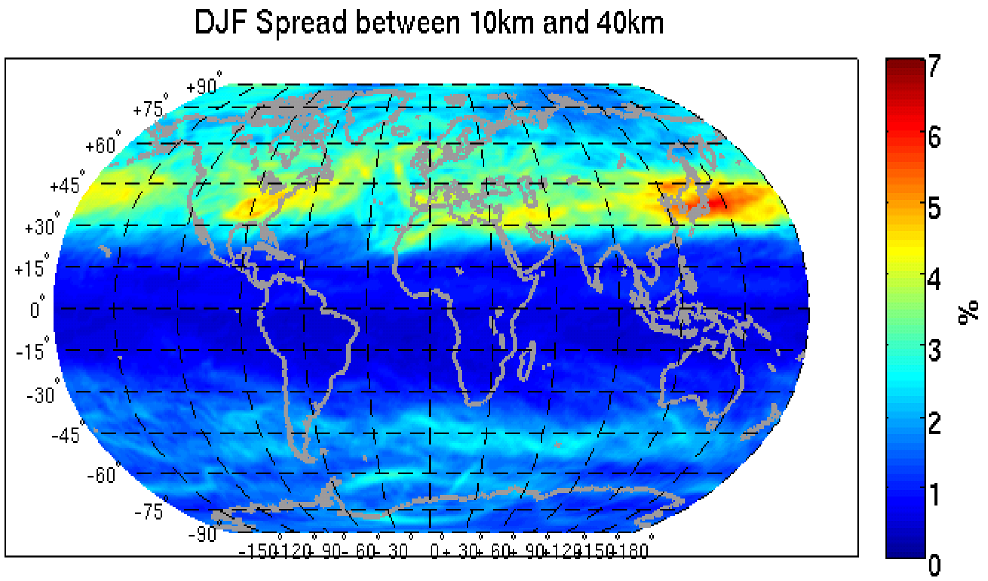


Figure 5.2. Variability in the UT/LS ozone column during northern winter months (December, January, and February) as caught by the IFS-MOZART reanalysis performed for MACC (now CAMS).

Combining this graph with Figure 5.1, we can expect very little impact due to balloon drift on the ozone measurements in equatorial regions, in particular in South America, while Japan shows both large drift distances and strong spatial variability in the ozone field, which will most likely lead to large differences between the measured ozone profile and the vertical profile at a (hypothetical) launch station. It should be kept in mind though that the variability in the ozone field is mostly due to advection and thus correlated with the local wind field. As such, the actual impact on the representativity of the measurements may be more moderate than estimated from this qualitative exercise.

Another aspect, not yet touch upon, is the time it takes the balloon to climb up to its burst point. This duration is typically of the order of 60-90 minutes which may in some cases also introduce significant errors w.r.t. the ozone field at the time of the sonde launch.

The above considerations and concerns warrant a more in-depth quantification of the impact of balloon drift and ascent duration on the measured ozone profile. In the next section, we present an OSSE-based calculation of the errors and uncertainties that result from the balloon trajectory that takes into account the actual interplay between balloon trajectory and local ozone field.

OSSSMOSE simulation of ozonesonde balloon drift errors

To arrive at a detailed, quantified estimate of the difference between the profile measured by an ozonesonde and that above the launch station, we used the OSSSMOSE system (Verhoelst et al., 2015) to trace actual ozonesonde trajectories in 4-D reanalysis fields and extract the corresponding simulated “measured” profile, which can then be compared to the reanalysis profile above the launch station. This is illustrated in Figure 5.3.

In essence, simulating the measured profile comes down to interpolating the reanalysis on the 4-D (lat,long,altitude, time) coordinate of every measurement during an ozonesonde ascent. For every simulated ascent, this allows us to calculate the drift "errors", i.e. the differences w.r.t the vertical profile at the launch time. By analyzing a large set of such difference profiles, e.g. covering several years of actual ozonesondes launches at a monitoring station, we can deduce both random and systematic statistical properties of these drift errors, as a function of altitude, season and particular atmospheric phenomena.

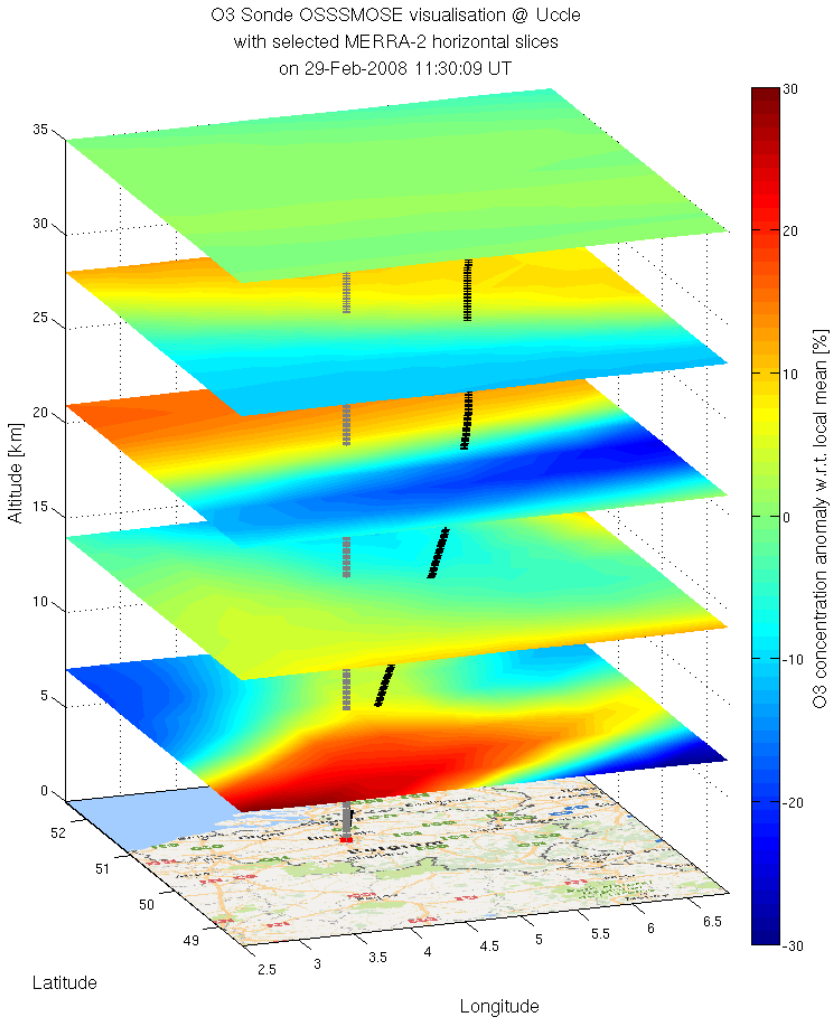


Figure 5.3: Illustration of the difference between the station zenith(grey markers) and the actual trajectory of an ozonesonde launched by the Belgian Met Office (KMI) from Uccle (Brussels) on 29 February 2008. Also a few slices of the corresponding MERRA-2 ozone reanalysis are shown to illustrate the spatial inhomogeneity of the ozone field at these scales and altitudes.

Based on the information in Figures 5.1-6.2 and the locations of the NDACC stations with well-sampled time series of ozonesonde launches, two distinct case studies were chosen here: Paramaribo, Suriname (5.75˚N, 55.2˚W, weekly launches by KNMI) as representative for small drift distances and low ozone field variability, and Hohenpeißenberg, Germany (47.8˚N, 11.0˚W, 3 launches per week by DWD) as representative for larger drift distances and significant ozone field variability. Results for 4 years of launches (2008-2011) at these 2 stations are shown in Figures 5.4-6.5. Averaged over these years, no large systematic errors are observed and the maximum uncertainty is of the order of 2% and 4% respectively (in the upper troposphere). In fact, the maximum spread in the upper troposphere is due to the combination of atmospheric variability and drift distance being largest at these altitudes. Going higher up, in the lower stratosphere, the drift distances are even larger, but the ozone field becomes much smoother and consequently the errors due to the drift decrease. In both cases, the probability density distribution (PDF) of the differences is fairly Gaussian, with non-negligible tails up to differences of 10% for launches at Hohenpeißenberg.

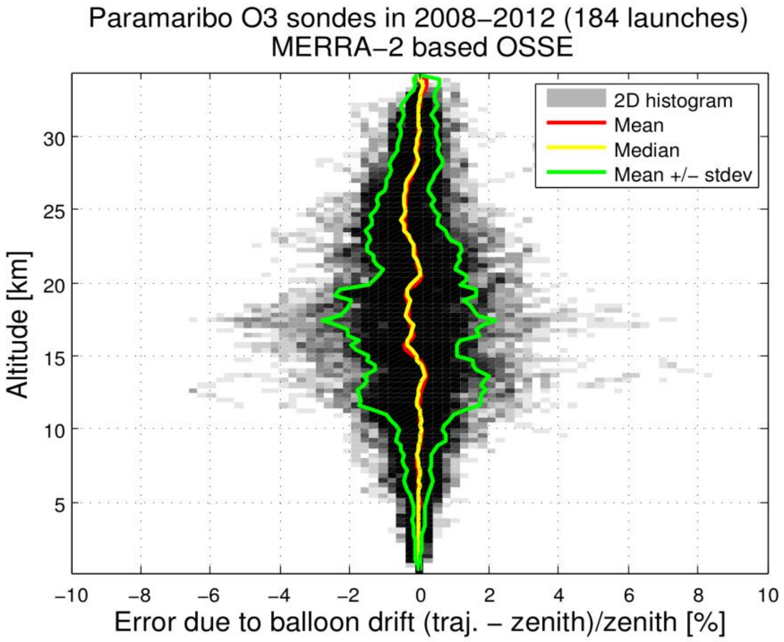


Figure 5.4: OSSSMOSE simulation of errors in ozone concentrations due to balloon drift and elapsed time since launch, at the equatorial NDACC station of Paramaribo, Suriname (5.75˚N, 55.2˚W, weekly launches by KNMI.

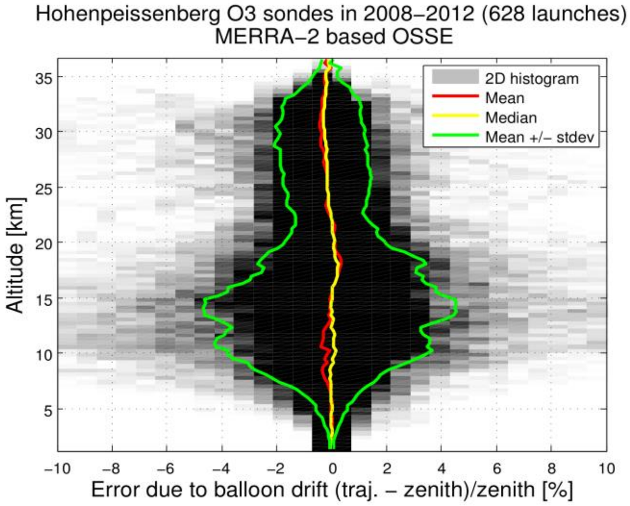


Figure 5.5: OSSSMOSE simulation of errors in ozone concentrations due to balloon drift and elapsed time since launch, at the northern mid-latitude NDACC station of Hohenpeißenberg, Germany (47.8˚N, 11.0˚W, 3 launches per week by DWD).

Implications, conclusions, and caveats

The two cases studies detailed in the previous section illustrate that the errors due to balloon drift depend strongly on launch location. When the drift distances are small and the ozone field relatively stable, e.g. at Paramaribo, the simulated errors remain below typical measurement uncertainties, which are estimated to be of the order of 3-5%. Indeed, when standard operating procedures are followed, the three most commonly used sonde types[1] produce consistent results between the tropopause and 28 km, with biases smaller than 5% and precisions better than 3% (Smit and ASOPOS panel, 2014). At higher and lower altitudes the data quality degrades somewhat, and the differences between the sonde types become more clear. Overall, ECC-type sondes perform best with a bias of 5–7% and a precision of 3–5% in the troposphere. Consequently, in this case, the errors due to balloon drift also remain below the accuracy targets for the use of ozonesondes as reference data in satellite validation work. On the other hand, when drift distances are >100km and the ozone field is inhomogeneous, the errors do become comparable to the measurement uncertainty and will contribute to the error budget of a comparison with other measurements when not minimized, e.g. by doing a height-dependent co-location, or taken into account by adding an additional uncertainty term in the consistency test (σ in Eq. 1.1).

It is important to be aware of the limitations of the OSSSMOSE simulations behind these results, in particular the horizontal and temporal resolution of the reanalysis fields that were used (typically of the order of 1°x1° and 3 hours respectively). Some preliminary comparison between the variability caught by the reanalysis and that observed in in-situ aircraft data by Sparling et al. (2006) suggests that the intra-pixel variability can not be neglected, and the simulations here are to be taken as lower limits on the actual drift errors. In view of this limitation, it is recommended to address balloon drift in a validation exercise by doing an altitude-dependent co-location, rather than by including an additional uncertainty term in the comparison uncertainty budget. For satellite sounders with high horizontal but poor vertical resolution, some satellite pixel averaging may be required when the balloon traverses multiple satellite pixels in a single vertical layer.

5.2 Ozone MWR: elevation angle of the line-of-sight

Because of its low sensitivity to weather conditions and aerosol load, ground-based millimeter radiometry is well suited for observations of stratospheric and mesospheric ozone. The technique was first explored in the early 80s (e.g. Wilson & Schwarz, 1981) and nowadays six micro-wave radiometers (MWR) routinely contribute O_3 profile measurements to the NDACC, with a further 3 instruments in the "candidate" phase. These instruments measure the emission of a thermally excited rotational transition at either 110 or 142 GHz. From these line spectra, O_3 volume mixing ratios are derived using an optimal estimation retrieval technique, relying on the pressure broadening of the line to disentangle different altitudes (i.e. different atmospheric pressures). A valuable proxy of the measurement sensitivity and vertical resolution is provided by the averaging kernel (AK). An example

---

[1] Nowadays more than 80% of the stations launch an electrochemical concentration cell (ECC) sonde (Komhyr, 1969). The Brewer–Mast sonde has mostly been used by the early sounding stations with long data records (Brewer and Milford, 1960), while the Japanese stations fly a carbon iodine cell sonde (Kobayashi and Toyama, 1966).

AK, corresponding to a night-time measurement at the NDACC station of Mauna Loa, Hawai'i (19.54°N, 155.58°W, 3397m a.s.l.), is shown in Figure 5.6. While the profile is sampled on a relatively fine grid (49 levels between 12km and 108km) the actual vertical resolution (taken here to be the full-width-at-half-max of the AK rows) ranges from a best of 6km at approx. 30km to 14km and worse above 65km.

One of the key applications of MWR measurements is the study of the diurnal cycle of O_3, which is very strong above about 45km/1hPa (e.g. Connor et al., 1994) but not negligible in the middle stratosphere either (e.g. Parrish et al., 2014). These studies are facilitated by the near-continuous measurement cycle, with retrievals performed on data integrated over a (few) hour(s) or so. Operational products are typically provided on 6-hourly intervals, but much higher sampling is possible as well, e.g. to study the abrupt changes at sunrise/sunset in the mesosphere.

In the context of validation of satellite O_3 profiles, MWR play a key role in extending the ground-based reference to altitudes not accessible by balloon sondes or LIDAR measurements. A recent application in the context of stability validation of limb sounders was presented by Hubert et al. (2016). A critical step in such validation work is the spatio-temporal co-location between ground-based and satellite measurement. While the ground-based measurements is usually associated with the location of the station, this assumption may not be optimal for MWR measurements, given their particular viewing geometry and measurement principle. This is the topic of this section.

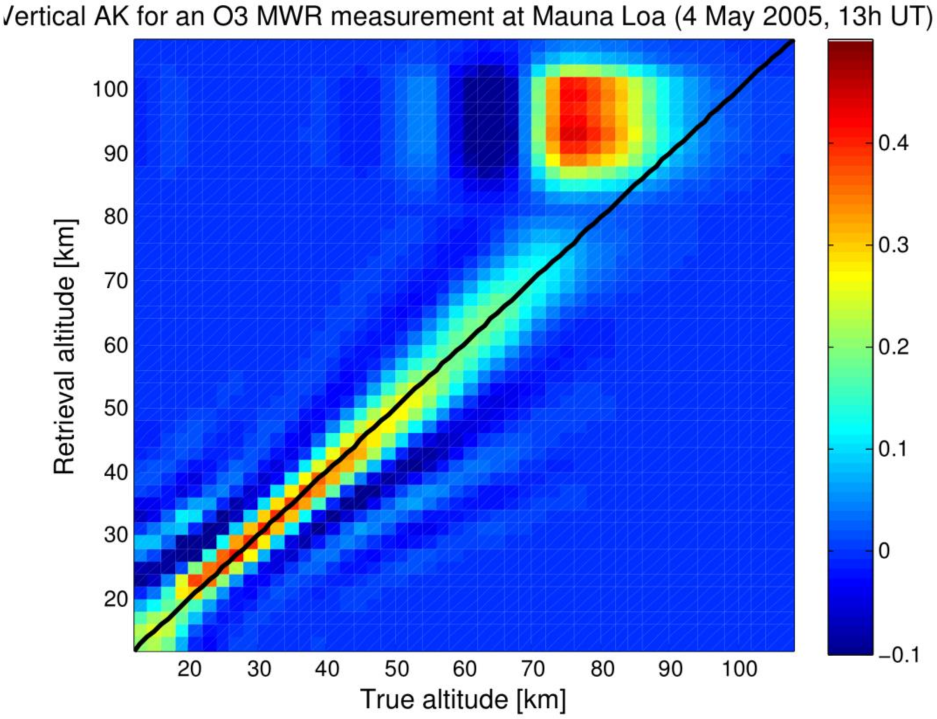


Figure 5.6: Typical vertical averaging kernel for an O_3 MWR measurement during night time at NDACC station of Mauna Loa, Hawai'i (19.54°N, 155.58°W, 3397m a.s.l.). To guide the eye, the black solid line represents the diagonal, where sensitivity ideally peaks. The apparent thermospheric sensitivity (at retrieval altitude 95km but true altitude 80km so upper mesosphere in reality) is not present during day time when most of the ozone at this altitude is destroyed more rapidly than it can be reformed because of the low local air density.

The line-of-sight and its projection on the ground

We can expect horizontal smoothing and sampling errors in O_3 MWR measurements for two reasons: (1) the non-zenith viewing angle, and (2) the sometimes significant integration times (several hours) which combine with the local wind speed to smear out the small-scale structure of the O_3 field. Below, we look in more detail at the viewing angle and how this combines with the measurement sensitivity to cause an offset between the location of actual measurement sensitivity and that of the station.

As a representative illustration, we show in Figure 5.7 the distribution of the elevation angles of 6 years of measurements with the O3 MWR at the NDACC site of Mauna Loa. At this station, a fixed azimuth angle is used (5° East-of-North) but the elevation angle is adjusted to compensate for weather-induced variations in the tropospheric opacity. In case of high tropospheric humidity elevation angles up to 24° are used to minimize the tropospheric path length. In very clear conditions, observations are done as low as 10° above the horizon. On average, the elevation angle is 14.1°.

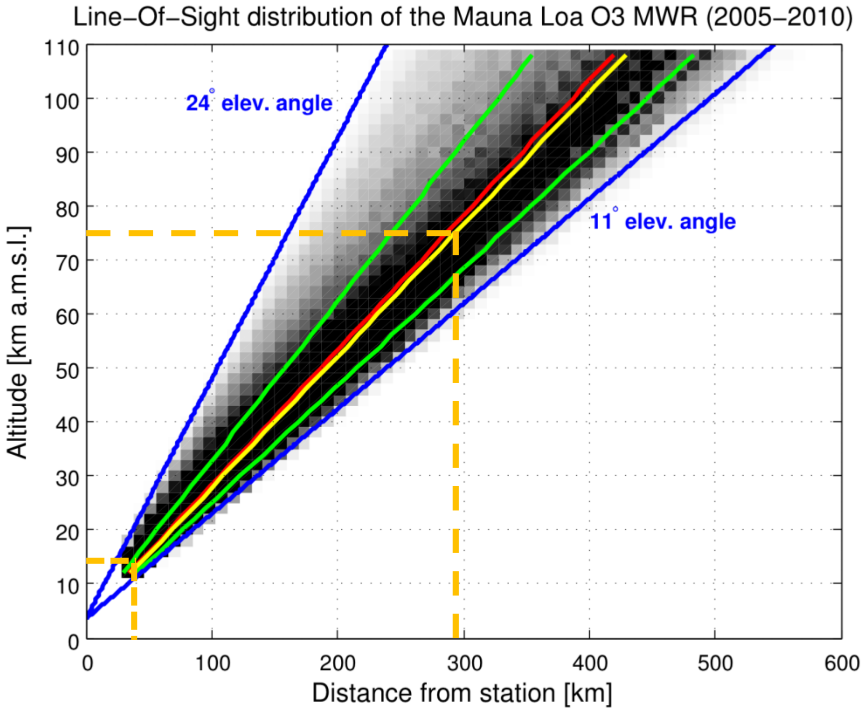


Figure 5.7: 2D histogram of the LOS of all measurements made with the O3 MWR at Mauna Loa between 2005 and 2010 (grayscale). The mean and median LOS are indicated in red and yellow respectively. The green lines represent the 1-sigma standard deviation. No measurements are made below about 11° or above 24°. The orange dashed lines indicate the altitude range for which this geometric projection is a reliable estimate of the true horizontal offset of the measurement sensitivity.

Figure 5.7 also shows the corresponding horizontal distance from the station for each altitude level, by pure geometric projection. To what extent this also represents the horizontal offset of the actual measurement sensitivity depends on the vertical distribution of the measurement sensitivity, which is quantified in the vertical averaging kernel. From Figure 5.6, and in agreement with Parrish et al. (2014), we conclude that the bulk of the sensitivity is close to the true altitude from the lowest levels up to about 70km. Hence, for this altitude range, a pure geometric projection as in Figure 5.7 can be used to estimate the horizontal offset of the measurement sensitivity w.r.t. the station location. Mathematically:

$$\mathrm{d}r = \frac{z}{\mathrm{tg}\gamma},$$

where $\mathrm{d}r$ is the distance from the station, z the altitude, and γ the elevation angle. Or, expressed as offsets in latitude and longitude (similar to Lambert et al., 2011):

$$\mathrm{d}\theta = \frac{dr}{\mathrm{R_T}}\cos\psi\,,$$

$$\mathrm{d}\varphi = \mathrm{tg}\psi.\ln\left(\frac{\mathrm{tg}\left(\frac{\theta_0}{2}+\frac{\pi}{4}+\frac{\mathrm{d}r}{2.\mathrm{R_T}}.\cos\psi\right)}{\mathrm{tg}\left(\frac{\theta_0}{2}+\frac{\pi}{4}\right)}\right),$$

where $\mathrm{d}\theta$ is the offset in latitude, $\mathrm{R_T}$ the radius of the Earth, ψ the viewing azimuth angle (East-of-North), $\mathrm{d}\varphi$ the offset in longitude, and θ_0 the latitude of the station.

Higher up, night-time observations also contain significant sensitivity but at altitudes slightly below the retrieval altitude (see the caption of Figure 5.6). The offset of the actual sensitivity is therefore somewhat smaller at these altitudes than estimated from the geometric projection, and the extreme offset values (>450km) in Figure 5.7 are probably never reached.

OSSSMOSE simulation of the errors induced by the non-vertical line-of-sight

In the section above, the horizontal offset between station and the location of the actual measurement sensitivity was estimated to be of the order of several hundred km for the upper stratosphere and lower mesosphere. The assumption that a measurement was made at the station zenith may therefore introduce significant errors when the ozone field is variable on these scales. Below, we quantify these errors using the OSSSMOSE simulation system (Verhoelst et al., 2015). The approach consists in simulating the actual measurements by interpolating 4-D reanalysis fields either on a vertical at the station location or along the LOS described in the previous section, and this at the times of the actual measurements provided through the NDACC archive. This is illustrated in Figure 5.8 for a measurement with the MWR at Lauder, New Zealand (45.04˚S, 169.68˚E, 370m a.s.l.).

The reanalysis of choice here is MERRA-2, in particular because of its good temporal sampling (3-hourly) compared to other reanalyses, an important asset in view of the diurnal cycle of O_3 in the higher stratosphere and mesosphere. Above about 45km/1hPa the diurnal cycle is so non-linear that further interpolation of the reanalysis fields to the actual measurement time leads to aleatory behavior and we chose instead to use fields as close as possible to the measurement time, i.e. always within 1.5h.

Since the aim of a ground-based measurement is usually to derive the profile above the station, i.e. at zenith, we take the zenith profile to be the reference and look at the difference between the profile along the LOS and that at zenith:

$$\delta x(z_i) = x_{LOS}(z_i) - x_{zenith}(z_i)\text{, or as a relative quantity: } \delta x(z_i) = \frac{x_{LOS}(z_i) - x_{zenith}(z_i)}{x_{zenith}(z_i)},$$

where x refers to the ozone concentration and z_i to a certain altitude level.

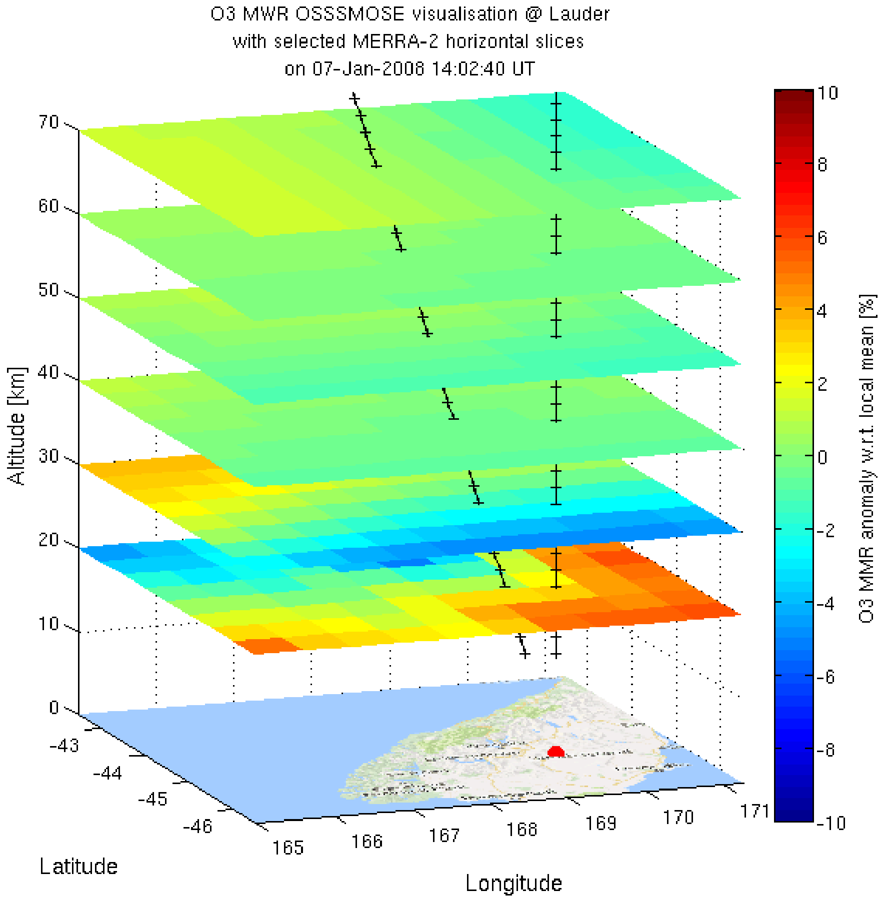


Figure 5.8: Visualisation of the OSSSMOSE simulation of both the zenith and actual LOS profiles at the time of a MWR measurement at Lauder, New Zealand. This graph is similar to the one in Figure 5.3, except for the model slices which are here shown on their native resolution, instead of the interpolated (smoothed) visualisation chosen for Figure 5.3. For clarity, only 6 MERRA-2 slices are shown, each converted to a local anomaly (difference w.r.t. the local horizontal mean at that altitude). The red marker indicates the station location and the black lines denote the zenith and actual LOS. Note that due to the different vertical and horizontal scales, the elevation angle appears much larger than it really is (19˚ in this case). Remark the extension to higher altitudes compared to Figure 5.3, and how much smoother the field is at these altitudes.

Two case studies will be elaborated here: the MWR operated by the Naval Research Laboratory at the tropical station of Mauna Loa, Hawai'i (19.54°N, 155.58°W, 3397m a.s.l.) and the MWR operated by NIWA at the mid-latitude station of Lauder, New Zealand (45.04°S, 169.68°E, 370m a.s.l.). From both instruments, extensive long-term data sets are available. Simulation results for 6 years of measurements at these two stations (covering the period 2005-2010) are shown in Figure 5.9. At

Lauder, differences remain in general (1-σ standard deviation) below 1%, except at the lowermost altitudes (below 25km) and above 55km. This is the result of the interplay between altitude-dependent atmospheric variability and increasing horizontal sensitivity displacement. Higher up (not shown), the differences increase rapidly, but this may be due to shortcomings in the simulation in dealing with the pronounced non-linear diurnal cycle of ozone at these altitudes so this needs further investigation. At Mauna Loa the differences are similar to those at Lauder in the middle to upper stratosphere, but they are larger at and below 20km, including some systematic effects. Identifying the cause of this requires further investigation as well. Typical measurement uncertainties (as provided with the data) at both stations are of the order of 4-5% random uncertainty and 6-8% systematic uncertainty. For a large part of the data sets, these measurement uncertainties can therefore be assumed to be larger than the “errors” due to the non-vertical LOS.

While the horizontal resolution of the MERRA-2 reanalysis is state-of-the-art for global chemical models (of the order of 80km), this is not necessarily sufficient to capture the small-scale variability of the ozone field, in particular at the lower levels. The significant integration times of the MWR measurements (of the order of an hour or more) however translate into horizontal smoothing because of the non-negligible wind velocities in the free troposphere and above. Consequently, the OSSSMOSE simulations performed here should not be affected much by their inability to resolve the smallest structures at the lowermost levels.

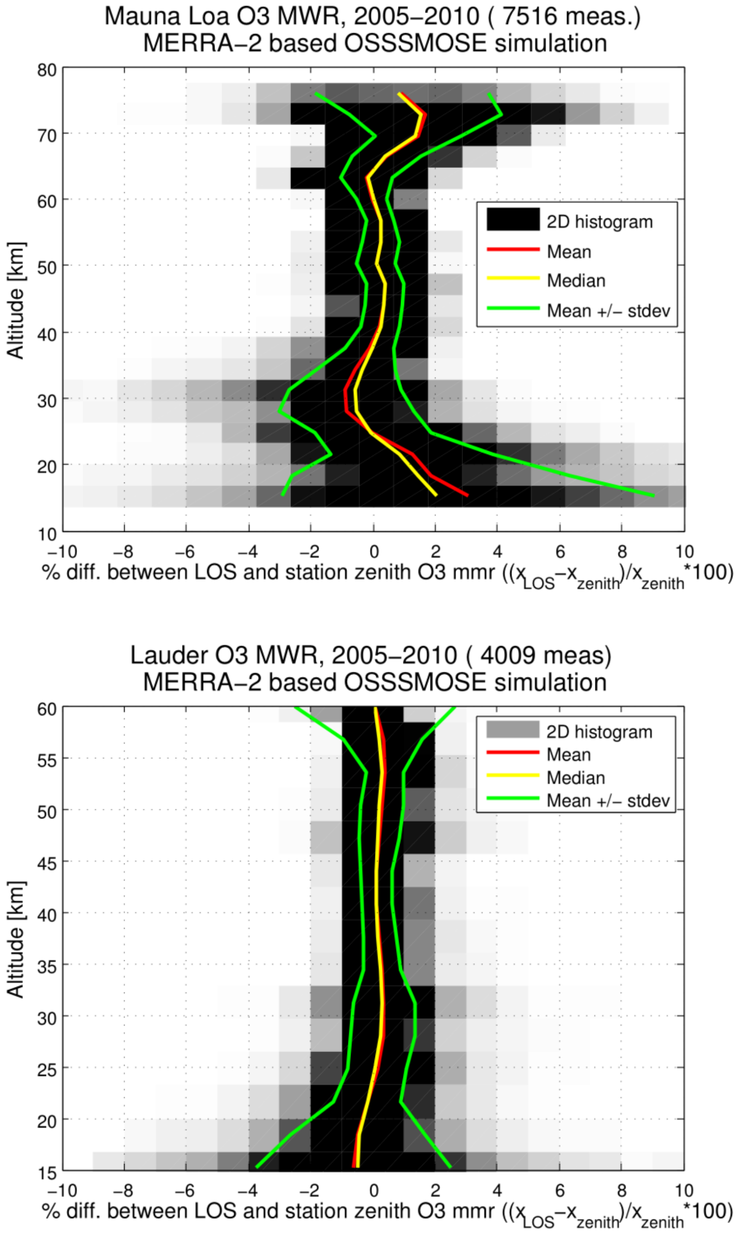


Figure 5.9. OSSSMOSE simulation of the difference in O_3 profile between the vertical at the station and the actual line-of-sight (LOS) of the O_3 MWR measurements at Mauna Loa, Hawai'i (upper panel) and Lauder, New Zealand (lower panel) . The grayscale represents the 2D histogram of the differences, the coloured lines some statistical properties. The mean and median are a proxy of the long-term systematic difference, the standard deviation (stdev) is a proxy of the random component.

An "effective" measurement location

Returning to Figure 5.7, it can be argued that the location of the instrument is not representative for the "mean" location of the measurement. When a single location is to be assigned to a measurement, e.g. in the context of comparisons with satellite data or models, it may be more meaningful to associate the measurement with an "effective" location taken to be the projection of the LOS on the ground at e.g. the 45km altitude level. A simulation similar to Figure 5.9 but using this effective location instead of the instrument location is shown in Figure 5.10. Comparing Figure 5.9 and Figure

5.10 reveals that indeed, differences w.r.t. the vertical at the effective location are noticeably smaller than those w.r.t. the instrument location, at least above 25km.

Based on these findings, this effective location was already used to co-locate MWR measurements with satellite limb measurements for recent validation work in the context of ESA's O3 CCI project and for stability analyses in SPARC's SI2N initiative (Hubert et al., 2016). Since the main driver for the use of ground-based MWR measurements in satellite validation is their sensitivity in the upper stratosphere and lower mesosphere, above the vertical range accessible by sondes, LIDAR, and FTIR, the large variability and potential bias at the lowermost measurement levels is not a concern as long as these issues are kept in mind.

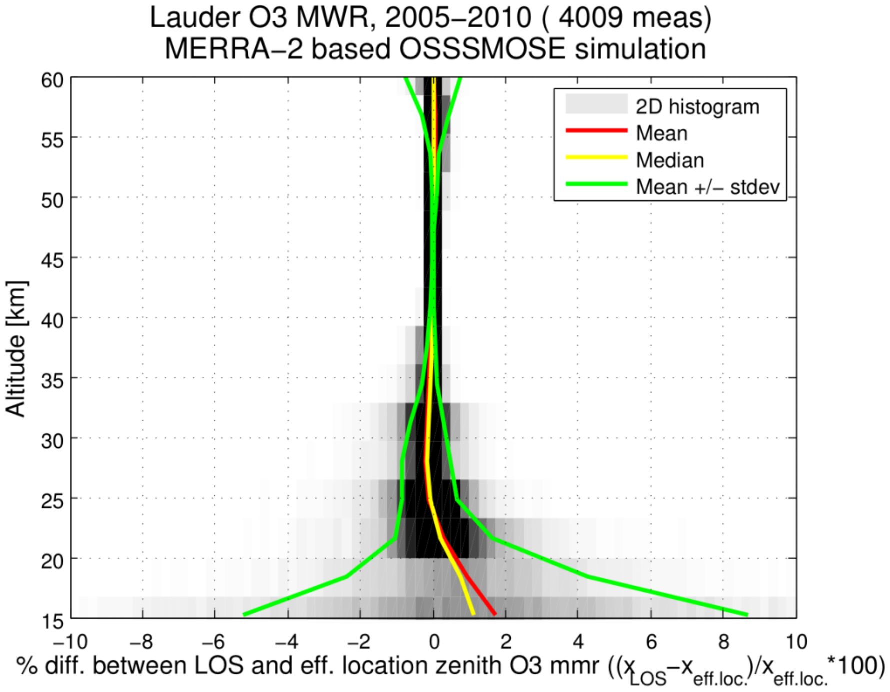


Figure 5.10. Similar to the upper panel of Figure 5.9, but the differences are calculated w.r.t. the zenith at the "effective location", here taken to be the ground projection (along the LOS) of the 45km altitude level, instead of the zenith at the instrument location.

Conclusions

MWR measurements of the vertical ozone profile are obtained using a line-of-sight (LOS) with a low elevation angle (usually between 10˚ and 20˚). This leads to a significant displacement of the location of actual sensitivity w.r.t the instrument location, reaching several 100km for the upper measurement levels. For that part of the altitude range for which the vertical averaging kernel of the MWR measurement is more or less diagonal, the projection on the ground of the LOS can be used to estimate this displacement. From an OSSSMOSE simulation, it was found that the resulting differences w.r.t. a zenith measurement at the instrument location are largest below 30km and above 65km. Instead of associating a measurement with the instrument location, it is shown that it may be better to use an "effective" location, corresponding to the projection on the ground of the 45km altitude level following the LOS. In general however, the measurement uncertainty remains the dominant source of uncertainty, leading to errors larger than the difference between LOS and zenith profile.

5.3 Stratospheric ozone LIDAR: integration time

Stratospheric O_3 profiles are measured with LIDAR (Light Detection and Ranging) instruments using the DIAL (Differential Absorption LIDAR) technique at 13 different stations (+ 1 mobile instrument) contributing to the NDACC (*http://ndacc-lidar.org/*). The DIAL technique relies on the difference in returned (backscatter) pulsed laser signal at two wavelengths for which the ozone cross sections are significantly different (Kobayashi & Toyama, 1966). Three of these instruments measure tropospheric ozone, the others cover the stratosphere (10-50km at best). We consider here only the stratospheric LIDARs as these play a crucial role in validating satellite ozone profile measurement above the burst point of ozonesonde balloons. The typical measurement uncertainty (from internal estimates, from network homogeneity, and from comparisons with other measurement techniques) is better than 5% up to the ozone maximum but degrades to 10-30% in the upper stratosphere (McDermid et al., 1990, Godin-Beekmann et al., 2003, and Keckhut et al., 2004). The vertical resolution is of the order of 300m at the bottom of the measured profiles, i.e. near the tropopause, but degrades to 3km and worse in the upper stratosphere (Godin et al., 1999). Note that at these higher altitudes, the profile is strongly oversampled, i.e. measurement levels are highly correlated due to the coarser actual measurement resolution.

LIDAR integration times

Depending on the power of the laser, the meteorological conditions, length of night, etc., the typical total integration time ranges from an hour up to an entire night. While an individual measurement may in fact be of much shorter duration (a few minutes), multiple consecutive measurements must be averaged to reach a sufficient SNR for a meaningful ozone profile retrieval. The distribution of integration times (per retrieval) for 3 well-established instruments is visualized in Figure 5.11.

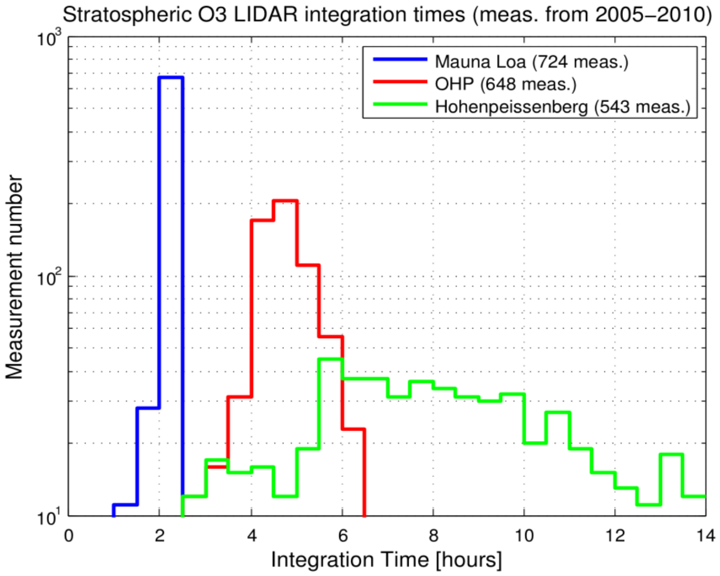


Figure 5.11: Histograms of the integration times for LIDAR measurements are 3 stations with extensive data sets between 2005 and 2010: Mauna Loa Observatory, Hawai'i (19.5°N, 155.6°W, operated by JPL), Observatoire de Haute Provence (OHP), France (43.9°N, 5.7°E, operated by LATMOS), and Hohenpeißenberg Observatory, Germany (47.8° N, 11.0° E, operated by DWD).

This integration time is reported in different ways across different stations, file types, and archives: for instance, total integration time can be reported along with the center time of the integration, or start and end times along with an effective mean measurement time, which takes into account gaps due to technical or meteorological difficulties.

Temporal smoothing and representativeness errors: definitions

At these temporal scales, we can expect variations in the ozone field during the integration of several 10s of percent, due to advection and photochemistry. This can affect the interpretation of the measurements in two ways:

1. If the evolution of the ozone concentration during the integration is non-linear, the mean value (i.e. the measurement) may not be equal to the value of the concentration at the claimed measurement time (typically the middle of the integration). This is a *temporal smoothing error*. Mathematically, we define it as follows:

$$\delta x_{temp.smoothing}(z_i,t_0)=\int_{t_0-\Delta t/2}^{t_0+\Delta t/2}x(z_i,t)dt\Big/\Delta t-x(z_i,t_0),$$

   or as a relative quantity:

$$\delta x_{temp.smoothing}(z_i,t_0)=\frac{\int_{t_0-\Delta t/2}^{t_0+\Delta t/2}x(z_i,t)dt\Big/\Delta t-x(z_i,t_0)}{x(z_i,t_0)}$$

   where x represents the ozone concentration, z_i refers to an altitude level, t_0 to the nominal measurement time, and Δt to the integration time.

2. The ozone concentrations may vary strongly during the integration, and the mean (i.e. the reported) value may be representative for only a small part of the covered measurement time. This is considered to be a *temporal representativeness error*:

$$\delta x_{temp.representativeness}(z_i,t)=x(z_i,t)-\int_{t_0-\Delta t/2}^{t_0+\Delta t/2}x(z_i,t)dt\Big/\Delta t,$$

   or as a relative quantity:

$$\delta x_{temp.representativeness}(z_i,t)=\frac{x(z_i,t)-\int_{t_0-\Delta t/2}^{t_0+\Delta t/2}x(z_i,t)dt\Big/\Delta t}{\int_{t_0-\Delta t/2}^{t_0+\Delta t/2}x(z_i,t)dt\Big/\Delta t}.$$

   Because this representativeness error is different for every time t within the integration window, it makes more sense to speak of a representativeness uncertainty, defined as the spread (standard deviation) on the representativeness errors (following the VIM and GUM on the definition of uncertainty as an estimate of a dispersion):

$$\sigma_{temp.representativeness}(z_i)=stdev_t(\delta x_{temp.representativeness}(z_i,t)),$$

where $t \in [t_0 - \Delta t/2, t_0 + \Delta t/2]$.

To what extent these errors can be comparable to (or larger than) the measurement uncertainty, is estimated in the following section.

OSSSMOSE simulation of the temporal smoothing and representativeness errors

To estimate temporal smoothing and representativeness errors, we extracted from MERRA-2 reanalysis fields the evolution of the ozone profile during the entire integration period, slightly oversampled to a 1-hour resolution and spatially interpolated to the exact instrument location and vertical grid. An example of such an extraction, for a measurement with the LIDAR installed at Hohenpeißenberg, is shown in Figure 5.12.

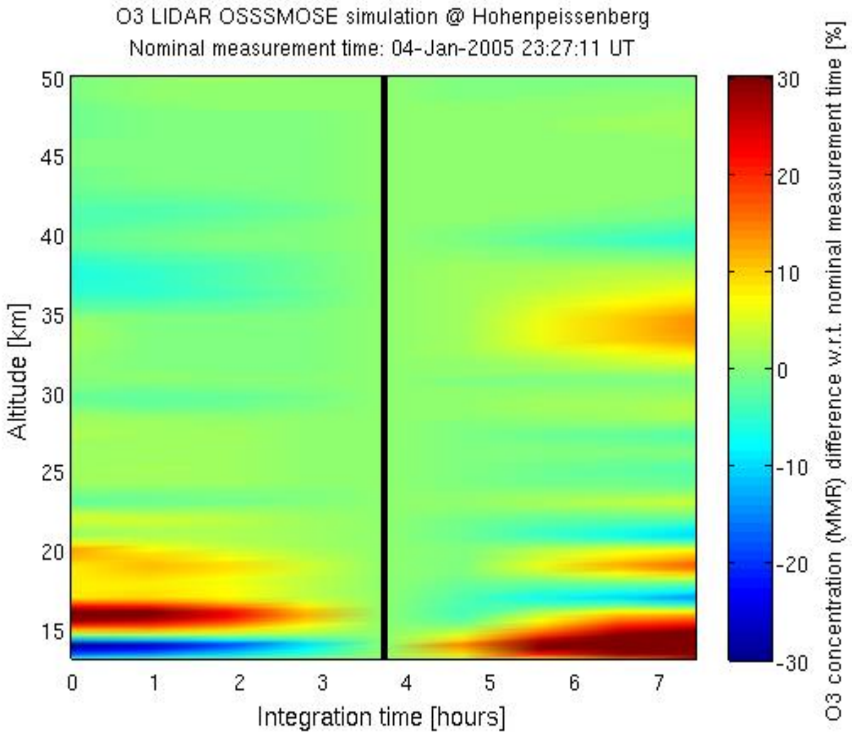


Figure 5.12: Evolution of the MERRA-2 ozone profile during a 7.5-hour LIDAR integration at Hohenpeißenberg, normalized to the profile at the nominal measurement time, indicated by the black solid line. This graph illustrates the potentially large temporal representativeness errors.

Some general features can already be seen in this graph: the decrease in variability with increasing altitude (except near the ozone maximum) and the presence of both linear changes (e.g. at the lowest altitude in this case) and non-linear changes (e.g. at 20km and 33km). It is also striking that the amplitude of the variations can substantially exceed the measurement uncertainty (better than 5% in the lower and middle stratosphere). In Figure 5.13, the same measurement is visualized in such a way that both the variability during the integration and the difference between the mean (i.e. the measured) profile and the profile at the nominal measurement time become clear. The former leads to a representativeness uncertainty, while the latter corresponds to the temporal smoothing error.

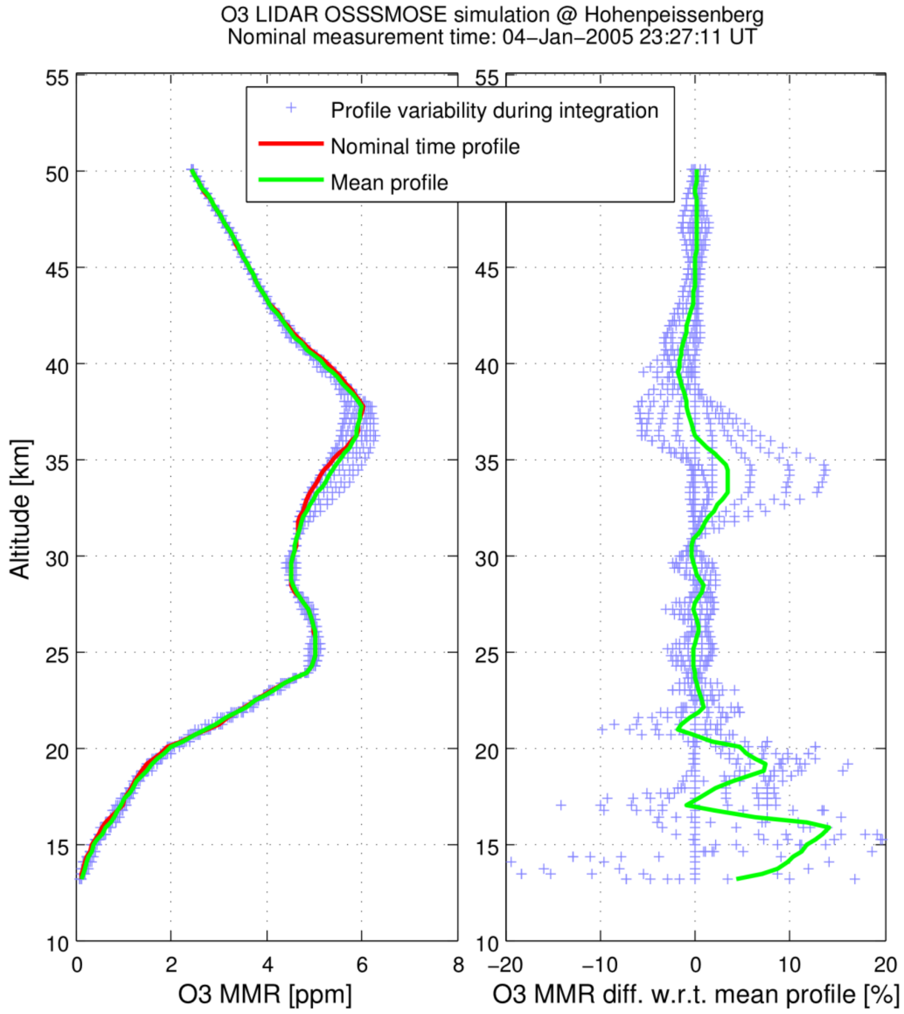


Figure 5.13. Variation of the ozone profile during a LIDAR integration at Hohenpeißenberg (the same measurement as in Figure 5.12), as captured by MERRA-2, both in absolute concentration values (left-hand panel), and relative to the profile at the nominal measurement time, i.e. the middle of the integration window (right-hand panel). The green curve in the right-hand panel corresponds to the (relative) temporal smoothing error.

Temporal smoothing errors were simulated as described above for all measurements between 2005 and 2010 obtained at the three NDACC stations already introduced in Figure 5.11. The results are shown in Figure 5.14 as a 2-D histogram and some summarizing statistics. The smallest smoothing errors (well below 0.5% for most of the altitude range) are estimated for the equatorial Mauna Loa observatory, where the ozone field is relatively stable (and varies linearly during the integration time) and the LIDAR integration times are fairly short (of the order of 2 hours). At the lower end of the altitude range, the smoothing errors are somewhat larger (occasionally up to 2%), but this remains at or below the measurement uncertainty. At the Observatoire de Haute Provence, integration times are longer (on average about 4 hours) and, moreover, the mid-latitude ozone field is more variable. Consequently, larger smoothing errors are estimated. Still, only below 20km altitude do they become comparable to the measurement uncertainty. At Hohenpeißenberg, a mid-latitude station where integration times up to 12 hours are used, the smoothing errors are again larger, in particular also in the middle of the altitude range. In general, they still do not dominate the measurement uncertainty, but further differentiation w.r.t. integration time and season may reveal regimes where smoothing errors become the dominant source of uncertainty. This is ongoing work. Note also that at none of the stations significant systematic errors (the mean or median of the smoothing errors, represented in red and yellow respectively) are observed. Again, this needs further differentiation, in particular looking for potential seasonalities.

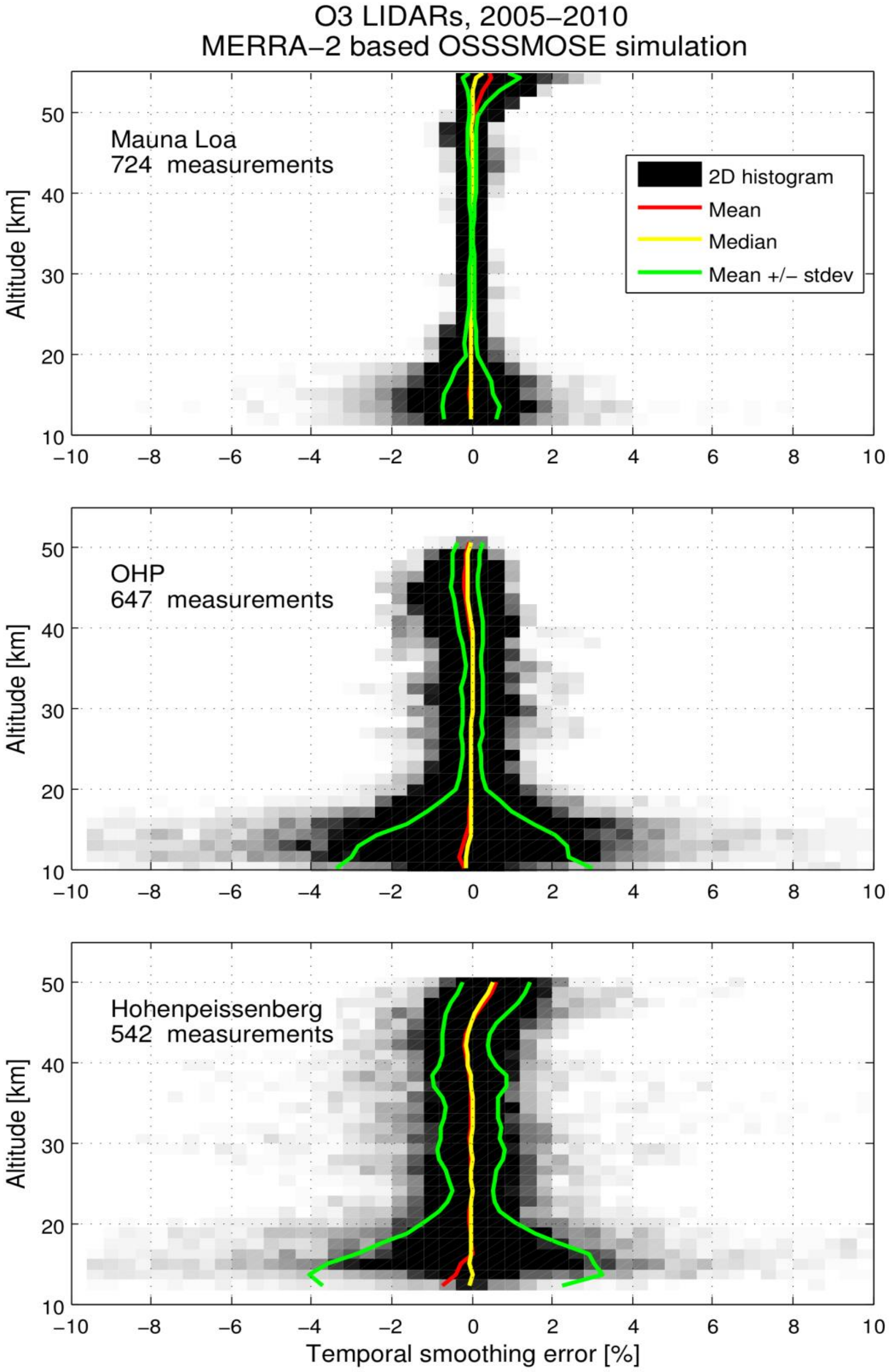


Figure 5.14: OSSSMOSE estimates of the temporal smoothing errors in 6 years of LIDAR measurements at the 3 stations for which the distribution in integration times is shown in Figure 5.11. See the text for an in-depth discussion.

While the temporal smoothing errors are found to be relatively modest, due mostly to the fact that linear changes in O3 concentration during the integration time do not lead to smoothing errors (if the nominal measurement time corresponds to the middle of the integration window), Figure 5.12 suggests that the temporal representativeness errors may in fact be much larger.

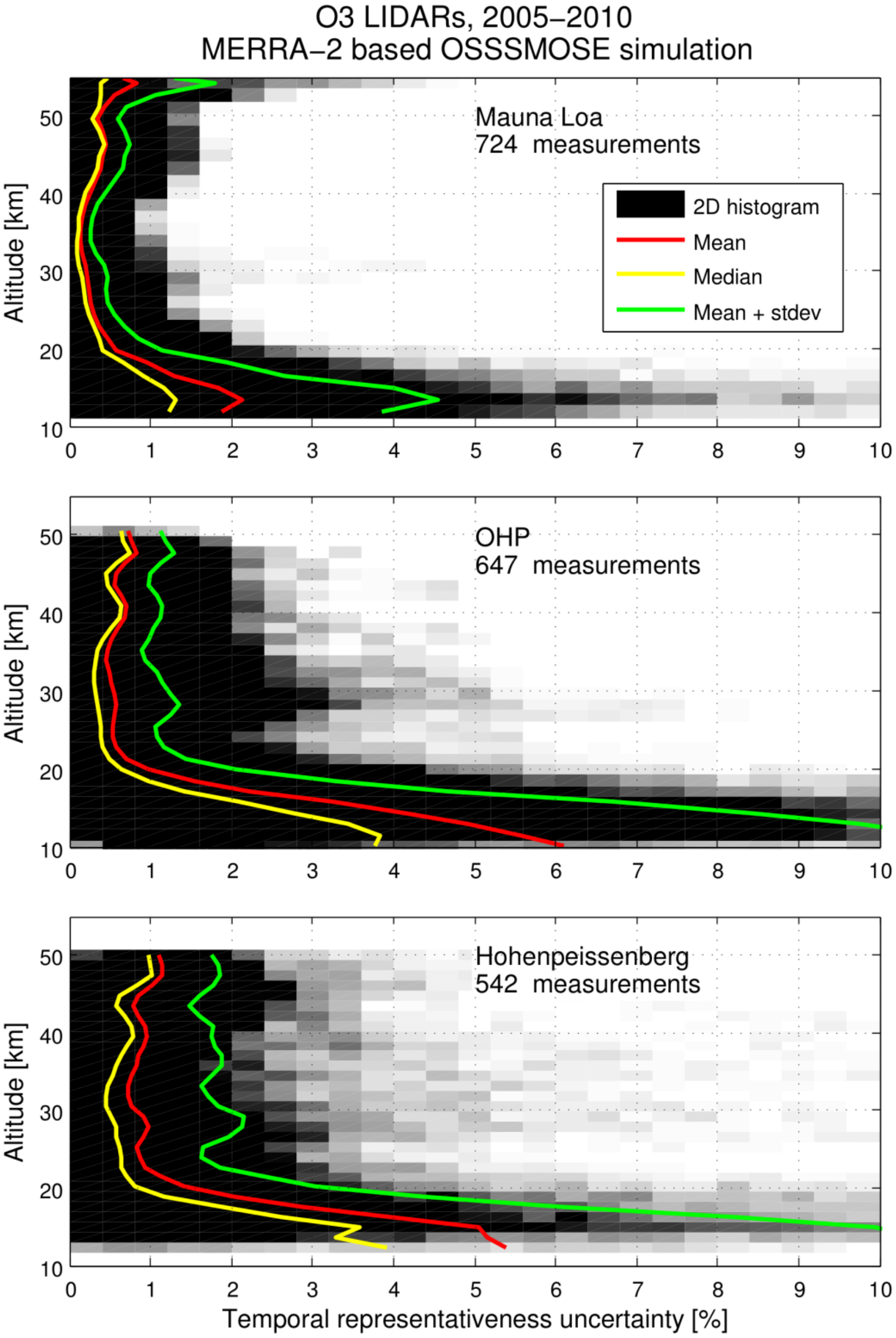


Figure 5.15: The distribution of temporal representativeness uncertainties, and some summarizing statistics as coloured lines. See text for the definition and estimation method. TO DO: remove mean and replace stdev with a quantile because this distribution is far from symmetrical.

Figure 5.15 demonstrates that this is indeed the case. The temporal representativeness uncertainties were calculated as the spread (standard deviation) of the temporal representativeness errors, calculated for a selection of times within the integration window, and relative to the mean concentration over the total integration window (i.e. the actual measurement). For every measurement, this yields an representativeness uncertainty profile. These uncertainty profiles are analyzed in a statistical way in Figure 5.15: their distribution is visualized with a 2-D histogram. From the different panels in this figure, we derive a similar dependence on integration time, latitude, and altitude as for the temporal smoothing errors. However, the amplitudes are clearly larger: the mean representativeness uncertainty at Hohenpeißenberg is of the order of 1%, but 2% is far from uncommon in the middle stratosphere, and at the lowermost levels uncertainties up to 10% are common. Note that these numbers are about the representativeness uncertainty of a single measurement, defined as the standard deviation (1σ) of the actual errors. In other words: individual errors larger than the numbers discussed here are still relatively common (at least 30% occurrence if the distribution is nicely Gaussian).

Conclusions

In this section, we investigated to what extent the potentially long integration times required for LIDAR measurements of the vertical O_3 profile affect (1) the accuracy of the measurement (i.e. does the long integration lead to a *temporal smoothing error* in the measurement when it is assigned to the mean measurement time?), and (2) the representativeness of the measurement for the time period covered by the integration (i.e. how much does the O_3 concentration vary during the integration?). We chose here not to visualize *individual representativeness errors*, but only the resulting *representativeness uncertainty*.

In general, the OSSSMOSE simulations indicate that the temporal smoothing errors are actually rather limited. This is due to the mostly linear variation in the ozone concentrations at the time scale of an integration. Only for the longest integration times and in variable atmospheric conditions do the smoothing errors become comparable to the measurement uncertainty (several percent).

On the other hand, the representativeness uncertainty is significantly larger. In the lower stratosphere, it can be up to three times as large as the measurement uncertainty. This is of particular importance when LIDAR measurements are co-located with other measurements that fall within the integration time, but can thus not necessarily be expected to agree within their measurement uncertainties.

Finally, we must repeat the caveat already expressed in previous sections: the OSSSMOSE simulations are inherently limited in their representation of small-scale variability by the resolution of the underlying model/reanalysis fields. As such, it may be that errors and uncertainties are somewhat underestimated, in particular in the lowermost atmospheric layers and when referencing against “singular” snap shots (e.g. the O_3 concentrations at the nominal measurement time).

6 Aerosol: Comparison of AATSR and AERONET

Retrieval algorithms based on passive satellite remote sensing instruments provide information on atmospheric aerosols on a global scale. While they are less accurate than methods based on active sensors (lidars), they can provide a global coverage every few days. The global AOD data can be used e.g. in air quality and climate studies.

We consider the comparison of the total atmospheric column aerosol optical depth (AOD) values obtained with AATSR and AERONET. AERONET provides point-like observations with several observations per day, while AATSR provides an instantaneous glimpse of a larger area (with a revisit time of 2-3 days). The focus is on assessing co-location mismatch uncertainties in comparison of the two datasets.

The uncertainties for the satellite based aerosol retrieval are much larger than for the AERONET observations. In addition, while the instrument uncertainties are propagated through the retrieval to produce AOD uncertainty estimates, various sources of uncertainties in the satellite retrievals remain unquantified. Therefore, the characterization of the co-location mismatch uncertainties is hampered by the incomplete uncertainty characterization of the satellite observations. Nevertheless, the relevant features are discussed in this section.

6.1 AATSR description

The European Space Agency's (ESA) Advanced Along Track Scanning Radiometer (AATSR) aboard ENVISAT measured the top of atmosphere (TOA) radiance at seven wavelengths (0.550, 0.659, 0.865, 1.61, 3.7, 10.85 and 12 µm). AATSR data record extends from 2002 to April 2012, when the connection with ENVISAT was lost. AATSR employed an along track scanning technique, where each surface pixel is viewed twice, first by the 55º forward view and some 150 seconds later by the nadir view. The AATSR Dual View (ADV) algorithm is used to retrieve AOD over land, and the AATSR Single View (ASV) algorithm is used over ocean. The ADV algorithm uses the stereo view to remove the surface reflectance contribution from the TOA reflectance, and retrieves the best fit aerosol model and AOD value using inversion techniques (Kolmonen et al. 2016).

Initial regridding

AATSR employs a conical scanning technique, where the sampling distance varies across track and depending on the view. These native data are regridded to a 1 km grid, where nadir and forward views are collocated (level 1B data). The regridding involves 'cosmetic filling' for grid pixels with no native data. This takes place in particular for the forward view, as the native pixel size in the forward direction is larger. The different smoothing between nadir and forward views is likely to cause smoothing difference errors. In the retrievals the L1B regridded data provided by ESA is used. Smoothing uncertainty estimates related to the regridding are not available.

The regridding differences can be seen as the difference in the level of details between nadir and forward view images (Figure 6.1). In the ADV/ASV algorithm, the ratio of the nadir and forward view surface reflectance, the k-ratio, is a crucial parameter. At coastlines, where the surface reflectance changes rapidly, the regridding difference errors are seen as erroneous k-ratio values, leading to overestimated AOD values (see Figure 6.1). The 3x3 pixel standard deviation of the k-ratio can be used as a qualitative indicator of the regridding difference uncertainty.

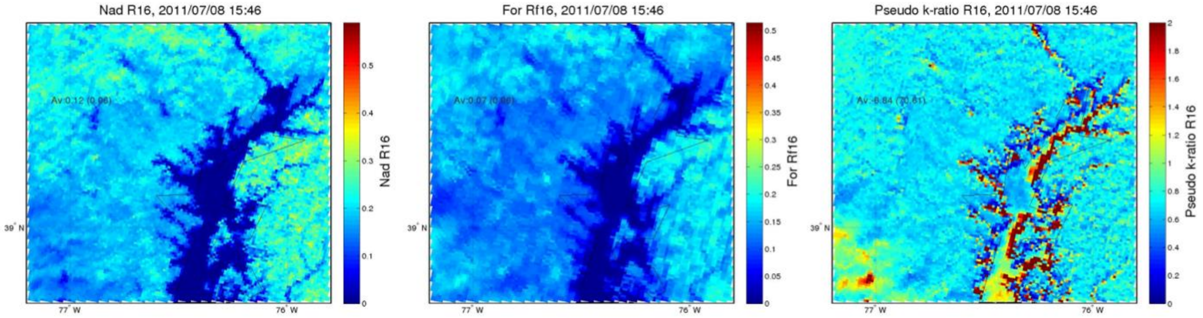


Figure 6.1 The regridding (smoothing) difference between the two AATSR views can be seen as a difference between the level of detail in the nadir view (left) and forward view (center) images of observed TOA reflectance at 1.6 µm over the Chesapeake bay area. The smoothing difference error leads to highly erroneous reflectance ratio (k-ratio) values at the coastlines (right).

Cloud screening

As part of the data sampling in ADV/ASV, cloud screening is applied. Aerosol retrieval can only be performed over cloud free areas. Four cloud tests are used in ADV/ASV to remove clouded pixels (Kolmonen et al. 2016). The tests are based on automatically adjusted thresholds for TOA reflectance and brightness temperature, and the relations of these at different wavelengths (Robles-Gonzales et al. 2003). The tests are applied separately for each viewing direction, and a pixel is considered cloud contaminated if any of the tests in either view flags it as cloud.

The cloud mask is not perfect, and residual (undetected) clouds cause overestimated AOD values. Also, the two AATSR views are collocated on the surface level, while residual clouds are higher. The parallax between the two views causes dislocation of clouds, leading to further uncertainty. Setting the cloud mask thresholds is a trade-off between the cloud mask uncertainty and coverage. The possible cloud contamination is seen as a major source of uncertainty in the AOD retrieval, but it remains unquantified as the full characterization of the cloud mask uncertainty is outside the scope of this exercise.

In addition to the cloud masks, an additional post-processing is applied to ADV/ASV data to reduce the effect of residual clouds (Kolmonen et al. 2016). For each retrieved 0.1º pixel the adjacent 8 pixels are inspected. If less than three adjacent pixels are retrieved (due to cloud screening) the center pixel is considered as cloud contaminated and removed. Also, if the standard deviation of AOD in the 3-by-3 pixel area is higher than 0.1, the center pixel is removed. Applying the cloud post-processing improves validation results against AERONET, but removes some of the high AOD cases (Kolmonen et al. 2016). An improved cloud post processing method which better takes into account high AOD cases has recently been developed (Sogacheva et al. 2017).

The parameters used in the cloud post-processing, the number of valid adjacent pixels and the 3x3 pixel standard deviation, can be used as qualitative cloud contamination uncertainty indicators. They do not depend on the cloud tests applied, and are easy to implement. As such, they might be well suited to the Virtual Observatory, with adjustable thresholds for the user.

ADV/ASV sampling

The ADV/ASV level 2 aerosol product is provided on a 0.1º grid (~10 km resolution), called 'the coarse resolution'. The retrieval is performed for an area of c. 100 subpixels (depending on latitude) using the observed TOA reflectance in both views (L1B data). Due to surface reflectance variability, a simple averaging of the reflectance over all subpixels might lead to a reflectance value that is not representative for the area. For example, exceptional surface conditions (e.g. sun glint from a small water body or urban structure) may cause extraordinarily high reflectance for one subpixel for one of the views. Therefore, a method to select a subset of representative subpixels is used.

In ADV/ASV a histogram method is used to select the representative pixels in both views from a 10-by-10 pixel retrieval area (Kolmonen et al. 2016). The reflectance values are binned and the mode reflectance bin is selected respectively for each view. Then the pixels which are in the mode bin for both views are selected for calculating the averaged reflectance for each view. A fully traceable sampling uncertainty of the histogram method is not calculated, but a quantitative measure is given by the standard deviation of the TOA reflectance of the selected 1-by-1 km pixels in the 0.1º-by-0.1º retrieval area. The effect of this parameter is shown in Figure 6.6 c).

The ADV/ASV algorithm can also be run in a special high resolution mode, where the AOD retrieval is run respectively for each 1 km pixel. This computationally heavy approach is not used for global retrievals, but can be applied for case studies. Here the high resolution retrieval is used to assess the smoothing method used in ADV/ASV (Figure 6.2).

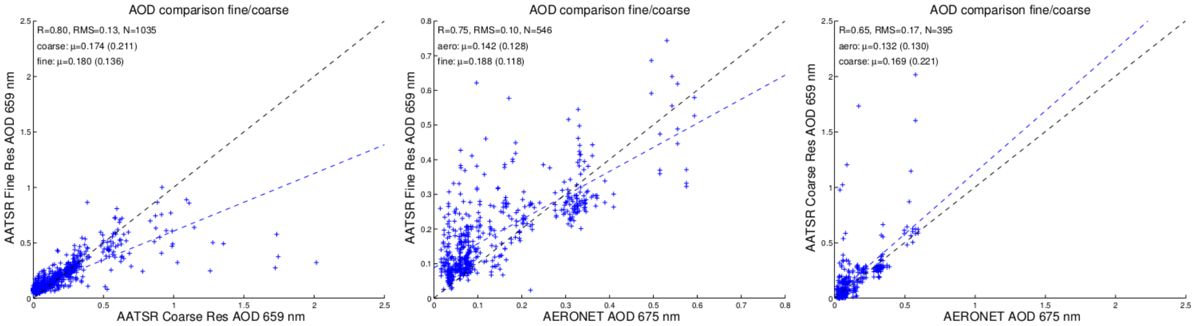


Figure 6.2 Comparison of the cloud post-processed high resolution ('fine') and low resolution ('coarse') ADV/ASV retrievals, averaged over each low resolution pixel (a), and comparison against AERONET for high (b) and low (c) resolution results. The low resolution retrieval has slightly more variance and slightly lower correlation with AERONET, but the overall agreement between the two retrieval modes is acceptable.

ADV/ASV uncertainty estimate

In ADV/ASV a static instrument uncertainty of 5% is assumed, and propagated through the retrieval to get an AOD uncertainty estimate for each retrieved pixel (Kolmonen et al. 2016). The uncertainties are typically of the order 0.01 – 0.1 (0.064 on the average) for the dataset used in this exercise. This estimate does not include sampling and smoothing uncertainties, uncertainties related to the selection of the best-fit aerosol model, or uncertainties related to the cloud screening.

6.2 AERONET description

AERONET (Aerosol Robotic Network) is a federated network of sun photometer instruments deployed at several hundred locations over the world for aerosol monitoring (Holben et al., 1998). The AERONET sun photometers measure solar irradiance at multiple wavelengths from UV to NIR to provide AOD with an uncertainty of 0.01-0.02 (Eck et al., 1999).

In this exercise we use the cloud-screened, quality-assured Level 2.0 AERONET AOD data for the wavelengths 440, 675, 870, and 1020 nm. Since the wavelengths do not match with those of AATSR, Ångström exponent is used to derive AERONET AOD values at 555 and 659 nm wavelengths.

AERONET DRAGON Campaign

In the summer 2011 AERONET deployed more than 40 CIMEL Sun-sky radiometers in the Baltimore-Washington DC region for the DRAGON (Distributed Regional Aerosol Gridded Observational Network) campaign (AERONET 2011). The campaign provides AOD results for a network of sites distributed on a roughly 10 km grid (see e.g. Schafer et al. 2014). The observations from this campaign are used here to study the natural AOD variability, and the related smoothing and sampling uncertainties in AATSR AOD retrievals.

6.3 Comparison methods

We use the AERONET quality-assured Level 2.0 AOD for validation of the satellite based AOD results, which is a common practice (see e.g. de Leeuw et al. 2015). For each AATSR orbit, spatially matching AERONET stations are searched for. For a matching site, the AATSR data for a fixed distance around the site is averaged (Figure 6.3). Correspondingly, a temporal average of the AERONET values in a fixed time frame around the AATSR overpass time is taken.

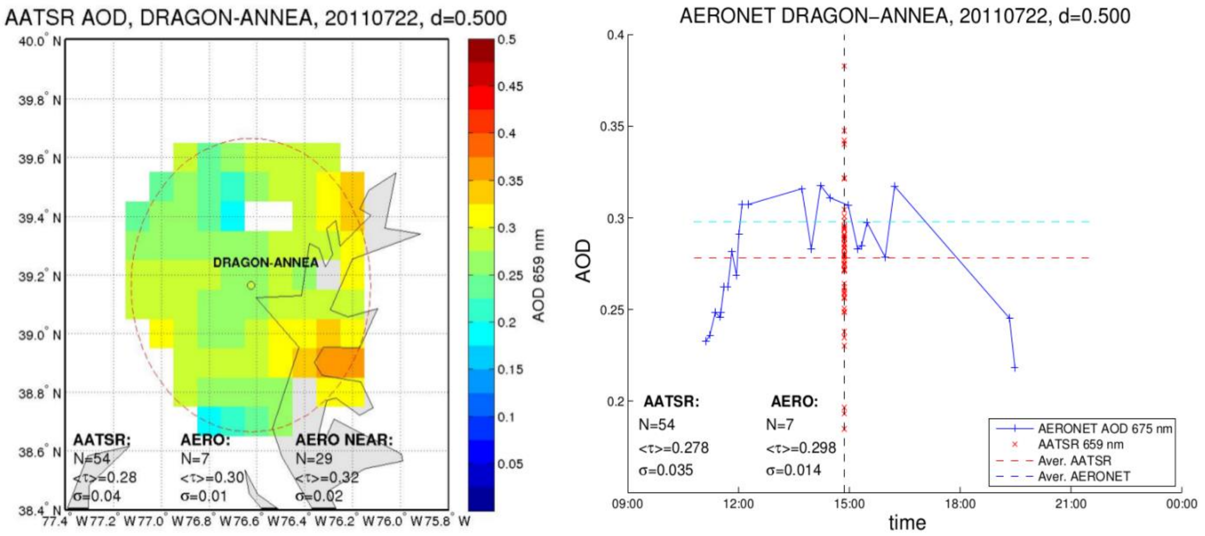


Figure 6.3 Illustration of the AOD comparison between AERONET and AATSR. **a)** The color shows the AATSR AOD values around an AERONET site. The dashed circle shows the sampling area. **b)** The blue curve shows the AERONET observations as function of time while the red dots show the AATSR observation at the overpass. The dashed lines show the obtained average values.

A measure of the agreement between the ADV/ASV AOD results and AERONET is obtained by calculating the correlation coefficient R between a number of matching AATSR and AERONET observations. A scatter plot of collocated AATSR and AERONET AOD results at 555 nm and 659 nm, is shown in Figure 6.4. Naturally, the validation results depend on the sampling parameters: the distance around the site for averaging AATSR data (Figure 6.5), and the time window used for averaging the AERONET data.

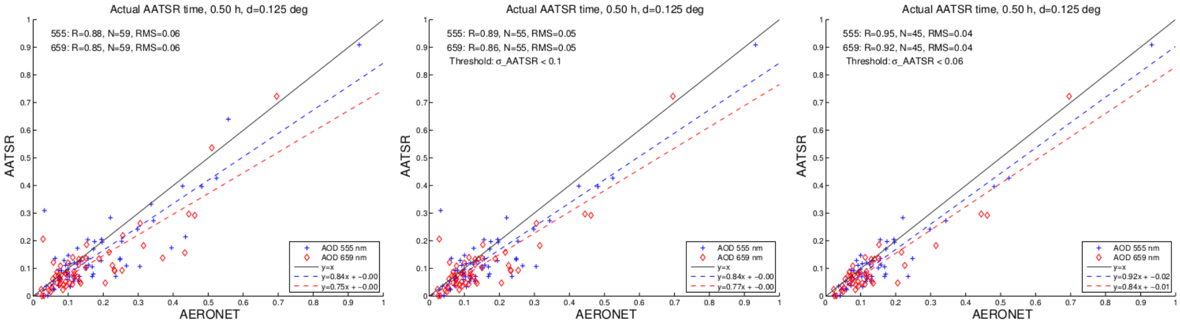


Figure 6.4 Scatter plots of AATSR and AERONET AOD at 555 nm (blue) and 659 nm (red). For this example global data from September 2008 are used. **a)** All matches included; **b)** pixels with σ_{AATSR}> 0.1 removed; **c)** pixels with σ_{AATSR}> 0.06 removed. The correlation coefficient is improved from 0.88 to 0.95 by applying the threshold (for 555 nm).

The AERONET data has a typical temporal sampling rate of ~15 minutes, and we use a one hour sampling window for the comparison as a default. The size of the temporal matching window does not have a large effect on the comparison results (not shown). The standard deviation of AOD values within the temporal sampling window for each match was calculated. The observed temporal variation of AOD data is of the order 0.01-0.02, and we perceive that the co-location mismatch is dominated by the spatial effects.

A simple measure of the representativeness of the point-like AERONET measurement for the larger area covered by the AATSR data is obtained from the standard deviation of the AATSR AOD (σ_{AATSR}) around the AERONET site. For highly varying AOD the point-like AERONET measurement is likely to be less representative. It is seen in Figure 6.4 that by applying a threshold to σ_{AATSR}, the correlation between AATSR and AERONET can be significantly improved. In this sense σ_{AATSR} serves as a quantitative measure of the co-location mismatch uncertainty.

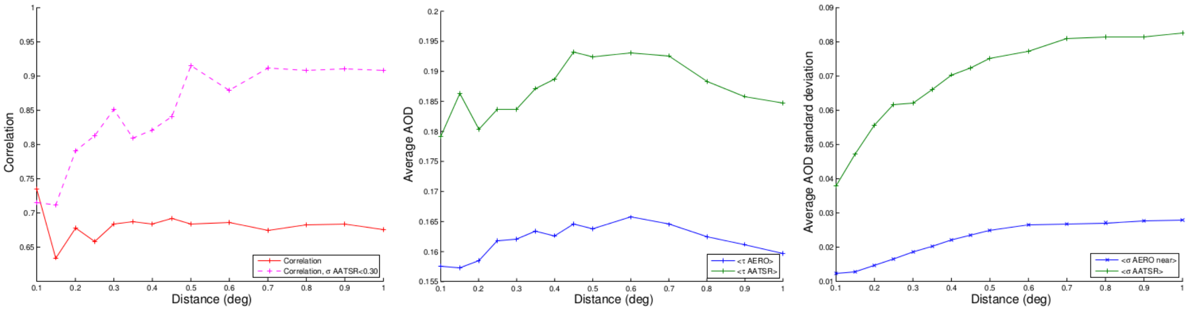


Figure 6.5 Effect of the sampling distance to the correlation between AATSR and AERONET. **a)** The correlation coefficient between AATSR and AERONET AOD (at 659 nm) for all data (red curve) and for cases limited to σ_{AATSR} < 0.3 (magenta curve). **b)** Average AOD as function of the sampling distance for AATSR (green curve) and for AERONET (blue curve), using the nearby AERONET sites with the same sampling radius (see Section 6.3). **c)** The corresponding (spatial) standard deviation of AOD.

However, it must be noted that the variation of the AATSR AOD values around a site is not only due to the natural variability of aerosol load, but is very likely dominated by ADV/ASV retrieval errors. Hence it is not a direct measure of the co-location mismatch uncertainty. The dominant error sources in the satellite aerosol retrieval are residual clouds (overestimated AOD due to clouds surviving the cloud screening), and varying surface reflectance in connection with the satellite dual view co-location uncertainties. The spatial variability of AOD is further studied in the following subsection.

Another simple measure of sampling uncertainty is the number of retrieved pixels in the fixed area around an AERONET site (N_{AATSR}). Pixels may be missing (retrieval not performed) due to cloud screening. Presence of (detected) clouds in the comparison area may suggest an elevated probability of residual clouds, and thus low N_{AATSR} indicates higher sampling uncertainty. Setting a higher threshold for N_{AATSR} improves the correlation slightly (not shown).

Spatial variability of AOD

We use the AERONET DRAGON campaign data to assess the spatial variability of AOD on scales similar to the AATSR AOD product grid (~10 km). Similar to the AATSR sampling, we average the AOD of the nearby AERONET sites (within the sampling distance), and calculate the corresponding standard deviation of AOD (σ_{AERO_NEAR}). We can then compare this to the corresponding σ_{AATSR} for each match between AATSR and AERONET during the DRAGON campaign. Initial comparison shows very weak correlation between σ_{AERO_NEAR} and σ_{AATSR} as seen in Figure 6.6 b).

Analyzing the AATSR results in more detail, using a special high resolution retrieval mode, reveals unrealistically high AOD values on the coastlines of the Chesapeake bay (see Figure 6.6 a). Such problems are typical for satellite based aerosol retrieval, and often the coastal areas are removed from AOD products (e.g. for MODIS). In the case of ADV/ASV, this results from the smoothing difference between the two views and the highly varying surface reflectance values between land and water surface. Another issue arises from overestimation of AOD in coastal waters: the ASV algorithm employs an ocean surface reflectance model which is valid over open ocean, but is less suitable for coastal waters with high sediment loads. To remove the coastal areas from the comparison, we apply an upper threshold (0.6) for the standard deviation of AATSR TOA reflectance within the retrieval area. This improves the correlation between σ_{AERO_NEAR} and σ_{AATSR} considerably. The AOD correlation is also improved, while the number of matches between AATSR and AERONET is reduced (not shown).

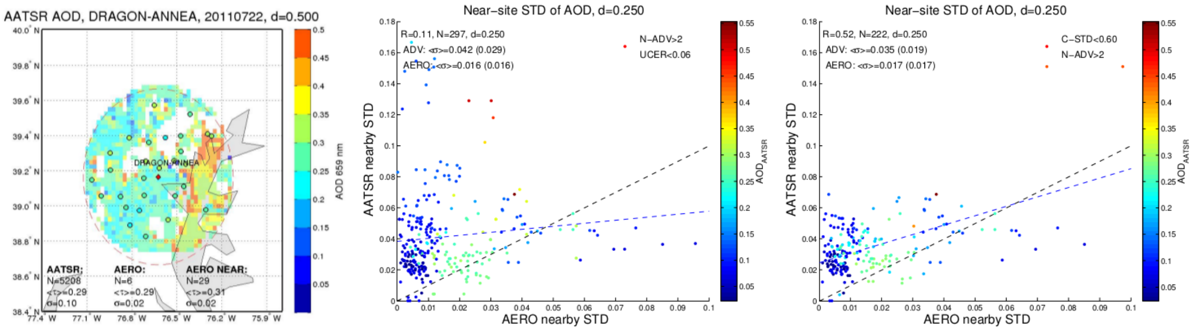


Figure 6.6. a) Illustration of the AOD spatial variability: the background color shows AATSR high resolution AOD values (without cloud post-processing) while the colored circles show the AERONET AOD at the nearby sites. **b)** Scatter plot of the standard deviation of (low res.) AOD within the sampling area for AATSR and for AERONET. **c)** Same as b), but with areas of high surface reflectance variability (coastal areas) removed

6.4 Conclusions on AATSR/AERONET comparison

Work has been done on identifying the main sources of co-location mismatch uncertainties in comparisons of AATSR and AERONET AOD retrievals. Quantitative indicators of these uncertainties have been developed and tested. In addition, the effect of comparison sampling parameters has been assessed.

The average standard deviation of AOD with a sampling distance of 0.25º is 0.042 for AATSR (0.035 when coastal points are excluded) and 0.016 for AERONET. For AERONET, this is of the order of the reported uncertainty of ~0.010 to 0.021 (Eck et al., 1999), and similar to the temporal variation scale within the sampling time window of 0.5 – 1.0 hours. The reported AATSR AOD uncertainty, based on the propagation of instrument uncertainty of 5%, is of the order 0.01 – 0.1 (average 0.064 for this data set). We note that this analysis may be specific to the 2011 DRAGON campaign region and more studies are needed before definite conclusions.

Overall, we conclude that more than one half of the observed AATSR spatial variability in the comparison sampling area is due to satellite retrieval uncertainties, rather than due to actual variability of the aerosol load. However, the standard deviation of AATSR AOD within the sampling area gives an approximate estimate of the co-location mismatch uncertainty between AATSR and AERONET aerosol retrievals.

Inputs to Virtual Observatory

As an input to the Virtual Observatory we provide description of methods that can be used to assess the co-location mismatch uncertainty in comparison of AOD retrievals made by satellite instruments and AERONET sun photometers. The methods are based on the knowledge of satellite product validation against AERONET in general, and on the case studies presented in this chapter. The description includes identification of parameters characterizing the uncertainties and recommendations for the default sampling parameters used in the comparison.

7 Aerosol profiles: comparison of CALIOP/CALIPSO and EARLINET

The high variability both in space and time of tropospheric aerosols is one of the main causes of the high uncertainty about radiative forcing related to tropospheric aerosols and their interactions with clouds (Forster et al., 2007). In particular, information about the vertical layering of aerosol and aerosol vertical distribution is a crucial point for aerosol-clouds interaction study. Moreover, the lack of information about the vertical mixing can lead also to significant horizontal inhomogeneities due to large vertical concentration gradients and it is therefore a large source of variability typically not considered in the models. Since 2006, CALIOP, the LIDAR onboard CALIPSO specifically designed for aerosol and clouds study, is providing high-resolution vertical profiles of aerosols and clouds on a global scale. However, because of the small footprint and the revisit time of 16 days, how well these CALIOP measurements represent the atmospheric conditions of a surrounding area over a longer time is an important issue to be investigated. An integrated study of CALIPSO and EARLINET correlative measurements opens new possibilities for spatial (both horizontal and vertical) and temporal representativeness investigation of this set of satellite measurements.

7.1 CALIOP/CALIPSO description

The NASA/CNES Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission is designed to study aerosols and clouds (Winker et al., 2010). Its aim is to provide profiling information at a global scale for improving our knowledge and understanding of their climatic role. The main instrument, CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization), is a dual wavelength (532 nm and 1064 nm) elastic backscatter LIDAR with the capability of polarization sensitive observations at 532 nm (Winker et al., 2006; Winker et al., 2007). The CALIPSO satellite was launched into a near sun-synchronous (SSO) and low earth orbit (LEO) at a 705-km altitude. Utilizing active remote sensing techniques, CALIOP observes aerosols during daytime and nighttime conditions, and therefore provides constant observations of aerosols and clouds.

In particular, CALIPSO mission offers unprecedented observations of day and night aerosol global optical properties profiles, vital for aerosol-radiation-cloud interaction studies to understand their climatic role (Winker et al., 2013).

Instrument data is transmitted from the satellite to the ground station once per day and transferred to the Level 0 processing facility to packetize, time order, and archive. The instrument data is combined with ancillary data sets such as meteorological, ephemeris, instrument status, and global reference products to enhance the quality and accuracy of the data products. The LIDAR Level 1 data (https://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-201v1.0.pdf) product contains a half orbit (day or night) of calibrated and geolocated LIDAR profiles. Apart from LIDAR data, satellite position data and viewing geometry are provided in the product. There are three types of Lidar Level 2 products: layer products (cloud and aerosol), profile products (backscatter and extinction) and a vertical feature mask (cloud and aerosol locations and the corresponding type). Details of the CALIOP instrument and algorithms can be found in the companion papers of the JTECH special issue (http://journals.ametsoc.org/topic/calipso). Additional details can be found in the CALIPSO algorithm theoretical basis documents (ATBDs; available online at http://www-calipso.larc.nasa.gov/resources/project_documentation.php). The aerosol related data are generated at a uniform horizontal resolution of 5 km. Finally, the Level 3 product reports monthly mean profiles of aerosol optical properties on a uniform spatial grid (Winker et al., 2013). All level 3 parameters are derived from the CALIPSO level 2 5-km Aerosol Profile products applying some additional quality screening filters (Winker et al. 2013).

The retrieval of optical profiles from CALIPSO observations is highly complex and its detailed description is out of the scope of this document. However, the whole procedure could be briefly summarized in the following manner. Aerosol extinction and backscatter coefficients are retrieved in three steps: (1) layers are searched in the LIDAR acquired profiles, with horizontal averaging varying from 1/3 km to 80 km; (2) these layers are flagged as clouds or aerosols; and (3) the aerosol extinction and backscatter profiles are retrieved. The succession of the above mentioned steps are described in detail in a special issue of the Journal of Atmospheric and Oceanic Technology (e.g. Winker et al., 2009).

Note that the above retrieval from CALIOP elastic backscatter LIDAR is underdetermined and an additional assumption is needed. In case of an elevated aerosol layer that lies in clear-air, the transmittance through the layer can be estimated from the clear-air signals (Young, 1995). This offers the needed constraint for the extinction retrieval, however the CALIOP SNR levels do no permit usually the application of the technique. Therefore, an algorithm (Omar et al., 2009) is used to estimate the extinction-to-backscatter ratio from the 532 nm depolarization and backscatter signals, which provides the above mentioned assumption (Young and Vaughan, 2009).

The signal calibration, that precedes the above chain of aerosol retrieval, along with the correct aerosol layer detection and the aerosol layer subtyping dictate the correct retrieval, and any errors in these parameters will lead to errors in the optical properties retrieved by CALIPSO. An extended error analysis of the aerosol extinction and backscatter retrieval can be found in Young and Vaughan (2013).

As only CALIPSO level 2 aerosol profiles are used, these data are described in the following subsections.

Sampling

The CALIOP sampling is dictated by the laser repetition rate, the detection configuration and the satellite-target geometry.

Vertical sampling

The fundamental sampling resolution of the LIDAR is 30 m vertically (https://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-201v1.0.pdf; p. 12).

Horizontal sampling

The fundamental sampling resolution of the LIDAR is 333 m horizontally (https://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-201v1.0.pdf; p. 12).

Temporal sampling

The firing rate of the laser is 20 Hz and according to the minimal resolution of 1/3 km (average of 15 single-shot profiles), leads to a temporal sampling of 0.75 s (https://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-201v1.0.pdf; Table 2.01, p. 9).

Smoothing

The SNR level of the CALIPSO raw signals at sampling could be very low because of many factors: CALIPSO's distance from the target, the high speed at which the LIDAR sweeps across the target space, constraints placed on the pulse energy of the laser transmitter by eye-safety requirements, the relatively low firing rate of the laser (20 Hz) relative to the velocity of the satellite, and vertical and horizontal variations in the composition of the layers being measured. Appropriate procedures are used for the CALIPSO satellite borne aerosol measurements for improving the SNR affecting the smoothing in the vertical, horizontal and temporal dimensions.

Vertical smoothing

There exists a multi-step averaging scheme that dominates the vertical and horizontal resolution. The spatial invariant resolution shown in Table 7.1 below is the resolution applied to raw data already in the on-board averaging scheme.

An altitude dependent averaging scheme is used by CALIPSO and provides higher resolution in the lower troposphere – where the spatial variability of cloud and aerosol is larger – and lower resolution above. The degree of averaging varies with the altitude, as detailed in the mentioned Table 7.1 (see https://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-202.Part1_v2-Overview.pdf). This scheme is performed before the data are downlinked to ground data processing stations and can be regarded as pre-processing.

Altitude Range (km)	Vertical resolution	Horizontal Resolution	Profiles per 5 km	Samples per profile
20.2 to 30.1	180 m	1.7 km	3	55
8.2 to 20.2	60 m	1.0 km	5	200
-0.5 to 8.2	30 m	1/3 km	15	290

Table 7.1. Spatial resolution for the CALIPSO on board averaging scheme (altitudes are with respect to mean sea level).

For the current study, only Level 2 aerosol profile data in the range -0.5-20.2 km are used. For these data the vertical resolution is made homogenous at 60 m by averaging consecutive points in the lower range, -0.5-8.2 km. As a result, we have 145 and 200 samples per profile in the 0.5-8.2 km and 8.2-20.2 km ranges, respectively. No further vertical averaging has been applied within this study to the original level 2 as released by the CALIPSO team following the ATBD reported on the CALIPSO website.

Horizontal smoothing

The CALIPSO algorithms perform a horizontal averaging to enhance the detection of aerosol layers. The averaging is performed for 1/3 km, 1 km, 5 km, 20 km, 80 km (https://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-202.Part1_v2-Overview.pdf).

For the current study, CALIPSO level 2 data are used which are spatially uniform and reported in 5 km segments. Details on how this is achieved are reported at: https://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-202.Part1_v2-Overview.pdf.

For the purpose of the current study, further horizontal averaging was applied on the Level 2 CALIPSO data (see Section9.3). Different horizontal averaging schemes for the CALIPSO data are used (Table 3) in order to investigate the influence of horizontal smoothing of CALIPSO data when compared against EARLINET data. CALIPSO data at different horizontal resolutions are obtained averaging the original 5 km ones, without applying any screening criteria on them.

Temporal smoothing

The varying horizontal averaging, as described in Table 7.2, changes the temporal resolution as well and therefore this could decrease the very high original uncertainties. Table 7.2 also reports the number of laser shots corresponding to the different horizontal smoothing windows.

Level 2 product post-processing

Horizontal averaging [km]	Temporal resolution [s]	Laser shots
25	3.75	75
45	6.75	135
75	11.25	225
105	15.75	315
125	18.75	375
155	23.25	465
175	26.25	525
205	30.75	615

Table 7.2. The horizontal averaging applied to CALIPSO data along with the corresponding temporal sampling and number of laser shots.

7.2 EARLINET description

EARLINET (European Aerosol Research Lidar NETwork) is the first LIDAR network for aerosol studies on continental scale. EARLINET comprises of different instrumental setups, specifics and team expertises. Building up on substantial measurement heritage, EARLINET worked on three main aspects: i) harmonization of the QA procedures (instrumental and algorithm), ii) establishing measurements schedule and iii) creating centralized dataset and homogeneous data format (Pappalardo et al., 2014 and references therein). The success of EARLINET, established in 2000 and currently part of ACTRIS, the European Research Infrastructure for Aerosol Clouds and Trace gases observations, paved the way for a further step in the global LIDAR aerosol monitoring even if starting up from heterogeneous LIDAR networks within the Global Aerosol Lidar Network (GALION) established by WMO.

Sampling

EARLINET is a network of different instruments with a wide variety of instrumental specifics, so that there is not a common vertical and temporal sampling overall within the network. On the other hand, differences in the instrument components result in different signal-to-noise ratio throughout the network, therefore different smoothings are needed station by station. Spatio-temporal resolution as well as sampling are established at station level and varies in a significant way among the network. This aspect could mean a loss of homogeneity in the considered dataset, but on the other hand it provides the opportunity for investigating how the different set-ups affect the EARLINET-CALIPSO comparison.

As illustrated in Table 7.3, EARLINET LIDAR signals considered here are acquired with vertical sampling between 3.75 and 60 m. Moreover, temporal sampling is characterized by the fact that, each LIDAR signal is acquired over a temporal window between 10 and 60s. Of course, horizontal sampling, can be considered as pointwise measurements.

Station	Lidar Name	Vertical sampling (m)	Temporal sampling (s)
Evora	Paoli	30	60
Granada	Raymetrics D400	7.5	10
Leipzig	Martha	60	30
Napoli	-	15	60
Potenza	MUSA	3.75	60

Table 7.3. Vertical and horizontal sampling of the EARLINET LIDAR systems considered in this study.

Smoothing

Vertical smoothing

After the LIDAR signal acquisition, the signals are integrated, thus modifying both the vertical and temporal resolution. Vertical resolution is decided at station level with the goal to improve the signal-to-noise (SNR) levels. It can be variable with the altitude range: typically a finer vertical resolution is set in the lowest altitude range where the aerosol load is high and a coarser one at the upper levels where the aerosol load is low for improving the SNR.

The aerosol extinction retrieval is numerically more complex with respect to aerosol backscatter because it involves the derivative of the signal. This complexity results in a coarser resolution. Even if the vertical extinction profiles are provided to the original raw resolution (typically 15m) the effective resolution is coarser: each point is provided because the ensemble of these “not-independent” points provides a better reconstruction of the real atmospheric feature (exactly as happens for image processing). The effective resolution is evaluated adopting interferometric criteria for peak discrimination (Iarlori et al., AMT 2015).

For the EARLINET stations, typically the resolution for the aerosol backscatter is 60 m, while the resolution typically ranges between 200-600 m for aerosol extinction, reaching values of 1.2-1.5 km at the highest altitude ranges when no aerosol layers are identified.

Temporal smoothing

After the LIDAR signal acquisition, the signals are integrated, thus modifying both the vertical and temporal resolution. With regards to the temporal resolution, the signals are averaged for increasing the SNR in such a way to cover the widest altitude range possible. Typically signals are averaged for about 30 min – 1 hour in homogeneous aerosol load conditions. However, this resolution depends also on the aerosol content: low aerosol content means low signal and therefore a longer temporal integration time is needed for obtaining high SNR. Table 7.4 reports different characteristic of LIDAR instruments for the different stations in the case study analyzed.

Station	Lidar Name	Laser Repetition Rate (Hz)	Laser shots
			Time window: 30 min – 60 min
Evora	Paoli	20	36000-72000
Granada	Raymetrics D400	10	18000-36000
Leipzig	Martha	30	~54000-108000
Napoli	-	20	36000-72000
Potenza	MUSA	20	36000-72000

Table 7.4. Repetition rate of the EARLINET LIDAR systems considered in this study and laser shots within a time period of 30min-1h

7.3 Comparison setup

Since June 2006, many EARLINET stations are providing measurements in correspondence to CALIPSO overpasses within 100 km (Pappalardo et al., 2010), according to CALIPSO validation plans. Additionally, simultaneous measurements are planned in order to study the aerosol temporal variability, or in the case of special events to study specific aerosol types and to investigate the geographical representativeness of the observations (Pappalardo et al., 2010). The measurement schedule is centrally distributed among the stations and measurements are performed under weather favorable conditions and conditioned by the station's manpower availability. Up to July 2016, the EARLINET database reports more than 9000 files related to CALIPSO overpasses (EARLINET publishing group 2014; https://data.earlinet.org). In the following, we consider only files related to overpasses within a 100 km radius from the station and excluding special events (Case B and Case C described in Pappalardo et al., 2010). In particular, 143 aerosol backscatter coefficient profiles are compared against their CALIPSO counterparts. This parameter has been selected for investigating the difference balance between the two observations because it was demonstrated that, among the CALIPSO optical properties, the aerosol backscatter is less affected by the inversion assumptions (Papagiannopoulos et al., 2016). Further only nighttime measurements are considered because of the larger calibration uncertainty for daytime CALIPSO measurements and because of the EARLINET capability of independently measuring the aerosol backscatter and extinction coefficients.

Regarding CALIPSO extinction coefficient, Section 7.1 highlighted that the retrieval is computed using an assumption on LIDAR ratio. Although It has been shown (e.g., Schuster et al., 2012; Papagiannopoulos et al., 2015; Amiridis et al., 2013) that this assumption may have an impact on EARLINET-CALIPSO comparison, this point is not considered in this deliverable, but it could be addressed by further research.

The comparison of EARLINET profiles and their CALIPSO counterpart is a straightforward procedure. Both EARLINET and CALIPSO make use of active remote sensing instruments, yet, the nature and the needs of the satellite mission require special care for any validation study. EARLINET is performing correlative measurements since CALIPSO started its life cycle (April 2006), based on a schedule established before the satellite mission. The strategy followed by the member stations is as follows: the observations occur during the satellite overflight within 100 km distance of the satellite ground-track from the station, and are performed for at least 60 min.

Only CALIPSO measurements synchronous to the EARLINET measurements are used here. The CALIPSO data are searched for the closest in distance point. This point corresponds to a 5-km CALIPSO profile, however, more 5-km CALIPSO profiles are also averaged in order to assess the spatio-temporal satellite's performance. Apart from the original 5-km profile further eight horizontal resolutions are used: 25-km, 45-km, 75-km, 105-km, 125-km, 155-km, 175-km and 205-km.

For example, in Figure 7.1, the "pin" symbol represents EARLINET Potenza site, while the white thick line, which is 46.7 km long, points to the center of the 5-km ground track for the CALIPSO overflight on 24/08/2010. Finally, the nine blue dots indicate the profiles used then for the calculation of the 45-km average CALIPSO profile.

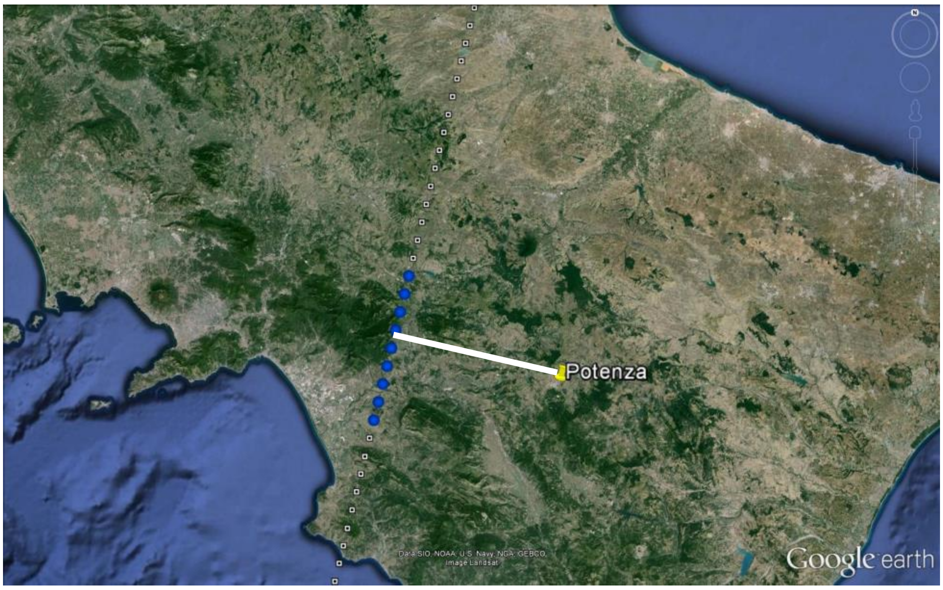


Figure 7.1. CALIPSO overpass on 24/08/2010 with a minimum distance of 46.7 km from EARLINET Potenza station.

To investigate dependence on the specific site, only the stations with a large enough number of co-located observations are considered and analyzed. Namely the stations are: Évora, Granada, Napoli, Potenza, Leipzig giving 19, 21, 40, 37, and 26 co-located observations, respectively. Prior to the aforementioned analysis, the following identified cirrus cases have been screened out from EARLINET data (CALIPSO aerosol data are already screened in this sense). The cloud screening of EARLINET data is not an automatic procedure for the current version of the database. This was done for this work taking advantage of the available labeling of the data. The cirrus category and the reported information about the cirrus cloud altitude range found in the comment field of the data file were used. However, in some cases manual inspection of the data was needed for identifying the cirrus cloud in the profiles, since the inclusion of this information within the EARLINET file is not mandatory at this stage. As a result the data reported in Table 7.5 have been removed from the analysis. In addition, the 27 uncertainties above 6923 m that resulted to be greater than 1 $m^{-1}sr^{-1}$ at Granada have been also removed from subsequent uncertainty analyses.

Station	Date	Altitude range
Naples – IT	19 set 2008 17 mag 2009 24 nov 2009 18 apr 2010	above 6.5km above 6 km above 8km above 5.4km
Potenza – IT	03 apr 2007 21 apr 2008 14 nov 2008 27 apr 2010	at 8-9.5km above 7.5 km above 7 km above 8 km

Table 7.5. Cirrus clouds detected and discarded in EARLINET stations involved in this study.

In order to homogenize the EARLINET aerosol profiles in terms of altitude levels (i.e. same altitude points for all the profiles provided by the same stations), an interpolation on two points has been applied to Évora profiles. In fact this station changed instrument configuration during the considered period so that an adaptation is needed to homogenize the data from the two periods. Finally, aerosol backscatter data from the station of Naples have been reconstructed to match the altitude points of the corresponding extinction profiles in order to allow in a second step to investigate the differences of the LIDAR ratio obtained from EARLINET and CALIPSO (i.e. EARLINET LIDAR ratio). The Naples aerosol backscatter profiles vertical resolution has been modified for the backscatter to fit the coarser extinction resolution. At each extinction altitude, it has been associated to the value of backscatter which is the closest in altitude (the difference is always within the vertical effective resolution).

The table below reports the data available from the original CALIPSO hdf and EARLINET netCDF files freely available at respective databases: https://eosweb.larc.nasa.gov/HBDOCS/langley_web_tool.html, and https://www.earlinet.org. Only the information considered relevant for the purposes of the work are extracted and reported in comma-separated-values (csv) file format.

EARLINET	CALIPSO
Altitude a.s.l.	Altitude a.s.l.
Aerosol Backscatter [$m^{-1}sr^{-1}$]	Aerosol Backscatter [$m^{-1}sr^{-1}$]
Aerosol Extinction [m^{-1}]	Aerosol Extinction [m^{-1}]
Station Latitude and Longitude [°]	Latitude and Longitude central point [°]
Start and Stop time of temporal smoothing [UT]	Time of central point [UT]
Aerosol Backscatter Uncertainty [$m^{-1}sr^{-1}$]	Aerosol Backscatter Uncertainty [$m^{-1}sr^{-1}$]
Aerosol Extinction Uncertainty [m^{-1}]	Aerosol Extinction Uncertainty [m^{-1}]

Table 7.6. Relevant parameters for this study as delivered from EARLINET and its CALIPSO counterpart.

7.4 Analysis and results

As reported above, the performed analysis is done on the backscatter variable, because it is the CALIPSO product less affected by retrieval assumptions. From equation (2) in Section 1.1 the uncertainty term σ^2 depends on the horizontal and temporal mismatch error between x_{sat} and x_{gnd}, the satellite and ground backscatter respectively. In order to investigate how the horizontal smoothing impact on the term σ^2 of the uncertainty budget, we consider the root mean squared error (*RMSE*), which is defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{l}(x_{sat} - x_{gnd})^2}.$$

In the above formula l is a multi-index $l = (h, s, d)$, where *h* indicates the altitude, *d* the day and *s* the measuring ground station; moreover N is the number of available co-located measurements.

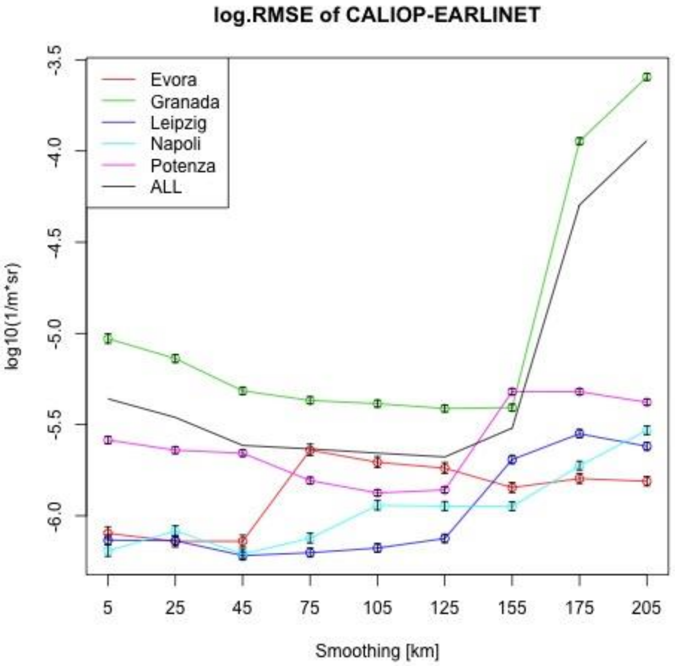


Figure 7.2. Co-location uncertainty (log.RMSE) for CALIOP and EARLINET backscatter mismatch by station and smoothing parameter. Vertical bars represent the 95% error intervals.

In this case study, we have *200* altitude levels*,* corresponding to CALIOP observations. The vertical range is 97.67 -12013 m, with a step of 60 m. The counter *s=1,…,5*, identifies the five EARLINET stations (Évora, Granada, Leipzig, Napoli and Potenza) and *d=1,…,*g_s, identifies the profile while g_s gives the number of profiles for station *s*. In total, the analysis considers 143 EARLINET profiles.

To understand how the co-location error depends on the horizontal smoothing, we have computed the *RMSE* of the eight different horizontal averaging schemes for CALIOP described in Table 7.2, which includes also the original CALIPSO 5-km data.

In Figure 7.2 and Table 7.7, the co-location uncertainty averaged by station and CALIOP horizontal smoothing is presented in order to understand how smoothing affects the comparison. The yellow highlighted values correspond to station minima. The minimal co-location uncertainty is obtained at 45 km for Évora, Leipzig and Napoli, which have small uncertainties already at 5km. This suggests that using 45 km as standard horizontal averaging is advisable for the comparison of pointwise ground based measurements with the CALIOP level 2 product. At Potenza and Granada instead the minimal co-location uncertainty is obtained for 105 and 155 km respectively. This different behavior of Granada and Potenza can be ascribed to the variant orography that affects the atmosphere sampled by the satellite LIDAR: compared to the other sites in fact, Granada and Potenza are the unique ones located at upper altitudes and surrounded by different areas (Alados-Arboledas 2013, Mona et al., 2009). From Figure 7.3, we see also that for all the stations the co-location uncertainty increases in the tails of smoothing range that is 175-205 km and 5-25 km. Figures 7.2 and 7.3 indicate that the uncertainty in Granada is larger than in all the other stations, with the exception of Potenza at 105 km smoothing, where however the uncertainties are comparable. This underlines the peculiarity of Granada comparison: in this case the co-location uncertainty is higher than all the other stations, independently from the horizontal smoothing. Granada station is located in a natural basin surrounded by mountains with the highest mountain range located to the southeast, with altitudes above 3000 m (Guerrero-Rascado et al., 2008). The presence of this mountains can act as a boundary for both local and free troposphere aerosol layers depending on the specific aerosol source, so that if the CALIOP track location is not favorable, very large differences are expected when compared to the ground-based measurements. Finally both Figure 7.2 and 7.3 report the mean co-location uncertainty, which has the same behavior of Granada values, because it is driven by these very large values and therefore cannot be regarded as representative of the ensemble of the considered locations.

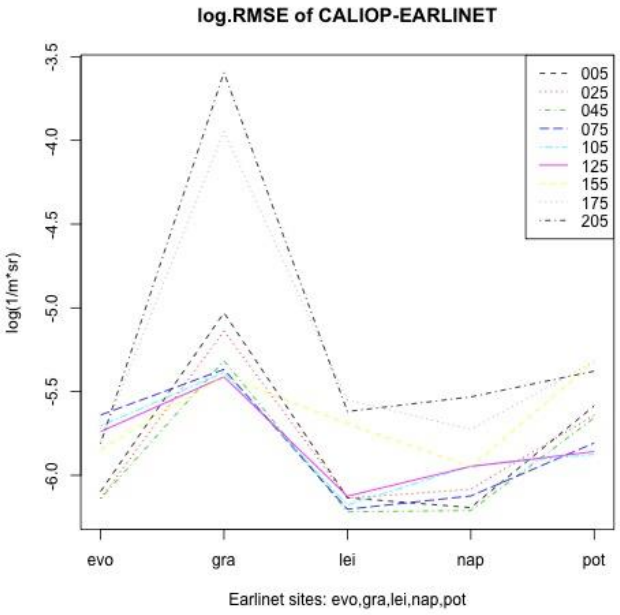


Figure 7.3. Co-location uncertainty (log.RMSE) for CALIOP and EARLINET backscatter mismatch by station and smoothing parameter.

	Horizontal smoothing (km)								
	5	25	45	75	105	125	155	175	205
Evora	8.01E-07	7.25E-07	7.25E-07	2.29E-06	1.97E-06	1.83E-06	1.43E-06	1.60E-06	1.55E-06
Granada	9.36E-06	7.28E-06	4.84E-06	4.29E-06	4.12E-06	3.88E-06	3.92E-06	1.13E-04	2.55E-04
Leipzig	7.35E-07	7.31E-07	6.05E-07	6.28E-07	6.65E-07	7.51E-07	2.03E-06	2.82E-06	2.41E-06
Napoli	6.42E-07	8.26E-07	6.17E-07	7.53E-07	1.14E-06	1.13E-06	1.13E-06	1.88E-06	2.94E-06
Potenza	2.60E-06	2.29E-06	2.21E-06	1.56E-06	1.33E-06	1.39E-06	4.79E-06	4.79E-06	4.19E-06
All stations	4.38E-06	3.46E-06	2.43E-06	2.33E-06	2.21E-06	2.10E-06	3.03E-06	5.06E-05	1.14E-04

Table 7.7. Co-location uncertainty (RMSE) by smoothing and Station [1/mr]. Yellow color highlights the station minima.

In order to better understand these results on co-location uncertainty, the measurement uncertainty has been investigated. In Figure 7.4, the solid horizontal lines represent the averaged measurement uncertainty of EARLINET while the dashed lines represent the uncertainties of CALIOP smoothed backscatter. As expected, the uncertainty of ground measurements is smaller than the satellite for all stations. Moreover, in Table 7.8, the mean of the measurement uncertainties by station and smoothing are reported, while in Table 7.9, the relative measurement uncertainty (coefficient of variation) for the backscatter values is reported.

Generally speaking, the satellite uncertainty is larger or equal to ground uncertainty. Granada has the largest measurement uncertainty, for both EARLINET and CALIOP. This appears also from Table 7.8 where Granada has the largest values of relative uncertainty both for EARLINET and CALIOP when averaged over smoothing. Comparing Figures 7.2 and 7.3 we can conclude that horizontal smoothing affects in a different way measurement uncertainty and co-location uncertainty.

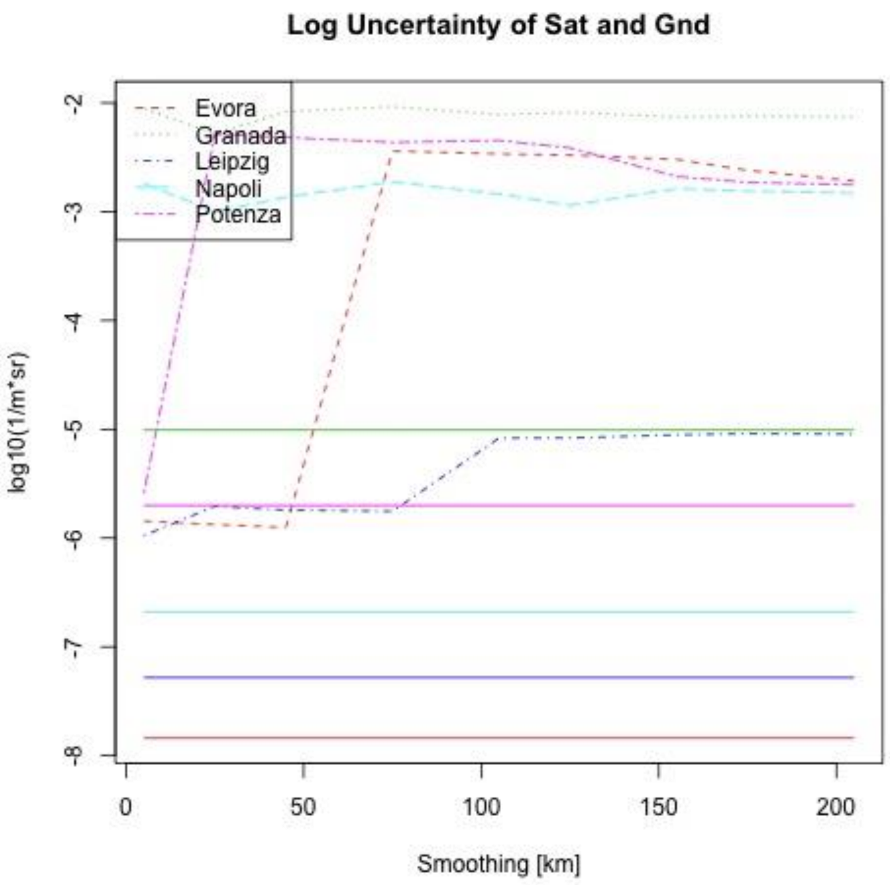


Figure 7.4. Measurement uncertainty by station. Dashed lines: CALIOP uncertainty averaged by smoothing parameter. Solid lines: EARLINET uncertainties.

	Earlinet	Caliop								
		Horizontal smoothing (km)								
		5	25	45	75	105	125	155	175	205
Evora	1.45E-08	1.43E-06	1.33E-06	1.25E-06	3.60E-03	3.40E-03	3.32E-03	3.02E-03	2.40E-03	1.92E-03
Granada	9.88E-06	8.90E-03	5.18E-03	8.30E-03	9.24E-03	7.78E-03	8.16E-03	7.41E-03	7.57E-03	7.44E-03
Leipzig	5.23E-08	1.05E-06	1.96E-06	1.81E-06	1.77E-06	8.29E-06	8.34E-06	8.88E-06	9.12E-06	9.02E-06
Napoli	2.09E-07	1.82E-03	1.00E-03	1.36E-03	1.88E-03	1.45E-03	1.15E-03	1.63E-03	1.55E-03	1.48E-03
Potenza	1.99E-06	2.58E-06	5.14E-03	4.82E-03	4.34E-03	4.52E-03	3.86E-03	2.10E-03	1.86E-03	1.76E-03

Table 7.8. Average measurement uncertainties for ground and smoothed satellite backscatter [1/msr].

	Earlinet	**Caliop**								
		Horizontal smoothing (km)								
		5	25	45	75	105	125	155	175	205
Evora	0.06	1.07	1.15	1.27	2106.5	2199.5	2247.1	2578.7	2273.1	2010.3
Granada	8.14	1422.1	1270.2	3325.4	4591.1	4897.1	4654.4	4381.3	1174.6	592.62
Leipzig	0.18	1.16	2.27	2.33	2.22	5.95	5.84	4.78	4.44	4.48
Napoli	0.59	1794.7	1031.5	1331.6	1674.1	1410.4	1278.1	1863.4	1741.8	1479.4
Potenza	5.06	1.87	2547.3	2499.9	2196.3	2237.2	2117.8	1468.9	1335.3	1327.9

Table 7.9. Relative measurement uncertainty of ground and smoothed satellite backscatter.

All values reported in the analysis until now are related to the whole column: uncertainties related to altitude regions with high aerosol content are mixed with altitude regions where the aerosol content is very low and uncertainties are expected to be high because of a low SNR. To take into account the differences in the vertical dimensions and to exploit the vertical profiling capability of both EARLINET and CALIPSO, we split the atmosphere into three zones: below 2.5 km (as representative of local aerosol conditions), between 2.5 and 5.5 km (middle troposphere with transport of aerosols) and above 5.5 km (free troposphere).

Figure 7.5 (left panel) corresponds to the lowest part of the troposphere, that is, the planetary boundary layer, where normally, lower agreement is expected between the ground station and the satellite swath. All the values are grouped around -6.0 $\log_{10}$ ($m^{-1}sr^{-1}$), with the exception of Granada probably because of the already mentioned variant topography surrounding this EARLINET station (Alados-Arboledas et al., 2013; Papagiannopoulos et al., 2016). Regarding Potenza station, there is a decrease in the RMSE for smoothing up to 100 km whilst it is approximately constant and comparable with Évora, Leipzig and Napoli stations for larger smoothing. This effect could be attributed on one hand to the mountainous area where Potenza is situated and, on the other hand, to the variety of surfaces that the satellite encounters (e.g land, sea). In particular, Potenza is on a mountain close to the sea but also to big cities, so the difference with smaller horizontal smoothing can be larger (e.g. CALIPSO ground track lies over sea) than at higher smoothing where smoothing procedure merges different conditions (comparing EARLINET mountain sampling versus an average of sea, cities and mountain). This finding is in agreement with the smaller discrepancies observed in the PBL at Potenza between EARLINET and corresponding CALIPSO observations for the overpasses at about 80km distance respect to the closer overpasses at 40km distance because of topographic and local effects (Mona et al., 2009).

The opposite behaviour is found for Leipzig because of the more homogeneous topography. For the remaining stations the RMSE is lower at original horizontal smoothing and gradually increases with increasing smoothing. Generally, the increasing smoothing tends to increase the RMSE as the satellite contains a vast geographic area and therefore the aerosol fields can be dramatically different. This effect has been documented and reported in several studies (Anderson et al., 2003; Pappalardo et al., 2010).

The mid-troposhere plot (Figure 7.5; middle panel) corresponds to the height range typically free from local sources and indicates the transboundary motion of aerosol, and therefore this plot is the most relevant to assess the CALIPSO representativeness. At first sight the lines are contained in the range between -6.5 and -5.0 $\log_{10}$ ($m^{-1}sr^{-1}$) which shows a better agreement with respect to the previous figure implying that the effect of local sources is greatly reduced, especially at Granada. For the first 50 km and for the stations of Napoli, Leipzig, and Évora the RMSE is the lowest and then either gradually or steeply increases for increasing smoothing. On the other hand, for the station of Potenza the situation is reversed in the range 0-50 km and behind that it shows a behavior which is similar to the other locations. For Granada, the situation is more complex as the RMSE peaks at 25 km and then decreases until 75 km, to follow for the next smoothing ranges the behavior of the other stations.

The free troposphere behaviour of Figure 7.5 (right panel) corresponds to predominantly aerosol-free area without large variations of RMSE w.r.t. smoothing. As a case in point, Évora, Potenza and Leipzig produce a constant value for the whole smoothing range, showing that in free troposphere the smoothing parameter loses its importance in explaining the co-location error. We observe a finer structure at Granada and Napoli that can be attributed to aerosol structures not observed by either instruments or in case of CALIPSO cloud misclassification (e.g. subvisual, thin cirrus clouds). These opaque clouds in the higher altitude levels frequently penetrate the CALIOP aerosol retrievals and alter the CALIPSO provided atmospheric description (Huang et al., 2011; 2012; 2013; Kittaka et al., 2011).

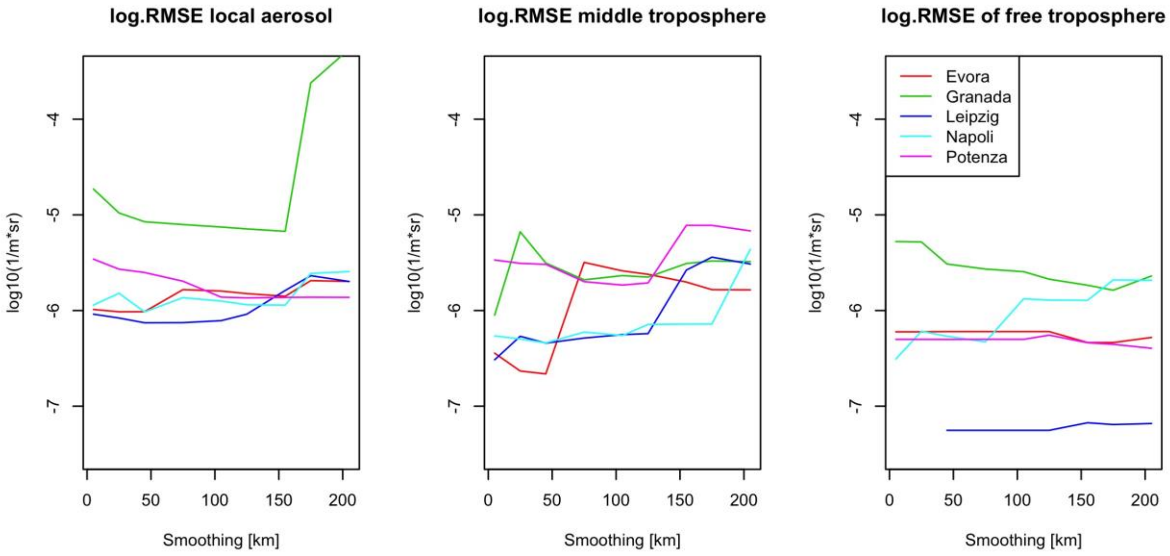


Figure 7.5. RMSE for CALIOP and EARLINET measure of aerosol backscatter at different smoothing parameter for different stations and for different zone of atmosphere.

7.5 Conclusions and inputs to Virtual Observatory

A first effort has been undertaken towards identifying the main contributions of co-location mismatch uncertainties in comparisons of CALIOP and EARLINET aerosol backscatter profiles. The comparison is not trivial because of two main reasons: 1) the small footprint of CALIPSO measurements comparing to the distance from EARLINET sites and 2) the high uncertainty of CALIPSO products (Young and Vaughan, 2013). The comparison is then even more complex because of the fine vertical structure of the aerosol field and its variability.

Co-location mismatch has been investigated as a function of the observational site and of the horizontal smoothing of the CALIOP data. Furthermore, the investigated altitude range [90m-12km asl] has been splitted in three regions corresponding to the PBL, the mid-troposphere and the free troposphere range.

The co-location mismatch decreases as altitude increases: 10^{-6}-$5 \cdot 10^{-5}m^{-1}sr^{-1}$ in the PBL, $5 \cdot 10^{-7}$-$5 \cdot 10^{-6}m^{-1}sr^{-1}$ in the middle troposphere and typically lower than $10^{-6}m^{-1}sr^{-1}$ in the free troposphere. An influence of the smoothing on co-location mismatch is found for the two lowest atmospheric ranges, while in the free troposphere its influence can be disregarded. This shows that above 5.5 km, LIDAR pointwise measurements can be typically considered representative even over horizontal scale as large as 200 km, because of the low variability of the aerosol field at these altitudes. In the middle troposphere the LIDAR data are representative for distances up to 100 km in agreement with Pappalardo et al. (2010). Finally in the lowest troposphere where the orography and local source play a relevant role, the representativeness strongly depends on the site characteristics. Typically the co-location mismatch has its minimum around 50 km, while for peculiar situations (mountain or region surrounded by mountains), this minimum is shifted around 100-150km.

Inputs to Virtual Observatory

The "Virtual Observatory" (VO) is being developed within WP5 to serve as a proof-of-concept facility in GAIA-CLIM. In the LIDAR framework described above, a tool will be offered to the end users through the VO, which implements an independent method able to co-locate LIDAR retrievals from EARLINET to the best horizontally smoothed CALIOP retrieval, for sites located in Évora, Granada, Napoli, Potenza and Leipzig.

To do this, datasets may be provided either in the original formats, namely hdf for CALIPSO and netCDF for EARLINET, or in csv format. In particular these may include original data, namely measurements and their uncertainties, and the estimated co-location uncertainties plus the optimally horizontally smoothed CALIOP retrieval (presently in csv format).

8 Conclusions and prospects

The metrology of an atmospheric measurement contains several aspects that are crucial when assuring the quality of the data and assessing their fitness-for-purpose with respect to user requirements, in particular those requirements expressed by the Copernicus Climate Change and Atmosphere Monitoring services (C3S and CAMS). In this context, quality assurance of satellite data sets involves comparison with ground-based reference measurements in order to obtain an independent indicator of the quality of the data and their uncertainties. These comparisons are affected by additional errors and uncertainties due to spatiotemporal co-location mismatch: different instruments and measurement methods have different sampling and smoothing properties, in the horizontal, vertical, and temporal domain. As the atmosphere is variable and inhomogeneous on a variety of scales, these differences can impact significantly the comparison results, depending on the co-location criteria and harmonization methods that were used to minimize or limit spatiotemporal mismatch.

D3.2 considered generic metrology aspects of an atmospheric data comparisons. Along those lines, Task 3.2 and Sections 3-7 above, developed a number of case studies of data comparisons covering temperature and humidity profiles, ozone profiles and aerosol, both columnar and profile.

The methods used here range from the classical statistical approach to explicit physic simulation, going through advanced statistical methods. The first simpler approach is essentially based on means, standard deviations and correlations or similar statistics capable to describe central tendency, spread and association. The latter approach is based on advanced statistical methods and includes functional data approach; this has the advantage of handling each entire atmospheric profile as a single (functional) data object, which is represented by splines or other basis functions. Moreover functional regression and isotonic regression allow us to understand relationship involving errors and or uncertainties. Finally physics simulation is based on OSSSMOSE software and allows to simulate realistic spatio-temporal atmospheric field at individual measurement level. Although this approach is clearly very powerful, it must be noted that the OSSSMOSE simulations are inherently limited in their representation of small-scale variability by the resolution of the underlying model/reanalysis fields. As such, it may be that errors and uncertainties are somewhat underestimated, in particular in the lowermost atmospheric layers and when referencing against “singular” snap shots.

This deliverable shows also that the work progress of Task 3.2 WP3 is in a very advanced stage and only little “refinements” are still to be completed. So that actions in Task 3.3 (WP3 results implementation in WP5) and inputs to VO may be developed.

To exemplify these conclusions an excerpt of the main results is summarized below, showing that Task 3.2 is made by a balanced blend of new methodological proposals and consolidated analysis techniques.

Satellite-radiosonde comparison of temperature and humidity

The comparison of temperature and humidity profiles from IASI mission and RAOB network has been considered in Central Europe (C-EU) using the NOAA’s NPROVS dataset, which is considered the major co-location software available. The aim was to understand the contribution of both the vertical smoothing of IASI and the spatio-temporal mismatch to the co-location uncertainty.

As a preliminary important step, the information content of RAOB data, based on Vaisala processing, has been assessed w.r.t. to vertical sampling using GRUAN data as high vertical resolution reference data based on GRUAN processing. This resulted in computing sparseness uncertainty and Vaisala-GRUAN processing uncertainty, which are useful to understand the behavior of total co-location uncertainty.

For example, considering temperature, Vaisala-GRUAN processing uncertainty resulted to be nearly constant, about 0.12 K up to 300 hPa and showed an increase, up to 0.39 K in the lower stratosphere, possibly due to solar radiation. After correction for Vaisala-GRUAN processing uncertainty, vertical sparseness uncertainty of RAOB profiles resulted to be smaller than 0.48 K below 300 hPa, but showed an increase to about 1 K in lower stratosphere.

Harmonization of RAOB and IASI has been made using a novel approach, independent on the IASI average kernels which were not available in NPROVS dataset. It is based on a flexible GEV weight function which is adapted to the data, and has the capability to estimate the vertical smoothing even in absence of detailed information about averaging kernels used for the satellite retrieval. As a result, vertical smoothing uncertainty, in comparison to total and co-location uncertainties resulted to be roughly constant, about 0.5 K, and slowly decreasing up to 300 hPa; above, it showed an increase more marked than co-location uncertainty.

The spatio-temporal mismatch uncertainty was computed using an isotonic regression approach applied to the co-location uncertainty adjusted for vertical smoothing and considered as a function of air distance and co-location delay. As a result, overall in C-EU, co-location uncertainty resulted to be smaller than 1.5 K, except near ground level where it reached 1.8 K. But large uncertainties (3.0-3.3K) have been observed when both air distance and delay are large.

Note that, although the algorithm used here considered a time invariant vertical smoothing, in the version under development for the Virtual Observatory of WP5, a dynamic procedure which covers seasonal and daily atmospheric variability will be used. As a consequence, the results of this D3.4 are expected to give an upper limit for the adjusted co-location uncertainty which will be made more tight in Virtual Observatory.

Radiosonde temporal mismatch of temperature and humidity

The work on temporal mismatch uncertainties among radiosonde profiles provided a direct means of estimating the temporal mismatch uncertainty in temperature as a function of altitude, season and time of day, for the selected sites using ERA Interim model data. This approach has been validated by comparing the ERA Interim results with GRUAN-processed radiosonde data from those sites where long-term high frequency data is available. Analysis of the results have shown that the differences between the radiosonde and ERA-Interim temperature data can be explained by the expected uncertainties in the different data sources.

Ozonesonde balloon drift

Ozone co-location case studies have been addressed using physics simulation system OSSSMOSE.

In the first case, considering ozonsondes, latest generation MERRA-2 and the MACC IFS-MOZART reanalysis have been used. According to data of Paramaribo, Suriname and Hohenpeißenberg, Germany, the errors due to balloon drift resulted strongly dependent on launch location. When the drift distances are small and the ozone field relatively stable, e.g. at Paramaribo, the simulated errors remain below typical measurement uncertainties, which are estimated to be of the order of 3-5%.

Indeed, when standard operating procedures are followed, the three most commonly used sonde types produce consistent results between the tropopause and 28 km, with biases smaller than 5% and precisions better than 3%. At higher and lower altitudes the data quality degrades somewhat, and the differences between the sonde types become more clear. Overall, ECC-type sondes perform best with a bias of 5–7% and a precision of 3–5% in the troposphere. Consequently, in this case, the errors due to balloon drift also remain below the accuracy targets for the use of ozonesondes as reference data in satellite validation work. On the other hand, when drift distances are >100km and the ozone field is inhomogeneous, the errors do become comparable to the measurement uncertainty and will contribute to the error budget of a comparison with other measurements when not minimized, e.g. by doing a height-dependent co-location, or taken into account by adding an additional uncertainty term in the consistency test (σ in Eq. 1.1).

Ozone profiles: difference between zenith and actual line-of-sight

Micro-wave radiometer (MWR) measurements of the vertical ozone profile are obtained using a line-of-sight (LOS) with a low elevation angle. This leads to a significant displacement of the location of actual sensitivity w.r.t the instrument location. From an OSSSMOSE simulation, it was found that the resulting differences w.r.t. a zenith measurement at the instrument location are largest below 30km and above 65km. Instead of associating a measurement with the instrument location, it is shown that it may be better to use an “effective” location, corresponding to the projection on the ground of the 45km altitude level following the LOS.

Ozone profiles from LIDAR: the effect of integration time

This case study investigated to what extent the integration time used for LIDAR measurements of the vertical O3 profile in the stratosphere may affect the accuracy of the measurement and its representativeness. In general, the OSSSMOSE simulations indicated that the temporal smoothing errors are actually rather limited. This is due to the mostly linear variation in the ozone concentrations at the time scale of an integration. Only for the longest integration times and in variable atmospheric conditions do the smoothing errors become comparable to the measurement uncertainty (several percent). The representativeness uncertainty on the other hand is significantly larger and can dominate the measurement uncertainty in the lower stratosphere. This is of particular importance when LIDAR measurements are co-located with other measurements that fall within the integration time, but could not agree within their measurement uncertainties.

Total aerosol comparison

Considering column aerosol optical depth (AOD) over the Chesapeake bay area, USA, the Pearson’s correlation coefficient has been used to assess co-location agreement of satellite (AATSR) and AERONET AOD retrievals. It has been shown that the correlation decreases when the spatial variability around the aeronet site increases. Overall, it was concluded that more than one half of the observed AATSR spatial variability in the comparison sampling area is due to satellite retrieval uncertainties, rather than due to actual variability of the aerosol load. However, the standard deviation of AATSR

AOD within the sampling area gives an approximate estimate of the co-location mismatch uncertainty between AATSR and AERONET aerosol retrievals.

Aerosol profile comparison

Considering LIDAR aerosol backscatter profiles from satellite (CALIOP/CALIPSO) and EARLINET, co-location mismatch has been investigated as a function of the observational site and of CALIOP horizontal smoothing. To do this the horizontal smoothing level has been changed from 5 to 205 km along the satellite trajectory and the optimal horizontal smoothing has been computed for each site and atmospheric region. This approach showed that above 5.5 km, LIDAR pointwise measurements can be typically considered representative even over horizontal scale as large as 200 km, because of the low variability of the aerosol field at these altitudes. In the middle troposphere the LIDAR data resulted to be representative for distances up to 100 km. Finally in the lowest troposphere where the orography and local source play a relevant role, the representativeness strongly depends on the site characteristics. Typically the co-location mismatch has its minimum around 50 km, while for peculiar situations (mountain or region surrounded by mountains), this minimum is shifted around 100-150km.

References and Annex abbreviations

Alados-Arboledas, L., Muller, D., Guerrero-Rascado, J.L., Navas-Guzman, F., Perez-Ramirez, D., and Olmo, F.J.: Optical and microphysical properties of fresh biomass burning aerosol retrieved by Raman LIDAR, and star-and sun-photometry, Geophys. Res. Lett., 38, doi:10.1029/2010GL045999, 2011.

Amiridis, V., Wandinger, U., Marinou, E., Giannakaki, E., Tsekeri, A., Basart, S., Kazadzis, S., Gkikas, A., Taylor, M., Baldasano, J., and Ansmann, A.: Optimizing CALIPSO Saharan dust retrievals, Atmos. Chem. Phys., 13, 12089-12106, doi:10.5194/acp-13-12089-2013, 2013

Anderson, T., Charlson, R., Winker,D., Ogren, J., and Holmén, K.: Mesoscale Variations of Tropospheric Aerosols. J. Atmos. Sci., 60, 119–136, doi: 10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2, 2003.

Blumstein D. Chalon G., Carlier T., Buil C. Hebert P., et al.: "IASI instrument: technical overview and measured performances", Proc. SPIE 5543, Infrared Spaceborne Remote Sensing XII, 196 (November 4, 2004); doi:10.1117/12.560907; http://dx.doi.org/10.1117/12.560907.

Bojinski, S., M. Vertraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp, 2014: The concept of essential climate variables in support of climate research, applications, and policy. Bull. Amer. Meteor. Soc., 95, 1431–1443, doi:10.1175/BAMS-D-13-00047.1.

Brewer, A. W. and Milford, J. R.: The Oxford-Kew Ozone Sonde, P. R. Soc. London, 256, 470–495, doi:10.1098/rspa.1960.0120, 1960.

Butterfield, D. and Gardiner, T, 2015: Determining the temporal variability in atmospheric temperature profiles measured using radiosondes and assessment of correction factors for different launch schedules; Atmos. Meas. Tech., 8, 463–470.

Calbet, X., Peinado-Galan, N., Ripodas, P., Trent, T., Dirksen, R., and Sommer, M.: Consistency between GRUAN sondes, LBLRTM and IASI, Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-344, in review, 2016.

Catmull, E., Rom, R. (1974), "A class of local interpolating splines", in Barnhill, R. E.; Riesenfeld, R. F., Computer Aided Geometric Design, New York: Academic Press, pp. 317–326.

CDRH. Center for Devices and Radiological Health (CDRH), General Principles of Software Validation; Final Guidance for Industry and FDA Staff, CBER CDRH/OC Doc. N. 938, January 11, 2002. Publicly available via http://www.fda.gov/RegulatoryInformation/Guidances/ucm085281.htm

CEOS Committee on Earth Observation Satellites (CEOS): Terms and Definitions and other documents and resources publicly available on http://calvalportal.ceos.org.
S. R. Deepika and N. Avinash, in “Proceedings of the Fourth International Conference on Signal and Image processing”, Volume 1, p. 576, 2012

Connor, B. J., D. E. Siskind, J. J. Tsou, A. Parrish, and E. E. Remsberg, Ground-based microwave observations of ozone in the upper stratosphere and mesosphere, J. Geophys. Res., 99(D8), 16757–16770, 1994

de Leeuw, G., T. Holzer-Popp, S. Bevan, W. Davies, J. Descloitres, R.G. Grainger, J. Griesfeller, A. Heckel, S. Kinne, L. Klüser, P. Kolmonen, P. Litvinov, D. Martynenko, P.J.R. North, B. Ovigneur, N. Pascal, C. Poulsen, D. Ramon, M. Schulz, R.Siddans, L. Sogacheva, D. Tanré, G.E. Thomas, T.H. Virtanen, W. von Hoyningen Huene, M.Vountas, S. Pinnock: Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis, Remote Sens. Env., 162, 295-315, doi:10.1016/j.rse.2013.04.023, 2015.

DIN 18716-3: 1997-07, Photogrammetry and remote sensing - Part 3: Remote sensing terms

Dirksen R.J., M. Sommer, F.J. Immler, D.F. Hurst, R. Kivi, H. Vömel, 2014, Reference quality upper air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmos. Meas. Tech., 7, 4462-4490, doi:10.5194/amt-7-4463-2014

EARLINET publishing group 2000–2010 (Adam, M., Alados-Arboledas, L., Althausen, D., Amiridis, V., Amodeo, A., Ansmann, A., Apituley, A., Arshinov, Y., Balis, D., Belegante, L., Bobrovnikov, S., Boselli, A., Bravo-Aranda, J. A., Bösenberg, J., Carstea, E., Chaikovsky, A., Comerón, A., D'Amico, G., Daou, D., Dreischuh, T., Engelmann, R., Finger, F., Freudenthaler, V., Garcia-Vizcaino, D., García, A. J. F., Geiß, A., Giannakaki, E., Giehl, H., Giunta, A., de Graaf, M., Granados-Muñoz, M. J., Grein, M., Grigorov, I., Groß, S., Gruening, C., Guerrero-Rascado, J. L., Haeffelin, M., Hayek, T., Iarlori, M., Kanitz, T., Kokkalis, P., Linné, H., Madonna, F., Mamouriat, R. E., Matthias, V., Mattis, I., Menéndez, F. M., Mitev, V., Mona, L., Morille, Y., Muñoz, C., Müller, A., Müller, D., Navas-Guzmán, F., Nemuc, A., Nicolae, D., Pandolfi, M., Papayannis, A., Pappalardo, G., Pelon, J., Perrone, M. R., Pietruczuk, A., Pisani, G., Potma, C., Preißler, J.,Pujadas, M., Putaud, J., Radu, C., Ravetta, F., Reigert, A., Rizi, V., Rocadenbosch, F., Rodríguez, A., Sauvage, L., Schmidt, J., Schnell, F., Schwarz, A., Seifert, P., Serikov, I., Sicard, M., Silva, A. M., Simeonov, V., Siomos, N., Sirch, T., Spinelli, N., Stoyanov, D., Talianu, C., Tesche, M., De Tomasi, F., Trickl, T., Vaughan, G., Volten, H., Wagner, F., Wandinger, U., Wang, X., Wiegner, M., and Wilson, K. M.): EARLINET Correlative Observations for CALIPSO (2006–2010), World Data Center for Climate (WDCC), Hamburg, Germany, doi:10.1594/WDCC/EN_Calipso_2006-2010, 2014.

Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, J. Geophys. Res., 104, 31333–31349, 1999.

EUMETSAT, IASI Level 2: Product Guide, EUM/OPS-EPS/MAN/04/0033, v3B, 15 October 2014.

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D.W., Haywood, J., Lean, J., Lowe, D.C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland R.: Changes in atmospheric constituents and in radiative forcing. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In: Solomon S., D. Qin, M. R. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, H.L. Miller (eds.), Climate Change 2007: The Physical Science Basis.. Cambridge University Press, Cambridge, United Kingdom / New York, NY, USA., pp.131-217, 2007.

Godin, S., Carswell, A. I., Donovan, D. P., Claude, H., Steinbrecht, W., McDermid, I. S., McGee, T. J., Gross, M. R., Nakane, H., Swart, D. P. J., Bergwerff, H. B., Uchino, O., von der Gathen, P., and Neuber, R.: Ozone differential absorption LIDAR algorithm intercomparison, Appl. Opt., 38, 6225–6236, doi:10.1364/AO.38.006225, 1999.

Godin-Beekmann, S.; Porteneuve, J. & Garnier, A. Systematic DIAL LIDAR monitoring of the stratospheric ozone vertical distribution at Observatoire de Haute-Provence (43.92N, 5.71E) J. Environ. Monit., 5, 57-67, 2003

Guerrero-Rascado, J. L., Ruiz, B., and Alados Arboledas, L.: Multi-spectral Lidar characterization of the vertical structure of Saharan dust aerosol over southern Spain, Atmos. Environ., 42, 2668–2681, doi:10.1016/j.atmosenv.2007.12.062, 2008.

GUM. Joint Committee for Guides in Metrology (JCGM/WG 1) 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in a measurement (GUM), http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer,A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: A ERONET- A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sens. Environ., 66, 1–16, 1998.

Huang, J., Hsu, N.C., Tsay, S.C., Liu, Z., Jeong, M.J., Hansell, R.A., and Lee, J.: Use of spaceborne lidar for the evaluation of thin cirrus contamination and screening in the Aqua MODIS Collection 5 aerosol products, J. Geophys. Res. Atmos., 118, 6444–6453, doi:10.1002/jgrd.50504, 2013.

Huang, J., Hsu,N.C., Tsay, S.C., Jeong, M.J., Holben, B.N., Berkoff, T.A., and Welton, E.J.: Susceptibility of aerosol optical thickness retrievals to thin cirrus contamination during the BASE-ASIA campaign, J. Geophys. Res., 116, D08214, 2012.

Huang, J., et al.: Evaluations of cirrus contamination and screening in ground aerosol observations using collocated lidar systems, J. Geophys. Res., 117, D15204, doi:10.1029/2012JD017757, 2011.

Hubert, D., Lambert, J.-C., Verhoelst, T., Granville, J., Keppens, A., Baray, J.-L., Bourassa, A. E., Cortesi, U., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Hoppel, K. W., Johnson, B. J., Kyrölä, E., Leblanc, T., Lichtenberg, G., Marchand, M., McElroy, C. T., Murtagh, D., Nakane, H., Portafaix, T., Querel, R., Russell III, J. M., Salvador, J., Smit, H. G. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Stübi, R., Swart, D. P. J., Taha, G., Tarasick, D. W., Thompson, A. M., Urban, J., van Gijsel, J. A. E., Van Malderen, R., von der Gathen, P., Walker, K. A., Wolfram, E., and Zawodny, J. M.: Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records, Atmos. Meas. Tech., 9, 2497-2534, doi:10.5194/amt-9-2497-2016, 2016.

Keckhut, P., McDermid, S., Swart, D., McGee, T., Godin-Beekmann, S., Adriani, A., Barnes, J., Baray, J.-L., Bencherif, H., Claude, H., di Sarra, A., Fiocco, G., Hansen, G., Hauchecorne, A., Leblanc, T., Lee, C., Pal, S., Megie, G., Nakane, H., Neuber, R., Steinbrecht, W., and Thayer, J.: Review of ozone and temperature lidar validations performed within the framework of the Network for the Detection of Stratospheric Change, J. Environ. Monit., 6, 721–733, doi:10.1039/b404256e, 2004.

Keppens, A., Lambert, J.-C., Granville, J., Miles, G., Siddans, R., van Peet, J. C. A., van der A, R. J., Hubert, D., Verhoelst, T., Delcloo, A., Godin-Beekmann, S., Kivi, R., Stübi, R., and Zehner, C.: Round-robin evaluation of nadir ozone profile retrievals: methodology and application to MetOp-A GOME-2, Atmos. Meas. Tech., 8, 2093-2120, doi:10.5194/amt-8-2093-2015, 2015.

Kobayashi, J. and Toyama, Y.: On various methods of measuring the vertical distribution of atmospheric ozone (III) – Carbon iodine type chemical ozonesonde, Pap. Met. Geophys., 17, 113–126, https://www.jstage.jst.go.jp/article/mripapers1950/17/2/17_97/_article/references, 1966.

Kolmonen, P., Sogacheva, L., Virtanen, T.H., de Leeuw, G., and Kulmala, M.: The ADV/ASV AATSR aerosol v2.30 retrieval algorithm: current status and presentation of a full-mission AOD data set. Int. J of Digital Earth, p.1-17, doi: 10.1080/17538947.2015.1111450, 2016.

Komhyr, W. D.: Electrochemical concentration cells for gas analysis, Ann. Geophys., 25, 203–210, 1969.

Kotz, S., and S. Nadarajah. *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press, 2000.

Iarlori, M., Madonna, F., Rizi, V., Trickl, T., and Amodeo, A.: Effective resolution concepts for lidar observations, Atmos. Meas. Tech., 8, 5157-5176, doi:10.5194/amt-8-5157-2015, 2015.

Immler, F. J., J. Dykema, T. Gardiner, D. N. Whiteman, P. W. Thorne, and H. Vömel, 2010: Reference quality upper-air measurements: Guidance for developing GRUAN data products. Atmos. Meas. Tech., 3, 1217–1231, doi:10.5194/amt-3-1217-2010.

ISO 9000:2015(en), Quality management systems - Fundamentals and vocabulary

ISO 19116:2004(en), Geographic information - Positioning services

ISO/TS 19130-2:2014(en), Geographic information - Imagery sensor models for geopositioning - Part 2: SAR, InSAR, lidar and sonar

ISO/TS 19159-1:2014(en), Geographic information - Calibration and validation of remote sensing imagery sensors and data — Part 1: Optical sensors

Kittaka, C., Winker, D. M., Vaughan, M. A., Omar, A., and Remer, L. A.: Intercomparison of column aerosol optical depths from CALIPSO and MODIS-Aqua, Atmos. Meas. Tech., 4, 131–141, doi:10.5194/amt-4-131-2011, 2011.

Lambert et al., "Multi-dimensional characterisation of remotely sensed data", EC FP6 GEOmon Technical Notes, 2011

Larssen, S., R. Sluyter, and C. Helmis, Criteria for EUROAIRNET – The EEA Air Quality Monitoring and Information Network, 1999.

MACC II Service Validation Protocol, Deliverable D153.4, May 2013, http://www.gmes-atmosphere.eu/documents/maccii/deliverables/man/MACCII_MAN_DEL_D_153.1_20130528_Lambert_V2.pdf

McDermid, S.; Godin, S. & Walsh, T. Lidar measurements of stratospheric ozone and intercomparisons and validation Appl. Opt., 29, 4914-4923, 1990

Meyer, M.C. (2013) A Simple New Algorithm for Quadratic Programming with Applications in Statistics, Communications in Statistics, 42(5), 1126-1139.

Mona, L., Pappalardo, G., Amodeo, A., D'Amico, G., Madonna, F., Boselli, A., Giunta, A., Russo, F., and Cuomo, V.: One year of CNR-IMAA multi-wavelength Raman lidar measurements in coincidence with CALIPSO overpasses: Level 1 products com-parison, Atmos. Chem. Phys., 9, 7213–7228, doi:10.5194/acp-9-7213-2009, 2009.

Nappo, C.J., Caneill J.Y., Furman R.W., Gifford F.A., Kaimal J.C., Kramer M.L., Lockhart T.J., Pendergast M.M, Pielke R.A., Randerson D., Shreffler J.H., and Wyngaard J.C., The Workshop on the Representativeness of Meteorological Observations, June 1981, Boulder, CO, Bull. Am. Meteorol. Soc. 63, 761-764, 1982.

NIST. Prokhorov, A. V., R. U. Datla, V. P. Zakharenkov, V. Privalsky, T. W. Humpherys, and V. I. Sapritsky, Spaceborne Optoelectronic Sensors and their Radiometric Calibration. Terms and Definitions. Part 1. Calibration Techniques, Ed. by A. C. Parr and L. K. Issaev, NIST Technical Note NISTIR 7203, March 2005.

Omar, A., Winker, D., Kittaka, C., Vaughan, M., Liu, Z., Hu, Y. X., Trepte, C., Rogers, R., Ferrare, R., Lee, K., Kuehn, R., and Hostetler, C.: The CALIPSO automated aerosol classification and lidar ratio selection algorithm, J. Atmos. Ocean. Tech., 26, 1994–2014, doi:10.1175/2009jtecha1231.1, 2009.

Papagiannopoulos, N., Mona, L., Alados-Arboledas, L., Amiridis, V., Baars, H., Binietoglou, I., Bortoli, D., D'Amico, G., Giunta, A., Guerrero-Rascado, J. L., Schwarz, A., Pereira, S., Spinelli, N., Wandinger, U., Wang, X., and Pappalardo, G.: CALIPSO climatological products: evaluation and suggestions from EARLINET, Atmos. Chem. Phys., 16, 2341-2357, doi:10.5194/acp-16-2341-2016, 2016.

Pappalardo, G., et al.: EARLINET correlative measurements for CALIPSO: First intercomparison results, J. Geophys. Res., 115, D00H19, doi:10.1029/2009JD012147, 2010.

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

Parrish, A., Boyd, I. S., Nedoluha, G. E., Bhartia, P. K., Frith, S. M., Kramarova, N. A., Connor, B. J., Bodeker, G. E., Froidevaux, L., Shiotani, M., and Sakazaki, T.: Diurnal variations of stratospheric ozone measured by ground-based microwave remote sensing at the Mauna Loa NDACC site: measurement validation and GEOSCCM model comparison, Atmos. Chem. Phys., 14, 7255-7272, doi:10.5194/acp-14-7255-2014, 2014.

QA4EO – A Quality Assurance framework for Earth Observation, established by the CEOS. It consists of ten distinct key guidelines linked through an overarching document (the QA4EO Principles) and more community-specific QA4EO procedures, all available on http://qa4eo.org/documentation.html A short QA4EO "user" guide has been produced to provide background into QA4EO and how one would start implementing it (http://qa4eo.org/docs/QA4EO_guide.pdf)

Reale, A., B. Sun, M.Pettey and F. Tilley, 2010: The NOAA products validation system (NPROVS) and archive summary system (NARCSS). 6th Annual Symposium on Future Operational Environmental Satellite Systems-NPOESS and GOES-R, 90th AMS Annual Meeting, Atlanta, Ga, 17-21 January.

Schafer, J. S., Eck, T. F., Holben, B. N., Thornhill, K. L., Anderson, B. E., Sinyuk, A., Giles, D. M., Winstead, E. L., Ziemba, L. D., Beyersdorf, A. J., Kenny, P. R., Smirnov, A., and Slutsker, I.:

Intercomparison of aerosol single-scattering albedo derived from AERONET surface radiometers and LARGE in situ aircraft profiles during the 2011 DRAGON-MD and DISCOVER-AQ experiments, J. Geophys. Res. Atmos., 119, 7439–7452, doi:10.1002/2013JD021166, 2014.

Seidel, D. J., and Coauthors, 2009: Reference upper-air observations for climate: Rationale, progress, and plans. Bull. Amer. Meteor. Soc., 90, 361–369, doi:10.1175/2008BAMS2540.1.

Seidel, D. J., B. Sun, M. Pettey, and A. Reale , Global radiosonde balloon drift statistics, J. Geophys. Res., 116, D07102, doi:10.1029/2010JD014891, 2011.

Smit, H. G. J. and ASOPOS-panel: Quality Assurance and Quality Control for Ozonesonde Measurements in GAW, WMO GAW report series 201, World Meteorological Organization, available at: http://www.wmo.int/pages/prog/arep/gaw/documents/FINAL_GAW_201_Oct_2014.pdf (last access: 27 April 2016), 2014.

Sofieva, V. F., F. Dalaudier, R. Kivi, and E. Kyrö (2008), On the variability of temperature profiles in the stratosphere: Implications for validation, Geophys. Res. Lett., 35, L23808, doi:10.1029/2008GL035539.

Sogacheva L., Kolmonen, P., Virtanen, T.H., Rodriguez, E., Saponaro, G., and de Leeuw, G.: Post-processing to remove residual clouds from aerosol optical depth retrieved using the Advanced Along Track Scanning Radiometer, Atmos. Meas. Tech., Atmos. Meas. Tech., 10, 491-505, doi:10.5194/amt-10-491-2017, 2017.

Schuster, G. L., Vaughan, M., MacDonnell, D., Su, W., Winker, D., Dubovik, O., Lapyonok, T., and Trepte, C.: Comparison of CALIPSO aerosol optical depth retrievals to AERONET measurements, and a climatology for the lidar ratio of dust, Atmos. Chem. Phys., 12, 7431-7452, doi:10.5194/acp-12-7431-2012, 2012.

VIM/ISO:99. Joint Committee for Guides in Metrology (JCGM/WG 2) 200:2012 & ISO/IEC Guide 99-12:2007, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), http://www.bipm.org/en/publications/guides/vim.html

Verhoelst, T., Granville, J., Hendrick, F., Köhler, U., Lerot, C., Pommereau, J.-P., Redondas, A., Van Roozendael, M., and Lambert, J.-C.: Metrology of ground-based satellite validation: co-location mismatch and smoothing issues of total ozone comparisons, Atmos. Meas. Tech., 8, 5039-5062, doi:10.5194/amt-8-5039-2015, 2015.

Wilson, W. J., and P. R. Schwartz, Diurnal variations of mesospheric ozone using millimeter-wave measurements, J. Geophys. Res., 86(C8), 7385–7388, 1981

Winker, D. M., Hunt, W.H., and McGill, M.J.: Initial performance assessment of CALIOP, Geophys. Res. Lett., 34, L19803, doi:10.1029/2007GL030135, 2007.

Winker, D. M., Vaughan, M. A., Omar, A. H., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H., and Young, S. A.: Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms, J. Atmos. Oceanic Technol., 26, 2310–2323, 2009.

Winker, D. M., Pelon, J., Coakley Jr., J. A., Ackerman, S. A., Charlson, R. J., Colarco, P. R., Flamant, P., Fu, Q., Hoff, R., Kittaka, C., Kubar, T. L., LeTreut, H., McCormick, M. P., Megie, G., Poole, L., Powell, K., Trepte, C., Vaughan, M. A., and Wielicki, B. A.: The CALIPSO Mission: A Global 3D View Of Aerosols And Clouds, Bull. Amer. Meteor. Soc., 91, 1211–1229, doi:10.1175/2010BAMS3009.1, 2010.

Winker, D. M., Tackett, J. L., Getzewich, B. J., Liu, Z., Vaughan, M. A., and Rogers, R. R.: The global 3-D distribution of tropospheric aerosols as characterized by CALIOP, Atmos. Chem. Phys., 13, 3345-3361, doi:10.5194/acp-13-3345-2013, 2013.

WMO. WMO Quality Management Framework (QMF), home page at https://public.wmo.int/en/programmes/wmo-quality-management-framework

WordNet. Princeton University "About WordNet." WordNet. Princeton University. 2010, http://wordnet.princeton.edu.

Young, S. A.: Analysis of lidar backscatter profiles in optically thin clouds, Appl. Optics, 34, 7019–7031, doi:10.1364/AO.34.007019, 1995.

Young, S. A., Vaughan, M. A., Kuehn, R. E., and Winker, D. M.: The retrieval of profiles of particulate extinction from CloudAerosol Lidar and Infrared Pathfinder Observations (CALIPSO) data: Uncertainty and error sensitivity analyses, J. Atmos. Ocean. Tech., 30, 395–428, 2013.

Young, S. A. and Vaughan, M. A.: The retrieval of profiles of particulate extinction from Cloud Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data: Algorithm description, J. Atmos. Ocean. Tech., 26, 1105–1119, doi:10.1175/2008JTECHA1221.1, 2009.

Annex A: QA4ECV recommended terms and definitions

TERM	DEFINITION	SOURCE
accuracy	closeness of agreement between a measured quantity value and a true quantity value of a measurand; note that it is not a quantity and it is not given a numerical quantity value	VIM/ISO:99, GUM
area (volume) of representativeness	the area (volume) in which the concentration does not differ from the concentration at the station by more than a specific range	Larssen
Bias	(1) systematic error of indication of a measuring system (2) estimate of a systematic measurement error (3) estimate of a systematic forecast error	(1) VIM/ISO:99 (2) VIM/ISO:99 (3) MACC
calibration	(1) the process of quantitatively defining the system responses to known, controlled signal inputs (2) operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication	(1) CEOS/ISO:19159 (2) VIM/ISO:99
dead band (or neutral zone)	maximum interval through which a value of a quantity being measured can be changed in both directions without producing a detectable change in the corresponding indication	VIM/ISO:99
detection limit	measured quantity value, obtained by a given measurement procedure, for which the probability of falsely claiming the absence of a component is β, given a probability α of falsely claiming its presence	VIM/ISO:99

error	(1) measured quantity value minus a reference quantity value (2) difference of quantity value obtained by measurement and true value of the measurand (3) difference of forecast value and a, estimate of the true value	(1) VIM/ISO:99 (2) CEOS/ISO:19159 (3) MACC
establish	define, document and implement	CDRH
instantaneous field of view (IFOV)	opening angle corresponding to one detector element	ISO:19130
fiducial	used as a fixed standard of reference for comparison or measurement (fiducial point)	WordNet
fiducial marker	refers to an object placed in the field of view of an imaging system which appears in the image produced, for use as a point of reference or a measure	
field-of-regard	an area of the object space scanned by the field-of-view of a scanning sensor	NIST
field-of-view	the solid angle from which the detector receives radiation	NIST
footprint	the area of a target encircled by the field-of-view of a detector of radiation, or irradiated by an active system	NIST
geometrical resolution	ability of a sensor system to record signals separately from neighboring object structures	DIN 18716-3: 1997-07
ground sampling distance (GSD)	linear distance between pixel centres on the ground	CEOS/ISO:19159
influence quantity	quantity that, in a direct measurement, does not affect the quantity that is actually measured, but affects the relation between the indication and the measurement result	VIM/ISO:99
in situ measurement	(1) a direct measurement of the measurand in its original place (2) any sub-orbital measurement of the measurand	(1) CEOS/ISO:19159 (2) GEOSS
measurand	quantity intended to be measured	VIM/ISO:99
metadata	data about the data; parameters that describe, characterise, and/or index the data	WMO

monitoring	(1) systematic evaluation over time of some quantity (2) by extension, evaluation over time of the performance of a system, of the occurrence of an event etc.	(1) NIST (2) MACC
point-to-area (point-to-volume) representativeness	the probability that a point measurement lies within a specific range of area-average (volume-average) concentration value	Nappo
positional accuracy	closeness of coordinate value to the true or accepted value in a specified reference system	ISO:19116
precision	(1) measure of the repeatability of a set of measurements. Note that precision is usually expressed as a statistical value based upon a set of repeated measurements such as the standard deviation from the sample mean (2) closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions	(1) ISO:19116 (2) VIM/ISO:99
procedure	specified way to carry out an activity or a process	ISO:9000
process	set of interrelated or interacting activities that use inputs to deliver an intended result	ISO:9000
process validation	establishing documented evidence of a high degree of assurance that a specific process will consistently produce a product meeting its pre-determined specifications and quality characteristics	CDRH
quality	degree to which a set of inherent characteristics of an object fulfils requirements	ISO:9000
quality assurance	part of quality management focused on providing confidence that quality requirements will be fulfilled	CEOS/ISO:19159, ISO:9000
quality assessment	term referring to the derivation of quality indicators providing sufficient information to assess whether quality requirements are fulfilled	CEOS
quality control (QC)	(1) QC refers to the activities undertaken to check and optimise accuracy and precision of the data after its collection (2) part of quality management focused on fulfilling quality requirements	(1) CEOS/ISO:19159 (2) ISO:9000

quality indicator (QI)	a means of providing a user of data or derived product with sufficient information to assess its suitability for a particular application. This information should be based on a quantitative assessment of its traceability to an agreed reference or measurement standard (ideally SI), but can be presented as a numeric or a text descriptor, provided the quantitative linkage is defined.	QA4EO
radiometric calibration	a determination of radiometric instrument performance in the spatial, spectral, and temporal domains in a series of measurements, in which its output is related to the true value of the measured radiometric quantity	NIST
random error	(1) component of measurement error that in replicate measurements varies in an unpredictable manner; note that random measurement error equals measurement error minus systematic measurement error (2) component of forecast error that varies in an unpredictable manner	(1) VIM/ISO:99 (2) MACC
relative standard uncertainty	standard measurement uncertainty divided by the absolute value of the measured quantity value	VIM/ISO:99
repeatability	measurement precision under set of conditions including the same measurement procedure, same operator, same measuring system, same operating conditions and same location, and replicated measurements over a short period of time	VIM/ISO:99
representativeness	the extent to which a set of measurements taken in a given space-time domain reflect the actual conditions in the same or different space-time domain taken on a scale appropriate for a specific application	Nappo
reproducibility	measurement precision under a set of conditions including different locations, operators, and measuring systems	VIM/ISO:99

resolution	(1) smallest change in a quantity being measured that causes a perceptible change in the corresponding indication (2) the least angular/linear/temporal/spectral distance between two identical point sources of radiation that can be distinguished according to a given criterion (3) the least vertical/geographical/temporal distance between two identical atmospheric features that can be distinguished in a gridded numerical product or in time series of measurements; resolution is equal to or coarser than vertical/geographical/temporal sampling of the grid or the measurement time series	(1) VIM/ISO:99 (2) NIST (3) MACC
stability	Property of a measuring instrument, whereby its metrological properties remain constant in time	VIM/ISO:99
systematic error	component of measurement error that in replicate measurements remains constant or varies in a predictable manner	VIM/ISO:99
system	set of interrelated or interacting elements	ISO:9000
traceability	(1) (*metrological traceability*) property of a measurement result relating the result to a stated metrological reference (free definition and not necessarily SI) through an unbroken chain of calibrations of a measuring system or comparisons, each contributing to the stated measurement uncertainty (2) ability to trace the history, application or location of an object, a product or a service	(1) VIM/ISO:99 (2) ISO:9000
traceability chain	sequence of measurement standards and calibrations that is used to relate a measurement result to a reference	VIM/ISO:99
uncertainty	non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used	VIM/ISO:99

validation	(1) the process of assessing, by independent means, the quality of the data products derived from the system outputs (2) verification, where the specified requirements are adequate for an intended use (3) confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled (4) the process of assessing, by independent means, the degree of correspondence between the value of the radiometric quantity derived from the output signal of a calibrated radiometric device and the actual value of this quantity. (5) confirmation by examination and provision of objective evidence that specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled	(1) CEOS/ISO:19159 (2) VIM/ISO:99 (3) ISO:9000 (4) NIST (5) CDRH
verification	(1) provision of objective evidence that a given item fulfils specified requirements; note that, when applicable, measurement uncertainty should be taken into consideration. (2) confirmation, through the provision of objective evidence, that specified requirements have been fulfilled (3) the provision of objective evidence that the design outputs of a particular phase of the software development life cycle meet all of the specified requirements for that phase	(1) VIM/ISO:99 (2) ISO:9000 (3) CDRH
vicarious calibration	post-launch calibration of sensors that make use of natural or artificial sites on the surface of the Earth	CEOS/ISO:19159