IDENTIFYING AND QUANTIFYING THE VALUE OF REDUNDANT MEASUREMENTS AT GRUAN SITES

Fabio Madonna

Consiglio Nazionale delle Ricerche
Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA)
C.da S. Loja, Tito Scalo, Potenza, I-85050
Tel. +390971427252, Fax +390971427271, e-mail: fabio.madonna@imaa.cnr.it

ABSTRACT

GATNDOR (GRUAN Analysis Team for Network Design and Operations Research), a research team supporting the development and implementation of GRUAN on scientifically sound foundations, is working for providing a quantification of the value of the so-called "complementary observations", identified as ground based remote sensing techniques available at GRUAN sites.
One of the key questions GATNDOR is trying to address is the quantification of the value of redundant measurements for the profiling of both temperature and humidity using data from highly-instrumented GRUAN sites (e.g., ARM site, Beltsville, Cabauw, Lindenberg, Potenza). The investigation is carrying out focusing on the most common instruments available at the GRUAN sites: radiosoundings and microwave profilers, for temperature; radiosoundings, Raman lidars, microwave profilers and GPS receivers, for moisture.
The aim of the investigation is the provision of recommendations for the establishment of an optimal observation strategy in GRUAN and for the reduction of uncertainties taking advantage of redundancy. Moreover, recommendations for the equipment to operate/acquire at the GRUAN sites will be also provided.
In this work, first results on the assessment of value of redundant measurements are provided and a possible metric for the quantification of redundancy with respect to measurement uncertainty, resolution and costs is proposed. Basic examples including radiosonde and Raman lidar observations collected at GRUAN stations are also discussed.

1. INTRODUCTION

The GCOS Reference Upper Air Network (GRUAN) is an international reference observing network, designed to meet climate requirements and to fill a major void in the current global observing system. Upper air observations within the GRUAN network will provide long-term high-quality climate records, will be used to constrain and validate data from space based remote sensors, and will provide accurate data for the study of atmospheric processes.
The GRUAN Analysis Team for Network Design and Operations Research (GATNDOR) is a research team, established in 2009, supporting the development and implementation of GRUAN on scientifically sound foundations. The team performs short-term focused research to address specific topics identified by the GRUAN science and management community. GATNDOR efforts are coordinated with other GRUAN Task Teams and with national GCOS programs when appropriate. One of the key questions GATNDOR is dealing with is how to reduce uncertainty taking advantage of measurement redundancy of GRUAN highly-instrumented sites.
Redundancy if often claimed as the best approach to study the atmosphere. This induced many atmospheric observatories to acquire multiples sensors measuring the same atmospheric variables though using different techniques and methodologies.
Redundancy in measurement systems at sites is typically used for:

► filling gaps, improving measurement continuity over the time in case of a failure or maintenance procedures;

► increasing the sampling of atmosphere, putting together measurements from different instruments;

► performing intercomparison among different techniques identifying unresolved changes in the systematic bias between observing systems;

► providing advanced products by exploiting synergies and integration of the observed data.

Cross-checking of redundant measurements for consistency is considered an essential part of the GRUAN quality assurance procedures. A fully equipped GRUAN site shall make at least double, and preferably triple, redundant measurements of all GCOS Essential Climate Variables.
The reduction of the uncertainty resulting from increasing the measurement redundancy is an important possibility to explore. GATNDOR aims at providing recommendations for the establishment of GRUAN observation strategy with the aim of reducing profiling uncertainty. The objective is to characterize redundancy level as a functions of instruments number, uncertainty and cost. GATNDOR work is currently focused on water vapour and on the most common instruments available at the GRUAN sites: radiosoundings, Raman lidars, infrared and microwave radiometers and GPS receivers.
This work introduces the statistical concept of mutual correlation as a possible "metric" to quantify the value of redundant measurements. Moreover, basic examples of the use of mutual correlation and conditional probability applied to datasets of radiosondes and Raman lidar measurements from ARM SGP, Oklahoma, US, and CIAO, Potenza Italy, GRUAN stations are reported and discussed.

2. METHODOLOGY AND EXAMPLES

An assessment and an approach for the quantification of value of redundant measurements with respect to measurement uncertainty, temporal and spatial resolution, and costs is missing. The elaboration of a such approach requires the identification of a metric to evaluate the redundancy of multiple measurements of the same atmospheric parameter.
Correlation is often used to study redundant measurements and their reliability. But correlation measures the linear relationship (Pearson's correlation) or monotonic relationship (Spearman's correlation) between two variables, X and Y. A more appropriate concept is mutual correlation (MC): this is more general and measures the reduction of uncertainty in Y after observing X. So MC can measure non-monotonic relationships and other more complicated relationships. Formally, the mutual information of two discrete random variables X and Y can be defined as:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)} \right)$$

where p(x,y) is the joint probability distribution function of X and Y, and p(x) and p(y) are the marginal probability distribution functions of X and Y respectively. In the case of continuous random variables, the summation is matched with a definite double integral. Redundancy concept is a generalization of mutual information to N variables (X_1. $X_{2,}$ ………, X_N). Mutual information is a quantitative measurement of how much one random variable (Y) tells us about another random variable (X). In this case, information is thought of as a reduction in the uncertainty of a variable. Thus, the more mutual information between X and Y, the less uncertainty there is in X knowing Y or Y knowing X. In the following of this section, two examples are proposed to make clearer the possible use of MC a possible metric for quantifying the redundancy.
In Figure 1, a comparison between the probability density functions (pdf) of integrated water vapour obtained using 14 months of radiosonde data from CIAO GRUAN station in Potenza, Italy, in the period September 2004 – October 2005 and the conditional pdfs obtained conditioning radiosonde distribution with Raman lidar measurements is reported.
Pdfs are calculated using three different atmospheric ranges from 1 to 3 km a.g.l. (upper panels), from 3 to 6km a.g.l. (middle panels),and from 6 to 8 km a.g.l. (lower panels). Moreover, two averaging times for the lidar data have been considered: 10 minutes (left panels) and 40 minutes

(right panels). All the measurements used for the analysis have been selected during both night time and clear sky conditions. Moreover, selected profiles have a maximum lidar range, corresponding to an error lower than 50%, higher than 8 km and has passed all the quality checks implemented by CIAO scientific team (Madonna et al., 2011). From the comparison, preliminary findings about the use of lidar measurements for conditioning the radiosonde pdf can be pointed out: in the range 1-3 km there is no significant impact on the radiosonde pdf, while in the free troposphere differences in the shape of pdf become more evident and, therefore, a reduction of uncertainty in the estimation of water vapour content might be achieved.

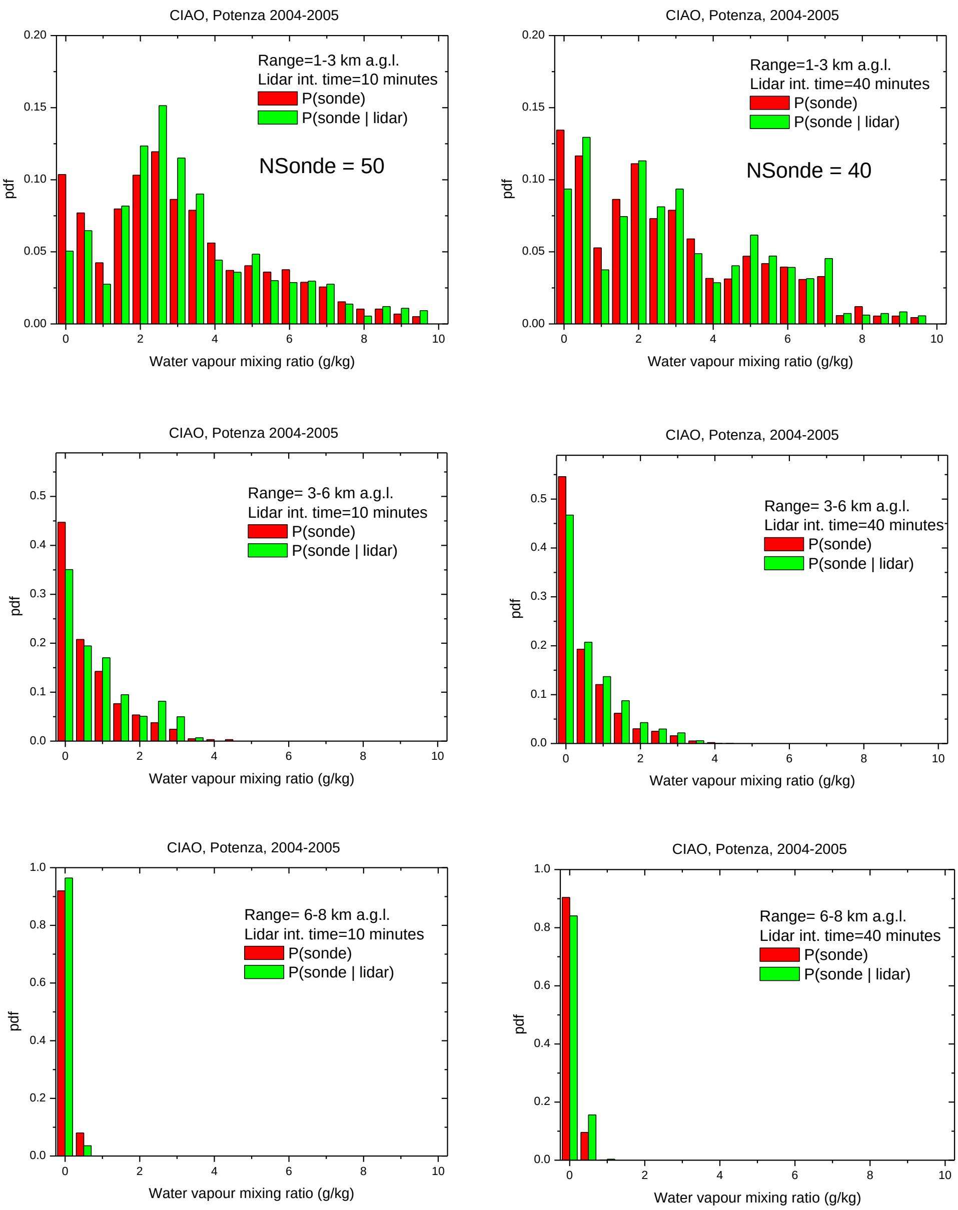

Figure 1: Comparison between the probability density functions of the integrated water vapour calculated using 14 months of radiosonde data from CIAO GRUAN station in Potenza, Italy, in the period September 2004 – October 2005 (red) and the conditional probability density functions (pdfs) obtained conditioning radiosonde distribution with Raman lidar measurements. Distributions are calculated using three different atmospheric ranges from 1 to 3 km a.g.l. (upper panels), from 3 to 6 km a.g.l. (middle panels),and from 6 to 8 km a.g.l. (lower panels). Moreover two averaging time periods for the lidar data have been considered: 10 minutes (left panels) and 40 minutes (right panels).

In Figure 2, it is reported the MC between radiosonde and Raman lidar data calculated over a period of one year (2010) of ARM SGP data from station in Oklahoma, US, as a function of lidar averaging time (upper panel) and an additional random uncertainty applied on the Raman lidar

data (lower panel). The upper panel of Figure 2 reports the MC for the ARM sonde and Raman lidar data calculated averaging lidar data over different time periods from 10 minutes (standard ARM average time) to 3 hours. In the lower panel of Figure 2, MC for the 2010 ARM data has been calculated adding an increasing random noise on Raman lidar profiles averaged over 2 hours, using a Monte Carlo approach. The additional random uncertainty on the lidar data has been obtained adding random noise to the Raman lidar data using a Monte Carlo number generator. A large number (>50-100) of extractions have been used to ensure the reliability of the analysis.
All the measurements used for the analysis have been selected during night time and clear sky conditions. Moreover, selected profiles have a maximum lidar range, corresponding to an error lower than 25%, higher than 7.5 km and have passed all the quality checks implemented by ARM scientific team (Miller et al., 2003). Moreover, both the radiosonde and Raman lidar time series of water vapour mixing ratio profiles have been preliminarily de-trended and normalized to zero mean and unit variance.

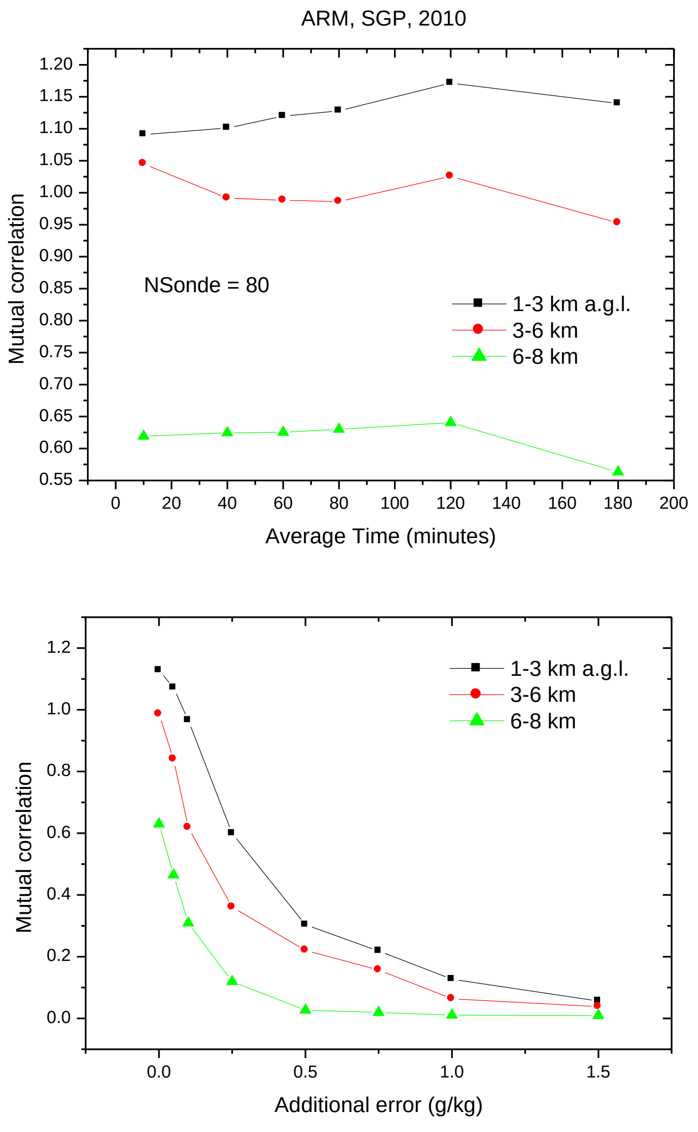

Figure 2: Mutual correlation between radiosonde and Raman lidar data calculated for one year (2010) of data from ARM SGP station in Oklahoma, US, as a function of lidar averaging time (upper panel) and an additional random uncertainty applied on the Raman lidar data (lower panel). The additional random uncertainty has been obtained using a Monte Carlo number generator.

The upper panel of Figure 2 shows the absence of a strong dependence of MC on time averaging, though the time averaged of 2 hours seems to be a threshold above which the MC in the free

troposphere decreases. This also decreases the possibility to reduce the uncertainty taking advantage of the mutual information from radiosonde and lidar measurements.
The lower panel of Figure 2 shows that at all the altitude levels, MC dramatically decreases with the uncertainty increase, revealing how much the impact of large uncertainty affecting one the redundant measurements can strongly decrease its power in reducing the overall uncertainty budget in the estimation of atmospheric water vapour.

3. SUMMARY AND NEXT STEPS

This work reports basic examples of the application of the concept of mutual correlation in the quantification of the value of redundant measurements. Mutual correlation looks an appealing tool to be used as a possible metric to quantify the value of redundant measurements, to reduce uncertainties and to optimize measurement scheduling and quality assurance in the frame of GRUAN.

Next steps of this work include:
► the study of mutual correlation on longer data records and at more GRUAN sites;
► the study of water vapour profiles from passive profiling techniques and the study of redundancy for the estimation of integrated water content;
► the use of new GRUAN radiosonde products;
► the provision of recommendations about the equipment to operate at GRUAN sites and about the elaboration of advanced (synergetic/integrated) products from redundant measurements.

4. ACKNOWLEDGEMENTS

ARM data is made available through the U.S. Department of Energy as part of the Atmospheric Radiation Measurement Program.
CIAO Potenza data is made available through CNR-IMAA (http://www.ciao.imaa.cnr.it).
Beltsville data is available through Howard University (http://meiyu.atmphys.howard.edu).

REFERENCES

1. Madonna, F., Amodeo, A., Boselli, A., Cornacchia, C., Cuomo, V., D'Amico, G., Giunta, A., Mona, L., and Pappalardo, G.: CIAO: the CNR-IMAA advanced observatory for atmospheric research, Atmos. Meas. Tech., 4, 1191-1208, doi:10.5194/amt-4-1191-2011, 2011.
2. Miller, M. A., K. L. Johnson, D. T. Troyan, E. E. Clothiaux, E. J. Mlawer, and G. G. Mace, 2003: ARM value-added cloud products: Description and status. Proc. of the 13th ARM Science Team Meeting, Broomfield, CO, ARM. [Available online at www.arm.gov/publications/proceedings/confl3/extended_abs/miller-ma.pdf.]