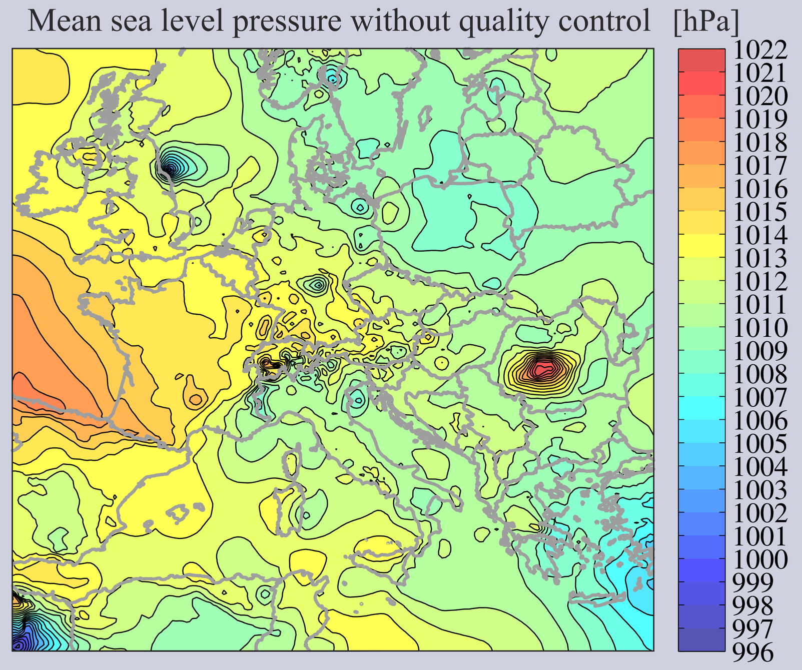
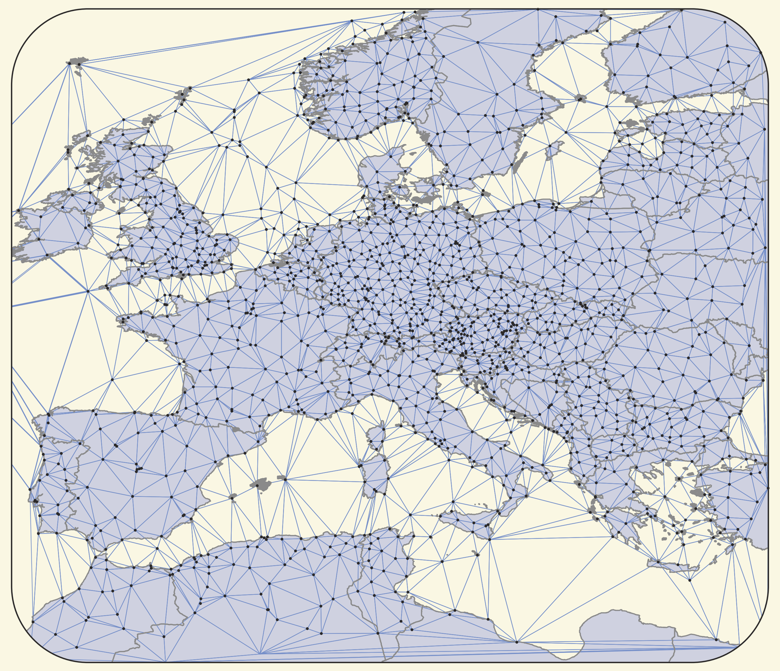
VERA-QC, approved Data Quality Control based on Self-Consistency

Andrea STEINER, Dieter MAYER & Reinhold STEINACKER
Department of Meteorology and Geophysics, University of Vienna
contact: andrea.steiner@univie.ac.at

TROUBLESOME DATA?



1) MOTIVATION



Positions and connecting lines of reporting GTS weather stations on 17th February 2012, 1800 UTC.

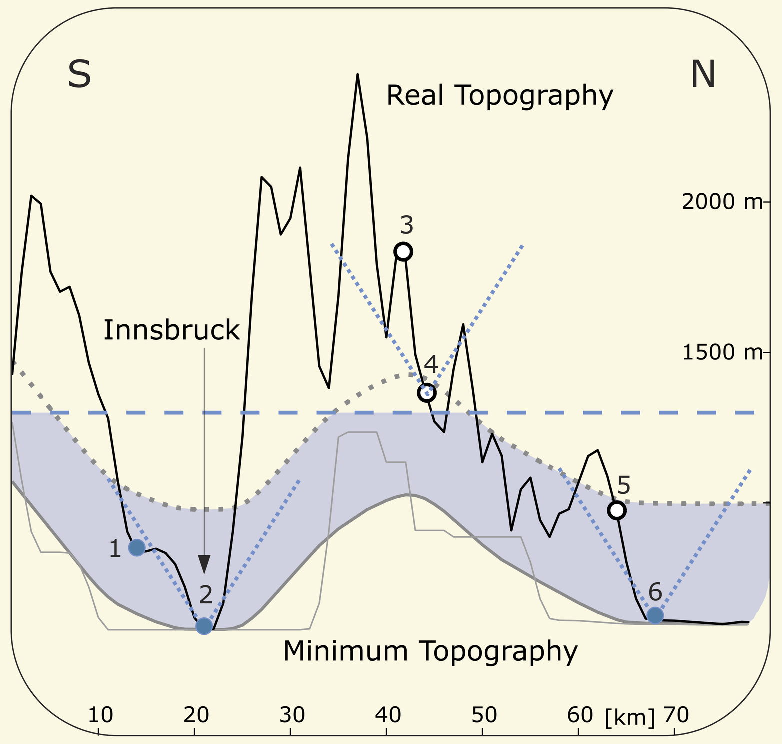
- Measurements are not always as accurate as required.
- Observations of high quality are indispensable for:
  - Generation of analyses
  - Nowcasting
  - Model verification
  - Collecting long-term time series
- Therefore, the quality control algorithm VERA-QC was developed.

2) METHOD

p_{red}, θ, θ_e, u, v, RR

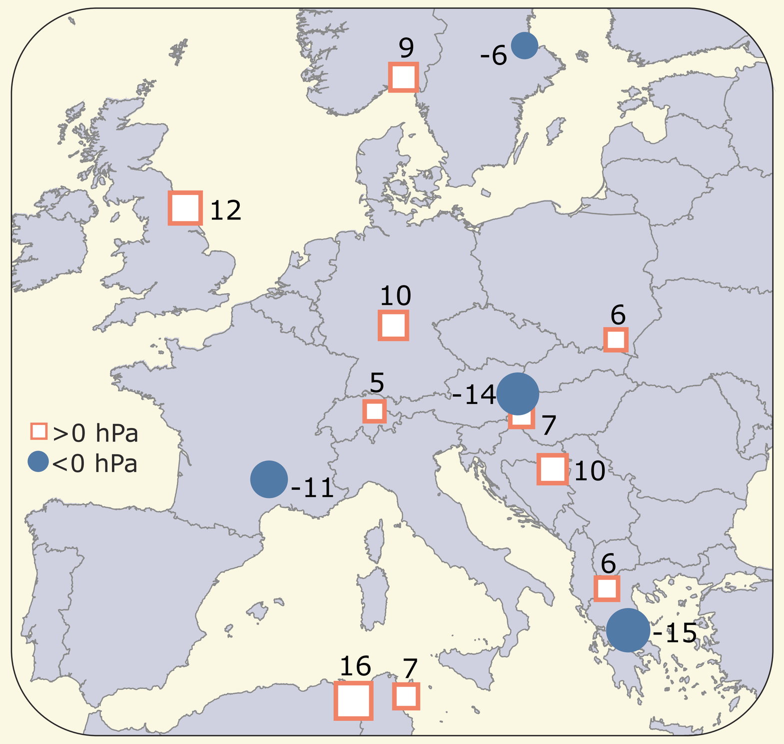
Station selection algorithm → Bias correction → Consideration of relative values → Limit check → Check for self-consistency (includes cluster treatment and gross error recognition) → Application of corrections → Statistics

2.1) Station selection algorithm



Only representative stations for valleys and plains are accepted (here: stations numbered 1, 2, and 6).

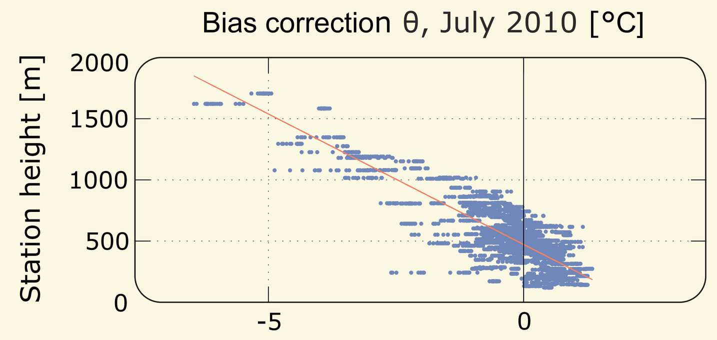
2.2) Bias correction



MSL-pressure bias for 31st July 2010. The bias correction is defined as median of a station's unweighted correction proposals collected within a parameter specific time interval in the past (e.g. the last 30 days).

2.3) Consideration of relative values

For avoiding a negative θ-bias for higher stations, relative values in comparison to the standard atmosphere are considered.



Strong dependence on height is due to the usage of dry-adiabatic lapse rate for computing θ.

2.4) Check for self-consistency

For parameters ψ featuring a high redundancy with regard to the observational network, the error free analysis field ψ_a should be smooth.

- A cost function J consisting of the squared curvatures C of the analysis field is to be minimized:

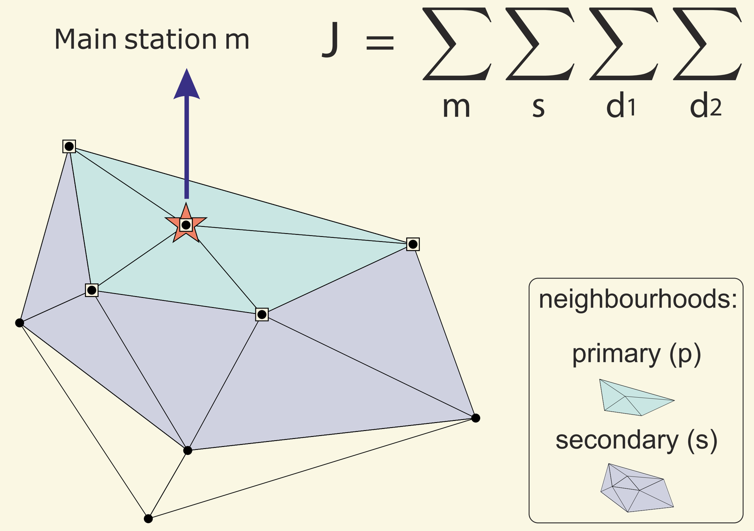
$$J(\Psi_a) = \sum_{n}^{N(G)} (C_{\Psi_a})_n^2 \rightarrow \min$$

Problem: ψ_a and its curvature are unknown

- $C(\psi_a)$ is approximated by a first order Taylor series around the curvature of the observed field $C(\psi_o)$:

$$J = \sum_m \sum_s \sum_{d_1} \sum_{d_2} \left[\frac{\partial^2 \Psi_o}{\partial d_1 \partial d_2} + \sum_p \frac{\partial}{\partial \Psi_p} \left(\frac{\partial^2 \Psi_o}{\partial d_1 \partial d_2} \right) \Delta\Psi_p \right]_s^2$$

unknown deviations



Exemplary station distribution illustrating the concept of finding natural neighbors

- For obtaining the unknown deviations, the cost function is differentiated with respect to all m observations ψ:

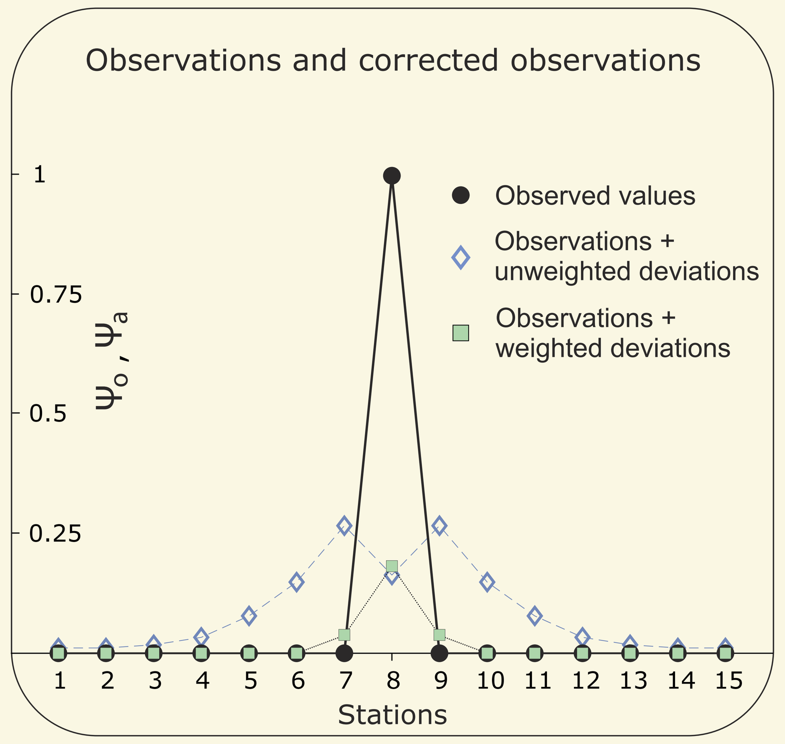
$$\frac{\partial J}{\partial \Psi_m} = 0$$

- Deviations are weighted according to the reduction of cost function they induce:

$$\Delta\Psi_{m,w} = W_m \cdot \Delta\Psi_m$$

$$W_m = \frac{J(\ldots,\Psi_m,\ldots) - J(\ldots,\Psi_m + \Delta\Psi_m,\ldots)}{J(\ldots,\Psi_m,\ldots)}$$

- A special cluster treatment allows applying the weighting method to station-clusters.

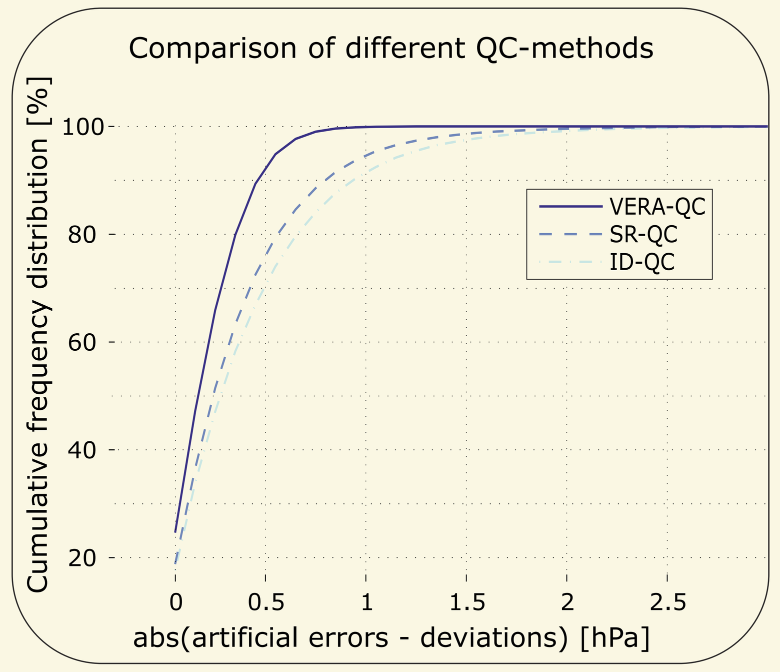


One-dimensional example of 15 stations, whereas no. 8 is an outlier. The application of weighted deviations results in satisfying error-correction without producing an error-propagation to neighboring stations.

- Weighted deviations constitute the basis for a decision tree:

$|\Delta\Psi_{m,w}| > th_1 \cdot \overline{|\Delta\Psi_w|}$ & $|\Delta\Psi_{m,w}| > th_2$ deviation statistics
- yes → gross error, discard observation
- no → $|\Delta\Psi_{m,w}| > th_3$
  - yes → correct observation by weighted deviations
  - no → accept observation as it is

3) EVALUATION OF METHOD



For comparison, a series of simulated pressure fields covering coastal, plain and alpine stations was generated. All three QC-methods were optimized in favor of best handling these cases.

VERA-QC is compared to two other spatial consistency checks using inverse distance (ID) and spatial regression (SR) interpolation.

Gross error recognition VERA-QC / SR-QC

		SIMULATED: GE	SIMULATED: no GE
RECOGNIZED	GE	689 / 680	3 / 13
RECOGNIZED	no GE	11 / 20	32497 / 32387

4) CONCLUSIONS

Primarily, VERA-QC proved itself as valuable preprocessing tool for the operational real time VERA analyses that also support forecasters of the Austrian aviation weather service (ACG, Austro Control GmbH).

References

Steinacker, R., D. Mayer, and A. Steiner, 2011: Data Quality Control Based on Self-Consistency. *Monthly Weather Review*, **139**, 3974–3991

Mayer, D., R. Steinacker and A. Steiner: Innovations and Applications of the VERA Quality Control. submitted to: *Geoscientific Instrumentation, Methods and Data Systems*

VERA analyses of quality controlled meteorological observations: http://www.univie.ac.at/amk/veraflex/test/public/

Acknowledgments

Thanks are due to the Austrian Science Fund (FWF): P19658 and to the Austrian research funding association (FFG; project 537003) for partial financial support of this work.

PROBLEM SOLVED!