Interpolation uncertainty of Vaisala RS41 temperature profiles

Alessandro Fassò - University of Bergamo, Italy

Based on https://doi.org/10.5194/amt-2020-161
In collaboration with Michael Sommer & Christoph von Rohden

ICM12 - Virtual meeting - 20 November 2020

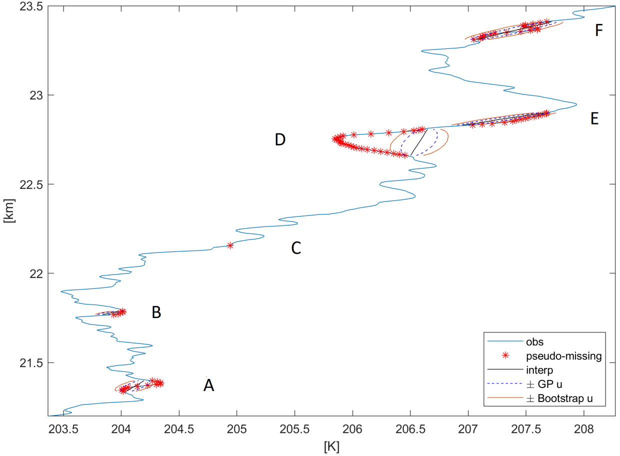
Motivation

- Although temperature readings made by Vaisala RS41 radiosondes at GRUAN sites are given at 1 s resolution, for various reasons, missing data are spread along the atmospheric profile, usually clustered in gaps of various lengths.
- There is some evidence that the missing fraction is higher at higher altitudes with an average around 0.13 for profiles in "bad" conditions and lower for profiles in "good" conditions.
- The RS41 Gruan Data Processing may benefit by linear interpolating small data gaps.
- In doing this the GDP should cover also for interpolation uncertainty.

Extreme Simulated Missing Data

Linear interpolation

RS41 T at Sodankylä, 2017-03-03 12:00, $\approx 22.5\,\mathrm{km}$ altitude.



Interpolation Uncertainty

Let t be the radiosonde flying time [s] and let (t^-, t^+) be a measurement gap.
For $t \in (t^-, t^+)$ we estimate the Interpolation Uncertainty by

$$IU_{Total}(t|s,l)^2 = U_{GP}(t|s,l)^2 + U_B(d, ALT)^2$$

where

- U_{GP} is the individual IU, based on a local Gaussian Process (GP) approximation of the specific profile (s, l).
- $U_B(d, ALT)$ is the average uncertainty not explained by the GP, at altitude ALT [Km] and "interpolation distance" $d = \sqrt{(t - t^-)(t^+ - t)}$ [s].
  It is estimated by a Bootstrap approach.

GRUAN RS41 Profile Data

Site	Code	Country	Selected Profiles
Beltsville	BEL	USA	15
Lauder	LAU	NZ	32
Lindenberg	LIN	DE	45
Ny-Ålesund	NYA	DE/FR	35
Payerne	PAY	CH	30
Lamont	SGP	USA	16
Sodankylä	SOD	FI	4
			177

The profiles selected for the analysis have gaps shorter than 5 s, and less then 10 missing values each.

$\Rightarrow$ more than 1 million measurements covering the various climatic regions of GRUAN.

GP Uncertainty of Linear Interpolation

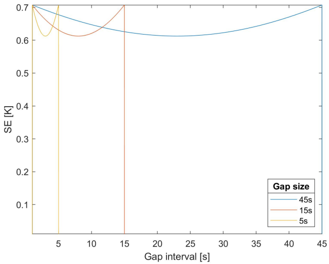
If the T-profile $y(t)$ is given by a GP with covariance function γ, we have

a closed form formula of individual uncertainty U_{GP}:

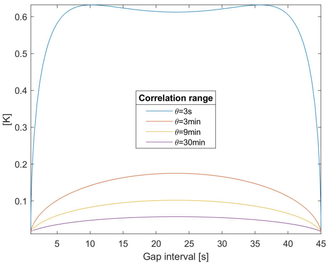
$$U_{GP}(t)^2 = 2\sigma_y^2 \left\{1 - \alpha + \alpha^2\right\} + 2\left\{\alpha(1-\alpha)\gamma(t^+ - t^-) - \alpha\gamma(t^+ - t) - (1-\alpha)\gamma(t - t^-)\right\} + \sigma_\epsilon^2$$

here, $\alpha = \alpha(t)$ is the linear interpolation weight: $\alpha(t) = \frac{t - t^-}{t^+ - t^-}$.

The "Shape" of Interpolation Uncertainty



White noise



Correlation effect

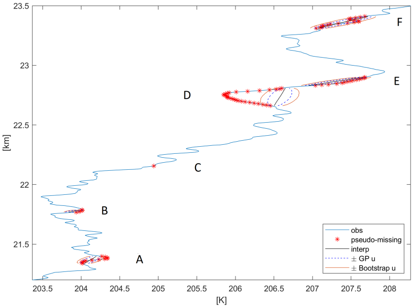
Block-bootstrap Cross-validation Scheme

Each fully observed T-profile $Y = (y(1), ..., y(n))$ is partitioned as follows:

$$Y \longrightarrow [Y^L, Y^*]$$

where

- t^L: set of learning times
- $Y^L = Y(t^L)$: learning set
- t^*: set of testing times
- $Y^* = Y(t^*)$: validation set made of n_G gap random sequences, each with Poisson duration.



Learning and Testing

To have abundant information in a dense vertical grid the above random extraction process is repeated B times:

$$Y \longrightarrow [{Y^L}_b, {Y^*}_b], \; b = 1, ..., B.$$

- Learning (for each s, l, b)
  - Linear interpolation: $Y^L \to \hat{Y}(t^*)$
  - GP: $Y^L \to GPmodel \to U_{GP}(t^*)$
- Testing
  - For each s, l, b: $e = \hat{Y}(t^*) - Y^*$
  - Pooling s, l, b: $MSE_B(d, ALT) = avg(e^2 | d, ALT)$
  - Bootstrap correction: $U_B^2 = MSE_B - avg(U_{GP}^2)$
  - Bootstrap corrected: $IU_{Total}^2 = U_{GP}^2 + U_B^2$

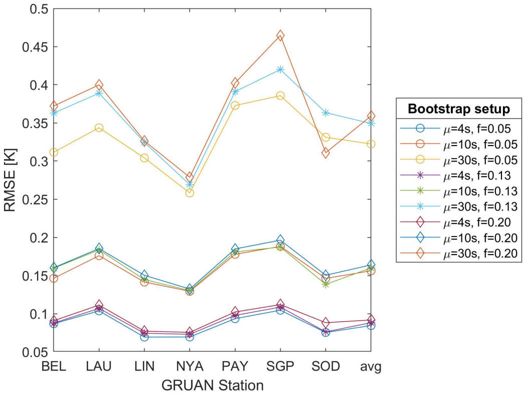
Preliminary Analysis

Overall $RMSE_B$ by station and avg gap length

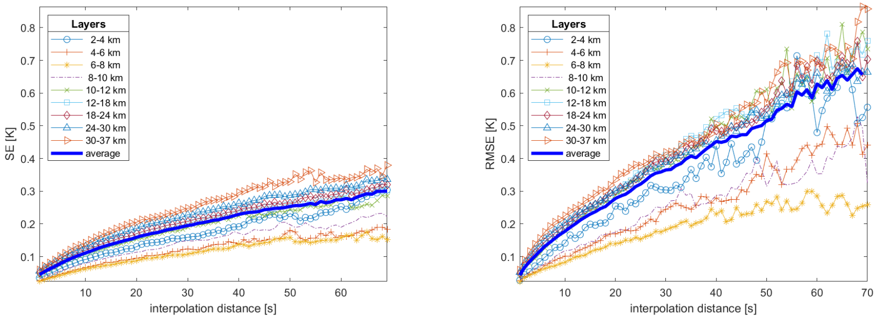
		Average gap length			
Site	Profiles	4 s	10 s	30 s	60 s
BEL	15	0.088	0.160	0.363	0.604
LAU	32	0.107	0.184	0.389	0.612
LIN	45	0.074	0.145	0.324	0.542
NYA	35	0.073	0.130	0.269	0.460
PAY	30	0.098	0.181	0.391	0.658
SGP	16	0.109	0.187	0.420	0.698
SOD	4	0.076	0.138	0.363	0.478
Average	177	0.088	0.160	0.349	0.576

Preliminary Analysis

$RMSE_B$ by avg gap length and missing fraction



Linear Interpolation Uncertainty Components



$$\Downarrow$$

$$IU_{Total}(t|s,l)^2 = U_{GP}(t|s,l)^2 + U_B(d,ALT)^2$$

Conclusions and Implementation Considerations

- We presented two tools for computing interpolation uncertainty, that may be integrated or used alone.
- The GP-based approach
  - provides uncertainty at the profile level;
  - it gives good results for small gaps but for larger ones underestimates the interpolation uncertainty;
  - it requires implementing GP fitting for each profile.
- The Bootstrap approach
  - provides an average uncertainty for fixed interpolation distance and altitude;
  - it may be implemented as a look-up table.

Thank You!