Atmos. Meas. Tech. Discuss., 8, C2423–C2427, 2015
www.atmos-meas-tech-discuss.net/8/C2423/2015/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Error estimation for localized signal properties: application to atmospheric mixing height retrievals" *by* G. Biavati et al.

G. Biavati et al.

gbiavati@bgc-jena.mpg.de

Received and published: 4 August 2015

Reply to Referee 2

We highly appreciate the work done by Anonymous Referee 2, who paid attention to the very details of our work. The comments and suggestions he provided were very seriously taken into account. Here we provide step by step answers that describe also the modifications done to the manuscript in order to fulfill the referee's suggestions.

The confidence z between two values is introduced by Eq. (2), and then the confidence neighborhood U is defined in Eq. (4). In Eq. (2) f is given as the argument of z, although it was never introduced before. I guess the authors would refer to y, instead. Please, clarify.

The fact that we referred to f(y) instead of y was a mistake coming from an earlier version of the manuscript. We corrected it removing the reference to f.

At 5115 – line 4 it is stated that U is "the contiguous set of locations surrounding the point xm that share the relation of confidence to the specific point f(y) with respect to the constant g". But, since the definition (4) only involves the difference between two any points (z_i and z_j) of the dataset, the contiguity is not guaranteed, since other points far from x_i or x_j might be associated to y values close to y_i or y_j. This may be particularly true for noisy data, where spurious values may result in points with high confidence to a given one, but which may far from it. This possibility is actually removed with the introduction of the strict confidence neighborhood, where a condition on the position of the points is posed. Please,

comment this fact.

In order to grant that the confidence neighborhood is a set of contiguous points, we have to remind that the confidence neighborhood is defined as a neighborhood (Eq. 3).

In order to make it more clear, we modified the text slightly. The manuscript now reads: "*The confidence neighborhood is defined as a neighborhood Eq.(3). So a confidence neighborhood must respect both Eq.(3) and Eq.(4). In this way contiguity and confidence relation are achieved together.*"

The authors say that $\gamma = 2$ produces reasonable uncertainties, providing reference of this statement.

We introduced this sentence in Sec. 3.2.1. We thought that by referring to Kretschmer et al. 2014 and looking just to Fig. 3, we could have a qualitative proof of this suggested γ value. To better prove this point, we made a number of Monte-Carlo tests. The results of these tests are provided in a supplementary document. However, we decided to not include these results in the manuscript, as it might cause more confusion, because the strict confidence neighborhood is a more relevant result.

In 3.3 the authors say that "the confidence neighborhood as the intersection of the respective confidence neighborhoods of each data series. So in the end, the confidence neighborhood for multiple data series is identical to the smallest confidence neighborhood of x_m for these data series".

It does not seems to me that this assertion is justified, since it may happen that if the neighborhood sets are not completely overlapped the intersection set might result smaller than the smaller one or, even, empty: $set_1 = (1, 2, 3, 4, 5, 6, 78)$, $set_2 = (7, 8, 9, 10)$, intersection set = $(7, 8)$

The definition of confidence neighborhood of x_i includes always x_{i-1} and x_{i+1} this property comes as heritage from the definition of discrete neighborhood. So when we

get x_m using multiple datasets, all the confidence neighborhoods of each dataset will contain these three points. In other words set1 and set2 of the example of the referee can not be confidence neighborhoods of the same x_m because they do not have at least 3 points in common.

In 3.4 it is not clear why increasing the resolution the algorithm tends to underestimate the 'right' value, although the uncertainty decreases, as expected. No explanation is given on this effect, which is counterintuitive. Please, try an explanation about that.

This interesting effect of increasing the resolution while keeping the noise unchanged depends mainly on the operative definition of the algorithm used to detect the threshold. As defined in Sec. 2.1, MH is located at the first point that respects the relation with the threshold. By using this definition, we can understand that by increasing the number of samples we also increase the possibility of detecting outliers below the "true" altitude. It is only an effect of the algorithm. If we had used an alternative definition like "MH is the last point lower than the threshold", we would have ended up with an overestimate of x_m. To make this clear in the text we added a sentence at the end of Sec. 3.4:

An interesting effect of increasing resolution can be seen in Fig. (6). Increasing the resolution provide results with mode lower than the expected result. This a combined effect of the choice of the algorithm described in Sec. 2.1 and of increasing the resolution. The algorithm points to the first point that satisfy the threshold relation. Increasing the number of points before the true one, we also increase the probability to have outliers before the true value. When choosing a different algorithm, e.g the last point that is lower than the threshold, we will end with a distribution of results that will overestimate the true value.

After Monte Carlo simulation on a large number of perturbed profiles, the authors find an analog y between the standard deviation of the distribution of the retrieved location and the confidence neighborhoods as from Eq. (4) for $\gamma \approx 2$.

The authors affirm that these quantities "correspond at least qualitatively". Finally, the uncertainty on the attribution of the local property is introduced by equation (6), using the definition of confidence neighborhood and the analogy to the standard deviation of the Monte Carlo simulation. The introduced formula (6) is interesting, but precisely for this reason a better quantification and analysis of the correspondence is desirable. I encourage the authors to better illustrate this point.

The choice of this formulation for the localization error is such that for uniform sampling and the smallest confidence neighborhood (x_{m-1},x_m,x_{m+1}) we obtain the resolution of the sampling as the localization error.

This can be seen easily applying Eq.(6) or Eq.(7). We added a sentence to explain this point (which originally was one of the reasons that pointed us to use Eq.(6)).

We found that Eq.(6) is a good choice for the definition of localization error, because uniformly sampled data series: in case of the smallest confidence neighborhood $U_{\gamma,y} = (x_{m-1}, x_m, x_{m+1})$*, the localization error is just the amplitude of the sampling, which in common practice is often used as uncertainty.*

The numerical proof provided in the supplement can be seen as a quantitative evaluation of the similarity of the results of Eq.(6) with the standard deviation of the distribution of the results, this when the measurement errors are reasonable small.

Last remark is about the captions. Almost all labels are too small; in A4 print they are almost illegible. Please, enlarge the labels. Furthermore, I encourage the authors to use more different colors so that also in BW prints the figure are comprehensible.

We followed the suggestion and increased the size of labels and annotations in figures as much as possible. In particular we increased the size of the symbols, so that they are visible when printing on A4.