

Point-by-point Responses to Reviewer Comments

Reviewer 3

This manuscript provides a rigorous probabilistic modeling study of numerical weather prediction (NWP) 10-m winds' error relative to independent ground-based wind speed measurements. A region-average wind error model is fitted as a Weibull distribution. Then, the errors are modeled marginally in space and time and combined using Gaussian coupling and spatiotemporal semivariograms. It seems technically sound and well written, with the following points to consider for further improvements.

We thank the reviewer for their detailed review of this manuscript and for identifying the rigour and technical quality of this work.

A main issue is that the connection between the core analysis/results of this work and the scope of AMT (specifically remote sensing of methane point source emissions) is relatively weak. It is really section 3.5 only. It may be helpful to provide more context on how this work fits in the methane quantification pipeline. Specifically, the 10-m wind that this study tries to model, as the ground truth, is measurement at (usually) 10 min interval. The 10-m wind->effective wind->methane emission pipeline is calibrated all in an LES model in Varon 2018. To what extent is the "ground-truth" wind speed relevant, if the wind-emission relationship is calibrated by a model? This study tackles the error from NWP to ground-measured 10 m wind, and it would be nice to have some discussion on the mapping from measured wind to effective wind and then to emissions. It almost makes me feel that targeting measured wind is a detour, and one should map NWP wind to LES wind, or whatever wind that calibrates the wind-emission relationship.

We believe that this manuscript is well within the scope of AMT, specifically (from AMT's stated aims and scope) the "validation ... of techniques of data processing" relevant to the retrieval, here quantification, of atmospheric gases.

We explicitly state in our introduction where wind quantification exists in the methane plume quantification pipeline. We would also argue that the ground-truth wind field, which is what has propagated an observed plume, is the exact wind data required to quantify methane flux through a control surface.

Varon et al. (2018) – which importantly is just one approach to bring wind speed information into a quantification pipeline as we note in our introduction – developed "calibration" functions that estimate an "effective wind speed" from a ground-truth wind speed. In their simulation work, the ground-truth wind speed was necessarily modelled using LES but, in practice, absent local measurements, this ground-truth wind speed would be estimated using an NWP model. Thus – even for this unique calibration approach that, to our knowledge, is not commonly used by others – evaluating the performance of NWP models vs. ground-truths is critical.

Fittings of the regional wind error model and the semivariograms are important for this work, so it is recommended to include details on how those fitting algorithms, specially the numerous constraints in fitting parameters, are implemented.

As noted by the reviewer below (see text following equations 17, 18, and 19 in the manuscript), we do explicitly note the constraints on fitted parameters. For optimization, we [use] MATLAB's *fmincon* (constrained minimization) function for the maximum likelihood estimation. We have revised section 2.2.2 to note this.

Technical comments:

Eq. 8: please double check as π seems to be reserved for a PDF, and dividing a PDF by u is unlikely to give another PDF.

The equation is correct as written. This is a change of variables to the distribution π_{cand} and the $1/\tilde{u}$ multiplier is the derivative of RER, necessary to ensure the law of total probability for the transformed distribution.

Page 9, lines 10-11: it is recommended to provide the formula of AIC as the objective function.

The AIC (Akaike Information Criterion) is a standard tool to evaluate relative model performance in a Bayesian setting and is essentially a parameter count-corrected negative log-likelihood. Rather than introduce the formulation for AIC, we have revised the text to include this description.

Page 10, line 1: should that be the "... the variance of the difference of the z_i at these positions"?

Yes, revised as requested.

Page 11, lines 9-10: double check the location and span of bins. Should that be 0-50 km and 25-75 km?

Yes, thank you for catching this. Revised.

Page 11, lines 24-25: H seems to be reserved for joint CDF. Consider another font/symbol for Heavyside step function.

Thank you for catching this. We have revised the equation to use a R for the Heaviside step function.

Page 11, lines 26: please confirm what $\gamma_k(b;b)$ is, and why it should be 0.95.

Since the monotonic semivariogram models employed only approach a value of unity asymptotically, their *range*, which defines a spatial limit of correlation must be defined using some threshold. By convention, this is typically taken as 0.95, which we employ here. Noting that a_k and b are always coupled, any value in (0, 1) could theoretically be chosen without affecting the underlying function; the fixed constants a_k and optimized range, b , would simply be different. We have added text to note that 0.95 is chosen by convention.

Page 12, lines 1-2: fixing b_3 to 0 contradicts the constraint ($b_3 > b_2 > b_1 \geq 0$) in the previous page. Please elaborate.

Thank you for catching this. We have revised to note that when each model is considered independently, b_3 is not required.

Page 12, lines 17-18: close the parenthesis.

Revised, thank you.

Page 21, line 5: should the RMS uncertainty be read as the "distance" of the upper and lower error bars?

As noted in the text, the RMS uncertainty is the root-mean-square of the upper and lower error bar lengths. We choose this parameter to collapse the asymmetric uncertainties into a single parameter.

Section 4.1: it reads a bit strange to have such a dominant subsection in conclusion section. Consider making it a dedicated section preceding the conclusion.

We have revised the structure as recommended to place this limitations subsection as section 4 preceding the conclusion.

References

Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O. A., Xia, Y. and Huang, Y.: Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes, *Atmos. Meas. Tech.*, 11(10), 5673–5686, doi:10.5194/amt-11-5673-2018, 2018.