

Reviewer 1

We thank the reviewer for his/her thoughtful and constructive comments.

General comments

1. p26085, l23-24: we have taken note of the reviewer's comments about the confusion with the use of the terms 'forward' and 'backward' (in relation to the STILT model) as well as about the terms 'surface influence' and 'footprint' and have now made the use of these terms more consistent throughout the manuscript. We have also added a definition of 'footprint' and how this differs from 'surface influence'.
2. optimal correlation scale lengths: the reviewer raises a very important point about how the optimal correlation scale length may be determined. We discuss this in point 14 under 'Specific comments'.
3. details of inversion set-up: we have now included the number of state variables (space and time) and the number of observations used in each inversion test in Table 2.

Specific comments

1. p.26076, l.11: I do not believe there is really a difference between "Ideally..." (relating to inversions) and "In an ideal inversion...". Preferring the latter, we have decided to leave this sentence as it is. The reviewer also makes the comment "if each state variable is observed at each resolved time-step, then one would need to ask oneself a question about what inversion has to bring to such a complete picture". This would be of course true if each state variable could be observed directly and independently from all others, then there would be no need for inversion, however, in this case, the observations (i.e. atmospheric concentrations) are related to the state variables via a physical model (i.e. atmospheric transport), thus, an inversion in this case is still pertinent to learn more about the state variables.
2. p.26078, l.20-21: a definition of "footprint" has been added
3. p.26079, l.3-4: this has been amended to "The STILT model provides surface influence functions with a dynamic resolution..."
4. p.26079, l.11-17: the first sentence of this paragraph has been removed and hereinafter the terminologies 'footprint' and 'surface influence function' are used more consistently.
5. p.26080, l.3: $c(x_i, y_j, z_k, t_0)$, represents the N₂O that is transported from the domain boundaries to the measurement site, where t_0 is defined as the time when the back-trajectory exits the domain. This has now been added to the text.

6. p.26081, l.13-14: the reviewer is correct in pointing out the possible confusion between “footprint” and “surface influence”. We have corrected the manuscript so that the term “surface influence” is used throughout when this is really what is meant.
7. p.26081, l.18: yes, n is the total number of state variables, i.e. the number of grid-cells multiplied by the number of time intervals. This has now been added to the text.
8. p. 26081, l. 21: it is not the boundary mixing ratio at the time of the observation that is considered here but the mixing ratio at the time when the back-trajectory exits the domain. The mixing ratio at the domain boundary is used to initialize the mixing ratio in the virtual air parcel as it enters the domain. The mixing ratio in this air parcel is then modified by the surface influence and by mixing inside the domain before arriving at the receptor. We have modified the sentence to make this clearer.
9. p.26083, l.2: by the “best estimate” we mean the optimum value of the state vector, i.e. that which results in the minimum of the cost function:

$$J(f) = (f-f_{\text{prior}})S_{\text{prior}}^{-1}(f-f_{\text{prior}}) + (Ff - y)S_{\varepsilon}^{-1}(Ff-y)$$

10. p.26085, l.23: “advected” in this case refers to the simulation of N2O using the STILT calculated surface influence functions.
11. p.26087, l.13: the number of observations was 2880 and the number of state variables (for 7-day resolution) was 15548. Following the definition of Tarantola (2005), the number of degrees of freedom for Bayesian Inversion is the dimension of the data space, thus it is 2880. We have added these values to Table 2.
12. p.26089, l.22: the temporal correlation scale length T is 30 days (this is given in Table 2).
13. p.26090, l.18: we presume that the reviewer is referring to the reference inversion (A) with a reduced chi-square value of 0.81. In this case, we do not consider the difference in reduced chi-square values between test A and the tests in which the aggregation error was accounted for (tests C) to be significant.
14. p.26091, l.24-26: the question of the correct spatial and temporal correlation scale lengths is one that pervades Bayesian inversions. It is not possible to determine the true values of the correlation scale lengths from these synthetic tests. Instead, one can only gauge the sensitivity of the results to the choice of correlation scale length. In order to derive the true correlation scale lengths, one would need to really look at the temporal and spatial correlations of flux measurements (provided that there are enough of these) and/or determine how these may be correlated temporally, e.g. on the basis of weather patterns, and spatially, e.g. on the basis of land-cover and climate

type. This is an area of ongoing research and is beyond the scope of this manuscript.

15. p.26092, l.1: a temporal resolution of 7 days was used for the fluxes, this has been clarified in the text
16. p.26092, l.3-4: the values of the errors in the observation space were chosen as follows:
 - i) measurement error: this estimate is based on the measured long-term precision of 0.18 ppb (calculated over 2.5 years of measurements) as well as on the comparisons of in-situ with independent flask measurements at Ochsenkopf (Thompson et al., 2009) in order to account for short term fluctuations in the precision and for the potential additional measurement errors associated with sampling air rather than a standard gas (which is used to determine the precision).
 - ii) transport error: this estimate was based on the mean of the 3-day standard deviations of the simulated N₂O mixing ratio at Ochsenkopf. The reason for this is that transport models are generally not able to represent synoptic scale variability well, which typically has time spans of circa 3-days. Therefore the mean of the standard deviations over this period was taken as the transport error.
 - iii) boundary error: this estimate was taken from the standard deviation of the boundary contribution to the mixing ratio observed at Ochsenkopf
 - iv) aggregation error: this value was calculated directly from the algorithm presented in the manuscriptA description of these estimates has been included in the manuscript on p.26092
17. p.26092, l.8-9: the posterior fluxes were coupled to STILT to simulate mixing ratios at both Ochsenkopf and an independent site (i.e. which was not included in the inversion) Bialystok. The simulated mixing ratios were then compared to the observations at each site as a way of verifying the posterior fluxes. This has been made clearer in the manuscript.
18. p. 26109, caption Fig.2: 'forward' has been removed from the caption
19. p.26110, caption Fig.3: For each inversion experiment we have used 3 criteria (or results) on which to assess the impact of changing e.g. the correlation scale length, these are i) RMSE of the true and posterior fluxes, ii) the reduced chi-square and iii) the total error reduction, thus, for each inversion run there are 3 points plotted (one for each criterion) and are black, red, and blue, respectively. The x-axis is for the parameter being tested in each inversion run e.g. correlation scale length. The caption has been amended according to the suggestions.

Technical corrections

We have taken note of the corrections recommended by the reviewer and have made changes to the manuscript where appropriate.