We would like to thank the referee for the helpful and constructive comments. In the following we list specific referee comments (in blue) together with author replies. Additions and changes to the paper text are written *italic*.

1. Referee general remarks:

Although the paper presents an interesting method to generate improved mixing height (MH) estimates as input for transport simulations, it has some important drawbacks that are not sufficiently acknowledged. Although the transport models may have vertically misplaced MHs, those MHs are internally consistent with the simulated vertical profiles of wind, temperature, humidity and other variables. Artificially changing the MH without adjusting other meteorological variables (notably the wind profile) will necessarily lead to inconsistencies, and such inconsistencies are likely contributing to the problems described in Section 4.2 where the STILT simulations based on MYJ versus YSU PBL parameterizations diverge even more after adjusting the MHs to more or less the same heights. Artificially displacing the MH may, for example, result in a situation where air parcels previously located in the free troposphere in a regime with strong wind speeds due to reduced drag, are suddenly located inside the PBL and thus able to interact with the surface. Signals from distant sources that would not reach a surface site within a short time given the low wind speeds in the PBL, could now reach the site more quickly due to fast transport aloft which is no longer decoupled from the PBL as it should.

Modern data assimilation methods such as 4D-VAR or Ensemble Kalman filters could be used to incorporate radiosonde observations for improved MH estimates similar to the method presented here, but with the advantage of not disturbing the internal consistency of the model (or at least much less). The above mentioned deficiencies of the method and the alternative of data assimilation should be better highlighted in the paper.

Author reply:

The referee has pointed at some of the known limitations of our method and we agree that these should be stated more clearly in the paper. Thus we extended the method section (Sec. 2.3, P4643, L8) to address these drawbacks and mention possible alternatives for more sophisticated assimilation method as follows:

The results of Kretschmer et al. (2013) indicate that the dominant effect MH errors on the transport simulation is the turbulent diffusion of tracer particles up to a wrong altitude, suggesting that potential physical inconsistencies and side effects affect the tracer concentrations in the mixing layer to a minor extend. Such physical inconsistencies involve other meteorological input variables used for the turbulence calculations. The profiles of vertical velocity variance σ_w, which determines the amount of random deviation from the mean vertical wind for a given particle, and the Lagrangian time scale T_L, which describes the decorrelation in the particles movement [Lin et al., 2003], depend not only on the MH, but also on roughness length, Monin-Obukhov length, convective velocity scale, and

frictional velocity, following Hanna [1982].
However, only the mixing height determines the altitude, at which strong turbulent mixing changes from high values within the mixing layer to lower values for the free troposphere. Here our assumption is, that the potential impact of the other meteorological input variables on resulting tracer profiles is small. To further support this assumption, we would need to assimilate these additional variables, but in comparison to MHs these are harder to measure.
Another side effect is the de- or entrainment of air particles to or from the layer above the mixing layer when the MH is changed, in combination with wind shear typically present in that region, which could lead to an alteration of the particles' trajectories within the mixing layer, resulting in modified surface influence. Lowering the MH leads to a more local flux influence since horizontal wind speed decreases with decreasing distance to the surface, thus we can expect a minor alteration of mixing ratios. More sensitivity has to be expected in case of strong wind shear near the MH together with an increase in the MH. This is likely to happen during nighttime when low level jets can develop near the MH. For CO2 we can expect minor negative impact on mixing ratios, because the dominating respiration flux is spatially rather homogeneous. During daytime the NEE is less homogeneous, but the mixing layer is generally deeper and thus the impact of a slight increase of the MH is expected to cause only small alterations of the mean horizontal trajectories. These assumptions are supported by the results of Kretschmer et al. (2013).

We added the following sentences to the discussion (Sec. 4, P4652, L5):
In addition, transport model errors could be due to physical inconsistencies of the presented method for MH optimization (c.f. Sec. 2.3). An improvement would be an assimilation of the MH directly in the meteorological model, which could be achieved by more classical assimilation techniques (e.g. 4D-VAR) and the introducing a new observational operator that relates prognostic variables (e.g. temperature) with MH observations, for instance one could use a Ri-method such as presented in Eq. (1).

2. Referee comment:
P4629, L7: What are "continental point observations"?

Author reply:
The sentence was changed as follows:

Within the top-down approach dispersion models are used to close the scale gap between global models and point observations by simulating regional greenhouse gas transport (Dolman et al., 2009; Gerbig et al., 2009).

3. Referee comment:
P4632, L5-L10: Should be reformulated. First the problem of prior flux uncertainties

is mentioned while the next sentence explains that simulations with two different PBL parameterizations were used. What is the connection between these two?

Author reply:
The intention of our approach is to isolate transport related differences in transport simulation by changing the transport simulation (two different PBL schemes) but keeping the CO2 surface fluxes identical.
We changed the sentences to make the connection more clear as follows:

The second complication is more difficult to tackle, because, uncertainties in prior fluxes were shown to have substantial impact on simulated CO2 concentrations (Peylin et al., 2011). To isolate the effect of transport errors on the CO2 concentrations we prescribed the same CO2 fluxes for all simulations, more specifically we compare results of two model setups with conceptually different PBL parametrizations, the Yonsei University Scheme (YSU, K-diffusion, Hong et al., 2006) and the Mellor–Yamada–Janjic scheme (MYJ, Turbulent Kinetic Energy, Janjic, 2002), prescribing the same vegetation and anthropogenic CO2 fluxes. In addition, we utilize the independent auxiliary tracer CO, to assess the model performance in simulating trace gas transport.

4. Referee comment:
P4639, L9: Why do you say "KED solves for the weights .."? Isn't this a general feature of Kriging not specific to KED?

Author reply:
Yes, it is a general feature of Kriging. We changed the sentence accordingly:
Kriging methods like KED solve for the weights wi such that interpolation error is minimal (Best Linear Unbiased Estimator, BLUE).

5. Referee comment:
P4639, L17: Shouldn't it be "linear function" rather than "linear combination"?

Author reply:
No, the specific term in linear algebra for such an equation is linear combination.

6. Referee comment:
P4639, L20: Why are the coefficients a and b without asterisks here?

Author reply:
That is a mistake. The asterisks have been added.

7. Referee comment:
P4641, L6: Why did you perform a weighted linear regression here? To get rid of the

biases? Should be explained more clearly.

Author reply:
In geostatistics observations are conceived as a combination of deterministic mean and spatio-temporally coloured random deviation of this mean. Fitting a linear regression was done to determine the first guess deterministic component of the observations in accordance with Eq. 6. The deterministic part is then subtracted from the observations yielding the random component of the observations that are the basis for the variogram analysis, i.e. input for Eq. 7 (the Res_{zi} terms).

We changed the sentence to improve understandability:
We estimated the deterministic component of the observations in accordance to Eq. 6 by fitting a weighted linear regression model to the observed MH as a function of the WRF MHs, taking the reciprocal of the estimated MH uncertainty from Eq. (2) as weights.

8. Referee comment:
P4641, L12-13: I didn't understand this sentence. What do you mean by "true variability"?

Author reply:
We changed the sentence to improve clarity:

Since the 12 h resolution of the IGRA data is too coarse to constrain the variogram model sufficiently, we make use of the hourly MHs from the WRF simulations, assuming that the resulting semivariance closely resembles the temporal auto-correlation properties of the observed signal adequately.

9. Referee comment:
P4642, Eq 8: In my view it should be f(xi,yi,tm | xr,tr), since the right hand term is a scalar product of two functions at the locations (xi,yi,tm) not at (xr,tr)

Author reply:
The notation was adopted from Gerbig et al. (2003b) and Lin et al. (2003) and was not changed for the sake of consistency.

10. Referee comment:
P4643, L8: What do you mean by "side effects"? Please explain.

Author reply:
Please see answer to referee comment 1.

11. Referee comment:
P4644, L7: To my knowledge EDGAR does not provide any time factors. Please pro-

vide more details.

Authore reply:
Additional information was added to the sentence:
The time factors are based on the step-function time profiles published on the EDGAR website (http://themasites.pbl.nl/tridion/en/themasites/edgar). These were modified before they were applied to yearly fluxes in order to resolve the daily cycle. The modification of the temporal factors involves a better global representation and a smoothing of the monthly transitions (c.f. Steinbach, 2010 for further details).

12. Referee comment:
P4645, L14: Explain how CO loss by reaction with OH is simulated. With prescribed 3D/4D OH field or just a constant OH value? What about CO production from VOCs?

Author reply:
We added this information after the sentence as follows:
Similar to Gerbig et al.(2003b) we estimate the OH on a given particle location based on a a climatological OH field.

CO production from non-methane hydrocarbon oxidation is neglected, which is is also not part of the EDGAR data set. However, CO from methane oxidation is accounted for in STILT.

13. Referee comment:
P4647, L14: "Random errors slightly decrease". Relative to what? Where or how do I see that?

Author reply:
The comparison is made between the two statistics in Table 2 (IGRA vs. WRF MH and IGRA vs. WRF optimized MH). The sentence has been changed to state this clearer:

Compared to the random errors of the unoptimized WRF MH (c.f. Table 2 upper part) the random errors of the optimized WRF MH are slightly smaller during day (c.f. Table 2 lower part), but become notably smaller when considering MH uncertainty, too.

14. Referee comment:
P4650, L28: Isn't the receptor located "above" the MH rather than "below"?

Author reply:
Yes, that is a mistake. The sentence has been changed accordingly.

15. Referee comment:
P4654, L5-8: How do you deal with potential biases in the MACC CO data? My own

experience with this data set is that it can have very large biases (though larger in winter than summer).

Author reply:
Potential biases in the MACC data were not accounted for in our study. However, CO is used only as auxiliary tracer to do a relative inter-comparison of the simulations and since the same background fields were used for all of these simulations the results of the comparison are not affected by such biases.

Other technical corrections to the paper have been addressed as suggested by the reviewer.