Dear Reviewer,

Thank you for undertaking a thorough review our study. We have addressed your concerns in order to improve the paper. The new version has benefited from your suggestions

The review starts with "*two main concerns that are key elements in the inverse systems, i.e. the observation and prior flux error covariance matrices*". We have included a better description of the observation error covariance matrix (2.3.1) and of the prior flux error covariance matrix (2.3.2), that was maybe not clear enough in the original version of the manuscript.

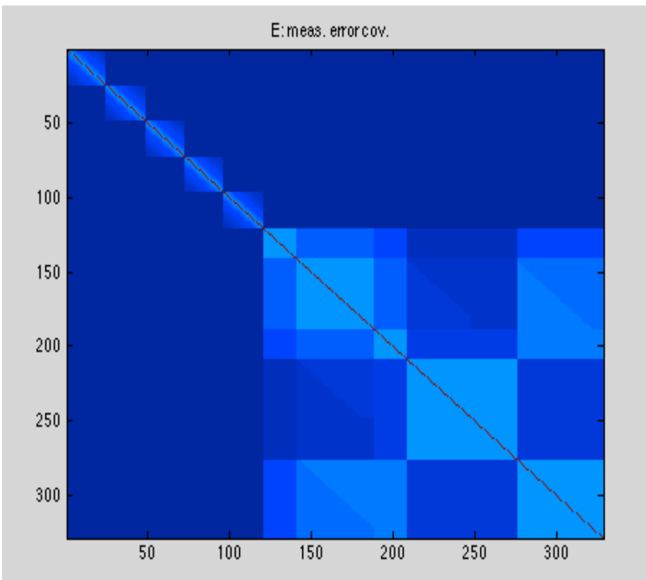
The review continues stressing that "*posterior uncertainties are not presented, and only sensitivity tests provide a crude approximation of the uncertainty from one component of the inversion, i.e. the prior fluxes*". We have included in the revised version the posterior uncertainties section (4.3) and a number of discussions that didn't find its place in the originally submitted version.

The final part of the opening paragraph reads: *"the boundary conditions, which are one of the major source of uncertainties at fine scale, are provided by a station in the middle of the Pacific Ocean. Considering the amplitude of the CO2 plume from large urban areas, and the presence of large sources in North East Asia, the boundary inflow can be highly influential and become even more important than in larger scale studies".* We have added sections 2.2, 3.3 and 5.1 and have justified our initial choice and added sensitivity studies to support it. Our initial choice of background data was motivated by the fact that global models are nevertheless scaled to match observations using station data (such as Mauna Loa or Yonaguni Island). It had been agreed that the definition of the "background" value depends on the scale and can be regarded as subjective. Nevertheless, the concern of the referee is justified at the small scale aimed in this work and therefore we have included some sensitivity tests using the global model AGCM as background. A number of technical details were omitted in the first version because a GMD paper was thought to be the natural place for them.

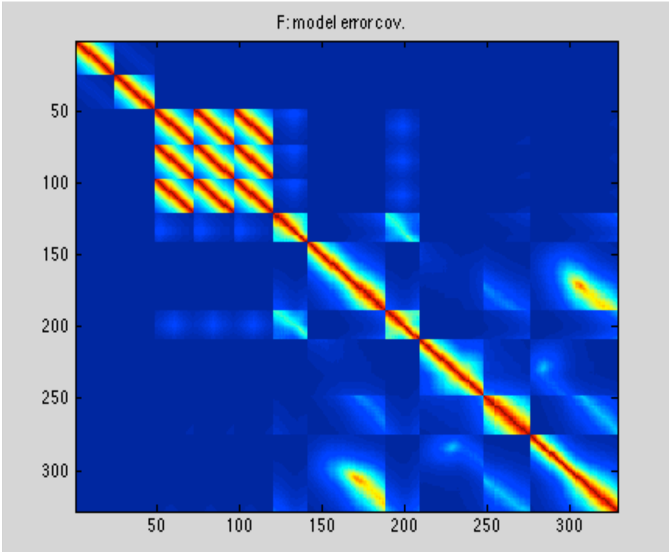
In the following paragraph of the review four main raised points are summarised: transport errors (1), prior flux errors (2), posterior uncertainties (3) and boundary conditions (4). We have addressed these major points as well as the specific points. The answers to these specific points can be found below.

1) Transport errors.
We do not disagree with the reviewer about the fact that transport errors are still large and not easy to quantify. We have included an explicit caveat in the text. However, we can justify our choice of the

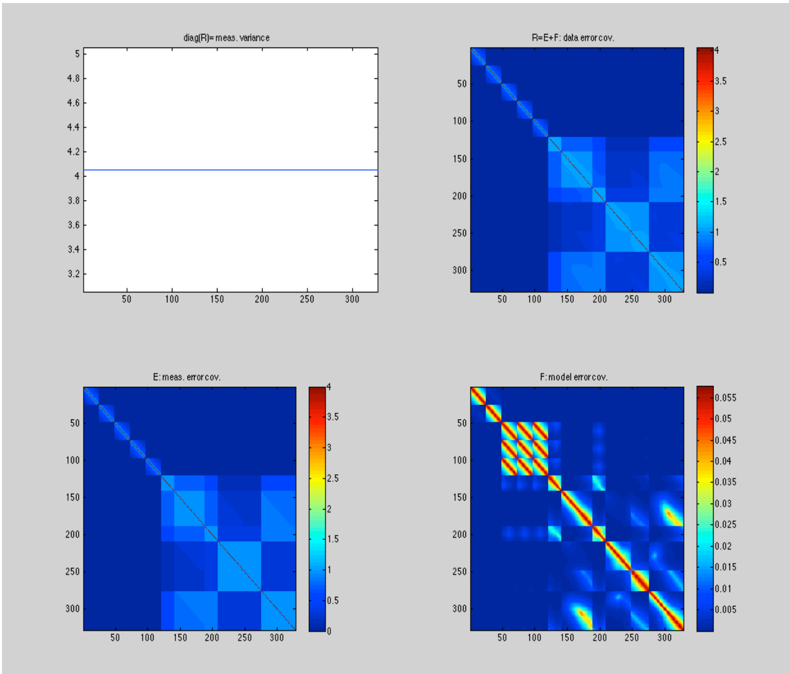
model and comment on the transport errors. We have added a new subsection "mesoscale meteorological model" based on the replies to these remarks.
WRF-CO2 has been chosen to be a state-of-the-art mesoscale model (Takigawa et al. (SOLA), Ballav et al. (JMS 2012), de Foy et al.). There exists literature referring to the assessment of the model developed by and active community coordinated by NCAR. The development of the Eulerian component of the model is beyond the scope of the paper, aimed to develop the Lagrangian component. A recent study supports our choice of using instantaneous winds (Brioude 2012, http://www.geosci-model-dev.net/5/1127/2012/gmd-5-1127-2012.html)
Also we have investigated previously the city plume calculating the proportion of particles in contact with Tokyo boundary layer, the "χ" number, inspired in the study of Berthet et al., (JGR 2007). A preliminary sensitivity study showed that in January 2007, 9 out of 11 major peaks of CO2 were captured. These results where presented at the AGU fall meeting in 2009 at San Francisco). Also, results obtained using ERA Interim and WRF meteorological winds are consistent with each other, as shown by Fig. 2.
In particular, a number of PBL uncertainty impact tests were performed during the preparation of the manuscript, but not included in the final version. We have re tested several aspects of the transport and we include some additional discussion about the PBL height in the revised manuscript.
We have also clarified our use of observation error covariance (see below). Often in data assimilation studies, transport uncertainty in included in this step. We have added a new subsection: "**Observations error covariance matrix"** based on the replies to these remarks.



Example of the structure of the measurement error covariance matrix E for one day worth of measurements.



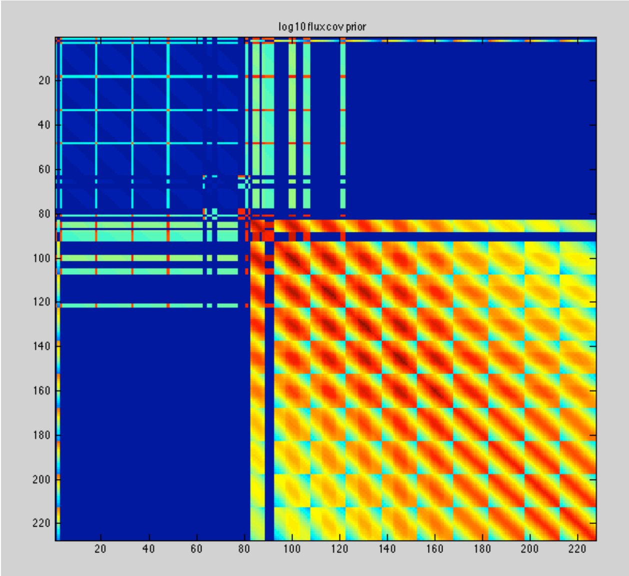
Example of the structure of the model error covariance matrix F for one day worth of measurements.



The R matrix is the sum of the measurement error covariance matrix E and the model error covariance matrix F (in this case for a constant standard error)

2) Prior flux errors:

The generalities of the calculation of the prior fluxes uncertainty is set out in the methodology section. We have added a new subsection: “Prior flux error covariance matrix” based on the replies to your remarks containing additional information.



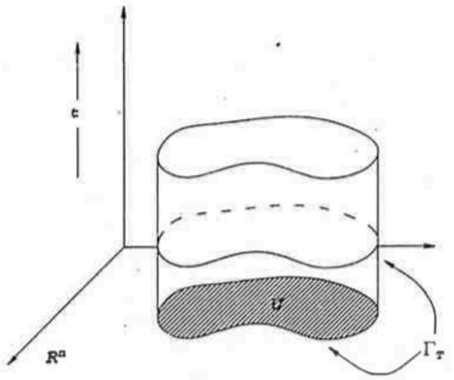
Structure of the flux error covariance matrix
for a single emission period.

3) Posterior fluxes and posterior flux error correlations are dealt with in section 4.3.
We have calculated the posterior fluxes following Tarantola (2005) and Gerbig (2006) and calculated error reduction following Lauvaux (2008). We have compared prior and posterior fluxes and their covariance in the text of the paper. As an example, the error reduction can reach 90% in the central area of Tokyo.

4) Boundary conditions (see also reply to Ref. 1 Pg. 10632, lines 1-4.): This point was not overlooked, although the explanation given in the manuscript was somewhat incomplete. We have developed a sophisticated adaptation of a previously developed method to calculate the background with very high resolution and accuracy (Pisso 2006). Is has been shown elsewhere (Legras 2005) that whenever a quality of the input intial 3D fields is good enough, the reconstructions are accurate and stable.

We proceed, following Pisso (2006), to estimate for every individual measurement the background CO2 concentration defined as the solution of the backwards transport-diffusion (AKA heat) equation.

The background composition is defined exactly as the integral on the first term of RHS in Eq. 1. Formally speaking, y belongs to the parabolic boundary Γ of the heat equation domain. Γ is defined in the extract from Evans (1998).



The region U_T

(ii) *The* parabolic boundary *of* U_T *is*

$$\Gamma_T := \bar{U}_T - U_T.$$

We interpret U_T as being the *parabolic interior* of $\bar{U} \times [0,T]$: note carefully that U_T includes the top $U \times \{t = T\}$. The parabolic boundary Γ_T comprises the bottom and vertical sides of $U \times [0,T]$, but not the top.

Numerically, Γ can be approximated in this as the whole atmosphere 3 days before the measurement has taken place. 3 days are chosen as a time that ensures a reasonable targeted coverage of the flux region under study. It is a tradeoff: less than a day would allow little time for the trajectories to spread backwards and “collect” relevant information, more than a week would let them spread too much far away from the measurement point.

$$= \int_y \mathcal{G}(x,t;y,s)\chi(y,s)dy$$

G is approximated with the density of a cloud of trajectories advected backwards from a given measurement point. Chi is approximated by a tree dimensional output of a chemical model of the CO2 concentration. We have used AGCM for initializing the background (Zhang 2008).

The justification of using a clean air site for the background (beyond the plain search for simplicity when calibrating a new methodology) is based on the very high level of Tokyo concentration vs. other cities and biogenic fluxes. A number of sensitivity tests not included in the initial submission were performed. The conclusion was that other factors such as the PBL height parameterization could impact the results much more than the few ppm that could arrive from the continent (China + Korea). One of the strongest reasons to use the clean air background was that it was already used implicitly with AGCM because of the scaling, and we thought that it was honest to acknowledge such a situation, but there is no explanation in the text.

A time series plot shown at the AGU fall meeting in San Francisco in 2009 compares the CO2 peaks respect to the background concentration with the origin of air masses (TBA ves non TBA). Only a few trajectories are coming from outside the TBA. We conclude that the bulk of the signal (the peaks) is coming from the TBA. Signal from abroad (mainland China, other Japanese populated areas) is very diffused. Therefore the effect of neglecting peaks from beyond the TBA is not significant for the purposes of this paper.

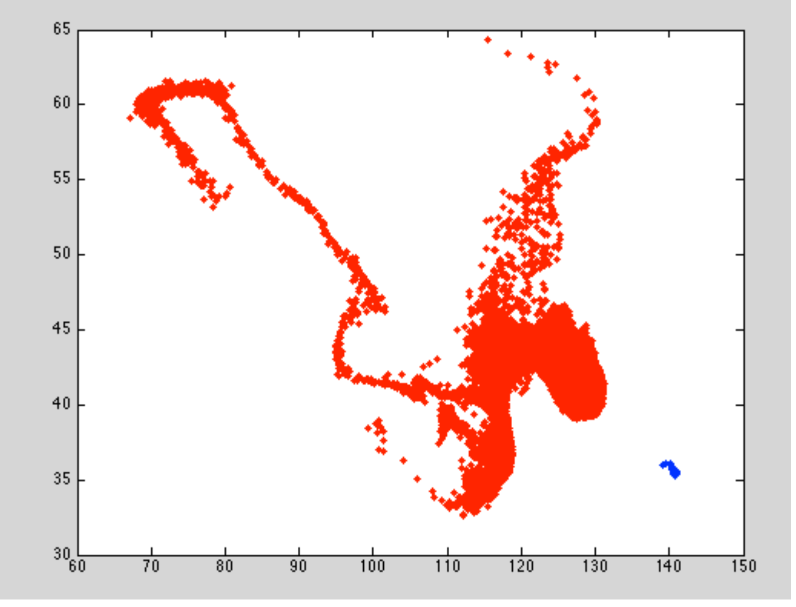
In addition, as shown by Tojima et al. 2010 (Figs 3 and 8) from both Lagrangian and Eulerian transport representation, the North-East Asian plume has a relatively little impact on Japan in general and Tokyo Bay Area in particular.

Description of the experiment (in pseudo code)

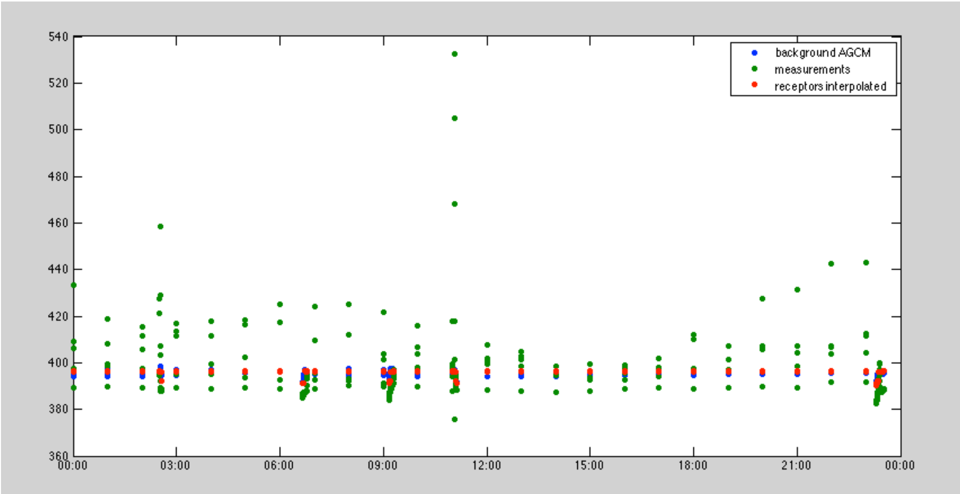
To Loop on days

- For a given day all measurements are loaded
- Select a measurement
- Load trajectory positions 3 days before
- Load corresponding AGCM global CO2 field
- Interpolate on trajectory positions CO2 concentration
- Average concentration
- Assign CO2 concentration so calculated as the background of the selected measurements

In this way all measurements will be assigned a background based on AGCM values 3 days before.



Blue: Original positions (one day worth of data) at measurement points in the TBA.
Red: Positions after 3 day backward calculation.



Measured, interpolated at measurement positions and interpolated in the swarm cloud 3 days before. From the relative magnitude apparent from the picture it can be deduced why the error in the background value used to rescale AGCM raw output has a larger impact on the inversion result than the little extra variability provided by AGCM higher resolution.

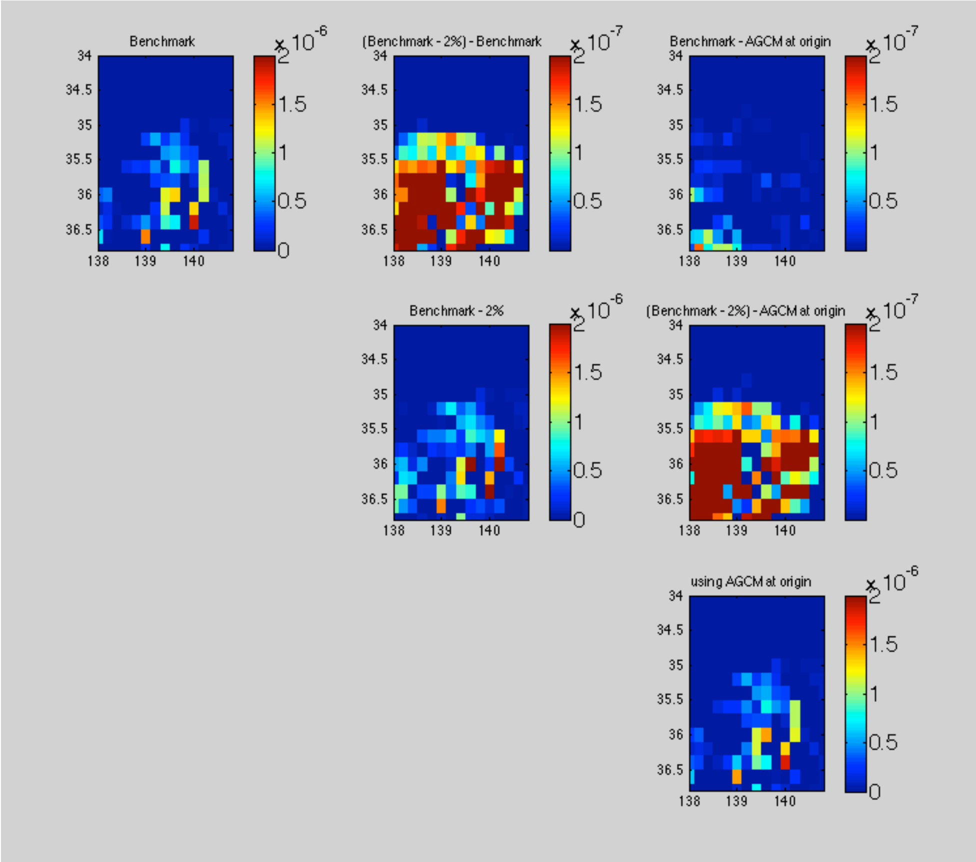
The problem is that, in spite of the increased resolution of AGCM versus a single values hemispheric background, AGCM itself relies on a hemispheric background needed to rescale the raw output in order to fit observations. So using AGCM instead of a single valued hemispheric background accounts for the extra variability represented by the impact of the East Asian continent on the TBA. Is it always the case that this is indispensable? Our tests indicate that not necessarily.

The calculated background was used to invert the fluxes. The result was compared with the inversion with a flat background, and the same flat background perturbed by 2%.

In the Figure, the plots in the diagonal show the inverted fluxes for a day in Jan 2007: (1,1) the benchmark Mauna Loa background; (2,2) the benchmark Mauna Loa background – 2%; (3,3) an inversion performed with the backward Lagrangian diffusive ensembles (BLDE; Legras et al. (2005), Pisso (2006)) connecting with AGCM data (necessarily rescaled to fit the background observations).

It can be observed that the difference between the benchmark Mauna Loa background and the benchmark – 2% (2,1); is relatively large and similar to the difference between the inversion with BLDE and AGCM output with the benchmark – 2% (2,3). In contrast, the difference between (1,1) and (3,3) is small, suggesting that the 2% modification in the benchmark background data prduces a much larger effect than the gain in accuracy provided by the more sophisticated BLDE method.

We argue therefore that in this particular case, the use of an ocean clean air site can be tolerated in terms of error and offers the advantage of simplicity. For this reason we suggested that it was preferable in the ACPD manuscript. We have clarified the choice in the discussion of the final version of the paper.



Technical corrections:

Abstract: During the whole development of this model we have worked with kg $CO2\ m^{-2}.s^{-1}$as many other studies in the literature. The conversion factor $10^9/44$ has been added for those that are less familiar with kg $CO2\ m^{-2}s^{-1}$and prefer $umol.m^{-2}s^{-1}$.

Introduction

The reviewer refers to studies over urban areas that constrain urban emissions using aircraft or surface tower data. We have added additional references (Lauvaux, Gerbig, Lin, etc.). Although the list may not be exhaustive, we have included every reference we have used in this study. We have also added meteorological modeling studies over urban surfaces. (Ballav, Tolk, etc.)

2-25: The text was rephrased.

2-2 There is no statement calling eddy flux an inverse approach ("…inverse methods… . …the other approach consists of local micro-scale flux measurements.").

2-11 The idea we want to convey is that just a large number of observations is not necessarily sufficient to ensure enough constrains. It also depends on the location of the sampling with respect to the current meteorological situation. Even at regional scale, when it seems to be an over determined system it may not be so because of the correlated information of the measurements.

4-1: re-phrased

4-26 We agree with the reviewer.

5-4 The authors are well aware of the relevance of the turbulence in the atmosphere (not only in the lower part). Single particle Gaussian plumes are crude approximations compared to ensembles. Single trajectory calculations are often not completely representative of the flow, to say the least.

5-10 With "data block" we refer to a subset of measurements, typically a day. In our setting all measurements are processed together irrespective of the origin (aircraft, tower or station). We have changed the wording. The section has been rewritten and a more detailed paper is in preparation for GMD.

7-Equation is referred in the text, i.e. when explaining the approximation to G and the transition probability.

8-24 This layer is in the literature defined to be a subset of the boundary layer (Gerbig, Seibert). We say that lies within the PBL, not that it is equivalent. We have included the referee's caveat inn the text.

9-19: We have used the Ide et al., (1997) convention on data assimilation (H, R, B). H=(linearized) observation operator,
R=observational error covariance,
R=E+F instrument + representativeness

B=approximate error covariance of x^b (background state of the model) in variational assimilation

10-2

It may not have been clear in the ACPD version but extensive tests where performed using different error covariance assumptions for the priors. Values of the order of 50-100% variance for inventories such as EDGAR where tested. Values for land fluxes like 100% can be found in previous literature (Thompson 2011). Cross correlations explained above in section 3 (Description of the data and numerical models). We have removed the reference to global estimates criticized by the referee.

Figure 1

We have added an additional figure with information on the vertical distribution of the measurements used in this study.
When a measurement is in the free troposphere it is filtered and not used in the inversion. It simply has zero SRR and therefore doesn't affect the results. It may even be filtered in the SRR matrix in order to reduce the computational burden. As a convective event may happen any day, it is sensible to consider a safe maximal height.

12-3: "... fluxes [the] represent..." OK

12-Section Models:

We have added an additional figure showing the domain of simulation for WRF. The domain used for the inversion can be seen in figures 3 and 4 of the ACPD manuscript.

14-8 Thanks for your suggestion. We will use the difference between models to assess the representation error in the measurement error covariance matrix approximation B. We have also performed a series of tests for PBL height sensitivity.

14-25: This is not a global statement and does not refer to seasonal variability, but rather to the absence of a noticeable inter-annual/long-term trend in Tokyo specifically. It is known that the emission trend has changed significantly in Japan in the past: during the 70s the economic growth was large, whereas in recent years the emissions evolve slowly. These features are robust and correlated to/explained by economic activity and population growth.

Inventories have of course very thoroughly prepared but many do not have any inter-annual variability. This one of the may reasons for doing inversion from atmospheric observations. Because for making inventories we need to rely on third party information, this kind of inversions provide a means to investigate estimates of CO2 fluxes on real time basis.

14-27 As suggested by the referee, we have added a caveat in the text: "If the pixels are independent, the signals will affect singular pixels. But if spatial correlations exist, the prior flux structure will remain unchanged"

Fig 2
The statistical distribution of SSR in time does not provide the special information we want to show in the figure. The point we want to make is that the description of the atmospheric transport is, although uncertain, robust as shown by the consistent description given by two independent meteorological models. The problem of the error correlation is matrix is more related with the inversion technique than with the mere transport modeling. We have not run WRF with ERA Interim. WRF is forced with NCEP FNL global winds. We have run alternatively FLEXPART 8 and Traczilla driven with ERA Interim winds and FLEXPART 6.2 driven with WRF winds.

16-6 Aircraft data are more sensitive to vertical mixing errors and may be affected by different errors than surface measurements.
The measurement uncertainty used in the inversion algorithm is larger for aircraft data, in order to account for the model error also. This has been added to the text.

Fig 4:
The caption was modified.

17-10: We have re-phrased the sentence.

17-14 The figure has been corrected. In effect the vertical axis corresponds to fluxes and not to time. Time is the horizontal axis.

17-18 We have removed the sentence and the reference to target uncertainties.

17-25/27 We have removed the sentence and the reference to global models.

20-20 We have added a caveat to the sentence and the reference to the Essen park. Not a direct comparison, but range of a priori fluxes in mixed urban areas.

20 (21) -25: The 2D is a valid formulation, we agree. The representation of the height of the PBL has a large impact.

21 (22) -6: We mean that we do not have enough PBL height measurements over the whole Kanto region to replace the WRF modeled PBL. Although some studies show the use of PBL height to correct model biases the WRF development is beyond the scope of this work. We follow previous works choosing a 'mixed layer' included in the PBL. Several studies have shown the potential of PBL height measurements in atmospheric modeling applied
to mesoscale inversions. In our case the available measurements to follow this approach are not available.

21-8/12 : We have performed extensive and thorough comparisons with the standard flexpart output based on a 15 minute advection time step and a turbulent perturbation of 18 seconds. The difference was lower than 5%, which is small considering the other sources of incertitude. We conclude that storing the trajectories hourly is sufficient and not significant biases result from that in this case.

21-2/8 (actually 22-2/8):
We refer to an example in which this kind of methodology has been applied to show discrepancies in reported data (Italy underreported HFCs). We are discussing the applicability of this technique to a different gas, CO2.
The comparison paragraph refers to the different initialization between different versions of FLEXPART.

21-9: OK, one "that" removed.

21-9/18: We have removed the paragraph. The results illustrating the point will rather be developed in the future.

References

Ballav, S., Patra, P. K., Takigawa, M., Ghosh, S., De, U. K., Murayama, S., Mukai, H., Hashimoto, S.: Simulation of CO_2 concentration over east asia using regional transport model WRF-CO_2 in revision, J. Meteorol. Soc. Jpn., 2012.

de Foy, B., Zavala, M., Bei, N., and Molina, L. T.: Evaluation of WRF mesoscale simulations and particle trajectory analysis for the MILAGRO field campaign, Atmos. Chem. Phys., 9, 4419-4438, 2009.

Takigawa, M., Niwano, M., Akimoto, H., and Takahashi, M.: Development of a one-way nested global-regional air quality forecasting model, SOLA, 3, 81-84, 2007.

Temporal and spatial variations of the atmospheric CO2 concentration in China. Dongqi Zhang, Jie Tang, Guangyu Shi, Takakiyo Nakazawa, Shuji Aoki, Satoshi Sugawara, Min Wen, Shinji Morimoto, Prabir K. Patra, Tadahiro Hayasaka, and Tazu Saeki, GRL 2008.