We thank Dr Levin for her comments. Here we will attempt to identify the key criticisms, and address each one separately. Dr Levin’s comments are given in italics followed by our replies.

Dr Levin: “Only a few months ago there was a paper published in ACP by Levin et al. Which, based on a completely independent data set of SF₆ background observations starting in 1978, came essentially to the same conclusions as Rigby et al. in the present work. At this point the present work is just a confirmation by independent data.”

Author response: We would like to acknowledge the pioneering work on atmospheric SF₆ that has been performed by Dr Levin and her colleagues in recent years, and have cited several publications from the Heidelberg group in the paper.

Whilst we do share some conclusions with Levin et al. (2010), we do not agree with Dr Levin’s assessment of our paper. There are several unique aspects to our work. Firstly, we are able to extend the Northern Hemisphere (NH) SF₆ record 15-20 years further back than previous studies. Secondly, we derive hemispheric emissions using this record, explicitly incorporating three-dimensional transport information, and using a Bayesian inverse method, neither of which have been done before for SF₆. Thirdly, we use this state-of-the-art chemical transport model and inverse method to quantitatively examine regional SF₆ emissions, using the most extensive networks of SF₆ measurements currently available. Whilst the latter estimates are found to be somewhat poorly resolved, they allow us to quantitatively examine some of the conclusions that were qualitatively discussed in Levin et al (2010). In particular, we determine, using the atmospheric measurements, that it is likely that UNFCCC reported emissions are underestimated. Levin et al. (2010) also suggest that the recent increase in growth “maybe linked to rising emissions from Non-Annex I countries, which qualitatively agree with their economic growth”. We agree with this assessment based on a quantitative top-down methodology, and go further by concluding that it is likely that Asian non-Annex I countries are largely responsible for the recent growth.

We have modified the introduction to ensure that the Levin et al. (2010) conclusions are properly represented in our paper:

“Most recently, Levin et al. (2010) extended the pioneering work of Maisis and Levin (1994), by reporting Southern Hemisphere (SH) measurements beginning in 1978 and NH measurements beginning in 1981, showing a renewed increase in the rise rate from 1997 to 2008. They inferred global SF₆ emissions by estimating the atmospheric burden and taking its derivative with respect to time. A two-dimensional atmospheric box model was then used to simulate atmospheric mole fractions based on these emissions, which were compared to the measurements to check that the derived emissions were reasonable. By studying inventory emissions estimates and economic factors, they postulated that emissions were likely to be under-reported to the United
Nations Framework Convention on Climate Change (UNFCCC, 2009), and that a recent emissions increase was probably driven by non-reporting countries.”

Dr Levin: “Rigby et al. are not able to pin down the emissions in the large continental regions and provide the irrefutable proof that the increase of global SF₆ emissions after about 2000 is mainly happening in the newly developing countries in Asia and that emissions reported to UNFCCC by industrialized countries are significantly too low.”

Author response: We have used the most extensive network of atmospheric SF₆ measurements so far compiled, a three-dimensional transport model and a sophisticated inverse method to derive emissions. We have also provided a detailed assessment of the potential uncertainties in our derived estimates. Crucially we also include a discussion of the covariances that were found between regional emissions estimates. We have attempted to frame the discussion and conclusions in light of these uncertainties and resolution issues. We acknowledge that we cannot ‘irrefutably prove’ that emission growth is mainly happening in developing Asian countries or that UNFCCC reported emissions are too low. Our assessment is that this is a highly likely scenario, given the uncertainties, and we feel that it will not be possible to make stronger statements without a much expanded monitoring network.

Dr Levin: “The reader would thus have expected to see now, after the paper by Levin et al. which was dealing mainly with global emissions, significant progress in the top-down method. This means, a more dedicated modeling framework should have been applied that would be able to “digest” ALL the continuous observations now available at the AGAGE sites, and not only deploying a subset measured during supposedly background conditions, as done in the present study.”

Author response: It is unclear what Dr Levin refers to in her criticism, since she acknowledges in her summary paragraph that non-background measurements were used for the 2004 – 2008 high-resolution inversion. To clarify, we performed two inversions: one at low resolution, using only background measurements, to determine hemispheric annual emissions, and a second in which regional emissions were inferred from 2004 – 2008 using all available AGAGE and NOAA in situ measurements with no “pollution filtering” (the auxiliary material shows all measurements used in this inversion). The reason why only background measurements were used for the first inversion was because the incorporation of pollution events provides little additional information on hemispheric-scale emissions. Further, at the 5x5 degree resolution required for this inversion (for computational efficiency, given the long simulation time), we would have very little confidence in the ability of the model to resolve small-scale pollutant transport.

To clarify that all available AGAGE and NOAA surface measurements were used in the second inversion, we have added the following to the manuscript:

Abstract: “By estimating continent-scale emissions using all available AGAGE and NOAA surface measurements covering the period 2004–2008, with no pollution filtering, we find...”

Page 13531, line 27: “Regional emissions estimation for 2004–2008 were also obtained using all available AGAGE and NOAA observations with no pollution filtering...”

Page 13536, line 16: “Here we ask whether this increase can be attributed to specific regions using all available (unfiltered) data from AGAGE and NOAA networks and the three-dimensional transport model.”

Conclusions: second paragraph: “Regional emissions estimates were obtained for the period 2004–2008 using all AGAGE and NOAA surface measurements.”

Dr Levin: “One of my additional concerns relates to the authors’ estimate of the un-
The uncertainty of EDGAR emissions (global totals as well as the spatial distribution) which was assumed to be only 10%. What is the basis of this assumption and to what extent are the inverse model estimates by Rigby et al. possibly biased by this (small) a-priori uncertainty of EDGAR emissions? I guess that this concerns not only the regional inversion but also the hemispheric and possibly even the global results (Figure 2b). I understand that it is not easy to determine uncertainties of bottom-up emissions inventories, and I suggest that the authors contact EDGAR people to confirm the 10% uncertainty as well as its probable range (for the global total but also for the spatial distribution).

Author response: A 10% uncertainty was obtained by an approximate comparison of EDGAR and Levin et al. (2010) top-down estimates, who show an average discrepancy in the annual release rates of the order of 10%. We thank Dr Levin for pointing out that this was not explained in the manuscript. We have since contacted the EDGAR team, and Jos Olivier, who was involved in the compilation of the EDGAR SF₆ estimates, is now included as a co-author and has provided the following discussion of the development of and uncertainties in the EDGAR inventory:

Section 3.2: “The SF₆ consumption data used in EDGAR v4 is based on the global sales and emissions dataset constructed by Maiss and Brenninkmeijer (1998) for the period 1953–1995. To compile this dataset they considered production data, sales into six end-use categories, other end-use estimates, and atmospheric observations. For EDGAR v4 (1948–2005), global sales data were used that were collected by the RAND corporation through surveys of six major producers of SF₆ (SPS, 1997; Smythe, 2004; Knopman and Smythe, 2007), and modified as described in Maiss and Brenninkmeijer (1998). The data cover all producing countries except for Russia and China (and possibly a very small contribution from India). From this dataset, supplemented with estimates for Russia and China (Bitsch, 1998; Cheng, 2006), annual emissions were estimated as the sum of prompt releases and delayed emissions from banked SF₆. For recent years, national consumption data have also been incorporated, including for example, SF₆ use in semiconductor manufacture, the magnesium industry, sound-insulated windows, soles of sport shoes and automobile tires (UNFCCC, 2010; ESIA, 2007; SIA, 2006; Nike, 2005). For end-use applications where SF₆ is stored in products, and for semiconductor manufacture, which exhibits reduced emissions due to SF₆ destruction during the manufacturing process, default emission factors and banking times were used as recommended by the Intergovernmental Panel on Climate Change (IPCC, 2006). The regional amounts of SF₆ in switchgear in 1995 and the identification of countries within each region that use SF₆-containing switchgear was based on industry estimates (Bitsch, 1998). Per-country estimates of annual stock changes of SF₆ in switchgear were based on their relative share in regional electricity consumption changes in 1995, while the trend in this proxy was used to estimate stock changes in other years. Regional and global total stock build-up over time was estimated such that regional stocks and stock emissions matched industry estimates for 1995 (Bitsch, 1998).

Estimates of the time delay between the sale of SF₆ and its release to the atmosphere from insulated electrical equipment, and the global fraction of sales that were banked in such equipment in 1995, were obtained by Maiss and Brenninkmeijer (1998) through comparison with atmospheric measurements. Independent estimates of these quantities by equipment manufacturers in Europe and Japan, which were used in EDGAR v4 to calibrate the accumulated regional SF₆ stock through 1995, were found to agree well with the top-down values derived. The Maiss and Brenninkmeijer (1998) measurements and calibration scale have not been used in our inversions.

In their analysis, Maiss and Brenninkmeijer (1998) observed a discrepancy between SF₆ sales and end-use estimates in 1995 for two regions. In North America and Europe (including Russia) these discrepancies totaled 1.2 ± 0.4 Gg/yr and 0.4 ± 0.4 Gg/yr respectively, and were identified as unaccounted-for sales to utilities (SPS, 1997). These quantities have been added as unknown sources in the USA, Canada and Russia, and represent about 20% of global reported sales in 1995.
From the size of the adjustments made in the RAND data due to incomplete reporting, the uncertainty in global total production and sales data is estimated at 5 to 10% for the period 1970-2000 and could be as high as 15% in 2005 (2-σ interval). Additional uncertainty in global emissions arise from sources where SF₆ is partially banked, mainly in switchgear stocks, but also, from the 1990s onwards, in soundproof windows, soles of sport shoes and car tires. Taking into account the estimated uncertainty in the emission factors of switchgear, in other applications with delayed emissions and in the factors for Chinese and Russian consumption, we obtain an average global emission factor uncertainty of about 10% in 1970, about 20% in the 1990s, and more than 25% in 2005 (2-σ interval). The resulting uncertainty in global EDGAR v4 emissions is estimated at about 10% for the period 1970-1995, increasing to over 15% in 2005 (1-σ interval). This uncertainty was incrementally increased to 20% between 2005 and 2008, since we expected that our simple extrapolation of the EDGAR emissions was an increasingly poor approximation of the ‘true’ emissions in later years. Global EDGAR v4 and projected emissions are shown in Figure 2.

Compared to global total SF₆ emissions, regional and national estimates are more uncertain. Due to uncertainty introduced by the proxies used and differences in equipment and maintenance practices, uncertainty in regional emissions may be twice as high as for global estimates. For the regional emissions inversion presented in the second part of our analysis (Section 5), we therefore assume a 40% error on the EDGAR v4 emissions (1-σ interval). It is estimated that on national scales EDGAR v4 emissions will have an uncertainty of up to 100% or more (2-σ interval)."
emissions, small biases will be adjusted in the inversion. This point is acknowledged on line 18 of page 13540.

Secondly, it is possible that Dr Levin is referring to “aggregation error”, in which uncertainties are introduced by our solving for continental-scale regions, rather than emissions at each grid point. In this case, errors in the EDGAR estimates within each region cannot be corrected in the inversion, and therefore lead to errors in the derived emissions. This aggregated region approach has been used extensively in previous inversions, and is adopted here, rather than e.g. a variational assimilation method, or ensemble filtering approach, because we feel that the measurement network is too sparse to warrant the use of a more fine-scale inverse method.

Thirdly, Dr Levin may be referring to model transport error. We have attempted to incorporate elements of model transport error into our uncertainty estimates in the following ways. Firstly, we estimate a model-measurement ‘mismatch term’ in equation 1. This term attempts to account for random errors in the short-range simulated mole fraction transport by comparing model mole fractions in grid cells surrounding the measurement site. Secondly, the influence of small biases related to site location are incorporated as described in Section 5. Whilst these methods will estimate uncertainty in short-range model transport, they will not address any larger-scale error sources (e.g. stratosphere-troposphere exchange rate errors). This is a problem common to any inversion using only one transport model, and work is required by the community to determine ways to quantify this error.

We add the following paragraph to section 3.4 to the manuscript to highlight these last two limitations: “Two potential sources of error are unaccounted for in the emissions derived below. Firstly, by solving for emissions from aggregated continental regions, we must assume that the EDGAR spatial distribution within each region is correct. This leads to ‘aggregation’ errors that cannot be quantified here, but may be substantial, given the large estimated national-level uncertainty derived above. Secondly, whilst we attempt to account for random short-range transport uncertainties through the mismatch term in Equation 1, transport model biases and large-scale transport errors cannot be fully estimated here, since only one transport model is used.”

Dr Levin: “In summary, although the present study represents some progress compared to the earlier Levin et al. paper, it is disappointing that Rigby et al. were not able to really proof the earlier suggestions made by Levin et al. which eventually warrants a follow-up paper. What is needed instead would be the application of a more appropriate modeling framework which could take real advantage of the wealth of the new continuous SF$_6$ measurements.”

Author response: We do not understand what Dr Levin means by this comment, and what improvements she envisages to take “real advantage” of the measurements further than some of the comments we make in the paper. We reiterate that continuous measurements were used in the second part of this work. MOZART v4.5, which was released last year, represents a state of the art transport model. The use of 1.8x1.8 degree NCAR/NCEP reanalysis is currently common for this type of work given existing computation speeds, and produces excellent simulations at some monitoring sites. However, as we make clear in the paper, higher resolution meteorology may be needed to accurately simulate transport at some measurement sites (e.g. Gosan, Korea). New reanalysis products are currently coming on line which will be useful for future work. However, as far as we are aware, NCEP/NCAR 1.8 degree is the highest-resolution data set currently available for MOZART for the whole of 2004 – 2008 (the period studied here).

Dr Levin: “What makes this data set so unique? Please substantiate.”

Author response: This data set is unique due to the extent of the measurements of archived air in the northern hemisphere and their combination with measurements of the CGAA and in situ data. The archived NH data have not been presented before and
they extend previous NH records by one and a half to two decades back in the past (starting in 1973) as pointed out by the anonymous referee 2.

The sentence has been changed to:

“These measurements were combined with modern high-frequency GC-MS and GC-electron capture detection (ECD) data from AGAGE monitoring sites, to produce a unique 35-year atmospheric record of this potent greenhouse gas.”

Dr Levin: “Last two sentences: These main findings (i.e. still assumptions) are not new but were Discussion Paper first expressed by Levin et al., 2010. Adequate acknowledgment to these suggestions made in this earlier work thus needs to be given, either here in the Abstract or at least in the Introduction.”

Author response: We do not understand why Dr Levin refers to these statements as “assumptions”. We have quantitatively investigated them using an inverse modeling framework and found strong evidence that both are likely, albeit with large uncertainty. To be clear, other than through the EDGAR prior, with its estimated uncertainty, we have not “assumed” either non-UNFCCC Asian emissions growth or UNFCCC under-reporting in the inversion.

To make sure that adequate acknowledgment of the Levin et al. (2010) work is given, as mentioned above, we have expanded our discussion of the Levin et al (2010) conclusions in the introduction (see response to first comment).

Dr Levin: “Page 13523, line 18: Reference should be given here to the - to my knowledge first - application of SF₆ in modeling studies which is the TransCom2 paper by Denning et al. Tellus B, 1999.”

Author response: We thank Dr Levin for bringing this to our attention. The citation has been added.

Dr Levin: “The second half of the page presenting the comparison results is confusing. I suggest listing these results in a Table.”

Author response: Similar descriptions of the agreement between the two instruments used for the measurements of archived air samples have been previously published (O’Doherty et al., 2009) (Mühle et al., 2010) and we think the information is best represented in the current form.

Dr Levin: “Page 13526, line 3: What were the explicit criteria to reject/accept data from the NH air archive (why should polluted air which is not really representative for a large region be archived)?”

Author response: In contrast to the Cape Grim Air Archive samples which were filled to create an air archive for the SH, a similar archive for the NH is not accessible to us. To address this, we collected a large suite of NH air samples from various sources in the scientific community and measured a wide suite of trace gases in these samples. Given the various purposes of the filled air samples, many samples have to be rejected as only air samples which represent background mixing ratios for a particular gas, here SF₆, are of interest for estimating hemispheric emissions. Samples were rejected in an iterative process based on their deviation from a polynomial fit through all data. We have modified the existing explanation on page 13526 and 7 to reflect this: “In contrast to the CGAA samples, which were filled to created an air archive for the SH, the NH samples were filled during periods when the sites intercepted background air, but with various filling techniques and for different purposes. Samples were rejected in an iterative process based on their deviation from a polynomial fit through all data.”

Dr Levin: “Page 13527, line 25: I cannot find “materials of the surface network” be mentioned in the manuscript.”
Author response: The materials for the network flasks are given in the reference Dlugokencky et al, 1994. We mention that they are identical to indicate we don’t have a potential bias because we are using different flask materials for our cooperative network and our tall towers. The sentence has been clarified in the manuscript.

Dr Levin: “Page 13529, line 1: Please give a number of the agreement (what means "well" in this context?).”

Author response: Changed to: "...also agree with each other to better than 0.1 pmol mol⁻¹"

Dr Levin: “Page 13535, line 28: It is possible to receive information from the EDGAR scientists on how many "top-down" results from atmospheric observations were included in their estimates (and what the uncertainties of the distribution actually are, see general comments).”

Author response: We have added the following to section 4: “The EDGAR inventory has been compiled using 'bottom-up' methods, and draws on practical experience of European and Japanese switchgear manufacturers on the fraction of SF₆ lost during manufacture, commissioning and maintenance. Atmospheric measurements have been used to confirm global estimates by manufacturers of the banked fraction of SF₆ in insulated electrical equipment in 1995 (Maiss and Brenninkmeijer, 1998)”

Dr Levin: “Page 13537, line 8ff: What is the temporal resolution of the meteorological fields driving model transport?”

Author response: Line changed to: “NCEP/NCAR reanalyses were available for use with MOZART from 1990–2008 at 6-hourly intervals at 1.8° x1.8° resolution, with 28 vertical sigma levels extending from the surface to approximately 3 hPa.”

Dr Levin: “Lines 12-13: This sentence is unclear: If I want to “extract” pollution events I would need higher than weekly resolution.”

Author response: This line has been changed to: “Weekly averages were used in order to extract emissions information from synoptic-scale ‘pollution events’ at the high-frequency measurement sites.”

Dr Levin: “Pages 13538, lines 8-11: Where are these uncertainty reductions shown?”

Author response: Perhaps Dr Levin means “aren’t shown”? We have now changed the figure showing the regional inversion results, so that it now highlights the error reduction produced in the inversion. This has meant no longer using the 'bootstrap' uncertainty estimate used in the discussion paper, which was an attempt to account for the regional correlations obtained. We now present traditional inversion uncertainties (augmented by site relocation and scale uncertainty), and rely on the reader noting the correlations presented in Figure 4 when interpreting the uncertainty in Figure 5. We also reduce the scope of the regional emissions estimate figure by comparing only two periods: 2004–2005 and 2006–2008, instead of annual averages (See Figure 1 below). We feel that this presents the information in a more concise way than in the discussion paper.

Dr Levin: “Lines 16-19: This is unclear: Why can’t we be sure to see regional pollution events (or the regional influence) at a (flask) sampling site in the middle of a continent?”

Author response: This is because the sampling strategy at these sites attempts to avoid polluted air. Therefore, the flasks tend to sample the lowest possible mole fractions, even at the continental sites. Hence, we use the weekly minimum, rather than the model mean, which would contain much ‘polluted’ air. Note that if the
background mole fraction at the flask sampling site is higher due to the surrounding continent, this will also be picked up by taking the weekly minimum at the site since this increase in background will be simulated by the model.

Dr Levin: “What must have happened in the respective industrial applications to make GLOBAL SF₆ emissions suddenly increasing by 20% (i.e. from 2001 to 2003) and then suddenly decreasing by 10% in the next year (2004)? This source variation is not visible in the estimates made by Levin et al., 2010. Are the authors sure that this strange behavior is not based on a measurement/calibration artifact?”

Author response: We have attempted to estimate measurement uncertainties and calibration scale uncertainty etc., all of which are carried through to our final estimates. However, whilst every care has been taken, we cannot of course rule out model transport error, or calibration errors. We would point out that the drop from 2003 to 2004, for example, is 0.48±0.53 Tg/yr. Therefore, whilst it may seem like “strange behavior”, such a drop is not statistically significant according to our uncertainty estimates.

Dr Levin: “I think the global source estimates by Levin et al. show higher emissions in 2008 than in 2007. Please check.”

Author response: We thank Dr Levin for pointing out this mistake. It has been corrected in the figure.

References

SIA (Semiconductor Industry Association): Total Semiconductor World Market Sales and
Fig. 1. Regional SF6 emission rates and 1-sigma uncertainty range for 2004-2005 (blue) and 2006-2008 (red) from a) Major emissions regions, b) minor regions and c) aggregated UNFCCC regions.