In the Conclusions section, it would be helpful to have a paragraph describing the implications of the results for climate forcing. For example, the models tend to underestimate the BC concentrations in spring, but overestimate BC in summer. What are the implications of this finding for climate simulations attempting to show the influence of Arctic aerosols on regional climate (e.g., Najafi et al., 2015). In fact, even over mid-latitudes the differences between models are large, as seen in Figure 1. Readers are curious what difference this makes to radiative forcing calculations, and to our understanding of the role of aerosol trends in driving climate change.

Najafi, M.R., F.W. Zwiers, N.P. Gillett, Attribution of Arctic temperature change to greenhouse-gas and aerosol influences, Nature Climate Change, 5, 246–249, 2015.

The reviewer is right in pointing out that our results have implications for radiative forcing calculations. However, it is beyond the scope of this study to quantify these implications in detail. Firstly, because we only evaluate the models for two aerosol components, whereas radiative forcing also depends on many other aerosol components. Secondly, our data set especially above the surface is quite limited and does not cover the full year, so it is difficult to quantify effects for the full year.

However, we added:
Our finding that Arctic BC concentrations in the spring tend to be underestimated by our models implies that these models would also underestimate radiative forcing by BC in the Arctic. This is particularly important because spring is the season when both aerosol concentrations are large and solar radiation is abundant. Furthermore, it is the season when feedback processes, e.g., via ice and snow melting, are most important (Quinn et al., 2008). The concentrations of BC in summer are much lower than in spring, so even with more abundant solar radiation modelling problems in summer would have a relatively small effect on radiative forcing.

1. Page 10437. The authors state that the modeled representations of aerosol loading agree "fairly well" over source regions, but even here there are large differences, with some models yielding double the loading of others. The authors should acknowledge this discrepancy.

Yes, correct, we reworded to:

This approximately covers the latitude range with the highest global emissions where the models agree at least within a factor of two in their simulated column loadings. In contrast, larger differences between the models are found in the Arctic, where column mass loadings vary by more than an order of magnitude

2. Do the authors have any clues about why the Canadian model CanAM4.2 appears to outperform the others? This would be useful information.

This question was also asked by Referee 1 – here is a copy of the explanation we gave:

Unfortunately, we don't know why CanAM tends to produce better agreement with near-surface observations than other models. The group running CanAM made a more thorough investigation on what might cause the differences in the simulated BC concentrations. This analysis has been submitted to JGR (Mahmood et al.; 2015 submitted). CanAM produces higher near-surface and lower tropospheric concentrations in winter and spring than NorESM, CESM, or SMHI. Sensitivity tests with CanAM and CESM give evidence for a strong sensitivity of these results to parameterizations of stratiform wet deposition and aging processes. Stratiform wet deposition in

CanAM is relatively inefficient in winter and spring compared to the other models, which explains the higher surface concentrations in CanAM. We have not determined the cause of these differences but it seems likely that these differences are related to parameterizations of wet deposition scavenging efficiencies in the models.

We added to the manuscript:

The reason why CanAM4.2 captures the spring peak better might be that this model has a less efficient removal through wet deposition under stratiform condition compared to the other models (Mahmood et al., 2015 submitted).

3. Section on sulfate/ BC correlations. The author should begin this section stating why examining such correlations could be helpful.

We added some sentences at the beginning of this section

In this section, we perform a correlation analysis of BC and sulfate. Such an analysis allows some insights into the mixing state of the Arctic aerosol. BC and sulfate largely originate from different sources (although some sulfate is co-emitted with BC by combustion processes). A poor correlation between BC and sulfate means that BC and sulfate either arrive at the measurement stations in distinct air masses or that at least the different aerosol types (even if the air masses mix) remain externally mixed and thus are affected to a different and varying extent by removal processes. On the other hand, a strong correlation implies that BC and sulfate arrive in air masses where contributions from their different emission sources are mixed and that, furthermore, also the aerosol must be internally mixed, as otherwise different removal efficiency for BC and sulfate would lead to decorrelation between the two species. Such a correlation analysis has in fact recently also been performed with measurement data from Station Nord (Massling et al., 2015). In our case, we can furthermore compare measured and modeled correlations, allowing some insights into how models treat the mixing of different aerosol types compared to reality.

4. The discussion of internally vs. externally mixed aerosols comes up late in the paper. It would be helpful to learn earlier in the paper more about how the different models treated aerosol mixtures.

We now specify in table 1 how the aerosols are treated in the respective models:

Model Name	Model Type[1]	Horizontal/vertical resolution Model domain	Meteorological fields; treatment of aerosol mixtures	Periods simulated / output temporal resolution	References
FLEXPART	LPDM	Met. Input data: 1° x1°92L global	ECMWF Operational Analyses; none	2008-2009 3h	Stohl et al. (1998, 2005)
OsloCTM2	CTM	2.8°x2.8°, 60L global	ECMWF IFS Forecasts ; aerosol externally mixed	2008-2009 3h	Myhre et al. (2009), Skeie et al. (2011a, 2011b)
NorESM	CCM	1.9°x2.5°, 26L global	Internal, observed SST prescribed; BC internally mixeed	2008-2009 3h	Kirkevåg et al. (2013), Bentsen et al. (2013)

TM4-ECPL	CTM	2°x3°, 34L global	ECMWF ERA-interim; aerosols externally mixed	2008-2009 24h	Myriokefalitakis et al. (2011); Kanakidou et al. (2012); Daskalakis et al. (2014)
ECHAM6-HAM2	ACM	1.8°x1.8°, 31L global	ECMWFReanalysis; aerosols internally mixed	March-August, 2008, 1h	Stevens et al. (2013), Zhang et al. (2012)
SMHI-MATCH	CTM	0.57°x0.75°, 38L 20-90°N	ECMW – ERA-Interim; BC internally mixed	2008, 2009 1h	Andersson et al. (2007), Robertson et al. (1999)
CanAM4.2	ACM	2.8°x2.8°, 49L, global	Nudged to ECMWF temp.and winds; aged BC internally, near emission externally	2008-2009 3h	Von Salzen et al. (2013), von Salzen (2006)
DEHM	CTM	150km <60° 50km >60°N, 29L 0-90°N	NCEP; internally mixed aerosols	2008-2009 3h	Christensen (1997), Brandt et al. (2012)
CESM1/CAM5.2	CCM	1.9°x2.5°, 30L global	Internal, observed SST prescribed; internally mixed aerosols	2008-2009 1h	Liu et al. (2012), Wang et al. (2013)
WRF-Chem	RCM	100kmx100km 38L 27-90° N	Nudged every 6h to FNL to all levels above the PBL; internally mixed aerosols	March-July 2008 3h	Grell et al. (2005), Zaveri et al. (1999), Zaveri et al. (2008)
HadGEM3	CCM	1.9°x1.3°, 63L global	ECMWF ERA-interim; internally mixed aerosol	March-June, November 2008, January, May and November 2009 2h	Hewitt et al. (2011), Mann et al. (2010)

[1]Chemistry transport model (CTM), Lagrangian particle dispersion model (LPDM), chemistry climate model (CCM), aerosol climate model (ACM), regional climate model coupled with a chemistry module (RCM)

5. It's not clear what is meant by "SO2 (converted to sulfate) to BC emission ratio." Please reword. Also are these mass ratios?

We reworded to: "Based on the ECLIPSE inventory which is available for BC and for SO2, we estimated ratios between those two substances under the assumption that all SO2 is converted to sulfate. The SO2 to BC emission ratio of anthropogenic emissions in the ECLIPSE inventory is 25 globally and 40 north of 50°N. For the GFED biomass burning emissions the emission ratio is only 1.7 …."

Table 3. The practice is to put significant correlations in boldface.

We changed to formatting to significant correlation in bold.

Figure 4 caption. What do red boxes show? What do grey whiskers show?

The red boxes show the observations and the grey whiskers the 1.5 fold interquartile range – we added the information to the figure caption.

Figure 5. Make clear that these are surface concentrations.

We changed to caption so it reads:

Surface concentrations of monthly (month is displayed on the abscissa) median observed eBC/EC and modeled BC. Each row represents one station: (from top) Alert, Nord, Zeppelin, Tiksi, Barrow and Pallas, for late winter/spring (left column) and summer/fall (right column).

Also in figure 6 we added "surface"

Figure 6. What do grey and red boxes show? Perhaps text could say that the boxes and whiskers are same as in Figure 4?

Red are the observations, grey the model results, we added an explanation what the red boxes and grey boxes show.

Figures 7 and 10. Red bar seems unnecessary.

Yes I agree, that the red bar provides redundant information, however I think it makes the figure better readable and doesn't use much space.

Figure 9. As in Figure 6, the description of the Figure is incomplete.

We added the missing information there as well.

Figure 13. It would be helpful to indicate which of the correlations are statistically significant, either by providing text on the plots or in the caption.

In the revised paper version, we plot significant correlation with bold lines, and the ones which are not significant with thin lines.

Page 10433, line 9: "Shall" should be "should." Also, there's a typo further along in this paragraph. Page 10435, line 9: Errant comma.

Thanks, we corrected above mistakes!

Captions for Figures 4-6 have run-on sentences.

We split the first sentences in figure caption 4 into 2 sentences and also made improvements in figure 5 and 6.

Caption for Figure 7. Collapsed sentences are hard to read. E.g. "The top (second from top): : : below (above): : :."

We split the mentioned collapsed sentences into 2 sentences!