We thank Reviewer 3 for the constructive comments and suggestions. We have addressed all comments in our response and feel these revisions improve the manuscript. The reviewer comments are copied here in italics, our response is in plain text, and revised statements from the manuscript are underlined. We also submit a "marked up" version of the revised manuscript.

Reviewer 3

This manuscript describes using SP2 measurements of black carbon aerosol and its mixing state to constrain model predictions of direct radiative forcing in the Arctic region. The methods employed seem to be fairly unique; but I wonder if the results produced are valuable. Basically, there are two separate model runs tested (with appropriate base cases). One run constrains the coating thicknesses on black carbon aerosol with SP2 measurements while allowing the total mass of black carbon to be adjusted to whatever the model simulates. This results in fewer particles containing BC in the model, because the model predicts a smaller mass of BC than the SP2 measurements do but larger non-BC mass. In the second model run, the fraction of BC containing particles relative to all particles is constrained by SP2/UHSAS measurements while the coating thicknesses are allowed to be adjusted to whatever the model simulates. This results in thicker coated BC particles because the model predicts more non-BC mass than the measurements show. My major concern with the manuscript is what does this actually tell us? If the magnitude of BC and non-BC aerosol is 'fixed' in the model (to match observations), either through improved emissions inventories or better transport, scavenging, etc., would that make both of these model runs more closely match each other? If the model isn't getting BC measurements right in any sense (mass or mixing state), then why is constraining just one of these at a time useful? Why not constrain both to the measurements?

We appreciate the reviewer's feedback. As it is important to directly communicate our research goals in this project, we have added several statements to the manuscript to better state the value of our results. These statements are copied below, and we will briefly elaborate on our goals here.

The main objective of this work is to explore the possibility of reducing the uncertainty range in model estimates of the direct radiative effect (DRE) by constraining only the BC mixing state through observations. Previous research has demonstrated that the mixing state of BC is a major uncertainty in estimating the DRE; however, many chemical-transport models (and climate models) do not explicitly simulate mixing state. Instead, modeling studies typically assume entirely internal or external mixing-state assumptions (for references and further discussion see the Introduction). To improve upon this assumption, we use observations of two population mixing-state parameters (coating thickness and fraction of particles containing BC). A main finding of this work is that when using measurement constraints, we reduce the range in estimated DRE by more than a factor of two relative to estimates of DRE using the bounding mixing-state assumptions. The remaining range in DRE is due to model bias in the BC-to-scattering aerosol mass fraction (not the total mass concentration).

Of course, constraining both mass and mixing state would further improve model estimates of the DRE. However, it is not always possible to constrain aerosol mass concentrations everywhere. Aerosol mass concentrations tend to be quite variable spatially and temporally. Conversely, our measurements seem to indicate that BC mixing state is roughly constant throughout the Arctic in Spring (more on this in a later comment). As aerosol models do explicitly track aerosol mass concentration, we choose instead to study the possibility of constraining a parameter models do not typically track.

If the model does reproduce the observed BC-to-scattering mass *ratio* then the two mixing-state constraints would produce identical DRE estimates.

To summarize, this study accomplishes the following:

1. Presents measurements of BC mixing state in the Arctic spring.
2. Combines these measurements with model simulations to reduce the uncertainty range in DRE due to mixing state of BC by more than a factor of 2.
3. Demonstrates how the uncertainty in mixing state of BC is not entirely independent of the ratio of BC to non-BC mass (again the ratio not absolute concentrations).

Overall, this work presents a first step in constraining an important parameter in model estimates of the DRE by combining measurements from a commonly used instrument (the SP2) with aerosol mass concentrations tracked in most models.

We have added the following statements to help address our response to this reviewer comment in the manuscript:

In the abstract:

We find that the pan-Arctic average springtime DRE ranges from -1.65 W m^{-2} to -1.34 W m^{-2} when assuming entirely externally or internally mixed BC. This range in DRE is reduced by over a factor of two (-1.59 to -1.45 W m^{-2}) when using the observed mixing-state constraints.

In Section 3.3 (Results):

The range in the pan-Arctic mean DRE using the bounding mixing-state assumptions (*external, allCoreShell*) is 0.3 W m^{-2} (or about 20%). Conversely, using the measurement constraints on BC mixing state (r_{shell}*-constrained, fBC-constrained*) results in a range of pan-Arctic mean DRE of 0.14 W m^{-2} (or about 9%). Thus, using the measurement constraints on BC mixing state reduces the range in DRE from assumptions of BC mixing state by a factor of two.

In Section 4 (Conclusions):

<u>The estimated range of the DRE using the measurement-constrained mixing states (-1.45 to -1.59 W m^{-2}) is approximately a factor of two less than the estimated range when using the bounding mixing-state assumptions (-1.34 to -1.65 W m^{-2}).</u>

What is the direct radiative forcing if the model is constrained to both BC mass and coating thickness as measured by the SP2? The answer to that relative to the base modeling case might be more useful than the two model runs described here.

We do not have sufficient observations to constrain BC mass concentration horizontally, vertically, and temporally throughout the Arctic spring. In general, we found that the mixing state of BC in the Arctic exhibit less variability than absolute BC mass concentrations (the latter will be discussed in detail in Schulz et al., 2018). Further, the DRE is also dependent on non-BC aerosol mass concentrations.

We do find in our work that constraining the BC mixing state in models is not entirely independent of the ratio of BC-to-scattering mass. If the model predicted the same ratio as observed, than the estimates of DRE using the two mixing-state constraints would be the same.

We have added the following sentences stating this:
In Section 3.2: <u>If the BC mass to total aerosol mass ratio in GEOS-Chem-TOMAS matched the observations, both the shell thickness and fraction of BC-containing particles could be constrained simultaneously.</u>

In Section 4: <u>If the BC mass fraction in TOMAS does match the observations, both mixing state parameters (shell thickness and fraction of BC-containing particles) can be constrained simultaneously and the DRE in both cases would be identical.</u>

A big assumption made in this work is that data from 7 flights over 1 week can be averaged and used as a monthly mean for the whole Arctic region. This is a big assumption that is not fully justified. The flights do not cover a significant region of the Arctic, so where do you get confidence that a campaign average of BC mass and coating thickness is valid for the whole Arctic for the whole month? How uniform are the SP2 measurements along the transects of each flight? Does the spatial variability in SP2 measurements match at all to the variability of the base model case in Figure 7? (It might be helpful to zoom in on the model to cover the flight region, which is very small on this map.) I think it is important to prove that this type of averaging gives useful data.

Thank you for pointing this out. We have added several statements to justify this assumption. First, we would like to clarify that we do not take a campaign average BC mass concentration.

We use simulated BC (and other aerosol species) mass and number concentrations from GEOS-Chem-TOMAS. We do use average coating thickness across the campaign. Overall, we did not find substantial variability in observed coating thickness across flights. This is likely due to relatively long transport (~days) of most aerosol sampled during the campaign. We have added the following statements:

Across all flights (and altitudes), we do not find substantial variability in measured coating thickness. This lack of variability is likely due to atmospheric processing over several days during transport of the air mass to the Arctic region. In the SI, we plot coating thickness for each flight in Figure S3. The low degree of variability across flights justifies the use of campaign averages in our model analysis.

Figure 7 shows the DRE due to all aerosol species, and not only BC. Aerosol optical depth has a strong dependency on non-BC aerosol mass concentration, and so we do not expect variability in BC mass concentration to explain all of the variability in simulated DRE. As discussed in the main text, some of the variability seen in Figure 7 (notably the portions of positive DRE) are likely caused by variability in underlying surface albedo. This is explored further in the Supplemental Material.

The manuscript is very well written and is a nice presentation of what was done. However, there is very little analysis of what was done. The work here needs to be put in context with other modeling methods and other observations. Spend some time telling me what these results mean.

We feel we have added a number of statements further elaborating on our results. In addition, we make a number of comparisons to past observations in the Introduction, Section 3.1, and Section 4 of the main text. Copied below are several of the additions:

We find that the pan-Arctic average springtime DRE ranges from -1.65 W m^{-2} to -1.34 W m^{-2} when assuming entirely externally or internally mixed BC. This range in DRE is reduced by over a factor of two (-1.59 to -1.45 W m^{-2}) when using the observed mixing-state constraints.

The range in the pan-Arctic mean DRE using the bounding mixing-state assumptions (*external, allCoreShell*) is 0.3 W m^{-2} (or about 20%). Conversely, using the measurement constraints on BC mixing state (*r_{shell}-constrained, fBC-constrained*) results in a range of pan-Arctic mean DRE of 0.14 W m^{-2} (or about 9%). Thus, using the measurement constraints on BC mixing state reduces the range in DRE from assumptions of BC mixing state by a factor of two.

The estimated range of the DRE using the measurement-constrained mixing states (-1.45 to -1.59 W m^{-2}) is approximately a factor of two less than the estimated range when using the bounding mixing-state assumptions (-1.34 to -1.65 W m^{-2}).

If the BC mass to total aerosol mass ratio in GEOS-Chem-TOMAS matched the observations, both the shell thickness and fraction of BC-containing particles could be constrained simultaneously.

If the BC mass fraction in TOMAS does match the observations, both mixing state parameters (shell thickness and fraction of BC-containing particles) can be constrained simultaneously and the DRE in both cases would be identical.

Specific comments
The last 2 sentences of the Abstract confuse me a bit - if the measurements of mixing state are so important then why do the differences in the methods seem to be entirely from an underestimation of BC mass fraction?

We have re-written these sentences to better reflect our results. The measurement constraints on mixing state reduce the range in estimated DRE from only using bounding mixing-state assumptions by over a factor of two. The remaining range in DRE across the two measurement constrained mixing states is due to interaction with the uncertainty in the BC to total aerosol mass fraction.

Revised sentences:
We find that the pan-Arctic average springtime DRE ranges from -1.65 W m^{-2} to -1.34 W m^{-2} when assuming entirely externally or internally mixed BC. This range in DRE is reduced by over a factor of two (-1.59 to -1.45 W m^{-2}) when using the observed mixing-state constraints. The difference in DRE between the two observed mixing-state constraints is due to an underestimation of BC mass fraction in the springtime Arctic in GEOS-Chem-TOMAS compared to POLAR6 observations.

Because not much is really said about the field mission as a whole, I wonder if Table 1 and Figure 1 could be moved to Supplemental?

This is a fair suggestion. As the level of detail in Table 1 does not add to the paper, we have moved this to the Supplemental Material. However, as we do not feel that the paper has an excess of figures, we feel it is helpful to retain Figure 1 in the main text. Figure 1 provides context for the range of vertical levels shown in Figures 5 and 6.

Revisions:
The reference to Table 1 in the main text has been changed to Table S1 (Supplemental Table 1). In the Supplemental Material, we have added this Table with a short description.

Section 2.4 is a little confusing and I had to read it several times to fully understand it. I wonder if an illustration or schematic of some kind explaining this procedure should appear in the Supplemental? I don't think combining data from two SP2's in this way is common, so more explanation is warranted. Regarding this procedure, was the core-shell Mie model used to relate the core BC with coating to the scattering signals that the UHSAS measures? As described, it seems just diameter derived by the SP2 measurements was used, but this is not quite right - the light scattering signal is what should be compared.

The reviewer has raised a good point. We have added a section to the Supplemental material giving details and an example of how the data from the two SP2s and the UHSAS was combined. In this added section in the Supplemental, we provide enumerated steps to clearly demonstrate the procedure. Additionally, Supplemental Figure S1 provides a graphical example on this process.

Does the pattern of coating thicknesses in Fig 3 follow any specific functional form (e.g. diffusion-controlled growth laws)? Could a functional form be used to derive coating thickness over the whole range of BC cores in Fig 4a?

A parameterization based on the fit line in Figure 4b (or perhaps the results in Figure 3) is a really interesting idea; however, we feel it is beyond the scope of the current manuscript. The main reason is that it is not clear from the observations what underlying microphysical processes may control the fraction of BC containing particles (e.g., coagulation time scale, condensation rate, cloud processing), and these processes will influence the mixing state outside of the measurement-constrained size range. While it does not explitictly include physical processes, the fit line in Figure 4b is a polynomial fit that best matches the observations. As this may be helpful for future studies we have included the form of the fit in the caption to Figure 4b. We have added the following statements to the main text to clarify this:
The solid line in Figure 4b is the best fit to the experimental data shown in Figure 4b. A polynomial function was used because this gave a good fit to the experimental data. The coefficients for this fit are provided in the caption to Figure 4; however, we note this fit is derived only from measurements in the springtime Arctic and is not based on any underlying understanding of the chemical or physical processes involved. With this polynomial fit, we…

And in the caption to Figure 4b:
The solid black line is a polynomial fit to the observations following the form (with x in nm): y = 1.25 – (1.52*10-2)*x + (8.05*10-5)*x2 – (2.00*10-7)*x3 +(2.31*10-10)*x4 – (1.00*10-13)*x5.

Fig 4, bottom panel, seems to show a single flight that was very different than the others. Can you explain this? If there is an outlier flight, is the average fit line really useful, or should the outlier be excluded?

There is not an obvious reason why this flight is different than the others. As we do not have a reason for the difference, we do not feel justified in removing it from the average.

Fig 5 Observations line has a weird squiggly part at the upper end - what is this?

This is most likely caused by a lower number of particle counts in this size range leading to a noiser fit.

The DRE numbers need some context. Are these significant changes? How do these number compare to other forcing mechanisms? Is BC the biggest forcer in the region?

In this work, we are really interested in the range of DRE due to uncertainty in BC mixing state. To add some context to the change in this range due to measurement constraints on mixing state, we have rephrased our results referring to a factor of two reduction in the range of estimated DRE (these statements have been shared in response to earlier comments). This reduction in uncertainty in DRE is the context we wish readers to take away from this study.

A factor of two reduction is certainty significant. A comparison of our simulated DRE in terms of absolute numbers is difficult, as there are not many studies providing a similar estimate of the direct radiative effect (as opposed to a direct radiative forcing). In this work, we follow the terminology described in Heald et al. (2014), where a forcing is relative to some time period (usually 1750 or 1850) and an effect is relative a case with no aerosol. We do not feel other effect mechanisms in the Arctic (such as the cloud effect, generally 10s of W m^{-2}) are relevant comparisons. Generally, a change in DRE on the order of 0.15 W m^{-2} (as we see for mixing state) is about on the order we may expect to see in the global mean due to turning on and off a source sector (such as biomass burning). However, as we do not explore this in this work, we feel this would be a distracting comparison to make.

With regard to the question about BC being the biggest forcer in the region, in the first paragraph of this manuscript we cite studies suggesting BC may be an important pollutant contributing a positive radiative forcing; however, the total magnitude of this forcing is uncertain. Overall, BC is likely not the biggest total forcer in the region. Only considering aerosol, most of the aerosol optical extinction is likely due to non-BC aerosol. Quantifying the contribution to DRE from the various aerosol chemical species in a model with size-resolved aerosol microphysics is challenging and beyond the scope of this study, as removing BC from the model would lead to non-linear effects in the optics calculation (for example, in the core-shell assumption removing BC would alter the total size of particles). However, it is likely that most of the aerosol absorption is due to BC. Assuming an external population, we can estimate that roughly 1-5% of

the total aerosol optical extinction in the Arctic is due to BC. We have added the following statement to Section 3.2:
In the model, these BC mass fractions translate to a contribution of roughly 1-5% of optical extinction from BC to the total aerosol optical extinction in the Arctic (calculated assuming fully externally mixed BC).

The Conclusion states that this method should be applied globally, but I'm not yet convinced that this method provides any valuable insight. How do your model results compare to satellite remote sensing measurements? That would give an indication of how your new modeling method matches the real world.

While we do not claim to perfectly constrain the DRE, we do feel we make valuable progress in reporting measurements of BC mixing state and integrating those into the model to reduce the range of estimates of DRE. Comparing to optical measurements (satellite or perhaps AERONET), is an interesting idea but beyond the scope of the present manuscript. As we do not include any direct comparisons that show our method improves model estimates of the DRE, we have removed the sentence from the Conclusions recommending this method be applied globally.

Technical corrections

Page 6 line 15 "and, and"

Thank you, this has been fixed.

Page 7 line 2 "can results"

Thank you, fixed.

Page 11 line 6 should be "Figure 4b"

Fixed.

Page 13 line 32 "0.11" needs units

Thank you, this has been fixed.