Thank you to all reviewer's for your suggestions. We respond to each comment below in italics. We will also submit a version of the article with all the changes marked to the editor.

**Response to Andrew Sawyer**

**This is not a review of the paper, only a short comment about the MODIS aerosol data used.**

**In this study, the authors use MODIS Collection 6 data for MODIS Aqua, but MODIS Collection 5 data from MODIS Terra. The latest data version, which is expected to remain the standard for at least the next few years, is Collection 6; older versions should generally be considered obsolete and avoided where possible. In case the authors are unaware, I thought I should mention that Collection 6 data are also available for MODIS Terra. In the 'Dark Target' data set which it looks like the authors are using, there were several important bug fixes and updates going from Collection 5 to Collection 6 (see the Levy et al., 2013 paper cited in the manuscript). It may therefore be advisable to repeat the Terra portion of the analysis with Collection 6 data, if this is practical. My expectation is that results for eastern North America may not change too much, but western North America and China may change more significantly. This is consistent with what the authors see in e.g. Figure 6 (Terra/Aqua differences are not consistent with diurnal sampling effects), and so using the latest data version for both platforms may simplify the analysis somewhat. I realise, though, that this may be quite a computational burden to update the data set use at this stage. MODIS Collection 6 also includes Deep Blue aerosol data covering all land surfaces, and therefore may be of additional interest for this type of study. Our experience suggests that the two data sets are quite similar over North America, though, so it might be that not much is gained from using both Deep Blue and Dark Target as model constraints (they would probably have a similar effect on the model over North America).**

*Thank you for this comment. We started this work using Collection 5, but upon release of Collection 6, we redid all of our analysis with the updated product for Aqua MODIS, but did not reprocess MODIS-Terra due to the timeline of release (and the need for many years of data). However, we kept the discussion of Collection 5 because previous studies using a satellite-based $PM_{2.5}$ method have relied on Collection 5 and there is substantial difference (as shown) between Collection 5 and Collection 6, which could be a significant source of uncertainty in those previous estimates. We have now removed any discussion of Terra from our results, ensuring that we are using the latest data products in our analysis.*

**As a minor unrelated point, I notice that C. A. Pope's surname is typeset as "Pope III"**

**in the paper and reference list. The surname should just read "Pope". If the authors are using LaTeX/BibTeX then I think the correct formatting can be achieved by writing "Pope, III, C. A." rather than "Pope III, C. A.**

*This is a typesetting issue which we will be sure to catch in the final version.*

**Response to Anonymous Referee #1**

**General comments: This paper estimates sources of uncertainty in exposure estimates by analyzing differences in model versus satellite-derived PM2.5 and various concentration-response functions. I think that this is an important contribution to the field because it compares the influence of individual assumptions on estimated mortality and compares results to different studies.**

*Thank you for your review. Responses are in italics below each comment.*

**Specific comments:**

**-Pg. 25333, line 14: Did the five-year population estimates indicate that a linear interpolation was appropriate in China?**

*Population growth is likely not linear in China over the time period, but it is outside the scope of this paper (beyond acknowledging this as a source of uncertainty) to determine the actual annual change in population given the projections for 2005, 2010, and 2015. Additionally, we calculate the average annual mortality over a 8 year time period, so this should reduce some of the uncertainty.*

**-Pg. 25334, line 4: How would you expect different spatial resolutions of baseline mortality data to influence your results and comparisons with the studies mentioned?**

*Spatial resolutions of all variables will impact local results, including the baseline mortality, specifically in regions where grid boxes might straddle two countries or states with very different mortality rates. Many previous studies use a country-wide regional baseline mortality rate; we go one step beyond by using state-specific mortality for the US, however we acknowledge that further spatially resolved mortality data (particularly in China) would provide more accurate local estimates of the burden. To address this concern, we had added discussion throughout the*

*text about resolution and referenced Punger and West (2013) and Thompson et al. (2014) both of which discuss the impact of resolution for estimating health impacts.*

**-Pg. 25338, line 18: Specify here that satellite-based estimates are gridded at the same spatial resolution as the unconstrained model instead of (or in addition to) further down in this section.**

*Done.*

**-Pg. 25339, line 5: Explain why MODIS and MISR were both used (strengths/weakness of each dataset), and why collection 6 for Aqua and 5 for Aqua and Terra. Also, how does this compare with also using SeaWiFS in the more recent van Donkelaar et al. (2014) work?**

*We used MISR and MODIS to highlight their differences because previous studies have used either or both of these for a combined AOD product. Additionally, MISR is generally biased low and MODIS is biased high (in comparisons with AERONET). This has been noted in other studies, and we address this discrepancy in the text and reference these other studies.*

*We started this work using Collection 5, but upon release of Collection 6, we redid all of our analysis with the updated product for Aqua MODIS, but did not reprocess MODIS-Terra due to the timeline of release. However, we retained the discussion of Collection 5 in our submitted text because previous studies using this method have relied on Collection 5 and there is substantial difference (as shown) between Collection 5 and Collection 6, which could be a significant source of uncertainty in those previous estimates. However, given concerns about the data quality of Collection 5 (particularly raised by Andrew Sawyer) and that this comparison is not central to our study, we have now removed all MODIS-Terra data from the analysis to ensure clarity.*

*van Donkelaar et al. (2014) use a combined product of MODIS, MISR, and SeaWiFs. Using this product would likely also provide different results because it is a different product. The discussion however would remain the same, that different satellites products have different biases and therefore would result in different estimates. We want to stress that the goal of this paper is not to design a $PM_{2.5}$ product as with the series of papers by van Donkelaar et al. (2010; 2013; 2014; 2015) but to discuss uncertainties in these products and stress the necessity of understanding the data used in assessing health impacts.*

**-Pg. 25341, line 22: This is mentioned briefly later in the manuscript, but do you have any indication of how MODIS and MISR compare with observations at shorter timescales (daily)? Are the satellites overestimating or underestimating peaks and how could this impact exposure estimates?**

*The satellites both underestimate and overestimate peaks in AOD depending on the time and location. Compared to AERONET sites, the mean normalized gross error in daily AOD for MODIS is 75% in the western U.S., 50% in western China, 35% in the eastern U.S. and in eastern China. Determining chronic exposure from long-term averages should reduce some of the uncertainty from not capturing daily variability (unless there is a systematic bias). Our discussion in the sensitivity analyses section of using average AODs to compute $PM_{2.5}$ alludes to some of the uncertainty in this, that daily variability can influence the annual means.*

**-Pg. 25342, line 17: Is there a difference in PM2.5 components between China and the U.S. that could influence results?**

*Yes, this is why we do some sensitivity tests examining the potential biases in aerosol composition (i.e. the sulfate and BC only sensitivity tests). As to the toxicity of different components, this is still an open area of research (i.e. Chung et al., 2015), therefore we do not estimate or discuss speciated $PM_{2.5}$ estimates here.*

Chung Y, Dominici F, Wang Y, Coull BA, Bell ML. 2015. Associations between long-term exposure to chemical constituents of fine particulate matter ($PM_{2.5}$) and mortality in Medicare enrollees in the eastern United States. Environ Health Perspect 123:467–474; http://dx.doi.org/10.1289/ehp.1307549.

**-Pg. 25344, line 21: Is there a figure or table with these AERONET results?**

*We have added a figure with these results.*

**-Pg. 25344, line 25: Better in the eastern US and eastern China than the western parts of each country? Please clarify.**

*We have added to the text:"than in the western U.S. and western China"*

**-Pg. 25348, lines 21-22: What are examples of some of these regional sources?**

*We have added to the text: "e.g biogenic aerosol in the Southeastern U.S."*

**-Pg. 25352, line 4: Implications of a study that is smaller and using only white participants?**

*The implications are that it may not be representative of the larger population that might not have the same demographics.*

**-Pg. 25353, line 3: Can you comment on how the results of Chen et al. (2013) (or another China-specific study) would impact your results?**

*We cannot make a quantitative comparison with the Chen et al. (2013) study because they used total suspended particles (TSP) rather than $PM_{2.5}$ (and most epidemiological studies agree that the most harmful constituent is the fine fraction). Pope and Dockery, 2013 do compare the Chen et al. (2013) results with other studies and find that the elevated risk is lower than found in the Laden et al. (2006) and Pope et al. (2002) studies which is in line with Aunan and Pan (2004)as mentioned in the text. We have added the Chen et al. (2013) reference to the text.*

**-Pg. 25355, line 17: What was the spatial resolution of the Lelieveld et al. (2013) study? As the authors mentioned earlier in the text, spatial resolution might be driving some differences in more populated grid cells.**

*Lelieveld et al. (2013) uses a model with a horizontal resolution of ~1.1° x 1.1° which is somewhat coarser than our model resolution.*

**-Conclusions: I think that the discussion of Figure 9 needs to be expanded, which could also include a brief discussion of how different spatial resolutions (among different models and between model and satellite-based estimates), emissions inventories, region-specific health data, etc. impact these estimates. And, if possible, it might be helpful if the authors could give some sort of general recommendations regarding the "best practices" of the factors that are most important for future authors to consider when estimating exposure at either at a global scale or for China and the U.S. specifically.**

*We have moved Figure 9 to section 4 and expanded the discussion. We have added more comments regarding resolution throughout the paper and added comments about other sources of uncertainty to the conclusion. As stated, the CRF seem to be the most important for attributing mortality, as such, we prefer to leave the recommendation about the "best practices" to experts in that field and instead suggest that a range of results are presented for comparison to other studies which also use a range of different methods.*

**-Table 1: Would it make sense to include the updated estimates of Lelieveld et al. (2015)? Also, doesn't Evans et al. (2013) provide estimates with different CRFs?**

*Lelieveld et al., 2015 had not been released at the time of submission but has now been added to the text along with several other estimates. Yes, Evans et al. (2013) does provide estimates with different CRFs which we have now added along with several other studies.*

**-Table 1: Does the heading mean that some of the U.S. estimates are for all of North America and China for all of Asia/Western Pacific? Please clarify in the caption because it's unclear if these refer to studies that were specific to the countries or to regions. You mention that the Anenberg study is regional in the text but it's unclear about the others. Also, are all of these studies for similar years?**

*We have clarified the region for the estimate in the table next to the study and added an extra column in Table 1 noting the year of the estimate.*

**-Figure 4: Anything available for China? It would be helpful to see a plot based on any available data, maybe AERONET as mentioned in the text?**

*There are now a significant number of surface monitoring sites in China, but there were none with long-term measurements (that were publically accessible) for the period of 2004-2011. We have added a figure with AERONET sites and comparisons (Figure 7).*

**-Figure 9: Are the previous estimates including only country-based US and China specific studies, or are some regional? This figure is very helpful and I would appreciate a more in-depth discussion.**

*Yes, there are. We have tried to clarify this in Table 1 and have added to the discussion on Figure 9.*

**Technical corrections:**

**-Pg. 25330, line 19: "on the order of.."** *fixed*

**-Pg. 25344, line 20: Are you referring to Fig. 5a (exposure plot) or to Fig. 6a (AOD plot)?** *Figure 6, fixed*

**-Pg. 25345, line 2: Missing a period or this is a run-on sentence.** *This is a typesetting error (we prefer U.S. to US) which we look for in the final version.*

**-Pg. 25345, line 10: Fig. 6b?** *yes.*

**-Pg. 25346, line 17: "Requires model output," that is unnecessary as written.** *Fixed.*

**-Pg. 25351, lines 19-23: Much of this is repeated from the introduction and could likely be cut or shortened if the word limit is an issue.** *We chose to leave it for clarification.*

**-Table 1: Do you mean that Table 4 provides additional information?**

*Yes, thank you for noting this mistake.*

**-Table 3: Define threshold abbreviations in caption.**

*We have altered the table.*

**-Figure 2: Can you change the font size of the individual studies? This figure is difficult to read, and the information might make more sense in a table.**

*We have made the font bigger.*

**-Figure 7: Please clarify that abbreviations are also defined in Table 3.**

*Done.*

**-Figure 8: Shouldn't this refer to the last column in Table 4?**

*Yes. This has been fixed.*

**References mentioned above: Chen, Y., Ebenstein, A., Greenstone, M. and Li, H.: Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy, Proc. Natl. Acad. Sci. USA, 110(32), 12936–12941, doi:10.1073/pnas.1300018110, 2013. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525(7569), 367–371, doi:10.1038/nature15371, 2015.**

**<u>Response to Anonymous Referee #2</u>**

**"Exploring the Uncertainty Associated with Satellite-Based Estimates of Premature Mortality due to Exposure to Fine Particulate Matter" by Ford and Heald. This study estimated the premature mortality in US and China by using satellite data and GEOS-Chem model simulations, and quantified the uncertainties of the results caused by different methods and dataset used to derive. The study is useful to constrain the estimated health effect due to increased concentrations of fine particulate matter with satellite-based observations. I have a few major concerns and some specific comments as below. Firstly, the relationship η, which links PM2.5 and AOD, is derived from the GEOS-Chem simulation in this study, although the authors have conducted a couple of sensitivity experiments to understand how much difference would be caused due to the uncertainty in η, I am curious that how these would be different from the real η if directly linking the surface PM2.5 and satellite AOD. Secondly, the relative risk (RR in the paper), which is a key factor to determine the premature mortality due to exposure to PM2.5, differs significantly because the pathophysiological mechanisms are currently unclear. The authors assessed the uncertainty of the estimated mortality rate by using different PM2.5 concentration response function. I wonder is it possible to give us a "better choice" for the study region such as US and China? Finally, the authors have conducted a few sensitivity experiments to test how different factors impact their estimations, which is a good attempt to improve our understanding. The disadvantage is lacking of the detailed explanations and discussions on theses sensitivity results.**

*For comparison of the actual η vs. model η using ground-based measurements, we refer the reviewer to Snider et al., 2015 and add this reference to the text and for discussion of the satellite η, we refer the reviewer to van Donkelaar et al., 2012. We have added text to the sensitivity discussion to better address these questions. As the CRF seem to be the most important for attributing mortality, we prefer to leave the recommendation about the "best choice" to experts in that field and instead suggest that a range of results are presented for comparison to other studies which also use a range of different functions.*

Snider, G., Weagle, C. L., Martin, R. V., van Donkelaar, A., Conrad, K., Cunningham, D., Gordon, C., Zwicker, M., Akoshile, C., Artaxo, P., Anh, N. X., Brook, J., Dong, J., Garland, R. M., Greenwald, R., Griffith, D., He, K., Holben, B. N., Kahn, R., Koren, I., Lagrosas, N., Lestari, P., Ma, Z., Vanderlei Martins, J., Quel, E. J., Rudich, Y., Salam, A., Tripathi, S. N., Yu, C., Zhang, Q., Zhang, Y., Brauer, M., Cohen, A., Gibson, M. D., and Liu, Y.: SPARTAN: a global network to evaluate and enhance satellite-based estimates of ground-level particulate matter for global health applications, Atmos. Meas. Tech., 8, 505-521, doi:10.5194/amt-8-505-2015, 2015.

**Specific comments:**

**p.25333, at top of the page, it is difficult to see here how PM2.5 contributes to health from the equations Eq.1 and 2; please add an equation to describe the link between PM2.5 and RR, if possible.**

*These are standard equations to determine the attributable fraction of mortality due to a specific factor, not just for $PM_{2.5}$ exposure. Therefore, we prefer to leave the equations as they are in order to align with what is standard in the literature. The concentration response functions are described in Section 2.4 which we refer to in this section.*

**p.25333 last paragraph, You use crude death rates, instead respiratory disease, to determine baseline mortalities, which will overestimate the burden of death due to air pollution. Can you find and use the death rates from non-accidental death? In China, it is even cruder as population rather than death rate is used to estimate. Can the authors estimate the biases caused by this?**

*We do not use overall crude death rates (or "all cause"), but the death rates specific for each disease (respiratory, heart, and lung cancer) for each state and each year. Same for China, these are not crude death rates; they are year-specific age-standardized mortality rates by cause from the WHO for China as described in Section 2.2.*

**Fig.2. the text is too small to see, I suggest the authors to make this figure bigger.**

*This may be related largely to ACPD formatting but we have made the font bigger and will ensure that this is legible in the final version.*

**Table 2: table caption, "… in Eq. (8)…", should be Eq.(7).**

*Yes. Thank you.*

**p.25339, line 5-20, You removed the satellite observations with high AOD (>2.0), can you explain how do you decide this threshold? since AOD could be very high (over 2.0) in some cases, e.g. heavy pollution?**

*We choose this threshold to attempt to take care of cloud contamination as discussed in the methods from our previous work (Ford and Heald, 2012, 2013). We acknowledge this and note that this could remove high pollution events, particularly in China.*

**Fig.5. How do you compute the values shown in Fig.5? Which field in Eqs corresponds to the results shown here? can you clarify that if the results are P in Eq. (2), or others?**

*This is a cumulative distribution plot showing the percent of the population exposed to different annually-averaged $PM_{2.5}$ concentrations calculated using the population (which is P in Equation 2) and the concentration (which is C in the RR equations). It is calculated as a sum of the population in each grid box which has an annual average concentration at or above each concentration on the x-axis. We have added this to the text to clarify.*

**p.25344, bottom paragraph. It would be good to give a plot to show the AERONET sites used in the comparisons in both US and China. The quantitative comparison of AOD between satellite and AERONET is not shown in a plot and/or table, and not even given in the text. Please include these comparisons.**

*A figure has been added (now Figure 7).*

**p.25347 and Fig.7: As I can see the NMB is apparently largest in Southeastern China from the experiment vertical profile, but there are no explanation in the text. For the test Relative humidity, there are positive NMB in Southeastern and Northeaster China, but negative NMB in western and Central China. The necessary explanations and discussions are needed in this sensitivity tests.**

*We have added more to this discussion to clarify these results.*

**Figure 8, figure caption "… in Table 3", should be Table 4.**

*Fixed.*

**Figure.9, I suggest to move the Figure 9 and associated text into section 4, rather than last section, i.e. section 5.**

*Done.*

**Response to Anonymous Referee #3**

**The authors present an interesting paper in which they estimate the health burden of PM2.5 in the US and China, compare those estimates with previous studies, and then explore uncertainties in the calculation due to satellite estimates of PM2.5, health function parameters, etc. The paper is unusual in its detailed treatment of atmospheric science and satellite retrievals, as well as concentration-response functions within a single paper. To my mind, this is both a strength of the paper – as different uncertainties are addressed within a single paper – and a weakenss, as the discussion ranges over a wide body of literature and can be hard to follow at times.**

**Overall, my sense is that the paper is a worthwhile addition to the existing literature.**

*Thank you for your review.*

**However, I feel that the presentation of the complex discussion can be improved and I have some general questions or concerns about the approach:**

**1) It seems that the main points of the paper are summarized in Figure 9. Differences in health burden are presented when exposure estimates are driven by the model vs. two satellite estimates, and then uncertainty analysis is performed on 3 parameters individually. Given that the uncertainty due to individual parameters has been estimated by the authors, I am surprised that they did not take the next step to estimate an overall uncertainty given uncertainty in those parameters individually. Also, is the uncertainty in CRF in Figure 9 a simple uncertainty given the confidence intervals from a single study, or does it somehow account for uncertainty as illustrated in Figure 8 or Table 4?**

*Our goal was to show a range of uncertainty due to the specific parameters that we explored. We did not examine every sources of uncertainty, especially with regards to the model. Estimating an overall uncertainty would be a much more complicated process that would likely require a much more thorough examination of the parameter space (along the lines of Lee et al., 2013). For this reason, we do not want to provide an overall uncertainty that might misrepresent the analysis done here. The grey lines show the uncertainty from the confidence intervals of the Krewski et al. study, the colored bar shows the uncertainty from Figure 8 (now Figure 9).*

**2) The goal as stated p. 25354 is "to explore how mortality burden estimates are made and how choices within this methodology can explain some of [the discrepancies among previous**

**studies].” The authors have succeeded in estimating how different modeling choices or parameters contribute to the overall uncertainty. But as there are many differences among many different studies, I don’t know that this paper helps to clarify those differences in results – or it certainly does not explain why a particular study is high or low vs. others. The results shown are not surprising given the current literature, and since previous studies have often included analysis and discussion comparing their results with others, I’m not sure that the authors add something new here. The results are interesting and seem to add to the literature, but I’d encourage the authors to think harder about what is new and present that more clearly.**

*We have added to the discussion of Figure 9 (now Figure 10) and the conclusions which we believe does a better job clarifying the differences.*

**3) Related to #2, despite the complexity of the paper and its extensive discussion, I thought the bottom-line messages were rather few. The authors should consider reorganizing in places to reduce repetition, and/or removing excessive discussion.**

*Thank you for this suggestion. As the reviewer highlighted, this was a study with a wide-ranging discussion, and we have endeavored in this final version to streamline the discussion. In particular, we believe that moving our discussion of the overall uncertainty from the conclusions into a separate section (4.4) clarifies the conclusions and contributions of the study.*

**More specific modeling questions & concerns:**

**1) I might be wrong, but I’m not aware that anyone uses a linear function as described in equation 3.**

*It is not commonly used for the reason stated in the text, that it produces very large estimates for high concentrations. However, it is used as an alternative concentration-response curve in Cohen et al. (2004; 2005), and the Hoek et al. 2013 is presented as linear by Pope et al. (2015). Our goal was to start with the simplest form and then explain alternate functions and the impact this has on the estimates. We have tried to clarify this is the text. Additionally, in order to be more in line with recent literature, we have chosen to use the Burnett et al. (2014) as our baseline function and discuss other forms as sensitivity tests.*

**2) They assume the C0 to be zero. I don’t think that there are other studies that use zero as C0, and I am concerned that it requires a significant leap of faith to assume that the same concentration-response relationships hold at concentrations below which we have observations. If the authors keep this assumption, they should do more to discuss and justify this choice.**

*Thank you for raising this point. Sun et al. (2015) uses a threshold of zero and many studies which estimate the change in mortality from different sensitivity simulations, which use a base case concentration (for changes over time or comparing natural and anthropogenic sources) as the $C_0$ value are not accounting for a threshold if those values are less than the threshold of the given CRF (e.g.. Anenberg et al., 2010; Silva et al., 2013). Many other studies also test using threshold values below the observations (e.g. Johnston et al., 2012; Evans et al., 2013).*

*The reason one might not include a threshold is that, as stated in the text, most experts in health impacts of ambient air quality agree that there is no population-level threshold (Roman et al., 2008). Additionally, the literature is full of assumptions about the shape and magnitude of the CRF above which there were no observations in the original studies (it is relatively linear in the range of observations, which is why there is so much discussion on the shape of the CRF at high concentrations) which is even more uncertain than at lower concentrations (see confidence intervals in Burnett et al., 2014).*

*However, we would like to point out that we are not trying to defend this as the "correct" approach, we instead stress that we explain different assumptions and how this impacts the results. We have tried to clarify this in the text.*

**More specific comments: - The title focuses on satellite-based estimates, but model estimates are also used here, and uncertainty in concentration-response factors are also a major focus of the analysis.**

*As stressed in the paper, satellite-based estimates are in many ways also model estimates, they are just constrained model estimates. Additionally, estimating premature mortality due to exposure (as in the title) requires use of a CRF so this is implicit. Furthermore our objective is to provide context for the interpretation of satellite-derived $PM_{2.5}$ health estimates, and we feel that our title accurately reflects this.*

**- Given the interest in models, satellites, and ground observations in the paper, I am surprised the authors didn't mention "data fusion" types of approaches such as Brauer et al (EST, 2012), who did data fusion to underlie the Lim et al. global burden estimates. Do data fusion studies reduce these uncertainties? The question may be beyond the scope of the paper, but I thought it deserved at least a little qualitative discussion.**

*Thank you for this suggestion. Data fusion methods can indeed reduce some of the uncertainty. We have added a reference to Brauer et al., (2012; 2015) and van Donkelaar et al. (2015b) in the conclusion.*

**p. 25331, l. 8-19 – The discussion here seems to mix up estimates of concentration to drive epidemiological studies vs. concentrations to drive risk assessments. This also seems to be confused a few times later in the paper. I would think that using concentrations to drive**

**risk assessments would be the main purpose here. The last sentence of this paragraph I don't think is true – many epidemiological studies consider health effects for whole populations using monitors as imperfect estimates of concentration (and not as estimates of personal exposure).**

*Our goal here was not to confuse the two, but to discuss them together as both rely on accurate PM2.5 observations. Risk assessments use CRFs from epidemiological studies, so the difference in how concentration is determined in the original epidemiological study versus how the concentration is determined in a risk assessment using that CRF could be important, and the difference in the population included in the study (gender, age, socioeconomic status, etc.) versus the whole population is likely very important. While monitors can be used to estimate population-level exposure, health effects are still an individual response (specific individuals with specific characteristics died or had an asthma attack) and the available health data may not be representative of the whole population (if there is some sort of bias in the percent of susceptible people or confounding variables not accounted for) but the response is still aggregated to be applicable to the whole population when a relative risk is determined. We have clarified this in the text.*

**p. 25332, l. 8 – "both" is ambiguous here since you've just talked about monitoring, satellites, and models.**

*We removed "both of"*

**p. 25334 top – it's not clear to me whether one value is used for the whole US or if different values are used in different states. If the first, then why is it important to start with state level data and use gridded population?**

*Different values are used for different states. We have clarified this in the text.*

**p. 25336 – The authors are correct that different studies use the terms linear and log-linear in different ways. But the discussion here doesn't quite clarify how the authors are using these terms.**

*These are both log-linear and are now referred to by the equation number later in the text.*

**p. 25338 top – what is the source for emissions for the rest of the world?**

*EDGAR and GEIA for anthropogenic emissions, but that is overwritten by regional inventories, such as BRAVO for Mexico, CAC for Canada, EMEP for Europe. We have added these references to the text.*

**p. 25340 bottom shows model performance compared to IMPROVE and AQS. In contrast, p. 25341 top discusses AERONET AOD, but presents no model performance evaluation.**

*The goal of satellite-based PM2.5 estimates is to improve surface PM2.5 estimates, not AOD (although inherently the method assumes that the two scale the same). As stated, we use AERONET to discuss the uncertainty in satellite AOD and have added a figure to show this.*

**p. 25341, l. 23 – what does "initial fraction" mean?**

*This should be "attributable fraction" and has been changed in the text.*

**p. 25346, l. 12-15 – This discusses how compensating errors may be hidden by NMB.**

**For that reason, it is common to also present Normalized Mean Error.**

*Yes, compensating errors can be hidden by the NMB and we have also investigated the NME. However, as discussed in the text, we use NMB here as appropriate for errors in annual mean values (since these are used for chronic exposure).*

**p. 25352 bottom – this long discussion of low thresholds might seem more appropriate to present in methods (there is some discussion there) or in a discussion sector.**

*We have removed this paragraph from this section and moved some of it to the methods section.*

**p. 25354 – Is this the first time Figure 9 is referenced? I find it a little strange to present a new figure in the conclusions section.**

*We moved this figure and associated text to section 4 following the suggestion of another reviewer.*

**p.25355, l. 14 – "correctly applying response functions is a determining factor" Are the authors claiming that some previous studies have done this incorrectly? I would think that there may be better or worse choices to make, but that authors may have reasons for choosing the approach that they do. I also wouldn't call these "epidemiological approaches" because these are risk assessment studies, not epidemiology.**

*Thank you for pointing out this potential misinterpretation of our text! We have removed the words "correctly" and "epidemiological."*

**p. 25355, l. 17 – "using only populated places" I don't understand what this means.**

**There should be no health effects in unpopulated places.**

*The populated places data set gives values for a point location rather than a grid and therefore has values for all major cities and town, but only some of the smaller towns in sparsely inhabited regions. We clarify this in the text.*

**Table 1 – I'm surprised that uncertainty is shown for only one study – certainly at least some of the other studies also estimated uncertainty. Also, Punger & West 2013 estimated US PM2.5 burden. Zheng et al. isn't listed in the references.**

*We didn't include the confidence intervals, just the different estimates from different CRFs. We have added Punger and West, 2013 and a reference to Zheng et al. 2014.*

**Figure 2 is pretty small and difficult to read. Is it true that all of these studies are chronic PM2.5?**

*We believe that this is largely an issue related to ACPD formatting, however we have made the font bigger and will ensure legibility in the final version. Yes, these are all for chronic exposure as stated in the Figure caption.*

**Figure 8 & 9 – units should be "deaths per year"**

*Fixed.*

**List of significant changes:**

- We have changed our base CRF from the Krewski et al. (2009) linear function to the Burnett et al. (2014) IER in order to be more in line with the literature published in 2015.

- As per the request of the reviewers, we have added in a new figure showing the location of the AERONET sites and the calculated NMB.

- Also following the request of the reviewers, we have moved some of our discussion and final figure into the new section 4.4.

# Exploring the Uncertainty Associated with Satellite-Based Estimates of Premature Mortality due to Exposure to Fine Particulate Matter

**Bonne Ford[1,*] and Colette L. Heald[2]**

[1]{Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA}

[2]{Department of Civil and Environmental Engineering and Department of Earth, Atmospheric and Planetary Sciences, MIT, Cambridge, MA, USA}

Correspondence to: B. Ford (bonne@atmos.colostate.edu)

## Abstract

The negative impacts of fine particulate matter ($PM_{2.5}$) exposure on human health are a primary motivator for air quality research. However, estimates of the air pollution health burden vary considerably and strongly depend on the datasets and methodology. Satellite observations of aerosol optical depth (AOD) have been widely used to overcome limited coverage from surface monitoring and to assess the global population exposure to $PM_{2.5}$ and the associated premature mortality. Here we quantify the uncertainty in determining the burden of disease using this approach, discuss different methods and datasets, and explain sources of discrepancies among values in the literature. For this purpose we primarily use the MODIS satellite observations in concert with the GEOS-Chem chemical transport model. We contrast results in the United States and China for the years 2004-2011. Using the Burnett et al. (2014) integrated exposure response function, ~~a simple linear concentration response function with no threshold, W~~we estimate that in the United States, exposure to $PM_{2.5}$ accounts for approximately 24% of total deaths compared to 1422% in China (using satellite-based exposure), which falls within the range of previous estimates. The difference in estimated mortality burden based solely on a global model vs. that derived from satellite is approximately ~~9~~1420% for the U.S. and ~~4~~2% for China on a nationwide basis, although regionally the differences can be much greater. This difference is overshadowed by

the uncertainty in the methodology for deriving $PM_{2.5}$ burden from satellite observations, which we quantify to be on the order of 20% due to uncertainties in the AOD-to-surface-$PM_{2.5}$ relationship, 10% due to the satellite observational uncertainty, and 30% or greater uncertainty associated with the application of concentration response functions to estimated exposure.

## 1 Introduction

By 2030, air pollution will be the leading environmentally-related cause of premature mortality worldwide (OECD, 2012). The World Health Organization (WHO) estimates that exposure to outdoor air pollution resulted in 3.7 million premature deaths in 2012. Many epidemiological studies have shown that chronic exposure to fine particulate matter ($PM_{2.5}$) is associated with an increase in the risk of mortality from respiratory diseases, lung cancer, and cardiovascular disease, with the underlying assumption that a causal relationship exists between PM and health outcomes (Dockery et al., 1993; Jerrett et al., 2005; Krewski et al., ~~2000~~2009; Pope et al., 1995; 2002; 2004; 2006). This has been shown through single and multi-population time series analyses, long-term cohort studies, and meta-analyses.

In order to stress the negative impacts of air pollution on human health and inform policy development (particularly with regard to developing strategies for intervention and risk reduction), many studies have calculated the total number of premature deaths each year attributable to air pollution exposure or the "burden of disease," through health impact assessment methods. One of the main obstacles in attributing specific health impacts of $PM_{2.5}$ is determining ~~personal~~ exposure and linking this to specific health outcomes. Jerrett et al. (2005) suggest personal monitors would be the optimal method because it would be easier to attribute individual recorded health outcomes to specific particulate levels, but point out that the financial costs and time-intensiveness limit widespread use. Many studies have instead relied on fixed-site monitors within a certain radius to estimate population-level exposure. However, these monitoring networks are generally located in urban regions and provide no information on concentration gradients between sites. Thus, epidemiological studies typically have to quantify the aggregate population response to an area-average concentration. Additionally, health data can be limited and therefore the responses may be determined from a subset of individuals that may not be representative of the wider population.

Estimating the burden of disease associated with particulate air pollution requires robust estimates of $PM_{2.5}$ exposure. Fixed-site monitoring networks can be costly to operate and maintain, and the sampling time period for many of these monitors in the United States is often only every third or sixth day. Due to the high spatial and temporal variability in aerosol concentrations, this makes it difficult to determine exposure and widespread health impacts. Worldwide, monitoring networks are even scarcer, with many developing countries lacking any long-term measurements. "Satellite-based" concentrations are now used extensively for estimating mortality burdens and health impacts (e.g. Crouse et al., 2012; Evans et al., 2013; Fu et al., 2015; Hystad et al., 2012; Villeneuve et al., 2015) ~~(e.g. Crouse et al., 2012; Hystad et al., 2012; Evans et al., 2013).~~ Satellite observations of aerosol optical depth (AOD) can offer ~~much needed~~ observational constraints for population-level exposure estimates in regions where surface air quality monitoring is limited; however they represent the vertically-integrated extinction of radiation due to aerosols, and thus additional information on the vertical distribution and the optical properties of particulate matter is required (often provided by a model) to translate these observations to surface air quality (van Donkelaar et al., 2006, 2010; Liu et al., 2004, 2005). Alternatively, studies have relied on model-based estimates of $PM_{2.5}$ exposure. Table 1 shows that the resulting estimates of premature mortality vary widely. Here, we discuss ~~both of~~ these different methods and contrast the uncertainty in these approaches for estimating exposure for both the U.S., where air quality has improved due to regulations and control technology, and China, where air quality is a contemporary national concern. Our objective is to investigate the factors responsible for uncertainty in chronic $PM_{2.5}$ burden of disease estimates, and use these uncertainties to contextualize the comparison of satellite-based and model-based estimates of premature mortality with previous work. As health impact assessment methods are becoming more popular in the scientific literature, a greater understanding of the uncertainties in these methods and the datasets that are used is important.

## 2 Methods and Tools

### 2.1 General formulation to calculate the burden of disease

To estimate the burden of premature mortality due to a specific factor like $PM_{2.5}$ exposure, we rely on equations 1 and 2 (equations 6 and 8 in Ostro, 2004 and as previously used in van Donkelaar et al., 2011; Evans et al., 2013; Marlier et al., 2013; Zheng et al., 2015). The attributable fraction (AF) of mortality due to $PM_{2.5}$ exposure depends on the relative risk value (RR), which here is the ratio of the probability of mortality (all-cause or from a specific disease) occurring in an exposed population to the probability of mortality occurring in a non-exposed population. The total burden due to $PM_{2.5}$ exposure ($\Delta M$) can be estimated by convolving the AF with the baseline mortality (equal to the baseline mortality rate $M_b$ x exposed population P). The relative risk is assumed to change ($\Delta RR$) with concentration, so that, in general, exposure to higher concentrations of $PM_{2.5}$ should pose a greater risk for premature mortality (section 2.4).

$$AF = (RR-1) / (RR) \quad \text{(or the alternate form of } AF = \Delta RR / (\Delta RR+1)) \tag{1}$$

$$\Delta M = M_b \, x \, P \, x \, AF \tag{2}$$

Application of this approach requires information on the baseline mortality rates and population, along with the RR, which is determined through a concentration response function (including a shape and initial relative risk, section 2.4), and ambient surface $PM_{2.5}$ concentrations.

### 2.2 Baseline mortality and population

For population data, we use the Gridded Population of the World, Version 3 (GPWv3), created by the Center for International Earth Science Information Network (CIESIN) and available from the Socioeconomic Data and Applications Center (SEDAC). This gridded dataset has a native resolution of 2.5 arc-minutes (~5km at the equator) and provides population estimates for 1990, 1995, and 2000, and projections (made in 2004) for 2005, 2010, and 2015. We linearly interpolate between available years to get population estimates for years not provided. Population density for China and the United States for the year 2000 are shown in Figure 1 along with the projected change in population density by the year 2015, illustrating continued growth of urbanized areas (at the expense of rural regions in China). We also compare mortality estimates

using only urban area population (similar to Lelieveld et al., 2013 which estimates premature mortality in mega-cities). For this, we rely on the populated places dataset (provided by Natural Earth, which gives values for a point location rather than a grid and includes all major cities and towns along with some smaller towns in sparsely inhabited regions) which is determined from LandScan population estimates (Bright et al., 2008). In the U.S., approximately 80% of the population lives in urban areas. For China, 36% of the population lived in urban areas in 2000, but this number rose to 53% in 2013 (World Bank, 2015).

To determine baseline mortalities in the U.S. for cardiovascular disease (ischemic heart disease and stroke), lung cancer, and respiratory disease, we use mortality~~death~~ rates for each cause of death for all ages from the Center for Disease Control (cdc.gov) for each year and each state. ~~which w~~We ~~then~~ multiply ~~by~~ the gridded population by these state-level mortality rates to obtain the ~~total~~ baseline mortality in each grid box. Other studies have also used country-wide (or regional) (e.g. Evans et al., 2013) or county-level (e.g. Fann et al., 2013) average deaths rates. Some studies use the mortality rate for all cardiovascular diseases, which would produce larger estimates than just using ischemic heart disease and cerebrovascular disease (stroke). Additionally, some studies also only consider respiratory deaths related to ozone exposure.

Mortality values are not as readily available for China, so we rely on country-wide values for baseline mortality (WHO age-standardized mortality rates by cause). Therefore, in China spatial variations in $M_b$ are only due to variations in population and not regional variations in actual death rates (i.e. we do not account for death-specific mortality rates varying between provinces). In order to account for some regional variability in mortality rates, we use a population threshold to distinguish between urban and rural regions for lung cancer mortality rates (Chen et al., 2013a)~~(Chen et al., 2013)~~.

## 2.3 Relative Risk

The relative risk (RR) is a ratio of the probability of a health endpoint (in this case premature mortality) occurring in a population exposed to a certain level of pollution to the probability of that endpoint occurring in a population that is not exposed. Values greater than one suggest an increased risk, while a value of one would suggest no change in risk. These values are determined through epidemiological studies which relate individual health impacts to changes in

concentrations, and literature values span a large range (Figure 2). While these studies attempt to account for differences in populations, lifestyles, pre-existing conditions, and co-varying pollutants, relative risk ratios determined from each study still differ. This is likely due to variables not taken into consideration, errors in exposure estimates ("exposure misclassification") (Sheppard et al., 2012), and because, although the long-term effects of exposure to atmospheric pollutants have been well-documented, the pathophysiological mechanisms linking exposure to mortality risk are still unclear (Chen and Goldberg, 2009; Pope and Dockery, 2013; Sun et al., 2010), which makeing it difficult to determine how transferable results are from the context in which they were generated.

For our initial estimates, Wwe use the integrated risk function from Burnett et al. (2014) risk ratios for heart disease, respiratory disease, cardiovascular and lung cancer premature mortality due to chronic exposure. We also compare our results to premature mortality estimates using risk ratios determined by Krewski et al. (2009), which is an extended analysis of the American Cancer Society study (Pope et al., 1995), and for respiratory disease, from Laden et al. (2006) which is an updated and extended analysis of the Harvard Six Cities study (Pope et al., 2002). The We use the updated Krewski et al. (2009) risk ratios have beenas they are widely used in similar studies due to the large study population with national coverage, 18 year time span, and extensive analysis of confounding variables (ecological covariates, gaseous pollutants, weather, medical history, age, smoking, etc.). However, the Burnett et al. (2014) function is becoming more widely used in the literature (e.g. Lelieveld et al., 2015; Lim et al., 2012, Apte et al., 2015) because it provides the shape of the mortality function for the global range of exposure concentrations. Using these same different risk ratios also can makes our results more directly comparable to studies in Table 1 which rely on the risk ratios from these fourthree studies (Burnett et al., 2014; Krewski et al., 2009; Laden et al., 2006; Pope et al., 2002).

## 2.4 Concentration response function

In order to determine an attributable fraction, it is necessary to understand how the response changes with concentration (i.e. does the relative risk increase, decrease, or level off with higher concentrations?). The shape of this concentration response function is an area of on-going epidemiological research (e.g. Burnett et al., 2014; Pope et al., 2015).

~~For our initial results, we rely on equation 4, where~~ In the simplest form, it might be assumed that the change in relative risk (RR, given as per 10 µg/m$^3$) linearly depends on the surface $PM_{2.5}$ concentration (C, in µg/m$^3$) as given in equation 4 (and as presented as an alternate form in Cohen et al., 2004; 2005).

$$\Delta RR = (RR-1) \times (C-C_0)/10 \quad (4)$$

In this equation, $C_0$ can be considered the "policy relevant (PRB)/target", "natural background" or "threshold"/ "counterfactual"/ "lowest effect level" surface $PM_{2.5}$ concentration. ~~As s~~Studies have shown that there is not a concentration level below which there is no adverse health effect for PM (e.g. Pope et al., 2002; Shi et al., 2015) and most experts in health impacts of ambient air quality agree that there is no population-level threshold (although there may be individual-level thresholds, e.g. Roman et al., 2008).~~,~~ However, there are few epidemiological studies in regions with very low annual average concentrations (Crouse et al., 2012 does records a 1.9 µg m$^{-3}$ annual concentration in rural Canada) making it difficult to determine the health risks in relatively clean conditions. How to extrapolate the relationship out of the range of observed measurements is uncertain. ~~iIn our initial analysis, we assume this value is zero. However,~~ Therefore, rather than assuming that the function is linear down to zero, ~~other~~ studies often set $C_0$ to the value of the lowest measured level (LML) observed in the epidemiology study from which the RRs are derived [e.g. *Evans et al.*, 2013 use 5.8 µg/m$^3$ with the RR from *Krewski et al.* 2009] or use the "policy relevant" background (PRB, generally 0-2 µgm$^{-3}$) concentration.~~,~~ This is the level to which policies might be able to reduce concentration and ~~which~~ is generally determined from model simulations in which domestic anthropogenic emissions have been turned off (e.g. Fann et al., 2012). Similarly, some studies have set this value to preindustrial (1850) pollution levels (e.g. Fang et al., 2013; Silva et al., 2013).

~~$\Delta RR = (RR-1) \times (C-C_0)/10 \quad (4)$~~

Linear response functions are generally a good fit to observed responses at lower concentrations (Pope et al., 2002). However, ~~some~~ studies ~~have~~ suggest~~ed~~ that linear response functions can greatly overestimate RR at high concentrations (e.g. Pope et al., 2015), where responses may start to level off. There is ~~is~~ uncertainty at high concentrations because~~, as~~ most epidemiology studies of the health effects of air pollution exposure have generally been conducted under lower

concentrations (i.e. in the U.S.). In order to determine the shape of this response at higher concentrations, smoking has been used as a proxy (Burnett et al., 2014; Pope et al., 2011, 2009)~~(Pope et al., 2011, 2009)~~, which does show a diminishing response at higher concentrations. Therefore, both log-linear (Equations 5 and 6, where $\beta = 0.15515/0.23218$ for heart disease/lung cancer from Pope et al. (2002) or $\beta = 0.18878/0.21136$ for heart disease/lung cancer from Krewski et al., 2009 in Equation 6 and $\beta = 0.01205/0.01328$ for heart disease/lung cancer from Krewski et al., 2009 in Equation 5) and power law (Equation 7, where I is the inhalation rate of $18 m^3 day^{-1}$, $\beta = 0.2730/0.3195$, $\alpha = 0.2685/0.7433$ for heart disease/lung cancer from Pope et al., 2011 and as used in Marlier et al., 2013) functions have been also been explored in this study.

$$\Delta RR = exp[\beta (C - C_0)] - 1 \quad (5)$$

$$\Delta RR = [(C+1)/(C_0+1)]^{\beta} - 1 \quad (6)$$

$$\Delta RR = \alpha (I \times C)^{\beta} \quad (7)$$

We note that Cohen et al. (2005) and Anenberg et al. (2010) reference Equation 5 as a log-linear function ~~(and take the log of $C-C_0$, also referred to as a log-log model)~~, while Ostro (2004), ~~and~~ Evans et al. (2013), and Giannadaki et al. (2014) use this as their linear function and instead use Equation 6 as their log-linear function, we will refer to these equation numbers for clarity in other sections. Another method to limit the response at high concentrations is to simply use a "ceiling," "maximum exposure/high-concentration threshold," or "upper truncation" value in which it is assumed that the response remains the same for any value above it (e.g. Anenberg et al., 2012; Cohen et al., 2005; Evans et al., 2013). This can be a somewhat arbitrary value or the highest observed concentration in the original epidemiological study.

~~$\Delta RR = exp[\beta (C - C_0)] - 1$ (5)~~

~~$\Delta RR = [(C+1)/(C_0+1)]^{\beta} - 1$ (6)~~

~~$\Delta RR = \alpha (I \times C)^{\beta}$ (7)~~

Recently, Burnett et al. (2014) fit an integrated exposure response (IER) model using RRs from a variety of epidemiological studies on ambient and household air pollution, active smoking, and second hand tobacco smoke in order to determine RR functions over all global $PM_{2.5}$ exposure ranges for ischemic heart disease, cerebrovascular disease, chronic obstructive pulmonary disease, and lung cancer (Equation 8). Monte Carlo simulations were conducted in order to derive

the one thousand sets of coefficients for the IER function (the coefficients are available at http://ghdx.healthdata.org/record/global-burden-disease-study-2010-gbd-2010-ambient-air-pollution-risk-model-1990-2010). ~~The fitted values for each cause of death are given in Table 2~~.

$$for\ C < C_0,\ \Delta RR = 0$$

$$for\ C \geq C_0,\ \Delta RR = \alpha\{1-exp[-\gamma(C-C_0)^{\rho}]\} \quad (8)$$

This form is now being widely used (Apte et al., 2015; Lelieveld et al., 2015; Lim et al., 2012) and we use it here for our baseline estimates. In the following sections, we will discuss the uncertainty on the burden of disease associated with the shape of the concentration response function and threshold concentration.

## 2.5 Estimating surface $PM_{2.5}$

We use both a global model and satellite observations to estimate surface $PM_{2.5}$ concentrations and translate these to $PM_{2.5}$ exposure and health burden. In addition, we use surface measurements of $PM_{2.5}$ to test the accuracy of these estimates.

### 2.5.1 Unconstrained Model Simulation

We use the global chemical transport model GEOS-Chem (geos-chem.org) to simulate both surface $PM_{2.5}$ and AOD. We use v9.01.03 of the model, driven by GEOS-5 meteorology, in the nested grid configuration over North America and Asia (0.5°x0.667° horizontal resolution) for 2004-2011. Using this longer time period gives greater confidence in our uncertainty results. The GEOS-Chem aerosol simulation includes sulfate, nitrate, ammonium (Park et al., 2004), primary carbonaceous aerosols (Park et al., 2003), dust (Fairlie et al., 2007; Ridley et al., 2012), sea salt (Alexander et al., 2005), and secondary organic aerosols (SOA) (Henze et al., 2008). There are several regional anthropogenic emission inventories used in the model, such as BRAVO over Mexico (Kuhns et al., 2003), EMEP over Europe (Vestreng et al., 2007), CAC for Canada (http://www.ec.gc.ca/pdb/cac/cac_home_e.cfm), the EPA NEI05 inventory (Hudman et al., 2007, 2008) over the U.S., and Streets et al. (2006) over Asia. Any location not covered by one of these regional inventories relies on the GEIA (Benkovitz et al., 1996) and EDGAR (Olivier and Berdowski, 2001; Olivier et al., 2001] inventories. ~~Most anthropogenic and biofuel~~ Biofuel emissions over the U.S. ~~A~~ are also from the EPA NEI05 inventory (Hudman et al., 2007, 2008)

and ~~from over Asia; although~~ anthropogenic emissions of black and organic carbon over North America follow Cooke et al. (1999) with the seasonality from Park et al. (2003). Biogenic VOC emissions are calculated interactively following MEGAN (Guenther et al., 2006), while year-specific biomass burning is specified according to the GFED2 inventory (van der Werf et al., 2006). Surface dry $PM_{2.5}$ is calculated by combining sulfate, nitrate, ammonium, elemental carbon, organic matter, fine dust, and accumulation mode sea salt concentrations in the lowest model grid box. In the following discussion, these values are referred as the "unconstrained model." Simulated AOD is calculated at 550 nm based on aerosol optical and size properties as described in Ford and Heald (2013).

### 2.5.2 Satellite-based

We also derive a satellite-based surface $PM_{2.5}$ using satellite observed aerosol optical depth, with additional constraints from the GEOS-Chem model, in a similar manner to Liu et al. (2004, 2007) and van Donkelaar et al. (2006, 2010, 2011). This method relies on the following relationship:

$$PM_{2.5,surface} = \eta \; x \; AOD_{satellite} \qquad (9)$$

Where the satellite-derived $PM_{2.5}$ is estimated at the resolution of the unconstrained model by multiplying the satellite observed AOD by the value η, which is the ratio of model simulated surface $PM_{2.5}$ to simulated AOD at the time of the satellite overpass. This is then a combined product which relies on a chemical transport model to simulate the spatially and temporally varying relationship between AOD and surface $PM_{2.5}$ by accounting for all the aerosol properties and varying physical distribution and then constraining these results by "real" (i.e. satellite) measurements of AOD. Using the satellite to constrain the model concentrations is extremely useful in regions where emissions inventories and model processes are less well known.

For AOD, we use observations from both of the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments and from the Multi-angle Imaging SpectroRadiometer (MISR) instrument. For this work we use MODIS 550 nm Level 2, Collection 6, Atmosphere Products for Aqua as well as Level 2, Collection 5 for ~~Terra and~~ Aqua. We filter these data for cloud fraction (CF < 0.2) and remove observations with high AOD (>2.0), as cloud contamination causes known biases in the AOD (Zhang et al., 2005) as in Ford and Heald (2012) although we note that this

could remove high pollution observations, particularly in China. For MISR, we also use the Level 2 AOD product (F12, version 22, 500nm). We note that this is a different wavelength than from the MODIS instrument, but we neglect that difference for these comparisons. We use both of these observations for comparison as MODIS has a greater number of observations while MISR is generally considered to better represent the spatial and temporal variability of AOD over China (Cheng et al., 2012; Qi et al., 2013; You et al., 2015). Satellite observations are gridded to the GEOS-Chem nested grid resolution. We sample GEOS-Chem to days and grid boxes with valid satellite observations to calculate the η used to translate the AOD to surface $PM_{2.5}$.

In Figure 3, we show the long-term average (2004-2011) of satellite-based $PM_{2.5}$ for the U.S. and China using MODIS Aqua Collection 6 and compare this to model-only estimates. In the following sections, most of our results will be shown using Collection 6; but reference and comparisons will be made to other products as a measure of uncertainty. In general the unconstrained model and satellite-based estimates show similar spatial features and magnitudes, with stronger local features apparent in the satellite-based $PM_{2.5}$. The satellite-based estimate suggests that concentrations should be higher over much of the western U.S., particularly over California, Nevada, and Arizona (comparisons with surface measurements are discussed in section 2.5.3). In China, the satellite-derived $PM_{2.5}$ is higher in Eastern China, around Beijing and the Heibei province, Tianjin, and Shanghai, but lower in many of the central provinces. While many previous studies suggest that MODIS may be biased high (and MISR biased low) over China (e.g. Cheng et al., 2012; Qi et al., 2013; You et al., 2015) and the Indo-Gangetic Plain (Bibi et al., 2015); Wang et al. (2013) note that the GEOS-Chem model underestimates $PM_{2.5}$ in the Sichuan basin, suggesting that the MODIS satellite-based estimate could reduce the bias in this province.

### 2.5.3 Surface-based Observations

We use observations of $PM_{2.5}$ mass from two networks in the United States (where long-term values are more readily available than in China) to evaluate the model and satellite-derived $PM_{2.5}$: the Interagency Monitoring of Protected Visual Environments (IMPROVE) and the EPA Air Quality System (AQS) database. The IMPROVE network measures $PM_{2.5}$ over a 24-hour period every third day and these measurements are then analyzed for concentrations of fine, total,

and speciated particle mass (Malm et al., 1994). We use the reconstructed fine mass (RCFM) values, which are the sum of ammonium sulfate, ammonium nitrate, soil, sea salt, elemental carbon and organic matter.

Previous studies have generally shown good agreement between measurements and GEOS-Chem simulations of $PM_{2.5}$ (e.g. Ford and Heald, 2013; van Donkelaar et al., 2010). In Figure 4, we show the long-term average of $PM_{2.5}$ at AQS and IMPROVE sites in the U.S. overlaid on simulated concentrations. In general, GEOS-Chem agrees better with measurements at IMPROVE sites, likely because these are located in rural regions where simulated values will not be as impacted by the challenge of resolving urban plumes in a coarse Eulerian model. There are noted discrepancies in California (Schiferl et al., 2014) and the Appalachia/Ohio River Valley region where the model is biased low. The model has a low mean bias of -25% compared to measurements at the EPA AQS sites and a bias of -6% compared to measurements at IMPROVE sites. Annual mean bias at individual sites ranges from -100 to 150%. At these same AQS sites, the satellite-derived $PM_{2.5}$ is less biased (-12% using MODIS C6 or -8% using MISR).

To estimate the uncertainty in satellite AOD, we also rely on surface-based measurements of AOD from the global AErosol RObotic NETwork (AERONET) of sun photometers. AOD and aerosol properties are recorded at eight wavelengths in the visible and near-infrared (0.34-1.64μm) and are often used to validate satellite measurements (e.g. Remer et al., 2005). AERONET AOD has an uncertainty of 0.01-0.015 (Holben et al., 1998). For this work, we use hourly Version 2 Level 2 measurements sampled to two hour windows around the times of the satellite overpasses. We also perform a least-square polynomial fit to interpolate measurements to 550 nm.

## 3 Estimated health burden associated with exposure to $PM_{2.5}$

We compare national exposure estimates for the U.S. and China using unconstrained and satellite-based (MODIS and MISR) annual average $PM_{2.5}$ concentrations in (Figure 5), which is a cumulative distribution plot that is calculated as the sum of the population in each grid box which has an annual average concentration at or above each concentration level. For the U.S., satellite-based estimates suggest a slightly greater fraction of the population is exposed to higher annual

average concentrations, while in China, the satellite-based estimates suggest a lower fraction. Using MISR AOD suggests higher annual average concentrations in the U.S. and much lower in China, as MISR has a high bias in regions of low AOD and a low bias in regions of high AOD (Jiang et al., 2007; Kahn et al., 2010)~~(Kahn et al., 2010)~~. The large discrepancy between results from MISR and MODIS could be due to differencing in sampling, but studies have also shown that MODIS is biased high in China and MISR is biased low (Cheng et al., 2012; Qi et al., 2013; You et al., 2015). We discuss the uncertainties in these estimates in section 4.

These exposure estimates are used to calculate an attributable~~initial~~ fraction of mortality associated with heart disease, lung cancer, and respiratory disease attributable to chronic exposure using both model and satellite-based annual average concentrations for the U.S. and China (Table 1). In the U.S., we estimate that exposure to $PM_{2.5}$ accounts for approximately ~~4~~2% of total deaths (~~13~~6% of heart diseases and ~~12~~5% of respiratory diseases) compared to ~~22~~14% (33~~40~~% of heart and ~~25~~22% of respiratory) in China using satellite-based concentrations. The Global Burden of Disease estimates for 2010, that 10% of total deaths in China and 3% of total deaths in the U.S. are attributable to exposure to $PM_{2.5}$ (Lim et al., 2012). We present these as an average over the 2004-2011 time period in order to provide more robust results that are not driven by an outlier year, as there is considerable year-to-year variability in AOD and surface $PM_{2.5}$ concentrations (for example, heavy dust years in China). However, there are trends in population (Figure 1) and surface concentrations that can influence these results. For example, there is a significant decreasing trend in AOD over the northeastern U.S. simulated in the model which is also noticeable in the satellite observations and the surface concentrations (Hand et al., 2012). This decreasing trend can be attributed to declining $SO_2$ emissions in the U.S. as noted in Leibensperger et al. (2012). Trends in China are more difficult to ascertain as emissions have been variable over this period in general [*Lu et al.*, 2011; *Zhao et al.*, 2013] with widespread increases from 2004 to 2008 followed by variable trends in different regions through 2011. The difference between mortality burden estimates using model or satellite concentrations is approximately 20~~9~~% for the U.S. and ~~4~~2% for China on a nationwide basis, although regionally the difference can be much greater. A question we aim to address here is whether these model and satellite-based estimates are significantly different.

We compare our results to premature mortality burden estimates from other studies in Table 1. In general, our estimates for ~~the U.S. and~~ China are higher than most previous ~~global~~ estimates, except for Lelieveld et al. (2015) and Rohde and Muller (2015). ~~However,~~ these studies provide estimates for 2010 and 2014, respectively, and we did find an increasing trend in our mortality estimates over the study time period. For the U.S., our estimates are in the lower range of previous studies. ~~Anenberg et al. (2010) (although their estimates are for all of North America and all of Asia) and Fann et al. (2012). In particular, our estimates appear to be twice those found by Lelieveld et al. (2013).~~ The spread among these studies can be attributed to the data used (i.e. MODIS Collection 5 rather than Collection 6 or unconstrained model concentrations, ~~and~~ choice of baseline mortality rates, and population), the resolution of the data, the years studied, as well as the risk ratios and response functions and the resoultion of these products. For example, Evans et al. (2013) also use satellite-based concentrations (using MISR/MODIS Collection 5 and GEOS-Chem), ~~but their resulting mortality estimates are much lower~~ but use a different concentration response function. ~~However, unlike this study, they use a lower threshold value and a different response function with lower baseline mortalities~~. In the following sections, we delineate the uncertainty in ~~our~~ these results and reasons for differences from previous studies.

## 4 Uncertainty in Satellite-based $PM_{2.5}$

Uncertainties in the $PM_{2.5}$ concentrations derived from satellite observations arise from the two pieces of information which inform this estimate: (1) satellite AOD and (2) model η. Here we explore some of the limitations and uncertainties associated with each of these inputs.

### 4.1 Uncertainty Associated with Satellite AOD

While satellite observations of aerosols are often used for model validation (e.g. Ford and Heald, 2012), these are indirect measurements with their own limitations and errors. The uncertainty in satellite AOD can be due to a variety of issues such as the presence of clouds, the choice of optical model used in the retrieval algorithm, and surface properties (Toth et al., 2014; Zhang and Reid, 2006). For validation of satellite products, studies have often relied on comparisons against AOD measured with sun photometers at AERONET ground sites (e.g. Kahn et al., 2005; Levy et al., 2010; Remer et al., 2005, 2008; Zhang and Reid, 2006). The uncertainty in AOD over land

from MODIS is estimated as 0.05 ± 15% (Remer et al., 2005), while Kahn et al. (2005) suggest that 70% of MISR AOD data are within 0.05 (or 20 % ×AOD) of AERONET AOD.

There are also discrepancies between AOD measured by the different instruments due to different observational scenarios and instrument design. The Aqua platform has an afternoon overpass while the Terra platform has a morning overpass. It might be expected that there would be some differences in retrieved AOD associated with diurnal variations in aerosol loading. However, the difference of 0.015 in the globally averaged AOD between MODIS onboard Terra and Aqua (Collection 5), although within the uncertainty range of the retrieval, is primarily attributed to uncertainties and a drift in the calibration of the Terra instrument, noted in Zhang and Reid (2010) and Levy et al. (2010). Collection 6 (as will be discussed further) reduces the AOD divergence between the two instruments (Levy et al., 2013). MISR employs a different multi-angle measurement technique with a smaller swath width; as a result the correlation between MISR AOD and MODIS AOD is only 0.7 over land (0.9 over ocean) (Kahn et al., 2005).

Not only are there discrepancies in AOD between instruments, there are also differences between product versions for the same instrument. The MODIS Collection 6 Level 2 AOD is substantially different from Collection 5.1 (Levy et al., 2013) and Figure 6. In general, AOD decreases over land and increases over ocean with Collection 6. These changes are due to a variety of algorithm updates including better detection of thin cirrus clouds, a wind speed correction, a cloud mask that now allows heavy smoke retrievals, better assignments of aerosol types, and updates to the Rayleigh optical depths and gas absorption corrections (Levy et al., 2013). These differences can also impact the derived $PM_{2.5}$ (and can explain some differences between our results and previous studies). In particular, because Collection 6 suggests higher AOD over many of the urbanized regions, the derived $PM_{2.5}$ and resulting exposure estimates (all other variables constant) are greater. The difference between these two retrieval products, given the same set of radiance measurements from the same platform, gives a sense of the uncertainty in the satellite AOD product (Figure 65a).

We estimate the uncertainty in satellite AOD used here by comparing satellite observations to AERONET and determining the normalized mean bias (NMB) between AOD from each satellite instrument and AERONET for the U.S. and China (Figure 7). Although there a very limited number of sites in China, from these comparisons, we find that the satellites generally agree with

AERONET better in the eastern U.S. and northeastern China than in the western U.S. and western and southeastern China.~~and have~~ There are larger biases in the west near deserts and at coastal regions where it may be challenging to distinguish land and water in the retrieval algorithm. NMBs at each AERONET site are generally similar among the instruments (MISR comparison not shown), with greater differences at these western sites. While Collection 6 does reduce the bias at several sites along the East Coast in the U.S., it is generally more biased at the Four Corners region of the U.S. We use these NMBs to regionally "bias correct" our AOD values and estimate the associated range of uncertainty in our premature mortality estimates. Compared to the standard MODIS AOD retrieval uncertainty, our overall NMB is less in the eastern U.S. (-1%) and western China (11%) and higher in the western U.S. (40%) and eastern China (18%).

There may also be biases associated with the satellite sampling, should concentrations on days with available observations be skewed. In order to assess the sampling bias, we use the model and compare the annual mean to the mean of days with valid observations (Figure 6~~5~~b). In general, sampling leads to an underestimation in AOD (average of 20% over the U.S.). This can partly be attributed to the presence of high aerosol concentrations below or within clouds which cannot be detected by the satellite, the mistaken identification of high aerosol loading as cloud in retrieval algorithms, as well as the removal of anomalously high AOD values (>2.0) from the observational record. This suggests that the average AOD values can also be influenced by the chosen filtering and data quality standards. Analysis of the impact of satellite data quality on the AOD to $PM_{2.5}$ relationship is discussed in Toth et al. (2014). They find that using higher quality observations does tend to improve correlations between observed AOD and surface $PM_{2.5}$ across the U.S. though in general correlations are low (<0.55).

## 4.2 Uncertainty Associated with Model η

In general, the model simulates $PM_{2.5}$ well (Figure 4) and represents important processes; but, satellite AOD can help to constrain these estimates to better represent measured concentrations (van Donkelaar et al., 2006). However, in specific regions or periods of time, errors in η could lead to discrepancies between satellite-derived and actual surface mass and~~.~~ Snider et al. (2015) does show some regional biases in the GEOS-Chem model η compared to η determined from collocated surface measurements of AOD and $PM_{2.5}$. In order to assess the potential uncertainty

in model-based η, we perform multiple sensitivity tests to determine the impact that different aerosol properties, grid-size resolution and time scales will have on η and, ultimately, on the resulting satellite-based $PM_{2.5}$ (listed in Table ~~3~~2). These sensitivity tests are performed solely with model output, which can provide a complete spatial and temporal record, and results from the modified simulations are compared to the standard model simulation. We note that these are "errors" only with respect to our baseline simulation; we do not characterize how each sensitivity simulation may be "better" or "worse" compared to true concentrations of surface $PM_{2.5}$, but rather how different they are from the baseline, thus characterizing the uncertainty in derived $PM_{2.5}$ resulting from the model estimates of η. We make these comparisons for both the U.S. and China and show results in Figure 8~~7~~. Because mass concentrations in China are generally much higher, the absolute value of potential errors can also be much greater.

The timescale of the estimated $PM_{2.5}$ influences the error metric we choose for this analysis. We use the NMB for estimating error associated with annual $PM_{2.5}$ exposure (the metric of interest for chronic exposure). This allows for the possibility that day-to-day errors may compensate, resulting in a more generally unbiased annual mean value. The error on any given day of satellite-estimated $PM_{2.5}$ is likely larger, and not characterized by the NMB used here.

Our first sensitivity tests relate specifically to the methodology. To derive a satellite-based $PM_{2.5}$ with this method requires ~~that~~ model output for every day and that there are valid satellite observations. Running a model can be labor intensive, at the same time there are specific regions and time periods with poor satellite coverage. Therefore, it might be beneficial to be able to use a climatological η or a climatological satellite AOD. To test the importance of daily variability in AOD, we compute daily η values and then solve for daily surface $PM_{2.5}$ values using a seasonally averaged model simulated AOD (AvgAOD). This mimics the error introduced by using seasonally averaged satellite observations, an attractive proposition to overcome limitations in coverage. This approximation often produces the greatest error (~20% in the U.S. and 0-50% in China) especially in regions where AOD varies more dramatically and specifically where transported layers aloft can significantly increase AOD (Figure 8~~7~~). For the seasonally averaged η test (AvgEta), we estimate daily $PM_{2.5}$ values (which are averaged into the annual concentration) from the seasonally averaged η and daily AOD values. As regional η relationships can be more consistent over time than $PM_{2.5}$ or AOD, this test evaluates the necessity of using

daily model output to define the η relationship. The error in the annual average of daily $PM_{2.5}$ values determined using a seasonally averaged η creates results that are very similar to the error found calculating an annual average of daily $PM_{2.5}$ values calculated using a seasonally averaged AOD.

The model η also inherently prescribes a vertical distribution of aerosol, which may be inaccurately represented by the model and introduce errors in the satellite-derived $PM_{2.5}$. Previous studies have shown that an accurate vertical distribution is essential for using AOD to predict surface $PM_{2.5}$. (e.g. Li et al., 2015; van Donkelaar et al., 2010). We test the importance of the variability of the vertical distribution in the η relationship for predicting surface $PM_{2.5}$ concentrations by comparing values from the standard simulation against using an η from a seasonally averaged vertical distribution (AvgProf). For this comparison, we allow the column mass loading to vary day-to-day, but we assume that the profile shape does not change (i.e. we re-distribute the simulated mass to the same seasonally averaged vertical profile). We note that this is not the same as assuming a constant η, as relative humidity and aerosol composition are allowed to vary. Additionally, this differs from other studies (van Donkelaar et al., 2010; Ford and Heald, 2013) in that we are not testing the representativeness of the seasonal average profile, but testing the importance of representing the daily variability in the vertical profile. From Figure 7, we see that using a seasonally averaged vertical distribution (AvgProf) can lead to large errors in surface concentrations. Information on how the pollutants are distributed is extremely important because changes in column AOD can be driven by changes in surface mass loading, but also by layers of lofted aerosols that result from production aloft or transport (and changes in the depth of the boundary layer). This is important in areas that are occasionally impacted by transported elevated biomass burning plumes or dust. ~~In China, l~~Large errors often occur in ~~western and central~~ China, especially during the spring when these regions are influenced by transported dust from the Taklamakan and Gobi Deserts (Wang et al., 2008). Southeastern China has the largest NMB due to not only transport from interior China, but also from other countries in Southeast Asia. There is a positive bias in most regions, because on average, most of the aerosol mass is located at the surface; therefore, using an average profile will overpredict the surface concentrations. Similar to the average AOD and η (AvgAOD and AvgEta), average vertical distributions generally overpredict~~s~~ $PM_{2.5}$ due to the presence of outliers. This stresses

the importance of not only getting the mean profile correct, but the necessity of also simulating the variability in the profile on shorter timescales.

We also test the sensitivity of derived $PM_{2.5}$ to aerosol water uptake. This is done by recalculating η using a seasonally averaged relative humidity (RH) profile (AvgRH). This generally reduces the seasonally averaged AOD (less water uptake) in every season (because hygroscopic growth of aerosols is non-linear with RH). This leads to an overestimate of η that when applied to the AOD values from the standard simulation and generally overestimates surface $PM_{2.5}$ in regions with potentially higher RH and more hygroscopic aerosols (eastern U.S. and eastern China). This is because, for the same AOD, a higher η value would suggest more mass at the surface in order to compensate for optically smaller particles aloft. Western China (and some of central China) has a negative bias, suggesting that using a mean relative humidity actually underestimates $PM_{2.5}$. However, this is because the RH is generally low but can have large variability, and concentrations (outside of the desert regions) are also low so that the NMB may be large although the absolute error is not.

A higher resolution model, although more computationally expensive, will likely better represent small scale variability and is better suited for estimating surface air quality. Punger and West (2013) find that coarse resolution models often drastically underestimate exposure in urban areas. We therefore investigate the grid-size dependence of our simulated η. For this, we determine the η values from a simulation running at 2°x 2.5° grid resolution (with the same emission inputs and time period), re-grid these values to the nested grid resolution (0.5°x0.666°) and solve for the derived $PM_{2.5}$ concentrations using the AOD values from the nested simulation (noted as 2x2.5 in Figure ~~7~~8). From Figure ~~7~~8, we see larger discrepancies in regions which are dominated by more spatially variable emissions (Northeastern U.S. and China) rather than areas with broad regional sources (Southeastern U.S.). This is line with Punger and West (2013) who show smaller differences due to resolution in estimated premature mortality due to $PM_{2.5}$ exposure in rural areas than in urban areas. ~~However, c~~Compared to the other sensitivity tests, using the coarser grid leads to mean errors of only 10-15% in the U.S. and in China, which suggests that spatially averaged η are potentially more useful than temporally averaged η for constraining surface $PM_{2.5}$. ~~This is in line with the conclusions from~~ Thompson and Selin (2012) and Thompson et al. (2014)~~who~~ show that coarse grids can overpredict pollutant concentrations and consequently

health impacts ~~(associated with exposure to ozone)~~, but using very fine grids does not significantly decrease the error in simulated concentrations compared to observations. . This effect was more pronounced with ozone. Additionally, their coarsest grid resolution is 36 km which they compare to results at 2, 4, and 12 km. Punger and West (2013) compare health impacts at a variety of resolutions out to several 400 km and show that coarser resolutions underestimate health impacts because concentrations are diluted over larger areas rather than high concentrations being co-located with large urban populations.

The GEOS-Chem simulation of surface nitrate aerosol over the U.S. is biased high (Heald et al., 2012). This can be an issue in regions where nitrate has a drastically different vertical profile (or η) from other species. To test how this nitrate bias could impact η and the derived $PM_{2.5}$, we compute η without nitrate aerosol, and then derive $PM_{2.5}$ using the standard AOD (No $NO_3$). This is not a large source of potential error (<15%), with only slightly larger errors in winter and in regions where nitrate has a significant high bias (central U.S.). Furthermore, these errors are less than the bias between the model and surface observations of nitrate in the U.S. (1-2 $\mu gm^{-3}$ compared to 2-7 $\mu gm^{-3}$), suggesting that even though there is a known bias in the model, using satellite observations may largely correct for this by constraining the total AOD when estimating satellite-derived $PM_{2.5}$. We also did this comparison for China. Measured nitrate concentrations are not widely available for evaluation, but Wang et al. (2014) suggests that model nitrate is also too high in eastern China. The NMB is even less in regions in China (<10%), with negative values in eastern China (where nitrate concentrations are high) and positive values in western and central China (where nitrate concentrations are lower and have less bias compared to observations).

To further explore the role of aerosol composition (and possible mischaracterization in the model), we take the simulated mass concentrations and compute the AOD assuming that the entire aerosol mass is sulfate ($SO_4$ in Figure ~~7~~8) or, alternatively, hydrophobic black carbon (BC in Figure ~~7~~8). Black carbon has a high mass extinction efficiency, which is constant with RH given its hydrophobic nature; while sulfate is very hygroscopic, resulting in much higher extinction efficiencies at higher relative humidity values. Overall, assuming that all the mass is sulfate leads to low biases on the order of 15-20% as the AOD in many regions in the U.S. is dominated by inorganics. Errors are largest in regions and seasons with larger contributions of

less hygroscopic aerosols (organic carbon and dust) and/or high relative humidity. Assuming the entire aerosol mass is black carbon can lead to greater errors than sulfate because BC has a larger mass extinction at lower relative humidity values and hydrophobic black carbon generally makes up a small fraction of the mass loading in all regions in the U.S. and China. When RH is low, this assumption increases the AOD, which leads to an under prediction in the derived $PM_{2.5}$. When RH is high, this decreases the AOD and leads to an over prediction in derived $PM_{2.5}$. The largest percentage changes occur in the southwestern U.S. and western China (~-30%) due to the low relative humidity, low mass concentrations, and large contribution of dust.

We also compare these sensitivity tests on daily timescales. We do not show the results here because we rely on chronic exposure (annual average concentrations) for calculating mortality burdens. The normalized mean biases in annual average concentrations (Figure 8~~7~~) are generally much less (range of ±20% in U.S. and ±50% in China) than potential random errors in daily values as many of these daily errors cancel out in longer term means. This is the case for our sensitivity tests regarding the vertical profile and relative humidity, which have much larger errors on shorter timescales. However, because our method to test the sensitivity to aerosol type assumes that all aerosol mass is black carbon or sulfate, we introduce a systematic bias that is not significantly reduced in the annual NMB. This highlights the differing potential impacts due to systematic and random errors, which is an important distinction for determining the usefulness of this method. Systematic errors may not be as obvious on short timescales compared to random errors (related to meteorology and/or representation of plumes) that can lead to large biases in daily concentrations. However, these random errors have less impact when we examine annual average concentrations and mortality burdens. Systematic errors, potentially related to sources or processes, may be harder to counteract even on longer timescales and even when the model is constrained by satellite observations. However, we also show that random daily errors can bias the long term mean, stressing the importance of not only correcting regional biases, but also in accurately simulating daily variability.

We translate this potential uncertainty in $\eta$ to potential uncertainty in mortality estimates determined from the satellite-based $PM_{2.5}$. We use the normalized mean bias in annual $PM_{2.5}$ determined from the sensitivity tests for RH, the vertical profile, grid resolution, and aerosol composition for each grid box and then use these values to "bias correct" our satellite-based

annual $PM_{2.5}$ concentrations and re-calculate exposure (shown in Figure 5) and mortality (discussed in Section 6). From Figure 5, we see that the uncertainty in η, when translated to an annual exposure level, are larger than the differences in exposure levels estimated from model and satellite-based $PM_{2.5}$, suggesting that satellite-based products which rely strongly on the model or which do not account for the variability in the aforementioned variables, does not necessarily provide a definitively better estimate of exposure. Secondly, these uncertainties in many regions are greater than the difference between both the model and surface $PM_{2.5}$ and the satellite-based and surface observations. While these comparisons are limited spatially and temporally, this highlights that constraining the model with the satellite observations can improve estimates of $PM_{2.5}$ but there remains a large amount of uncertainty in these estimates.

### 4.3 Selection of concentration response function and relative risk

The choice of the shape of the concentration response function (CRF) and relative risk ratio value explains much of the difference in burden estimated in different studies listed in Table 1. In general, it is difficult to determine risks at the population level and ~~ambient air quality measurements may not necessarily be representative of an individual's actual exposure. S~~studies have found that using ambient concentrations tend to under predict health effects (e.g. Hubbell et al., 2009). However, personal monitoring is costly and time-intensive, and therefore, epidemiology studies generally rely on determining population~~-~~level concentration response functions rather than personal-level exposure responses. However, populations also respond differently; and therefore the shape and magnitude of this response varies among studies.

For an initial metric of the uncertainty in the risk ratios, studies often include estimates generated using the ~~that we used, we can include the range using the~~ 95% confidence intervals of the RR determined in the original study (as shown in Figure 2 ~~in the Laden et al. (2006) and Krewski et al. (2009) studies~~. A confidence interval shows the statistical range within which the true PM coefficient for the study population is likely to lie, which could be a single city, region, or population group. The Krewski et al. [2009] study, which is a reanalysis of the American Cancer Society (ACS) Cancer Prevention Study II (CPS-II), included 1.2 million people in the Los Angeles and New York City regions, whereas the Laden et al. (2006) study, an extended analysis of the Harvard Six Cities Studies, included 8,096 white participants. Using just these confidence

intervals as a measure of uncertainty suggests that there exists a large range of uncertainty in population-level health responses to exposure and caution should be exercised when attempting to transfer these values beyond the population from which they were determined in order to estimate national-level mortality burdens based on ambient concentrations. The IER coefficients from Burnett et al. (2014) are generated using the risk ratios, threshold values, and confidence intervals from previous studies and therefore also provide a large range in premature mortality estimates. To depict this range, we also include the 5th and 95th percentile estimates in addition to the mean estimate and show the maximum value in our sensitivity tests.

To test the impact of methodological choices associated with the burden calculation, we compare results using different concentration response functions and relative risk ratios that previous studies have used. Table ~~4~~ 3 lists the different choices that we explore regarding the CRF and relative risk, the study that used these values, and the resulting percent change in burden compared to our initial estimates using the IER from Burnett et al. (2014). In particular we compare our results using risk ratio values from Krewski et al. (2009), Pope et al. (2002) and Laden et al. (2006), and log-linear and power law relationships. Figure ~~8~~ 9 shows that the largest difference in burden is associated with using the higher risk ratios from Laden et al. (2006) vs. using Krewski et al. (2009) or the mean estimates determined using the IER coefficients from Burnett et al. (2014), the former suggest a much greater mortality response to $PM_{2.5}$ exposure.

~~We also test the use of a "threshold" value. Most experts in health impacts of ambient air quality agree that there is no population-level threshold (although there may be individual-level thresholds, e.g. Roman et al., 2008); however, there are few epidemiological studies in regions with very low annual average concentrations (Crouse et al., 2012 records a 1.9 $\mu gm^{-3}$ annual concentration in rural Canada) making it difficult to determine the health risks in relatively clean conditions. Therefore, the lowest measured concentration is often used to account for uncertainty in the shape of the response function at very low concentrations. Other studies have not used a threshold value, but a PRB concentration (generally 0-2 $\mu gm^{-3}$) to calculate mortality attributable to concentrations due to national anthropogenic emissions (not transport or natural sources, i.e. the concentration levels to which policies might be able to reduce, e.g. Fann et al., 2012).~~

Our estimates of Section 3 also use the same relative risk values for every location. However, studies have found that different populations have varied responses to exposure (potential for

"effect modification") (Dominici et al., 2003). One of the main uncertainties in ~~our~~ these methods is relying on risk ratios that are primarily determined from epidemiology studies conducted in the United States, which may not represent the actual risks for populations in China. Long-term epidemiology studies examining exposure to $PM_{2.5}$ across broad regions of China are scarce, but studies using acute exposure to $PM_{2.5}$ or chronic exposure to $PM_{10}$ or total suspended particles have suggested lower exposure-response coefficients than determined by studies conducted in the U.S. and Europe (Aunan and Pan, 2004; Chen et al., 2013b; Shang et al., 2013), indicating that assessments which use CRFs from studies conducted in the U.S. might overestimate the health effects in China.

We also explore using different "threshold" values. The IER function uses threshold values between 5.8 $\mu gm^{-3}$ and 8.8 $\mu gm^{-3}$. In the U.S., ~~using the lowest measured value of 5.8 $\mu gm^{-3}$ from the Krewski et al. (2009) study has a significant impact,~~ higher threshold values can significantly reduc~~e~~ing ~~our~~ burden estimates. ~~by more than half (which consequently would put our estimates into agreement with Lelieveld et al., 2013 as it reduce the impact to zero in any rural region with low concentration)~~. When we compare sensitivity tests that use the same CRF (Krewski et al., 2009) but ~~Using~~ with a ~~CF value of 4 $\mu gm^{-3}$ or~~ a regional PRB concentration instead of the lowest measured level (5.8 $\mu gm^{-3}$), ~~also reduces~~ the premature mortality ~~our~~ estimates are reduced, ~~s but not by as much,~~ suggesting that the choice of this value is very important in the U.S. where annual mean concentrations are relatively low. ~~-~~However, in China these threshold values have less ~~than a 10%~~ impact on our results because annual mean concentrations are high enough that subtracting a threshold makes little difference. Conversely, using a ceiling value of 30 $\mu gm^{-3}$ or ~~(~~50 $\mu gm^{-3}$~~)~~ produces no difference in the U.S. (0% of the population experiences annual concentration values greater than 30 $\mu gm^{-3}$), while strongly reducing burden estimates in China.

We also see that the shape of the CRF produces different results between the U.S. and China. Using a power law or log-linear (Equation 6) function increase relative risks at low concentrations and decreases risk ratios at high concentrations such that total disease burden estimates increase in the U.S. and decrease in China. In the U.S., a log-linear CRF is almost equivalent to a linear response because of the low concentrations. In general, the shape of the concentration response function is more important at low or very high concentrations.

## 4.4 Comparison of uncertainty

Figure 10 provides a summary of the different sources of uncertainty discussed here is shown in Figure 10. We show the mortality burdens for respiratory disease, lung cancer and heart disease associated with chronic exposure to ambient $PM_{2.5}$ and calculated using annual average model-based and "satellite-based" values (from MISR and MODIS) for both the U.S. and China. We show here that the satellite-based estimates suggest slightly higher national burdens in the U.S. and slightly lower in China. However, our values using these different annual average concentrations fall within the range of values found in the literature (Table 1).

We further contrast these estimates to the range in uncertainty associated with our observations and methodology. The difference between the burden calculated using strictly the model or the satellite-based approach is greater than the uncertainty range in the satellite AOD, suggesting that this difference is outside of the scope of measurement limitations and errors. However, the potential uncertainty in the satellite-based estimate due to the conversion from AOD to surface $PM_{2.5}$ (represented by the model η) is substantially larger, larger even than the difference between model-derived and satellite-derived estimates. Therefore, while constraining the model estimate of $PM_{2.5}$ by actual observations should improve our health effect estimates, the uncertainty in the required model information may limit the accuracy of this approach. Again, we stress that these are "potential" model uncertainties which may overestimate the true uncertainty in regions where the model accurately represents the composition and distribution of aerosols. We also acknowledge that we have investigated a limited set of factors; additional biases may exacerbate these uncertainties.— However, adding additional observational data and model estimates can also help to better constrain these satellite-based $PM_{2.5}$ estimates (Brauer et al., 2012, 2015; van Donkelaar et al., 2015a, 2015b).

Figure 10 also conveys the range in mortality estimates for the U.S. and China that can result from varying choices for the risk ratio or shape of the concentration response. While epidemiology studies attempt to statistically account for differences in populations and confounding variables, there is still a large spread in determined risk ratios. Just as important, or perhaps more so than determining ambient concentrations, applying response functions is a determining factor in quantifying the burden of mortality due to outdoor air quality. Differences in exposure estimates can be overshadowed by these different approaches. As an added example,

we calculated the mortality burden using only populated places, similar to Lelieveld et al. (2013) and Cohen et al. (2004) and find that for the U.S. this decreased the burden by 13%, (satellite-based, 18% for model). For China, this reduces the burden estimate by 72%. Differences in our estimates here and those found in the literature can be partly attributed to differences in application of the CRF function, along with differences in baseline mortalities and population estimates. Disease burdens estimated in various studies can therefore only be truly compared when the methodology is harmonized.

## 5 ~~Discussion and~~ Conclusions

Calculating health burdens is an extremely important endeavor for informing air pollution policy, but literature estimates cover a large range due to differences in methodology regarding both the measurement of ambient concentrations and the health impact assessment. Satellite observations have proved useful in estimating exposure and the resulting health impacts (van Donkelaar et al., 2015b; Yao et al., 2013)~~(e.g. van Donkelaar et al., 2015; Yao et al., 2013)~~. However, there remain large uncertainties associated with these satellite measurements and the methods for translating them into surface air quality that needs to be further investigated. Our goal with this work is to explore how mortality burden estimates are made and how choices within this methodology can explain some of these discrepancies. We also aim to provide a context for interpreting the quantification of $PM_{2.5}$ chronic exposure health burdens.

While we have discussed several potential sources for uncertainty in calculating health burdens with satellite-based $PM_{2.5}$, there are still a significant number of other sources of uncertainty that we did not explore. There are processes that could impact the AOD to $PM_{2.5}$ relationship in the model, such as different emissions and removal processes. Additionally, our sensitivity test results are likely partly tied to the spatial resolution of the model and~~,~~ the satellite AOD, and their ability to capture finer spatial variations in pollution in regions with high populations. ~~Although~~However, ~~as~~ Thompson et al., (2014) suggest that~~,~~ uncertainty in the CRF will likely still have a larger impact than uncertainties in population-weighted concentrations due to model resolution.

In Figure 910, we show the mortality burdens for respiratory disease, lung cancer and heart disease associated with chronic exposure to ambient $PM_{2.5}$ and calculated using annual average model-based and "satellite-based" values (from MISR and MODIS) for both the U.S. and China. We show here that the satellite-based estimates suggest slightly higher national burdens in the U.S. and slightly lower in China. However, our values using these different annual average concentrations fall within the range of values found in the literature (Table 1).

We further contrast these estimates to the range in uncertainty associated with our observations and methodology. The difference between the burden calculated using strictly the model or the satellite-based approach is greater than the uncertainty range in the satellite AOD, suggesting that this difference is outside of the scope of measurement limitations and errors. However, the potential uncertainty in the satellite-based estimate due to the conversion from AOD to surface $PM_{2.5}$ (represented by the model η) is substantially larger, larger even than the difference between model-derived and satellite-derived estimates. Therefore, while constraining the model estimate of $PM_{2.5}$ by actual observations should improve our health effect estimates, the uncertainty in the required model information may limit the accuracy of this approach. Again, we stress that these are "potential" model uncertainties which may overestimate the true uncertainty in regions where the model accurately represents the composition and distribution of aerosols. We also acknowledge that we have investigated a limited set of factors; additional biases may exacerbate these uncertainties.

Figure 910 also conveys the range in mortality estimates for the U.S. and China that can result from varying choices for the risk ratio or shape of the concentration response. While epidemiology studies attempt to statistically account for differences in populations and confounding variables, there is still a large spread in determined risk ratios. Just as important, or perhaps more so than determining ambient concentrations, correctly applying response functions is a determining factor in quantifying the burden of mortality due to outdoor air quality. Differences in exposure estimates can be overshadowed by these different epidemiological approaches. As an added example, we calculated the mortality burden using only populated places, similar to and find that for the U.S. this decreased the burden by 13 %, (satellite-based,

~~18% for model). For China, this reduces the burden estimate by 72%. Disease burdens estimated in various studies can therefore only be truly compared when the methodology is harmonized.~~

Satellite measurements have provided great advancements in monitoring global air quality, providing information in regions with previously few measurements. However, further progress still needs to be made in determining how to characterize exposure to ambient $PM_{2.5}$ using these satellite observations, especially as they are becoming more widely used in epidemiological studies and health impact assessments. Reducing uncertainty, even at the lower concentrations observed in the U.S., is important if these methods and datasets are to be used for policy assessment or air quality standards. However, ~~A~~as air pollution is a leading environmentally-related cause of premature mortality, the difficulties in applying this data should not negate the importance of this endeavor. Overcoming sampling limitations in satellite observations and better accounting for regional biases could help to reduce the uncertainty in satellite-retrieved AOD and adding additional observational data and model estimates can help to better constrain satellite-based $PM_{2.5}$ estimates (Brauer et al., 2012, 2015; van Donkelaar et al., 2015a, 2015b). Future geostationary satellites will also be critical to advance this methodology and will provide extremely valuable information for daily monitoring and tracking of air quality. Furthermore, these geostationary observations, in concert with greater surface monitoring, will offer new constraints for epidemiological studies to develop health risk assessments and lessen the uncertainty in applying concentration-response functions and determining health burdens.

## Acknowledgements

This work was supported in part by the MIT Center for Environmental Health Science (via NIH grant P30-ES002109). We acknowledge useful discussions with Dr. Jennifer Peel.

## References


Alexander, B., Park, R. J., Jacob, D. J., Li, Q. B., Yantosca, R. M., Savarino, J., Lee, C. C. W. and Thiemens, M. H.: Sulfate formation in sea-salt aerosols: Constraints from oxygen isotopes, J. Geophys. Res., 110(D10), D10307, doi:10.1029/2004JD005659, 2005.

Anenberg, S. C., Horowitz, L. W., Tong, D. Q. and West, J. J.: An Estimate of the Global Burden of Anthropogenic Ozone and Fine Particulate Matter on Premature Human Mortality Using Atmospheric Modeling, Environ. Health Perspect., 118(9), 1189–1195, doi:10.1289/ehp.0901220, 2010.

Apte, J. S., Marshall, J. D., Cohen, A. J. and Brauer, M.: Addressing Global Mortality from Ambient PM2.5 - Environmental Science & Technology (ACS Publications), Environ. Sci. Technol., 49(13), 8057–8066, 2015.

Aunan, K. and Pan, X.-C.: Exposure-response functions for health effects of ambient air pollution applicable for China -- a meta-analysis, Sci. Total Environ., 329(1-3), 3–16, doi:10.1016/j.scitotenv.2004.03.008, 2004.

Bibi, H., Alam, K., Chishtie, F., Bibi, S., Shahid, I. and Blaschke, T.: Intercomparison of MODIS, MISR, OMI, and CALIPSO aerosol optical depth retrievals for four locations on the Indo-Gangetic plains and validation against AERONET data, Atmos. Environ., 111, 113–126, doi:10.1016/j.atmosenv.2015.04.013, 2015.

Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., Ezzati, M., Henderson, S. B., Krzyzanowski, M., Martin, R. V., Van Dingenen, R., van Donkelaar, A. and Thurston, G. D.: Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution, Environ. Sci. Technol., 46(2), 652–660, doi:10.1021/es2025752, 2012.

Brauer, M., Freedman, G., Frostad, J., van Donkelaar, A., Martin, R. V., Dentener, F., Dingenen, R. van, Estep, K., Amini, H., Apte, J. S., Balakrishnan, K., Barregard, L., Broday, D., Feigin, V., Ghosh, S., Hopke, P. K., Knibbs, L. D., Kokubo, Y., Liu, Y., Ma, S., Morawska, L., Sangrador, J. L. T., Shaddick, G., Anderson, H. R., Vos, T., Forouzanfar, M. H., Burnett, R. T. and Cohen, A.: Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013, Environ. Sci. Technol., doi:10.1021/acs.est.5b03709, 2015.

Burnett, R. T., Pope, C. A., III, Ezzati, M., Olives, C., Lim, S. S., Mehta, S., Shin, H. H., Singh, G., Hubbell, B., Brauer, M., Anderson, H. R., Smith, K. R., Balmes, J. R., Bruce, N. G., Kan, H., Laden, F., Prüss-Ustün, A., Turner, M. C., Gapstur, S. M., Diver, W. R. and Cohen, A.: An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure, Environ. Health Perspect., doi:10.1289/ehp.1307049, 2014.

Cheng, T., Chen, H., Gu, X., Yu, T., Guo, J. and Guo, H.: The inter-comparison of MODIS, MISR and GOCART aerosol products against AERONET data over China, J. Quant. Spectrosc. Radiat. Transf., 113(16), 2135–2145, doi:10.1016/j.jqsrt.2012.06.016, 2012.

Chen, H. and Goldberg, M. S.: The effects of outdoor air pollution on chronic illnesses, McGill J. Med. MJM, 12(1), 58–64, 2009.

Chen, W., Zheng, R., Zhang, S., Zhao, P., Li, G., Wu, L. and He, J.: The incidences and mortalities of major cancers in China, 2009, Chin. J. Cancer, 32(3), 106–112, doi:10.5732/cjc.013.10018, 2013a.

Chen, Y., Ebenstein, A., Greenstone, M. and Li, H.: Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy, Proc. Natl. Acad. Sci., 110(32), 12936–12941, doi:10.1073/pnas.1300018110, 2013b.

Cohen, A. J., Anderson, H. R., Ostro, B., Pandey, K. D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope III, C. A., Romieu, I., Samet, J. M. and Smith, K. R.: Urban air pollution, in Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors, vol. 1, edited by M. Ezzati, A. D. Lopez, A. Rodgers, and C. J. L. Murray, pp. 1353–1434, World Health Organization, Geneva., 2004.

Cohen, A. J., Ross Anderson, H., Ostro, B., Pandey, K. D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J. M. and Smith, K.: The global burden of disease due to outdoor air pollution, J. Toxicol. Environ. Health A, 68(13-14), 1301–1307, doi:10.1080/15287390590936166, 2005.

Cooke, W. F., Liousse, C., Cachier, H. and Feichter, J.: Construction of a 1° × 1° fossil fuel emission data set for carbonaceous aerosol and implementation and radiative impact in the ECHAM4 model, J. Geophys. Res., 104(D18), 22137–22,162, doi:10.1029/1999JD900187, 1999.

Crouse, D. L., Peters, P. A., van Donkelaar, A., Goldberg, M. S., Villeneuve, P. J., Brion, O., Khan, S., Atari, D. O., Jerrett, M., Pope, C. A., Brauer, M., Brook, J. R., Martin, R. V., Stieb, D. and Burnett, R. T.: Risk of Nonaccidental and Cardiovascular Mortality in Relation to Long-term Exposure to Low Concentrations of Fine Particulate Matter: A Canadian National-Level Cohort Study, Environ. Health Perspect., 120(5), 708–714, doi:10.1289/ehp.1104049, 2012.

van Donkelaar, A., Martin, R. V. and Park, R. J.: Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing, J. Geophys. Res. Atmospheres, 111(D21), n/a–n/a, doi:10.1029/2005JD006996, 2006.

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C. and Villeneuve, P. J.: Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application, Environ. Health Perspect., 118(6), 847–855, doi:10.1289/ehp.0901623, 2010.

van Donkelaar, A., Martin, R. V., Levy, R. C., da Silva, A. M., Krzyzanowski, M., Chubarova, N. E., Semutnikova, E. and Cohen, A. J.: Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the Moscow fires in 2010, Atmos. Environ., 45(34), 6225–6232, doi:10.1016/j.atmosenv.2011.07.068, 2011.

van Donkelaar, A., Martin, R. V., Spurr, R. J. D. and Burnett, R. T.: High-Resolution Satellite-Derived PM2.5 from Optimal Estimation and Geographically Weighted Regression over North America, Environ. Sci. Technol., 49(17), 10482–10491, doi:10.1021/acs.est.5b02076, 2015a.

van Donkelaar, A., Martin, R. V., Brauer, M. and Boys, B.: Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter, Environ. Health Perspect., 123, 135–143, doi:http://dx.doi.org/10.1289/ehp.1408646, 2015b.

Evans, J., van Donkelaar, A., Martin, R. V., Burnett, R., Rainham, D. G., Birkett, N. J. and Krewski, D.: Estimates of global mortality attributable to particulate air pollution using satellite imagery, Environ. Res., 120, 33–42, doi:10.1016/j.envres.2012.08.005, 2013.

Fairlie, D. T., Jacob, D. J. and Park, R. J.: The impact of transpacific transport of mineral dust in the United States, Atmos. Environ., 41(6), 1251–1266, doi:10.1016/j.atmosenv.2006.09.048, 2007.

Fang, Y., Naik, V., Horowitz, L. W. and Mauzerall, D. L.: Air pollution and associated human mortality: the role of air pollutant emissions, climate change and methane concentration increases from the preindustrial period to present, Atmos Chem Phys, 13(3), 1377–1394, doi:10.5194/acp-13-1377-2013, 2013.

Fann, N., Lamson, A. D., Anenberg, S. C., Wesson, K., Risley, D. and Hubbell, B. J.: Estimating the National Public Health Burden Associated with Exposure to Ambient PM2.5 and Ozone, Risk Anal., 32(1), 81–95, doi:10.1111/j.1539-6924.2011.01630.x, 2012.

Fann, N., Fulcher, C. M. and Baker, K.: The Recent and Future Health Burden of Air Pollution Apportioned Across U.S. Sectors, Environ. Sci. Technol., 47(8), 3580–3589, doi:10.1021/es304831q, 2013.

Ford, B. and Heald, C. L.: An A-train and model perspective on the vertical distribution of aerosols and CO in the Northern Hemisphere, J. Geophys. Res., 117(D6), D06211, doi:10.1029/2011JD016977, 2012.

Ford, B. and Heald, C. L.: Aerosol loading in the Southeastern United States: reconciling surface and satellite observations, Atmos Chem Phys, 13(18), 9269–9283, doi:10.5194/acp-13-9269-2013, 2013.

Forouzanfar, M. H., Alexander, L., Anderson, H. R., Bachman, V. F., Biryukov, S., Brauer, M., Burnett, R., Casey, D., Coates, M. M., Cohen, A., Delwiche, K., Estep, K., Frostad, J. J., KC, A., Kyu, H. H., Moradi-Lakeh, M., Ng, M., Slepak, E. L., Thomas, B. A., Wagner, J., Aasvang, G. M., Abbafati, C., Ozgoren, A. A., Abd-Allah, F., Abera, S. F., Aboyans, V., Abraham, B., Abraham, J. P., Abubakar, I., Abu-Rmeileh, N. M. E., Aburto, T. C., Achoki, T., Adelekan, A., Adofo, K., Adou, A. K., Adsuar, J. C., Afshin, A., Agardh, E. E., Al Khabouri, M. J., Al Lami, F. H., Alam, S. S., Alasfoor, D., Albittar, M. I., Alegretti, M. A., Aleman, A. V., Alemu, Z. A., Alfonso-Cristancho, R., Alhabib, S., Ali, R., Ali, M. K., Alla, F., Allebeck, P., Allen, P. J., Alsharif, U., Alvarez, E., Alvis-Guzman, N., Amankwaa, A. A., Amare, A. T., Ameh, E. A., Ameli, O., Amini, H., Ammar, W., Anderson, B. O., Antonio, C. A. T., Anwari, P., Cunningham,

S. A., Arnlöv, J., Arsenijevic, V. S. A., Artaman, A., Asghar, R. J., Assadi, R., Atkins, L. S., Atkinson, C., Avila, M. A., Awuah, B., Badawi, A., Bahit, M. C., Bakfalouni, T., Balakrishnan, K., Balalla, S., Balu, R. K., Banerjee, A., Barber, R. M., Barker-Collo, S. L., Barquera, S., Barregard, L., Barrero, L. H., Barrientos-Gutierrez, T., Basto-Abreu, A. C., Basu, A., Basu, S., Basulaiman, M. O., Ruvalcaba, C. B., Beardsley, J., Bedi, N., Bekele, T., Bell, M. L., Benjet, C., Bennett, D. A., et al.: Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013, The Lancet, 386(10010), 2287–2323, doi:10.1016/S0140-6736(15)00128-2, 2015.

Fu, J., Jiang, D., Lin, G., Liu, K. and Wang, Q.: An ecological analysis of PM2.5 concentrations and lung cancer mortality rates in China, BMJ Open, 5(11), e009452, doi:10.1136/bmjopen-2015-009452, 2015.

Giannadaki, D., Pozzer, A. and Lelieveld, J.: Modeled global effects of airborne desert dust on air quality and premature mortality, Atmos Chem Phys, 14(2), 957–968, doi:10.5194/acp-14-957-2014, 2014.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I. and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos Chem Phys, 6(11), 3181–3210, doi:10.5194/acp-6-3181-2006, 2006.

Hand, J. L., Schichtel, B. A., Pitchford, M., Malm, W. C. and Frank, N. H.: Seasonal composition of remote and urban fine particulate matter in the United States, J. Geophys. Res., 117(D5), D05209, doi:10.1029/2011JD017122, 2012.

Heald, C. L., Collett, J. L., Lee, T., Benedict, K. B., Schwandner, F. M., Li, Y., Clarisse, L., Hurtmans, D. R., Van Damme, M., Clerbaux, C., Coheur, P.-F. and Pye, H. O. T.: Atmospheric ammonia and particulate inorganic nitrogen over the United States, Atmospheric Chem. Phys. Discuss., 12(8), 19455–19498, doi:10.5194/acpd-12-19455-2012, 2012.

Henze, D. K., Seinfeld, J. H., Ng, N. L., Kroll, J. H., Fu, T.-M., Jacob, D. J. and Heald, C. L.: Global modeling of secondary organic aerosol formation from aromatic hydrocarbons: high- vs. low-yield pathways, Atmos Chem Phys, 8(9), 2405–2420, doi:10.5194/acp-8-2405-2008, 2008.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I. and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sens. Environ., 66(1), 1–16, doi:10.1016/S0034-4257(98)00031-5, 1998.

Hubbell, B., Fann, N. and Levy, J. I.: Methodological considerations in developing local-scale health impact assessments: balancing national, regional, and local data, Air Qual. Atmosphere Health, 2(2), 99–110, doi:10.1007/s11869-009-0037-z, 2009.

Hudman, R. C., Jacob, D. J., Turquety, S., Leibensperger, E. M., Murray, L. T., Wu, S., Gilliland, A. B., Avery, M., Bertram, T. H., Brune, W., Cohen, R. C., Dibb, J. E., Flocke, F. M., Fried, A., Holloway, J., Neuman, J. A., Orville, R., Perring, A., Ren, X., Sachse, G. W., Singh, H. B.,

Swanson, A. and Wooldridge, P. J.: Surface and lightning sources of nitrogen oxides over the United States: Magnitudes, chemical evolution, and outflow, J. Geophys. Res., 112(D12), D12S05, doi:10.1029/2006JD007912, 2007.

Hudman, R. C., Murray, L. T., Jacob, D. J., Millet, D. B., Turquety, S., Wu, S., Blake, D. R., Goldstein, A. H., Holloway, J. and Sachse, G. W.: Biogenic versus anthropogenic sources of CO in the United States, Geophys. Res. Lett., 35(4), L04801, doi:10.1029/2007GL032393, 2008.

Hystad, P., Demers, P. A., Johnson, K. C., Brook, J., Donkelaar, A. van, Lamsal, L., Martin, R. and Brauer, M.: Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study, Environ. Health, 11(1), 22, doi:10.1186/1476-069X-11-22, 2012.

Jiang, X., Liu, Y., Yu, B. and Jiang, M.: Comparison of MISR aerosol optical thickness with AERONET measurements in Beijing metropolitan area, Remote Sens. Environ., 107, 45–53, 2007.

Kahn, R. A., Gaitley, B. J., Martonchik, J. V., Diner, D. J., Crean, K. A. and Holben, B.: Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations, J. Geophys. Res. Atmospheres, 110(D10), D10S04, doi:10.1029/2004JD004706, 2005.

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T. F., Smirnov, A. and Holben, B. N.: Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, J. Geophys. Res. Atmospheres, 115(D23), D23209, doi:10.1029/2010JD014601, 2010.

Krewski, D., Jerrett, M., Burnett, R. T., Ma, R., Hughes, E., Shi, Y., Turner, M. C., Pope, C. A., 3rd, Thurston, G., Calle, E. E., Thun, M. J., Beckerman, B., DeLuca, P., Finkelstein, N., Ito, K., Moore, D. K., Newbold, K. B., Ramsay, T., Ross, Z., Shin, H. and Tempalski, B.: Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality, Res. Rep. Health Eff. Inst., (140), 5–114; discussion 115–136, 2009.

Laden, F., Schwartz, J., Speizer, F. E. and Dockery, D. W.: Reduction in Fine Particulate Air Pollution and Mortality, Am. J. Respir. Crit. Care Med., 173(6), 667–672, doi:10.1164/rccm.200503-443OC, 2006.

Leibensperger, E. M., Mickley, L. J., Jacob, D. J., Chen, W.-T., Seinfeld, J. H., Nenes, A., Adams, P. J., Streets, D. G., Kumar, N. and Rind, D.: Climatic effects of 1950–2050 changes in US anthropogenic aerosols – Part 1: Aerosol trends and radiative forcing, Atmos Chem Phys, 12(7), 3333–3348, doi:10.5194/acp-12-3333-2012, 2012.

Lelieveld, J., Barlas, C., Giannadaki, D. and Pozzer, A.: Model calculated global, regional and megacity premature mortality due to air pollution, Atmos Chem Phys, 13(14), 7023–7037, doi:10.5194/acp-13-7023-2013, 2013.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525(7569), 367–371, doi:10.1038/nature15371, 2015.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R. and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmospheric Chem. Phys., 10(21), 10399–10420, doi:10.5194/acp-10-10399-2010, 2010.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F. and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmos Meas Tech, 6(11), 2989–3034, doi:10.5194/amt-6-2989-2013, 2013.

Li, J., Carlson, B. E. and Lacis, A. A.: How well do satellite AOD observations represent the spatial and temporal variability of PM2.5 concentration for the United States?, Atmos. Environ., 102, 260–273, doi:10.1016/j.atmosenv.2014.12.010, 2015.

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., Amann, M., Anderson, H. R., Andrews, K. G., Aryee, M., Atkinson, C., Bacchus, L. J., Bahalim, A. N., Balakrishnan, K., Balmes, J., Barker-Collo, S., Baxter, A., Bell, M. L., Blore, J. D., Blyth, F., Bonner, C., Borges, G., Bourne, R., Boussinesq, M., Brauer, M., Brooks, P., Bruce, N. G., Brunekreef, B., Bryan-Hancock, C., Bucello, C., Buchbinder, R., Bull, F., Burnett, R. T., Byers, T. E., Calabria, B., Carapetis, J., Carnahan, E., Chafe, Z., Charlson, F., Chen, H., Chen, J. S., Cheng, A. T.-A., Child, J. C., Cohen, A., Colson, K. E., Cowie, B. C., Darby, S., Darling, S., Davis, A., Degenhardt, L., Dentener, F., Des Jarlais, D. C., Devries, K., Dherani, M., Ding, E. L., Dorsey, E. R., Driscoll, T., Edmond, K., Ali, S. E., Engell, R. E., Erwin, P. J., Fahimi, S., Falder, G., Farzadfar, F., Ferrari, A., Finucane, M. M., Flaxman, S., Fowkes, F. G. R., Freedman, G., Freeman, M. K., Gakidou, E., Ghosh, S., Giovannucci, E., Gmel, G., Graham, K., Grainger, R., Grant, B., Gunnell, D., Gutierrez, H. R., Hall, W., Hoek, H. W., Hogan, A., Hosgood, H. D., Hoy, D., Hu, H., Hubbell, B. J., Hutchings, S. J., Ibeanusi, S. E., Jacklyn, G. L., Jasrasaria, R., Jonas, J. B., Kan, H., Kanis, J. A., Kassebaum, N., Kawakami, N., Khang, Y.-H., Khatibzadeh, S., Khoo, J.-P., Kok, C., et al.: A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010, Lancet, 380(9859), 2224–2260, doi:10.1016/S0140-6736(12)61766-8, 2012.

Liu, Y., Park, R. J., Jacob, D. J., Li, Q., Kilaru, V. and Sarnat, J. A.: Mapping annual mean ground-level PM2.5 concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States, J. Geophys. Res. Atmospheres, 109(D22), D22206, doi:10.1029/2004JD005025, 2004.

Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J. and Koutrakis, P.: Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing, Environ. Sci. Technol., 39(9), 3269–3278, 2005.

Liu, Y., Koutrakis, P. and Kahn, R.: Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: part 1--method development, J. Air Waste Manag. Assoc. 1995, 57(11), 1351–1359, 2007.

Malm, W. C., Sisler, J. F., Huffman, D., Eldred, R. A. and Cahill, T. A.: Spatial and seasonal trends in particle concentration and optical extinction in the United States, J. Geophys. Res. Atmospheres, 99(D1), 1347–1370, doi:10.1029/93JD02916, 1994.

Marlier, M. E., DeFries, R. S., Voulgarakis, A., Kinney, P. L., Randerson, J. T., Shindell, D. T., Chen, Y. and Faluvegi, G.: El Nino and health risks from landscape fire emissions in southeast Asia, Nat. Clim. Change, 3(2), 131–136, doi:10.1038/nclimate1658, 2013.

Park, R. J., Jacob, D. J., Chin, M. and Martin, R. V.: Sources of carbonaceous aerosols over the United States and implications for natural visibility, J GEOPHYS RES, 108, 2002–3190, 2003.

Park, R. J., Jacob, D. J., Field, B. D., Yantosca, R. M. and Chin, M.: Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy, J. Geophys. Res. Atmospheres, 109(D15), n/a–n/a, doi:10.1029/2003JD004473, 2004.

Pope, C. A. and Dockery, D. W.: Air pollution and life expectancy in China and beyond, Proc. Natl. Acad. Sci., 110(32), 12861–12862, doi:10.1073/pnas.1310925110, 2013.

Pope, C. A., Bates, D. V. and Raizenne, M. E.: Health effects of particulate air pollution: time for reassessment?, Environ. Health Perspect., 103(5), 472–480, 1995.

Pope, C. A., Burnett, R. T., Turner, M. C., Cohen, A., Krewski, D., Jerrett, M., Gapstur, S. M. and Thun, M. J.: Lung Cancer and Cardiovascular Disease Mortality Associated with Ambient Air Pollution and Cigarette Smoke: Shape of the Exposure-Response Relationships, Environ. Health Perspect., 119(11), 1616–1621, doi:10.1289/ehp.1103639, 2011.

Pope, C. A., Cropper, M., Coggins, J. and Cohen, A.: Health benefits of air pollution abatement policy: Role of the shape of the concentration–response function, J. Air Waste Manag. Assoc., 65(5), 516–522, doi:10.1080/10962247.2014.993004, 2015.

Pope, C. A., 3rd, Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K. and Thurston, G. D.: Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, JAMA J. Am. Med. Assoc., 287(9), 1132–1141, 2002.

Pope, C. A., 3rd, Burnett, R. T., Krewski, D., Jerrett, M., Shi, Y., Calle, E. E. and Thun, M. J.: Cardiovascular mortality and exposure to airborne fine particulate matter and cigarette smoke: shape of the exposure-response relationship, Circulation, 120(11), 941–948, doi:10.1161/CIRCULATIONAHA.109.857888, 2009.

Punger, E. M. and West, J. J.: The effect of grid resolution on estimates of the burden of ozone and fine particulate matter on premature mortality in the USA, Air Qual. Atmosphere Health, 6(3), 563–573, doi:10.1007/s11869-013-0197-8, 2013.

Qi, Y., Ge, J. and Huang, J.: Spatial and temporal distribution of MODIS and MISR aerosol optical depth over northern China and comparison with AERONET, Chin. Sci. Bull., 58(20), 2497–2506, doi:10.1007/s11434-013-5678-5, 2013.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E. and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, J. Atmospheric Sci., 62(4), 947–973, doi:10.1175/JAS3385.1, 2005.

Remer, L. A., Kleidman, R. G., Levy, R. C., Kaufman, Y. J., Tanré, D., Mattoo, S., Martins, J. V., Ichoku, C., Koren, I., Yu, H. and Holben, B. N.: Global aerosol climatology from the MODIS satellite sensors, J. Geophys. Res. Atmospheres, 113(D14), D14S07, doi:10.1029/2007JD009661, 2008.

Ridley, D. A., Heald, C. L. and Ford, B.: North African dust export and deposition: A satellite and model perspective, J. Geophys. Res., 117(D2), D02202, doi:10.1029/2011JD016794, 2012.

Rohde, R. A. and Muller, R. A.: Air Pollution in China: Mapping of Concentrations and Sources, PLoS ONE, 10(8), e0135749, doi:10.1371/journal.pone.0135749, 2015.

Schiferl, L. D., Heald, C. L., Nowak, J. B., Holloway, J. S., Neuman, J. A., Bahreini, R., Pollack, I. B., Ryerson, T. B., Wiedinmyer, C. and Murphy, J. G.: An investigation of ammonia and inorganic particulate matter in California during the CalNex campaign, J. Geophys. Res. Atmospheres, 119(4), 2013JD020765, doi:10.1002/2013JD020765, 2014.

Shang, Y., Sun, Z., Cao, J., Wang, X., Zhong, L., Bi, X., Li, H., Liu, W., Zhu, T. and Huang, W.: Systematic review of Chinese studies of short-term exposure to air pollution and daily mortality, Environ. Int., 54, 100–111, doi:10.1016/j.envint.2013.01.010, 2013.

Sheppard, L., Burnett, R. T., Szpiro, A. A., Kim, S.-Y., Jerrett, M., Pope, C. A. and Brunekreef, B.: Confounding and exposure measurement error in air pollution epidemiology, Air Qual. Atmosphere Health, 5(2), 203–216, doi:10.1007/s11869-011-0140-9, 2012.

Shi, L., Zanobetti, A., Kloog, I., Coull, B. A., Koutrakis, P., Melly, S. J. and Schwartz, J. D.: Low-Concentration PM2.5 and Mortality: Estimating Acute and Chronic Effects in a Population-Based Study, Environ. Health Perspect., doi:10.1289/ehp.1409111, 2015.

Silva, R. A., West, J. J., Zhang, Y., Anenberg, S. C., Lamarque, J.-F., Shindell, D. T., Collins, W. J., Dalsoren, S., Faluvegi, G., Folberth, G., Horowitz, L. W., Nagashima, T., Naik, V., Rumbold, S., Skeie, R., Sudo, K., Takemura, T., Bergmann, D., Cameron-Smith, P., Cionni, I., Doherty, R. M., Eyring, V., Josse, B., MacKenzie, I. A., Plummer, D., Righi, M., Stevenson, D. S., Strode, S., Szopa, S. and Zeng, G.: Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change, Environ. Res. Lett., 8(3), 034005, doi:10.1088/1748-9326/8/3/034005, 2013.

Snider, G., Weagle, C. L., Martin, R. V., van Donkelaar, A., Conrad, K., Cunningham, D., Gordon, C., Zwicker, M., Akoshile, C., Artaxo, P., Anh, N. X., Brook, J., Dong, J., Garland, R. M., Greenwald, R., Griffith, D., He, K., Holben, B. N., Kahn, R., Koren, I., Lagrosas, N., Lestari, P., Ma, Z., Vanderlei Martins, J., Quel, E. J., Rudich, Y., Salam, A., Tripathi, S. N., Yu, C., Zhang, Q., Zhang, Y., Brauer, M., Cohen, A., Gibson, M. D. and Liu, Y.: SPARTAN: a global

network to evaluate and enhance satellite-based estimates of ground-level particulate matter for global health applications, Atmos Meas Tech, 8(1), 505–521, doi:10.5194/amt-8-505-2015, 2015.

Streets, D. G., Zhang, Q., Wang, L., He, K., Hao, J., Wu, Y., Tang, Y. and Carmichael, G. R.: Revisiting China's CO emissions after the Transport and Chemical Evolution over the Pacific (TRACE-P) mission: Synthesis of inventories, atmospheric modeling, and observations, J. Geophys. Res. Atmospheres, 111(D14), D14306, doi:10.1029/2006JD007118, 2006.

Sun, Q., Hong, X. and Wold, L. E.: Cardiovascular Effects of Ambient Particulate Air Pollution Exposure, Circulation, 121(25), 2755–2765, doi:10.1161/CIRCULATIONAHA.109.893461, 2010.

Thompson, T. M. and Selin, N. E.: Influence of air quality model resolution on uncertainty associated with health impacts, Atmos Chem Phys, 12(20), 9753–9762, doi:10.5194/acp-12-9753-2012, 2012.

Thompson, T. M., Saari, R. K. and Selin, N. E.: Air quality resolution for health impact assessment: influence of regional characteristics, Atmos Chem Phys, 14(2), 969–978, doi:10.5194/acp-14-969-2014, 2014.

Toth, T. D., Zhang, J., Campbell, J. R., Hyer, E. J., Reid, J. S., Shi, Y. and Westphal, D. L.: Impact of data quality and surface-to-column representativeness on the PM2.5 / satellite AOD relationship for the contiguous United States, Atmos Chem Phys, 14(12), 6049–6062, doi:10.5194/acp-14-6049-2014, 2014.

U.S. Environmental Protection Agency: Risk Assessment to Support the Review of the PM Primary National Ambient Air Quality Standards (External Reivew Draft), U.S. EPA, Research Triangle Park, NC., 2009.

U.S. Environmental Protection Agency: Quantitative health risk assessment for particulate matter, U.S. Environmental Protection Agency, Research Triangle Park, NC., 2010.

Villeneuve, P. J., Weichenthal, S. A., Crouse, D., Miller, A. B., To, T., Martin, R. V., van Donkelaar, A., Wall, C. and Burnett, R. T.: Long-term Exposure to Fine Particulate Matter Air Pollution and Mortality Among Canadian Women, Epidemiol. Camb. Mass, 26(4), 536–545, doi:10.1097/EDE.0000000000000294, 2015.

Wang, X., Huang, J., Ji, M. and Higuchi, K.: Variability of East Asia dust events and their long-term trend, Atmos. Environ., 42(13), 3156–3165, doi:10.1016/j.atmosenv.2007.07.046, 2008.

Wang, Y., Zhang, Q. Q., He, K., Zhang, Q. and Chai, L.: Sulfate-nitrate-ammonium aerosols over China: response to 2000–2015 emission changes of sulfur dioxide, nitrogen oxides, and ammonia, Atmos Chem Phys, 13(5), 2635–2652, doi:10.5194/acp-13-2635-2013, 2013.

Wang, Y., Zhang, Q., Jiang, J., Zhou, W., Wang, B., He, K., Duan, F., Zhang, Q., Philip, S. and Xie, Y.: Enhanced sulfate formation during China's severe winter haze episode in January 2013

missing from current models, J. Geophys. Res. Atmospheres, 119(17), 2013JD021426, doi:10.1002/2013JD021426, 2014.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S. and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos Chem Phys, 6(11), 3423–3441, doi:10.5194/acp-6-3423-2006, 2006.

World Bank: World Development Indicators 2015, World Bank, Washington, DC., 2015.

Yao, J., Brauer, M. and Henderson, S. B.: Evaluation of a wildfire smoke forecasting system as a tool for public health protection, Environ. Health Perspect., 121(10), 1142–1147, doi:10.1289/ehp.1306768, 2013.

You, W., Zang, Z., Pan, X., Zhang, L. and Chen, D.: Estimating PM2.5 in Xi'an, China using aerosol optical depth: A comparison between the MODIS and MISR retrieval models, Sci. Total Environ., 505, 1156–1165, doi:10.1016/j.scitotenv.2014.11.024, 2015.

Zhang, J. and Reid, J. S.: MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals, J. Geophys. Res. Atmospheres, 111(D22), D22207, doi:10.1029/2005JD006898, 2006.

Zhang, J. and Reid, J. S.: A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and Level 2 MISR aerosol products, Atmos Chem Phys, 10(22), 10949–10963, doi:10.5194/acp-10-10949-2010, 2010.

Zhang, J., Reid, J. S. and Holben, B. N.: An analysis of potential cloud artifacts in MODIS over ocean aerosol optical thickness products, Geophys. Res. Lett., 32(15), L15803, doi:10.1029/2005GL023254, 2005.

**Table 1** Premature mortality from $PM_{2.5}$ exposure by all-cause (All), heart disease (heart), and lung cancer (LC) as estimated in other studies for the globe, U.S. (or North America), and China (or Asia). Values are for (x 1000 deaths per year). All cause values for this study are calculated as the sum of heart disease, lung cancer, and respiratory disease deaths (as opposed to calculating this based on an all-cause CRF). *Study provides several estimates determined using different CRFs. **Study provides several estimates from 14 different models. Table 3 provides additional information on the data sources and concentrations response functions used in these studies.

| **Study** | **U.S. (North America)** | | | **China (Asia/Western Pacific, East Asia)** | | | **Global** | | | **Year for estimate** |
|---|---|---|---|---|---|---|---|---|---|---|
| | All | Heart | LC | All | Heart | LC | All | Heart | LC | |
| *Evans et al.*, 2013* (WHO region) | | | | | | | 3313~~2~~640-4220 | 12~~56~~1123-1669 | 22~~2~~176-264 | 2004 |
| *Fann et al.*, 2012* (U.S.) | 130-320 | | | | | | | | | 2005 |
| *Anenberg et al.*, 2010* (continents) | 141 | 124 | 17 | 2736 | 2584 | 152 | ~~3381 (~~2077-7714~~)~~ | 3499 ~~(~~1800-4549~~)~~ | 222 ~~(~~39-336~~)~~ | 2000 |
| *Lelieveld et al.*, 2013(U.S. and China) | 55 | 46 | 9.1 | 1006 | 898 | 108 | 2200 | 2000 | 186 | 2005 |
| *Cohen et al.*, 2004* (WHO region) | 28 | 3-55 | 1-12 | ~~487~~355 | 192-504 | 22-53 | ~~800~~799 | ~~712~~474-1132 | 39-105 | 2000 |
| *Lim et al.*, 2012 (GBD 2010, U.S. and China) | 86 | 58 | 20 | 858 | 563 | 185 | 3100 | | | 2010 |
| *Forouzanfar et al.*, 2015 (GBD 2013, U.S. and China) | 78 | 54 | 17 | 916 | 600 | 201 | 2900 | | | 2013 |
| *WHO*, 2014 (WHO region) | 152 | | | 1669 | | | 3700 | 1505 | 227 | 2012 |
| *Fang et al.*, 2013 (North America and East Asia) | | 38 | 4.4 | | 661 | 53 | | 1532 | 95 | 2000 |
| *Silva et al.*, 2013** (continents) | 12.2-77 | | | 908-1240 | | | 1880-2380 | | | 2000 |
| *U.S. EPA*, 2010* (U.S.) | 26-360 | | | | | | | | | 2005 |
| *U.S. Environmental Protection Agency*, 2009 (U.S.) | 144 | | | | | | | | | |
| *Punger and West*, 2013 (U.S.) | 66 | 61 | 9.9 | | | | | | | 2005 |
| *Lelieveld et al.*, 2015 (U.S. and China) | 55 | | | 1357 | | | 3297 | | | 2010 |
| *Sun et al.*, 2015 (U.S.) | 103.3 | 68.3 | 15.4 | | | | | | | 2000 |
| *Rohde and Muller*, 2015 (China) | | | | 1600 | | | | | | 2014 |
| ~~This Study: Satellite (U.S. and China)~~ | ~~113~~ | ~~82~~ | ~~15~~ | ~~2084~~ | ~~1512~~ | ~~220~~ | - | - | - | |
| ~~This Study: Model (U.S. and China)~~ | ~~104~~ | ~~75~~ | ~~14~~ | ~~2171~~ | ~~1570~~ | ~~231~~ | - | - | - | |
| This Study: Satellite (U.S. and China) with *Burnett et al.*, 2014 | ~~49~~50 | 38 | 5 | 127~~14~~ | 9~~20~~ | 138 | | | | 2004-2011 |
| This Study: Model (U.S. and China) with *Burnett et al.*, 2014 | 4~~0~~3 | 3~~0~~2 | 4 | 1300~~297~~ | 931~~0~~ | 14~~34~~ | | | | 2004-2011 |

Table 2. Parameters for the IER model given in Equation 8 determined in *Burnett et al.*, 2014 and as in *Zheng et al.*, 2014.

| Disease | $\alpha$ | $\gamma$ | $C_0$ | $\rho$ |
|---|---|---|---|---|
| Ischemic Heart Disease | 1.65 | 0.0483 | 7.45 | 0.467 |
| Cerebrovascular Disease | 1.31 | 0.0120 | 7.36 | 1.274 |
| Chronic Obstructive Pulmonary Disease | 22.16 | 0.00110 | 7.34 | 0.697 |
| Lung Cancer | 159.22 | 0.00020 | 7.35 | 0.759 |

Table ~~3~~2. List of model sensitivity tests and descriptions with results shown in Figure ~~7~~8.

| Sensitivity Test | Description |
|---|---|
| AvgAOD | AOD is held constant through season while Ƞ varies daily. |
| AvgEta | AOD varies daily, while Ƞ is held constant through season. |
| AvgProf | Column mass varies daily, but shape of vertical profile is held constant for season. AOD and Ƞ vary daily but are re-calculated for redistributed mass. |
| AvgRH | AOD and Ƞ vary daily but are re-calculated assuming relative humidity remains constant throughout season. |
| 2x2.5 | Ƞ values are calculated for simulation run at coarser (2°x2.5°) resolution and then regrid to nested resolution (0.5°x0.666°). |
| $SO_4$ | Assume all mass in column is sulfate and recalculate Ƞ. |
| BC | Assume all mass in column is black carbon and recalculate Ƞ. |
| No $NO_3$ | Calculate AOD and Ƞ without the contribution of nitrate. |

Table 43. Input for premature mortality burden estimate sensitivity tests and the resulting percent change in mortality due to chronic exposure determined from satellite-based concentrations. Parentheses are for values determined from model simulated concentrations.

| RR source | Threshold | CRF shape | % Change USA | % Change China | Study using method | In Fig. 8 |
|---|---|---|---|---|---|---|
| *Krewski et al.*, 2009 | No | Linear | base | base | | K-L |
| *Krewski et al.*, 2009 | ~~Yes, LML~~lowest measured level | Linear | -56 (-62) | -8 (-8) | ~~*Annenburg*~~ *Anenberg et al.*, 2010 | K-L-LM |
| *Krewski et al.*, 2009 | ~~Yes, CF~~counterfactual | Linear | -38 (-43) | -6 (-5) | | K-L-CF |
| *Krewski et al.*, 2009 | ~~Yes, PRB~~policy relevant background | Linear | -10 (-11) | | | K-L-PRB |
| *Krewski et al.*, 2009 | No | Linear to 30 $ugm^{-3}$ | 0 (0) | -30 (-33) | ~~*Anenburg*~~ *Anenberg et al.*, 2010 | K-L-30 |
| *Krewski et al.*, 2009 | No | Linear to 50 $ugm^{-3}$ | 0 (0) | -4 (-8) | | K-L-50 |
| *Krewski et al.*, 2009 | ~~Yes, LML~~lowest measured level | Log-Linear (Eq. 6) | -23 (-30) | -13 (-15) | *Evans et al.*, 2013; *Fann and Risley*, 2013; *U.S. EPA*, 2010 | K-LL-LML |
| *Krewski et al.*, 2009 | ~~Yes, CF~~counterfactual | Log-Linear (Eq. 6) | 28 (23) | -2 (-4) | | K-LL-CF |
| *Krewski et al.*, 2009 | No | Log-Linear (Eq. 5) | 10 (10) | 26 (26) | | K-LL |
| *Krewski et al.*, 2009 | lowest measured level~~Yes, LML~~ | Log-Linear (Eq. 5) | -53 (-59) | 14 (16) | *U.S. EPA*, 2010 | K-LL-LML |
| *Pope et al.*, 2002 | No | Power Law | 44 (52) | -17 (-18) | *Marlier et al.*, 2013 | K-PL |
| *Pope et al.*, 2002 | lowest measured level~~Yes, LML~~ | Power Law | -8 (-8) | -44 (-46) | | K-PL-LML |
| *Pope et al.*, 2002 | ~~Yes, PRB~~policy relevant background | Power Law | 21 (35) | | | K-L-PRB |
| *Laden et al.*, 2006 | No | Linear | 105 (108) | 47 (45) | | L-L |
| *Laden et al.*, 2006 | lowest measured level~~Yes, LML~~ | Linear | -14 (-24) | 38 (36) | *An~~n~~enburg et al.*, 2010; *U.S. EPA*, 2010 | L-L-LML |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Laden et al.*, 2006 | ~~Yes,~~ ~~CF~~counterfactual | Linear | 32 (24) | 51 (49) | | L-L-CF |
| *Burnett et al.*, 2014 | ~~Yes,~~ fitted | IER | base ~~26 (-33)~~ | base ~~18 (-20)~~ | *Lim et al.*, 2012; *Zheng et al.*, 2014; *Lelieveld et al.*, 2015 | B-IER |
| *Burnett et al.*, 2014 | fitted | IER | 167 (167) | 65 (64) | *maximum value detemined from set of coefficeints | B-IER$_{max}$ |
| *Krewski et al.*, 2009 | Lowest measured level (5.8 µg m$^{-3}$) | Equation 5 | 18 (15) | 18 (21) | *Evans et al.*, 2013; *Lelieveld et al.*, 2013 | K-L$_{5.8}$ |
| *Krewski et al.*, 2009 | Lowest measured level (5.8 µg m$^{-3}$), ceiling (30 µg m$^{-3}$) | Equation 5 | 18 (15) | -24 (-26) | *Anenberg et al.*, 2010 | K-L$_{c30}$ |
| *Krewski et al.*, 2009 | Lowest measured level (5.8 µg m$^{-3}$), ceiling (50 µg m$^{-3}$) | Equation 5 | 90 (93) | 6 (4) | *Cohen et al.*, 2004 | K-L$_{c50}$ |
| *Krewski et al.*, 2009 | Lowest measured level (5.8 µg m$^{-3}$) | Equation 6 | 143 (167) | -6 (-7) | *Evans et al.*, 2013 | K-LL$_{5.8}$ |
| *Krewski et al.*, 2009 | Policy Relevant Background | Equation 5 | 134 (158) | | *U.S. EPA*, 2010 | K-L$_{PR}$ |
| *Krewski et al.*, 2009 | No threshold | Equation 5 | 169 (200) | 52 (55) | *Silva et al.*, 2013 | K-L$_{0}$ |
| *Pope et al.*, 2002 | Lowest measured level (5.8 µg m$^{-3}$), ceiling (30 µg m$^{-3}$) | Power Law | 134 (158) | -26 (-28) | *Marlier et al.*, 2013 | P-PL$_{5.8c30}$ |
| *Pope et al.*, 2002 | Lowest measured level (7.5 µg m$^{-3}$) | Power Law | 102 (105) | -15 (-15) | *Pope et al.*, 2002 | P-PL$_{7.5}$ |
| *Laden et al.*, 2006 | lowest measured level (10 µg m$^{-3}$) | Equation 5 | -58 (-68) | 126 (130) | *Anenburg et al.*, 2010; *U.S. EPA*, 2010 | L-L$_{10}$ |
| *Laden et al.*, 2006 | lowest measured level (10 µg m$^{-3}$) | Equation 5 | 239 (275) | | *U.S. EPA*, 2010 | L-L$_{PR}$ |
| *Laden et al.*, 2006 | lowest measured level (10 µg m$^{-3}$); ceiling (30 µg m$^{-3}$) | Equation 5 | | 38 (33) | *Anenburg et al.*, 2010 | L-L$_{c30}$ |
| *Pope et al.*, 2002 | lowest measured level (7.5 µg m$^{-3}$) | Equation 5 | -55 (-58) | -25 (-27) | | P-L$_{7.5}$ |
| *Pope et al.*, 2002 | lowest measured level (7.5 µg m$^{-3}$) | Equation 6 | -29 (-28) | 1 (1) | | P-LL$_{7.5}$ |

1

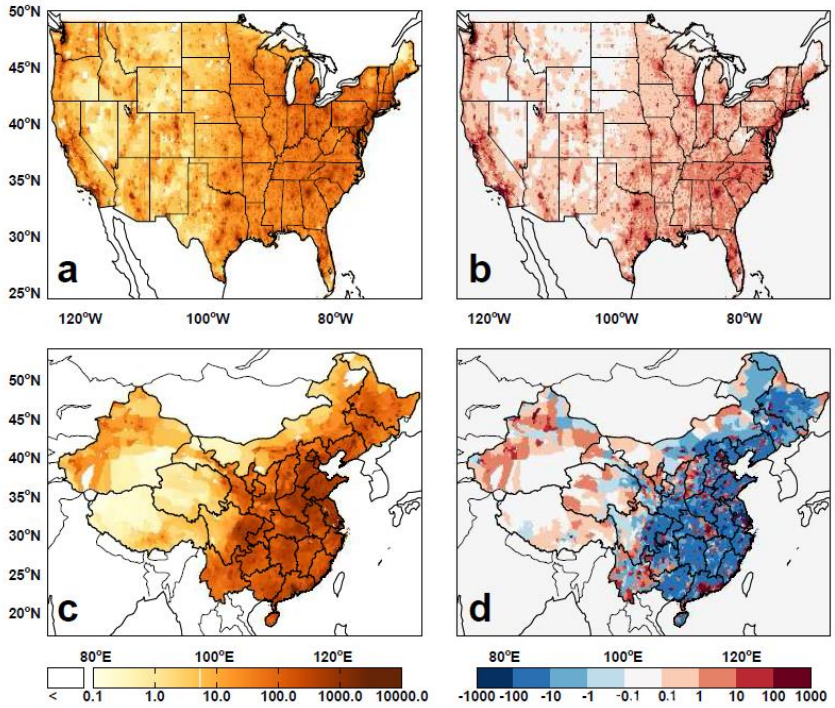
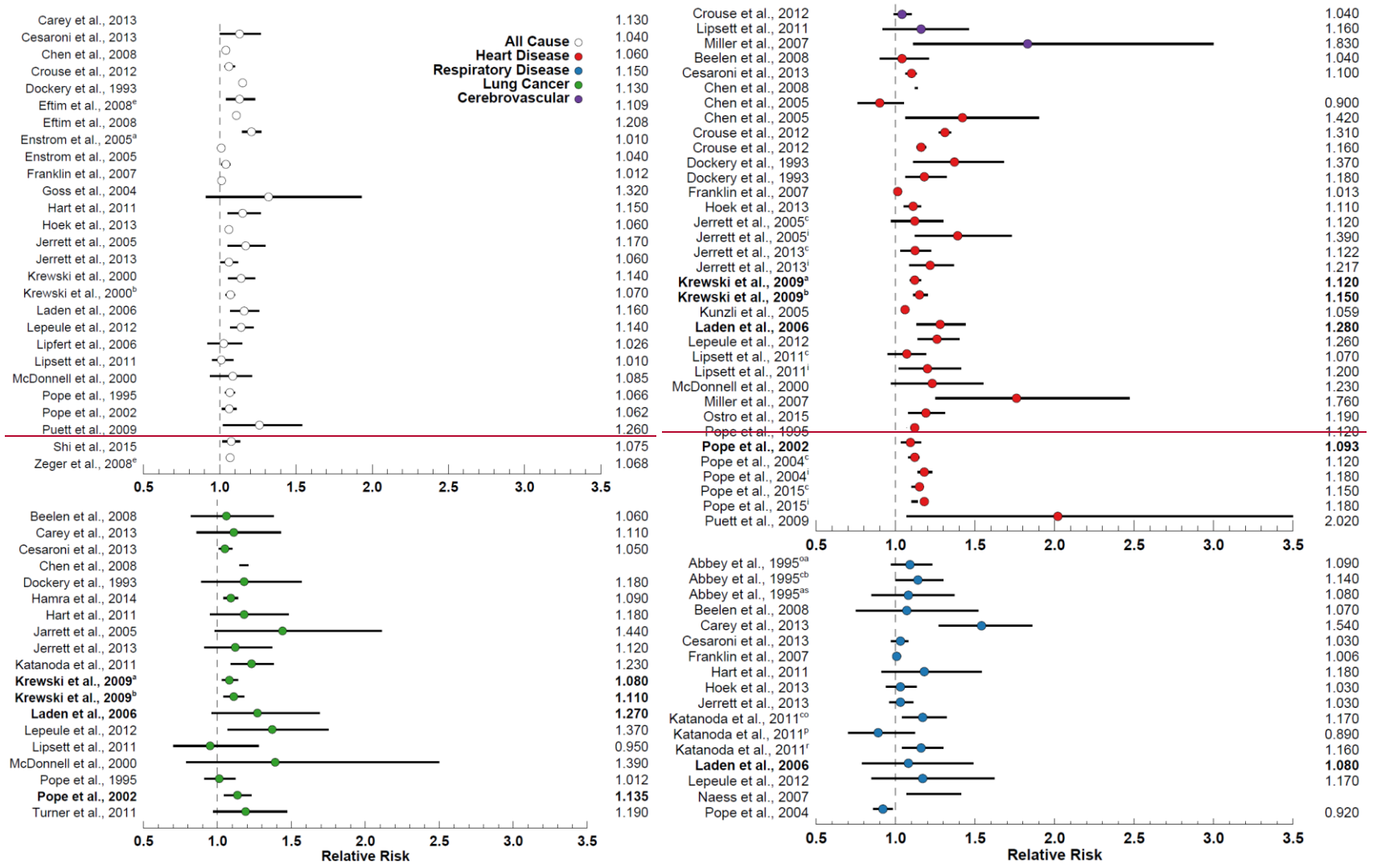
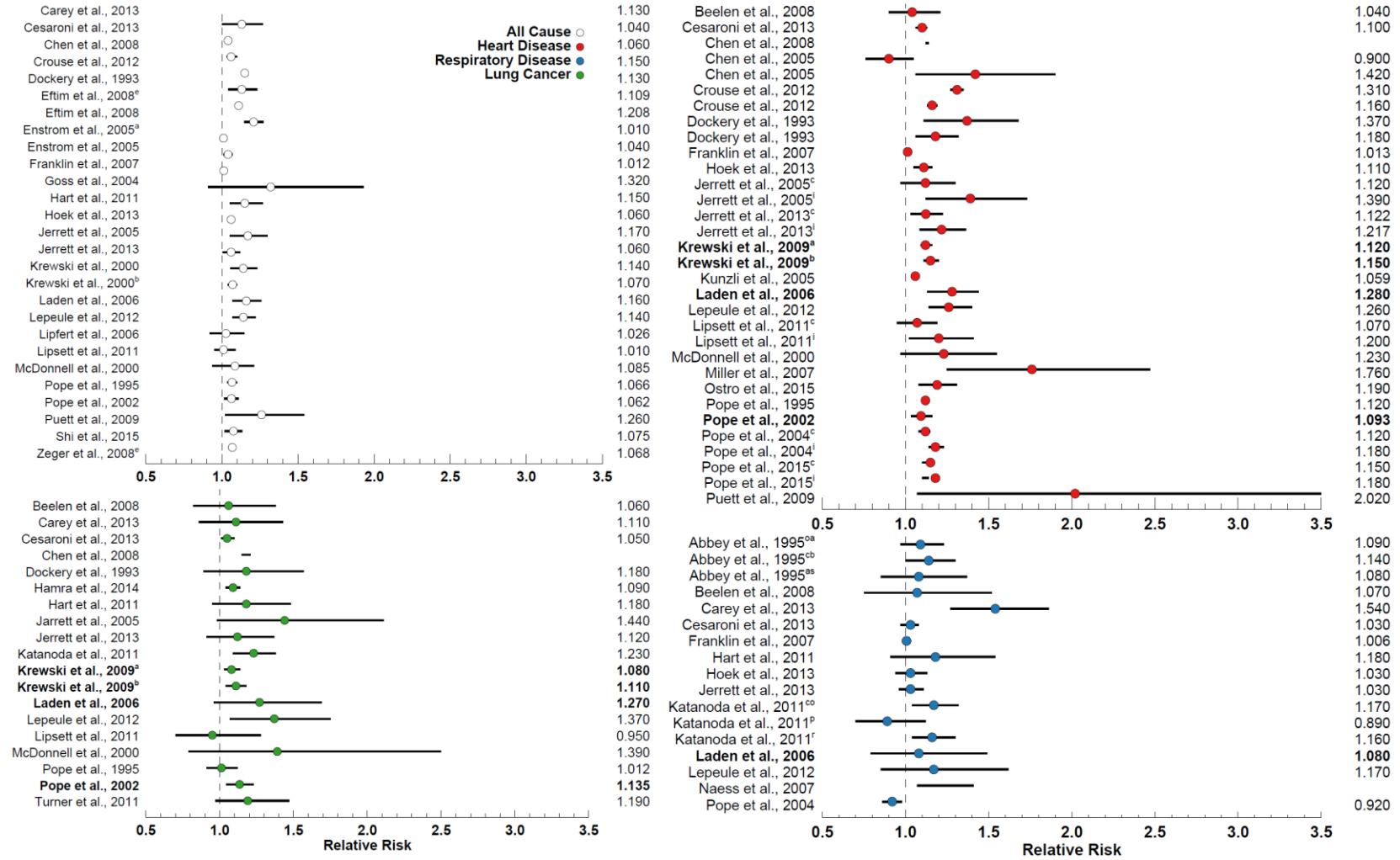



Figure 1. Population density [per $km^2$] for the year 2000 from the GPWv3 data for (a) the continental U.S. and (c) China. The projection for increase in population density by the year 2015 for (b) the continental U.S. and (d) China.


All Cause
Heart Disease
Respiratory Disease
Lung Cancer
Cerebrovascular
Carey et al., 2013 1.130
Cesaroni et al., 2013 1.040
Chen et al., 2008 1.060
Crouse et al., 2012 1.150
Dockery et al., 1993 1.130
Eftim et al., 2008[e] 1.109
Eftim et al., 2008 1.208
Enstrom et al., 2005[a] 1.010
Enstrom et al., 2005 1.040
Franklin et al., 2007 1.012
Goss et al., 2004 1.320
Hart et al., 2011 1.150
Hoek et al., 2013 1.060
Jerrett et al., 2005 1.170
Jerrett et al., 2013 1.060
Krewski et al., 2000 1.140
Krewski et al., 2000[b] 1.070
Laden et al., 2006 1.160
Lepeule et al., 2012 1.140
Lipfert et al., 2006 1.026
Lipsett et al., 2011 1.010
McDonnell et al., 2000 1.085
Pope et al., 1995 1.066
Pope et al., 2002 1.062
Puett et al., 2009 1.260
Shi et al., 2015 1.075
Zeger et al., 2008[e] 1.068
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Beelen et al., 2008 1.060
Carey et al., 2013 1.110
Cesaroni et al., 2013 1.050
Chen et al., 2008
Dockery et al., 1993 1.180
Hamra et al., 2014 1.090
Hart et al., 2011 1.180
Jarrett et al., 2005 1.440
Jerrett et al., 2013 1.120
Katanoda et al., 2011 1.230
Krewski et al., 2009[a] 1.080
Krewski et al., 2009[b] 1.110
Laden et al., 2006 1.270
Lepeule et al., 2012 1.370
Lipsett et al., 2011 0.950
McDonnell et al., 2000 1.390
Pope et al., 1995 1.012
Pope et al., 2002 1.135
Turner et al., 2011 1.190
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Relative Risk
Crouse et al., 2012 1.040
Lipsett et al., 2011 1.160
Miller et al., 2007 1.830
Beelen et al., 2008 1.040
Cesaroni et al., 2013 1.100
Chen et al., 2008
Chen et al., 2005 0.900
Chen et al., 2005 1.420
Crouse et al., 2012 1.310
Crouse et al., 2012 1.160
Dockery et al., 1993 1.370
Dockery et al., 1993 1.180
Franklin et al., 2007 1.013
Hoek et al., 2013 1.110
Jerrett et al., 2005[c] 1.120
Jerrett et al., 2005[i] 1.390
Jerrett et al., 2013[c] 1.122
Jerrett et al., 2013[i] 1.217
Krewski et al., 2009[a] 1.120
Krewski et al., 2009[b] 1.150
Kunzli et al., 2005 1.059
Laden et al., 2006 1.280
Lepeule et al., 2012 1.260
Lipsett et al., 2011[c] 1.070
Lipsett et al., 2011[i] 1.200
McDonnell et al., 2000 1.230
Miller et al., 2007 1.760
Ostro et al., 2015 1.190
Pope et al., 1995 1.120
Pope et al., 2002 1.093
Pope et al., 2004[c] 1.120
Pope et al., 2004[i] 1.180
Pope et al., 2015[c] 1.150
Pope et al., 2015[i] 1.180
Puett et al., 2009 2.020
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Abbey et al., 1995[oa] 1.090
Abbey et al., 1995[ob] 1.140
Abbey et al., 1995[as] 1.080
Beelen et al., 2008 1.070
Carey et al., 2013 1.540
Cesaroni et al., 2013 1.030
Franklin et al., 2007 1.006
Hart et al., 2011 1.180
Hoek et al., 2013 1.030
Jerrett et al., 2013 1.030
Katanoda et al., 2011[co] 1.170
Katanoda et al., 2011[p] 0.890
Katanoda et al., 2011[r] 1.160
Laden et al., 2006 1.080
Lepeule et al., 2012 1.170
Naess et al., 2007
Pope et al., 2004 0.920
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Relative Risk

1


All Cause
Heart Disease
Respiratory Disease
Lung Cancer
Carey et al., 2013 1.130
Cesaroni et al., 2013 1.040
Chen et al., 2008 1.060
Crouse et al., 2012 1.150
Dockery et al., 1993 1.130
Eftim et al., 2008[e] 1.109
Eftim et al., 2008 1.208
Enstrom et al., 2005[a] 1.010
Enstrom et al., 2005 1.040
Franklin et al., 2007 1.012
Goss et al., 2004 1.320
Hart et al., 2011 1.150
Hoek et al., 2013 1.060
Jerrett et al., 2005 1.170
Jerrett et al., 2013 1.060
Krewski et al., 2000 1.140
Krewski et al., 2000[b] 1.070
Laden et al., 2006 1.160
Lepeule et al., 2012 1.140
Lipfert et al., 2006 1.026
Lipsett et al., 2011 1.010
McDonnell et al., 2000 1.085
Pope et al., 1995 1.066
Pope et al., 2002 1.062
Puett et al., 2009 1.260
Shi et al., 2015 1.075
Zeger et al., 2008[e] 1.068
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Beelen et al., 2008 1.060
Carey et al., 2013 1.110
Cesaroni et al., 2013 1.050
Chen et al., 2008
Dockery et al., 1993 1.180
Hamra et al., 2014 1.090
Hart et al., 2011 1.180
Jarrett et al., 2005 1.440
Jerrett et al., 2013 1.120
Katanoda et al., 2011 1.230
Krewski et al., 2009[a] 1.080
Krewski et al., 2009[b] 1.110
Laden et al., 2006 1.270
Lepeule et al., 2012 1.370
Lipsett et al., 2011 0.950
McDonnell et al., 2000 1.390
Pope et al., 1995 1.012
Pope et al., 2002 1.135
Turner et al., 2011 1.190
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Relative Risk
Beelen et al., 2008 1.040
Cesaroni et al., 2013 1.100
Chen et al., 2008
Chen et al., 2005 0.900
Chen et al., 2005 1.420
Crouse et al., 2012 1.310
Crouse et al., 2012 1.160
Dockery et al., 1993 1.370
Dockery et al., 1993 1.180
Franklin et al., 2007 1.013
Hoek et al., 2013 1.110
Jerrett et al., 2005[c] 1.120
Jerrett et al., 2005[l] 1.390
Jerrett et al., 2013[c] 1.122
Jerrett et al., 2013[l] 1.217
Krewski et al., 2009[a] 1.120
Krewski et al., 2009[b] 1.150
Kunzli et al., 2005 1.059
Laden et al., 2006 1.280
Lepeule et al., 2012 1.260
Lipsett et al., 2011[c] 1.070
Lipsett et al., 2011[l] 1.200
McDonnell et al., 2000 1.230
Miller et al., 2007 1.760
Ostro et al., 2015 1.190
Pope et al., 1995 1.120
Pope et al., 2002 1.093
Pope et al., 2004[c] 1.120
Pope et al., 2004[l] 1.180
Pope et al., 2015[c] 1.150
Pope et al., 2015[l] 1.180
Puett et al., 2009 2.020
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Abbey et al., 1995[oa] 1.090
Abbey et al., 1995[cb] 1.140
Abbey et al., 1995[as] 1.080
Beelen et al., 2008 1.070
Carey et al., 2013 1.540
Cesaroni et al., 2013 1.030
Franklin et al., 2007 1.006
Hart et al., 2011 1.180
Hoek et al., 2013 1.030
Jerrett et al., 2013 1.030
Katanoda et al., 2011[co] 1.170
Katanoda et al., 2011[p] 0.890
Katanoda et al., 2011[r] 1.160
Laden et al., 2006 1.080
Lepeule et al., 2012 1.170
Naess et al., 2007
Pope et al., 2004 0.920
0.5 1.0 1.5 2.0 2.5 3.0 3.5
Relative Risk

1

2

Figure 2. Relative risk ratios from select previous studies for mortality due to chronic exposure to $PM_{2.5}$ (given as per $10\mu gm^{-3}$ increase) colored by cause of death. Studies applied in this work are highlighted in bold.

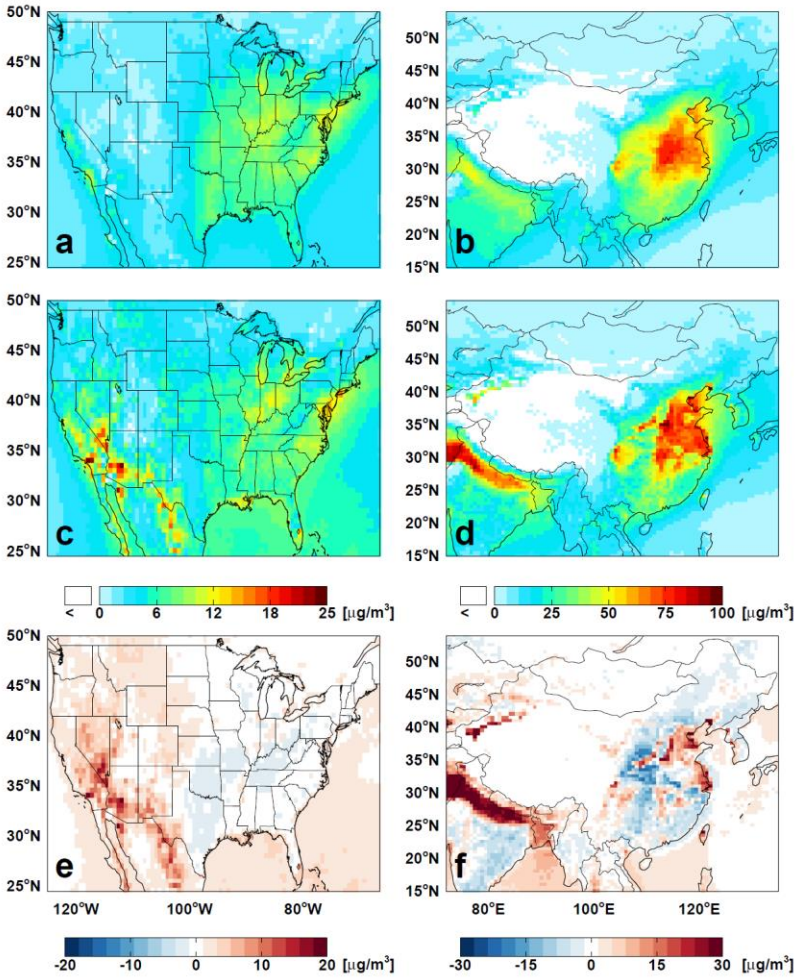



Figure 3. Long-term average (2004-2011) unconstrained model simulation of $PM_{2.5}$ for the (a) continental U.S. and (b) China, along with the (MODIS-Aqua Collection 6) satellite-based $PM_{2.5}$ for the (c) continental U.S. and (d) China, and the difference between the satellite-constrained and unconstrained model $PM_{2.5}$ concentrations.

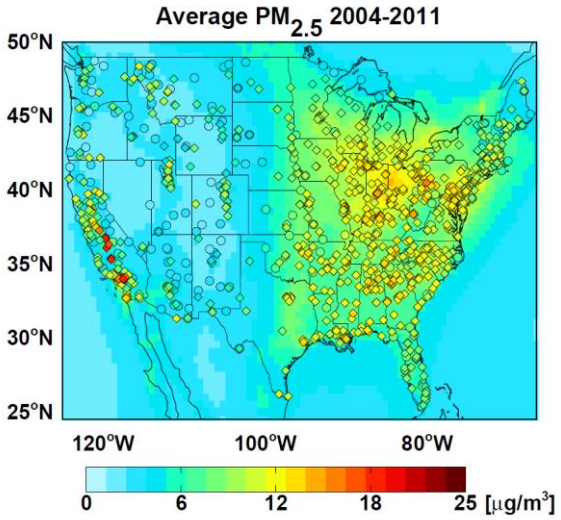



Figure 4. GEOS-Chem simulated average surface $PM_{2.5}$ mass for years 2004-2011 overlaid with measurements at IMPROVE (circles) and AQS sites (diamonds).

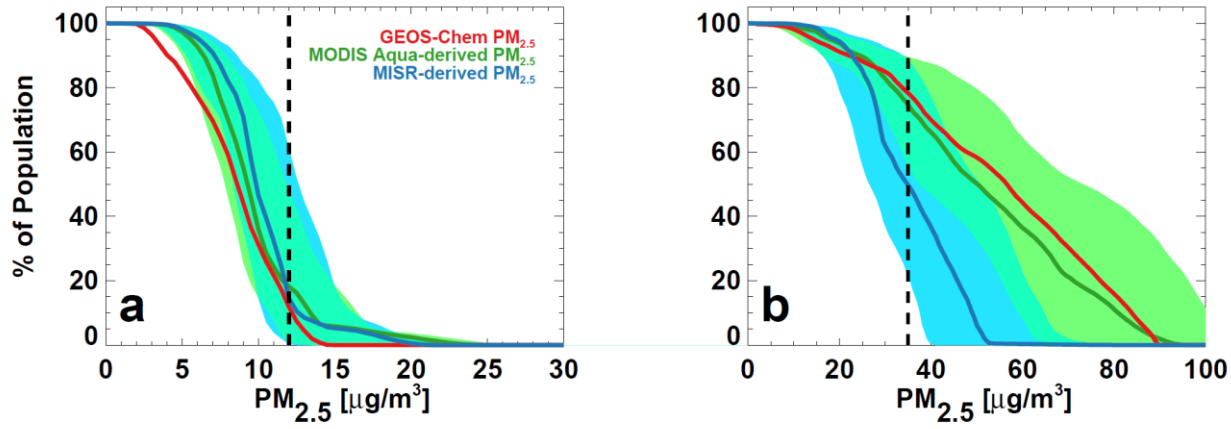



Figure 5. Percent of the population exposed to different annual $PM_{2.5}$ concentrations in the U.S. (a) and China (b). Lines denote estimates using the unconstrained GEOS-Chem simulation (red) or using satellite-based estimates with MODIS (green) and MISR (blue). Shading represents potential uncertainty associated with the model η (described in Section 4.2) and dashed black lines represent national annual air quality standards.

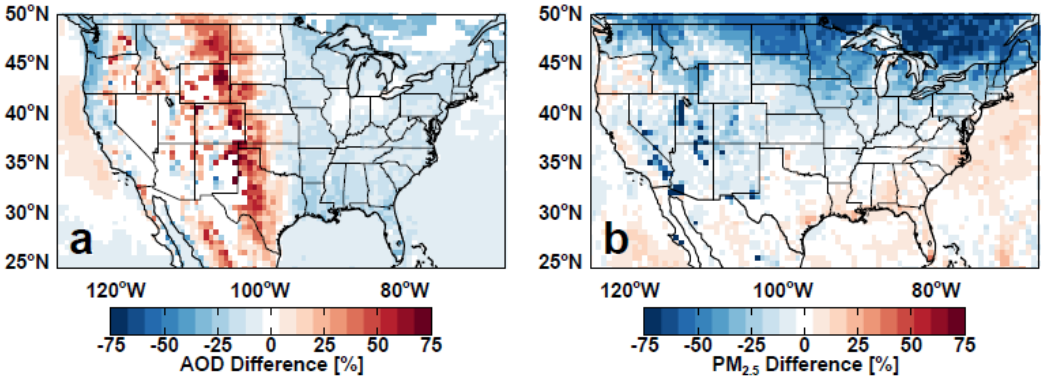



Figure 6. (a) Percent difference between annual mean AOD from MODIS Collection 6 and Collection 5 and (b) simulated bias in satellite-derived annual average surface $PM_{2.5}$ associated with satellite sampling.

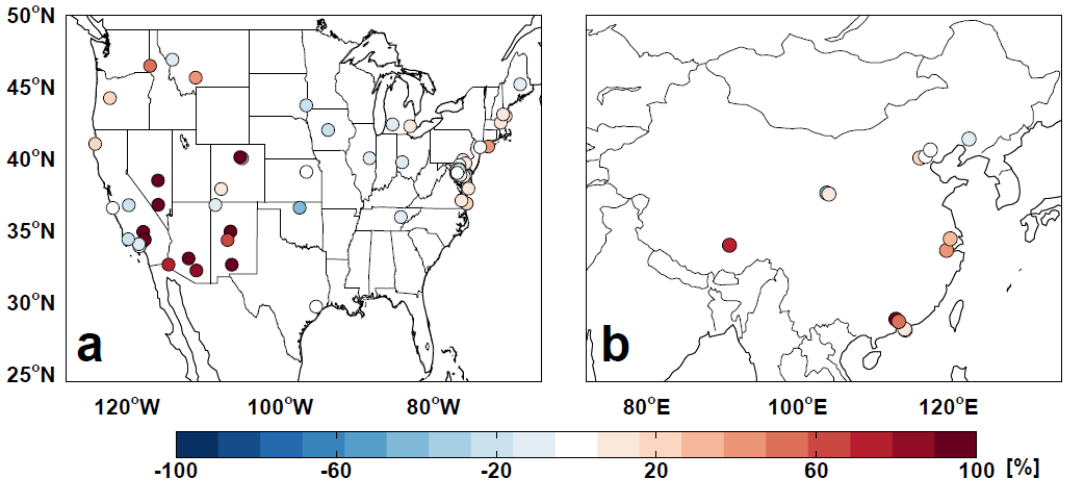



Figure 7. Normalized mean bias in AOD between MODIS-Aqua Collection 6 and AERONET sites for (a) the U.S. and (b) China.

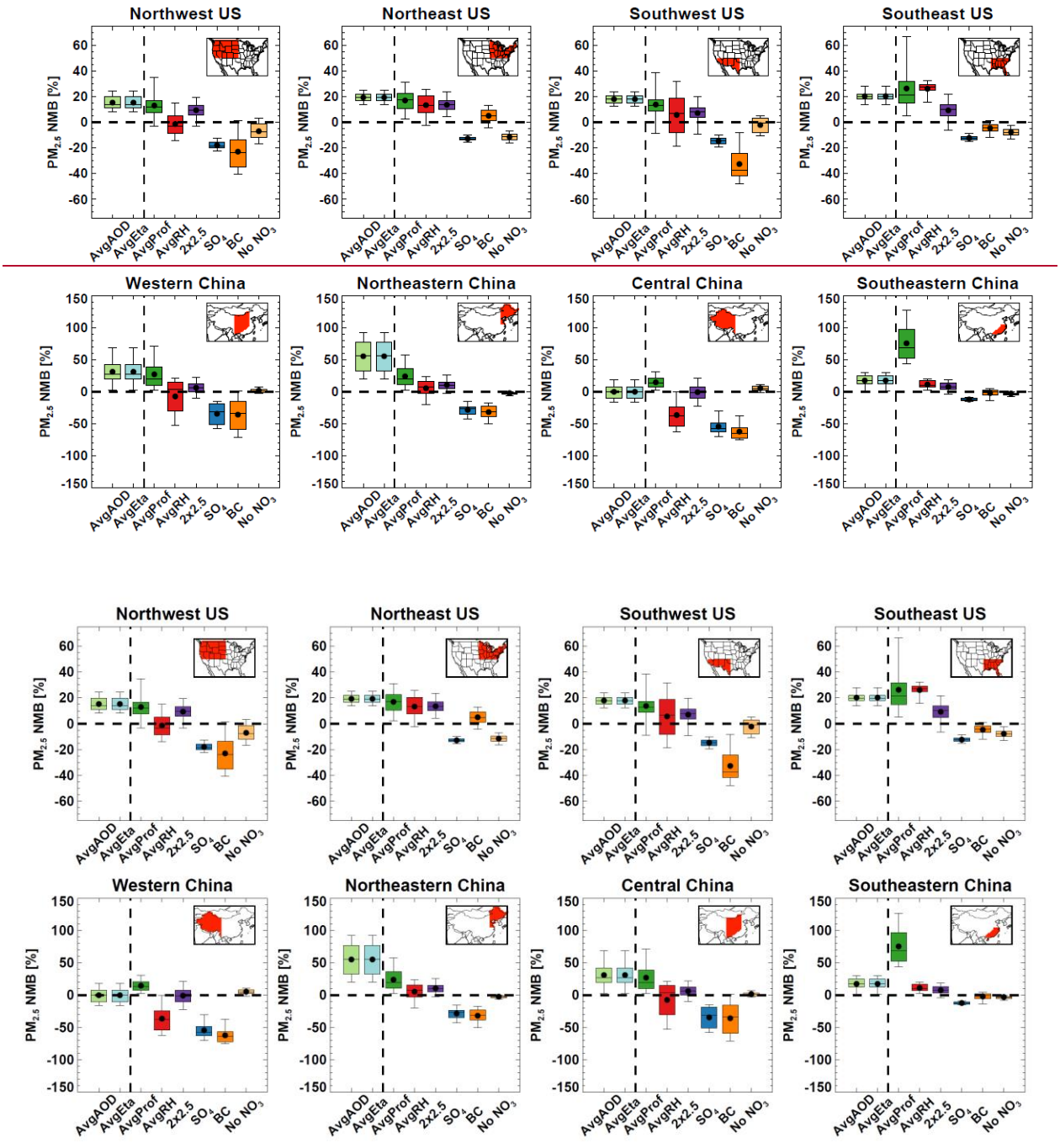
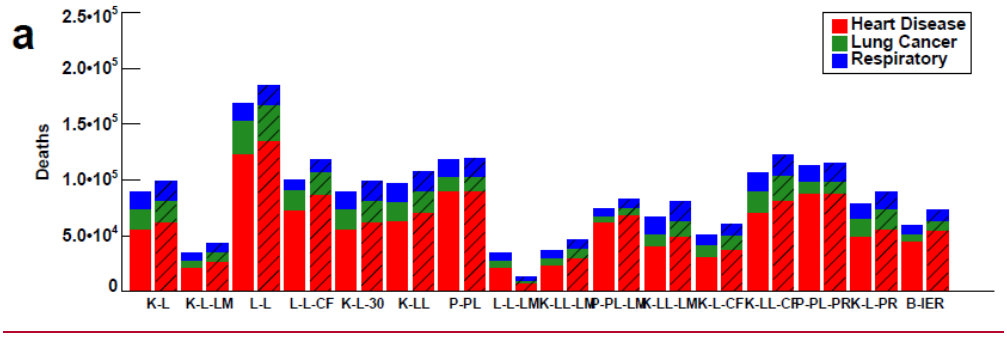
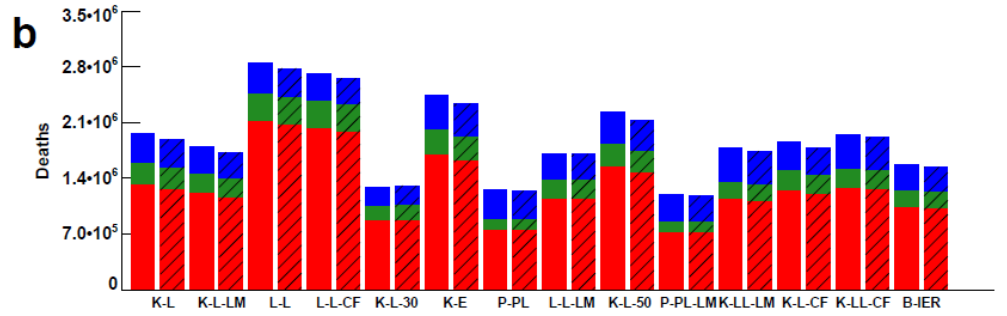


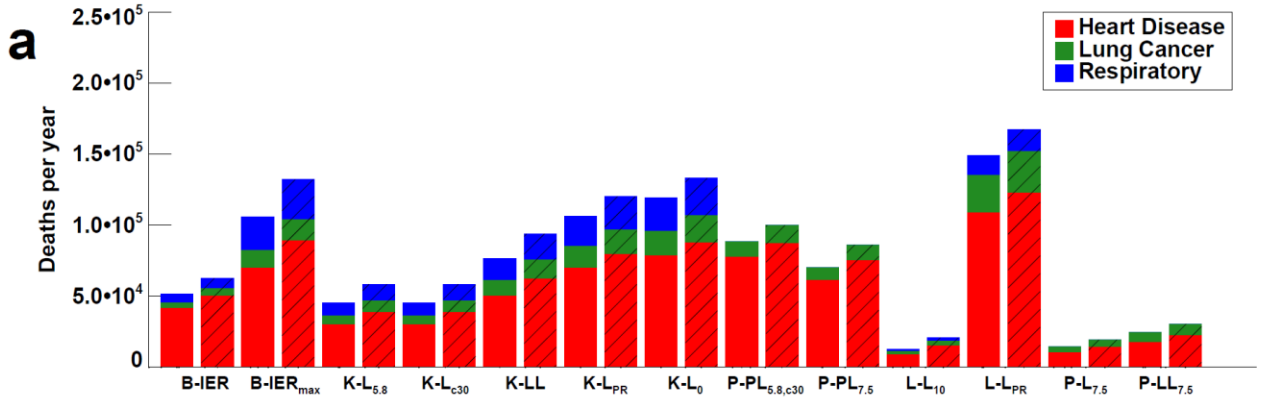

Figure ~~7~~8. Distribution of normalized mean biases in annual average $PM_{2.5}$ for grid boxes in different regions of the U.S. (top row) and China (bottom row) determined from sensitivity tests to investigate the uncertainty in η. Sensitivity tests are described (and abbreviations defined) in Table 3.

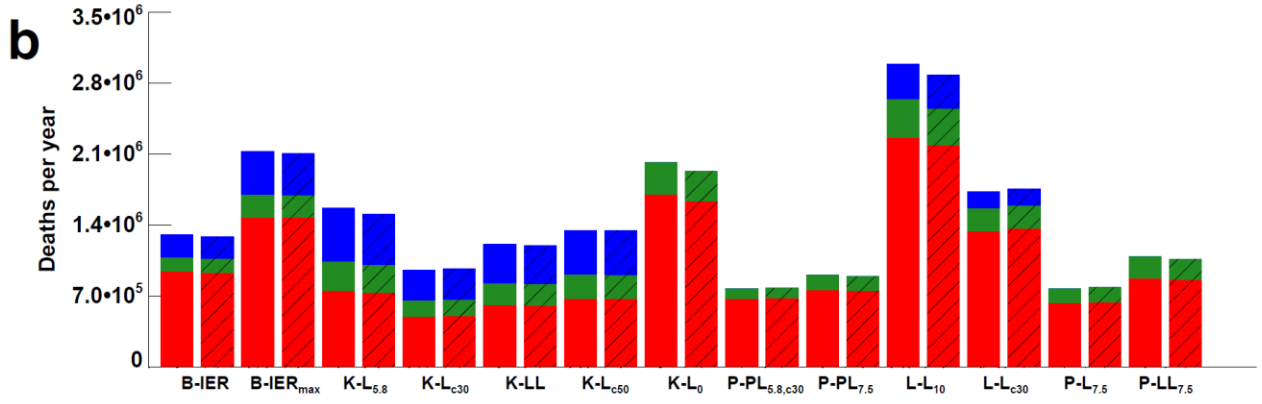
1


a
Heart Disease
Lung Cancer
Respiratory
Deaths
2.5•10^5
2.0•10^5
1.5•10^5
1.0•10^5
5.0•10^4
0
K-L
K-L-LM
L-L
L-L-CF
K-L-30
K-LL
P-PL
L-L-LM
K-LL-LM
P-PL-LM
K-LL-LM
K-L-CF
K-LL-CF
P-PL-PR
K-L-PR
B-IER


b
Deaths
3.5•10^6
2.8•10^6
2.1•10^6
1.4•10^6
7.0•10^5
0
K-L
K-L-LM
L-L
L-L-CF
K-L-30
K-E
P-PL
L-L-LM
K-L-50
P-PL-LM
K-LL-LM
K-L-CF
K-LL-CF
B-IER

2


a
Heart Disease
Lung Cancer
Respiratory
Deaths per year
2.5•10^5
2.0•10^5
1.5•10^5
1.0•10^5
5.0•10^4
0
B-IER
B-IER_max
K-L_5.8
K-L_c30
K-LL
K-L_PR
K-L_0
P-PL_5.8,c30
P-PL_7.5
L-L_10
L-L_PR
P-L_7.5
P-LL_7.5


b
Deaths per year
3.5•10^6
2.8•10^6
2.1•10^6
1.4•10^6
7.0•10^5
0
B-IER
B-IER_max
K-L_5.8
K-L_c30
K-LL
K-L_c50
K-L_0
P-PL_5.8,c30
P-PL_7.5
L-L_10
L-L_c30
P-L_7.5
P-LL_7.5

3

Figure ~~8~~9. Premature mortality estimates for (a) the U.S. and (b) China determined using different RR, CRFs, and threshold/ceiling values, as described in Table 3. Colors represent cause of death estimated using $PM_{2.5}$ concentrations from unconstrained model simulations (solid) and satellite-based estimates (hatched).

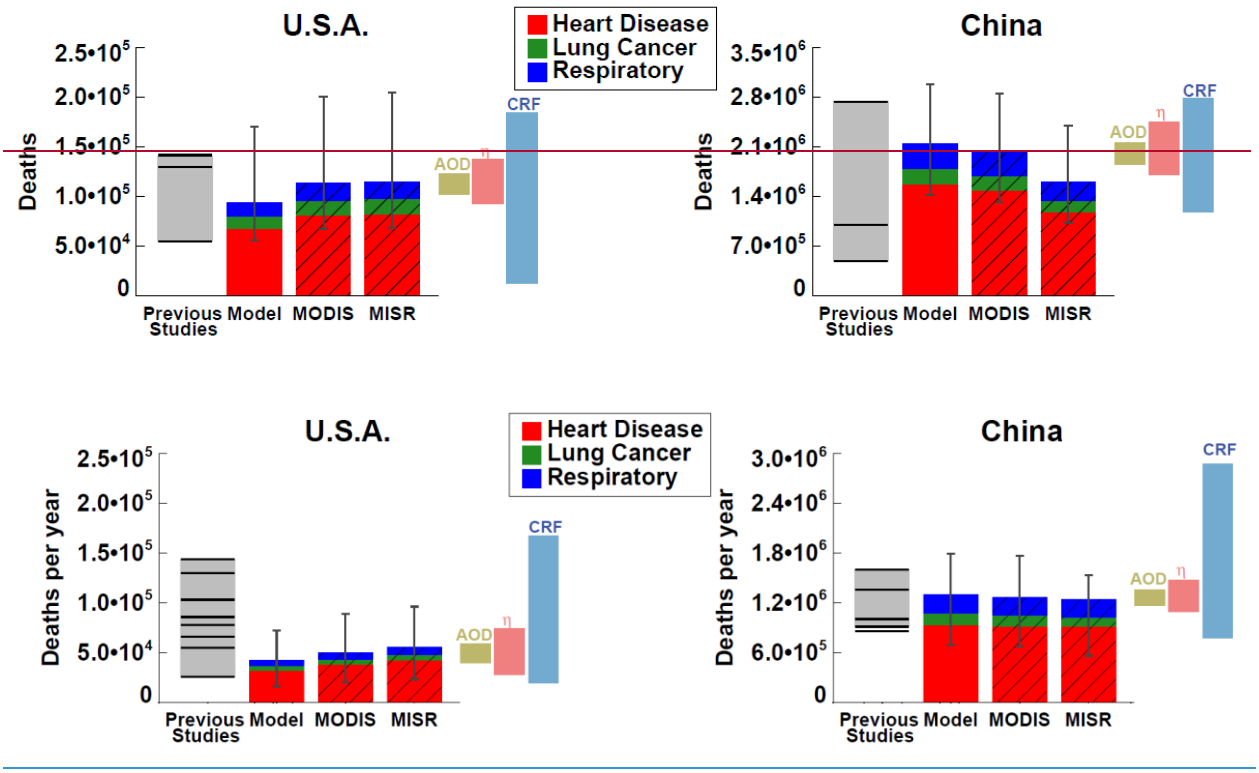



Figure ~~9~~10. Burden of mortality due to outdoor exposure to fine particulate matter as determined in previous studies (Table 1, gray bars with values from individual studies designated by black lines), calculated using model (GEOS-Chem, solid) and satellite-based (hatched) annual concentrations (colored by disease, whiskers denote 5th and 95th percentile estimates generated using the Burnett et al., 2014 coefficients~~95% confidence intervals on RRs~~). The uncertainty range on the MODIS-based estimates due to satellite AOD (taupe), model η (coral), and CRF (blue) are shown on the right.