Note:
This file includes the responses to two referees.
Response to referee #1: Page 1-13.
Response to referee #2: Page 14-25.

Response to referee #1

We would like to thank the referee #1 for his/her thoughtful and constructive comments that have helped us to improve the quality of this manuscript. Below are our responses (in normal text) to the comments from referee #1 (**in bold**):

General comments:
Pan et al., 2014 discuss the multi-model evaluation of aerosol distributions over the South Asian region. The focus is on understanding the common problems in model-simulated aerosol properties and possible causes of underestimation of model-simulated aerosol properties. Even though model underestimation of aerosol properties over South Asia are previously reported in regional-scale analysis (e.g. Reddy etal., 2004; Ganguly et al., 2009), the multi-model evaluation of aerosol distributions over South Asia could be useful for the scientific community.

The paper is generally well written and the possible causes of underestimation of AOD such as relative humidity and emission amount are quite interesting. The multi-model simulated aerosol properties are evaluated using different observation data sets (e.g. ISRO-GBP, ICARB and satellites).

Dust aerosol underestimation is previously reported as one of the possible causes in AOD underestimation especially in pre-monsoon season over South Asia. However, the inter-model differences and pre-monsoon underestimations are not well described in the manuscript. These points need to be addressed in the manuscript in context with the existing literature. The following comments should be addressed before the manuscript would be satisfactory for publication in ACP

Yes, dust is the dominant aerosol during the pre-monsoon season over South Asia. We add more information about dust, please refer to the reply to the Specific comments 2) below.

Specific comments:
1) **In Section 4.1 and 4.2, large diversity among model-simulated AOD is visible. The possible causes of inter-model differences over IGP region are not clear from the manuscript. Varying wet /dry deposition rates and emission fluxes do cause significant variations in a single model, but these uncertainties do not explain most of the inter-model differences. Textor et al. [2007] also found that inter-model differences were only partially explained by differences in emission inventories. Bond et al., [2013] pointed out large differences in modelled horizontal and vertical transport are largely responsible for the inter-model diversity for BC distributions. It could be useful if authors highlight the most significant parameter in the model need to be focused for improving the aerosol**

distributions over South Asia instead of one general sentence that mentioned the manuscript (Page 19119, lines 15-19).

Thanks for the suggestion. We have added a new section in the manuscript to address the inter-model diversity issue with a new table 3 below.

Table 3. The statistics of the aerosol parameters over South Asia (61.5°-90.0°E, 5.0°-36.0°N)

Parameter	Unit	#	Mean	Median	Min	Max	Stdev	Diversity [a]
				SO_4				
Emi [b]	$Tg(SO_2)/a$	7	8.42	8.49	6.93	9.79	0.84	10%
Cheaq[c]	$Tg(SO_4)/a$	4	0.36	0.39	0.19	0.48	0.13	36%
Cheg[d]	$Tg(SO_4)/a$	4	0.33	0.23	0.16	0.71	0.26	**77%** [h]
Wet	$Tg(SO_4)/a$	7	6.47	5.97	5.38	8.58	1.21	19%
Dry	$Tg(SO_4)/a$	7	1.02	0.92	0.32	1.48	0.43	**42%**
Dry/Dry+Wet	%	7	17	17	6	22	6	35%
Life time	days	7	5.17	4.64	3.71	9.27	1.90	37%
Load	$Tg(SO_4)$	7	0.09	0.08	0.05	0.15	0.03	34%
MEE[e]	$m^2/g(SO_4)$	4	8.81	9.16	5.53	11.39	2.43	28%
AOD	Unitless	4	0.07	0.07	0.04	0.08	0.02	27%
				BC				
Emi	Tg/a	7	0.68	0.68	0.62	0.71	0.04	5%
Wet	Tg/a	7	0.36	0.36	0.29	0.42	0.04	12%
Dry	Tg/a	7	0.17	0.20	0.06	0.22	0.06	35%
Dry/Dry+Wet	%	7	33	37	15	41	10	29%
Life time	days	7	7.98	6.89	4.48	14.35	3.31	**42%**
Load	Tg	7	0.01	0.01	0.007	0.020	0.004	38%
MEE	m^2/g	4	7.16	7.62	2.77	10.60	3.72	**52%**
AOD	Unitless	4	0.008	0.009	0.003	0.01	0.003	44%
				OA				
Emi[f]	Tg/a	7	3.19	3.12	2.05	4.07	0.64	20%
Load	Tg	7	0.05	0.04	0.03	0.07	0.01	27%
Wet	Tg/a	7	2.17	1.98	1.50	3.26	0.56	26%
Dry	Tg/a	7	0.80	0.86	0.30	1.37	0.36	**45%**
Dry/Dry+Wet	%	7	29	32	15	38	9	**31%**
Life time	days	7	6.20	5.95	4.56	9.20	1.60	26%
MEE	m^2/g	4	5.33	5.28	3.61	7.14	1.47	28%
AOD	Unitless	4	0.020	0.019	0.016	0.024	0.004	18%
				DUST				
Emi	Tg/a	7	96.34	39.21	6.42	356.46	125.33	**130%**
Load	Tg	7	1.28	1.25	0.25	2.51	0.68	53%
Wet	Tg/a	7	72.78	67.62	20.58	171.48	47.29	65%
Dry + Sed[g]	Tg/a	7	100.84	45.97	1.72	330.85	115.95	**115%**
Dry/Dry+Wet	%	7	53	60	11	78	25	46%
Life time	days	7	4.34	4.25	1.48	8.34	2.25	52%
MEE	m^2/g	4	0.67	0.60	0.54	0.92	0.18	27%
AOD	Unitless	4	0.09	0.09	0.06	0.14	0.04	41%

[a]. The diversity is defined as the ratio of mean and standard deviation (i.e. mean/stdev). The largest and second largest diversities in each species are highlighted in bold.
[b]. The emission of so_2, including anthropogenic and biomass burning emission.
[c]. The chemical production of SO_4 in aqueous phase reaction (i.e. SO_2 reacts with H_2O_2).
[d]. The chemical production of SO_4 in gaseous phase reaction (i.e. SO_2 reacts with OH).
[e]. Mass extinction efficiency, defined as the ratio of AOD and load (i.e. AOD/load).

[f]. Sum of anthropogenic emission, biomass burning emissions and secondary organic aerosol.
[g]. Dry deposition plus sedimentation.
[h]. The top two largest diversities in each species are highlighted in bold.

We found (1) for aerosols with dominant anthropogenic origin (i.e. BC, OA and SO_4), the largest diversity among models occurs in the treatment of dry deposition, with diversities of dry deposition amount ranging 35-45% across three species. The diversity of wet deposition is relatively smaller with a range of 12-26%. (2) The chemical production of sulfate in gaseous phase among models (4 models) also has large diversity (about 77%); (3) BC has the largest diversity of mass extinction efficiency (i.e. MEE) among models compared to other species, with a diversity of 52% compared to 27-28% for other species.

2) **In Section 5, different possible cause for AOD and AAOD underestimation is described. The underestimation of natural aerosols (e.g. dust) emission flux may also lead to error in total aerosol distributions. Previous studies reported that dust emission flux underestimation can lead to underestimation of model-simulated AOD over South Asia/IGP during pre-monsoon (MAM) season (e.g. Cherian et al., 2012). Few information are reported in the manuscript Section 2.2 (Page 19103, lines 1-5). In Fig.5, the pre-monsoon season (MAM) AOD is not well captured by most of models over Kanpur. The spatial distribution of AOD is also partially captured by all models during this season (Section 4.3). This could be due to missing dust transport to Kanpur from dust source regions. It could be useful if authors provide more information about the inter-model.**

Yes, the underestimation of dust emission flux may also lead to error in total aerosol distribution in models during pre-monsoon season, in particular the model ECH. We have added information of dust size distribution (highlighted in gray) in the existing Table 1 (see below). In addition, we examine the dust budgets of the monthly emission, dry and wet depositions, load, lifetime, MEE and AOD (new Table 3). The dust emission itself has very large diversity among the models (about 130%), and the diversity of dry (and settling) deposition is as large as 115%. The difference in treatment of dust size bin in models significantly contributes to these diversities among models (see Table 1).

Table 1. General information of multi-models.

Model	HadGEM2	GOCART-v4	ECHAM5-HAMMOZ	GISS-modelE	GISS-MATRIX	SPRINTARS	GEOS5-GOCART
ID	HAD	GOC	ECH	GIE	GIM	SPR	GE5
Time range	2000-2006	2000-2007	2000-2005	2000-2008	2000-2007	2000-2008	2000-2008
Res. [a]	1.8× 1.2×38	2.5× 2×30	2.8× 2.8×31	2.5× 2×40	2.5× 2×40	1.1× 1.1×56	2.5× 2×72
Anthrop. Emi. [b]	A2-MAP	A2-MAP	A2-MAP	A2-ACCMIP	A2-ACCMIP	A2-ACCMIP	A2-ACCMIP
BB Emi. [c]	GFED2	GFED2	GFED2	GFED2	GFED2	GFED2	GFED2
Met. Field	ERA- Interim	GEOS-DAS	ECMWF analysis	NCEP wind	NCEP-wind	NCEP/ NCAR	MERRA
Refractive index 550nm SO_4^{2-}	1.53 – 1e-7*i*	1.43-1e-8 *i*	1.43-1e-8*i* [e]	1.528-1e-7*i*	1.528-1e-7*i*	1.43-1e-8*i*	1.43-1e-8*i*
BC	1.75 – 0.44*i* (FF) [d]	1.75-0.44*i*	1.85-0.71*i*	1.85-0.71*i*	1.85-0.71*i*	1.75-0.44*i*	1.75-0.44*i*
OA	1.54 – 0.006*i* (FF)	1.53-0.006*i*	1.53-0.0055*i*	1.527-0.014*i*	1.527-0.014*i*	1.53-0.006*i*	1.53-0.006*i*
Dust	1.52 – 0.0015*i*	1.53-0.0055*i*	1.517-0.0011*i*	1.564-0.002*i*	1.564-0.002*i*	1.53-0.002*i*	1.53-0.008i
SS	1.55 – 1e-7*i* Aged BB: 1.54 – 0.018*i*	1.50-1e-8 *i*	1.49-1e-8*i*	1.45-0.*i*	1.45-0.*i*	1.38-4.26e-9*i*	1.50- 1e-8*i*
Additional Species [f]	NO_3	-	-	NO_3	NO_3	-	-
Dust Size distribution (μm) [g]	6 bins 0.0316-0.1-0.316-1.0-3.16-10-31.6	8 bins 0.1-0.18-0.3-0.6-1.0-1.8-3.0-6.0-10.0	Accum. mode: 0.05<r_m<0.5 coarse mode: r_m>0.5	5 bins 0.1-1-2-4-8-16	4 bins 0-1-2-4-8	6 bins 0.1-0.22-0.46-1.0-2.15-4.64-10.0	8 bins 0.1-0.18-0.3-0.6-1.0-1.8-3.0-6.0-10.0

[a] Spatial resolutions (°latitude × °longitude × number of vertical levels).
[b] Anthropogenic emission data are from either A2-ACCMIP or A2-MAP (refer to Diehl et al. 2012).
[c] Biomass burning emission data (refer to Diehl et al. 2012).
[d] FF is fossil fuel and BB is biomass burning.
[e] As for EHCAM5-HAMMOZ model with a mixed aerosol scheme, the refractive index for each of the 7 modes is calculated as the volume weighted average of the refractive indices of the components of the mode, including the diagnosed aerosol water.
[f] Additional aerosols besides commonly included aerosol species, i.e. SO_4^{2-} (sulfate), Dust, SS (sea salt), BC (black carbon), and OA(organic aerosol). Here NO_3^- is nitrate.
[g] Listed are the edges of bin size range in all models except for ECH in which r_m is modal radii.

3) **The information about aerosol refractive index for each species used in the different models is missing from the paper. For example, Black Carbon and Dust AAODs are strongly depending on refractive indices used in the model. The refractive index information could be useful for understanding the inter-model differences in AAODs. It could also useful for providing the future model improvements of AAOD distributions over South Asia**

We have added information of aerosol refractive index at 550nm for each species in the existing Table 1 (highlighted in yellow). The real part of refractive indices of each aerosol at 550nm is similar among these seven models, and they are based on Optical Properties of Aerosols and Clouds (OPAC) (Hess et al., 1998) with slight modifications in some models. However, the imaginary parts are varying among models even by a factor of 10 for species SO_4 and SS. In particular for BC, the most absorbing aerosol, the imaginary part of refractive index (representing light absorption) is 0.44 in four models and 0.71 in three models. This large diversity of BC refractive indices among models is reflected in BC mass extinction efficiency (i.e. MEE) as shown in the new table 3 with a diversity of 52% compared to 27-28% for other species. Bond and Bergstrom (2006) recommended a higher value of 0.79 for the imaginary part of BC refractive index at 550nm, based on agreement between measured real and imaginary parts of the refractive index of light absorbing carbon. Therefore, we suggest implementing the BC refractive indices from Bond and Bergstrom (2006). For dust, the light absorption is less than that of BC. The imaginary of dust refractive index ranges from 0.0011 (ECH) to 0.008 (GE5).

Technical comments:

4) **Page 19099, line 8: aboved?**

Corrected. Modifying "aboved" to "above".

5) **Page 19110, line15: It is interesting to see nitrate rather than dust dominate AOD over northern India in the Had Gem model. Whether this is the reason for satellite observed AOD well simulated during winter by this model over South Asia?**

Yes, in the HadGEM2 model, the nitrate aerosol AOD contributes significantly to the feature of high AOD along IGP during winter (DJF), see the Fig. S1. Wintertime surface observations also showed that the surface nitrate concentrations are comparable to the sulfate (Fig.10), which was not represented by most of the models. Therefore, we think that it is important to have nitrate in a model to reproduce the satellite retrievals over South Asia. However, HadGEM2 likely overestimates nitrate in April and October, and underestimates dust in pre-monsoon and monsoon season (March-July).

6) **Page 19112, Section 4.5: Moorthy et al., [2013] pointed out that improvement in the atmospheric boundary layer (ABL) parametrization in GOCART model over tropical region might improve the model-simulated BC distributions. How the authors comment about this?**

The PBL is an important factor to determine the surface concentration of aerosols including BC, besides the factor of strength of emission sources. In winter, the averaged ABL is 400-500 meters in the model GOCART (similar meteorological data used by GEOS5), about the double of the observed ABL, thus a better-constrained ABL in GOCART and GEOS5 could be helpful. Unfortunately we don't have ABL information from other models, so it is hard to address this point in general in the previous ACPD manuscript.

In addition, we would like to add that the column AAOD during wintertime is underestimated as well, although in a less degree than surface concentration (by a factor of 3 verse 10). Considering the fact that the aerosol is confined to near surface due to the low ABL in winter, the underestimation of both surface concentration and AAOD together indicates a fundamental problem – winter time BC emissions might be underestimated in these models. The biofuel emission, the major emission source of BC, is supposed to be higher in the winter due to a higher demand of heating. A constant anthropogenic emission amount throughout of a year, however, is used by the model in this study, which makes the winter underestimation worse.

7) **Page 19113, line 24: It is not clear what "low bias" means? Rewrite the sentence.**

Corrected. We have changed "despite the low bias of BC concentration" to "despite the underestimation of BC surface concentration".

8) **Page 19114, line 7: Only BC surface concentrations are severely underestimated over IGP? All the models underestimate sulphate by 5-50% (Page 19115, line 15).**

"BC surface concentrations are severely underestimated over IGP". This sentence is a short summary based on the preceding analysis in section 4. The referee is right that the models also underestimate sulfate by 5-50%. But this point is discussed later in the following sections, and thus it is inappropriate to address this in the BC section.

9) **Page 19114, line 17: Correct the sentence.**

Rewrite the sentence of "Figure 10 shows the comparisons between seven models and in-situ measurements from the ISRO-GBP land campaign at four stations in the IGP for December 2004." to

" Figure 10 shows the comparisons among seven models and between models and in-situ measurements (ISRO-GBP land campaign) for December 2004 at four stations which are located in the IGP region."

10) **Page 19115, line 15: What is the reason for sulphate underestimation? Is it related to the sulphate chemistry scheme used in each models?**

Thanks you for bringing up this question. Yes, the underestimation of sulfate surface concentration might be caused by not accounting for aqueous phase oxidation in the models. Observations show that foggy weather is very common in IGP during winter, which favors the formation of sulfate in aqueous phase. However, the much lower relative humidity than observation (Fig. 10) thus the drier condition in models (except for the HAD) would inhibit this reaction. Sulfate concentration, however, is found low in HAD as well although with high relative humidity. There could be additional reasons, such as low concentration of oxidants (H_2O_2 and OH). Unfortunately, it is difficult to quantify here because the information of chemical productions of sulfate of this model was not available in the AeroCom database.

11) **In the Conclusion section, point 2 (Page 19120): BC concentrations are better captured by models over Kharagpur. This should be mentioned in this section.**

Thanks for pointing out this, we have added the sentence of "BC concentrations are better captured by models over Kharagpur, where the BC emissions are mostly from the burning of coals in power plants."

12) **In the Conclusion section, point 4 (Page 19121): Better represent nitrate in the models is not clear. Whether Nitrate emissions or chemistry scheme used in the models?**

Clarify as this "However, NO_3^- is either not considered in 4 out of 7 models or significantly lower than observations in other 2 models, suggesting a need to add the NO_3^- aerosol component or improve the chemistry scheme in these models."

13) **Figures 4, 5 and 6: Quality is poor. Lines and colours are not clearly visible.**

All three figures are modified according to referee's request. All modified figures are listed in the end of this file.

14) **Figure 7a-d: Its very difficult to compare spatial distribution of AOD in different seasons. Better provide mutli-model mean and deviation against observations and move these figures into supplementary information.**

Considering these figures are only figures in this manuscript to demonstrate the unique spatial distribution of AOD (various locations of maxima AOD in different seasons), we prefer to keep these figures in the main text. In order to make the comparison of different seasons easier, we have rearranged the figure 7 with grouping all three satellites into Figure 7a and all seven models into Figure 7b. In this way, four seasons from one model/satellite are shown in one figure.

Reference:

Bond, T. C., and Bergstrom, R. W.: Light absorption by carbonaceous particles: An investigative review, Aerosol Sci. Technol., 40(1), 27–67, doi:10.1080/02786820500421521, 2006.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., K¨archer, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, J. Geophys. Res., 118, 5380–5552, doi:10.1002/jgrd.50171, 2013.

Cherian, R., Venkataraman, C., Ramachandran, S., Quaas, J., and Kedia, S.: Examination of aerosol distributions and radiative effects over the Bay of Bengal and the Arabian Sea region during ICARB using satellite data and a general circulation model, Atmos. Chem. Phys., 12, 1287-1305, doi:10.5194/acp-12-1287-2012, 2012.

Hess, M., K¨opke, P., and Schult, I.: Optical properties of aerosols and clouds: The software package OPAC, B. Am. Meteorol. Assoc., 79, 831–844, doi: http://dx.doi.org/10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2, 1998.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Feichter, J., Fillmore, D., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I. S. A., Iversen, T., Kloster, S., Koch, D., Kirkevåg, A., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J. E., Pitari, G., Reddy, M. S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: The effect of harmonized emissions on aerosol properties in global models – an AeroCom experiment, Atmos. Chem. Phys., 7, 4489-4501, doi:10.5194/acp-7-4489-2007, 2007.

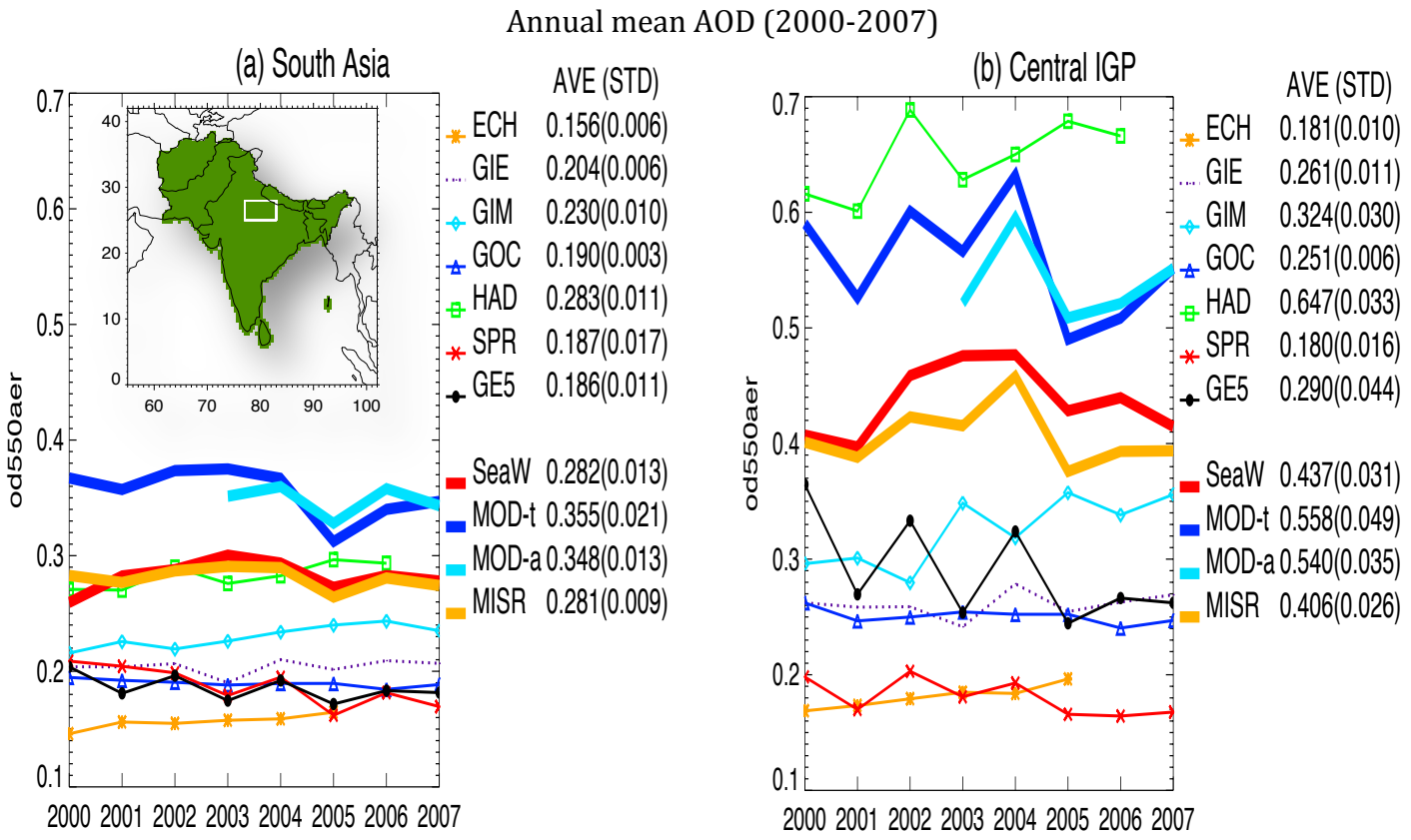

Figure 4. The annual averaged mean AOD for 2000-2007 over (a) South Asia (the green area in the map); (b) Central IGP (77°-83°E; 25°-28°N, the white box in that map). The thin curves with symbols represent seven models, and the thick curves represent four NASA remote sensors, with corresponding multi-year averaged annual mean AOD and the standard deviation followed.

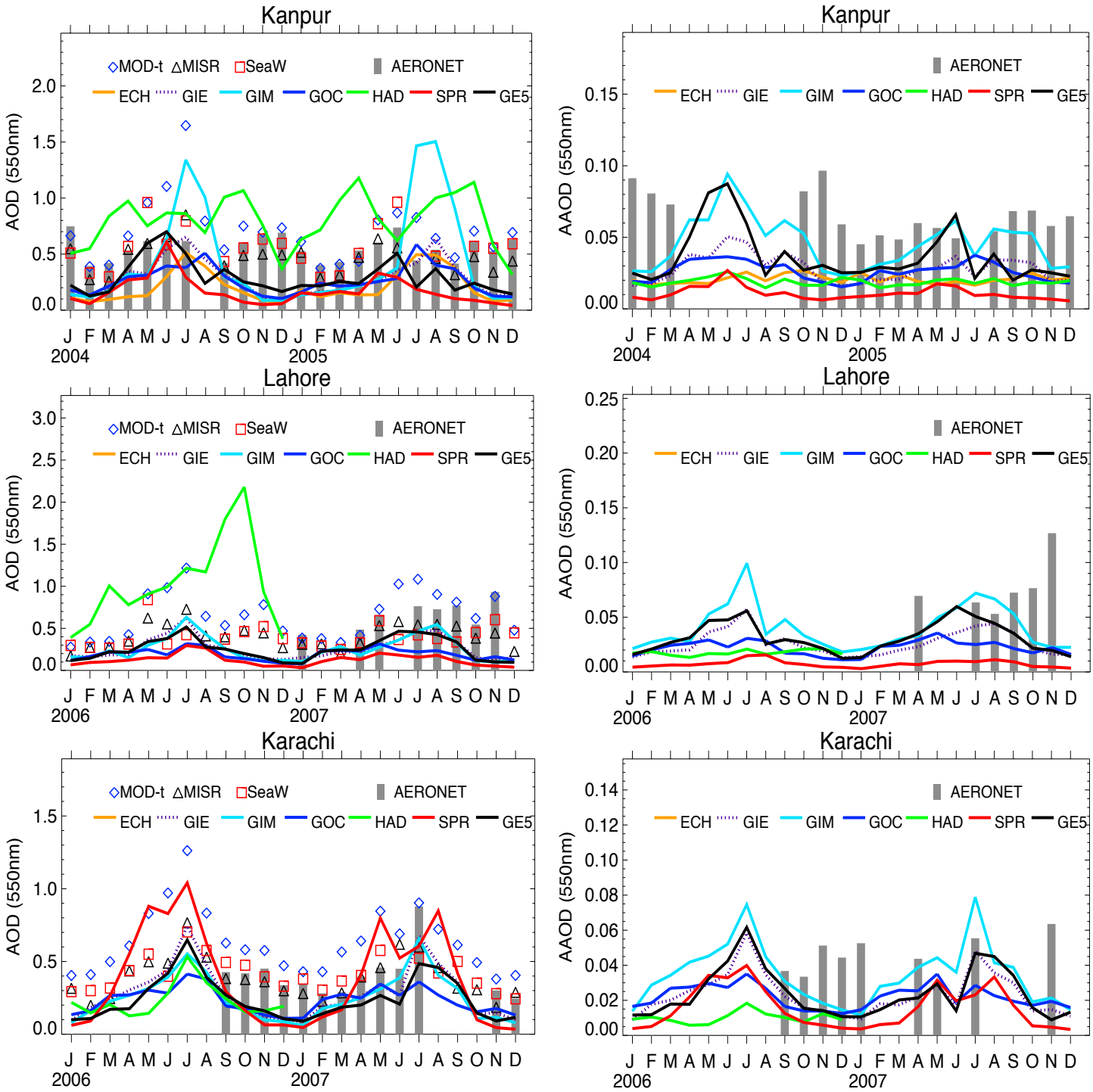

Figure 5. Monthly mean AOD (left column) and AAOD (right column) in a two-year period over 3 AERONET stations in South Asia. The gray bar represents measurement from AERONET. The thin curves represent seven models, and symbols represent three NASA remote sensors.

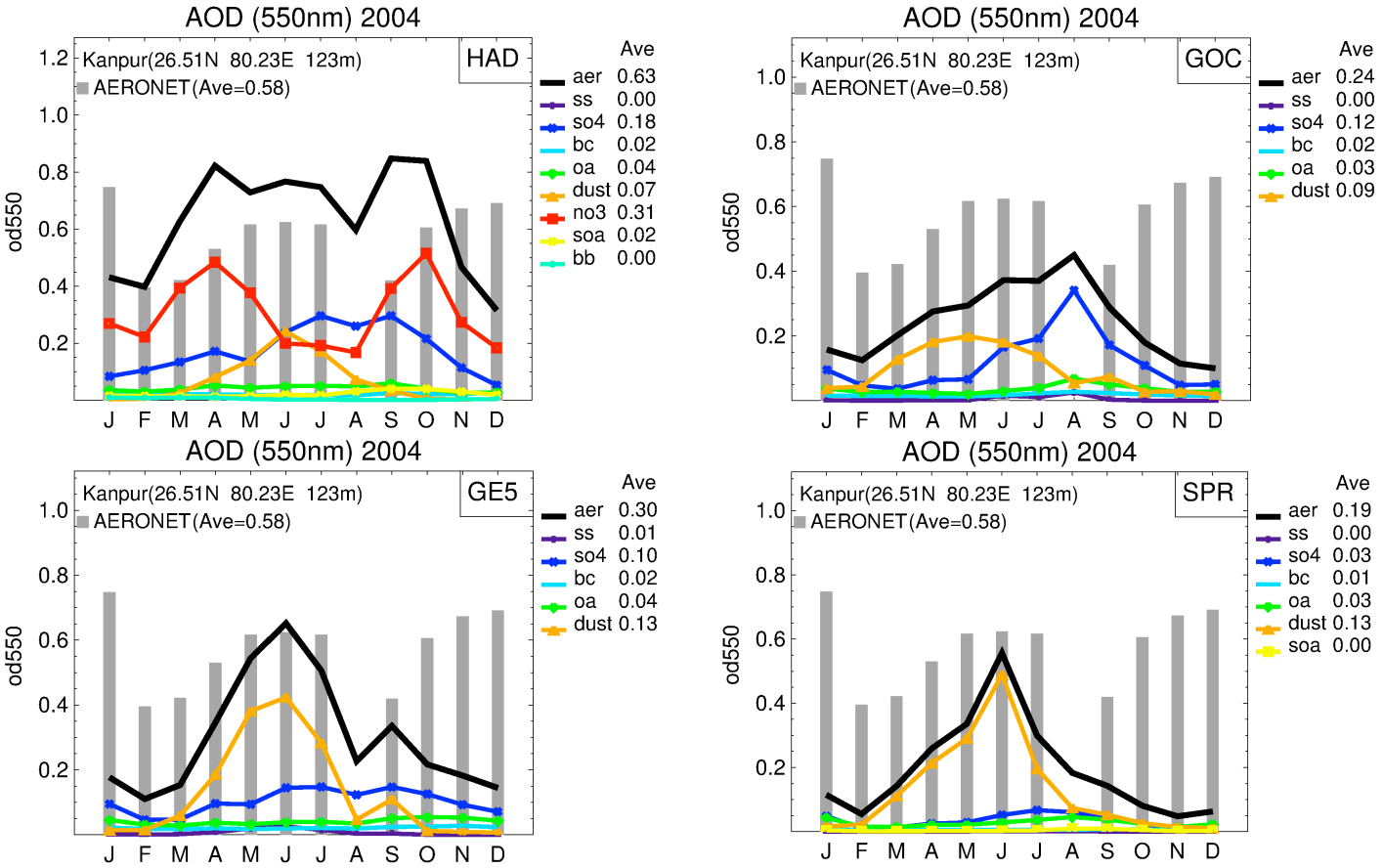

Figure 6. AOD of total aerosol (aer) and components (ss, so4, bc, oa, dust, no3, soa and bb) at Kanpur for 2004 in 4 models, HAD (upper left), GOC(upper right), GES (lower left) and SPR (lower right). The gray bar represents measurement from AERONET. The annual mean AOD value is followed after the name of each symbol. NOTE: bc and oa represent emission from fossil fuel only and bb represents emission from biomass burning only).

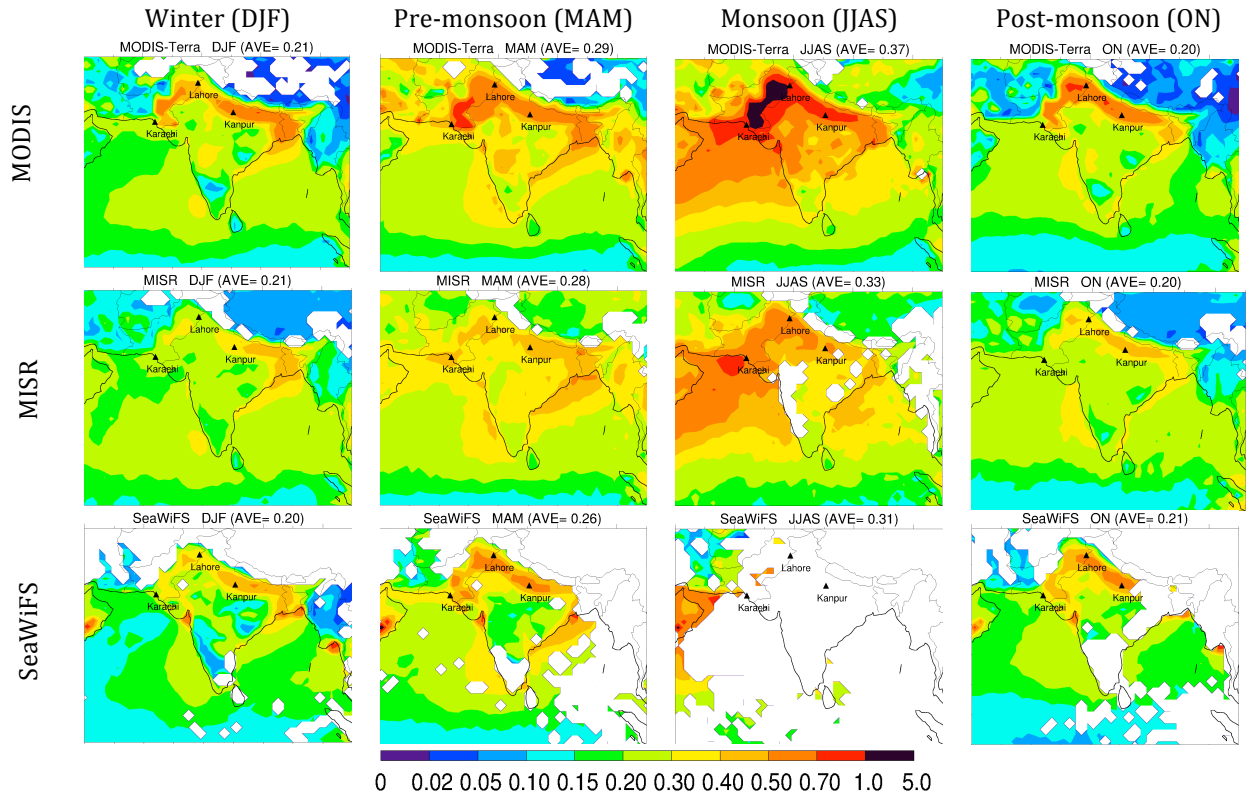

Figure 7a. Spatial distribution of AOD over South Asia in 4 seasons averaged for 2000–2007 in three satellite observations. The corresponding area averaged annual mean AOD value is listed in each panel (domain:0–36°N; 55°E–100°E). Three AERONET stations used in this study are labeled in the maps. Regions in white indicate insufficient sampling sizes of aerosol retrievals due to the presence of bright surface or frequent cloud cover in satellite data.

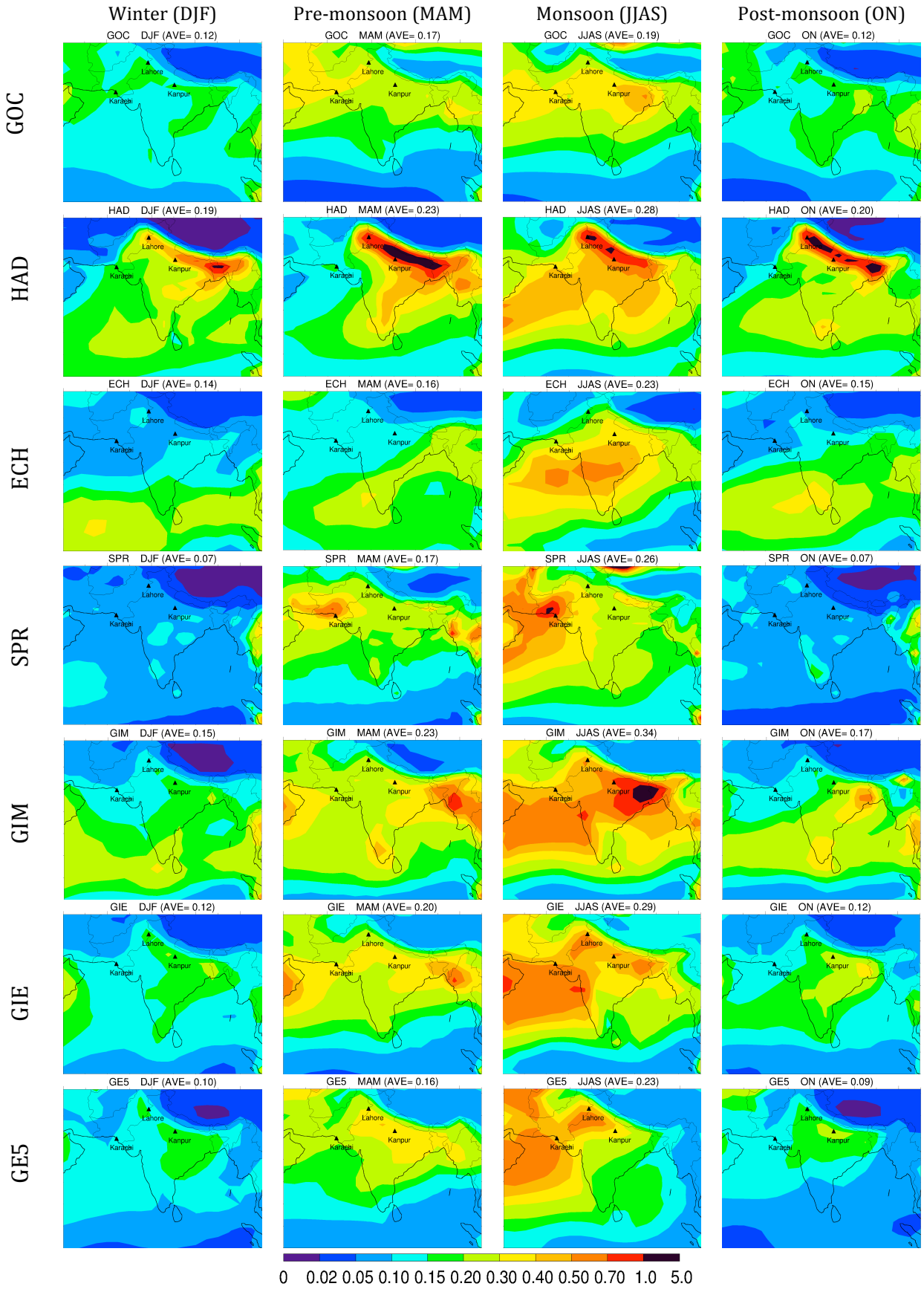

Figure 7b. Spatial distribution of AOD over South Asia in 4 seasons averaged for 2000–2007 in seven models (the first three models with the anthropogenic emissions from A2-MAP and the rest with A2-ACCMIP). The corresponding area averaged annual mean AOD value is listed in each panel (domain:0–36°N; 55°E–100°E). Three AERONET stations used in this study are labeled in the maps.

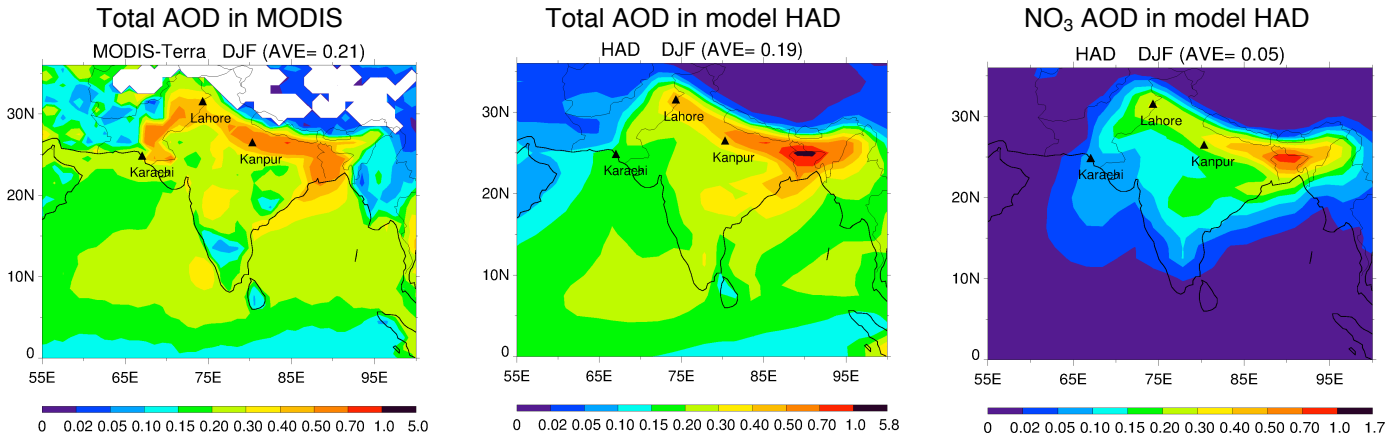

Fig. S1. The total AOD and NO_3 AOD during the winter of 2000-2007.

Response to referee # 2

The authors would like to thank the reviewer for constructive comments and guidance on improvement of this paper. Below are <our responses> to the [comments from referee #2]:

[COMMENT FROM REFEREE]: This manuscript compares observations with modeled aerosol properties in South Asia (primarily India), with a focus on the Indo-Gangetic Plain (IGP), from 7 global models. There are a number of strengths of this manuscript. The first is that it addresses a region of clear low bias in the models and seeks to better understand the source of the bias. That this region is home for a large population makes the study even more compelling. The second is that it brings a variety of observations (satellite and ground- based remote-sensing, as well as in situ) to compare with model products. A third is the use of a range of global models, which permits comparing different models with different capabilities (e.g. those that include nitrate aerosol versus those that do not), makes the results more robust than if only a single model were used. That said, there are a number of ways that the manuscript could and should be improved. Broadly, two main issues are (1) improving comparisons of model output and observations and (2) quantifying the various explanations for the model low biases.

1. Use of observations

(a) The authors accurately state (p. 13, 20-22) "It should be noted that it is difficult for a global model with a coarse spatial resolution to reproduce pollutant concentrations measured in an urban environment...", which I agree with. However, recognition of this is not, I believe, sufficient. Given this known scale issue, what would constitute "agreement" between model and the point observation? Presumably if the model out- put exactly matched the point observation that would not imply a perfect model. So without some clear idea of what a perfect model would do, how do you know there is a "low bias" in the model? It's possible that the entire mismatch is due to scaling, right? I don't think that this is the case, but it seems that quantifying this issue is required. What if, say, CALIPSO or some other satellite data were used to try to get some sense of the spatial distribution in this particular grid box?

<RESPONSE> The underestimation of BC found in the urban city (e.g. Delhi) could partly attribute to the fact that a global model with a coarse spatial resolution is difficult to reproduce pollutant concentrations measured in a point station under urban environment. However, the underestimations of BC surface concentration are found in those background stations as well (e.g. over the mountain site of Nainital and the island sites of Minicoy and Port Blair), in Figure 9. In addition, the conclusion that the modeled AODs are too low is based on the comparisons not only with AERONET point observations, but also with the level-3 multiple satellite data from MODIS, SeaWiFS (both 1°x1° resolution) and MISR (0.5°x0.5° resolution) on regional scales, as shown in Figure 5 and Figure 7. Therefore, the underestimation of modeled BC and AOD in the wintertime is more likely due to other factors, as discussed in Section 5, than scaling. We have modified the text to clarify the discussion on model low bias in Section 4.5.

[COMMENT FROM REFEREE]: (b) All observations have their uncertainties and, most importantly for this study, biases. To conclude that the model biases are large, one should probably quantitatively evaluate the observational biases. How much of the model/observation discrepancies might be a result of the observations?

For example, my understanding is that AERONET has a very strict cloud-screening requirement. I did not see details on how AERONET data are compared with models. Was it assumed that AERONET is representative of all conditions, regardless of cloud cover? This could lead to substantial biases if there is some correlation between meteorology and aerosol. Or was there a cloud-screening criterion applied to the model output as well? If so, how does one reconcile the model scale with the AERONET scale?

<RESPONSE>: We agree with the reviewer that there is uncertainty to compare cloud-screened AOD with the modeled AOD. AERONET AOD data are only under clear-sky conditions, while the model output are under all-sky conditions, except two models (GISS-modelE and GISS-MATRIX) that also provided clear-sky AOD. As shown in the paper, we used clear sky AOD from these two models in the model-data comparison. Considering the fact that the clear-sky AOD is generally lower than its corresponding all-sky AOD (e.g. by 60% based on GISS-modelE at Kanpur), the low biases in other five models, especially during the winter, would be more pronounced if clear-sky AOD were present in these models. We now have added the discussion on the difference of all-sky and clear-sky AOD in Section 2.1.

[COMMENT FROM REFEREE]: I don't know much about satellite remotely-sensed aerosol products, but I suspect there are a number of potential biases. One obvious one would be the late-morning/early afternoon timing of the overpasses not accurately reflecting a daily average in aerosol.

<RESPONSE>: We have used monthly mean AOD from several satellite products (MODIS, MISR, SeaWiFS) to compare with the models. Although the satellite data are averaged over the "snap shot" observations at the local overpassing time (varying between 10:30AM to 1:30PM) and the model results are diurnally averaged, previous studies compared model simulated AOD sampled at MODIS/MISR overpass times with that averaged over diurnal time steps and found the differences were small on monthly mean AOD, only about 10% in south America and southern Africa (i.e. biomass burning regions) and smaller elsewhere (Colarco et al., 2010). Thus, since we are using monthly mean satellite data products in comparison to monthly mean model AOD simulations, the bias caused by time difference is expected to be small. We will note these discussions in the revised manuscript in Section 3.1, per reviewer's comment.

[COMMENT FROM REFEREE]: Also, my understanding is that some (if not all) of the passive sensors used (MISR, MODIS, SeaWIFS) require a surface albedo in order to make certain retrievals. If so, what albedo product was used? Is there, say, an annual cycle in albedo (perhaps due to vegetation or agricultural cycles) that is not properly represented in this region that causes an observational bias? Is there an issue with retrievals of external aerosol mixtures (e.g. mixed absorbing and scattering aerosol)? As I said, this is not my area but I believe this should be explored much more carefully.

<RESPONSE>: Yes, the satellite-based aerosol retrievals require information about the underlying surface reflectance for different surface types. However, the surface reflectance parameterizations are generally well established in the respective aerosol retrieval algorithms, and have improved significantly in the past decade (e.g. Levy et al. 2007, 2010; Hsu et al. 2006; Kahn et al., 2007, 2010; Sayer et al., 2012, 2013). These aerosol products (from MODIS, MISR and SeaWiFS) are regionally validated retrievals with reference to AERONET sites located worldwide, and include uncertainties (e.g. due to surface reflectance) as part of each product's accuracy assessment. For example, MODIS dark-target aerosol product has an improved surface reflectance parameterization introduced in collection 5.1 AOD dataset (Levy et al. 2007), which is used in our paper, with its overall uncertainty over land reported to be within ±(0.05±0.15%) AOD and better for oceanic regions (Levy et al., 2010). Whereas, about 70% of the MODIS Deep Blue (aerosol retrievals over bright reflecting surfaces such as desert/arid regions) and SeaWiFS AOD (over both bright desert/arid regions and vegetated surface) retrievals fall within an expected absolute uncertainty of 0.05 ± 20% (for the wavelength of 550nm AOD used in our paper) (Sayer et al. 2012, 2013). It should also be noted that only the best-quality aerosol retrievals are aggregated to form the Level-3 gridded monthly mean AOD dataset, which is being used in our paper. Similarly, aerosol retrievals from MISR have comparable or better accuracy assessment as part of their overall uncertainty (Kahn et al. 2010). Therefore, per the extensive validation and improved parameterization of surface reflectance in satellite aerosol retrievals, any large biases or seasonal influences of surface albedo variations on our intercomparison study between satellite/AERONET/model AOD, is unlikely. We have added the aforementioned uncertainties of various AOD products in Section 3.1,

{MC: The retrieval uncertainties and error estimates of the satellite products we used in this study have been

published extensively, including addressing the impact of errors in the surface reflectance on aerosol retrieval qualities (e.g., Levy et al., 2007, 2010, 2014; Remer et al., 2008; Kahn et al., 2010, 20xx; Sayer et al., 2012, 2013; Hsu et al., 2006; Drury et al., 20xx). For example, the MODIS AOD has shown high bias over relatively bright surfaces, such as the western US, and errors in the urban areas mostly because of the surface reflectance that was inadequately characterized (ref.) Thorough assessment of the surface-reflectance induced error in satellite AOD retrieval is an area of the AERO-SAT (the international satellite aerosol science network) activities, a topic which is beyond the scope of this paper. It should be noted that only quality-assured aerosol retrievals are aggregated to form the Level-3 (gridded) monthly mean AOD dataset, which is being used in our paper.}

Regarding Reviewer's comment related to issues with retrievals of external aerosol mixtures: satellite-based aerosol retrievals surely take into account external aerosol mixtures (such as varying degrees of mixtures of absorbing and scattering aerosol types). For all three satellite retrievals used in this study, MODIS, MISR and SeaWiFS, they use a lookup table approach including several aerosol optical models consisting of varying degrees of aerosol absorption/scattering and various size bins. Additionally, MODIS aerosol retrievals benefit from a clustering approach based on dominant aerosol types/mixtures assigned to a specific region depending on regional aerosol characteristics compiled from AERONET data.

[COMMENT FROM REFEREE]: (c) Uncertainty/variability

Most of the figures showing observations lack any indication of uncertainties or variability (whichever is larger). This should be included to aid in comparing observations with models.

<RESPONSE>: We have added the correlation, relative mean bias and root mean square error of each model in Fig.5, and one standard deviation in Figure 7 and 8 to show the uncertainty/variability. Please check the improved figures at the end of this file. In addition, variability of AOD from multiple models and satellite (using one standard deviation) was also shown earlier in Figure 4 (in parentheses, alongside mean values).

[COMMENT FROM REFEREE]: 2. Quantifying causes of model biases

(a) While the manuscript lists all the potential sources of biases in models, it would be a lot more satisfying if you could actually quantify these bias sources in some way. I understand that it's not easy to do with high precision for a variety of reasons (e.g. model dependence), but even a ranking or sorting the bias sources into tiers (e.g. Tier 1: dominant bias source; Tier 2: major bias source; Tier 3: minor bias source) seems like it would be very useful. Such quantification (or semi-quantification) would be a much more satisfying product of this research than the mostly qualitative statements that are currently provided. In some cases, it seems like it wouldn't take much work to actually provide quantitative estimates, but maybe for others it will require some new analyses.

(a) A related issue is that the manuscript addresses the bias sources somewhat superficially. You broadly describe what the problem is, but don't really do a good job of analyzing more carefully what the specific issue is. Here are some examples:

*** The low bias in relative humidity is described, and there is speculation that the cause is a high bias in temperature. Well, why isn't this checked? It would be quite easy to take the model output, apply a more appropriate T, and see if the humidity bias disappears. Or if it corrects a small fraction of the bias, then one would conclude that it's actually an absolute humidity bias.**

<RESPONSE>: This is a good point, although it cannot be shown using a simple sensitivity calculation. We have done some further investigation on the potential sources of biases with one of these seven models, NASA GEOS5, in our subsequent IMPROVING PROJECT. Based on the ongoing work and preliminary results, we

found the ranking of the bias sources to be, in the order of importance: Tier 1 – poor representation of meteorological fields in these models (e.g. wind, relative humidity and temperature); Tier 2 – insufficient anthropogenic emission and lack of nitrate; Tier 3 – insufficient model spatial resolution. Considering this as a subsequent study in itself, which involves extensive analysis of various parameters/factors and their sensitivities, we have decided to investigate/solve the "puzzle" of underestimation of South Asia aerosol simulation in two steps – first with the current study to identify the problem in multi-models and then to quantify the contribution of each potential source of biases in IMPROVING PROJECT. However, per Reviewer's suggestion, we will briefly allude to our ongoing follow-up study here and briefly mention about the first results of the assessment of relative impacts of various source biases on modeled AOD.

<RESPONSE>: South Asia is a difficult region for global models to reproduce the aerosol observations, and our focus in this paper mainly includes to evaluate the performance of the multiple global models participating in the AeroCom Phase II model experiments with satellite and ground-based data, to find common problems and model diversity, and to suggest the possible causes of the problems. Because of the limited model output fields in the AeroCom protocol, there are simply no enough information to further investigate the source of errors and rank them accordingly across the multiple models. Realizing the importance of understanding the source of the bias, we are currently working on quantifying the problems with ranks of importance via a series model sensitivity studies using our own model (GEOS5), including change the model spatial resolution, emission strength, additional species, meteorological variables, etc. These sensitivity simulations will allow us to rank the importance of the bias sources, which is not possible to do with the AeroCom models but will definitely provide insights to diagnose the model problems and directions of improvements for all models. We will report the findings in our future publications. The above discussion has been added to the Section 6.

[COMMENT FROM REFEREE]: * Boundary layer depth is mentioned as a source of bias in comparing surface observations. There must be some measure of boundary layer thickness in this region, either in situ or remotely-sensed, that can be used to evaluate this idea quantitatively.

<RESPONSE>: Right, the atmospheric boundary layer (ABL) plays an important role in modulating the surface concentration including BC. In winter, the averaged ABL is 400-500 meters in the GOCART model (similar meteorological data is used by GEOS5, one of the models used in our paper), which is about twice thicker than the observed ABL (Tripathi et al., 2006; Nair et al. 2007), thus a better-constrained ABL in models could be helpful (Moorthy et al. 2013). Unfortunately we don't have ABL information from other models, so it is difficult to address this point in detail. We have added this discussion in Section 5.5.

[COMMENT FROM REFEREE]: * A low-bias in sulfate aerosol is found. Wouldn't it be interesting to try to isolate this problem? Determine whether it is, say, a result of gas-to-particle conversion that is too slow or in the sulfur emission inventory. The former could be diagnosed if *total* sulfur was accurately represented in the model, but the ratio of gas phase to particle phase sulfur was too high. Similarly for organics and nitrate, at least for those models that actually have nitrate.

<RESPONSE>: It is a good suggestion. However, unfortunately there is no observed SO_2 concentration or nitrate precursors available for investigating the gas-to-particle conversion. The sulfur emission inventories used by the models were very similar.

[COMMENT FROM REFEREE]: I've provided a number of other comments in an attached PDF file. Some may overlap with the above and can be ignored. Most identify areas where the wording is awkward, ambiguous or otherwise requiring editing.

<RESPONSE>: We have incorporated your comments in a marked-up manuscript in the supplement.

References:

Colarco, P., A. da Silva, M. Chin, and T. Diehl (2010), Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, J. Geophys. Res., 115, D14207, doi:10.1029/2009JD012820.

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

Levy, R. C., L. A. Remer, S. Mattoo, E. F. Vermote, and Y. J. Kaufman (2007), Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, J. Geophys. Res., 112, D13211, doi:10.1029/2006JD007811.

Sayer, A. M., Hsu, N. C., Bettenhausen, C., Jeong, M.-J., Holben, B. N., and Zhang, J.: Global and regional evaluation of over-land spectral aerosol optical depth retrievals from SeaWiFS, Atmos. Meas. Tech., 5, 1761-1778, doi:10.5194/amt-5-1761-2012, 2012.

Sayer, A. M., N. C. Hsu, C. Bettenhausen, and M.-J. Jeong (2013), Validation and uncertainty estimates for MODIS Collection 6 “Deep Blue” aerosol data, J. Geophys. Res. Atmos., 118,7864–7872, doi:10.1002/jgrd.50600.

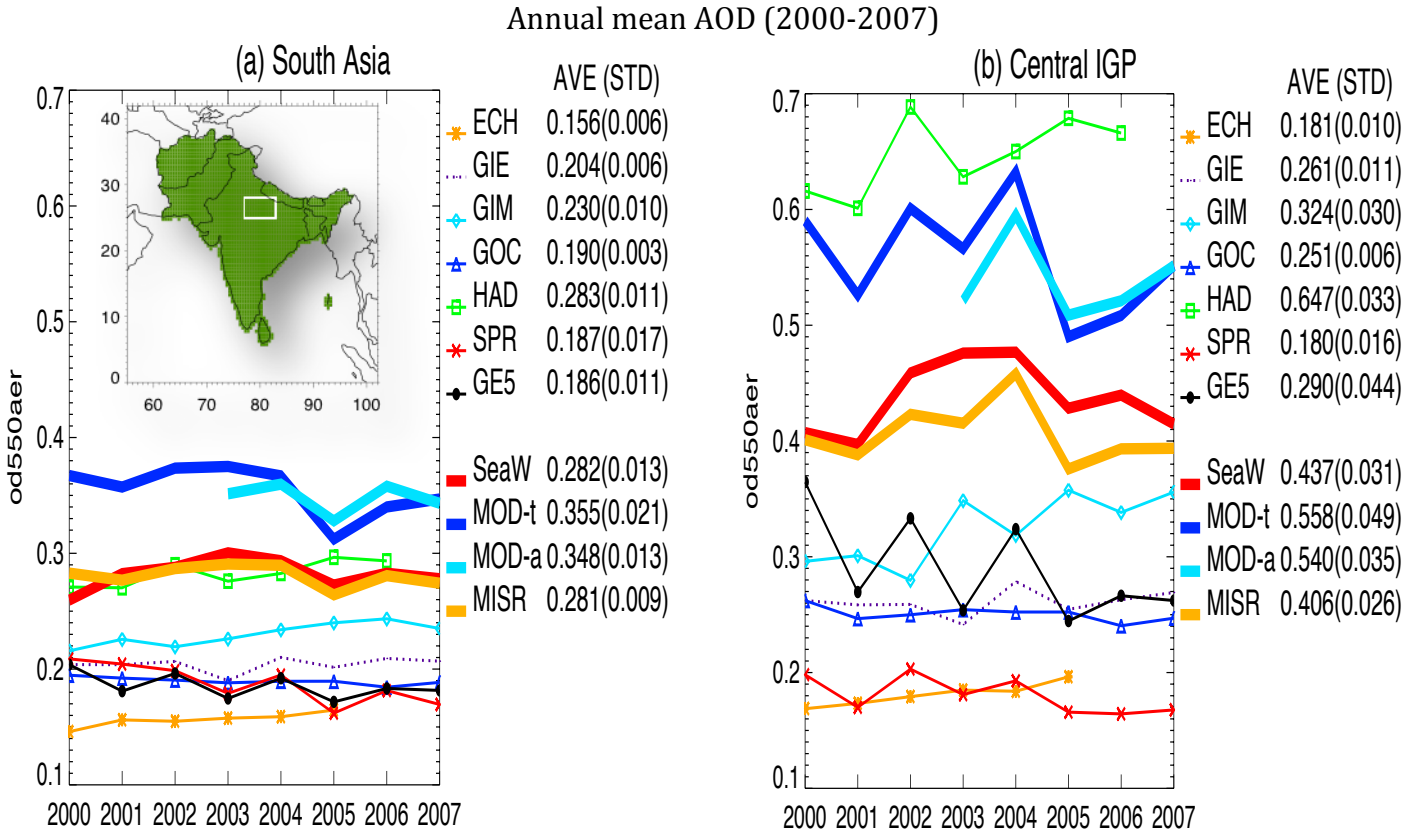

Figure 4. The annual averaged mean AOD for 2000-2007 over (a) South Asia (the green area in the map); (b) Central IGP (77°-83°E; 25°-28°N, the white box in that map). The thin curves with symbols represent seven models, and the thick curves represent four NASA remote sensors, with corresponding multi-year averaged annual mean AOD and the standard deviation followed.

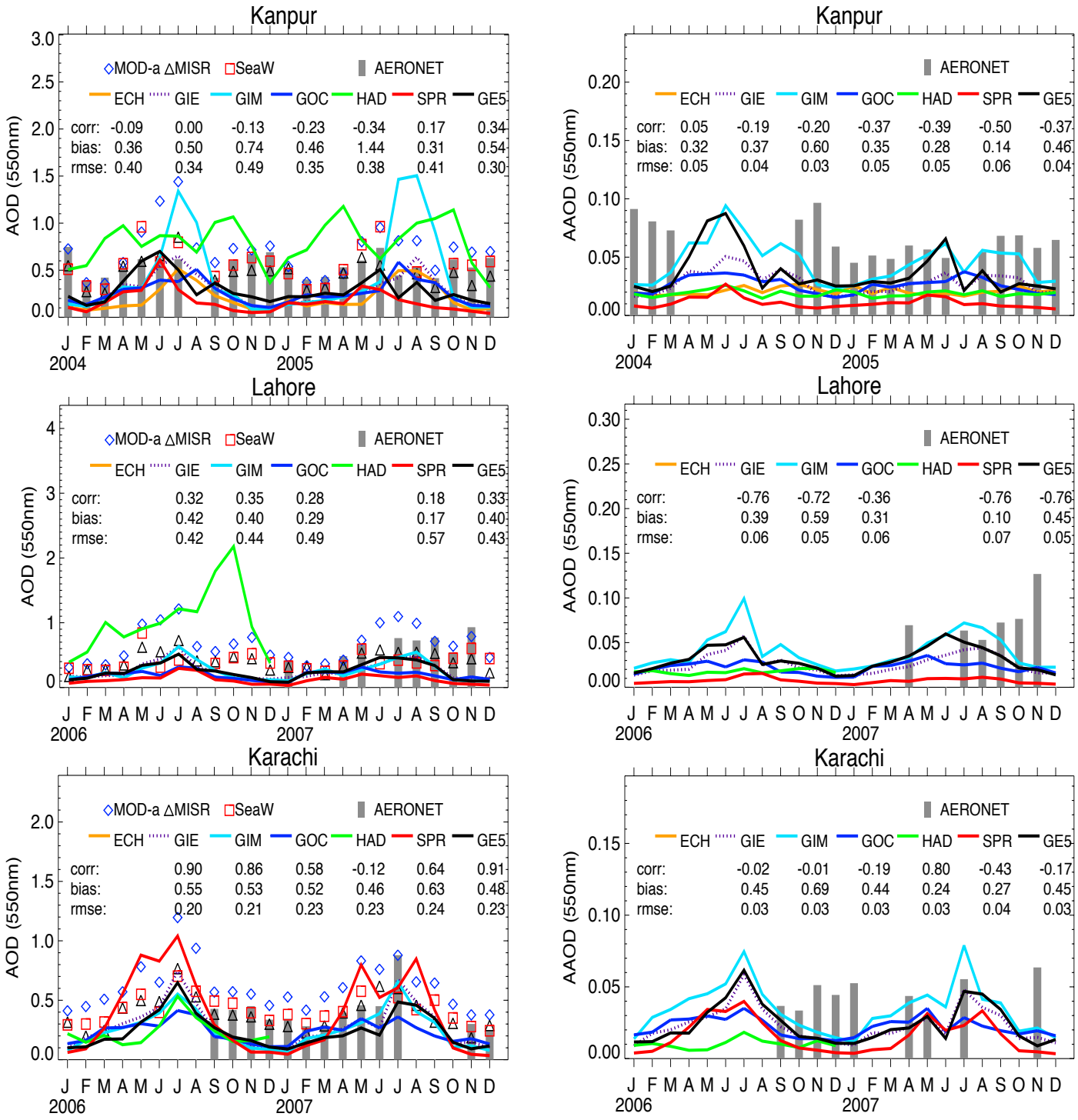

Figure 5. Monthly mean AOD (left column) and AAOD (right column) in a two-year period over 3 AERONET stations in South Asia. The gray bar represents measurement from AERONET. The thin curves represent seven models, and symbols represent three NASA remote sensors. On each panel, corr=correlation coefficient of a model with AERONET, bias=relative mean bias, i.e. $\Sigma(\mathrm{MODEL}_i)/\Sigma(\mathrm{AERONET}_i)$, rmse=root-mean-square error relative to AERONET.

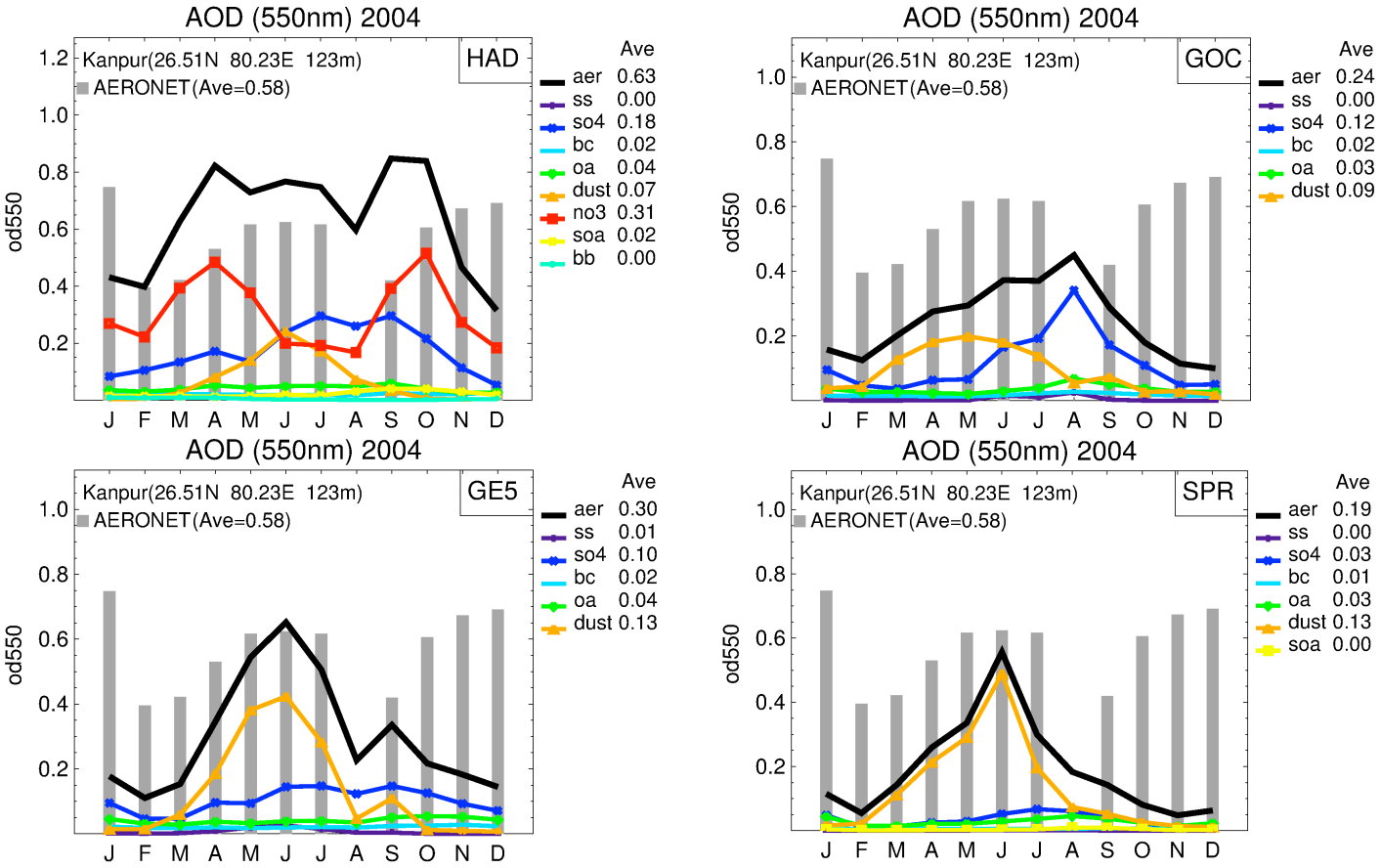

Figure 6. AOD of total aerosol (aer) and components (ss, so4, bc, oa, dust, no3, soa and bb) at Kanpur for 2004 in 4 models, HAD (upper left), GOC(upper right), GES (lower left) and SPR (lower right). The gray bar represents measurement from AERONET. The annual mean AOD value is followed after the name of each symbol. NOTE: bc and oa represent emission from fossil fuel only and bb represents emission from biomass burning only).

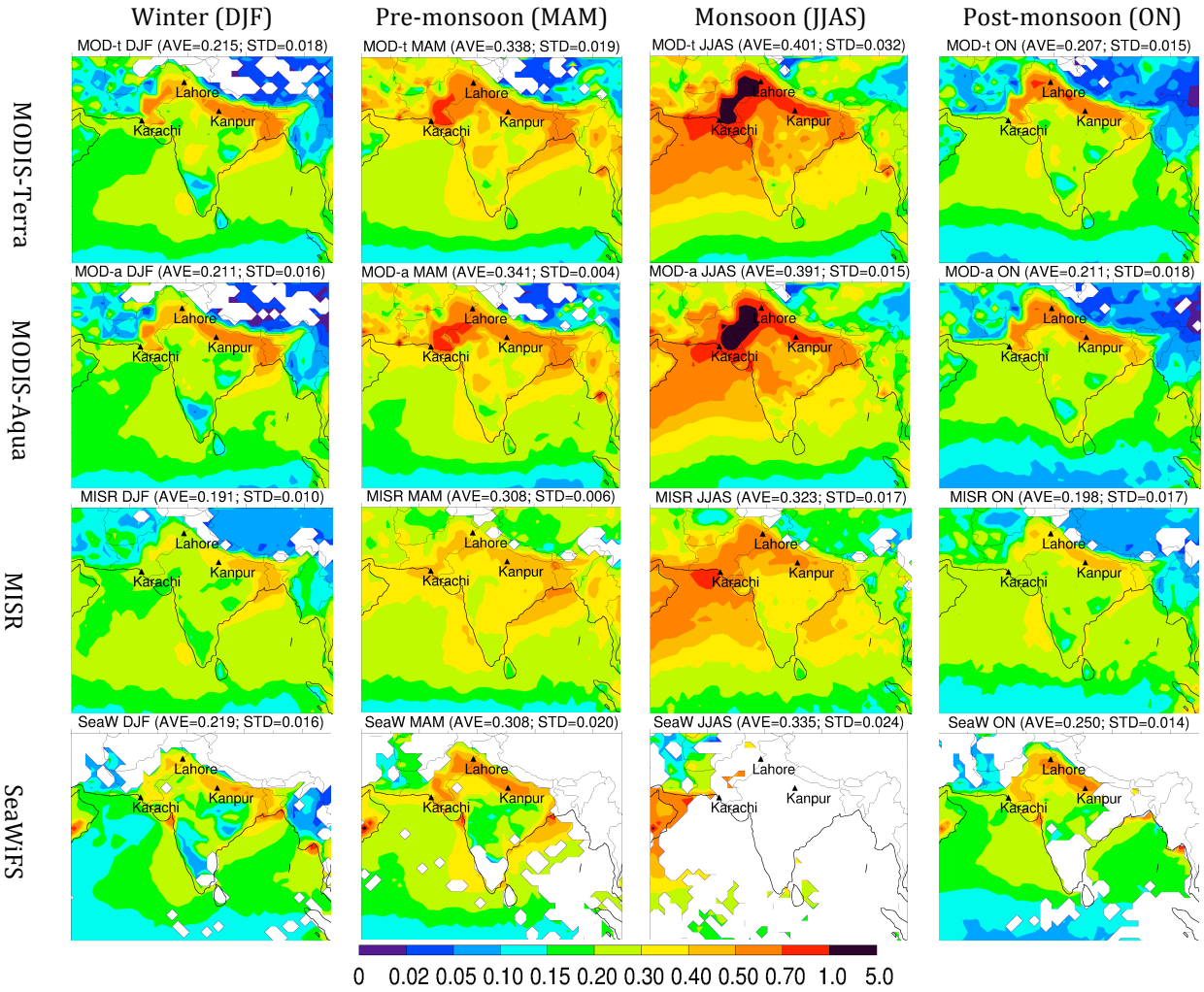

Figure 7a. Spatial distribution of AOD over South Asia in 4 seasons averaged for 2000–2007 in three satellite observations (two from MODIS, MISR and SeaWiFS). The corresponding area averaged annual mean AOD value is listed in each panel (domain:0–36°N; 55°E–100°E). Three AERONET stations used in this study are labeled in the maps. Regions in white indicate insufficient sampling sizes of aerosol retrievals due to the presence of bright surface or frequent cloud cover in satellite data.

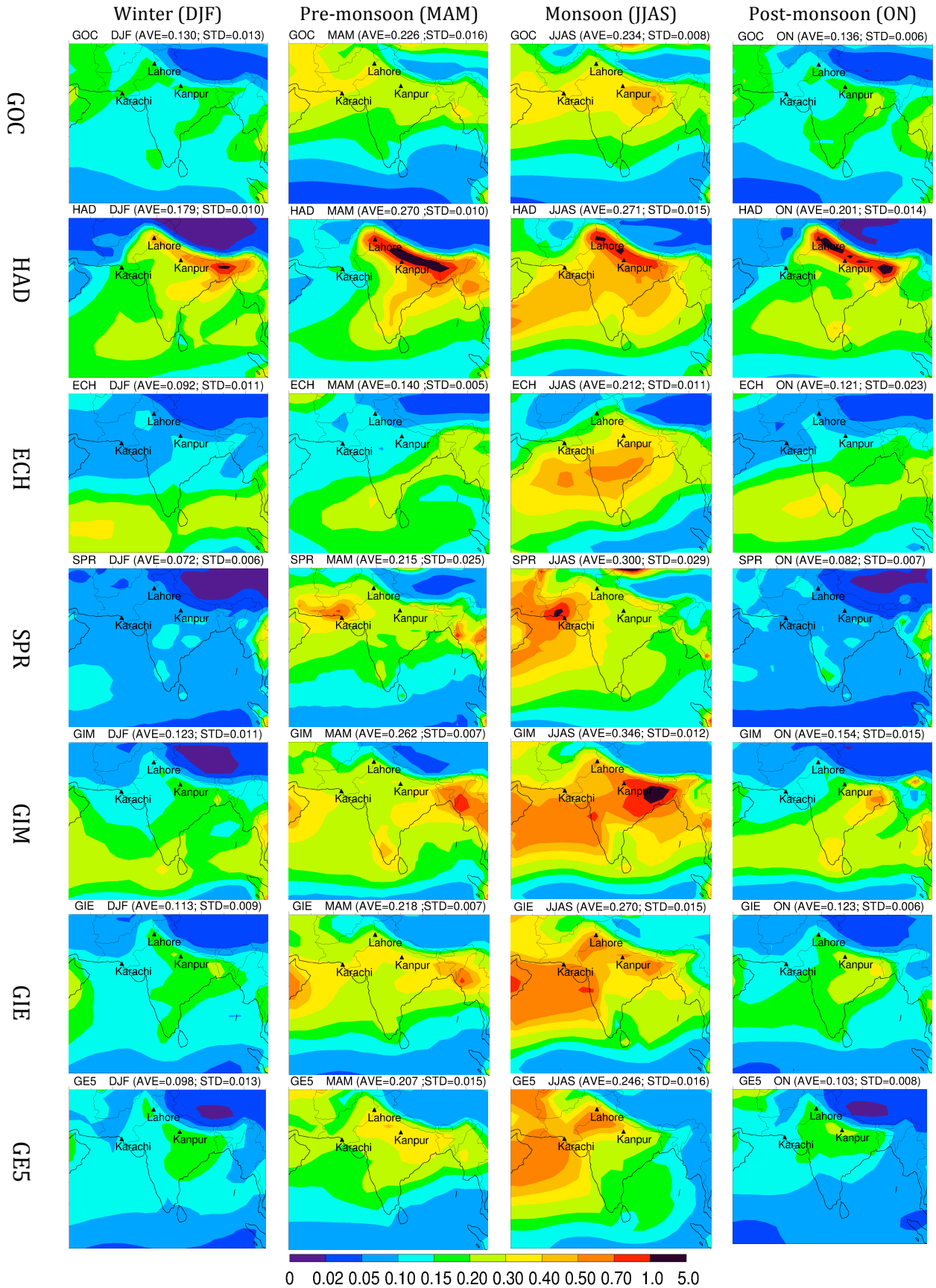

Figure 7b. Spatial distribution of AOD over South Asia in 4 seasons averaged for 2000–2007 in seven models (the first three models with the anthropogenic emissions from A2-MAP and the rest with A2-ACCMIP). The corresponding area averaged annual mean AOD value is listed in each panel (domain:0–36°N; 55°E–100°E). Three AERONET stations used in this study are labeled in the maps.

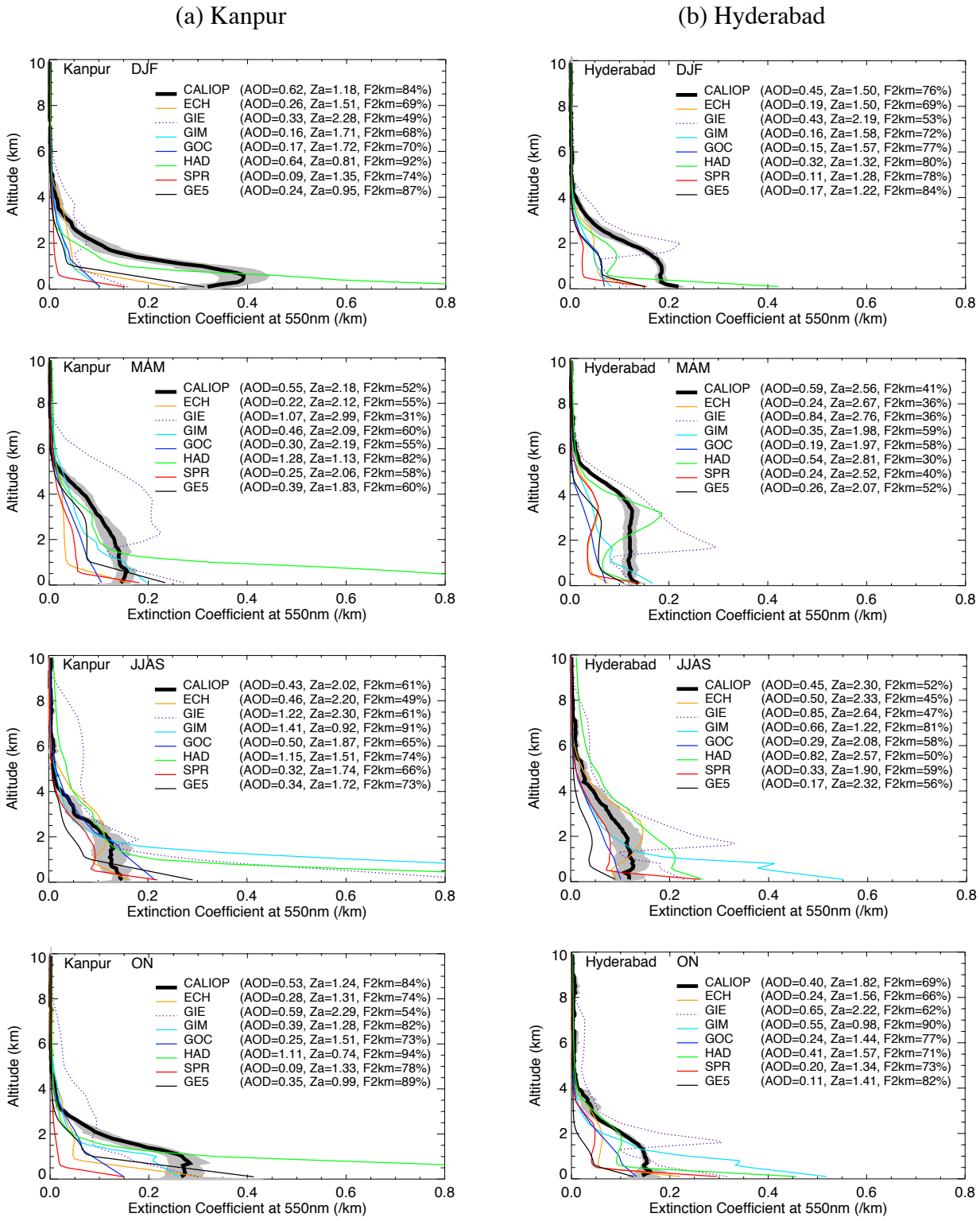

Figure 8. The seasonal mean of vertical profile of extinction coefficient (units: 1/km) at (a) Kanpur, and (b) Hyderabad from CALIOP and seven models. Units of Za is km. The corresponding averaged AOD, Za and F_{2km} are listed after each symbol name. The gray shaded area in CALIOP shows one standard deviation relative to 2006-2011 averages.