THE UNIVERSITY *of* EDINBURGH

Understanding particulate matter pollution and photovoltaic power output using data and models

Fei Yao

THE UNIVERSITY
of EDINBURGH

Thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy
to the
University of Edinburgh — 2022

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

Fei Yao

March 2022

Abstract

Particulate matter (PM) both in the atmosphere and deposited on solar photovoltaic (PV) panels reduce PV energy generation efficiency. Atmospheric PM near the surface, particularly those with an aerodynamic diameter $\leq$ 2.5 μm ($PM_{2.5}$), have well-documented deleterious impacts on human health. It is imperative to develop models to accurately simulate PM impacts on PV efficiency in order to enlighten relevant policy-making aimed at reducing these impacts. It is important to develop models to accurately capture magnitudes and variabilities of ground-level $PM_{2.5}$ concentrations in order to inform follow-on epidemiological studies. In this thesis I use a combination of satellite observations, *in situ* measurements, process-driven models, and data-driven models to understand PM pollution and PV power output with a focus on their links from global to regional scales.

In Chapter 2, I integrate the GEOS-Chem global 3-D model of atmospheric composition, equipped with online radiative transfer calculations, with PVLIB-Python which is a solar PV performance model, to quantify PV efficiency losses due to atmospheric and deposited PM. I calculate three PV efficiencies: 1) real PV efficiency considering atmospheric and deposited PM; 2) hypothetical PV efficiency considering atmospheric PM only; and 3) hypothetical PV efficiency assuming no PM. By comparing these PV efficiencies, I find that regions with low PV efficiency are typically associated with high PM-induced PV efficiency losses, and that the losses due to deposited PM far exceed those due to atmospheric

PM, with the maximum magnitude of the former almost eight times that of the latter. Desert regions including the Sahara, Arabian-Peninsula, Central Asia, and Southern South America are most susceptible to deposited PM which causes PV efficiency losses that are comparable to the maximum PV efficiency close to 0.3 achieved elsewhere. Coastal regions are also significantly affected by deposited PM, e.g. countries around the Caribbean and the Mediterranean, and over New-Zealand. The main regions where PV efficiency losses due to atmospheric PM are East and South Asia, particularly over highly polluted regions such as North China and the Indo-Gangetic Plain.

In Chapter 3, I focus on developing strategies to mitigate PM-induced PV efficiency losses. From the perspective of reducing emissions, I calculate the aforementioned three PV efficiencies by halving anthropogenic source sector emissions. By subtracting each of these quantities from baseline values and comparing these differences across source sectors, I find that that reducing residential emissions is the most effective approach to reduce PM-induced PV efficiency losses, and that the biggest PV efficiency gains are over East and South Asia. Using 2019 PV capacities as a baseline, I find that a 50% reduction in residential emissions would lead to an additional 7,687 and 1,823 GWh yr^{-1} produced in China and India, respectively. The corresponding economic benefits would amount to US$653 million and US$144 million per year, respectively. From the perspective of removing deposited PM from solar panels either manually or by robots, I calculate the aforementioned first PV efficiency by taking into account yearly, quarterly, monthly, weekly, and daily sweeping of panels. By subtracting the baseline value from this quantity and comparing these differences across frequencies, I find that routine sweeping of panels is effective at reducing PV efficiency losses due to deposited PM, and that even an annual sweeping routine could lead to an approximately 40% PV efficiency recovery in Central Asia and Southern South America. The work described in previous and current chapters

are collectively and succinctly summarized into a manuscript that is under review for *Environmental Science & Technology*.

In Chapter 4, I go a step further by fully quantifying the benefits to air quality and PV power output from reducing residential fuel emissions. I run the integrated model at a finer resolution of 0.5° latitude × 0.625° longitude with a focus on an Asian wintertime (January 1 to February 29, 2008) which is the main spatiotemporal domain where reducing residential emissions substantially benefit PV power output. I run the integrated model with original emissions to output baseline ground-level $PM_{2.5}$ concentrations, column-integrated aerosol optical depth (AOD) at wavelength of 550 nm (AOD_{550nm}), and the aforementioned three PV efficiencies. I then run the integrated model again but with reduced residential fuel emissions to determine the benefits to air quality and PV power output from reducing these emissions by taking the difference of baseline and sensitivity simulations. I find that Eastern and Northeastern China, and the Indo-Gangetic Plain are the three key regions where reducing residential particularly solid biofuel emissions leads approximately linearly to reductions in PM pollution and improvements in PV efficiency. Completely removing residential emissions would result in 20–30% and 30–40% reductions in column-integrated AOD_{550nm} and ground-level $PM_{2.5}$ concentrations, respectively, and an approximately 30% improvement in PV efficiency. I attribute these approximately linear benefits to the large volume of carbon emissions, primarily owing to the low combustion and thermal efficiencies of residential devices and the general absence of any end-of-pipe controls, that typically form into carbon aerosols in an approximately linear way in the atmosphere after aggregating to temporal and spatial mean values. This work is being prepared for submission.

In Chapter 5, I use the GEOS-Chem 3-D model of atmospheric composition to improve our ability to use satellite-retrieved column-integrated AOD and statistical and machine learning models to infer ground-level $PM_{2.5}$ concentrations. In

particular, I use the GEOS-Chem global 3-D model of atmospheric composition to explore how changes in the vertical distribution of aerosol extinction coefficients affects the relationship between ground-level $PM_{2.5}$ and column-integrated AOD and how I can use that information to improve the robustness of ground-level $PM_{2.5}$ estimates inferred from satellite-retrieved AOD over eastern China. I define a metric, Γ_{PBL}^{AOD}, that describes the fraction of AOD that resides in the planetary boundary layer compared to the total columnar AOD. I determine physically-meaningful $PM_{2.5}$:AOD relationships using data for which $\Gamma_{PBL}^{AOD} \geq 50\%$, a criterion based on sensitivity analyses on data clusters that I identify using a hierarchical clustering method. I use statistical and machine learning models to describe these $PM_{2.5}$:AOD relationships, and use a Monte Carlo approach to quantify the improvement after the selection of more physically relevant data records. Benefiting from the improved representativeness of AOD for ground-level $PM_{2.5}$, the new method effectively reduces bias in inferred estimates of ground-level $PM_{2.5}$ by 10–15% and 9–12% for space-borne sensors passing over in the morning and afternoon, respectively. It also captures more variations in ground-level $PM_{2.5}$ by up to 8% and 5% for space-borne sensors passing over in the morning and afternoon, respectively. This work is published in *Atmospheric Environment*.

In closing, the above findings advance understanding of PM pollution and PV power output particularly their links, and provide motivation for future studies of evaluating impacts of real emission controls and climate change on past, current, and future power output and intermittency of more types of PV panels and vice versa using improved models.

Lay Summary

The Sun provides us with free energy which we can convert to usable electricity. We have invented various technologies to enable that conversion, but currently the most common approach is to use solar photovoltaic (PV) panels. However, particulate matter (PM), a mixture of solid particles and liquid droplets, jeopardizes our ability to exploit fully solar energy. Atmospheric PM scatters and absorbs sunlight that would otherwise reach the solar panels. PM deposited on the solar panels further impedes sunlight being received by the solar cells. Atmospheric PM near the surface, particularly those with an aerodynamic diameter $\leq$ 2.5 μm ($PM_{2.5}$), also represents a major threat to human health.

A chemical transport model is a type of computational model which typically simulates atmospheric chemistry and transport, and is widely used to study PM pollution. A solar PV performance model is a type of computational model which simulates the performance of PV energy systems. Usually, these two categories of models are developed, calibrated, validated, and used in their respective fields. Nonetheless, they, once well-evaluated against observations, could be integrated to explore PM impacts on PV power output and more importantly how to reduce these impacts. On the other hand, chemical transport models can be potentially combined with statistical and machine learning models to produce a value-added ground-level $PM_{2.5}$ data product. The overarching goal of my thesis is to use data

and models to understand PM pollution and PV power output with a focus on their links.

I begin by linking a chemical transport model to a PV performance model. I demonstrate that the integrated model can well reproduce observed variations in surface solar radiation and levels of atmospheric and deposited PM with relatively small bias and errors. I thereby have confidence to run the integrated model to provide ten years of simulations from solar radiation to PV power output with and without the influence of atmospheric and/or deposited PM. I find that PV power output over desert regions including the Sahara, Arabian-Peninsula, Central Asia, and Southern South America are most vulnerable to PM pollution. The followed are coastal regions such as countries around the Caribbean and the Mediterranean, and over New-Zealand. The highly polluted regions such as North China and the Indo-Gangetic Plain are also significantly affected.

By perturbing the integrated model, I then focus on developing strategies to mitigate PM damages to PV power output. I find that deep cuts in residential emissions substantially benefit Asian PV power output, with a 50% reduction in residential emissions leading to an additional 7,687 GWh yr^{-1} (US\$653 million yr^{-1}) and 1,823 GWh yr^{-1} (US\$144 million yr^{-1}) produced in China and India, respectively. I also find that routine sweeping of panels is effective at improving PV power output, and that even an annual sweeping routine could lead to an approximately 40% PV power output recovery in Central Asia and Southern South America.

I subsequently go a step further by fully quantifying the benefits to air quality and PV power output from reducing residential fuel emissions over Asian domain during a wintertime (January 1 to February 29, 2008). I identify Eastern and Northeastern China, and the Indo-Gangetic Plain as the three key regions where

reducing residential particularly solid biofuel emissions leads approximately linearly to reductions in PM pollution and improvements in PV power output. I attribute these approximately linear benefits to the large volume of carbon emissions, primarily owing to the low combustion and thermal efficiencies of residential devices and the general absence of any end-of-pipe controls, that typically form into carbon aerosol in an approximately linear way in the atmosphere after aggregating to temporal and spatial mean values.

I finally turn to improving current methods of estimating spatially continuous maps of ground-level $PM_{2.5}$ that is most relevant to human health. I develop a framework to combine chemical transport, statistical, and machine learning models by maximizing their merits via consideration of the suitability of satellite observations in order to improve inferred estimates of ground-level $PM_{2.5}$. In comparison to conventional models, I find that my method can effectively reduce bias in inferred estimates of ground-level $PM_{2.5}$. It can also capture more variations in ground-level $PM_{2.5}$.

In closing, the above findings further understanding of PM pollution and PV power output particularly their links. Looking into the future, studies can focus on evaluating impacts of real emission controls and climate change on past, current, and future power output and intermittency of more types of PV panels and vice versa using improved models.

Acknowledgements

After a whole summer of fighting with my keyboard, I finally come to the most exciting part that I would like to write from the bottom of my heart.

Foremost I would like to thank the the China Scholarships Council/University of Edinburgh Scholarships for giving me this great opportunity to get deeply involved in a field that I am really interested in. The scholarships actually provides financial funding for three years, so I also owe many thanks to the School of GeoSciences, the University of Edinburgh for giving me an additional financial support lasting half a year to compensate for the slightly delayed work unfortunately due to the COVID-19 pandemic.

I first met with my supervisor, Professor Paul Palmer, when our group socialised in an Edinburgh restaurant. I clearly remember how I felt nervous and awkward at the time, particularly when I spared no effort to remember the names and the faces of the warmhearted group members. My journey then started. Paul, efficient, energetic, enthusiastic, and full of wisdom, is definitely on the top list of my respected scholars. His constant support, both in academy and life, led me through countless barriers, and eventually to where I am now. His supervision and care, are absolutely precious to me. I do not feel the word "thanks" is of any help to express my deep appreciation here. I also clearly remember that I told Paul that I would like to become a good scientist like him during our group

trip to Highland in 2019, and importantly this idea has not changed but instead further enhanced since then.

I do not think I would have managed to finish this thesis without the help from the scientific communities. In particular, I would like to express my sincere gratitude to the GEOS-Chem Support Team who could always get back to me very quickly regarding any issues I have. I must admit that some of my questions might sound a little bit 'stubborn', but they consistently provide me with nice replies.

I will never be able to thank my parents enough for their endless love and support. I think they are doing a quite good job by not blaming me too much in the context of my peers getting married and settling down. Ten years have passed since I began my college life in 2011, but for a better career development, I need and would like to stay abroad for an additional period to obtain some working experiences. That being said, I strongly believe that I can now better support and care them than I could before.

Finally, I would like to extend my sincere gratitude to all the folks I have made friends with through this journey. It is them who made this journey so much fun. The list of the names unfortunately ended up too long to be appropriately presented here, but wholeheartedly, thank you.

@ 2 Woodstock Place, Edinburgh EH16 6AY
Friday, 27 August 2021

Contents

Declaration iii

Abstract v

Lay Summary ix

Acknowledgements xiii

List of Figures xix

List of Tables xxvii

Acronyms xxxi

1 Introduction 1
- 1.1 Motivation and scientific questions . . . 1
- 1.2 PM pollution . . . 4
  - 1.2.1 Sources and sinks of PM pollution . . . 4
  - 1.2.2 Measures of PM pollution . . . 6
  - 1.2.3 Measuring PM pollution from ground and space . . . 9
  - 1.2.4 Modelling PM pollution . . . 21
- 1.3 Photovoltaics . . . 29
  - 1.3.1 PV basics . . . 29
  - 1.3.2 Factors affecting PV performance . . . 32
  - 1.3.3 Modelling PV performance . . . 35
- 1.4 Thesis layout . . . 40

2 Global reduction of photovoltaic power output due to particulate matter pollution 41
- 2.1 Abstract . . . 42
- 2.2 Introduction . . . 42
- 2.3 Data and methods . . . 45
  - 2.3.1 GEOS-Chem model coupled with rapid radiative transfer model for GCMs . . . 46
  - 2.3.2 Linking GCRT to PVLIB-Python . . . 49

2.3.3 PVLIB-Python model 50
2.3.4 Experimental design 51
2.3.5 Model evaluation 53
2.4 Results 54
2.4.1 Results of model evaluation 54
2.4.2 Results of PV efficiency and PM impacts 61
2.4.3 Results of PM dimming versus soiling impacts 62
2.4.4 Role of precipitation 66
2.5 Discussion and concluding remarks 67

3 Deep cuts in residential emissions substantially benefit Asian photovoltaic power output 71
3.1 Abstract 72
3.2 Introduction 72
3.3 Methods 74
3.4 Results 76
3.4.1 Benefits of reducing emissions 76
3.4.2 A clear priority of sectors to target in South Asia 85
3.4.3 Benefits of sweeping panels 91
3.5 Discussion and concluding remarks 92

4 Approximately linear benefits to Asian wintertime air quality and photovoltaic power output from reducing residential fuel emissions 99
4.1 Abstract 100
4.2 Introduction 100
4.3 Methods 102
4.4 Results 104
4.4.1 Air quality improvements 104
4.4.2 PV efficiency improvements 109
4.5 Discussion and concluding remarks 115

5 A model framework to reduce bias in ground-level $PM_{2.5}$ concentrations inferred from satellite-retrieved AOD 119
5.1 Abstract 120
5.2 Introduction 121
5.3 Data and methods 123
5.3.1 Ground-level $PM_{2.5}$ measurements 123
5.3.2 MODIS MAIAC and AERONET AOD retrievals 123
5.3.3 Framework 127
5.3.4 GEOS-Chem Model of Atmospheric Chemistry and Transport 128
5.3.5 Data clustering 130
5.3.6 Data suitability 135
5.3.7 Data-driven model development 136

5.3.8 Data-driven model validation and comparison 138
5.3.9 Ground-level $PM_{2.5}$ mapping 139
5.4 Results 140
5.4.1 Results of data clustering 140
5.4.2 Results of data suitability 142
5.4.3 Results of data-driven model development, validation, and comparison 143
5.4.4 Ground-level $PM_{2.5}$ maps 146
5.5 Discussion and concluding remarks 149

6 Conclusions and prospects **151**
6.1 Thesis overview 151
6.2 Summary of key findings 152
6.3 Significance and implications 154
6.4 Suggestions for future research 156

Appendices **159**
A: Dry deposition calculations in GEOS-Chem "Classic" 159
B: Estimating DNI and DHI from GHI using the Erbs model 163
C: Transposing irradiance to solar panels 165
D: Calculations of DC, AC, and CF in PVLIB-Python model 166

References **168**

List of Figures

1.1 Extinction of incident parallel beam solar radiation as it passes through an infinitesimally thin atmospheric layer containing scattering and absorbing PM. The figure is taken from Wallace and Hobbs (2006). 7

1.2 Schematic of a) downward direct solar radiation; b) downward diffuse solar radiation; c) multiple times scattered solar radiation due to the trapping mechanism; d) upward direct solar radiation; e) upward diffuse solar radiation; and f) path radiance. θ_s and θ_v denote solar and view zenith angle, respectively. I_λ is the monochromatic intensity at wavelength λ at the top of atmosphere. 11

1.3 The main satellite platforms and sensors for global PM observations. The figure is taken from Zhang et al. (2021). 18

1.4 The number of papers published on the topic of ground-level $PM_{2.5}$ estimations inferred from satellite-retrieved AOD in the past 17 years. The figure is plotted using Matplotlib with data from Web of Science by searching "$PM_{2.5}$" AND "AOD" AND "Satellite". . 19

1.5 Schematic of a solar cell, module, and array. The figure is taken from SamlexAmerica (2021). 30

1.6 Diagrams of (a) a typical non-grid-tied PV system and (b) a typical grid-tied PV system. The figure is adapted from SamlexAmerica (2021). 31

1.7 Interaction between atmospheric PM and solar radiation. The figure is taken from Ekins-Daukes and Kay (2019). 33

1.8 Interaction between deposited PM and solar radiation. The figure is taken from Bergin et al. (2017). 34

1.9 Workflow to estimate solar photovoltaic energy generation efficiency in this study. Full definitions for abbreviations include: 1) global horizontal irradiance (GHI) simulated from GEOS-Chem coupled with rapid radiative transfer model for GCMs (RRTMG), and it is subsequently decomposed to direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI); 2) beam (E_b), ground-reflected (E_g), and sky-diffuse (E_d) components of transposed irradiance; 3) PM gravitational (V^g) and turbulent (V^t) velocities; 4) PM dry mass concentrations (C^{dm}) and PM mass extinction coefficients (Q^m); 5) direct (DC) and alternating (AC) current powers. 36

2.1 Workflow to estimate solar photovoltaic energy generation efficiency in this study. Full definitions for abbreviations include: 1) global horizontal irradiance (GHI) simulated from GEOS-Chem coupled with rapid radiative transfer model for GCMs (RRTMG), and it is subsequently decomposed to direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI); 2) beam (E_b), ground-reflected (E_g), and sky-diffuse (E_d) components of transposed irradiance; 3) PM gravitational (V^g) and turbulent (V^t) velocities; 4) PM dry mass concentrations (C^{dm}) and PM mass extinction coefficients (Q^m); 5) direct (DC) and alternating (AC) current powers. 46

2.2 Definitions of 46 AR6 land regions used in this thesis to calculate regional area-weighted mean variables. 52

2.3 Site-scale evaluation of simulated GHI against BSRN GHI observations between 2008 and 2017. Pearson correlation coefficients with $p \geq 0.05$ are marked as grey dots. 56

2.4 Site-scale evaluation of simulated AOD against AERONET GHI observations between 2008 and 2017. Pearson correlation coefficients with $p \geq 0.05$ are marked as grey dots. 57

2.5 Site-scale evaluation of simulated $PM_{2.5}$ against CNEMC and OpenAQ $PM_{2.5}$ observations between 2008 and 2017. Pearson correlation coefficients with $p \geq 0.05$ are marked as grey dots. . . 58

2.6 Evaluation of simulated $PM_{2.5}$ chemical composition concentrations against ground-based observations collected from the literature (Zhang et al., 2019). The solid line corresponds to the 1:1 line, and the dashed lines correspond to the 1:2 and 2:1 lines. Statistically significant ($p < 0.05$) Pearson correlation coefficients are marked with a $*$. The bottom right subplot shows the number of observations of at least one $PM_{2.5}$ chemical component over each monitoring location. 59

2.7 Geographical distributions of decadal mean (2008–2017) (a–c) PV efficiency and (d–f) its losses due to atmospheric and deposited PM for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panel. 62

2.8 Comparison of regional area-weighted averages of decadal mean (2008–2017) PV efficiency and PM total impacts for (a) flat, (b) tilt, and (c) one-axis tracking panels. Please refer to Figure 2.2 for definitions of regions that are used here for taking regional area-weighted averages. 63

2.9 Geographical distributions of decadal mean (2008–2017) PV efficiency losses due to (a–c) atmospheric and (d–f) deposited PM for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels. 64

2.10 Comparisons of regional area-weighted averages of decadal mean (2008–2017) PM dimming and soiling impacts for (a) flat, (b) tilt, and (c) one-axis tracking panels. The blue lines are the regression lines between PM dimming and soiling impacts, with R denoting the Pearson correlation coefficients. The red lines are the 1:1 lines. The vertical colorbar represents PM total impacts. The percentages following E.Asia and S.Asia texts are the non-negligible shares of PM dimming impacts over these two regions. Please refer to Figure 2.2 for definitions of regions that are used here for taking regional area-weighted averages. 65

2.11 Geographical distributions of decadal mean (2008–2017) deposition fluxes ($\mu g/m^2/hr$) for each PM species: organic aerosol (OA), black carbon (BC), sea salt (SS), dust (DU), and sulfate-nitrate-ammonium (SNA) at 2° latitude × 2.5° longitude resolution. . . . 66

2.12 Frequency of precipitation rates of < 1 mm/hr, 1–3 mm/hr, 3–5 mm/hr, and > 5 mm/hr from MERRA-2 between 2008 and 2017. 67

2.13 Geographical distributions of decadal mean (2008–2017) cleaning benefits resulting from precipitation for (a) flat, (b) tilt, and (c) one-axis tracking panels. 68

2.14 Geographical distributions of decadal mean (2008–2017) PV efficiency losses due to (a–c) deposited PM updated with the equal weights for POAI_{out} and (d–f) its changes from Figure 2.9d–f for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels. 70

3.1 Illustration of strategy used in this chapter to identify leading source sectors where reducing emissions greatly benefit photovoltaic power output. 75

3.2 Geographical distributions of (a-c) source sectors from which halving emissions provide (d-f) maximum decadal mean (2008–2017) brightening benefits for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels. Full definitions for source sectors are non-combustion agriculture (AGR), energy generation (ENE), industrial processes (IND), on-road (ROAD) and non-road (NRTR) transportation, separate residential (RCOR), commercial (RCOC), and other (RCOO) sectors, waste (WST), solvent use (SLV), and international shipping (SHP). 77

3.3 Geographical distributions of (a-c) source sectors from which halving emissions provide (d-f) maximum decadal mean (2008–2017) cleaning benefits for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels. Full definitions for source sectors are non-combustion agriculture (AGR), energy generation (ENE), industrial processes (IND), on-road (ROAD) and non-road (NRTR) transportation, separate residential (RCOR), commercial (RCOC), and other (RCOO) sectors, waste (WST), solvent use (SLV), and international shipping (SHP). 80

3.4 Geographical distributions of (a-c) source sectors from which halving emissions provide (d-f) maximum decadal mean (2008–2017) total benefits for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels. Full definitions for source sectors are non-combustion agriculture (AGR), energy generation (ENE), industrial processes (IND), on-road (ROAD) and non-road (NRTR) transportation, separate residential (RCOR), commercial (RCOC), and other (RCOO) sectors, waste (WST), solvent use (SLV), and international shipping (SHP). 83

3.5 Geographical distributions of decadal mean (2008–2017) cleaning benefits resulting from various frequencies of panel sweeping for flat, tilt, and one-axis tracking panels. 92

3.6 Regional-area-weighted averages of decadal mean (2008–2017) percent cleaning benefits (relative to PM soiling impacts) resulting from various frequencies of panel sweeping, in several regions of interest. Please refer to Figure 2.2 for definitions of regions that are presented here. 93

3.7 Consistent benefits to PV efficiency throughout years particularly those from stringent residential emission reductions over Asia with respect to the proportion of occupied areas. Please refer to Figure 2.2 for definitions of regions that are presented here. 94

3.8 Geographical distributions of (a) sectors from which halving emissions provide (b) maximum decadal mean (2008–2017) co-benefits to surface air quality with respect to PM with an aerodynamic diameter $\leq$ 2.5 μm. Full definitions for source sectors are non-combustion agriculture (AGR), energy generation (ENE), industrial processes (IND), non-road (NRTR) and on-road (ROAD) transportation, separate residential (RCOR), commercial (RCOC), and other (RCOO) sectors, waste (WST), solvent use (SLV), international shipping (SHP), and agricultural waster burning (AWB). 95

4.1 Mean spatial distribution of ground-level $PM_{2.5}$ concentrations and column-integrated AOD_{550nm} over Asian domain from January 1 to February 29, 2008 in the baseline simulation. The three key regions with high PM pollution levels are marked with red polygons: (A) Eastern China, (B) Northeastern China, and (C) Indo-Gangetic Plain. 105

4.2 Mean spatial distribution of decreases in ground-level $PM_{2.5}$ concentrations over Asian domain from January 1 to February 29, 2008 as a result of reducing residential fuel emissions. 106

4.3 Mean spatial distribution of decreases in column-integrated AOD_{550nm} over Asian domain from January 1 to February 29, 2008 as a result of reducing residential fuel emissions. 107

4.4 Absolute benefits to regional-scale air quality from reducing residential fuel emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangeric Plain (IGP). 108

4.5 Percentage benefits (relative the baseline case) to regional-scale air quality from reducing residential fuel emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP). 109

4.6 Mean spatial distribution of PV efficiency and its losses due to PM total, soiling, and dimming impacts over Asian domain from January 1 to February 29, 2008 for flat, tilt, and one-axis tracking panels. 110

4.7 Mean spatial distribution of brightening benefits over Asian domain from January 1 to February 29, 2008 as a result of completely reducing residential fuel emissions for flat, tilt, and one-axis tracking panels. Similar spatial patterns (not shown) are found for other extents (25%, 50%, and 75%) to which residential emissions are reduced. 111

4.8 Mean spatial distribution of cleaning benefits over Asian domain from January 1 to February 29, 2008 as a result of completely reducing residential fuel emissions for flat, tilt, and one-axis tracking panels. Similar spatial patterns (not shown) are found for other extents (25%, 50%, and 75%) to which residential emissions are reduced. 112
4.9 Mean spatial distribution of total benefits over Asian domain from January 1 to February 29, 2008 as a result of completely reducing residential fuel emissions for flat, tilt, and one-axis tracking panels. Similar spatial patterns (not shown) are found for other extents (25%, 50%, and 75%) to which residential emissions are reduced. . 113
4.10 Absolute benefits to regional-scale PV efficiency of flat, tilt, and one-axis tracking panels from reducing residential fuel emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP). 114
4.11 Percentage benefits (relative to the baseline case) to regional-scale PV efficiency of flat, tilt, and one-axis tracking panels from reducing residential fuel emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP). 115
4.12 Total emissions of seven major air pollutants from the residential sector from January 1 to February 29, 2008 over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP). 117
4.13 Absolute reductions in regional-scale levels of (upper) ground-level $PM_{2.5}$ and (bottom) column-integrated AOD_{550nm} major chemical components (POA: primary organic aerosol, SOA: secondary organic aerosol, SNA: sulfate - nitrate - ammonium, BC: black carbon) from reducing residential emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP). 118
4.14 Total emissions of seven major air pollutants from the residential sector from January 1 to February 28, 2017 over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP). 118

5.1 Spatiotemporal window determination for matching MAIAC and AERONET AOD from 2000 to 2018 in China. Solid lines denote the change of statistics and dashed lines denote the change of the sample size. 125
5.2 Density scatterplots of AERONET and (left) Terra and (right) Aqua MAIAC AOD from 2000 to 2018 in China (bottom) with and (top) without quality assurance flags filtering under the spatiotemporal window of 60 minutes and 25 km. 126

5.3 A brief description of the original method of inferring ground-level $PM_{2.5}$ concentrations from satellite observations of AOD, and the revised method that has developed in this study and most importantly takes into account the fraction of columnar AOD that resides in the planetary boundary layer. 127
5.4 Temporal Pearson correlation coefficient (R) (top panel) between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and (bottom panel) between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values using spatially collocated data on a daily basis at the (rows 1 and 3) Terra and (rows 2 and 4) Aqua overpass times in 2014 over eastern China. Gray pixels denote statistically insignificant R values at $\alpha = 0.05$. . . . 131
5.5 Temporal mean percentage error (top panel) between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and (bottom panel) between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values using spatially collocated data on a daily basis at the (rows 1 and 3) Terra and (rows 2 and 4) Aqua overpass times in 2014 over eastern China. 132
5.6 Spatial Pearson correlation coefficient (R) (left panel) between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and (right panel) between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values using daily collocated data on a spatial basis at the (columns 1 and 3) Terra and (columns 2 and 4) Aqua overpass times in 2014 over eastern China. Gray pixels denote statistically insignificant R values at $\alpha = 0.05$. 133
5.7 Spatial mean percentage error (left panel) between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and (right panel) between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values using daily collocated data on a spatial basis at the (columns 1 and 3) Terra and (columns 2 and 4) Aqua overpass times in 2014 over eastern China. 134
5.8 Temporal clusters determined by spatially collocated individual chemical components of AOD simulated from the GEOS-Chem model using the UPGMA hierarchical clustering method at the (top) Terra and (bottom) Aqua overpass times in 2014 over eastern China. That is we calculate the correlation distance between two day of years using their spatially collocated data in step 1 of the UPGMA hierarchical clustering method. 137
5.9 Spatial clusters of ground-level $PM_{2.5}$ monitoring locations over eastern China during 2014 for (left) Terra and (right) Aqua overpass times. Clusters are determined by one year of GEOS-Chem 3-D model fields of individual chemical components of AOD using the UPGMA hierarchical clustering method. 140

5.10 Spatial clusters of ground-level $PM_{2.5}$ monitoring locations over eastern China in each month of 2014 for (top panel) Terra and (bottom panel) Aqua overpass times. Clusters are determined by monthly GEOS-Chem 3-D model fields of individual chemical components of AOD using the UPGMA hierarchical clustering method. Number shown inset denote how many clusters determined for a specific month and satellite overpass time. 142
5.11 (Orange) Pearson and Spearman rank correlation coefficients between ground-level $PM_{2.5}$ and AOD, and the (green) corresponding number of observations, as a function of Γ_{PBL}^{AOD}, in a specific monthly cluster. We use a Γ_{PBL}^{AOD} value of 0.5 (red vertical lines) as a threshold, below which we discard the corresponding data. . . . 143
5.12 Seasonal ground-level $PM_{2.5}$ maps in 2014 over eastern China inferred from the RF2 using the data from the Aqua overpass time at 0.25° latitude × 0.3125° longitude coincide with observed ground-level $PM_{2.5}$ measurements. The first row presents results using the RF2 trained by the full data. The second row presents results using the RF2 trained by the reduced data where $\Gamma_{PBL}^{AOD} \geq 0.5$. The third row presents their differences. Two red circles in the last two subplots highlight areas where large differences are located. 148

List of Tables

1.1 Growth factor values under RH = 35% and 50%. 25

1.2 Solar panel settings currently supported by the PVLIB-Python model. 37

2.1 A summary of GCRT model configurations. 48

3.1 The proportion of occupied areas, and regional-area-weighted mean brightening benefits and PM dimming impacts and their ratios of sectors from which halving emissions provide the largest decadal and corresponding seasonal mean brightening benefits, in various regions of interest. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq$ 75%. Please refer to Figure 2.2 for definitions of regions that are presented here. 78

3.2 The proportion of occupied areas, and regional-area-weighted mean cleaning benefits and PM soiling impacts and their ratios of sectors from which halving emissions provide the largest decadal and corresponding seasonal mean cleaning benefits, in various regions of interest. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq$ 75%. Please refer to Figure 2.2 for definitions of regions that are presented here. 81

3.3 The proportion of occupied areas, and regional-area-weighted mean total benefits and PM total impacts and their ratios of sectors from which halving emissions provide the largest decadal and corresponding seasonal mean total benefits, in various regions of interest. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq$ 75%. Please refer to Figure 2.2 for definitions of regions that are presented here. 84

3.4 The proportion of occupied areas, and regional-area-weighted mean brightening benefits and PM dimming impacts and their ratios of top three sectors from which halving emissions provide the largest decadal and corresponding seasonal mean brightening benefits, in S.Asia. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for the definition of S.Asia that is presented here. 86

3.5 The proportion of occupied areas, and regional-area-weighted mean cleaning benefits and PM soiling impacts and their ratios of top three sectors from which halving emissions provide the largest decadal and corresponding seasonal mean cleaning benefits, in S.Asia. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for the definition of S.Asia that is presented here. 88

3.6 The proportion of occupied areas, and regional-area-weighted mean total benefits and PM total impacts and their ratios of top three sectors from which halving emissions provide the largest decadal and corresponding seasonal mean total benefits, in S.Asia. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for the definition of S.Asia that is presented here. 89

3.7 Potential additional PV electricity bonus (GWh yr^{-1}) and economic benefits (US$ million yr^{-1}) by combining decadal mean (2008–2017) capacity factor improvements (ΔCF) from halving residential emissions with the installed capacities as of 2019 in China. The last row provides the national total values for all variables except ΔCFs. 97

3.8 Potential additional PV electricity bonus (GWh yr^{-1}) and economic benefits (US$ million yr^{-1}) by combining decadal mean (2008–2017) capacity factor improvements (ΔCF) from halving residential emissions with the installed capacities as of 2019 in India. The last row provides the national total values for all variables except ΔCFs. 98

5.1 The fraction of individual chemical componenets to the annual AOD of 2014 in each cluster of eastern China derived from the GEOS-Chem model. 141

5.2 Overall model fitting and cross validation results of the PooledOLS, TFEM, RF1, and RF2 models in 2014 over eastern China. N, R^2, and MPE represent statistics of the model trained by the full data, while N$'$, $R^{2'}$, and MPE$'$ represent those of the model trained by the suitable data. ${R^2}_p$ and MPE_p represent the possibility of obtaining a model performance no worse than that of our approach by randomly selecting a subset of the full data that matches the length of the suitable table to train the model. 145
5.3 Cross validation results of the PooledOLS, TFEM, RF1, and RF2 models in each cluster of eastern China in 2014. 147

Acronyms

AC Alternating Current.
AERONET AErosol RObotic NETwork.
AOD Aerosol Optical Depth.
AOI Angle of Incidence.

BC Black Carbon.
BCPI Hydrophilic BC.
BCPO Hydrophobic BC.
BSRN Baseline Surface Radiation Network.

CEDS Community Emissions Data System.
CEDS$_{GBD-MAPS}$ CEDS updated for the Global Burden of Disease - Major Air Pollutant Sources project.
CF Capacity Factor.
CNEMC China National Environmental Monitoring Center.
CO Carbon Monoxide.
CTM Chemical Transport Model.
CV Cross Validation.

DC Direct Current.
DHI Diffuse Horizontal Irradiance.
DNI Direct Normal Irradiance.

GCM General Circulation Model.
GCRT GEOS-Chem coupled with RRTMG.
GEOS Goddard Earth Observing System.
GEOS-FP GEOS Forward Processing.
GHI Global Horizontal Irradiance.
GMAO Global Modeling and Assimilation Office.

HEMCO Harvard-NASA Emissions Component.
HEMCO Harmonized Emissions Component.

MAIAC Multi-Angle Implementation of Atmospheric Correction.
MERRA-2 Modern-Era Retrospective analysis for Research and Application, Version 2.

MODIS Moderate Resolution Imaging Spectroradiometer.

NASA National Aeronautics and Space Administration.
NMVOC Non-Methane Volatile Organic Compounds.

OA Organic Aerosol.
OC Organic Carbon.
OCPI Hydrophilic OC.
OCPO Hydrophobic OC.

PBL Planetary Boundary Layer.
PM Particulate Matter.
PM_{10} Particles with aerodynamic diameter $\leq$ 10 μm.
PM_1 Particles with aerodynamic diameter $\leq$ 1 μm.
$PM_{2.5}$ Particles with aerodynamic diameter $\leq$ 2.5 μm.
POA Primary OA.
POAI Plane of Array Irradiance.
PV Photovoltaic.

RH relative Humidity.
RRTMG Rapid Radiative Transfer Model for GCMs.

SALA Accumulation mode SS.
SALC Coarse mode SS.
SAPM Sandia Array Performance Model.
SF Soiling Factor.
SNA Sulfate - Nitrate - Ammonium.
SOA Secondary OA.
SS Sea Salt.

TOA Top of Atmosphere.

UPGMA Unweighted Pair Group Method with Arithmetic mean.

Chapter 1

Introduction

1.1 Motivation and scientific questions

The Sun provides us with free, clean energy which we can convert to usable electricity to support our producing and living activities. Various technologies exist to enable that conversion, but currently the most common approach is to use solar photovoltaic (PV) panels. Particulate matter (PM) is a mixture of solid particles and liquid droplets consisting of a variety of components such as organic compounds, metals, acids, soil, and dust (Agency, 1996; Ciencewicki and Jaspers, 2007) that can be suspended in the air or deposited on solar panels. These PM absorb and scatter incoming solar radiation that would otherwise reach the solar panels then the solar cells to be converted to usable electricity. As such, they represent a major barrier for maximizing the solar PV performance (Sarver et al., 2013; Bergin et al., 2017; Peters et al., 2018; Sweerts et al., 2019; Li, Wagner, Peng, Yang and Mauzerall, 2017; Li et al., 2020). Atmospheric PM near the surface, particularly those with aerodynamic diameter ≤ 2.5 μm ($PM_{2.5}$), also

represent a major threat to human health as well-documented by epidemiological evidences (Lelieveld et al., 2015; Burnett et al., 2018).

To reduce PM damage to solar PV energy generation, we must identify the major sources of PM pollution severely affecting solar PV energy generation. This, however, remains unresolved, despite being urgently needed. Accurately capturing magnitudes and variabilities of ground-level $PM_{2.5}$ concentrations is the basis for understanding its impacts on human health, but to date reliable, physically-relevant methods have yet to be well established. In this thesis, the scientific questions that I address mainly focus on two aspects: 1) integrating models of atmospheric composition and solar PV performance to quantify PM impacts on PV power output and more importantly develop strategies to mitigate these impacts, and 2) integrating models of atmospheric composition, statistics, and machine learning to produce a value-added ground-level $PM_{2.5}$ data product. To do this, I integrate different models in two ways: 1) taking one model's outputs as inputs of another model and 2) using one model's outputs to enlighten the construction of another model. Moreover, I use a series of satellite observations and *in situ* measurements to facilitate the integration.

The main scientific questions that I address through this thesis are:

1. **What are the impacts of PM pollution on PV power output and how do we mitigate these impacts?**
   (a) To what extent does PM pollution damage PV power output and what are the relative importance of atmospheric and deposited PM?
   (b) What is the relative importance of reducing emissions from individual anthropogenic source sectors for improving PV power output?
   (c) To what extent can we reduce PM damage to PV power output by routine sweeping of PM deposited on solar panels?
   (d) Over highly polluted regions, what are the relative importance of

reducing emissions from individual fuels within the sector where reducing emissions mostly improves PV power output?

2. **How do we produce a value-added ground-level $PM_{2.5}$ data product by using data and models?**

   (a) How do we improve the robustness of ground-level $PM_{2.5}$ estimates by integrating models of atmospheric composition, statistics, and machine learning?

   (b) How do we assess the improved robustness of ground-level $PM_{2.5}$ estimates?

Broadly speaking, I address the first of these questions in Chapters 2–4, and the second in Chapter 5. Spatially, I examine these questions from global to regional scales: a global domain with a spatial resolution of 2° latitude × 2.5° longitude in Chapters 2 and 3, an Asian domain with a spatial resolution of 0.5° latitude × 0.625° longitude in Chapter 4, and eastern China domain (95–140°E and 15–55°N) with a spatial resolution of 0.25° latitude × 0.3125° longitude in Chapter 5. I zoom into Asia and eastern China mainly because of the severe PM pollution I identify here. Temporally, I examine these questions from decadal to daily scales: decadal scale in Chapters 2 and 3, seasonal scale in Chapter 4, and daily scale in Chapter 5. I choose these varying spatial and temporal scales to keep a good balance of analysis requirements and computational costs.

In this chapter, I begin with a brief introduction to PM pollution in section 1.2, including its sources and sinks, measures and associated measurements from ground and space, and how it is modelled. I then describe solar PV energy systems in section 1.3, with an emphasis on PV basics, factors affecting PV performance, and how PV performance is modelled. I close this chapter by describing the thesis layout in section 1.4.

1.2 PM pollution

I refer to PM as a mixture of solid particles and liquid droplets that can be suspended in the air or deposited on solar panels in section 1.1. In this section, I follow the mainstream practice to limit PM pollution to atmospheric PM only, and assume PM deposition on solar panels to be a loss process. Below I mainly describe PM sources and sinks, measures and available ground and space measurements, and process modelling.

1.2.1 Sources and sinks of PM pollution

PM either has natural or anthropogenic sources and is either emitted as primary particles or formed by secondary processes (Fuzzi et al., 2015). Natural sources of PM include desert dust, sea salt, volatile organic compounds from vegetation, wildfires, volcanic eruptions, and others. Anthropogenic sources of PM include burning fuels, industrial and agricultural activities, waste incineration, and others. Primary PM is directly released to the atmosphere from sources such as incomplete combustion and desert dust. Secondary PM is formed in the atmosphere almost exclusively by condensation of low volatility gases such as sulfuric acid, ammonia and functionalised organic compounds onto pre-existing PM. While PM largely originates from natural particularly primary sources on the global scale (Andreae, 1995), the contribution from anthropogenic particularly secondary sources is significant in populous urban and surrounding regions (Wang et al., 2016, 2020, 2021).

Primary PM dominates the "coarse" fraction of the PM (the particles with diameter greater than 1 μm). Desert dust, sea salt, and volatile organic compounds from vegetation typically have radius of 1–10 μm. This is simply because particles finer than 1 μm are difficult to generate mechanically due to

their large area-to-volume ratios, while particles coarser than 10 μm are difficult to lift but easy to sediment.

Primary PM is mainly removed by dry and wet deposition. Dry deposition is the removal of PM by sedimentation under gravity, diffusion processes (i.e. Brownian motion), or by turbulent transfer resulting in impaction and interception (Beckett et al., 1998). Wet deposition is the washout of PM during precipitation (but may also occur during dew formation, mists, fog, and snow), which will be dependent on the air–water partition coefficient and the particle scavenging efficiency of precipitation, respectively (Barber et al., 2004).

Secondary PM constitutes most of the "fine" PM (particles with diameter typically below 1 μm). This is mainly due to its origin from condensation of low volatility gases (Jacob, 1999). Generally, clustering of gas molecules typically in the 10^{-4}–10^{-3} μm size range (nucleation) produces ultrafine PM in the 10^{-3}–10^{-2} μm size range. These ultrafine PM grow rapidly to the 0.01–1 μm fine PM size range by condensation of gases and by coagulation (collisions between particles during their random motions). Growth beyond 1 μm is much slower because large particles formed by then have slower growth rate by condensation of gases and slower random motion that reduces the coagulation rate. As a consequence, secondary PM that almost originates from condensation of gases tend to accumulate in the 0.01–1 μm size range.

Similar to primary PM, dry and wet deposition are main removal mechanisms for secondary PM. The sole difference is that secondary PM is too small to sediment at a significant rate under gravity and thus turbulent deposition likely controls the dry deposition process.

1.2.2 Measures of PM pollution

One of most widely-used measures for non-gaseous PM pollutants are the number and mass concentrations. The number concentration n_i of a PM species i is defined as the number of molecules of that species per unit volume of air. It is commonly expressed in units of molecules cm^{-3}. The mass concentration C_i^m of a PM species i is defined as the mass of that species per unit volume of air. It is commonly expressed in units of $\mu\text{g m}^{-3}$ or mg m^{-3}. Per appropriate unit conversion, the number and mass concentrations can be linked by equation 1.1.

$$C_i^m = \frac{n_i}{A_v} \times MW_i, \tag{1.1}$$

where $A_v = 6.022 \times 10^{23}$ molecules mol^{-1} is the Avogadro constant, and MW_i is the molecular weight of PM species i.

As mentioned in section 1.1, PM scatters and absorbs solar radiation. Hence, another widely-adopted measure for PM is the extent to which PM scatters and absorbs solar radiation. Figure 1.1 depicts a beam of solar radiation passing through an infinitesimally thin atmospheric layer which contains scattering and absorbing PM. For each PM species i that the beam encounters, its monochromatic intensity at wavelength λ at that point along the ray path (I_λ) experiences a decrease ($\text{d}I_\lambda$) that is proportional to 1) I_λ, 2) the local abundance of PM species i, and 3) the effectiveness of the scattering and absorption caused by PM species i. For illustration, I simply assume a fixed radius for PM species i and use equation 1.2 to formulate this incremental decrease:

$$\text{d}I_\lambda = -I_\lambda K_\lambda^i \sigma_i n_i \text{d}s, \tag{1.2}$$

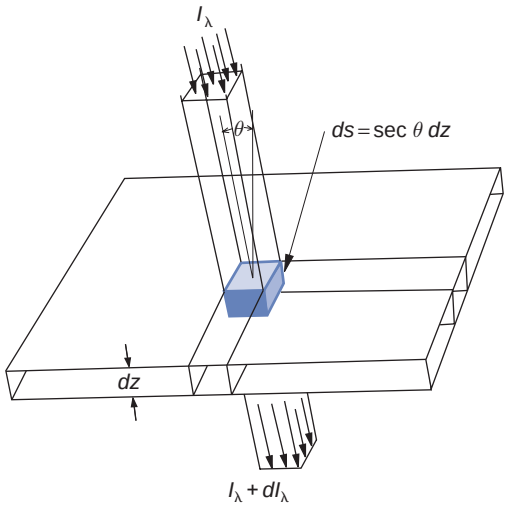


Figure 1.1: Extinction of incident parallel beam solar radiation as it passes through an infinitesimally thin atmospheric layer containing scattering and absorbing PM. The figure is taken from Wallace and Hobbs (2006).

where K_λ^i is the combined, dimensionless scattering and absorption efficiency of PM species i for wavelength λ, σ_i is the areal cross section of PM species i, and thus $K_\lambda^i \sigma_i$ is the combined scattering and absorption cross section of PM species i for wavelength λ, or, in other words, the effectiveness of the scattering and absorption caused by PM species i. In practice, we generally term the combined scattering and absorption as extinction, so K_λ^i and $K_\lambda^i \sigma_i$ are also called the extinction efficiency and extinction cross section of PM species i for wavelength λ, respectively. ds is the differential path length along the ray path of the incident radiation, and thus $n_i \mathrm{d}s(\times \frac{1m^2}{1m^2}) = \mathrm{d}n_i$ is the number of molecules of PM species i per unit area of the thin layer.

Equation 1.2 is readily solved by separation of variables:

$$\frac{\mathrm{d}I_\lambda}{I_\lambda} = -K_\lambda^i \sigma_i n_i \mathrm{d}s, \tag{1.3}$$

for which we integrate both sides over the altitude interval $[z_{in},\, z_{out}]$ to obtain:

$$\ln I_\lambda\Big|_{z_{in}}^{z_{out}} = \int_{z_{in}}^{z_{out}} -K_\lambda^i \sigma_i n_i \mathrm{d}s, \tag{1.4}$$

which gives:

$$I_\lambda\Big|_{z_{out}} = I_\lambda\Big|_{z_{in}} \times e^{-\int_{z_{in}}^{z_{out}} K_\lambda^i \sigma_i n_i \mathrm{d}s}, \tag{1.5}$$

from which we define the optical depth of PM species i for wavelength λ from z_{in} to z_{out} as:

$$\tau_\lambda^i\Big|_{z_{in}}^{z_{out}} = \int_{z_{in}}^{z_{out}} K_\lambda^i \sigma_i n_i \mathrm{d}s, \tag{1.6}$$

so that equation 1.5 can be re-written compactly as:

$$I_\lambda\Big|_{z_{out}} = I_\lambda\Big|_{z_{in}} \times e^{-\tau_\lambda^i\Big|_{z_{in}}^{z_{out}}} \tag{1.7}$$

It is convenient to define $K_\lambda^i \sigma_i n_i$ as the extinction coefficient, typically in units of m^{-1}, of PM species i for wavelength λ such that $\tau_\lambda^i\Big|_{z_{in}}^{z_{out}}$ denotes the integral of the extinction coefficient over the certain path from z_{in} to z_{out}. $\tau_\lambda^i\Big|_{z_{in}}^{z_{out}} = 1$ just means that I_λ falls to $1/e \approx 37\%$ of its initial value due to PM species i distributed between z_{in} and z_{out}.

Using $n_i \mathrm{d}s = \mathrm{d}n_i = \frac{\mathrm{d}m_i}{M_i} \times A_v$, where $\mathrm{d}m_i$ is the mass of PM species i per unit area of the thin layer, we can further write equation 1.6 to:

$$\tau_\lambda^i \Big|_{z_{in}}^{z_{out}} = \int_{z_{in}}^{z_{out}} K_\lambda^i \sigma_i \frac{A_v}{M_i} \mathrm{d}m_i, \tag{1.8}$$

from which we define $K_\lambda^i \sigma_i \frac{A_v}{M_i}$ as the mass extinction coefficient, typically in units of $\mathrm{m}^2\ \mathrm{g}^{-1}$, of PM species i for wavelength λ such that τ_λ^i can be also interpreted as the integral of the mass extinction coefficient over a certain "mass" path.

1.2.3 Measuring PM pollution from ground and space

Here I describe ground and satellite observations for the measures defined above. I focus on PM mass concentrations and PM optical depth that are extensively used throughout this thesis.

With respect to ground measurements of PM mass concentrations, routine monitoring is largely concentrated on ground-level concentrations of PM with aerodynamic diameter $\leq$ 10 μm (PM_{10}), $\leq$ 2.5 μm ($PM_{2.5}$), and $\leq$ 1 μm (PM_1) that are mostly relevant to human health. Generally, the smaller the threshold diameter of PM, the deeper their penetration into the human body, but the fewer resources allocated to its routine monitoring. For instance, as of 2018, China hosts more than 1,500 monitors for PM_{10} and $PM_{2.5}$, but less than 100 monitors for PM_1. Regarding the established monitoring networks, developed countries have obviously higher monitor densities than developing countries likely due to their stronger affordability for high cost of purchasing, operating, and maintaining monitors. For instance, Martin et al. (2019) revealed that North American countries have more than three $PM_{2.5}$ monitors per million inhabitants, as opposed to an average $PM_{2.5}$ monitor density of 0.03 per million inhabitants in Africa. More extensive sampling and analysis details are also reported from the North American networks than those from other regions (McDuffie et al., 2021).

To harmonize disparate air quality data including ground PM measurements from across the world and to make them readily available in a near real-time manner, initiatives such as OpenAQ (`https://openaq.org/`) emerge and play a role in fighting air inequality. Meanwhile, rapidly developed low-cost sensors are emerging with exciting prospects to augment regulatory monitoring networks (Snyder et al., 2013; Gao et al., 2015; Jiao et al., 2016; Crawford et al., 2021). Quality screening is suggested prior to using ground PM measurements from regulatory networks (Liu, Li and Li, 2016) and this should spontaneously apply to those collected from low-cost sensors. PM chemical component measurements are currently limited to research-oriented campaigns.

Ground and satellite observations of PM optical depth both involve modelling of atmospheric effects, so I give a joint description below.

Figure 1.2 depicts the various processes occurred when a beam of solar radiation passes through the atmosphere, reaches ground sensors, interacts with the surface, passes through the atmosphere again, and finally reaches satellite sensors. No cloud is assumed to exist during these processes. Following and extending the practice in Figure 1.1, I term the monochromatic intensity at wavelength λ at the top of atmosphere (TOA) as I_λ, and consider the atmosphere composed of many infinitesimally thin layers. In this sense, I focus on the retrieval of columnar PM optical depth τ_λ^a, which is widely called aerosol optical depth (AOD) partly due to the interchangeability of PM and aerosol. I use θ_s and θ_v as the solar zenith angle and the view zenith angle, respectively, with their cosines written as μ_s and μ_v, respectively. I use ϕ to denote the azimuth of the scattered radiation from the solar beam (not shown in Figure 1.2). Below I describe the fate of I_λ as it passes through the atmosphere and interacts with the surface and how we can use that information to retrieve columnar τ_λ^a from ground and satellite sensors.

A specific ground target receives solar radiation directly and indirectly. Based on

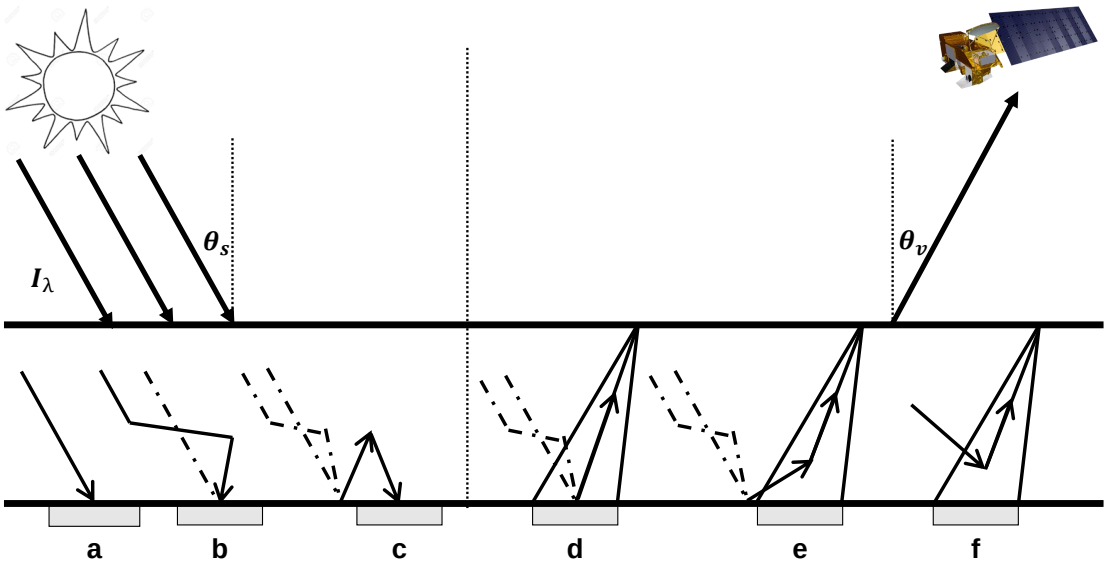


Figure 1.2: Schematic of a) downward direct solar radiation; b) downward diffuse solar radiation; c) multiple times scattered solar radiation due to the trapping mechanism; d) upward direct solar radiation; e) upward diffuse solar radiation; and f) path radiance. θ_s and θ_v denote solar and view zenith angle, respectively. I_λ is the monochromatic intensity at wavelength λ at the top of atmosphere.

the analysis of Figure 1.1, we can simply write the downward direct solar radiation attenuated by the atmosphere (I_λ^{dir}, Figure 1.2a) as:

$$I_\lambda^{dir} = \mu_s I_\lambda e^{-\frac{\tau_\lambda}{\mu_s}}, \tag{1.9}$$

where τ_λ is the total optical depth of the atmosphere and is the sum of optical depth from PM, molecular scattering, and others (e.g. gases such as O_3 and NO_2), i.e. $\tau_\lambda = \tau_\lambda^a + \tau_\lambda^m + \tau_\lambda^o$

Rearranging equation 1.9 gives τ_λ:

$$\tau_\lambda = -\mu_s \ln \frac{I_\lambda^{dir}}{\mu_s I_\lambda}, \tag{1.10}$$

from which τ_λ^a is derived:

$$\tau_\lambda^a = -\mu_s \ln \frac{I_\lambda^{dir}}{\mu_s I_\lambda} - \tau_\lambda^m - \tau_\lambda^o \tag{1.11}$$

This constitutes the basis of retrieving τ_λ^a from ground sensors that typically measure I_λ^{dir}. To calculate τ_λ^a using equation 1.11, all terms on the right side of the formula must be known. I_λ is actually the solar constant. τ_λ^m can be determined from equation 1.12. τ_λ^o can be derived from the total precipitatable water vapour information or equals zero in channels unaffected by water vapour absorption (e.g. 440, 870, and 1020 nm).

$$\tau_\lambda^m = A\lambda^{-B}, \tag{1.12}$$

where A and B are constants to be determined from a power-law fit and the equation is normalized to 1013.25 hPa pressure. An example was given by Dutton et al. (1994), who performed such a fit over the visible range and provided the equation:

$$\tau_\lambda^m = \frac{p}{p0} \times 0.0088\lambda^{-4.05}, \tag{1.13}$$

where p and $p0$ are actual and standard (1013.25 hPa) atmospheric pressures.

Ground-based retrievals of τ_λ^a from the National Aeronautics and Space Administration (NASA) AErosol RObotic NETwork (AERONET) are widely recognized

as the 'ground truth' to study spatiotemporal varibilities of τ_λ^a and to validate satellite-retrieved τ_λ^a (Holben et al., 1998). With the recognition of the significance of PM in climate and environment, more and more similar local networks are emerging. These include the European aerosol Lidar Network (Pappalardo et al., 2014), Canadian Sun Photometer Network (Bokoye et al., 2001), China Aerosol Remote Sensing Network (Che et al., 2009, 2015, 2019), Chinese Sun Hazemeter Network (Xin et al., 2007), Campaign on Atmospheric Aerosol Research Network of China (Xin et al., 2015), Sun–Sky Radiometer Observation Network (Li et al., 2018), and others.

The downward diffuse solar radiation (I_λ^{diff}, Figure 1.2b) is independent from surface properties and is noted by a diffuse transmittance factor $t_d(\theta_s)$:

$$I_\lambda^{diff} = \mu_s I_\lambda t_d(\theta_s) \tag{1.14}$$

The ground target also receives a second, third, and multiple times scattered solar radiation due to the trapping mechanism (I_λ^{multi}, Figure 1.2c). This is dependent on the environment of the ground target and corresponds to the successive reflections and scattering between the surface and the atmosphere. If the spherical albedo of the atmosphere is noted S_λ and ρ_λ^t is the surface reflectance averaged on the view and illumination angles, we can write this term as:

$$I_\lambda^{multi} = (I_\lambda^{dir} + I_\lambda^{diff}) \times [\rho_\lambda^t S_\lambda + (\rho_\lambda^t S_\lambda)^2 + ... + (\rho_\lambda^t S_\lambda)^n], \tag{1.15}$$

where $n \to \infty$.

Combining I_λ^{dir}, I_λ^{diff}, and I_λ^{multi}, we obtain the total radiation (I_λ^{tot}) received by the ground target:

$$I_\lambda^{tot} = (I_\lambda^{dir} + I_\lambda^{diff}) \times [1 + \rho_\lambda^t S_\lambda + (\rho_\lambda^t S_\lambda)^2 + ... + (\rho_\lambda^t S_\lambda)^n], \tag{1.16}$$

which can be readily simplified to:

$$I_\lambda^{tot} = (I_\lambda^{dir} + I_\lambda^{diff}) \times \frac{1 - (\rho_\lambda^t S_\lambda)^{(n+1)}}{1 - \rho_\lambda^t S_\lambda} \tag{1.17}$$

$\rho_\lambda^t S_\lambda < 1$ so when $n \to \infty$ then $(\rho_\lambda^t S_\lambda)^{(n+1)} \to 0$, therefore:

$$I_\lambda^{tot} = (I_\lambda^{dir} + I_\lambda^{diff}) \times \frac{1}{1 - \rho_\lambda^t S_\lambda} \tag{1.18}$$

Substituting equations 1.9 and 1.14 into equation 1.18 yields:

$$I_\lambda^{tot} = \mu_s I_\lambda [e^{-\frac{\tau_\lambda}{\mu_s}} + t_d(\theta_s)] \times \frac{1}{1 - \rho_\lambda^t S_\lambda} \tag{1.19}$$

It is convenient to define the downward transmittance as:

$$T(\theta_s) = e^{-\frac{\tau_\lambda}{\mu_s}} + t_d(\theta_s), \tag{1.20}$$

leading to a compact form for the total radiation received by the ground target:

$$I_\lambda^{tot} = \mu_s I_\lambda \times \frac{T(\theta_s)}{1 - \rho_\lambda^t S_\lambda} \tag{1.21}$$

A satellite sensor receives the solar radiation that has interacted with the ground

target directly (Figure 1.2d) and indirectly (Figure 1.2e), as well as the path radiance (Figure 1.2f).

Similar to equation 1.9, the upward direct solar radiation ($I_\lambda^{dir}sat$) is:

$$I_\lambda^{dir}sat = I_\lambda^{tot} \times \rho_\lambda^t(\theta_s, \theta_v, \phi) \times e^{-\frac{\tau_\lambda}{\mu_v}}, \tag{1.22}$$

where $\rho_\lambda^t(\theta_s, \theta_v, \phi)$ describes the surface bidirectional reflectance properties and in practice are approximated by the direct value of the surface reflectance for the illumination and view directions (i.e. ρ_λ^t), which results in small errors for areas with low values of surface reflectance (Lee and Kaufman, 1986), so:

$$I_\lambda^{dir}sat = I_\lambda^{tot} \times \rho_\lambda^t \times e^{-\frac{\tau_\lambda}{\mu_v}} \tag{1.23}$$

Similar to equation 1.14, the upward diffuse solar radiation ($I_\lambda^{diff}sat$) is:

$$I_\lambda^{diff}sat = I_\lambda^{tot} \times \rho_\lambda^t \times t_d(\theta_v) \tag{1.24}$$

The path radiance ($I_\lambda^{path}sat$) is:

$$I_\lambda^{path}sat = \mu_s I_\lambda \times \rho_\lambda^{atm}(\theta_s, \theta_v, \phi), \tag{1.25}$$

where $\rho_\lambda^{atm}(\theta_s, \theta_v, \phi)$ is the atmospheric reflectance and under single-scattering approximation is equivalent to:

$$\rho_\lambda^{atm}(\theta_s, \theta_v, \phi) = \rho_\lambda^m(\theta_s, \theta_v, \phi) + \frac{\omega_0 \tau_\lambda^a P_a(\theta_s, \theta_v, \phi)}{4\mu_s \mu_v}, \tag{1.26}$$

where $\rho_\lambda^m(\theta_s, \theta_v, \phi) = e^{-\tau_\lambda^m}$ is the path radiance due to molecular scattering, ω_0 is the single-scattering albedo, and $P_a(\theta_s, \theta_v, \phi)$ is the scattering phase function (Kaufman et al., 1997).

Combining $I_\lambda^{dir}sat$, $I_\lambda^{diff}sat$, and $I_\lambda^{path}sat$, we obtain the total radiation received by the satellite sensor ($I_\lambda^{tot}sat$):

$$I_\lambda^{tot}sat = I_\lambda^{tot} \times \rho_\lambda^t \times [e^{-\frac{\tau_\lambda}{\mu_v}} + t_d(\theta_v)] + \mu_s I_\lambda \times \rho_\lambda^{atm}(\theta_s, \theta_v, \phi) \tag{1.27}$$

By defining the upward transmittance as:

$$T(\theta_v) = e^{-\frac{\tau_\lambda}{\mu_v}} + t_d(\theta_v), \tag{1.28}$$

and by substituting equations 1.21 and 1.26 to equation 1.27, we obtain:

$$I_\lambda^{tot}sat = \mu_s I_\lambda [\rho_\lambda^t \frac{T(\theta_s)T(\theta_v)}{1 - \rho_\lambda^t S_\lambda} + \rho_\lambda^m(\theta_s, \theta_v, \phi) + \frac{\omega_0 \tau_\lambda^a P_a(\theta_s, \theta_v, \phi)}{4\mu_s \mu_v}] \tag{1.29}$$

By dividing $\mu_s I_\lambda$ on both sides, we obtain the apparent reflectance (ρ_λ^*) observed by satellite sensors:

$$\rho_\lambda^* = \rho_\lambda^t \frac{T(\theta_s)T(\theta_v)}{1 - \rho_\lambda^t S_\lambda} + \rho_\lambda^m(\theta_s, \theta_v, \phi) + \frac{\omega_0 \tau_\lambda^a P_a(\theta_s, \theta_v, \phi)}{4\mu_s \mu_v} \tag{1.30}$$

This constitutes the basis of satellite-based retrieval of τ_λ^a. To solve τ_λ^a from

equation 1.30, τ_λ^a should be the sole unknown variable in this equation. Compared to ground-based retrieval of τ_λ, the first variable additionally required to be known is the surface reflectance to achieve surface-atmosphere decoupling. Generally, the lower the surface reflectance, the larger contribution to ρ_λ^* from the path radiance, thus the higher accuracy of the retrieval of τ_λ^a. Geographical distributions of aerosol types and associated properties including ω_0 and P_a are the second variable additionally required to be known to make equation 1.30 solvable. Generally, they are taken from *a prior* global ground measurements. Also note that τ_λ^a is not explicitly solved from equation 1.30, but instead is typically found from a pre-prepared look-up table.

A variety of sensors with gradually enhanced τ_λ^a observation capabilities onboard polar-orbiting and geostationary satellite platforms have been launched since 1972 (Figure 1.3). Due to the difference in sensor design (e.g. spectral configurations) and observing geometries, many algorithms have been developed to accurately retrieve τ_λ^a from observations of these different sensors. The biggest difference among these algorithms lies in their strategies to decouple the contribution to ρ_λ^* from the atmosphere and the surface. For instance, the dark target algorithm (Kaufman et al., 1997) infers the surface reflectance, and thereby τ_λ^a, in the visible band from the reflectance in the near-infrared band which is hardly affected by PM based on a strong linear statistical relationship between these bands over densely vegetated regions, where the surface reflectance in the visible band tends to be small. In contrast, the deep blue algorithm (Hsu et al., 2004, 2006, 2013) relies on a reference image, usually derived from long-term climatological observations of TOA reflectance using the minimum value method (Herman and Celarier, 1997; Koelemeijer et al., 2003), to contrast with the PM-polluted images to initiate the τ_λ^a retrieval in the deep blue band ($\approx$ 0.41 μm) over arid/semi-arid areas, deserts, cities, and other bright surfaces, where the surface reflectance in that band tends to be small. The time sequence algorithm is one of the most promising

τ_λ^a retrieval approaches and works on the principal that the characteristics of the underlying surface do not change or change slightly during a certain period (e.g. 30 days); thus, the surface-atmosphere decoupling is achievable using the associated image series within the temporal window (Hagolle et al., 2008; Zhang et al., 2014; Lyapustin et al., 2018). τ_λ^a typically reported at $\lambda = 550$ nm is usually interpolated in logarithm scale from the values of other channels that these and other algorithms mainly utilize. It is expected that these and other algorithms will be continuously adapted and improved to work with future satellite missions that are likely characterized by high temporal, spatial, vertical, and spectral resolutions with multiple viewing angles and polarization to enable near real-time retrieval of different types (e.g. fine model, absorptive mode, and so forth) of τ_λ^a at unprecedented details.

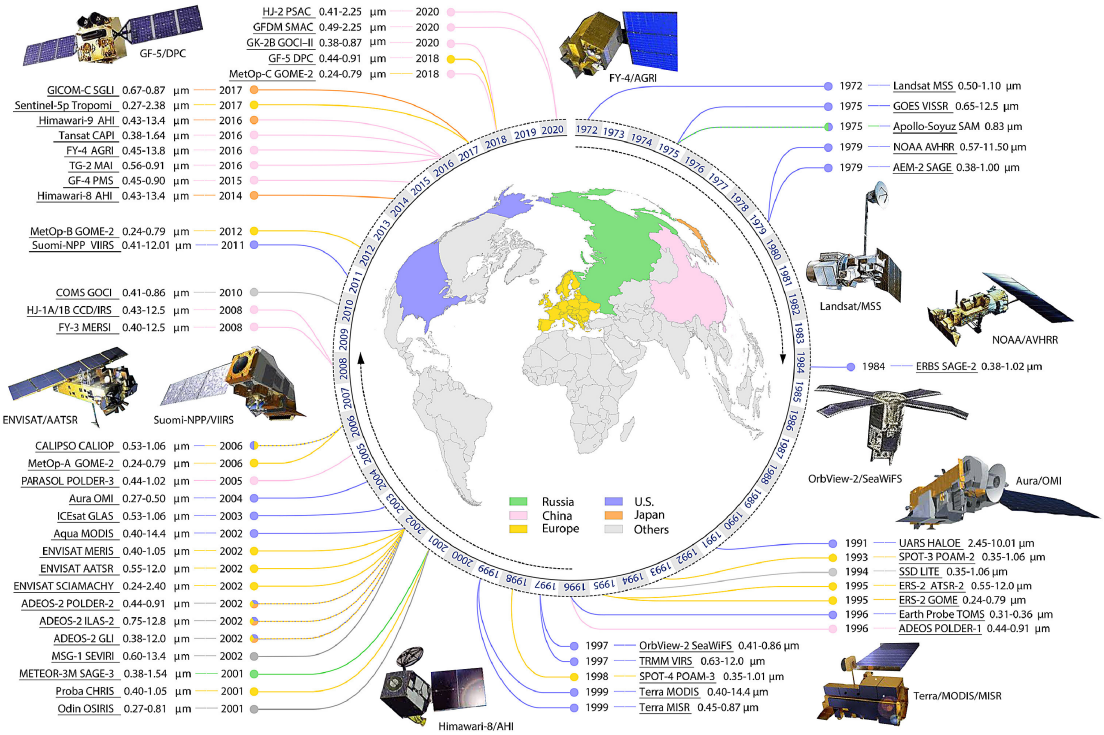


Figure 1.3: The main satellite platforms and sensors for global PM observations. The figure is taken from Zhang et al. (2021).

Like ground measurements, satellite observations largely focus on ground-level

PM mass concentrations. But unlike the former, the latter becomes harder to achieve because most satellites simply provide columnar observations. As a better strategy, the research community has instead developed a series of methods to convert satellite-retrieved AOD to ground-level PM. Because of the severe negative impacts of $PM_{2.5}$ on human health and because visible light is sensitive to $PM_{2.5}$, a proliferation of publications focusing on converting satellite-retrieved AOD at 550 nm (AOD_{550}) to ground-level $PM_{2.5}$ concentrations have emerged (Figure 1.4). Nonetheless, to convert satellite-retrieved AOD_{550} to ground-level $PM_{2.5}$ is non-trivial due to the inherent differences between the two. This thus leads to the development of a variety of data-driven and process-driven models.

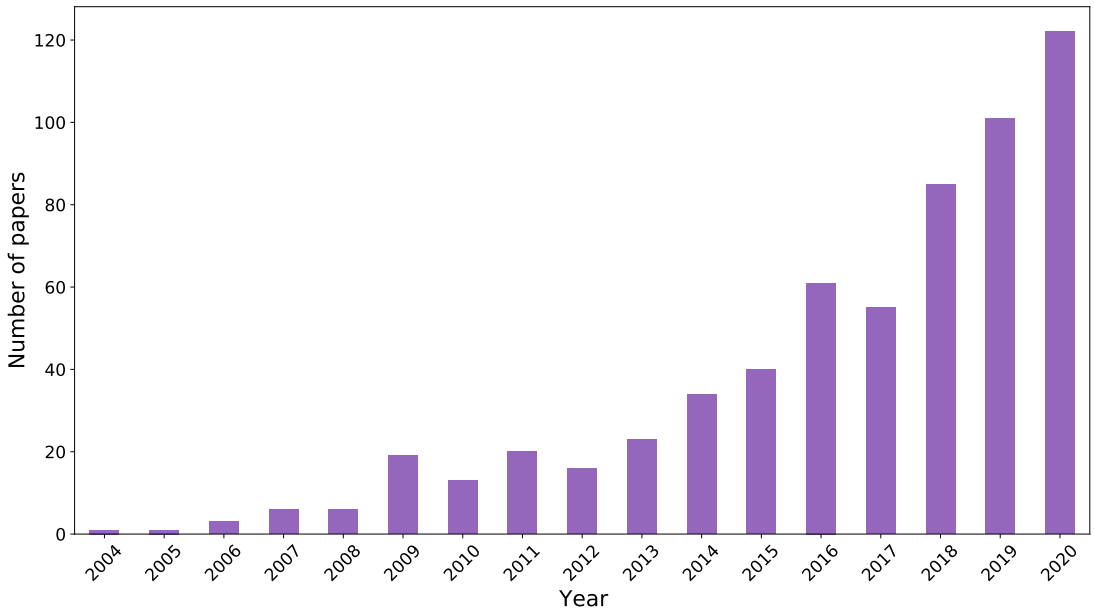


Figure 1.4: The number of papers published on the topic of ground-level $PM_{2.5}$ estimations inferred from satellite-retrieved AOD in the past 17 years. The figure is plotted using Matplotlib with data from Web of Science by searching "$PM_{2.5}$" AND "AOD" AND "Satellite".

Data-driven models have evolved from the use of AOD_{550} as the sole predictor (Wang and Christopher, 2003; Lee et al., 2011; Xie et al., 2015) to the incorporation of multiple additional predictors including meteorology and land use variables

(Liu et al., 2005; Ma et al., 2014), from the use of linear, parametric statistical methods (Liu et al., 2009; He and Huang, 2018*a*,*b*) to the employment of nonlinear, non-parametric machine learning methods (Di et al., 2016; Hu et al., 2017; Wei et al., 2019, 2021), and from one-stage (Yao et al., 2018) to multistage methods (Hu et al., 2014; Xiao et al., 2017, 2018; Yao et al., 2019). These approaches aim to take advantage of the significant heterogeneity of the relationship between AOD_{550} and $PM_{2.5}$. However, generally the more sophisticated the model, the less interpretable the model. This is particularly true for machine and deep learning models that are still regarded as black boxes and hence more efforts are required to make them more transparent (Liao et al., 2020).

Process-driven models generally utilize chemical transport models (CTMs) to simulate the spatiotemporal varying scaling factors between AOD_{550} and ground-level $PM_{2.5}$ concentrations, which are subsequently applied to satellite-retrieved AOD_{550} to derive satellite-based $PM_{2.5}$ (Van Donkelaar et al., 2010, 2015; Boys et al., 2014). Compared to data-driven models, these models are more complicated and computationally expensive but all the processes are transparent such that they are more physically coherent. The biggest deficiencies of process-driven models are the ubiquitous uncertainties embodied in the driven emission inventories and parameterization schemes. Consequently, process-driven models are broadly used to derive $PM_{2.5}$ estimates on long-term scales where satisfactory accuracy can be met. Accordingly, the progress on ground-level $PM_{2.5}$ from these models largely relies on the progress on the developments of these models themselves.

Based on the descriptions above, it is promising to combine data- and process-driven models to obtain better ground-level $PM_{2.5}$ estimates. Existing studies have explored several strategies including: 1) using data-driven models to fuse ground-level $PM_{2.5}$ measurements and process-driven ground-level $PM_{2.5}$ estimates (Van Donkelaar et al., 2019; Hammer et al., 2020) and 2) using

simulations from process-driven models as part of the predictors in building data-driven models (Xue et al., 2019; Geng et al., 2021). In this thesis, I propose a model framework that uses the process-driven models to enlighten the construction of data-driven models in order to improve the robustness of ground-level $PM_{2.5}$ estimates inferred from satellite-retrieved AOD, as described in Chapter 5.

1.2.4 Modelling PM pollution

Ground and satellite observations of PM pollution are essentially a series of snapshots. Neither of them would be sufficient to determine the sources and sinks of PM pollution, hindering policy-making aimed at reducing the PM pollution and PM-induced PV energy losses. Therefore, the use of process-driven models becomes critical in integrating and interpreting these observations. On the other hand, the process-driven models are evaluated by these observations before attaching confidence to the interpretation attained. A range of process-driven models could describe PM pollution. Here I describe the GEOS-Chem global 3-D model of atmospheric composition which is extensively used in this thesis. In particular, I use the 'Classic' version of GEOS-Chem with single-node shared-memory parallelization (OpenMP) and rectilinear latitude-longitude grids (Bey et al., 2001). It is an off-line CTM driven by Goddard Earth Observing System (GEOS) meteorological analyses, provided by Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center. Below I give a brief description of the features and very recent updates of the several main modules in GEOS-Chem 'Classic'.

Emissions are a central component of all CTMs and GEOS-Chem 'Classic' is of no exception. GEOS-Chem 'Classic' uses HEMCO to assemble a versatile of

natural and anthropogenic emission inventories to compute emission fluxes, online or off-line, for uses in relevant modules (Keller et al., 2014). HEMCO is the common abbreviation for its previous and current full names of Harvard-NASA Emissions Component and Harmonized Emissions Component, respectively. The change in the full names reflects its recent updates featuring an improved three-layer architecture to facilitate implementation into any atmospheric model (Lin et al., 2021). HEMCO comes with a default database library of emission inventories to which users can add their own. Guidance on assembling these emission inventories is specified through the HEMCO configuration file without changing the source code of the GEOS-Chem 'Classic'. Notably, HEMCO in GEOS-Chem 'Classic' is used not only for emissions but also as a general data broker to read and process all model input data including meteorological fields, chemical boundary conditions (for nested runs), initial conditions, and other environmental data. This has in particular enabled the FlexGrid capability in GEOS-Chem 'Classic' which facilitates the definition of custom grid for nested runs (Wang et al., 2004) at runtime rather than at compile time.

Meteorology plays a vital role in determining PM pollution in that it significantly influences the production and loss rates of PM both physically and chemically. The two GEOS assimilated meteorological data used by GEOS-Chem 'Classic' are: 1) the GEOS Forward Processing (GEOS-FP) meteorological analyses covering 2012–present with a horizontal resolution of 0.25° latitude × 0.3125° longitude (Lucchesi, 2018) and 2) the Modern-Era Retrospective analysis for Research and Application, Version 2 (MERRA-2) meteorological analyses covering 1970–present with a horizontal resolution of 0.5° latitude × 0.625° longitude (Gelaro et al., 2017). Both GEOS-FP and MERRA-2 meteorological analyses extend through 72 vertical terrain-following sigma levels that describe the atmosphere from the surface to 0.01 hPa. They also share an identical 3-hour temporal

resolution for 3-D fields and 1-hour temporal resolution for 2-D fields. GEOS-Chem 'Classic' simulations are generally conducted at the native horizontal of these GEOS fields in nested runs and at coarser resolutions, 2° latitude × 2.5° longitude or 4° latitude × 5° longitude, in global runs. GEOS-Chem 'Classic' simulations can also be conducted at the original vertical grid or at a reduced vertical grid containing 47 levels, of which about 30 lie below the dynamic tropopause, in both global and nested runs.

Chemistry is the selling point of GEOS-Chem. In this thesis, I consistently use the GEOS-Chem configuration that has full chemistry only in the troposphere with simple linear representation of stratospheric chemistry following the Linoz algorithm for ozone (McLinden et al., 2000) and 2-D (latitude-altitude) monthly mean sources and loss rate constants for other gases (Murray et al., 2012). The troposphere-only simulation contains a detailed NO_x-O_x-hydrocarbon-aerosol-bromine-chlorine-iodine chemical mechanism applied from the surface to the tropopause. The sulfate (SO4) - nitrate (NIT) - ammonium (NH4) secondary inorganic aerosol (SNA) was developed by Park et al. (2004), with thermodynamics computed by the ISORROPIA thermodynamic module (Fountoukis and Nenes, 2007). The simulation of primary (POA) and secondary (SOA) organic aerosol (OA) is by default to follow the simple, irreversible, direct yield scheme of Pai et al. (2020). The POA is a combination of hydrophilic (OCPI) and hydrophobic (OCPO) components. The SOA has an alternative to use a combination of explicit aqueous uptake mechanisms (Marais et al., 2016) with a standard volatility basis set scheme (Pye et al., 2010). Two choices of the OA to organic carbon (OC) ratios are available to calculate the former from the latter. The default is to use a global-mean ratio of 1.4 and 2.1 for OCPO and OCPI, respectively (Turpin and Lim, 2001). The other is to use the spatially and seasonally resolved OA/OC ratios which offers more information despite introduces some uncertainties (Philip et al., 2014). The black carbon (BC) simulation was described by Wang, Jacob,

et al. (2014), with an option (default is yes) of specifying absorption enhancement factors from coating for both hydrophilic (BCPI) and hydrophobic (BCPO) BC (Wang, Heald, et al., 2014). The simulation of dust distributed in seven size bins (Zhang et al., 2013) was described by Fairlie et al. (2007). The simulation of accumulation (SALA) and coarse (SALC) mode sea salt (SS) were described by Jaeglé et al. (2011). With simulated 3-D fields of dry mass concentrations of these individual PM species $C_i^{dm}(I,J,L)$s, $PM_{2.5}$ mass concentrations at a given box (I,J,L), $C^m_{PM_{2.5}}(I,J,L)$, can be properly derived using equation 1.31, which accounts for aerodynamic diameters of PM species, and controlled conditions characterized by most *in situ* $PM_{2.5}$ measurements.

$$
\begin{aligned}
C^m_{PM_{2.5}}(I,J,L) = \quad & \{[C^{dm}_{SO4}(I,J,L) + C^{dm}_{NIT}(I,J,L) + C^{dm}_{NH4}(I,J,L)] \times GF_{SNA} \\
& + C^{dm}_{BCPI}(I,J,L) + C^{dm}_{BCPO}(I,J,L) \\
& + [C^{dm}_{OCPI}(I,J,L) + C^{dm}_{SOA}(I,J,L)] \times GF_{ORG} \\
& + C^{dm}_{OCPO}(I,J,L) + C^{dm}_{SALA}(I,J,L) \times GF_{SSA} \\
& + C^{dm}_{DUST}(I,J,L,1) + C^{dm}_{DUST}(I,J,L,2) \\
& + C^{dm}_{DUST}(I,J,L,3) + C^{dm}_{DUST}(I,J,L,4) \\
& + C^{dm}_{DUST}(I,J,L,5) \times 0.38\} \\
& \times \frac{1013.25}{Air\_Pressure(I,J,L)} \\
& \times \frac{Air\_Temperature(I,J,L)}{298}, \qquad (1.31)
\end{aligned}
$$

where $C^{dm}_{DUST}(I,J,L,1-5)$ are the dry mass concentrations of dust distributed in the first five size bins. $C^{dm}_{DUST}(I,J,L,5)$ consists of PM with aerodynamic diameter smaller and larger than 2.5 μm and thereby is approximately multiplied by 0.38 to include only the former. For the same reason, SALA but not SALC is included into $PM_{2.5}$. The last two scaling factors are to use ideal gas law to convert

simulated $PM_{2.5}$ values to standard temperature pressure conditions which are usually characterized by $PM_{2.5}$ measurements. GF_{SNA}, GF_{ORG}, and GF_{SSA} are growth factors for SNA, OCPI and SOA, and SALA to account for aerosol water of $PM_{2.5}$ in a way that is consistent with *in situ* $PM_{2.5}$ measurements. No growth factor for BCPI simply because it only starts to grow when the relative humidity (RH) becomes larger than 80%. These growth factors are calculated using the change in radius between different RHs:

$$GF = 1 + [(\frac{radius_{RH\_wet}}{radius_{RH\_dry}})^3 - 1] \times \frac{Density_{Water}}{Density_{DrySpecies}}, \tag{1.32}$$

for which Table 1.1 lists calculated values under RH = 35% and RH = 50% that are operational conditions in the United States (U.S. Environmental Protection Agency, 1997) and in Europe (European Committee for Standardization (CEN), 1998) and China (HJ 618-2011, 2011), respectively.

Table 1.1: Growth factor values under RH = 35% and 50%.

RH	GF_{SNA}	GF_{ORG}	GF_{SSA}
35%	1.33	1.16	1.86
50%	1.51	1.24	2.42

PM optical depth at a given box (I, J, L) for wavelength λ, $\tau_\lambda^a(I, J, L)$, can be properly calculated by combining simulated 3-D fields of dry concentrations of individual PM species and their mass extinction coefficients as indicated in equation 1.8. However, the calculation of mass extinction coefficients is not that straightforward implied by equation 1.8. This is largely because the assumption of a fixed radius for a certain PM species i to facilitate the illustration of equation 1.8 does not hold true in reality. With the acknowledgement of the impossibility to model varying radius even for a PM species i which is countless, statistical

modelling starts to play a role. The foremost is to assume a distribution for PM species and a lognormal size distribution is a widely-used reasonable one. It has been shown that one key parameter best related to the radiative properties of a PM species i with a given size distribution is the cross section weighted effective radius r_e^i (Hansen and Travis, 1974; Lacis, 1995):

$$r_e^i = \frac{\int_r r\pi r^2 f(r)\mathrm{d}r}{\int_r \pi r^2 f(r)\mathrm{d}r}, \tag{1.33}$$

where $f(r)$ is the fraction of a PM species i with radius between r and $r + \mathrm{d}r$.

Together with other properties of PM species i, r_e^i can then be used to derive the mass extinction coefficient $Q_\lambda^i m$ for wavelength λ of this PM species based on a general formula (Lacis, 1995; Tegen and Lacis, 1996):

$$Q_\lambda^i m = \frac{3Q_\lambda^i}{4\rho^i r_e^i}, \tag{1.34}$$

where ρ^i is the density of this PM species. Q_λ^i is the extinction coefficient for wavelength λ of this PM species, which can be calculated from the Mie-scattering theory using the assumed lognormal size distributions.

Note that $Q_\lambda^i m$ must be combined with the mass concentration of PM species i C_i^m to compute τ_λ^i. To directly combine with simulated dry mass concentrations of PM species, equation 1.34 can be slightly modified to (Chin et al., 2002):

$$Q_\lambda^i m = \frac{3Q_\lambda^i}{4\rho^i r_e^i} \times \frac{C_i^m}{C_i^{dm}}, \tag{1.35}$$

where $\frac{C_i^m}{C_i^{dm}}$ are the ratio of the mass concentration of PM species i to its dry mass

concentration. This ratio equals one for hydrophobic PM and depends on ambient RH for hydrophilic PM.

Also note that all variables on the right hand of equation 1.35 have exclusive dependence on RH and/or wavelength. With hydrophilic PM optics from Latimer and Martin (2019) and dust optics from Ridley et al. (2012), $Q^i_\lambda m$ is immediately determined and applied in GEOS-Chem 'Classic' to derive $\tau^a_\lambda(I, J, L)$:

$$\begin{aligned} \tau^a_\lambda(I, J, L) &= \sum_i \tau^i_\lambda \\ &= \sum_i Q^i_\lambda m(I, J, L) \times C^{dm}_i(I, J, L) \times H(I, J, L), \end{aligned} \tag{1.36}$$

where $H(I, J, L)$ is the height of the given box (I, J, L) and its product with C^{dm}_i gives the dry mass loading of PM species i per unit area. Note that the summing practice here ignores the influence of aerosol mixing, which may commonly occur, on aerosol optical properties that are used to derive the optical depth. This is not ideal but adoptable considering the difficulty of obtaining such information and that the influence is fairly minor.

Deposition is the ultimate path by which PM is removed from the atmosphere and accumulated onto the surface where solar PV panels may be located. GEOS-Chem 'Classic' includes both wet and dry deposition. Wet deposition by rain is considered for both water-soluble PM (Liu et al., 2001) and gases (Amos et al., 2012), and scavenging by snow and cold/mixed precipitation is also considered for PM (Wang et al., 2011; Wang, Jacob, Spackman, Perring, Schwarz, Moteki, Marais, Ge, Wang and Barrett, 2014). Dry deposition generally follows the resistance-in-series scheme of Wesely (2007), with Zhang et al. (2001) describing PM dry deposition as a complex process depending on PM

physical and chemical properties, the underlying surface characteristics, and micro-meteorological conditions. Dry deposition is considered as a main source of pollution for solar PV panels, and I give a detailed description of its calculation in Appendix A.

Transport moves PM pollution from one place to another. In this sense, it stands for a source or a sink for a receptor region depending on meteorological conditions, of which are primarily wind speed and direction. For physical processes, GEOS-Chem 'Classic' uses the TPCORE advection algorithm (Lin and Rood, 1996) and computes convective transport from the convective mass fluxes in the meteorological data (Wu et al., 2007). For planetary boundary layer (PBL) mixing, GEOS-Chem 'Classic' uses either the non-local PBL mixing scheme as implemented by Lin and McElroy (2010), or the full PBL mixing scheme as briefly described in (Bey et al., 2001; Wu et al., 2007). In this thesis, I consistently choose the non-local PBL mixing scheme for all research chapters.

Radiation is additionally implemented into GEOS-Chem 'Classic' by coupling to the rapid radiative tranfer model for general circulation models (GCMs) (RRTMG) (Iacono et al., 2008), a configuration which is known as GCRT (Heald et al., 2014). This coupling with the troposphere-only simulations enables users to investigate the impact of PM species on the Earth's energy balance at both TOA and the surface for both longwave and shortwave under both all-sky and clear-sky conditions. In this thesis, I mainly use GCRT to compute all-sky, shortwave, surface solar radiation flux under control and no PM conditions for the uses of solar PV performance models. This flux is also widely known as surface solar radiation downwards or global horizontal irradiance GHI. GHIs are calculated over fourteen contiguous bands in the shortwave and subsequently combined for output. In this process, PM impacts can be included or not in these fourteen individual band ranges such that PM impacts can be readily obtained by taking

the differences. The individual band ranges (in wavenumbers, cm^{-1}) are: 2600–3250, 3250–4000, 4000–4650, 4650–5150, 5150–6150, 6150–7700, 7700–8050, 8050–12850, 12850–16000, 16000–22650, 22650–29000, 29000–38000, 38000–50000, and 820–2600. The last band is coded out of sequence to preserve spectral continuity with the longwave bands.

1.3 Photovoltaics

The word photovoltaic is derived from two different words. The word photos comes from the Greek and means light. The word voltaic comes from the name of an Italian scientist, Volta, who studied electricity. This explains what a PV system does: it converts light energy from the Sun into usable electrical energy. Studying PV essentially and inevitably involves interdisciplinary efforts considering the many processes occurring during the conversion from the original sunlight to the final electricity. In this section, I briefly introduce PV basics, factors affecting PV performance with a focus on those relevant to availability of solar radiation, and PV performance models used in this thesis.

1.3.1 PV basics

The basic element of a PV system is the solar cell which is generally made of semiconductor materials such as silicon. These semiconductor materials exhibit the photovoltaic effect that comprises the basis of conversion from solar light energy to electrical energy. Briefly, a solar cell is a thin semiconductor wafer that is specially treated to form an electric field, positive on one side and negative on the other. When the light is absorbed by the solar cells and the photon energy is large enough, electrons are knocked loose from the atoms in the semiconductor

material. If electrical conductors are attached to the positive and negative sides, forming an electrical circuit, the electrons can be captured in the form of an electric current – that is, electricity. This electricity can then be used to power a load or stored.

The power output from a solar cell is often of small magnitude. To increase the utility, a number of individual solar cells are interconnected together to a sealed, waterproof package called a panel or a module. For instance, a 12 V module will have 36 solar cells connected in series. Further, to achieve desirable voltage and current, modules are wired in series and in parallel into what is called an array. With these flexibility and modularity, designers are able to create solar PV systems that can meet a wide variety of electrical needs. Figure 1.5 shows the schematic of a solar cell, module, and array.

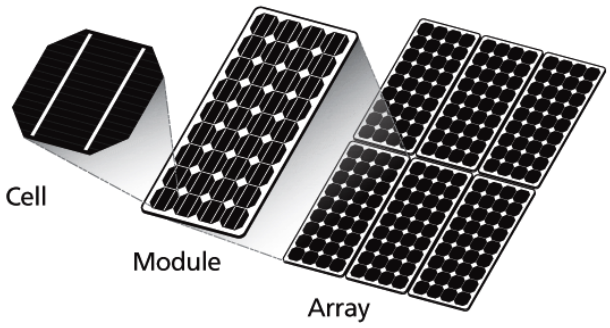


Figure 1.5: Schematic of a solar cell, module, and array. The figure is taken from SamlexAmerica (2021).

Based on whether connected to an electricity transmission and distribution system (hereafter referred to as the electricity grid), PV systems can be categoried into non-grid-tied and grid-tied ones.

Figure 1.6a depicts a typical non-grid-tied PV system. It consists of a PV module, charge controller, battery, and inverter. The PV module converts the solar light energy into direct current (DC) electrical energy. The charge controller conditions

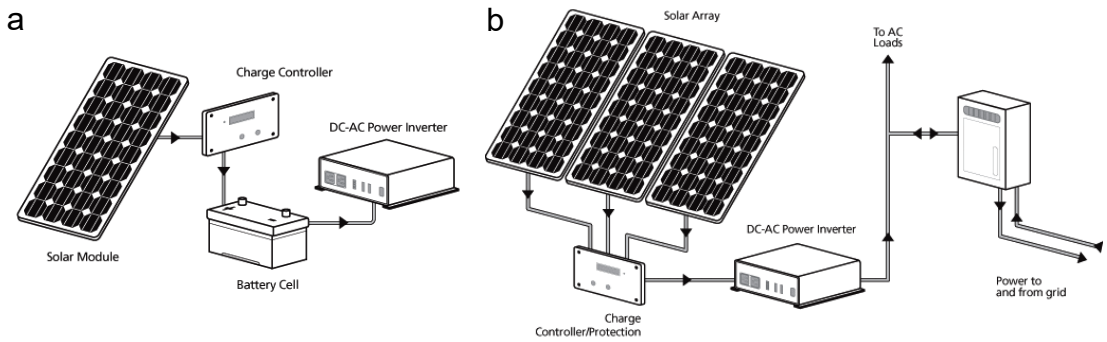


Figure 1.6: Diagrams of (a) a typical non-grid-tied PV system and (b) a typical grid-tied PV system. The figure is adapted from SamlexAmerica (2021).

the DC electrical voltage and current to safely charge the battery. The battery stores the DC electrical energy so that it can be used when there is no solar light energy available for instant conversion, e.g. night time and cloudy days. The inverter converts the DC electrical energy produced by the PV module or stored in the battery into alternating current (AC) electrical energy. DC loads are powered by the DC electrical energy produced by the PV module or stored in the battery, while AC loads are powered by the AC electrical energy produced from the inverter.

Figure 1.6b depicts a typical grid-tied PV system. Compared with the non-grid-tied PV system, it removes the battery but introduces an electrical grid. The PV array generates the DC electrical energy that is subsequently conditioned by the charge controller to feed into the inverter to produce the AC electrical energy. This AC electrical energy can be directly used to power AC loads or sent into the electrical grid which can power AC loads during unfavourable conditions such as when clouds prevent solar light energy reaching the solar PV panels to be instantly converted.

1.3.2 Factors affecting PV performance

A range of processes over the course of solar light energy to electrical energy means a range of factors affecting PV performance. These factors range from weather variations to properties and configurations of PV materials. With respect to the former, the three leading factors are GHI, which is in turn modulated by PM (Sarver et al., 2013; Bergin et al., 2017; Peters et al., 2018; Sweerts et al., 2019; Li, Wagner, Peng, Yang and Mauzerall, 2017; Li et al., 2020) and by clouds (Chen et al., 2016; Gil et al., 2019), air temperature (T) with cooler conditions generally improving PV performance and hotter conditions reducing it (Panagea et al., 2017; Chaichan and Kazem, 2016), and surface wind speed (v) with air flow typically cooling the PV module (Chenni et al., 2007). With respect to the latter, it includes degradation of PV modules and inverters, operation away from the maximum power point to adapt to the electrical grid, wiring losses, and so forth (Meral and Dincer, 2011; Vidyanandan, 2017). In this thesis, I simply study the impact of availability of GHI on the PV power output, with a focus on the role of anthropogenic PM and its sensitivity to upstream emissions.

As illustrated in Figure 1.7, a solar PV panel typically collects three components of sunlight: a direct component that passes unaffected through the atmosphere which is dependent on direct normal irradiance (DNI), solar positions, and panel configurations, a diffuse component arising from scattering which is largely dependent on diffuse horizontal irradiance (DHI) and panel configurations, and a component reflected from the ground which is dependent on GHI, visible surface albedo, and panel configurations. All but DNI interact with atmospheric PM either through absorption or scattering, thereby reducing the amount of solar radiation reaching the solar panels. GHI, DNI, and DHI are linked with each other by:

$$GHI = DHI + DNI \times \cos\theta_s, \tag{1.37}$$

where θ_s is the solar zenith angle, as similarly defined in section 1.2.3.

With this relationship, knowing two of the three terms would readily determine the remaining one considering that θ_s is generally solvable. Nonetheless, measurements, reanalyses, or models generally provides GHI only, as a result of which empirical models such as DISC (Maxwell, 1987; Skartveit et al., 1998) or DIRINT (Ineichen et al., 1992) are proposed to use GHI to estimate DNI and DHI. In this thesis, I simply choose the Erbs model (Erbs et al., 1982) to estimate the DHI and then the DNI from the GHI simulated from the GCRT model, as detailed in Appendix B.

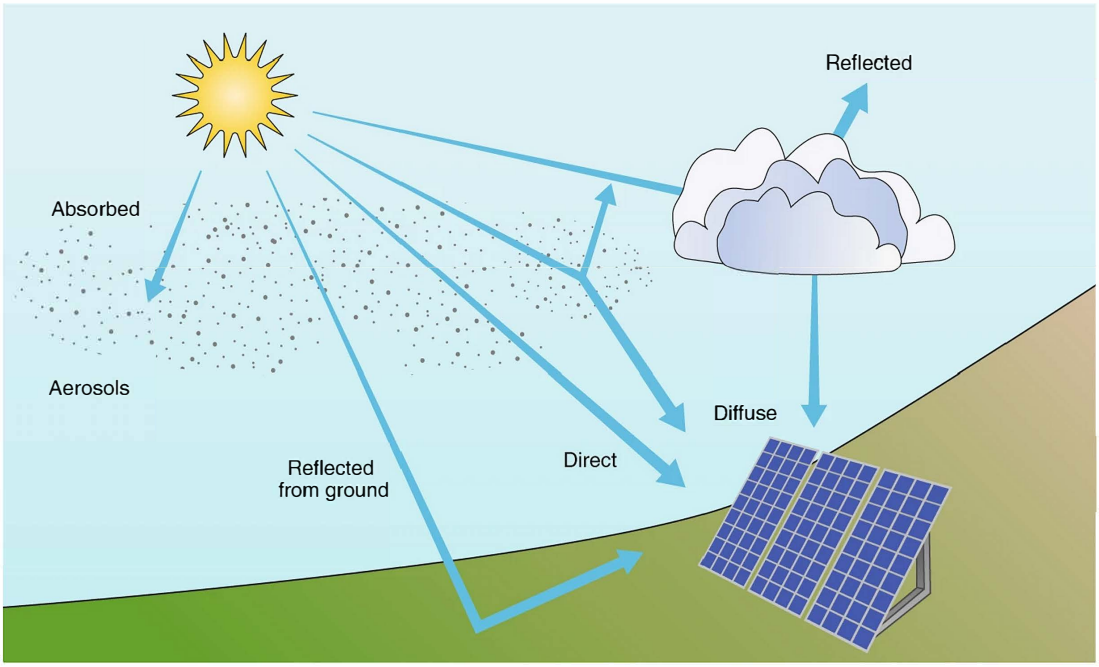


Figure 1.7: Interaction between atmospheric PM and solar radiation. The figure is taken from Ekins-Daukes and Kay (2019).

As illustrated in Figure 1.8, solar radiation received by the solar panels is further

reduced by PM deposited on solar panels before reaching the solar cells to be converted to usable electricity. To quantify the impacts of these deposited PM, we must know the mass loading of each PM species accumulated on the solar panels, which is subject to both accumulation (e.g. dry deposition) and removal (e.g. precipitation) processes. It can then be combined with speciated mass extinction coefficients to determine the optical depth of deposited PM and thereby the reduced solar radiation received by the solar cells.

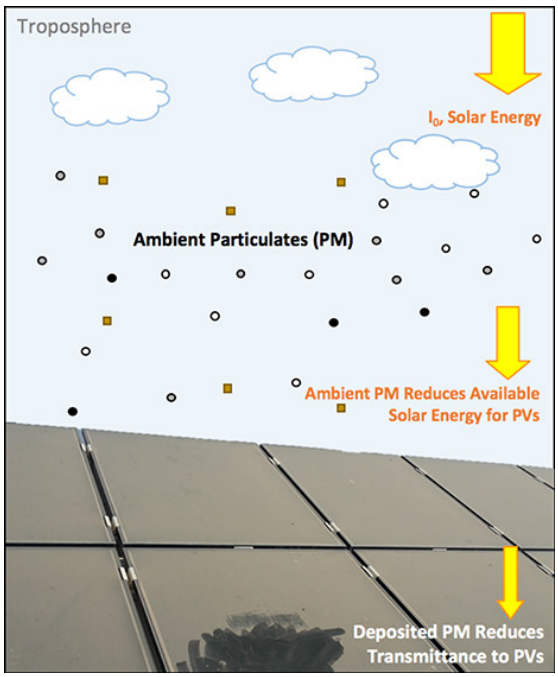


Figure 1.8: Interaction between deposited PM and solar radiation. The figure is taken from Bergin et al. (2017).

As described in section 1.2, atmospheric and thereby deposited PM are subject to a variety of factors. It thus logically follows that PV performance will also be impacted by these factors. Among them anthropogenic emissions are potentially controllable. While there are studies using GCMs to quantify PM-induced PV energy losses and PV energy gains due to certain emission reductions, to date no attempt has used CTMs to do so. Compared with GCMs, CTMs typically include a more detailed mechanism of coupled gas-aerosol phase chemistry and thereby

could potentially provide more accuracy in PM simulations. Moreover, previous studies have not explored the benefits to PV energy gains from emission reductions in every individual anthropogenic source sector. Consequently, no one knows which source sector where reducing emissions could provide the largest benefits to PV energy gains, not to say the contributions of different fuel types within each source sector. Gaining these knowledge, however, can provide critical information on addressing the trade-off between the tremendous efforts and limited resources that are needed to reduce PM pollution.

To remedy the above deficiencies, through chapters 2–4 in this thesis, I integrate models of solar PV performance and atmospheric composition to investigate to which extent PV energy generation is threatened by atmospheric and deposited PM, and more importantly to which extent we could mitigate such threat by reducing anthropogenic emissions. More specifically, I integrate GEOS-Chem 'Classic' as detailed in the previous section with the PVLIB-Python which is a solar PV performance model, as described in the following section.

1.3.3 Modelling PV performance

To study the impacts of the aforementioned factors on PV performance and in order to develop strategies to improve PV performance, the research community has developed, calibrated, validated, and applied various solar PV performance models. Typical examples include global solar energy estimator (Pfenninger and Staffell, 2016) and PVLIB-Python (Holmgren et al., 2018). The strength of the former lies in its sensible defaults to facilitate fast simulations and support for solar energy technologies other than PV. The strength of the latter lies in its easy control over the nuts and bolts of simulating PV performance. Both models can meet the research purpose set in this thesis. I choose the PVLIB-Python model to take advantage of its easy-to-modify capabilities. Figure 1.9 describes how I

use the PVLIB-Python to calculate capacity factor (CF) from GHI simulated from the GCRT model. The CF, defined as the ratio of the actual power output of a PV panel to the theoretical maximum power output, describes the solar PV electricity generation efficiency. Below I describe the detailed calculation process with specific configurations chosen in this thesis.

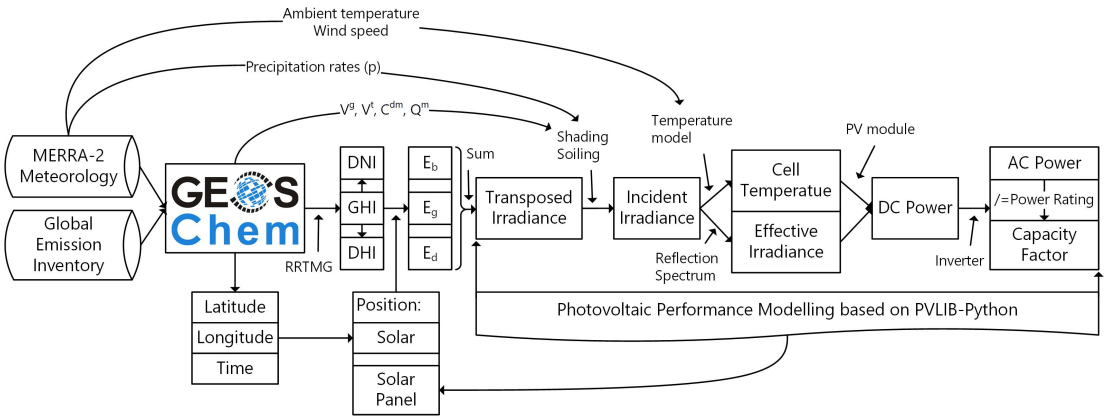


Figure 1.9: Workflow to estimate solar photovoltaic energy generation efficiency in this study. Full definitions for abbreviations include: 1) global horizontal irradiance (GHI) simulated from GEOS-Chem coupled with rapid radiative transfer model for GCMs (RRTMG), and it is subsequently decomposed to direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI); 2) beam (E_b), ground-reflected (E_g), and sky-diffuse (E_d) components of transposed irradiance; 3) PM gravitational (V^g) and turbulent (V^t) velocities; 4) PM dry mass concentrations (C^{dm}) and PM mass extinction coefficients (Q^m); 5) direct (DC) and alternating (AC) current powers.

As mentioned earlier, I use the Erbs model (Erbs et al., 1982) to estimate the DHI and then the DNI from the GHI simulated from the GCRT model (Appendix B). The GHI, DNI, and DHI constitute the most important part of the 'weather' data that drive the PVLIB-Python model. Other fields of the 'weather' include air temperature and wind speed for calculating cell temperature. In this thesis, I use the MERRA-2 meteorological analyses for these fields. More specifically, I employ surface temperature to represent the former, and take the square root

of the sum of the square of east-west and north-south winds at 10 m height for the latter. I also make use of surface pressure and visible surface albedo provided by the MERRA-2 meteorology. By doing this I acknowledge the slight but time-varying nature of these two attributes. I additionally require 'weather' contain a soiling factor (SF) field that describes the irradiance reduction by the deposited PM. Its calculation will be detailed in the following research chapter. In short, the 'weather' is a UTC time series data containing six fields, and in practice, each spatial grid will hold its own 'weather' data. To support the large-scale global and decadal analyese in this thesis, I have developed parallel computation capabilities for all spatial grids and temporal steps.

The PVLIB-Python model currently supports the three widely used panel settings: horizontal fixed (Flat), fixed with optimal tilt (Tilt), and one-axis tracking (OAT). Their characteristics are summarized in Table 1.2. For each panel, the PVLIB-Python model uses predefined PV modules, inverters, and cell temperature models to convert the UTC time series 'weather' data to the UTC time series CF data.

Table 1.2: Solar panel settings currently supported by the PVLIB-Python model.

Name	Abbreviation	Descriptions
Horizontal fixed panels	Flat	Panels are fix mounted and horizontal
Optimally tilted panels	Tilt	Panels are fixed mounted, tilted at the latitude angle, and oriented to the equator
One-axis tracking panels	OAT	Panels rotate around one axis from east to west to track the sun throughout the day

The PVLIB-Python model first calls relevant methods to calculate the UTC time series of solar position including (apparent) solar zenith (θ_s) and azimuth (θ_A). The surface temperature and pressure information are consistently used for better accuracy (Reda and Andreas, 2004). It then transposes DNI, GHI, and DHI

to the beam (E_b), ground-reflected (E_g), and sky-diffused (E_d) components of irradiance collected by the solar panels (Appendix C). E_b, E_g, and E_d collectively constitute the transposed irradiance received by the solar panels. They are multiplied by the SF field in the 'weather' data to form the irradiance received by the solar cells. The irradiance before and after this scaling are termed as 'in' ($POAI_{in} = E_b + E_g + E_d$) and 'out' ($POAI_{out} = POAI_{in} \times SF$) plane of array irradiance (POAI), respectively. The default value of SF is set to one, meaning $POAI_{in} = POAI_{out}$. The actual values of SF developed in this thesis based on the GCRT model outputs and MERRA-2 precipitation rates will be detailed in the following research chapter. Whether to use the default or the actual values of SF depends on research purpose and corresponding experimental designs, and will be thoroughly described in the following research chapters.

Next, the PVLIB-Python model calculates cell temperature and effective irradiance. Cell temperature is required to calculate the IV curve and thereby the DC power of the PV module. It is affected by the incident irradiance, weather conditions, module materials, and so forth. Various models exist to estimate cell temperature. Among them I simply choose to use the Sandia Array Performance Model (SAPM) (King et al., 2004) which gives cell temperature (T_c) as:

$$T_c = T_m + \frac{POAI_{out}}{E_0}\Delta T, \tag{1.38}$$

where $E_0 = 1000$ W m^{-2} is the reference irradiance. ΔT is the temperature difference parameter whose values are a function of module materials and module mounting configurations. T_m is the module temperature and is expressed as:

$$T_m = POAI_{out} \times e^{a+b\times WS} + T_a, \tag{1.39}$$

where WS and T_a are wind speed and air temperature stored in 'weather' provided by the MERRA-2 meteorology. a and b are parameters whose values are a function of module materials and module mounting configurations.

Effective irradiance is the other basis for calculating the IV curve and thereby the DC power of the PV module. It is defined as the $POAI_{out}$ further adjusted for reflection diffusion losses and losses due to spectral mismatch. In other words, it is the final irradiance that is "available" to the PV module for power conversion. I again choose the SAPM which calculates effective irradiance (E_e) as:

$$E_e = f_1(AM_a)[E_b f_2(AOI) + (E_g + E_d)f_d], \tag{1.40}$$

where f_1 is a fourth degree polynomial in absolute airmass $AM_a = \frac{1}{\cos\theta_s} \times \frac{P}{101325}$, where P is the surface pressure available in 'weather' provided by the MERRA-2 meteorology. f_1 describes the losses due to spectral mismatch. f_2 is a fifth degree polynomial in the angle of incidence (AOI, see equation C.2) and describes the reflection losses. f_d is the fraction of diffuse irradiance on the plane of array that it not reflected away. The value of f_d takes from the parameter list of the chosen PV module.

Using cell temperature and effective irradiance determined above and given a specified PV module, the PVLIB-Python model further calculates the IV curve, based on which I continue to take the SAPM to calculate the maximum-power and take it as the DC power. The PVLIB-Python model subsequently converts the DC power to AC power, for which I carry on with the SAPM inverter (Boyson et al., 2007; King et al., 2007). Finally, the CF is defined as the ratio of AC power to maximum AC power rating for the inverter at reference conditions. All these calculations are detailed in Appendix D.

1.4 Thesis layout

The overarching goal of my thesis is to use a mixture of data and models to understand PM pollution and PV power output with a focus on their links. The specific goals of my thesis, as outlined in section 1.1, are addressed in the following four research chapters. I present each research chapter in the format of a research article that contains abstract, introduction, (data and) methods, results, discussion and concluding remarks. The first three research chapters are strongly connected with each other, with the previous research chapter lays the foundation for the following one. The last research chapter is slightly independent.

Finally, Chapter 6 reflects on the findings of the separate research chapters and offer suggestions for future studies.

Chapter 2

Global reduction of photovoltaic power output due to particulate matter pollution

This chapter is adapted from part of a manuscript that is, at the time of writing, under review for *Environmental Science & Technology*. I designed and carried out the model experiments, analyzed the results, and wrote the first draft of the manuscript. Prof Paul Palmer supervised the study and provided feedback on subsequent manuscript revisions.

Yao, F. and Palmer, P.I., 2021. Deep cuts in residential emissions substantially benefit Asian photovoltaic power output. Under review for *Environmental Science & Technology*.

2.1 Abstract

Particulate matter (PM) is a mixture of solid particles and liquid droplets that can be suspended in the air and deposited on solar photovoltaic (PV) panels. It scatters and absorbs the solar radiation that would otherwise reach the solar PV panels, thereby compromising our ability to generate carbon-free, clean energy. Here I combine well-evaluated models of solar PV performance and atmospheric composition to provide a global picture of PM impacts on PV energy generation efficiency. Results show that regions with low PV efficiency are typically associated with high PM-induced PV efficiency losses, and that the losses due to deposited PM far exceed those due to atmospheric PM. Desert regions including the Sahara, Arabian-Peninsula, Central Asia, and Southern South America are most susceptible to deposited PM which causes PV efficiency losses that are comparable to the maximum PV efficiency close to 0.3 achieved elsewhere. Coastal regions are also significantly affected by deposited PM, e.g. countries around the Caribbean and the Mediterranean, and over New-Zealand. The main regions where PV efficiency losses due to atmospheric PM are East and South Asia, particularly over highly polluted regions such as North China and the Indo-Gangetic Plain.

2.2 Introduction

Particulate matter (PM) is a mixture of solid particles and liquid droplets and represents a major environmental risk to human health, causing about 3.3 premature deaths annually worldwide (Lelieveld et al., 2015). It is also widely known that PM attenuates incoming solar radiation by scattering and absorbing sunlight before it reaches the surface (Streets et al., 2006; Wild, 2009). It thus logically follows that the amount of solar radiation reaching the solar

photovoltaic (PV) panels will be subsequently decreased (Li, Wagner, Peng, Yang and Mauzerall, 2017; Sweerts et al., 2019). On the other hand, PM can deposit on solar PV panels, causing a further reduction in solar radiation reaching the solar PV cells (Sarver et al., 2013; Boyle et al., 2015; Bergin et al., 2017). In addition, some PM species absorb moisture from the air and can serve as the cloud condensation nuclei, thereby increasing cloud reflectivity (Twomey, 1974; Twomey et al., 1984) and lifetime (Albrecht, 1989) and decreasing the solar radiation reaching the solar PV panels.

Previous studies have attempted to quantify the impacts of PM pollution on PV power output. These studies either use *in situ* measurements (Peters et al., 2018; Sweerts et al., 2019), satellite observations (Li, Wagner, Peng, Yang and Mauzerall, 2017; Li et al., 2020), or general circulation models (GCMs) (Bergin et al., 2017; Labordena et al., 2018) to quantify part of the PM impacts on PV power output with a focus from regional to global scales.

Intuitively, *in situ* measurements simply provide truth of real PV power output. However, the limited spatial coverage of *in situ* measurements is insufficient to unveil the complete picture of real PV power output, and the lack of measurements of PV power output under no PM conditions inevitably leads to indirect experiments that are designed to quantify the impacts of PM pollution on PV power output. For instance, Sweerts et al. (2019) reasonably assumed that there is much less or no PM pollution in the past, so they compared PV power output between past and current atmospheric conditions to determine the solar energy generation losses attributable to air pollution in China. Satellite observations improve upon in situ measurements by filling the data gaps left by the latter, but also suffer from the issue of incapable of observing PV power output under no PM conditions.

By contrast, model simulations can provide both PV power output under real

and no PM conditions whose differences subsequently describe the impacts of PM pollution on PV power output. Moreover, model-based studies have the potential to identify the main sources of PM pollution that reduces PV power output. This is typically achieved by perturbing upstream emissions and comparing resulting PV power output. GCMs are among models that have been used to study the links between PM pollution and PV power output. The advantages of GCMs lie in the accuracy in reproducing large-scale meteorological features and the allowance of climate-chemistry interactions. Likely because of the general absence of radiation modules from the majority of chemical transport models (CTMs), to date limited attempts have been made to use CTMs to quantify PV energy losses attributable to PM pollution. However, the versatility of gas-aerosol chemistry in CTMs provides more flexibility for researchers to modify CTMs to narrow the uncertainty in corresponding PM simulations. Using CTMs to quantify the impacts of PM pollution on PV power output can bear out and refine those using GCMs.

As described in section 1.2.4, GEOS-Chem 'Classic' is a well-known global 3-D model of atmospheric composition (Bey et al., 2001) with a radiation module, which is implemented by coupling to the rapid radiative transfer model for GCMs (RRTMG) (Iacono et al., 2008). This configuration, known as GCRT (Heald et al., 2014), provides a good opportunity to investigate the impacts of PM pollution on PV power output taking into account of more physical and chemical details. The objective of this chapter is exactly to integrate the GCRT and the PVLIB-Python which is a solar PV performance model (Holmgren et al., 2018) to quantify the impacts of PM pollution on PV power output. In the next section I briefly describe the GCRT and the PVLIB-Python with a focus on how to integrate them together. In this section, I also describe the multiple sources of observations that I use to evaluate the integrated model, and the experimental design to isolate the impacts on PV power output due to atmospheric and deposited PM. In section 2.4, I report

the results of the integrated model evaluation and sensitivity experiments. In section 2.5, I conclude this chapter by reflecting our results through comparison to previous findings, outlining limitation, and discussing future research directions.

2.3 Data and methods

Figure 2.1 describes how I integrate v12.9.3 of the GCRT with v0.8.0 of the PVLIB-Python model to estimate PV energy generation efficiency.

Briefly, the GCRT simulates PM dry mass concentrations, and computes global horizontal irradiance (GHI) under control and no PM conditions. It provides PM gravitational and turbulent deposition velocities, and PM mass extinction coefficients. To calculate the irradiance reaching the solar panels and subsequently the solar cells, I combine the GCRT model outputs with surface albedo and precipitation rates (p) from MERRA-2 (Modern-Era Retrospective analysis for Reserach and Applications, Version 2) meteorological analyses, and with solar positions and solar panel configurations determined by the PVLIB-Python model. Using the PVLIB-Python model I calculate PV efficiency from the irradiance received by the solar cells together with ambient temperature and wind speed from MERRA-2 meteorological analyses. The PV efficiency is described by the capacity factor (CF) which is defined as the ratio of the actual power output of a PV panel to the theoretical maximum power output.

With this integrated model approach, it is feasible to investigate the impacts of PM pollution on PV energy generation efficiency by turning on and/or off certain processes and comparing corresponding results. Below I describe the specific configurations that are used by the integrated model, evaluations for the integrated model, and experimental design that quantifies PM impacts on PV energy generation efficiency.

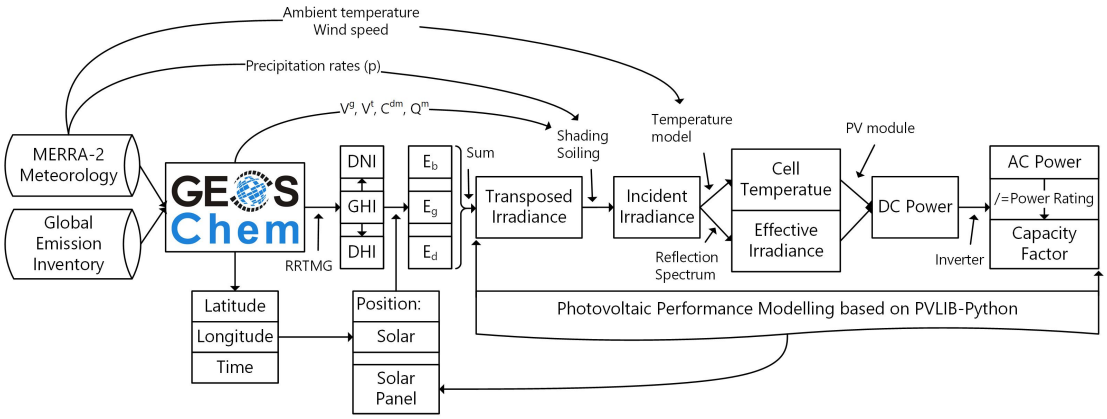


Figure 2.1: Workflow to estimate solar photovoltaic energy generation efficiency in this study. Full definitions for abbreviations include: 1) global horizontal irradiance (GHI) simulated from GEOS-Chem coupled with rapid radiative transfer model for GCMs (RRTMG), and it is subsequently decomposed to direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI); 2) beam (E_b), ground-reflected (E_g), and sky-diffuse (E_d) components of transposed irradiance; 3) PM gravitational (V^g) and turbulent (V^t) velocities; 4) PM dry mass concentrations (C^{dm}) and PM mass extinction coefficients (Q^m); 5) direct (DC) and alternating (AC) current powers.

2.3.1 GEOS-Chem model coupled with rapid radiative transfer model for GCMs

I configure the GCRT (Table 2.1), driven by MERRA-2 meteorological analyses, to provide three-hourly output at a horizontal resolution of 2° latitude × 2.5° longitude from July 2005 through December 2017, of which the first 2.5 years serve as the model spin-up period. As such, my simulations are representative of the 2008–2017 period. I set the atmospheric transport and chemistry time steps to 10 and 20 mins, respectively. I use 47 hybrid-sigma levels from the surface to 0.01 hPa, of which 30 lie below the dynamic tropopause.

I use global anthropogenic emissions from the Community Emissions Data System, updated for the Global Burden of Disease - Major Air Pollutant Sources

project ($CEDS_{GBD-MAPS}$) (McDuffie et al., 2020). $CEDS_{GBD-MAPS}$ provides the most contemporary global emission estimates to date for seven major air pollutants (NO_x; CO; SO_2; NH_3; non-methane volatile organic compounds, NMVOCs; black carbon, BC; organic carbon, OC) as a function of 11 detailed emission source sectors: agriculture (non-combustion sources only), energy generation, industrial processes, on-road and non-road transportation, separate residential, commercial, and other sectors, waste, solvent use, and international shipping. In addition to the $CEDS_{GBD-MAPS}$ anthropogenic emissions, I provide the GCRT simulations with necessary emissions from aircraft (Stettler et al., 2011), open biomass burning (Werf et al., 2017), soil (Hudman et al., 2012) and lightning (Murray et al., 2012) nitrogen oxides, biogenic volatile organic compounds (Guenther et al., 2012) and ammonia (Bouwman et al., 1997), volcanic sulfur dioxide (Carn et al., 2015), mineral (Zender et al., 2003) and anthropogenic (Philip et al., 2017) dust, oceanic sea salt (Gong, 2003; Jaeglé et al., 2011), and others (Riddick et al., 2012; Liang et al., 2010; Ordóñez et al., 2012; Millet et al., 2010; Johnson, 2010; Nightingale et al., 2000; Vinken et al., 2011; Holmes et al., 2014; Sherwen et al., 2016).

As described in section 1.2.4, the GCRT uses the troposphere-only simulation which includes a detailed NO_x-O_x-hydrocarbon-aerosol-bromine-chlorine-iodine chemical mechanism applied from the surface to tropopause. In this chapter, I simply adopt all the default settings to compute GHIs under control and no PM conditions to determine atmospheric PM impacts. The practice of adopting all default settings helps keep a good balance between model accuracy and computational costs which suits for the long study period of 2008–2017. Dry mass concentrations of a total of 17 PM species are simulated. They are sulfate - nitrate - ammonium secondary inorganic aerosol (SNA), hydrophilic (OCPI) and hydrophobic (OCPO) primary organic aerosol (OA), secondary OA (SOA), hydrophilic (BCPI) and hydrophobic (BCPO) BC, accumulation (SALA) and coarse (SALC) mode sea salt (SS), and dust (DU) distributed in seven

size bins. PM dry deposition velocities are outputted and combined with PM dry mass concentrations to determine PM dry mass fluxes on solar panels. PM mass extinction coefficients with humidification effects taken into account for hydrophilic PM (SNA, OCPI, SOA, BCPI, SALA, SALC) at 11 common wavelengths (300, 400, 440, 500, 550, 600, 675, 800, 870, 1,000, and 1,020 nm) are written out and combined with PM dry mass fluxes and elapsed time (relative to 00:00:00 UTC January 1, 2008) to determine the optical depth (τ) and thereby impacts of deposited PM.

Table 2.1: A summary of GCRT model configurations.

Module	Sub-module	Reference
Emissions	Anthropogenic	McDuffie et al. (2020)
	Aircraft	Stettler et al. (2011)
	Open biomass burning	Werf et al. (2017)
	Soil nitrogen oxides	Hudman et al. (2012)
	Lightning nitrogen oxides	Murray et al. (2012)
	Biogenic volatile organic compounds	Guenther et al. (2012)
	Biogenic ammonia	Bouwman et al. (1997)
	Volcanic sulfur dioxide	Carn et al. (2015)
	Mineral dust	Zender et al. (2003)
	Anthropogenic dust	Philip et al. (2017)
	Oceanic sea salt	Gong (2003); Jaeglé et al. (2011) [a]
	Others	Riddick et al. (2012); Liang et al. (2010) Ordóñez et al. (2012); Millet et al. (2010) Johnson (2010); Nightingale et al. (2000) Vinken et al. (2011); Holmes et al. (2014) Sherwen et al. (2016)
Meteorology	MERRA-2	Gelaro et al. (2017)
Chemistry	Secondary inorganic aerosols (SNA) [a]	Park et al. (2004)
	Organic aerosols [b]	Pai et al. (2020)
	Black carbon [c]	Wang, Jacob, et al. (2014) Wang, Heald, et al. (2014)
	Dust in seven size bins	Zhang et al. (2013); Fairlie et al. (2007)
	Sea salt [d]	Jaeglé et al. (2011)
Model output	PM dry mass concentrations [e]	
	PM dry deposition velocities [f]	
	PM mass extinction coefficients [g]	

[a] Sulfate + nitrate + ammonium.
[b] OCPI + OCPO + SOA.
[c] BCPI + BCPO.
[d] SALA + SALC.
[e] Standard GCRT codes output diagnostics for PM dry mass concentrations with some species combined. I have modified part of the GCRT codes (all physical and chemical processes won't be altered) to output separate dry mass concentrations for all PM species.
[f] Standard GCRT codes calculate both gravitational and turbulent deposition velocities but only output diagnostics for the latter. I have modified part of the GCRT codes (all physical and chemical processes won't be altered) to output gravitational dry deposition velocities, too.
[g] Standard GCRT codes calculate PM mass extinction coefficients to enable the calculation of aerosol optical depth but do not output diagnostics for the former. I have modified part of the GCRT codes (all physical and chemical processes won't be altered) to output PM mass extinction coefficients.

2.3.2 Linking GCRT to PVLIB-Python

As illustrated in Figure 2.1, the incident irradiance prepared from the GCRT model used by the PVLIB-Python model is the bridge of the two models. More specifically, the GCRT simulates GHI which is subsequently used to estimate the direct normal irradiance (DNI) and the diffuse horizontal irradiance (DHI). DNI, GHI, and DHI then derive the beam (E_b), ground-reflected (E_g), and sky-diffused (E_s) components of the irradiance transposed to the solar PV panels. More details of these calculations can be found in Appendices B and C.

E_b, E_g, and E_s collectively constitute the transposed or 'in' plane-of-array irradiance (POAI_{in}) that is subsequently reduced to incident or 'out' plane-of-array irradiance (POAI_{out}) received by the solar PV cells. POAI_{in} and POAI_{out} are linked by a soiling factor as mentioned in section 1.3.3. An intuitive way to represent this soiling factor is to use the deposited PM's τ. Inheriting from the wavelength dependence of PM mass extinction coefficients, I separately combine wavelength-specific τ with POAI_{in} that represents the entire shortwave flux to calculate wavelength-specific $POAI_{out} = POAI_{in} \times e^{-\tau}$. With the weights from a black body spectrum ($T = 5778\ K$), these POAI_{out}s are subsequently averaged to a single POAI_{out}, which constitutes the basis of final CF calculations.

For each wavelength, $\tau = \sum_{i=1}^{17}(Q_i^m \times PM_i)$, where i denotes a PM species. Q_i^m, extracted from the lowest GCRT model layer where solar PV panels are located, is the ground-level mass extinction coefficient for PM species i. PM_i is the accumulated deposited dry mass for PM species i and is the net combination of PM accumulation and removal processes occurred on solar PV panels, i.e. $PM_i = PM_i^{Accum} - PM_i^{Removal}$.

PM_i^{Accum} comes from PM dry deposition processes. $PM_i^{Accum} = \int_t (V_i^g cos(\theta_T) + V_i^t) C_i^{dm} \mathrm{d}t$, where V_i^g and V_i^t are gravitational and turbulent dry deposition

velocities for PM species i, respectively. C_i^{dm}, extracted from the lowest GCRT model layer where solar PV panels are located, is the ground-level dry mass concentration for PM species i. The gravitational velocity is vertical, so I reduce it on tilted solar panels by multiplying $cos(\theta_T)$, where θ_T is the tilt angle of solar PV panels.

$PM_i^{Removal}$, based on p which shows low bias, high correlation, and a realistic diurnal cycle compared with observations from the Global Precipitation Project version 2.2 (Reichle et al., 2017), follows Li et al. (2020):

- When $p < 1\ mm\ h^{-1}$, no PM removal occurs.
- When $1 < p < 3\ mm\ h^{-1}$, SNAs are entirely removed and half of the OAs are removed.
- When $3 < p < 5\ mm\ h^{-1}$, SNAs are entirely removed and half of all other PMs are removed.
- When $p > 5\ mm\ h^{-1}$, all PMs are removed.

2.3.3 PVLIB-Python model

The PVLIB-Python model is a community supported tool that provides a set of functions and classes for simulating the performance of solar PV energy systems. PVLIB-Python v0.8.0 currently supports performance modelling of horizontal fixed (Flat), fixed with optimal tilt (Tilt), and one-axis tracking (OAT) panels and they are simultaneously investigated in this chapter and this thesis. Their characteristics are given in Table 1.2 in section 1.3.3.

For each panel, the PVLIB-Python model applies different solar panel configurations (i.e. tilt angle and azimuth of solar panels) to transpose solar radiative fluxes to irradiance received by the solar PV panels, POAI_{in}.

The PVLIB-Python model takes the POAI_{out} reduced from POAI_{in} and ambient temperature and wind speed from the MERRA-2 meteorological analyses as inputs to calculate the cell temperature and effective irradiance, further uses the PV module (Canadian_Solar_CS5P_220M___2009_) to calculate the direct (DC) power (King et al., 2004), and finally applies the inverter (ABB__MICRO_0_25_I_OUTD_US_208__208V_) to calculate the alternating (AC) power.

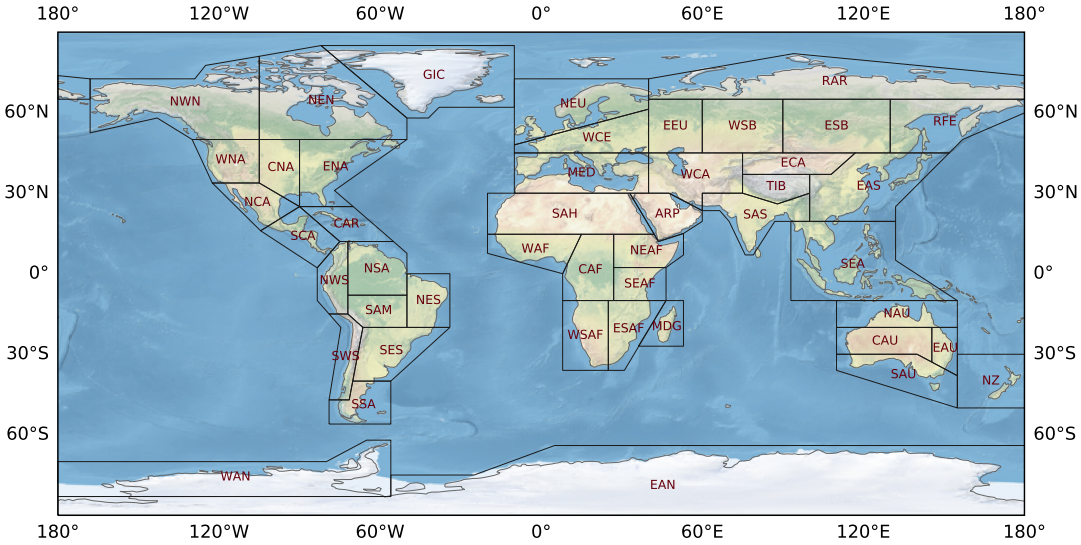
In this process, both the PV cell efficiency (solar energy to DC power, 12.94%) and the inverter efficiency (DC to AC power, 96%) are considered. I model a single module that contains 96 cells in series and ignore potential electricity losses due to, for example, degradation of modules and inverters. I use this approach in the absence of an established model describing these losses. I divide the AC power by the AC power rating of the inverter (250 W) to obtain CF that describes PV efficiency. More details of PVLIB-Python calculations can be found in section 1.3.3.

2.3.4 Experimental design

To quantify the reduction of PV efficiency caused by atmospheric and deposited PM, I calculate three CFs: 1) $CF1$ includes both atmospheric and deposited PM impacts, namely real PV efficiency; 2) $CF2$ includes atmospheric PM impacts only; and 3) $CF3$ that does not include atmospheric or deposited PM impacts. I am then able to isolate atmospheric PM impacts as $CF3 - CF2$, deposited PM impacts as $CF2 - CF1$, and total PM impacts as $CF3 - CF1$, respectively. For simplicity, I term them as PM dimming, soiling, and total impacts, respectively.

Throughout this chapter, I analyse the decadal mean (2008–2017) PV efficiency and PM impacts at each 2° latitude × 2.5° longitude grid for their spatial

distributions, and further calculate the regional area-weighted mean values for regional characteristics (and interannual variabilities). I use the latest IPCC climate reference regions (Iturbide et al., 2020) (Figure 2.2) for regional synthesis.



0	GIC	Greenland/Iceland	23	NEAF	N.Eastern-Africa
1	NWN	N.W.North-America	24	SEAF	S.Eastern-Africa
2	NEN	N.E.North-America	25	WSAF	W.Southern-Africa
3	WNA	W.North-America	26	ESAF	E.Southern-Africa
4	CNA	C.North-America	27	MDG	Madagascar
5	ENA	E.North-America	28	RAR	Russian-Arctic
6	NCA	N.Central-America	29	WSB	W.Siberia
7	SCA	S.Central-America	30	ESB	E.Siberia
8	CAR	Caribbean	31	RFE	Russian-Far-East
9	NWS	N.W.South-America	32	WCA	W.C.Asia
10	NSA	N.South-America	33	ECA	E.C.Asia
11	NES	N.E.South-America	34	TIB	Tibetan-Plateau
12	SAM	South-American-Monsoon	35	EAS	E.Asia
13	SWS	S.W.South-America	36	ARP	Arabian-Peninsula
14	SES	S.E.South-America	37	SAS	S.Asia
15	SSA	S.South-America	38	SEA	S.E.Asia
16	NEU	N.Europe	39	NAU	N.Australia
17	WCE	West&Central-Europe	40	CAU	C.Australia
18	EEU	E.Europe	41	EAU	E.Australia
19	MED	Mediterranean	42	SAU	S.Australia
20	SAH	Sahara	43	NZ	New-Zealand
21	WAF	Western-Africa	44	EAN	E.Antarctica
22	CAF	Central-Africa	45	WAN	W.Antarctica

Figure 2.2: Definitions of 46 AR6 land regions used in this thesis to calculate regional area-weighted mean variables.

2.3.5 Model evaluation

I evaluate the integrated model from three perspectives by using a range of *in situ* observations. They are: 1) hourly GHI observations from the Baseline Surface Radiation Network (BSRN) (Driemel et al., 2018), partly corresponding to real PV efficiency $CF1$; 2) column aerosol optical depth (AOD) observations at every 15 minutes interval from the NASA AErosol RObotic NETwork (AERONET) (Giles et al., 2019), partly corresponding to PM dimming impacts $CF3 - CF2$; and 3) hourly surface $PM_{2.5}$ (PM with aerodynamic diameter less than 2.5 μm) concentration observations falling within [0,3000] μg m^{-3} from the China National Environmental Monitoring Center (CNEMC) and OpenAQ, and its major chemical components observations mainly at monthly intervals from the literature (Zhang et al., 2019), partly corresponding to PM soiling impacts $CF2 - CF1$. I sample model values with the locations and periods of observed values, and for GHI, AOD, and $PM_{2.5}$, I further compile the paired model and observed values into monthly composites to make the comparisons. I compare AODs at 550 nm, to which I interpolate AERONET AOD using a second-order polynomial fit of $\ln(AOD) \sim \ln(wavelength)$ (Eck et al., 1999) that requires at least three data pairs encircling 550 nm. Model AODs at 550 nm are calculated based on dry mass concentrations of PM species and associated optical properties, as described in section 1.2.4. Model values of $PM_{2.5}$ and its major chemical components are calculated at the 35% relative humidity, 298 K, and 1013.25 hPa condition, as also described in section 1.2.4.

To describe the comparisons between model simulations and *in situ* observations I use the Pearson correlation coefficients, normalized mean bias $NMB = \frac{\sum_1^N (M-O)}{\sum_1^N O} 100\%$, and normalized root mean square error $NRMSE = \frac{\sqrt{\frac{1}{N}\sum_1^N (M-O)^2}}{O_{max}-O_{min}} 100\%$, where M and O denote model and observed values, respectively, O_{max} and O_{min} denote the maximum and minimum observed values, respectively, and N denotes the number of comparison points. I report model performance statistics on both annual and seasonal scales to evaluate the integrated model capability during both periods of high and low solar insolation.

2.4 Results

2.4.1 Results of model evaluation

Figure 2.3 shows the site-scale evaluation of simulated GHI against BSRN GHI between 2008 and 2017. For Pearson correlation coefficients, approximately 98%, 93%, 93%, 94%, and 95% of the sites report statistically significant ($p < 0.05$) values, 100%, 98%, 91%, 97%, and 97% of which are no less than 0.7, during the entire, spring, summer, autumn, and winter period, respectively. For NMB, approximately 94%, 98%, 87%, 85%, and 77% of the sites report values within ±30% during the entire, spring, summer, autumn, and winter period, respectively. For NRMSE, approximately 94%, 70%, 63%, 77%, and 69% of the sites report values no larger than 30% during the entire, spring, summer, autumn, and winter period, respectively.

Figure 2.4 shows the site-scale evaluation of simulated AOD against AERONET AOD between 2008 and 2017. For Pearson correlation coefficients, approximately 70%, 54%, 50%, 51%, and 46% of the sites report statistically significant ($p < 0.05$) values, 55%, 66%, 63%, 73%, and 62% of which are no less than 0.7, during

the entire, spring, summer, autumn, and winter period, respectively. For NMB, approximately 84%, 86%, 82%, 80%, and 77% of the sites report values within ±50% during the entire, spring, summer, autumn, and winter period, respectively. For NRMSE, approximately 81%, 72%, 59%, 75%, and 63% of the sites report values no larger than 50% during the entire, spring, summer, autumn, and winter period, respectively.

Figure 2.5 shows the site-scale evaluation of simulated $PM_{2.5}$ against CNEMC and OpenAQ $PM_{2.5}$ between 2008 and 2017. For Pearson correlation coefficients, approximately 86%, 19%, 40%, 35%, and 52% of the sites report statistically significant ($p < 0.05$) values, 51%, 46%, 55%, 58%, and 70% of which are no less than 0.7, during the entire, spring, summer, autumn, and winter period, respectively. For NMB, approximately 70%, 54%, 55%, 62%, and 79% of the sites report values within ±50% during the entire, spring, summer, autumn, and winter period, respectively. For NRMSE, approximately 87%, 33%, 33%, 53%, and 60% of the sites report values no larger than 50% during the entire, spring, summer, autumn, and winter period, respectively.

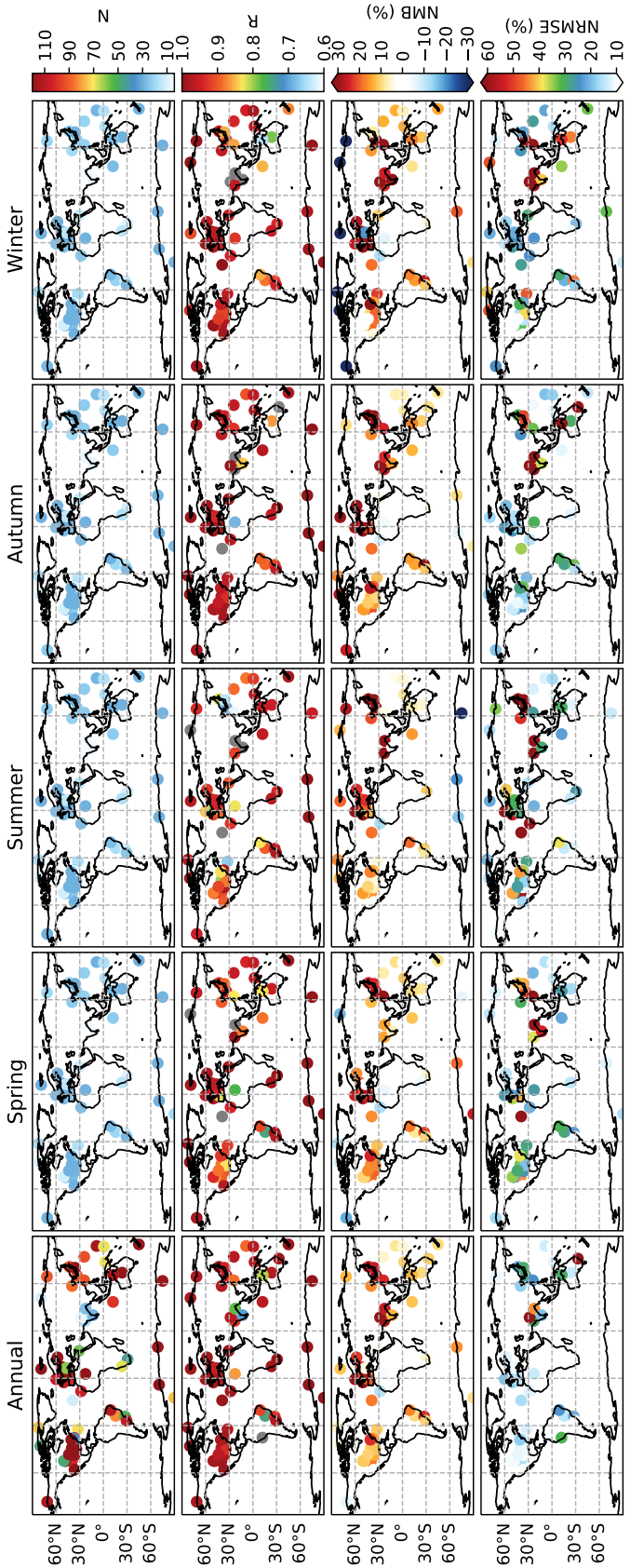


Figure 2.3: Site-scale evaluation of simulated GHI against BSRN GHI observations between 2008 and 2017. Pearson correlation coefficients with $p \geq 0.05$ are marked as grey dots.

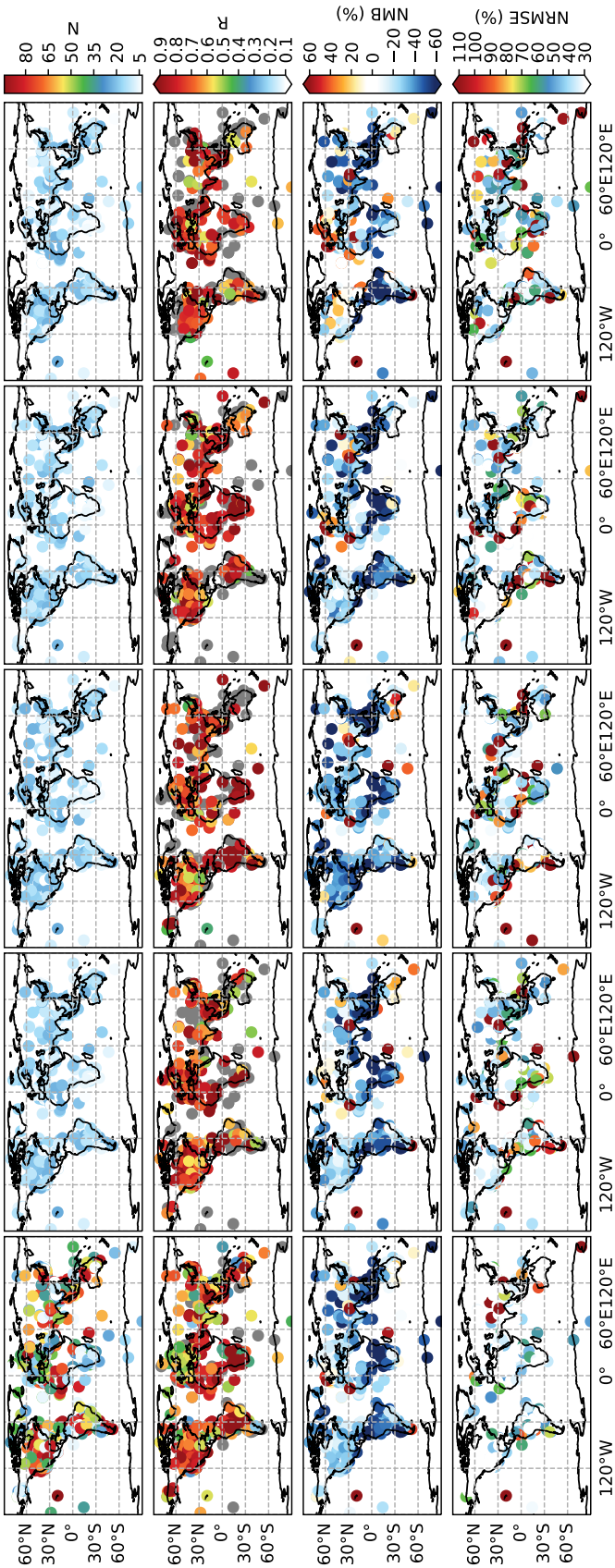


Figure 2.4: Site-scale evaluation of simulated AOD against AERONET GHI observations between 2008 and 2017. Pearson correlation coefficients with $p \geq 0.05$ are marked as grey dots.

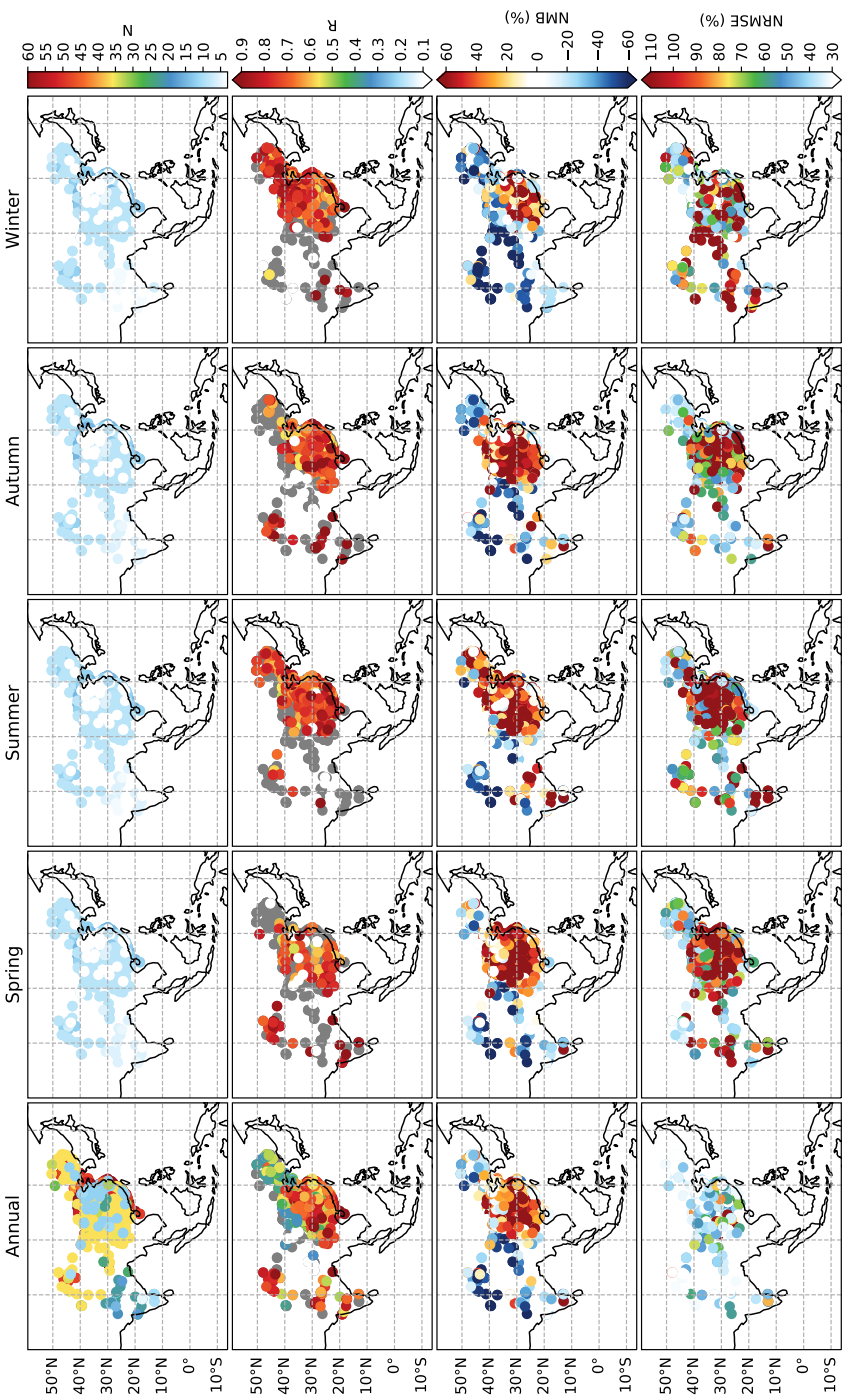


Figure 2.5: Site-scale evaluation of simulated $PM_{2.5}$ against CNEMC and OpenAQ $PM_{2.5}$ observations between 2008 and 2017. Pearson correlation coefficients with $p \geq 0.05$ are marked as grey dots.

Figure 2.6 shows the evaluation of simulated $PM_{2.5}$ chemical composition concentrations against ground observation data collected from the literature (Zhang et al., 2019). On the annual scale, the Pearson correlation coefficients, NMB, and NRMSE between model and observed values range from 0.42 to 0.59, from -29% to 14%, and from 9% to 16% for different species, respectively. Admittedly, the Pearson correlation coefficients are not particularly high but most of them are statistically significant ($p < 0.05$) and comparable to those reported in prior studies (Miao et al., 2020). On the seasonal scale, both superior and inferior statistics exist, reflecting the possible varying model capability in different seasons but note that the decreasing number of observations may also result in some fluctuations.

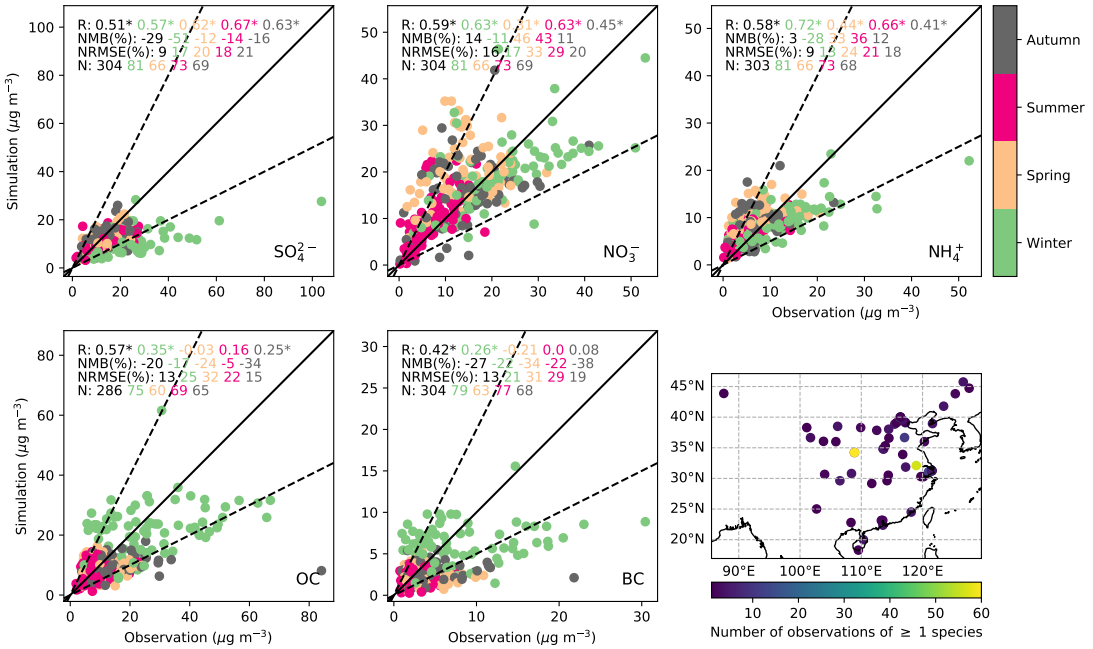


Figure 2.6: Evaluation of simulated $PM_{2.5}$ chemical composition concentrations against ground-based observations collected from the literature (Zhang et al., 2019). The solid line corresponds to the 1:1 line, and the dashed lines correspond to the 1:2 and 2:1 lines. Statistically significant ($p < 0.05$) Pearson correlation coefficients are marked with a $*$. The bottom right subplot shows the number of observations of at least one $PM_{2.5}$ chemical component over each monitoring location.

The above statistics suggest that the integrated model has generally well reproduced the observed variations in GHI, AOD, and $PM_{2.5}$ and its major chemical composition concentrations with relatively small bias and errors during both periods of high and low solar insolation. In particular, the integrated model has the best performance for GHI, followed by AOD then $PM_{2.5}$ and its chemical composition concentrations. This may suggest the higher accuracy in real PV efficiency than PM dimming and soiling impacts. Uncertainties in the bottom-up inventories and the underlying physical and chemical processes could be the plausible explanation, but arguably, this is somewhat ubiquitous in most process-based models.

Three points about the model evaluation need to be fairly highlighted. First, the model evaluation presented here is not particularly rigorous due to the discrepancies in temporal and spatial resolutions between simulations and observations that lead to the nearest neighbour-based data sampling and thereby inevitable associated uncertainties. Second, although each source of observations is temporally and spatially incomplete, it could be argued that the various sources of observations collectively supplement each other to evaluate the integrated model. Third, $PM_{2.5}$ and its major chemical components exclude those particles with an aerodynamic diameter above 2.5 μm, so may not be the most appropriate metric to evaluate the integrated model. But they appear to be the most available metric given that my immediate focus in this work is anthropogenic sources of PM which mainly contribute to $PM_{2.5}$, and that the well-documented harmful impacts of $PM_{2.5}$ and its major chemical components on human health lead to their abundant data sources.

2.4.2 Results of PV efficiency and PM impacts

Figure 2.7 shows the geographical distributions of decadal mean (2008–2017) PV efficiency and its losses due to atmospheric and deposited PM for flat, tilt, and OAT panels. Globally, flat panels have an area-weighted mean CF of 0.08, with high values distributed over North and South America, Eastern and Southern Africa, the Tibetan-Plateau, Southeast Asia, Australia, and Madagascar. Tilt panels improve the global area-weighted mean CF to 0.09 by enhancing values over high latitudes. The largest value enhancements are particularly over Greenland and Antarctic, where CFs are improved by more than 70%. OAT panels further improve the global area-weighted mean CF to 0.12, with relatively similar value enhancements everywhere.

Regions with low CFs (Figure 2.7a–c) are typically associated with high PM total impacts (Figure 2.7d–f). This is supported by the statistically significant ($p < 0.05$) negative Pearson correlation coefficients between regional area-weighted mean PV efficiency and PM impacts for flat (-0.7), tilt (-0.81), and OAT (-0.83) panels (Figure 2.8). This partly indicates that PM pollution is exactly a major factor negatively affecting PV efficiency.

The three panels present similar spatial distributions of PM total impacts, with OAT panels having the largest values, followed by tilt and flat panels. Desert regions including the Sahara, Arabian-Peninsula, Central Asia, and Southern South America report high PM total impacts that are comparable to the maximum PV efficiency which can be achieved elsewhere (e.g. ≈ 0.3 in OAT panels). Coastal regions are also significantly affected. These include countries around the Caribbean and the Mediterranean, and over New-Zealand. East and South Asia report a middle to large magnitude of PM total impacts, which are collectively contributed by PM dimming and soiling impacts, as described below.

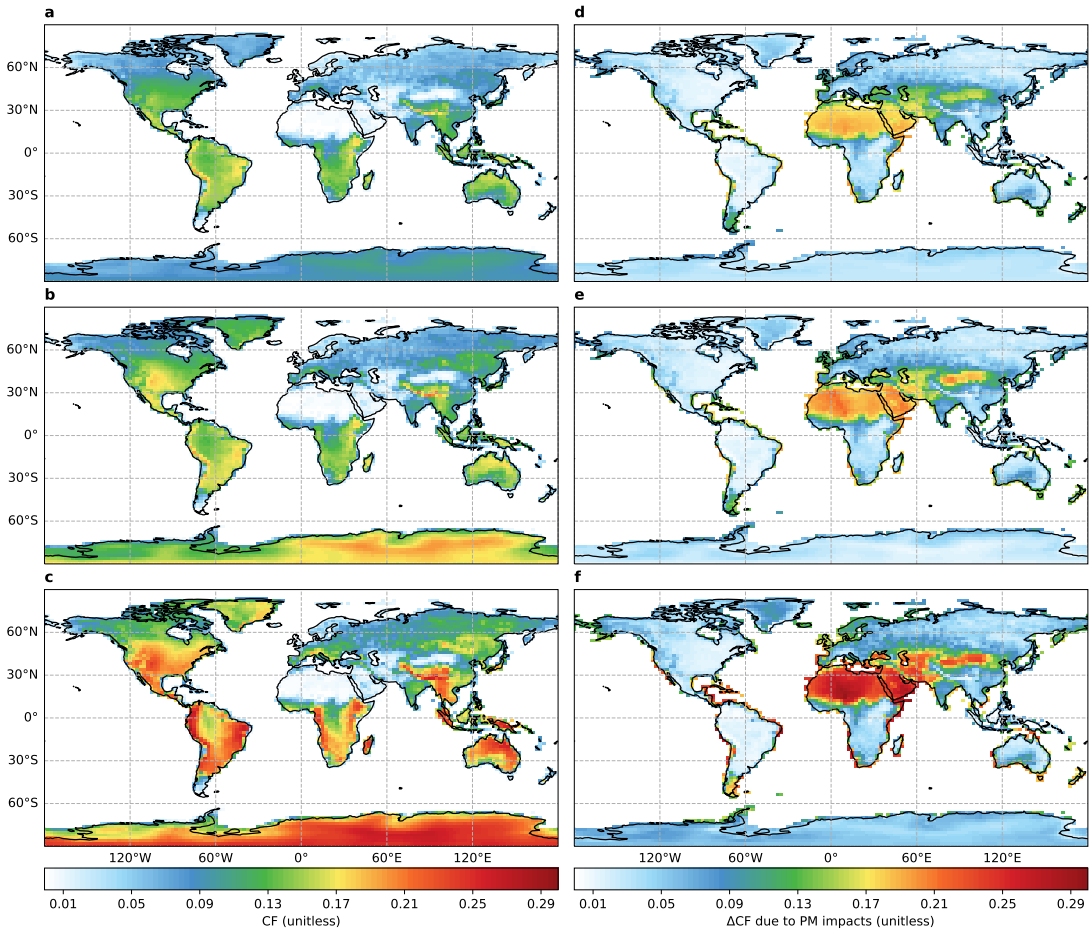


Figure 2.7: Geographical distributions of decadal mean (2008–2017) (a–c) PV efficiency and (d–f) its losses due to atmospheric and deposited PM for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panel.

2.4.3 Results of PM dimming versus soiling impacts

Figure 2.9a–c shows the geographical distributions of decadal mean (2008–2017) PM dimming impacts for flat, tilt, and OAT panels. The three panels present similar spatial patterns for PM dimming impacts, with larger values found for OAT than tilt or flat panels. The main regions where PM dimming impacts are as high as 0.04 are East and South Asia, particularly over highly-polluted regions such as North China and the Indo-Gangetic Plain, consistent with previous studies

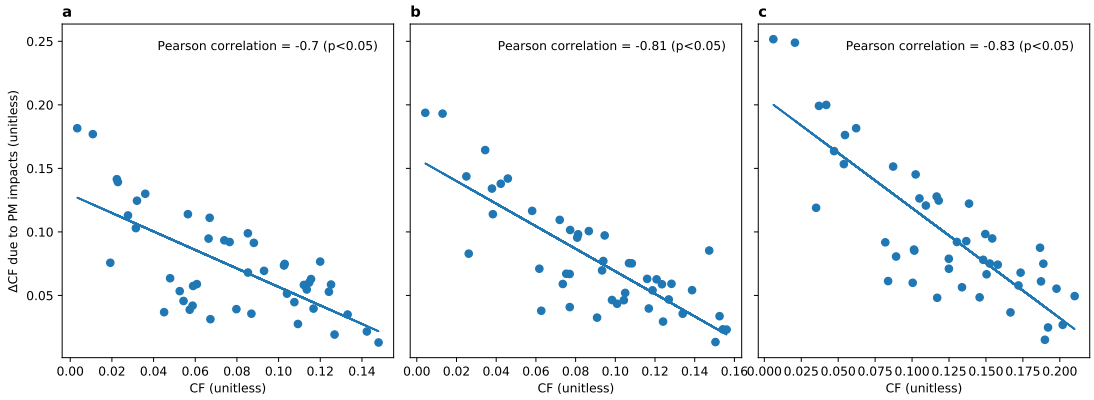


Figure 2.8: Comparison of regional area-weighted averages of decadal mean (2008–2017) PV efficiency and PM total impacts for (a) flat, (b) tilt, and (c) one-axis tracking panels. Please refer to Figure 2.2 for definitions of regions that are used here for taking regional area-weighted averages.

(Li, Wagner, Peng, Yang and Mauzerall, 2017; Li et al., 2020). Other regions where PM dimming impacts are moderate at ΔCF levels of 0.01 include West and Central Africa.

Figure 2.9d–f shows the geographical distributions of decadal mean (2008–2017) PM soiling impacts for flat, tilt, and OAT panels. Figure 2.9d–f quite resembles Figure 2.7d–f. This partly indicates that PM total impacts are mostly determined by PM soiling impacts in most regions around the world, in agreement with prior studies (Li et al., 2020). In this sense, I can readily report that the areas with large PM soiling impacts are desert and coastal regions. The former includes the Sahara, Arabian-Peninsula, Central Asia, and Southern South America. The latter includes countries around the Caribbean, the Mediterranean, and over New-Zealand.

While PM soiling impacts dominate over PM dimming impacts and the maximum magnitude of the former can be almost eight times that of the latter (see the upper value of the colorbar in Figures 2.9a–c and 2.9d–f), they can be

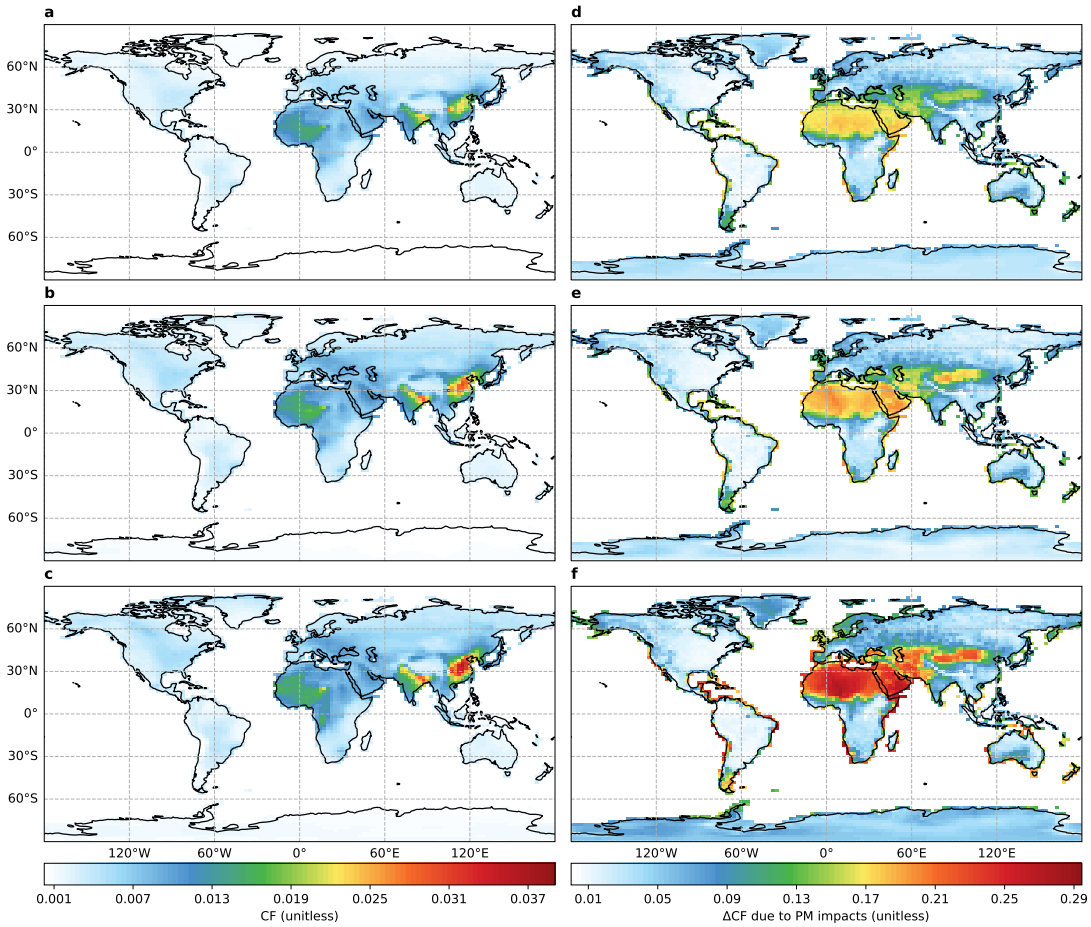


Figure 2.9: Geographical distributions of decadal mean (2008–2017) PV efficiency losses due to (a–c) atmospheric and (d–f) deposited PM for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels.

positively correlated with each other. This is theoretically because the larger PM dimming impacts, the higher probability of higher surface PM concentrations, meaning the higher probability of higher PM deposition fluxes, and thereby the higher probability of larger PM soiling impacts. The opposite is also logically reasonable. PM chemical and/or vertical distributions should be the main factors strengthening and/or weakening this theoretical reasoning.

Figure 2.10 shows the comparison of regional area-weighted averages of decadal mean (2008–2017) PM dimming and soiling impacts for the three panels. The

regression lines between PM dimming and soiling impacts is largely skewed from the 1:1 reference lines, once again suggesting the dominance of PM soiling over dimming impacts. Nonetheless, there are statistically significant ($p < 0.05$) positive Pearson correlations between regional area-weighted mean PM dimming and soiling impacts in falt, tilt, and OAT panels. The corresponding Pearson correlation coefficients are 0.44, 0.4, and 0.32, respectively. This suggests that PM dimming and soiling impacts are generally coincident, providing an opportunity to reduce them simultaneously by decreasing upstream emissions, as targeted in the next chapter. The decreasing order of the Pearson correlation coefficients from flat to tilt to OAT panels suggests that this coincidence is subject to panel configurations. Likely because flat panels collects more PM gravitational deposition masses than tilt or OAT panels, its coincidence of PM dimming and soiling impacts is stronger than that of tilt or OAT panels.

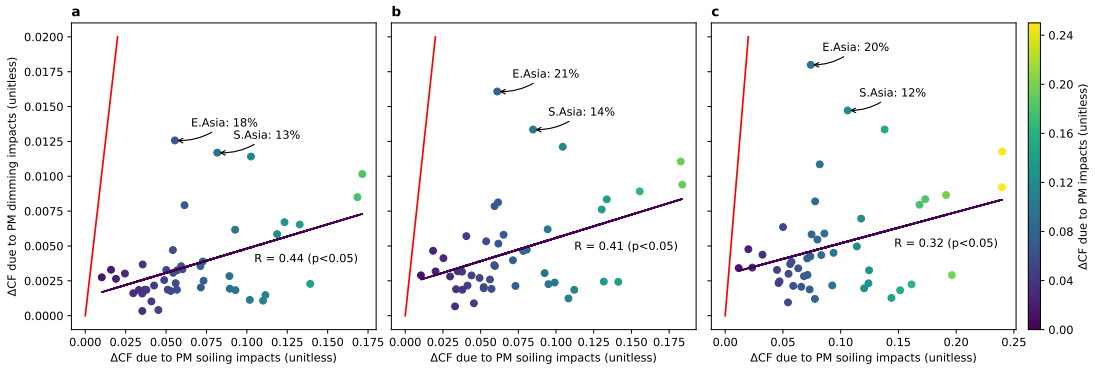


Figure 2.10: Comparisons of regional area-weighted averages of decadal mean (2008–2017) PM dimming and soiling impacts for (a) flat, (b) tilt, and (c) one-axis tracking panels. The blue lines are the regression lines between PM dimming and soiling impacts, with R denoting the Pearson correlation coefficients. The red lines are the 1:1 lines. The vertical colorbar represents PM total impacts. The percentages following E.Asia and S.Asia texts are the non-negligible shares of PM dimming impacts over these two regions. Please refer to Figure 2.2 for definitions of regions that are used here for taking regional area-weighted averages.

2.4.4 Role of precipitation

I follow Li et al. (2020) in using the precipitation as the sole natural mechanism to reduce the impacts of PM soiling. I find that the strongest PM soiling impacts over desert regions are a result of rapid accumulation of dust (Figure 2.11) deposited on the solar PV panels and of limited removal by precipitation (Figure 2.12).

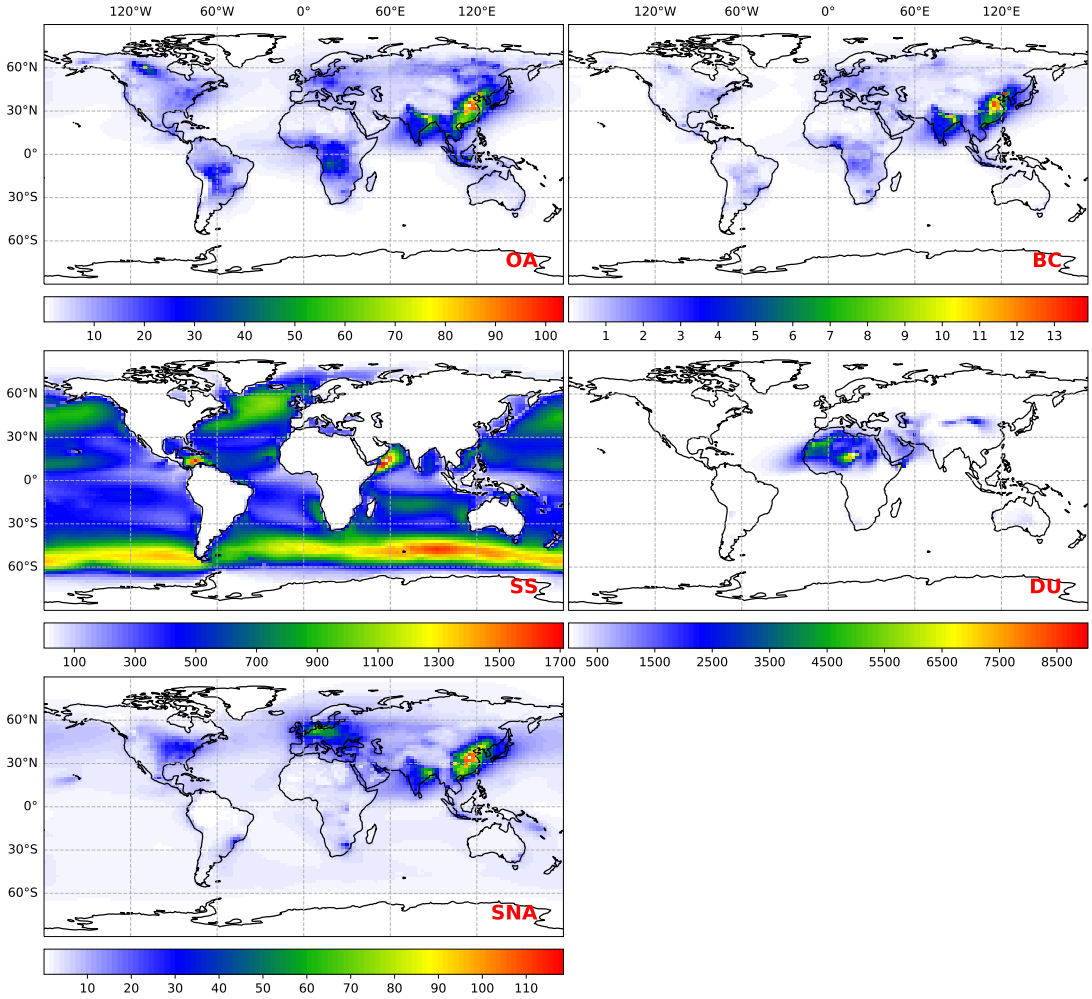


Figure 2.11: Geographical distributions of decadal mean (2008–2017) deposition fluxes ($\mu g/m^2/hr$) for each PM species: organic aerosol (OA), black carbon (BC), sea salt (SS), dust (DU), and sulfate-nitrate-ammonium (SNA) at 2° latitude × 2.5° longitude resolution.

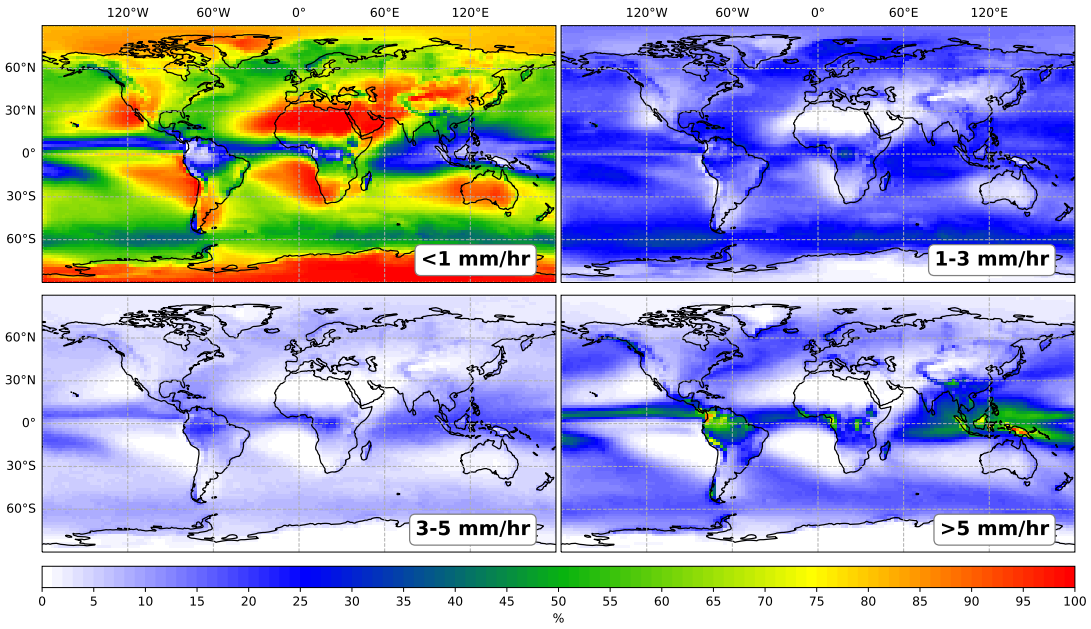


Figure 2.12: Frequency of precipitation rates of < 1 mm/hr, 1–3 mm/hr, 3–5 mm/hr, and > 5 mm/hr from MERRA-2 between 2008 and 2017.

Moreover, I determine the influence of precipitation on reducing the impacts of PM soiling by comparing PV efficiency in model runs with and without the influence of precipitation (Figure 2.13). I find that precipitation plays an important role in shaping the spatial pattern of current-level PV efficiency whose values would otherwise be reduced by more than 70% over resource-abundant regions, excluding Greenland and Antarctica.

2.5 Discussion and concluding remarks

To the best of my knowledge, this chapter is one of the earliest studies that integrates models of solar PV performance and atmospheric composition to quantify the impacts of PM pollution on PV power output on a decadal, global scale. The integrated model performs well in comparison with the several sources of *in situ* observations. The derived spatial patterns for PV efficiency and PM

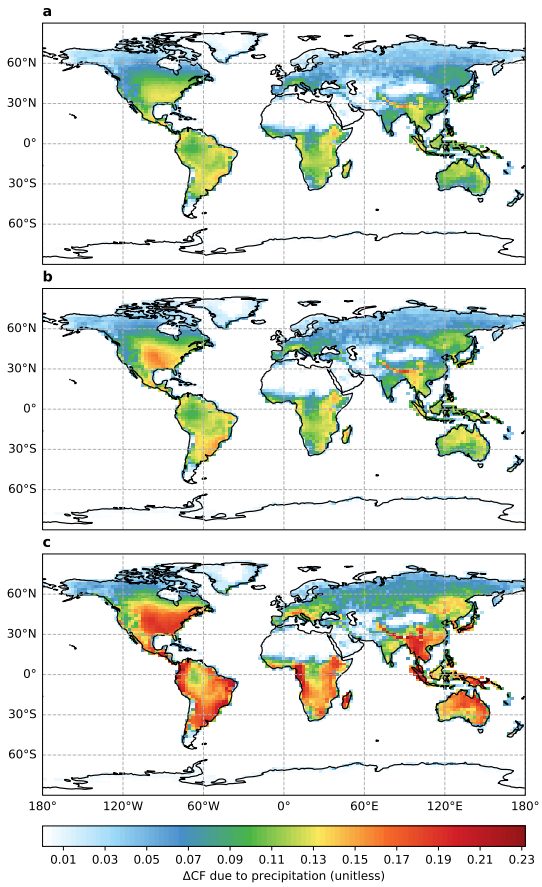


Figure 2.13: Geographical distributions of decadal mean (2008–2017) cleaning benefits resulting from precipitation for (a) flat, (b) tilt, and (c) one-axis tracking panels.

impacts are broadly consistent with recent studies such as Li et al. (2020). Nonetheless, coastal regions with high PM impacts missing from previous studies are additionally identified in this chapter. This is likely due to the use of the GCRT model which considers a wider range of PM constitutes, e.g. sea salt (Methods). Moreover, I spot Southern South Asia as another desert region where PV efficiency is impacted by PM pollution. All of these lay the foundation for the next chapter in which I further identify the leading anthropogenic sources of PM pollution that reduce PV power output.

My integrated model is subject to several important limitations and sources of uncertainty.

First, the use of the offline meteorological data particularly the offline clouds leads to the impossibility of isolating and quantifying the impacts of aerosol-cloud interactions on PV power output. This is not ideal considering the increasing importance of aerosol-cloud interactions in radiative forcing (Jia et al., 2021). This issue will extend to the next chapter where the altered emissions will have no impact on clouds. Nonetheless, the model evaluation against a wide range of *in situ* measurements and the associated statistics have provided me with confidence of the integrated model skill. With the development and release of WRF-GC (v2.0), an online two-way coupling of the Weather Research and Forecasting meteorological model and the GEOS-Chem model (Feng et al., 2021), it is expected that the integrated model, once updated to the combination of WRF-GC, RRTMG, and PVLIB-Python, will include more realism of the assessment by presenting the changes in PV power output within the context of co-varying aerosols and climate.

Second, the broadband irradiance intercepted by solar panels is combined with spectrally resolved optical depth of deposited PM to determine multiple reduced broadband irradiance that are subsequently averaged into a single mean value based on a black body spectrum before it is utilized by solar cells to calculate PV efficiency. This practice is adopted not only to directly use the broadband global horizontal irradiance the integrated model provided, but more importantly due to the general absence of well-established models that can estimate spectrally resolved direct normal irradiance and diffuse horizontal irradiance from global horizontal irradiance. However, the use of the black body spectrum as the weights may only be valid in clear-sky, low aerosol optical depth conditions when the majority of irradiance would be direct. To test whether the results would significantly change if other weights are used, I re-average the multiple reduced

broadband irradiance but with equal weights, and then repeated the calculation of PV efficiency. Figure 2.14 shows that on the decadal scale, the changes from the weighted to the unweighted PM soiling impacts are negative but tiny in most areas of the world. Part of the grids over East and South Asia, Greenland, and the Antarctic show slightly pronounced changes, but I have calculated that on the regional scale the changes are all within 6%.

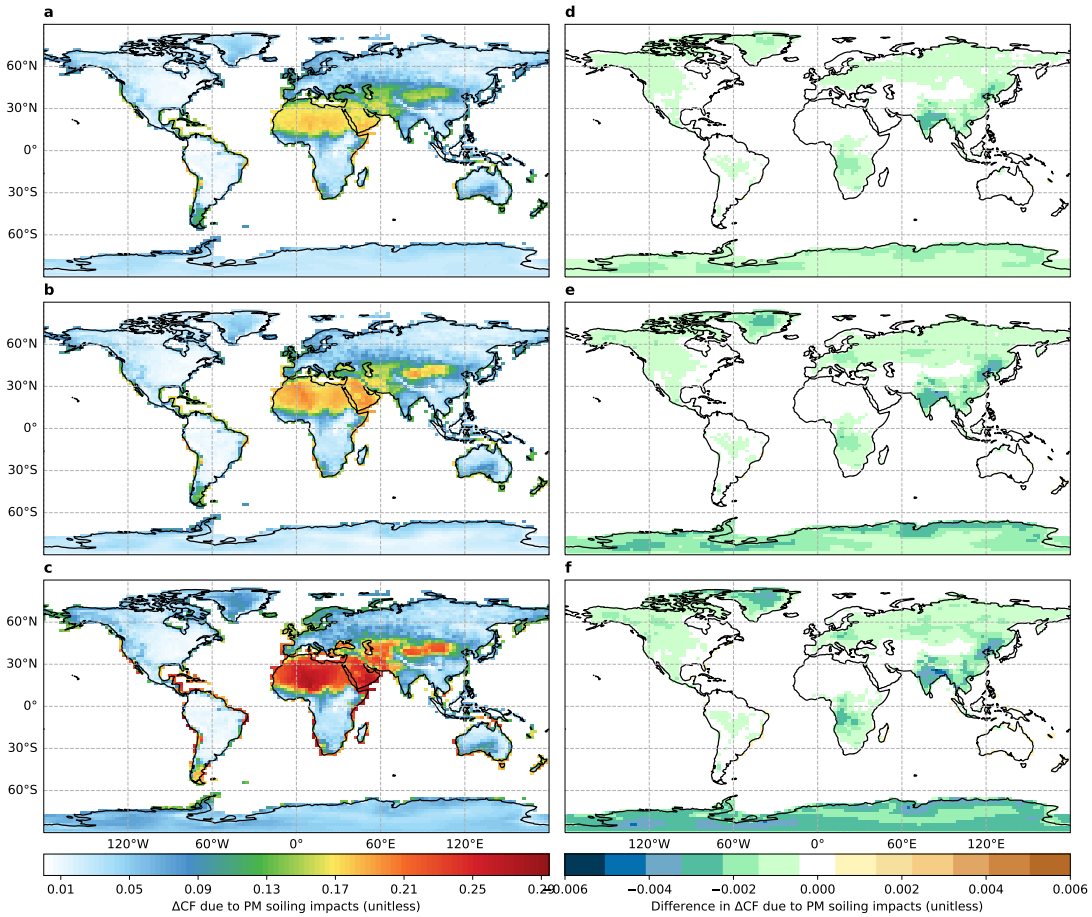


Figure 2.14: Geographical distributions of decadal mean (2008–2017) PV efficiency losses due to (a–c) deposited PM updated with the equal weights for $POAI_{out}$ and (d–f) its changes from Figure 2.9d–f for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels.

Chapter 3

Deep cuts in residential emissions substantially benefit Asian photovoltaic power output

This chapter is adapted from part of a manuscript that is, at the time of writing, under review for *Environmental Science & Technology*. I designed and carried out the model experiments, analyzed the results, and wrote the first draft of the manuscript. Prof Paul Palmer supervised the study and provided feedback on subsequent manuscript revisions.

Yao, F. and Palmer, P.I., 2021. Deep cuts in residential emissions substantially benefit Asian photovoltaic power output. Under review for *Environmental Science & Technology*.

3.1 Abstract

Particulate matter (PM) in the atmosphere and deposited on solar photovoltaic (PV) panels reduce PV energy generation. Reducing anthropogenic PM emissions will therefore increase carbon-free energy generation and as a co-benefit will improve surface air quality. However, we lack a global understanding of the sectors that would be the most effective at achieving the necessary reductions in PM emissions. Here I combine well-evaluated models of solar PV performance and atmospheric composition to show that reducing residential emissions is the most effective approach to mitigate PM-induced PV energy losses, and that the biggest PV energy gains are over East and South Asia. Using 2019 PV capacities as a baseline, I find that a 50% reduction in residential emissions would lead to an additional 7,687 GWh yr^{-1} (US\$653 million yr^{-1}) and 1,823 GWh yr^{-1} (US\$144 million yr^{-1}) produced in China and India, respectively.

3.2 Introduction

Our harnessing energy provided for free by the Sun has a low environmental footprint and will therefore play a role in reducing emissions of greenhouse gases and mitigating the harmful impacts of climate change (Hernandez et al., 2014; Creutzig et al., 2017). A variety of technologies convert sunlight to usable electricity, but currently the most common approach is to use solar photovoltaic (PV) panels. The last decade (2011 to 2020) has seen an enormous increase in the worldwide solar PV installed capacity, from 72 to 707 GW (IRENA, 2021). Solar PV is expected to dominate growth in the renewable energy sector for the foreseeable future (IEA, 2020). However, particulate matter (PM), a mixture of solid particles and liquid droplets suspended in the air, represents a major barrier to maximizing the performance of solar PV technologies and therefore

compromises our ability to generate clean energy. Atmospheric PM scatters and absorbs the solar radiation that would otherwise reach the solar panels (Li, Wagner, Peng, Yang and Mauzerall, 2017; Peters et al., 2018; Sweerts et al., 2019). PM deposited on the solar panels further impedes the solar radiation being received by the PV semiconductor material (Sarver et al., 2013; Bergin et al., 2017; Li et al., 2020).

PM is released directly into the atmosphere via processes such as combustion (primary source) and is formed in the atmosphere from the condensation of low volatility gases (secondary source) (Fuzzi et al., 2015; Weagle et al., 2018). These primary and secondary sources of PM are emitted from a wide range of anthropogenic activities, e.g. production of iron and steel, cement, and glass. Reductions in these emissions are required to improve the energy generation performance of solar PV cells (Sweerts et al., 2019; Labordena et al., 2018), but it remains unclear which are the most effective source sectors to target. Natural sources of PM can also be significant on a regional basis (Aguilera et al., 2021; Xue et al., 2021) but are not easily controlled and therefore not the subject of this chapter.

To calculate the benefits of stringent 50% global emission reductions from individual source sectors to solar PV electricity generation I integrate the GEOS-Chem global 3-D model of atmospheric composition, equipped with online radiative transfer calculations, with PVLIB-Python which is a solar PV performance model. I focus on source sectors to identify a systematic approach to prioritizing mitigation measures (Unger et al., 2020). The 50% source sector reductions I explore are large but plausible, with success previously reported in the US (Simon et al., 2015) and China (Zheng et al., 2018). Following previous studies (Sweerts et al., 2019; Li et al., 2020; Jerez et al., 2015; Feron et al., 2021), I use capacity factors (CFs) to describe solar PV electricity generation efficiency. I calculate CF values

as the ratio of the actual power output of a PV panel to the theoretical maximum power output.

I propose an experimental design (Methods) to distinguish PV efficiency losses attributable to atmospheric and deposited PM, which I term as PM dimming and soiling impacts, respectively. Conversely, abatements in PM dimming and soiling impacts resulting from emission reductions are call brightening and cleaning benefits, respectively. I run experiments for the three widely used panel settings: horizontal fixed (Flat), fixed with optimal tilt (Tilt), and one-axis tracking (OAT). Their characteristics are presented in Table 1.2 in section 1.3.3. I find that stringent reductions in residential emissions lead to substantial benefits to PV efficiency in East and South Asia.

3.3 Methods

I continue to use the integrated model developed, calibrated, and evaluated in the previous chapter to do model experiments in this chapter, so here I simply describe the experimental design additionally introduced for the research purpose of this chapter.

All the model experiments in the previous chapter use the standard emissions, so they are called control simulations (CTRL). PM dimming, soiling, and total impacts in the control simulations are accordingly written as $(CF3 - CF2)_{CTRL}$, $(CF2 - CF1)_{CTRL}$, and $(CF3 - CF1)_{CTRL}$, respectively. To identify the leading source sectors where reducing emissions greatly benefit PV power output, I perform mitigation simulations (0.5SECTOR) by halving emissions of all air pollutants sector by sector of the $\text{CEDS}_{GBD-MAPS}$ emission inventory. Corresponding PM dimming, soiling, and total impacts are thereby written as $(CF3 - CF2)_{0.5SECTOR}$, $(CF2 - CF1)_{0.5SECTOR}$, and $(CF3 - CF1)_{0.5SECTOR}$,

respectively. I then determine brightening, cleaning, and total benefits from halving source sector emissions by taking the difference in each of these quantities between the control and mitigation simulations, i.e. $(CF3 - CF2)_{CTRL} - (CF3 - CF2)_{0.5SECTOR}$, $(CF2 - CF1)_{CTRL} - (CF2 - CF1)_{0.5SECTOR}$, and $(CF3 - CF1)_{CTRL} - (CF3 - CF1)_{0.5SECTOR}$. Comparing these difference quantities across source sectors identifies which are the most effective ones to target to alleviate PM-induced PV efficiency losses (Figure 3.1).

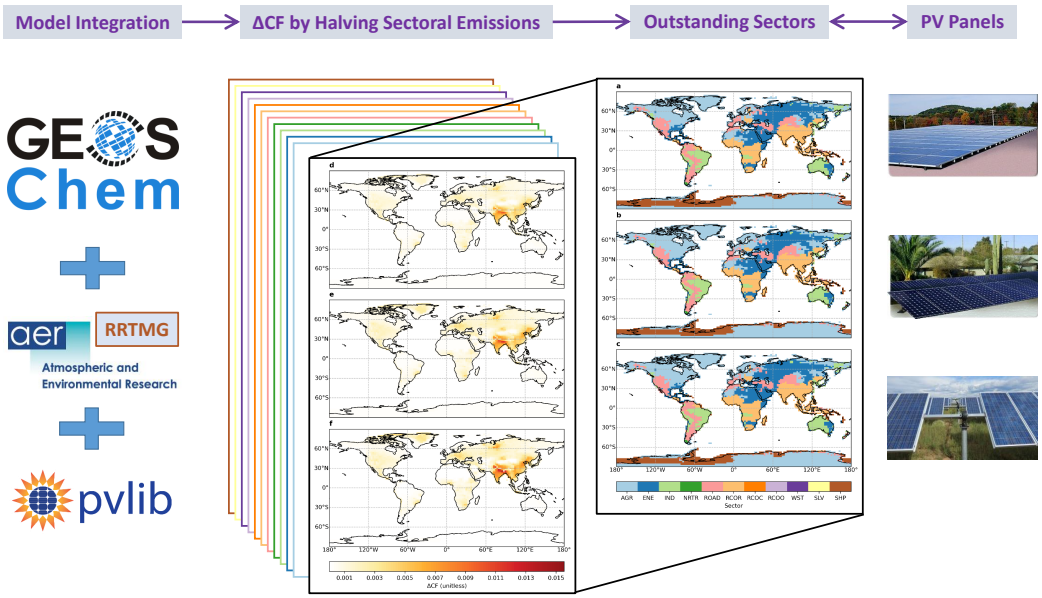


Figure 3.1: Illustration of strategy used in this chapter to identify leading source sectors where reducing emissions greatly benefit photovoltaic power output.

Although not the main focus of this chapter, sweeping panels either manually or by robots could, to some extent, recover PM-induced PV efficiency losses. But unlike reducing emissions, this practice simply affects PM soiling process. I thus evaluate the benefits of sweeping panels by directly comparing $CF1$s in the control simulations, i.e. $CF1_{CTRL+SWEEPING} - CF1_{CTRL}$. I stipulate that the accumulated dry mass of deposited PM becomes zeros at the beginning of each year, quarter, month, week, and day to correspond to $CF1_{CTRL+SWEEPING}$

of sweeping panels at yearly, quarterly, monthly, weekly, and daily frequencies, respectively.

As I do in the previous chapter, in this chapter, I analyse the decadal (and corresponding seasonal) mean (2008–2017) brightening, cleaning, and total benefits at each 2° latitude × 2.5° longitude grid for their spatial distributions, and further calculate the regional area-weighted mean values for regional characteristics. Absolute benefits are compared to relevant PM impacts to obtain percent benefits where applicable. I continue to use the latest IPCC climate reference regions (Iturbide et al., 2020) (Figure 2.2) for regional synthesis.

3.4 Results

3.4.1 Benefits of reducing emissions

Here I present the extent to which we can reduce PM impacts, as described above, by decreasing PM emissions. I quantify and determine the maximum benefits of halving emissions from all anthropogenic source sectors (Methods).

Figure 3.2 and Table 3.1 show that halving residential and agricultural emissions result in widespread decreases in PM dimming impacts. The proportion of areas occupied by the residential sector from which halving emissions provides the largest brightening benefits for flat, tilt, and OAT panels are 46%, 58%, and 50% and uniformly 93% over East and South Asia, respectively, and they are 43%, 37%, and 40% and uniformly 100% by the agricultural sector over East Asia and West and Central Europe, respectively. The brightening benefits for flat, tilt, and OAT panels of halving residential emissions are 8%, 9%, and 9% and identically 12% over East and South Asia, respectively, and they are equally 8%

and equally 13% of halving agricultural emissions over East Asia and West and Central Europe, respectively.

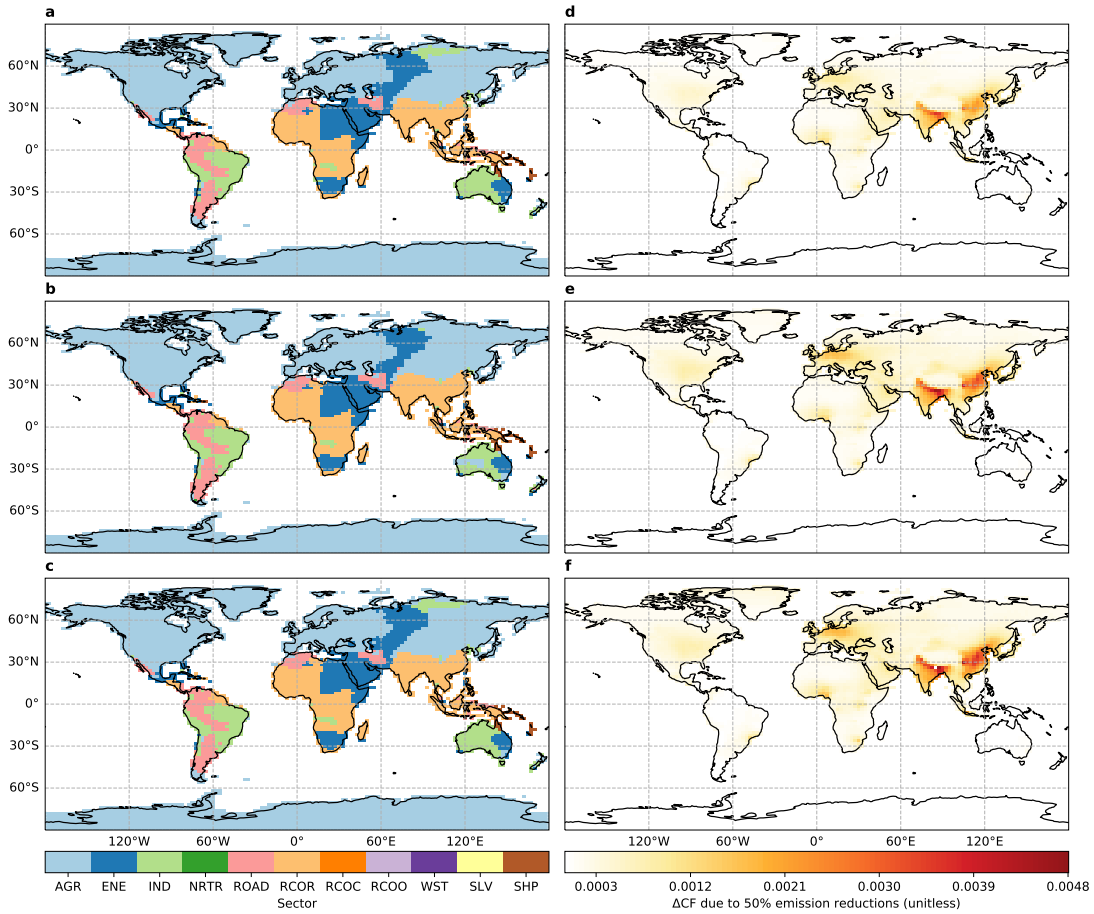


Figure 3.2: Geographical distributions of (a-c) source sectors from which halving emissions provide (d-f) maximum decadal mean (2008–2017) brightening benefits for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels. Full definitions for source sectors are non-combustion agriculture (AGR), energy generation (ENE), industrial processes (IND), on-road (ROAD) and non-road (NRTR) transportation, separate residential (RCOR), commercial (RCOC), and other (RCOO) sectors, waste (WST), solvent use (SLV), and international shipping (SHP).

Seasonal statistics (Table 3.1) show that over East Asia, halving residential, agricultural, industrial, and agricultural emissions result in the most widespread brightening benefits during DJF, MAM, JJA, and SON, respectively. The

proportion of areas occupied by these four sectors from which halving emissions provide the largest brightening benefits for the three panels during the four seasons range 99%, 79–80%, 68–72%, and 55–56%, respectively. The brightening benefits for the three panels from cuts to these four sectors during the four seasons range 14–15%, 10%, 10–11%, and 9–10%, respectively. Despite the seasonal nature (e.g. heating from November to March in the north of China) of emissions from the residential sector, the large and widespread brightening benefits by halving emissions from this sector during DJF dominate the annual results.

The seasonal results (Table 3.1) of halving emissions over South Asia and West and Central Europe are less complicated than those over East Asia. Except during JJA when there are significant brightening benefits from halving energy emissions, halving residential and agricultural emissions consistently dominate the brightening benefits over South Asia and West and Central Europe throughout the year. The largest brightening benefits for the three panels vary 15–16% for South Asia during DJF and 15% for West and Central Europe during SON.

Table 3.1: The proportion of occupied areas, and regional-area-weighted mean brightening benefits and PM dimming impacts and their ratios of sectors from which halving emissions provide the largest decadal and corresponding seasonal mean brightening benefits, in various regions of interest. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for definitions of regions that are presented here.

Period	Region	Sector	Panel setting	Area fractions (%)	Brightening benefits (unitless)	PM dimming impacts (unitless)	Brightening benefits (%)
Annual	E.Asia	RCOR	Flat	46	0.0010	0.0126	8
			Tilt	58	0.0015	0.0161	9
			OAT	50	0.0016	0.0180	9
		AGR	Flat	43	0.0010	0.0126	8
			Tilt	37	0.0013	0.0161	8
			OAT	40	0.0015	0.0180	8
	S.Asia	RCOR	Flat	93	0.0014	0.0117	12
			Tilt	93	0.0017	0.0134	12
			OAT	93	0.0018	0.0147	12
	West&Central-Europe	AGR	Flat	100	0.0006	0.0047	13
			Tilt	100	0.0011	0.0079	13
			OAT	100	0.0011	0.0082	13
DJF	E.Asia	RCOR	Flat	99	0.0017	0.0118	15

Table 3.1: Continued from previous page.

Period	Region	Sector	Panel setting	Area fractions (%)	Brightening benefits (unitless)	PM dimming impacts (unitless)	Brightening benefits (%)
			Tilt	99	0.0031	0.0216	14
			OAT	99	0.0028	0.0185	15
	S.Asia	RCOR	Flat	94	0.0018	0.0115	15
			Tilt	94	0.0026	0.0168	16
			OAT	94	0.0024	0.0152	16
	West&Central-Europe	AGR	Flat	100	0.0003	0.0023	12
			Tilt	100	0.0011	0.0085	13
			OAT	100	0.0007	0.0055	13
MAM	E.Asia	AGR	Flat	79	0.0016	0.0169	10
			Tilt	80	0.0017	0.0178	10
			OAT	80	0.0024	0.0228	10
	S.Asia	RCOR	Flat	82	0.0015	0.0146	10
			Tilt	82	0.0016	0.0147	11
			OAT	81	0.0019	0.0178	11
	West&Central-Europe	AGR	Flat	100	0.0010	0.0072	13
			Tilt	100	0.0013	0.0094	14
			OAT	100	0.0016	0.0115	14
JJA	E.Asia	IND	Flat	69	0.0012	0.0113	10
			Tilt	68	0.0010	0.0096	10
			OAT	72	0.0017	0.0163	11
		RCOR	Flat	14	0.0007	0.0113	6
			Tilt	14	0.0006	0.0096	6
			OAT	13	0.0010	0.0163	6
	S.Asia	RCOR	Flat	66	0.0008	0.0107	8
			Tilt	66	0.0007	0.0093	8
			OAT	65	0.0011	0.0144	8
		ENE	Flat	29	0.0007	0.0107	6
			Tilt	29	0.0006	0.0093	6
			OAT	29	0.0009	0.0144	6
	West&Central-Europe	AGR	Flat	67	0.0007	0.0058	12
			Tilt	67	0.0007	0.0055	12
			OAT	67	0.0011	0.0092	12
		ENE	Flat	33	0.0004	0.0058	7
			Tilt	33	0.0004	0.0055	7
			OAT	33	0.0007	0.0092	8
SON	E.Asia	AGR	Flat	56	0.0010	0.0103	9
			Tilt	55	0.0014	0.0153	9
			OAT	55	0.0014	0.0142	10
		IND	Flat	38	0.0009	0.0103	9
			Tilt	38	0.0012	0.0153	8
			OAT	39	0.0013	0.0142	9
	S.Asia	RCOR	Flat	92	0.0013	0.0099	13
			Tilt	93	0.0017	0.0127	14
			OAT	91	0.0016	0.0116	14
	West&Central-Europe	AGR	Flat	100	0.0005	0.0035	15
			Tilt	100	0.0012	0.0081	15
			OAT	100	0.0010	0.0065	15

Figure 3.3 and Table 3.2 show that halving residential emissions results in widespread decreases in PM soiling impacts. The proportion of areas occupied by the residential sector from which halving emissions provides the largest cleaning benefits for flat, tilt, and OAT panels are 62%, 63%, and 63%, 86%, 85%, and

88%, and uniformly 94% over East Asia, South Asia, and the Tibetan-Plateau, respectively. Halving residential emissions uniformly reduces PM soiling by 3–4% for the three panels over East and South Asia. The corresponding value is slightly higher at 4–5% over the Tibetan-Plateau.

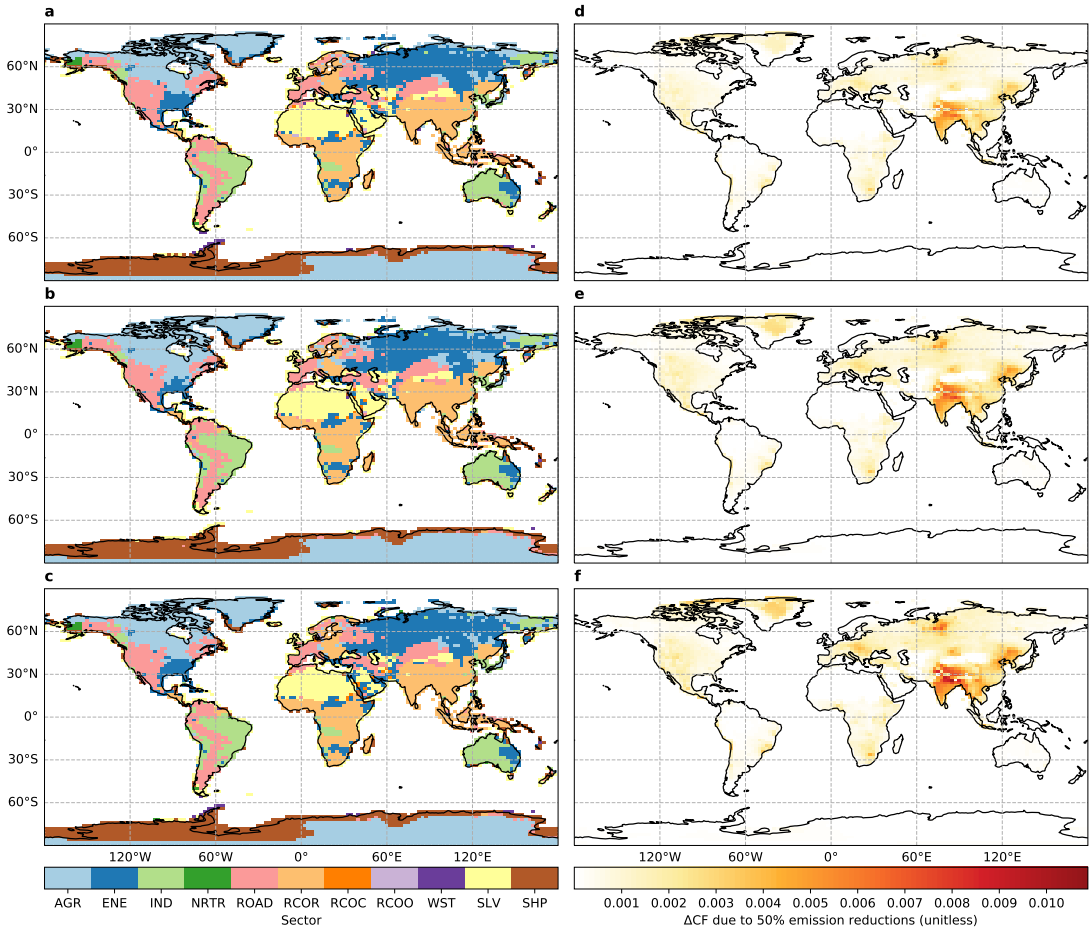


Figure 3.3: Geographical distributions of (a-c) source sectors from which halving emissions provide (d-f) maximum decadal mean (2008–2017) cleaning benefits for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels. Full definitions for source sectors are non-combustion agriculture (AGR), energy generation (ENE), industrial processes (IND), on-road (ROAD) and non-road (NRTR) transportation, separate residential (RCOR), commercial (RCOC), and other (RCOO) sectors, waste (WST), solvent use (SLV), and international shipping (SHP).

Seasonal statistics (Table 3.2) show that over East Asia, the proportion of areas occupied by the residential sector from which halving emissions provides the largest cleaning benefits for the three panels follows a descending order of DJF, MAM, JJA, and SON, which are 78–82%, 71–73%, 38–40%, and 11–13%, respectively. The cleaning benefits for the three panels from cuts to the residential sector during DJF and MAM (3–4%) are slightly higher than those during JJA and SON (2–3%). The industrial sector is another place from which halving emissions provides significant cleaning benefits of 2–4% for the three panels during JJA and SON.

Over South Asia and the Tibetan-Plateau, I find that the residential sector is always where halving emissions provides the largest cleaning benefits throughout the year (Table 3.2). The largest cleaning benefits of 5–6% for the three panels are during DJF for South Asia and during DJF and MAM for the Tibetan-Plateau.

Table 3.2: The proportion of occupied areas, and regional-area-weighted mean cleaning benefits and PM soiling impacts and their ratios of sectors from which halving emissions provide the largest decadal and corresponding seasonal mean cleaning benefits, in various regions of interest. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for definitions of regions that are presented here.

Period	Region	Sector	Panel setting	Area fractions (%)	Cleaning benefits (unitless)	PM soiling impacts (unitless)	Cleaning benefits (%)
Annual	Tibetan-Plateau	RCOR	Flat	94	0.0032	0.0728	4
			Tilt	94	0.0038	0.0806	5
			OAT	94	0.0044	0.0832	5
	S.Asia	RCOR	Flat	86	0.0026	0.0817	3
			Tilt	85	0.0029	0.0848	3
			OAT	88	0.0039	0.1060	4
	E.Asia	RCOR	Flat	62	0.0016	0.0555	3
			Tilt	63	0.0019	0.0610	3
			OAT	63	0.0024	0.0741	3
		IND	Flat	21	0.0010	0.0555	2
			Tilt	20	0.0012	0.0610	2
			OAT	22	0.0016	0.0741	2
DJF	Tibetan-Plateau	RCOR	Flat	94	0.0027	0.0513	5
			Tilt	94	0.0045	0.0797	6
			OAT	94	0.0043	0.0676	6
	S.Asia	RCOR	Flat	85	0.0035	0.0731	5
			Tilt	83	0.0044	0.0907	5

Table 3.2: Continued from previous page.

Period	Region	Sector	Panel setting	Area fractions (%)	Cleaning benefits (unitless)	PM soiling impacts (unitless)	Cleaning benefits (%)
			OAT	91	0.0056	0.0984	6
	E.Asia	RCOR	Flat	82	0.0021	0.0548	4
			Tilt	78	0.0029	0.0792	4
			OAT	80	0.0031	0.0781	4
MAM	Tibetan-Plateau	RCOR	Flat	94	0.0047	0.0940	5
			Tilt	94	0.0049	0.0926	5
			OAT	94	0.0062	0.1011	6
	S.Asia	RCOR	Flat	91	0.0044	0.1121	4
			Tilt	90	0.0044	0.1087	4
			OAT	91	0.0060	0.1411	4
	E.Asia	RCOR	Flat	71	0.0025	0.0813	3
			Tilt	72	0.0027	0.0797	3
			OAT	73	0.0038	0.1061	4
		AGR	Flat	13	0.0006	0.0813	1
			Tilt	13	0.0006	0.0797	1
			OAT	14	0.0010	0.1061	1
JJA	Tibetan-Plateau	RCOR	Flat	91	0.0031	0.0873	4
			Tilt	91	0.0029	0.0756	4
			OAT	91	0.0041	0.0990	4
	S.Asia	RCOR	Flat	81	0.0013	0.0819	2
			Tilt	81	0.0012	0.0723	2
			OAT	81	0.0017	0.1074	2
	E.Asia	IND	Flat	44	0.0010	0.0510	2
			Tilt	43	0.0009	0.0421	2
			OAT	45	0.0015	0.0655	2
		RCOR	Flat	40	0.0009	0.0510	2
			Tilt	40	0.0008	0.0421	2
			OAT	38	0.0014	0.0655	2
SON	Tibetan-Plateau	RCOR	Flat	91	0.0021	0.0584	4
			Tilt	91	0.0030	0.0746	4
			OAT	91	0.0031	0.0652	5
	S.Asia	RCOR	Flat	69	0.0013	0.0596	2
			Tilt	70	0.0015	0.0674	2
			OAT	73	0.0022	0.0773	3
		SLV	Flat	12	-0.0000	0.0596	0
			Tilt	12	-0.0000	0.0674	0
			OAT	11	-0.0000	0.0773	0
	E.Asia	IND	Flat	68	0.0012	0.0347	3
			Tilt	68	0.0015	0.0431	3
			OAT	71	0.0017	0.0466	4
		RCOR	Flat	11	0.0008	0.0347	2
			Tilt	13	0.0011	0.0431	3
			OAT	11	0.0013	0.0466	3

The combined benefits from brightening and cleaning (Figure 3.4 and Table 3.3) mainly follow the pattern of cleaning benefits, as expected. On the decadal timescale, I report that halving residential emissions uniformly results in total

benefits of 4–6% for the three panels over East Asia, South Asia, and the Tibetan-Plateau. Halving industrial emissions results in total benefits of 3% for the three panels over East Asia.

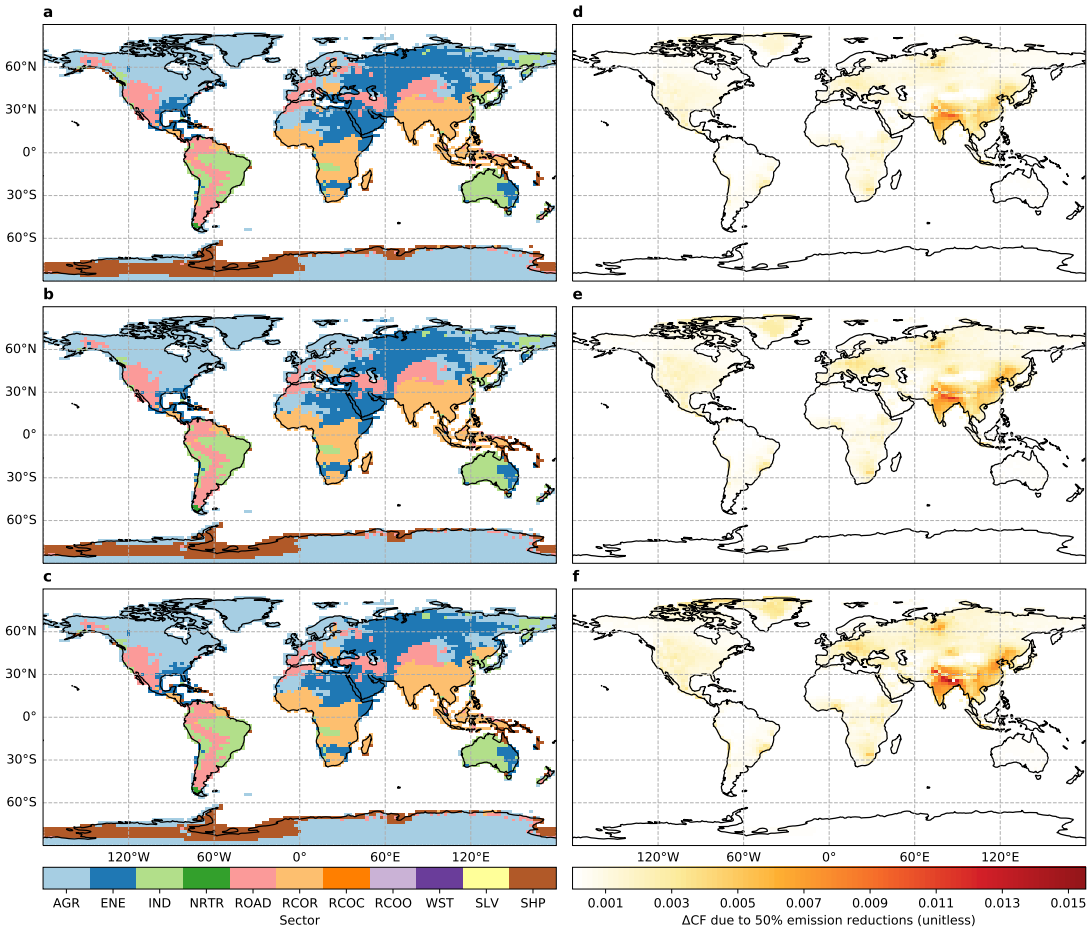


Figure 3.4: Geographical distributions of (a-c) source sectors from which halving emissions provide (d-f) maximum decadal mean (2008–2017) total benefits for (a and d) flat, (b and e) tilt, and (c and f) one-axis tracking panels. Full definitions for source sectors are non-combustion agriculture (AGR), energy generation (ENE), industrial processes (IND), on-road (ROAD) and non-road (NRTR) transportation, separate residential (RCOR), commercial (RCOC), and other (RCOO) sectors, waste (WST), solvent use (SLV), and international shipping (SHP).

Table 3.3: The proportion of occupied areas, and regional-area-weighted mean total benefits and PM total impacts and their ratios of sectors from which halving emissions provide the largest decadal and corresponding seasonal mean total benefits, in various regions of interest. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for definitions of regions that are presented here.

Period	Region	Sector	Panel setting	Area fractions (%)	Total benefits (unitless)	PM total impacts (unitless)	Total benefits (%)
Annual	S.Asia	RCOR	Flat	93	0.0040	0.0934	4
			Tilt	93	0.0045	0.0981	5
			OAT	93	0.0056	0.1208	5
	Tibetan-Plateau	RCOR	Flat	94	0.0036	0.0766	5
			Tilt	94	0.0044	0.0854	5
			OAT	94	0.0049	0.0876	6
	E.Asia	RCOR	Flat	61	0.0026	0.0680	4
			Tilt	67	0.0034	0.0771	4
			OAT	62	0.0040	0.0921	4
		IND	Flat	21	0.0020	0.0680	3
			Tilt	15	0.0024	0.0771	3
			OAT	19	0.0031	0.0921	3
DJF	S.Asia	RCOR	Flat	94	0.0052	0.0846	6
			Tilt	94	0.0070	0.1074	6
			OAT	94	0.0080	0.1135	7
	Tibetan-Plateau	RCOR	Flat	96	0.0031	0.0539	6
			Tilt	96	0.0053	0.0845	6
			OAT	96	0.0048	0.0709	7
	E.Asia	RCOR	Flat	92	0.0038	0.0666	6
			Tilt	93	0.0061	0.1008	6
			OAT	93	0.0059	0.0966	6
MAM	S.Asia	RCOR	Flat	93	0.0059	0.1267	5
			Tilt	93	0.0060	0.1233	5
			OAT	93	0.0079	0.1588	5
	Tibetan-Plateau	RCOR	Flat	94	0.0052	0.0991	5
			Tilt	94	0.0055	0.0979	6
			OAT	94	0.0067	0.1064	6
	E.Asia	RCOR	Flat	62	0.0036	0.0982	4
			Tilt	64	0.0038	0.0975	4
			OAT	60	0.0053	0.1290	4
		AGR	Flat	26	0.0023	0.0982	2
			Tilt	27	0.0023	0.0975	2
			OAT	31	0.0034	0.1290	3
JJA	S.Asia	RCOR	Flat	87	0.0021	0.0926	2
			Tilt	87	0.0020	0.0816	2
			OAT	87	0.0028	0.1218	2
	Tibetan-Plateau	RCOR	Flat	91	0.0036	0.0919	4
			Tilt	91	0.0033	0.0796	4
			OAT	91	0.0047	0.1044	4
	E.Asia	IND	Flat	63	0.0022	0.0623	3
			Tilt	61	0.0019	0.0517	4
			OAT	64	0.0032	0.0819	4
		RCOR	Flat	24	0.0016	0.0623	3
			Tilt	25	0.0014	0.0517	3
			OAT	22	0.0025	0.0819	3
SON	S.Asia	RCOR	Flat	88	0.0026	0.0695	4
			Tilt	89	0.0032	0.0801	4
			OAT	88	0.0038	0.0889	4
	Tibetan-Plateau	RCOR	Flat	91	0.0025	0.0617	4

Table 3.3: Continued from previous page.

Period	Region	Sector	Panel setting	Area fractions (%)	Total benefits (unitless)	PM total impacts (unitless)	Total benefits (%)
			Tilt	91	0.0036	0.0793	5
			OAT	91	0.0035	0.0684	5
	E.Asia	IND	Flat	65	0.0020	0.0450	5
			Tilt	65	0.0027	0.0584	5
			OAT	65	0.0030	0.0609	5
		AGR	Flat	25	0.0018	0.0450	4
			Tilt	25	0.0024	0.0584	4
			OAT	25	0.0026	0.0609	4

3.4.2 A clear priority of sectors to target in South Asia

Here I explore the second and third ranking of benefits resulting from halving sector emissions to identify a clear priority of sectors to target across South Asia, as described below.

For brightening benefits (Table 3.4), I find that the industrial and energy sectors are the next two important places to target after the residential one. The proportion of areas occupied by the industrial and energy sectors where halving emissions provide the second and third ranking of brightening benefits for the three panels are 64–65% and 62–63%, respectively. Halving industrial and energy emissions lead to similar brightening benefits of 7–8% across the three panels. Moreover, the brightening benefits of halving industrial and energy emissions are both approximately aseasonal, ranging 6–9% and 6–8% for the three panels, respectively.

Table 3.4: The proportion of occupied areas, and regional-area-weighted mean brightening benefits and PM dimming impacts and their ratios of top three sectors from which halving emissions provide the largest decadal and corresponding seasonal mean brightening benefits, in S.Asia. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for the definition of S.Asia that is presented here.

Period	Rank	Sector	Panel setting	Area fractions (%)	Brightening benefits (unitless)	PM dimming impacts (unitless)	Brightening benefits (%)
Annual	1	RCOR	Flat	93	0.0014	0.0117	12
			Tilt	93	0.0017	0.0134	12
			OAT	93	0.0018	0.0147	12
	2	IND	Flat	64	0.0009	0.0117	8
			Tilt	65	0.0010	0.0134	8
			OAT	65	0.0012	0.0147	8
		ENE	Flat	28	0.0008	0.0117	7
			Tilt	29	0.0009	0.0134	7
			OAT	28	0.0011	0.0147	7
	3	ENE	Flat	62	0.0008	0.0117	7
			Tilt	63	0.0009	0.0134	7
			OAT	63	0.0011	0.0147	7
		IND	Flat	23	0.0009	0.0117	8
			Tilt	22	0.0010	0.0134	8
			OAT	22	0.0012	0.0147	8
DJF	1	RCOR	Flat	94	0.0018	0.0115	15
			Tilt	94	0.0026	0.0168	16
			OAT	94	0.0024	0.0152	16
	2	IND	Flat	81	0.0009	0.0115	8
			Tilt	81	0.0013	0.0168	8
			OAT	81	0.0013	0.0152	8
	3	ENE	Flat	71	0.0007	0.0115	6
			Tilt	70	0.0010	0.0168	6
			OAT	70	0.0010	0.0152	7
		AGR	Flat	15	0.0006	0.0115	5
			Tilt	16	0.0008	0.0168	5
			OAT	16	0.0008	0.0152	5
MAM	1	RCOR	Flat	82	0.0015	0.0146	10
			Tilt	82	0.0016	0.0147	11
			OAT	81	0.0019	0.0178	11
	2	IND	Flat	56	0.0011	0.0146	8
			Tilt	57	0.0011	0.0147	8
			OAT	56	0.0014	0.0178	8
		ENE	Flat	27	0.0011	0.0146	8
			Tilt	26	0.0011	0.0147	8
			OAT	26	0.0014	0.0178	8
	3	ENE	Flat	56	0.0011	0.0146	8
			Tilt	57	0.0011	0.0147	8
			OAT	56	0.0014	0.0178	8
		IND	Flat	28	0.0011	0.0146	8
			Tilt	27	0.0011	0.0147	8
			OAT	28	0.0014	0.0178	8
JJA	1	RCOR	Flat	66	0.0008	0.0107	8
			Tilt	66	0.0007	0.0093	8
			OAT	65	0.0011	0.0144	8
		ENE	Flat	29	0.0007	0.0107	6
			Tilt	29	0.0006	0.0093	6
			OAT	29	0.0009	0.0144	6
	3	IND	Flat	39	0.0006	0.0107	6

Table 3.4: Continued from previous page.

Period	Rank	Sector	Panel setting	Area fractions (%)	Brightening benefits (unitless)	PM dimming impacts (unitless)	Brightening benefits (%)
			Tilt	41	0.0006	0.0093	6
			OAT	41	0.0009	0.0144	6
		ENE	Flat	35	0.0007	0.0107	6
			Tilt	35	0.0006	0.0093	6
			OAT	35	0.0009	0.0144	6
SON	1	RCOR	Flat	92	0.0013	0.0099	13
			Tilt	93	0.0017	0.0127	14
			OAT	91	0.0016	0.0116	14
	2	IND	Flat	69	0.0009	0.0099	9
			Tilt	70	0.0011	0.0127	9
			OAT	68	0.0011	0.0116	9
		ENE	Flat	25	0.0008	0.0099	8
			Tilt	25	0.0010	0.0127	8
			OAT	24	0.0010	0.0116	8
	3	ENE	Flat	63	0.0008	0.0099	8
			Tilt	65	0.0010	0.0127	8
			OAT	65	0.0010	0.0116	8
		IND	Flat	20	0.0009	0.0099	9
			Tilt	19	0.0011	0.0127	9
			OAT	20	0.0011	0.0116	9

Similar to brightening benefits, I find that the industrial and energy sectors are the next two important places to target after the residential one for cleaning benefits (Table 3.5). The proportion of areas occupied by the industrial and energy sectors where halving emissions provide the second and third ranking of cleaning benefits for the three panels are 65–68% and 58–61%, respectively. Halving industrial emissions leads to cleaning benefits of 1–2% for the three panels, comparable to the cleaning benefits of 1% by halving energy emissions. The cleaning benefits for the three panels of halving industrial and energy emissions both vary 1–3% throughout the year, suggesting limited seasonality.

Table 3.5: The proportion of occupied areas, and regional-area-weighted mean cleaning benefits and PM soiling impacts and their ratios of top three sectors from which halving emissions provide the largest decadal and corresponding seasonal mean cleaning benefits, in S.Asia. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for the definition of S.Asia that is presented here.

Period	Rank	Sector	Panel setting	Area fractions (%)	Cleaning benefits (unitless)	PM soiling impacts (unitless)	Cleaning benefits (%)
Annual	1	RCOR	Flat	86	0.0026	0.0817	3
			Tilt	85	0.0029	0.0848	3
			OAT	88	0.0039	0.1060	4
	2	IND	Flat	65	0.0012	0.0817	1
			Tilt	66	0.0013	0.0848	2
			OAT	68	0.0018	0.1060	2
		ENE	Flat	17	0.0010	0.0817	1
			Tilt	17	0.0011	0.0848	1
			OAT	17	0.0015	0.1060	1
	3	ENE	Flat	58	0.0010	0.0817	1
			Tilt	60	0.0011	0.0848	1
			OAT	61	0.0015	0.1060	1
		IND	Flat	19	0.0012	0.0817	1
			Tilt	19	0.0013	0.0848	2
			OAT	19	0.0018	0.1060	2
DJF	1	RCOR	Flat	85	0.0035	0.0731	5
			Tilt	83	0.0044	0.0907	5
			OAT	91	0.0056	0.0984	6
	2	IND	Flat	63	0.0016	0.0731	2
			Tilt	63	0.0020	0.0907	2
			OAT	65	0.0026	0.0984	3
		ENE	Flat	19	0.0014	0.0731	2
			Tilt	18	0.0018	0.0907	2
			OAT	24	0.0022	0.0984	2
	3	ENE	Flat	57	0.0014	0.0731	2
			Tilt	57	0.0018	0.0907	2
			OAT	57	0.0022	0.0984	2
		IND	Flat	24	0.0016	0.0731	2
			Tilt	24	0.0020	0.0907	2
			OAT	26	0.0026	0.0984	3
MAM	1	RCOR	Flat	91	0.0044	0.1121	4
			Tilt	90	0.0044	0.1087	4
			OAT	91	0.0060	0.1411	4
	2	IND	Flat	64	0.0018	0.1121	2
			Tilt	66	0.0019	0.1087	2
			OAT	66	0.0026	0.1411	2
		ENE	Flat	16	0.0015	0.1121	1
			Tilt	17	0.0016	0.1087	1
			OAT	20	0.0022	0.1411	2
	3	ENE	Flat	64	0.0015	0.1121	1
			Tilt	63	0.0016	0.1087	1
			OAT	62	0.0022	0.1411	2
		IND	Flat	20	0.0018	0.1121	2
			Tilt	18	0.0019	0.1087	2
			OAT	20	0.0026	0.1411	2
JJA	1	RCOR	Flat	81	0.0013	0.0819	2
			Tilt	81	0.0012	0.0723	2
			OAT	81	0.0017	0.1074	2
SON	1	RCOR	Flat	69	0.0013	0.0596	2

Table 3.5: Continued from previous page.

Period	Rank	Sector	Panel setting	Area fractions (%)	Cleaning benefits (unitless)	PM soiling impacts (unitless)	Cleaning benefits (%)
			Tilt	70	0.0015	0.0674	2
			OAT	73	0.0022	0.0773	3
		SLV	Flat	12	-0.0000	0.0596	0
			Tilt	12	-0.0000	0.0674	0
			OAT	11	-0.0000	0.0773	0

For total benefits (Table 3.6), it is within expectation that the industrial and energy sectors are the next two important places to target after the residential one. The proportion of areas occupied by the industrial and energy sectors where halving emissions provide the second and third ranking of total benefits for the three panels are 69–70% and 66–67%, respectively. Halving industrial and energy emissions uniformly provide total benefits of 1–3% for the three panels on the decadal and corresponding seasonal scales.

Table 3.6: The proportion of occupied areas, and regional-area-weighted mean total benefits and PM total impacts and their ratios of top three sectors from which halving emissions provide the largest decadal and corresponding seasonal mean total benefits, in S.Asia. For brevity, I sort sectors in descending order by their mean proportions of occupied areas of the three panels, and only keep to maximum two sectors whose cumulative mean proportions $\geq 75\%$. Please refer to Figure 2.2 for the definition of S.Asia that is presented here.

Period	Rank	Sector	Panel setting	Area fractions (%)	Total benefits (unitless)	PM total impacts (unitless)	Total benefits (%)
Annual	1	RCOR	Flat	93	0.0040	0.0934	4
			Tilt	93	0.0045	0.0981	5
			OAT	93	0.0056	0.1208	5
	2	IND	Flat	69	0.0021	0.0934	2
			Tilt	70	0.0024	0.0981	2
			OAT	69	0.0030	0.1208	2
		ENE	Flat	23	0.0018	0.0934	2
			Tilt	23	0.0020	0.0981	2
			OAT	23	0.0026	0.1208	2
	3	ENE	Flat	66	0.0018	0.0934	2
			Tilt	66	0.0020	0.0981	2
			OAT	67	0.0026	0.1208	2
		IND	Flat	28	0.0021	0.0934	2
			Tilt	27	0.0024	0.0981	2
			OAT	28	0.0030	0.1208	2
DJF	1	RCOR	Flat	94	0.0052	0.0846	6

Table 3.6: Continued from previous page.

Period	Rank	Sector	Panel setting	Area fractions (%)	Total benefits (unitless)	PM total impacts (unitless)	Total benefits (%)
			Tilt	94	0.0070	0.1074	6
			OAT	94	0.0080	0.1135	7
	2	IND	Flat	70	0.0025	0.0846	3
			Tilt	71	0.0033	0.1074	3
			OAT	71	0.0038	0.1135	3
		ENE	Flat	26	0.0022	0.0846	3
			Tilt	24	0.0028	0.1074	3
			OAT	24	0.0033	0.1135	3
	3	ENE	Flat	61	0.0022	0.0846	3
			Tilt	62	0.0028	0.1074	3
			OAT	62	0.0033	0.1135	3
		IND	Flat	21	0.0025	0.0846	3
			Tilt	20	0.0033	0.1074	3
			OAT	20	0.0038	0.1135	3
MAM	1	RCOR	Flat	93	0.0059	0.1267	5
			Tilt	93	0.0060	0.1233	5
			OAT	93	0.0079	0.1588	5
	2	IND	Flat	66	0.0030	0.1267	2
			Tilt	67	0.0030	0.1233	2
			OAT	66	0.0040	0.1588	3
		ENE	Flat	25	0.0027	0.1267	2
			Tilt	26	0.0027	0.1233	2
			OAT	26	0.0036	0.1588	2
	3	ENE	Flat	66	0.0027	0.1267	2
			Tilt	67	0.0027	0.1233	2
			OAT	66	0.0036	0.1588	2
		IND	Flat	29	0.0030	0.1267	2
			Tilt	27	0.0030	0.1233	2
			OAT	29	0.0040	0.1588	3
JJA	1	RCOR	Flat	87	0.0021	0.0926	2
			Tilt	87	0.0020	0.0816	2
			OAT	87	0.0028	0.1218	2
	2	IND	Flat	65	0.0012	0.0926	1
			Tilt	65	0.0011	0.0816	1
			OAT	64	0.0016	0.1218	1
		ENE	Flat	18	0.0010	0.0926	1
			Tilt	18	0.0009	0.0816	1
			OAT	19	0.0013	0.1218	1
	3	ENE	Flat	60	0.0010	0.0926	1
			Tilt	59	0.0009	0.0816	1
			OAT	58	0.0013	0.1218	1
		IND	Flat	30	0.0012	0.0926	1
			Tilt	31	0.0011	0.0816	1
			OAT	32	0.0016	0.1218	1
SON	1	RCOR	Flat	88	0.0026	0.0695	4
			Tilt	89	0.0032	0.0801	4
			OAT	88	0.0038	0.0889	4
	2	IND	Flat	65	0.0017	0.0695	2
			Tilt	66	0.0020	0.0801	3
			OAT	66	0.0025	0.0889	3
		ENE	Flat	24	0.0015	0.0695	2
			Tilt	23	0.0018	0.0801	2
			OAT	23	0.0022	0.0889	2
	3	ENE	Flat	60	0.0015	0.0695	2
			Tilt	62	0.0018	0.0801	2
			OAT	61	0.0022	0.0889	2
		IND	Flat	30	0.0017	0.0695	2
			Tilt	29	0.0020	0.0801	3
			OAT	29	0.0025	0.0889	3

3.4.3 Benefits of sweeping panels

Based on Figures 3.2d–f, 3.3d–f, and 3.4d–f, it is notable that halving anthropogenic PM emissions does not benefit desert and coastal regions where there are large PM soiling impacts on PV energy generation. These regions typically have a high abundance of natural PM including dust and sea salt, and a better strategy is to do routine sweeping of panels to overcome the majority of PM soiling impacts. In reverse, Figure 2.7d–f may be viewed as the maximum potential PV efficiency enhancements that can be achieved by routine sweeping of panels. Figure 3.5 exactly proves this point of view. The higher the frequency of routine sweeping of panels, the smaller differences of spatial patterns and magnitudes between cleaning benefits and PM soiling impacts.

Figure 3.6 further shows the regional synthesis. East and West Central Asia are similar, with yearly sweeping providing more than 40% cleaning benefits, which are further improved to around 80% by quarterly sweeping. The Arabian-Peninsula resembles the Mediterranean, with yearly sweeping providing around 30% cleaning benefits that are further improved to around 70% by quarterly sweeping. Quarterly sweeping provides around 60% for both the Sahara and Southern South America, with the former relying more on quarterly sweeping. The Caribbean and New-Zealand have pretty similar improved cleaning benefits with increased frequencies of panel sweeping practices. These numbers highlight that even a little sweeping effort (annual and quarterly sweeping) will remove a substantial proportion of PM soiling impacts in desert and coastal regions. A bigger energy return will result from regularly sweeping at higher frequencies, but will incur higher costs and higher risk damaging PV panels. Clearly, further regional cost-benefit analyses are needed to balance the value of increased energy production versus costs and risks associated with sweeping PV panels.

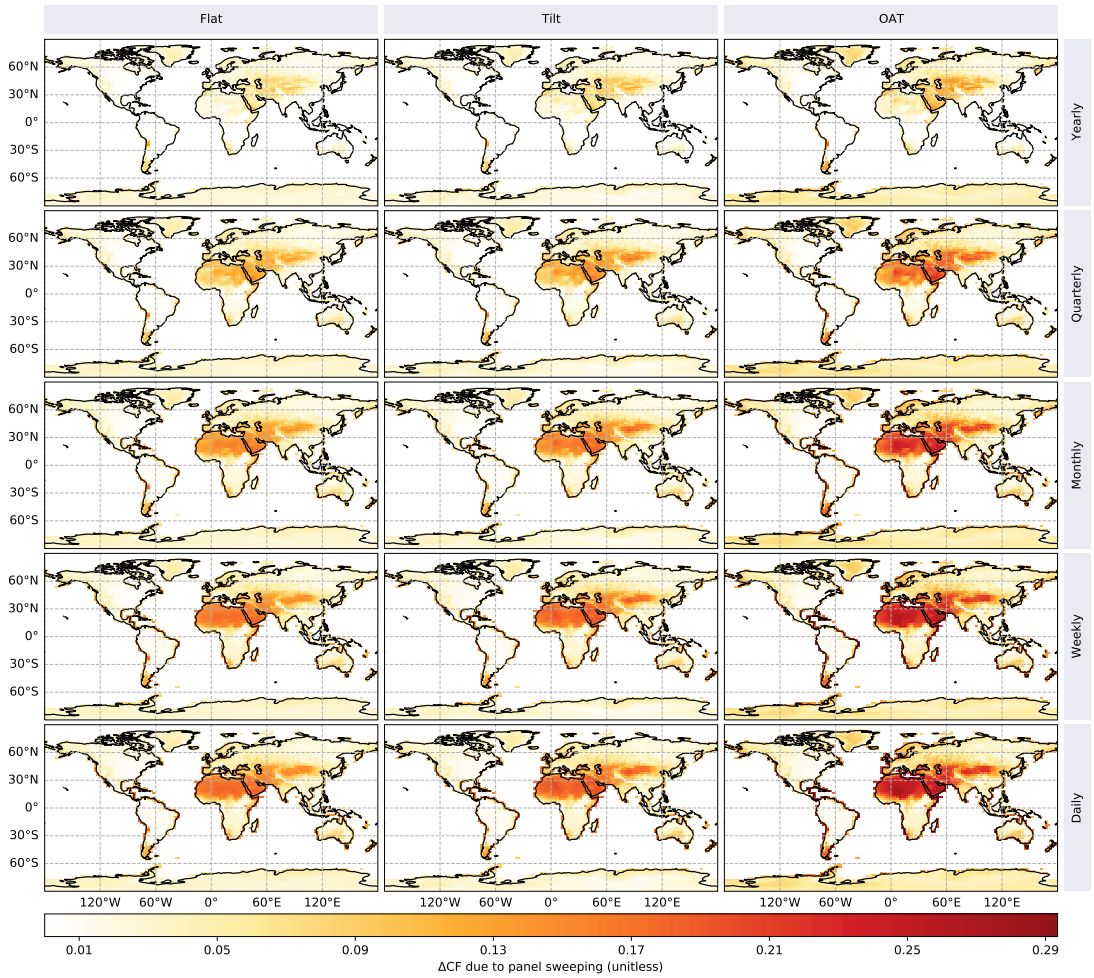


Figure 3.5: Geographical distributions of decadal mean (2008–2017) cleaning benefits resulting from various frequencies of panel sweeping for flat, tilt, and one-axis tracking panels.

3.5 Discussion and concluding remarks

To the best of my knowledge, this chapter provides the first quantitative assessment of the benefits to global and regional PV efficiency from halving air pollutant emissions in the anthropogenic source sectors. I present the results on decadal and corresponding seasonal scales because I find consistent benefits to PV efficiency throughout years particularly those from stringent residential emission

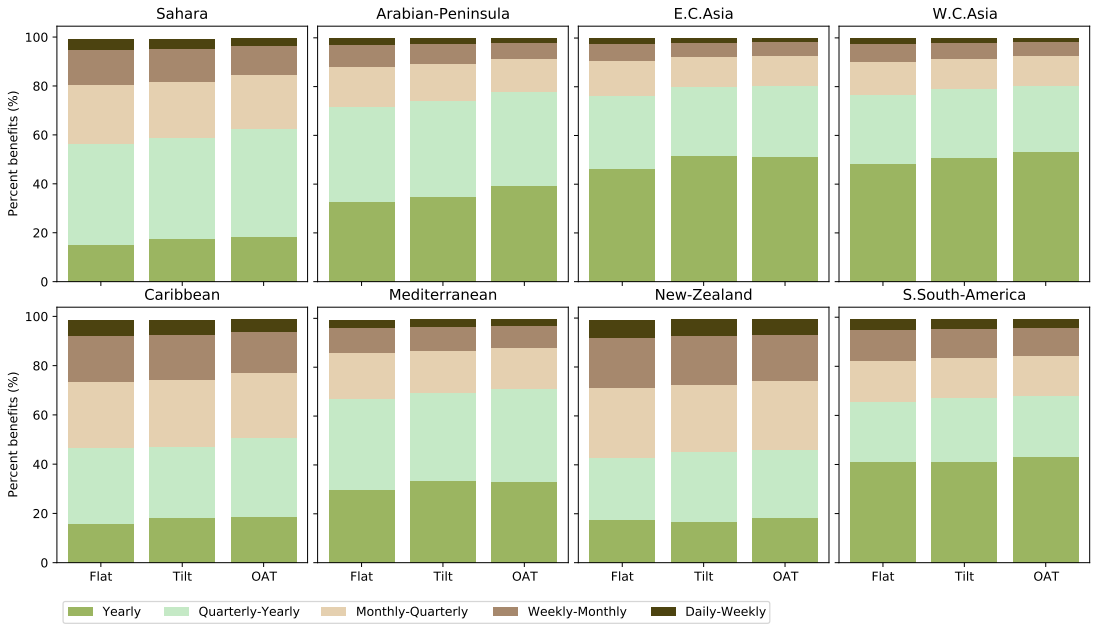


Figure 3.6: Regional-area-weighted averages of decadal mean (2008–2017) percent cleaning benefits (relative to PM soiling impacts) resulting from various frequencies of panel sweeping, in several regions of interest. Please refer to Figure 2.2 for definitions of regions that are presented here.

reductions over Asia with respect to the proportion of occupied areas (Figure 3.7). The uncontrolled and inefficient combustion of solid fuels in residential devices is likely the prime culprit. This is supported by previous studies that show that completely removing residential emissions can achieve considerable air quality benefits (McDuffie et al., 2021), particularly in East (Liu, Mauzerall, Chen, Zhang, Song, Peng, Klimont, Qiu, Zhang, Hu et al., 2016; Zhao et al., 2018) and South Asia (Chowdhury et al., 2019). My work highlights that the more realistic stringent reductions of residential emissions also lead to noticeable improvements in surface air quality with respect to PM with an aerodynamic diameter $\leq$ 2.5 μm (Figure 3.8), which will benefit human health. The resulting improvements of PV efficiency will subsequently reduce the dependency on conventional energy generation including the inefficient combustion of solid fuels over East and South Asia, which will further improve air quality, leading to a virtuous cycle.

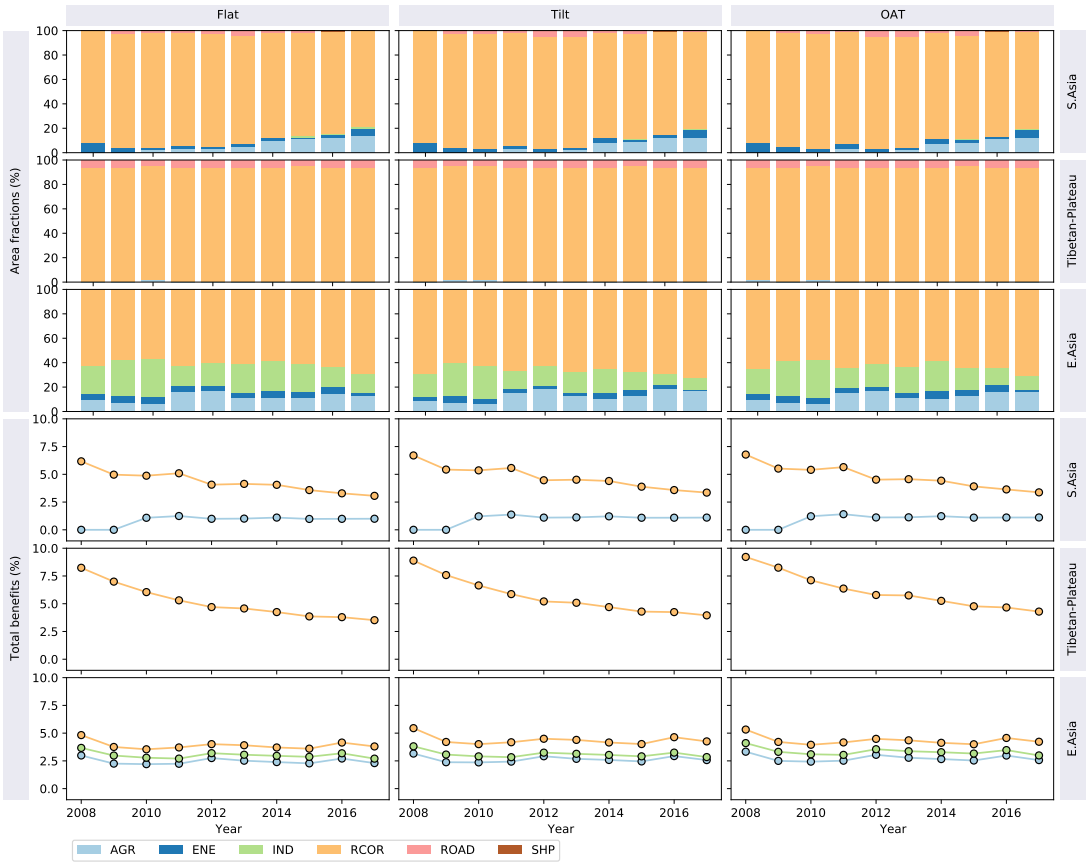


Figure 3.7: Consistent benefits to PV efficiency throughout years particularly those from stringent residential emission reductions over Asia with respect to the proportion of occupied areas. Please refer to Figure 2.2 for definitions of regions that are presented here.

The solar energy industry over East and South Asia stands to reap considerable rewards from halving residential emissions. To illustrate this point, I collect provincial and state-wise installed PV capacities as of 2019 from Chinese (National Energy Administration, 2020) and Indian (Ministry of New and Renewable Energy, 2020) national energy-related administrations, respectively, and combine them with the decadal mean CF improvements (ΔCFs) due to halving residential emissions to determine the energy and economic benefits. The installed capacities have a distinction between distributed and utility-scale PV installations, so I follow Sweerts et al. (2019) to assume Tilt ΔCFs for distributed and OAT ΔCFs

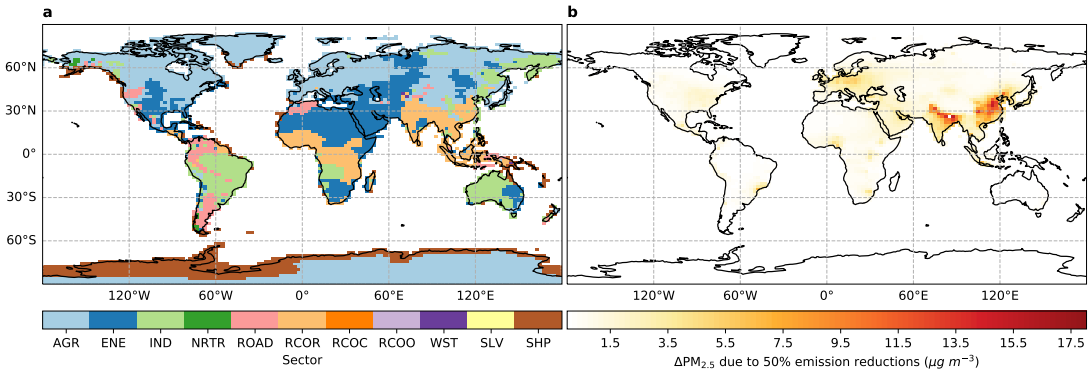


Figure 3.8: Geographical distributions of (a) sectors from which halving emissions provide (b) maximum decadal mean (2008–2017) co-benefits to surface air quality with respect to PM with an aerodynamic diameter $\leq$ 2.5 μm. Full definitions for source sectors are non-combustion agriculture (AGR), energy generation (ENE), industrial processes (IND), non-road (NRTR) and on-road (ROAD) transportation, separate residential (RCOR), commercial (RCOC), and other (RCOO) sectors, waste (WST), solvent use (SLV), international shipping (SHP), and agricultural waster burning (AWB).

for utility-scale PV installations. For each province/state, I extract ΔCFs in grid cells within that province/state. For grid cells overlapping more than one province/state, I split them along provincial/state boundaries with a geographical information system program. Further using areas of these (split) grid cells, I derive the provincial/state-wise area-weighted mean ΔCFs. Potential increases in PV electricity generation are then computed on the provincial/state level by multiplying ΔCFs for Tilt and OAT with the installed capacities of distributed and utility-scale PV, respectively. Economic benefits are further derived by multiplying these electricity bonus with the national mean electricity price for households in China ($\approx$ US\$0.085 kWh^{-1}) and India ($\approx$ US\$0.079 kWh^{-1}) in 2020, respectively (GlobalPetrolPrices.com, 2021). Tables 3.7 and 3.8 show that the energy benefits from halving residential emissions are 7,687 and 1,823 GWh yr^{-1} over China and India, respectively. The corresponding economic benefits

are US$653 million yr^{-1} and US$144 million yr^{-1}, respectively. Compared to the 2020 electricity generation of 260500 and 60400 GWh yr^{-1} from solar PV technology in China (National Energy Administration, 2021) and India (Central Electricity Authority, 2021), respectively, these energy and economic benefits all represent an approximately additional 3% improvement. Generally, regions where there are larger established PV installations will benefit more from stringent residential emissions. For example, four of the top five Chinese provinces with the largest PV installations benefit the most from halving residential emissions. Even regions with moderate PV installations benefit from large CF improvements due to halving residential emissions, e.g. Henan province in China and Madhya Pradesh state in India.

My experimental design is subject to several important limitations and sources of uncertainty. First, by simultaneously reducing emissions across all geographical regions, the role of long-range or regional transport (Meng et al., 2019; Lin et al., 2019) cannot be explicitly investigated. As the implementation of mitigation policies is typically constrained to political borders, specific policies may need to consider the long-range and regional influence on local PM pollution levels and PV power output. Second, while I focus on the large but plausible half elimination of emissions from each individual source sectors, the same results may not be expected from studies that test complete or other reduction strategies due to the non-linear relationship of emission – PM pollution – PV power output.

Table 3.7: Potential additional PV electricity bonus (GWh yr^{-1}) and economic benefits (US$ million yr^{-1}) by combining decadal mean (2008–2017) capacity factor improvements (ΔCF) from halving residential emissions with the installed capacities as of 2019 in China. The last row provides the national total values for all variables except ΔCFs.

Province	Installed capacities (10^4kW)			ΔCF		Electricity bonus (GWh yr^{-1})			Economic benefits (US$ million yr^{-1})		
	distributed PV	utility PV	total	Tilt	OAT	distributed PV	utility PV	total	distributed PV	utility PV	total
Xinjiang	14.0	1066.0	1080.0	0.0013	0.0014	1.56	132.6	134.15	0.13	11.27	11.4
Xizang	0.0	110.0	110.0	0.0051	0.0057	0.0	54.62	54.62	0.0	4.64	4.64
Inner Mongol	80.0	1001.0	1081.0	0.0016	0.0018	11.39	159.47	170.86	0.97	13.56	14.52
Gansu	72.0	836.0	908.0	0.0018	0.0021	11.04	151.54	162.58	0.94	12.88	13.82
Yunnan	25.0	350.0	375.0	0.0041	0.005	8.9	154.65	163.55	0.76	13.15	13.9
Heilongjiang	79.0	195.0	274.0	0.0038	0.004	26.04	68.65	94.68	2.21	5.83	8.05
Jilin	69.0	205.0	274.0	0.0052	0.0055	31.16	97.98	129.14	2.65	8.33	10.98
Liaoning	97.0	246.0	343.0	0.0048	0.0053	40.83	115.01	155.84	3.47	9.78	13.25
Guangxi	30.0	105.0	135.0	0.0046	0.0058	12.0	53.47	65.47	1.02	4.55	5.56
Guangdong	308.0	302.0	610.0	0.0038	0.0046	102.69	120.91	223.6	8.73	10.28	19.01
Hainan	13.0	127.0	140.0	0.0016	0.0022	1.87	24.17	26.05	0.16	2.05	2.21
Fujian	131.0	38.0	169.0	0.003	0.0034	34.23	11.44	45.67	2.91	0.97	3.88
Zhejiang	925.0	414.0	1339.0	0.0034	0.0039	275.48	141.23	416.71	23.42	12.0	35.42
Shanghai	103.0	6.0	109.0	0.002	0.0024	17.88	1.25	19.13	1.52	0.11	1.63
Jiangsu	665.0	821.0	1486.0	0.0049	0.0058	285.69	416.2	701.89	24.28	35.38	59.66
Shandong	942.0	677.0	1619.0	0.0059	0.0071	484.05	418.31	902.36	41.14	35.56	76.7
Hebei	512.0	962.0	1474.0	0.0051	0.0061	228.25	510.31	738.56	19.4	43.38	62.78
Tianjin	39.0	104.0	143.0	0.0063	0.0075	21.64	68.33	89.97	1.84	5.81	7.65
Beijing	46.0	5.0	51.0	0.0056	0.0066	22.5	2.88	25.38	1.91	0.24	2.16
Sichuan	19.0	169.0	188.0	0.0038	0.0046	6.31	67.61	73.92	0.54	5.75	6.28
Chongqing	7.0	58.0	65.0	0.0047	0.0059	2.91	29.89	32.8	0.25	2.54	2.79
Guizhou	19.0	491.0	510.0	0.0047	0.0059	7.83	253.66	261.49	0.67	21.56	22.23
Hunan	189.0	155.0	344.0	0.0045	0.0053	74.0	72.04	146.04	6.29	6.12	12.41
Ningxia	74.0	844.0	918.0	0.002	0.0023	12.94	172.89	185.83	1.1	14.7	15.8
Shaanxi	161.0	778.0	939.0	0.0038	0.0045	53.11	308.97	362.08	4.51	26.26	30.78
Qinghai	15.0	1086.0	1101.0	0.002	0.0024	2.66	226.03	228.69	0.23	19.21	19.44
Shanxi	231.0	857.0	1088.0	0.0034	0.0041	69.6	310.76	380.36	5.92	26.41	32.33
Jiangxi	263.0	367.0	630.0	0.004	0.0046	91.46	146.47	237.94	7.77	12.45	20.22
Henan	454.0	600.0	1054.0	0.0053	0.0065	212.24	342.9	555.13	18.04	29.15	47.19
Hubei	202.0	419.0	621.0	0.0048	0.0057	84.84	208.11	292.95	7.21	17.69	24.9
Anhui	481.0	773.0	1254.0	0.005	0.0059	211.52	398.25	609.77	17.98	33.85	51.83
Total	6265.0	14167.0	20432.0	NaN	NaN	2446.62	5240.6	7687.22	207.96	445.45	653.41

Table 3.8: Potential additional PV electricity bonus (GWh yr^{-1}) and economic benefits (US$ million yr^{-1}) by combining decadal mean (2008–2017) capacity factor improvements (ΔCF) from halving residential emissions with the installed capacities as of 2019 in India. The last row provides the national total values for all variables except ΔCFs.

State	Installed capacities (MW)			ΔCF		Electricity bonus (GWh yr^{-1})			Economic benefits (US$ million yr^{-1})		
	distributed PV	utility PV	total	Tilt	OAT	distributed PV	utility PV	total	distributed PV	utility PV	total
Gujarat	416.6	2346.9	2763.6	0.0045	0.0053	16.26	109.46	125.73	1.28	8.65	9.93
Jammu and Kashmir	10.8	8.5	19.3	0.0048	0.0057	0.45	0.42	0.88	0.04	0.03	0.07
Arunachal Pradesh	4.3	1.3	5.6	0.0025	0.0025	0.09	0.03	0.12	0.01	0.0	0.01
Assam	30.6	10.7	41.2	0.0066	0.0082	1.77	0.76	2.53	0.14	0.06	0.2
Sikkim	0.1	0.0	0.1	0.0058	0.0048	0.0	0.0	0.0	0.0	0.0	0.0
West Bengal	43.2	66.3	109.4	0.008	0.0105	3.03	6.11	9.14	0.24	0.48	0.72
Uttaranchal	75.7	239.8	315.5	0.0051	0.0055	3.41	11.47	14.88	0.27	0.91	1.18
Uttar Pradesh	146.1	899.0	1045.1	0.0092	0.0113	11.78	89.34	101.13	0.93	7.06	7.99
Bihar	10.4	138.9	149.4	0.0109	0.014	1.0	17.02	18.02	0.08	1.34	1.42
Nagaland	1.0	0.0	1.0	0.0048	0.006	0.04	0.0	0.04	0.0	0.0	0.0
Manipur	4.6	0.0	4.6	0.0056	0.007	0.22	0.0	0.22	0.02	0.0	0.02
Mizoram	1.4	0.1	1.5	0.0066	0.0085	0.08	0.01	0.09	0.01	0.0	0.01
Tripura	4.4	5.0	9.4	0.0069	0.0088	0.27	0.39	0.65	0.02	0.03	0.05
Meghalaya	0.1	0.0	0.1	0.0086	0.0107	0.01	0.0	0.01	0.0	0.0	0.0
Punjab	118.5	828.6	947.1	0.0052	0.0066	5.42	48.23	53.65	0.43	3.81	4.24
Rajasthan	206.7	4637.5	4844.2	0.0046	0.0057	8.33	230.74	239.07	0.66	18.23	18.89
Himachal Pradesh	15.6	17.0	32.6	0.0043	0.0045	0.58	0.67	1.25	0.05	0.05	0.1
Andhra Pradesh	88.0	3471.0	3559.0	0.0047	0.0058	3.6	177.65	181.25	0.28	14.03	14.32
Orissa	14.3	383.6	397.8	0.0056	0.0072	0.7	24.21	24.91	0.06	1.91	1.97
Daman and Diu	6.4	10.1	16.6	0.0026	0.003	0.14	0.27	0.41	0.01	0.02	0.03
Maharashtra	216.1	1447.3	1663.4	0.0058	0.0068	11.0	86.84	97.84	0.87	6.86	7.73
Goa	3.8	0.9	4.8	0.0052	0.0061	0.17	0.05	0.22	0.01	0.0	0.02
Karnataka	232.8	7042.2	7274.9	0.0048	0.0058	9.86	356.52	366.38	0.78	28.17	28.94
Kerala	41.8	100.0	141.8	0.0034	0.0041	1.24	3.58	4.83	0.1	0.28	0.38
Puducherry	5.5	0.0	5.5	0.0025	0.0032	0.12	0.0	0.12	0.01	0.0	0.01
Tamil Nadu	155.8	3632.5	3788.4	0.0036	0.0045	4.93	141.93	146.86	0.39	11.21	11.6
Lakshadweep	0.0	0.8	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Andaman and Nicobar	4.6	7.6	12.2	0.0006	0.001	0.03	0.07	0.09	0.0	0.01	0.01
Jharkhand	19.4	19.0	38.4	0.0079	0.0102	1.35	1.71	3.05	0.11	0.13	0.24
Delhi	147.2	9.0	156.1	0.0071	0.0087	9.09	0.68	9.77	0.72	0.05	0.77
Dadra and Nagar Haveli	3.0	2.5	5.5	0.0029	0.0033	0.07	0.07	0.15	0.01	0.01	0.01
Chandigarh	30.6	6.3	37.0	0.0077	0.0094	2.07	0.52	2.59	0.16	0.04	0.2
Madhya Pradesh	49.4	2188.1	2237.5	0.0071	0.0085	3.09	163.27	166.36	0.24	12.9	13.14
Chhattisgarh	15.5	215.8	231.4	0.0061	0.0075	0.83	14.24	15.07	0.07	1.13	1.19
Haryana	118.5	130.8	249.3	0.0057	0.0071	5.96	8.13	14.1	0.47	0.64	1.11
Telangana	90.5	3530.3	3620.8	0.0058	0.007	4.56	216.8	221.37	0.36	17.13	17.49
Total	2333.2	31397.3	33730.6	NaN	NaN	111.58	1711.21	1822.79	8.81	135.19	144.0

Chapter 4

Approximately linear benefits to Asian wintertime air quality and photovoltaic power output from reducing residential fuel emissions

For this chapter, I designed and performed analysis under the supervision of Prof Paul Palmer. I wrote the chapter with editing from Prof Paul Palmer.

Yao, F. and Palmer, P.I., 2021. Approximately linear benefits to Asian wintertime air quality and photovoltaic power output from reducing residential fuel emissions. In preparation for submission.

4.1 Abstract

Reducing residential emissions is critical to alleviate particulate matter (PM) pollution over Asian countries which are home to roughly 60% of the global population It logically follows that reducing residential emissions will lead to co-benefits to solar photovoltaic (PV) power output over Asian countries in that less PM pollution will allow more solar radiation to reach and be utilized by solar PV panels. A full evaluation of benefits to air quality and PV power output from reducing residential fuel emissions is critical, but inadequately addressed due to data and model gaps. Here I combine well-evaluated models of solar PV performance and atmospheric composition to show that Eastern and Northeastern China, and the Indo-Gangetic Plain are the three key regions where reducing residential particularly solid biofuel emissions leads approximately linearly to reductions in wintertime PM pollution and improvements in wintertime PV efficiency. Completely removing residential emissions would result in 20–30% and 30–40% reductions in column-integrated aerosol optical depth at wavelength of 550 nm and ground-level $PM_{2.5}$ (PM with an aerodynamic diameter $\leq$ 2.5 μm) concentrations, respectively, and an approximately 30% improvement in PV efficiency. I attribute these approximately linear benefits to the large volume of carbon emissions, primarily owing to the low combustion and thermal efficiencies of residential devices and the general absence of any end-of-pipe controls, that typically form into carbon aerosols in an approximately linear way in the atmosphere after aggregating to temporal and spatial mean values.

4.2 Introduction

Asia is the Earth's largest and most populous continent, located primarily in the Eastern and Northern Hemisphere. It is home to ~4.5 billion people (data as

of June 2019), representing roughly 60% of the global population (Population of Asia, 2019). However, it is generally troubled by particulate matter (PM) pollution (Van Donkelaar et al., 2010, 2015; Boys et al., 2014; Hammer et al., 2020), which are typically distributed proportionally to population. Identifying anthropogenic sources that are responsible for the highest PM exposures would help develop effective emission mitigation strategies beneficial for human health. There is growing evidence suggesting that residential emissions significantly impact PM pollution-associated human health over Asian countries (Lelieveld et al., 2015; McDuffie et al., 2021), particularly over China (Liu, Mauzerall, Chen, Zhang, Song, Peng, Klimont, Qiu, Zhang, Hu et al., 2016; Zhao et al., 2018; Shen et al., 2019; Yun et al., 2020) and India (Conibear et al., 2018; Chowdhury et al., 2019) during wintertime.

In addition to health benefits, reductions in PM pollution as a result of decreasing residential emissions will have side-benefits, such as an increase in electricity produced from solar photovoltaic (PV) panels as less PM pollution in the atmosphere allows more solar radiation to reach the surface and to be utilized by solar PV panels (Li, Wagner, Peng, Yang and Mauzerall, 2017; Labordena et al., 2018; Sweerts et al., 2019). Moreover, less PM pollution near the surface results in less PM fluxes deposited on solar PV panels, and thereby more solar radiation being received and utilized by PV semiconductor materials (Bergin et al., 2017; Li et al., 2020). While Chapters 2 & 3 have shown that halving residential emissions substantially benefits Asian PV power output and surface air quality with respect to PM with an aerodynamic diameter $\leq$ 2.5 μm ($PM_{2.5}$), it remains unclear how these benefits will change as residential emissions are decreased. Meanwhile, residential emissions are composed of burning several fuel types (e.g. coal, solid biofuel, and others which are primarily gaseous fuels) whose relative contributions to these benefits are unknown under different emission

reduction scenarios, either. Gaining these knowledge, however, is critical to develop optimized, realistic emission mitigation strategies.

To address this gap in knowledge, in this chapter I quantify the benefits of reducing residential fuel emissions to Asian wintertime air quality and PV power output. To this end, I continue to use the integrated model developed in Chapter 2 but at a finer horizontal resolution of 0.5° latitude × 0.625° longitude over the Asian nested domain. The higher resolution simulations and corresponding findings would be more appropriate to guide local actions. I simultaneously focus on reductions in ground-level $PM_{2.5}$ concentrations, column-integrated aerosol optical depth at wavelength of 550 nm (AOD_{550nm}), and PM damages to solar PV electricity generation efficiency. I use the capacity factor (CF) to describe the PV efficiency of the three widely used solar panel settings: horizontal fixed (Flat), fixed with optimal tilt (Tilt), and one-axis tracking (OAT). The CF is defined as the ratio of the actual power output of a PV panel to the theoretical maximum power output. Despite the many nonlinear physical and chemical processes from emission release to PM pollution formation, I find approximately linear benefits to Asian wintertime air quality and PV power output from reducing residential particularly solid biofuel emissions.

4.3 Methods

To facilitate the higher resolution simulations over the Asian nested domain in this chapter, I use the 2° latitude × 2.5° longitude global version of the GEOS-Chem model coupled with rapid radiative transfer model for general circulation models (GCRT) of the integrated model to provide lateral, time-dependent boundary conditions for its nested simulation. I define a 3-box "buffer zone" around my Asian study region to incorporate spurious edge effects from the lateral boundary

conditions for the nested simulation, and exclude data from this zone from further scientific analysis. I spin-up both models for > two years prior to my study period, January 1 to February 29, 2008, to minimize the impact of initial conditions. The atmospheric transport and chemistry steps are 10 and 20 minutes, respectively, for the global model, and 5 and 10 minutes, respectively, for the nested model.

I drive both models by the Modern-Era Retrospective analysis for Research and Application, Version 2 (MERRA-2) meteorological analyses (Gelaro et al., 2017), provided by the Global Modeling and Assimilation Office at the NASA Goddard Space Flight Center. I run both models over the reduced 47 vertical terrain-following sigma levels that describe the atmosphere from the surface to 0.01 hPa, of which 30 are below the dynamic tropopause. The nested model uses MERRA-2 meteorology at its native resolution of 0.5° latitude × 0.625° longitude, while the global model uses MERRA-2 meteorology at a coarse resolution of 2° latitude × 2.5° longitude.

The configuration for emission inventories follows that of the previous chapters except that I further fine-tune the global anthropogenic emissions from the Community Emissions Data System updated for the Global Burden of Disease - Major Air Pollutant Sources project ($\text{CEDS}_{GBD-MAPS}$). $\text{CEDS}_{GBD-MAPS}$ provides the most contemporary global emission estimates to date for seven major atmospheric pollutants as a function of 11 anthropogenic source sectors and four fuel types. In this chapter, I mainly focus on the residential sector which includes combustion and emissions of coal, solid biofuel, liquid-fuel, and natural-gas, and remaining process-level emissions. Given the relatively small amount of liquid-fuel, natural-gas, and remaining process-level emissions in the residential sector, I combine them as other fuels for investigation in this chapter.

I run the nested model with original emissions to output baseline ground-level $\text{PM}_{2.5}$ concentrations, column-integrated AOD_{550nm}, and PM damages to PV

efficiency. I then run the nested model again but with reduced residential fuel emissions. I can subsequently determine the benefits to air quality and PV power output from reducing residential fuel emissions by taking the difference of baseline and sensitivity simulations. To describe these benefits in a relatively full manner, I run a total of $4 \times 4 = 16$ sensitivity simulations corresponding to 25%, 50%, 75%, and 100% reductions in residential total, coal, solid biofuel, and other emissions. With respect to PM damages to PV efficiency, I follow the convention of Chapters 2 & 3 and report PM dimming, soiling, and total impacts. A reduction in these impacts as a consequence from reducing residential fuel emissions are accordingly termed as brightening, cleaning, and total benefits, respectively.

4.4 Results

4.4.1 Air quality improvements

Figure 4.1 shows similar distributions for ground-level $PM_{2.5}$ and column-integrated AOD_{550nm}, with high values distributed over Eastern China (region A) and the Indo-Gangetic Plain (region C). Nonetheless, it is worth noting that the eastern part of Eastern China has higher or comparable ground-level $PM_{2.5}$ concentrations than the western part of Eastern China, while the opposite is true for column-integrated AOD_{550nm}. Moreover, Northeastern China (region B) shows relatively high values for ground-level $PM_{2.5}$ but not column-integrated AOD_{550nm}. Although a slight deviation from the immediate focus of this chapter, the interpretation of these spatial patterns suggests not only the possibility but also the difficulty of correlating column-integrated AOD_{550nm} which can retrieved from satellite remote sensing observations to ground-level $PM_{2.5}$ concentrations, as investigated in Chapter 5.

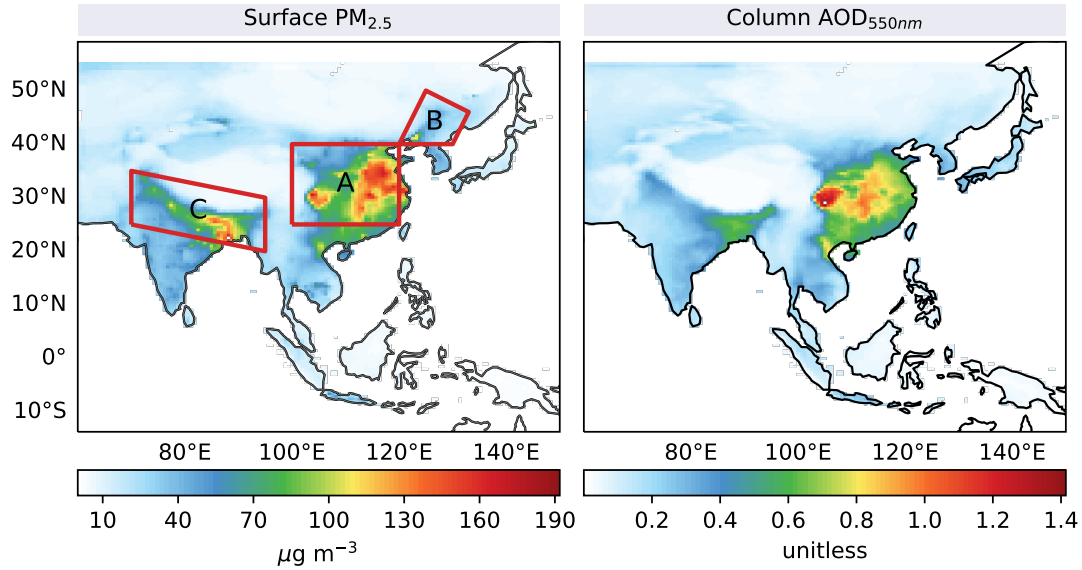


Figure 4.1: Mean spatial distribution of ground-level $PM_{2.5}$ concentrations and column-integrated AOD_{550nm} over Asian domain from January 1 to February 29, 2008 in the baseline simulation. The three key regions with high PM pollution levels are marked with red polygons: (A) Eastern China, (B) Northeastern China, and (C) Indo-Gangetic Plain.

Figure 4.2 shows that decreases in ground-level $PM_{2.5}$ concentrations as a result of reducing residential emissions are centred over Eastern and Northeastern China and the Indo-Gangetic Plain. The four residential emission reduction scenarios present similar spatial distributions of decreases in ground-level $PM_{2.5}$ concentrations, with higher emission reductions leading to more decreases. The solid biofuel emission reductions contribute to the majority decreases of ground-level $PM_{2.5}$ concentrations over the three regions. The coal emission reductions are also important in Eastern China.

Figure 4.3 shows that decreases in column-integrated AOD_{550nm} as a result of reducing residential emissions are centred over Eastern China and the Indo-Gangetic Plain. Spatial patterns of decreases in column-integrated AOD_{550nm} are similar

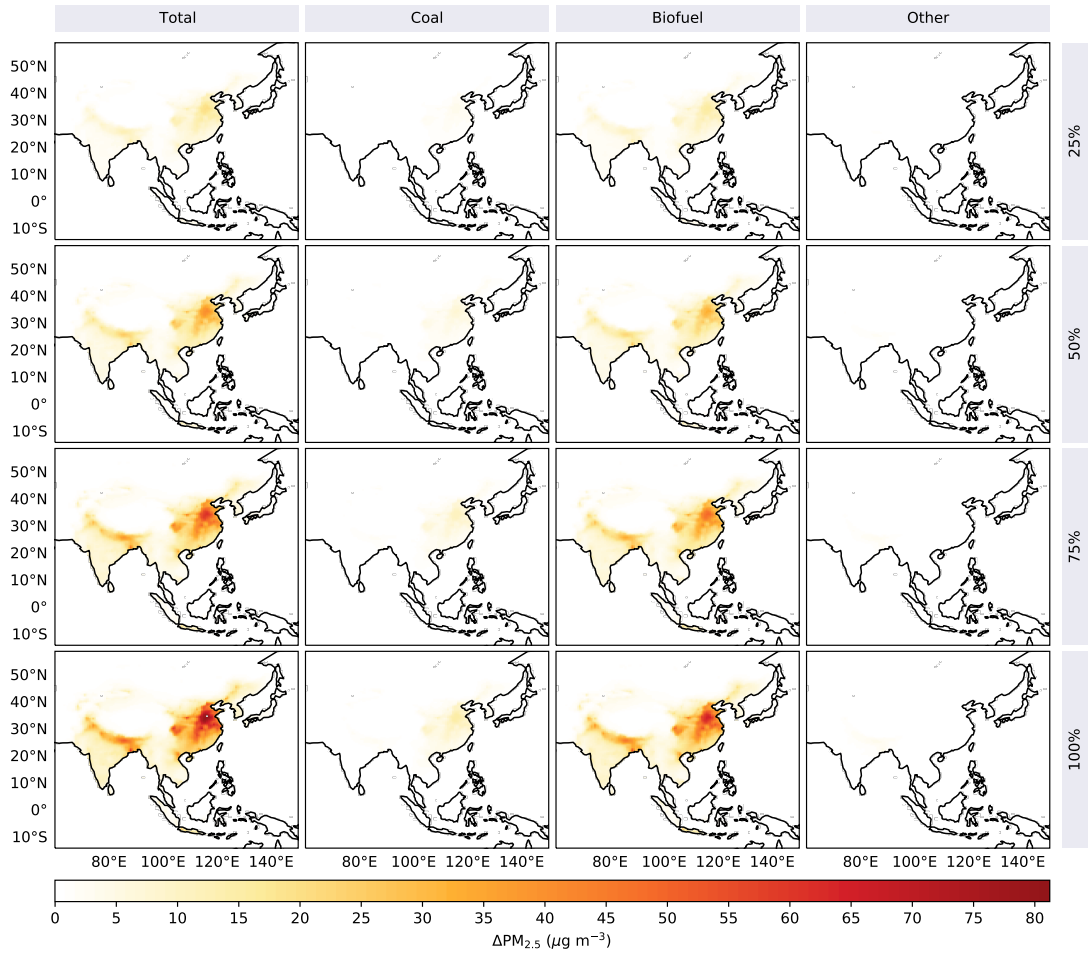


Figure 4.2: Mean spatial distribution of decreases in ground-level $PM_{2.5}$ concentrations over Asian domain from January 1 to February 29, 2008 as a result of reducing residential fuel emissions.

across the four residential emission reduction scenarios, with higher emission reductions leading to more decreases. The solid biofuel emission reductions lead the spatial pattern of the reductions in column-integrated AOD_{550nm} over the two regions. The coal emission reductions are also important in Eastern China.

To demonstrate the air quality benefits from reducing residential fuel emissions on the regional scale, I further calculate the area-weighted statistics over the three key regions defined in Figure 4.1.

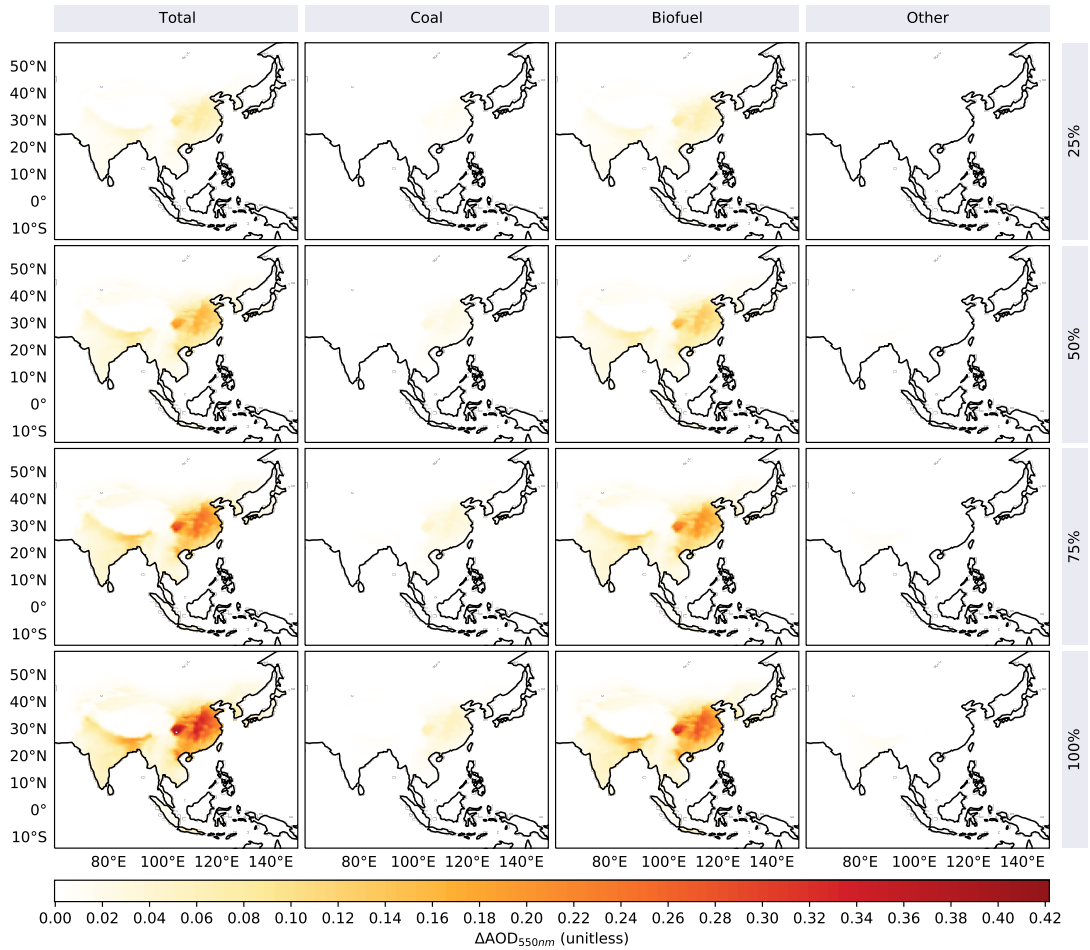


Figure 4.3: Mean spatial distribution of decreases in column-integrated AOD_{550nm} over Asian domain from January 1 to February 29, 2008 as a result of reducing residential fuel emissions.

Figure 4.4 shows that the benefits to regional-scale air quality follows a decreasing order of Eastern China, the Indo-Gangetic Plain, and Northeastern China. There exists an approximately linear relationship between the benefits to regional-scale air quality from reducing residential emissions and the extent to which these emissions are reduced. The sum of the benefits to regional-scale air quality from reducing individual residential fuel emissions approximately equal the benefits to regional-scale air quality from reducing residential total emissions. All of these

results point to one policy-relevant conclusion, i.e. reducing emissions from the residential sector will likely lead approximately linearly to air quality benefits.

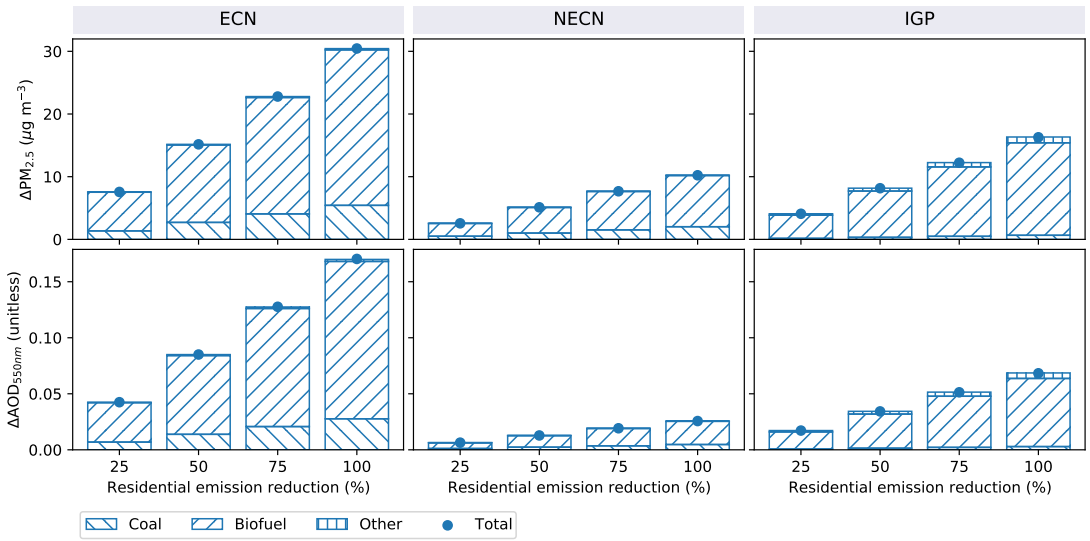


Figure 4.4: Absolute benefits to regional-scale air quality from reducing residential fuel emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangeric Plain (IGP).

While reducing residential emissions leads to varying absolute air quality benefits over different regions, Figure 4.5 shows that this spatial difference becomes small in terms of percentage values. A 100% residential emission reduction leads to 30–40% reductions in ground-level $PM_{2.5}$ concentrations across the three key regions. The corresponding range is 20-30% with respect to column-integrated AOD_{550nm}. Solid biofuel and coal are the leading fuels whose emission reductions in the residential sector provide considerable air quality benefits over Eastern and Northeastern China. The share of solid biofuel emission reductions dominates the air quality benefits from reducing residential emissions over the Indo-Gangetic Plain.

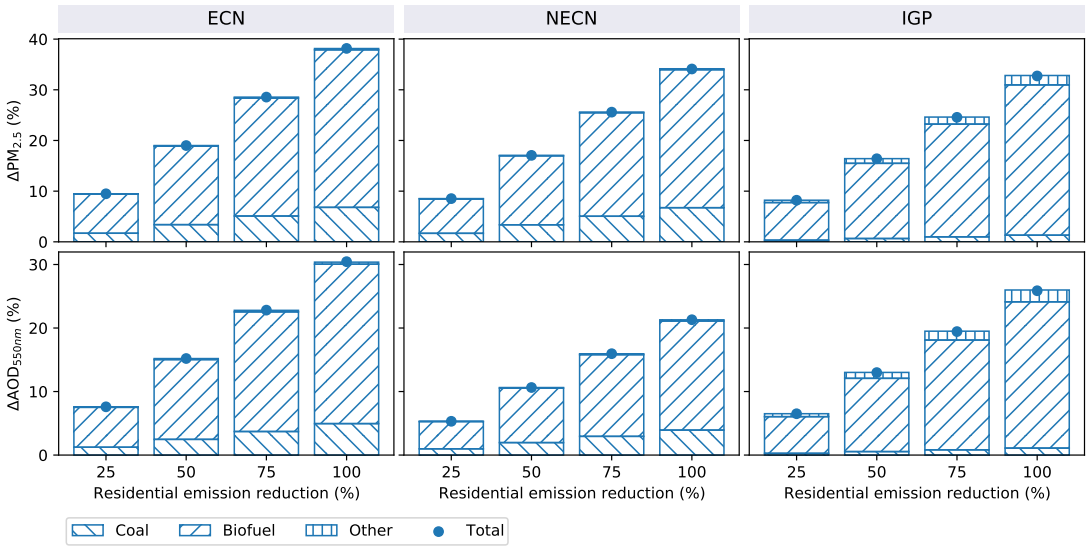


Figure 4.5: Percentage benefits (relative the baseline case) to regional-scale air quality from reducing residential fuel emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP).

4.4.2 PV efficiency improvements

Figure 4.6 shows that the Tibetan-Plateau and Southeast Asia feature high PV efficiency and low PM-induced PV efficiency losses. Tilt panels significantly improve the PV efficiency over high latitudes compared to flat panels. OAT panels significantly improve the PV efficiency over Southeast Asia compared to tilt panels.

The spatial pattern of PM dimming (soiling) impacts follows that of baseline column-integrated AOD_{550nm} (ground-level $PM_{2.5}$ concentrations), suggesting the feasibility of using the latter as the proxy for the former. Tilt and OAT panels report relatively higher PM dimming and soiling impacts than flat panels. The magnitude of PM soiling are broadly slightly higher than that of PM dimming.

PM dimming and soiling impacts jointly constitute the total PM impacts, so regions where PV efficiency severely affected by PM impacts are those

characterized by either high ground-level $PM_{2.5}$ concentrations or high column-integrated AOD_{550nm} levels, and they are Eastern and Northeastern China, and the Indo-Gangetic Plain.

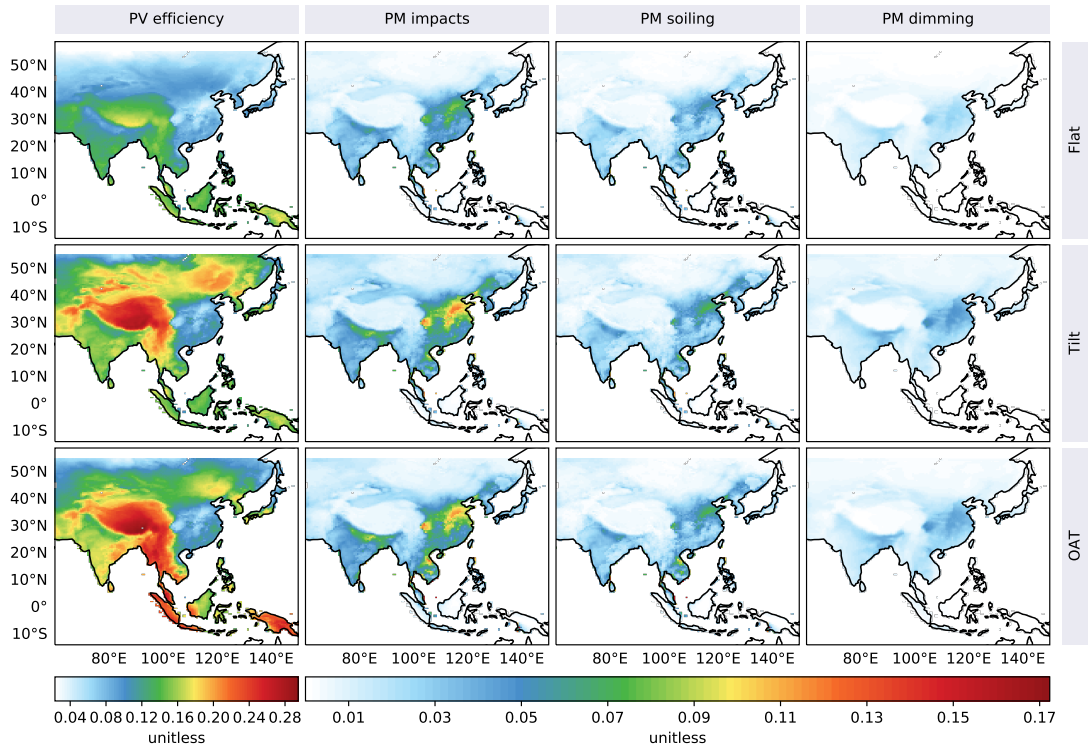


Figure 4.6: Mean spatial distribution of PV efficiency and its losses due to PM total, soiling, and dimming impacts over Asian domain from January 1 to February 29, 2008 for flat, tilt, and one-axis tracking panels.

Figure 4.7 shows that the spatial pattern of brightening benefits from completely reducing residential emissions follows that of PM dimming impacts. Over Eastern and the Indo-Gangetic Plain, reducing residential (particularly solid biofuel) emissions result in considerable brightening benefits, with higher values found for tilt and OAT than flat panels. Reducing residential coal and other fuel emissions are also beneficial for Eastern China and the Indo-Gangetic Plain, respectively.

Figure 4.8 shows that the spatial pattern of cleaning benefits from completely reducing residential emissions follows that of PM soiling impacts. Over Eastern

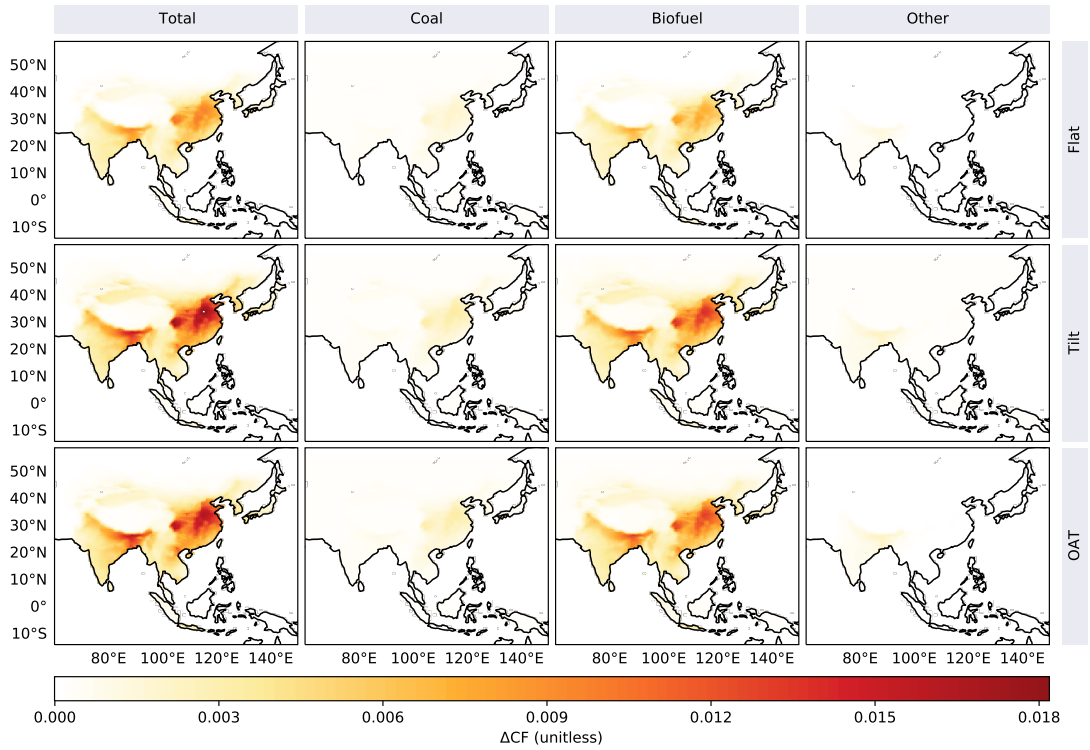


Figure 4.7: Mean spatial distribution of brightening benefits over Asian domain from January 1 to February 29, 2008 as a result of completely reducing residential fuel emissions for flat, tilt, and one-axis tracking panels. Similar spatial patterns (not shown) are found for other extents (25%, 50%, and 75%) to which residential emissions are reduced.

and Northeastern China, and the Indo-Gangetic Plain, reducing residential (particularly solid biofuel) emissions result in considerable cleaning benefits, with higher values found for tilt and OAT than flat panels. Reducing residential coal emissions is also beneficial for Eastern and Northeastern China.

Combining brightening and total benefits, Figure 4.9 shows that completely reducing residential (particularly solid biofuel) emissions leads to significant improvements in PV efficiency over Eastern and Northeastern China, and the Indo-Gangetic Plain. Reducing residential coal and other fuel emissions are also

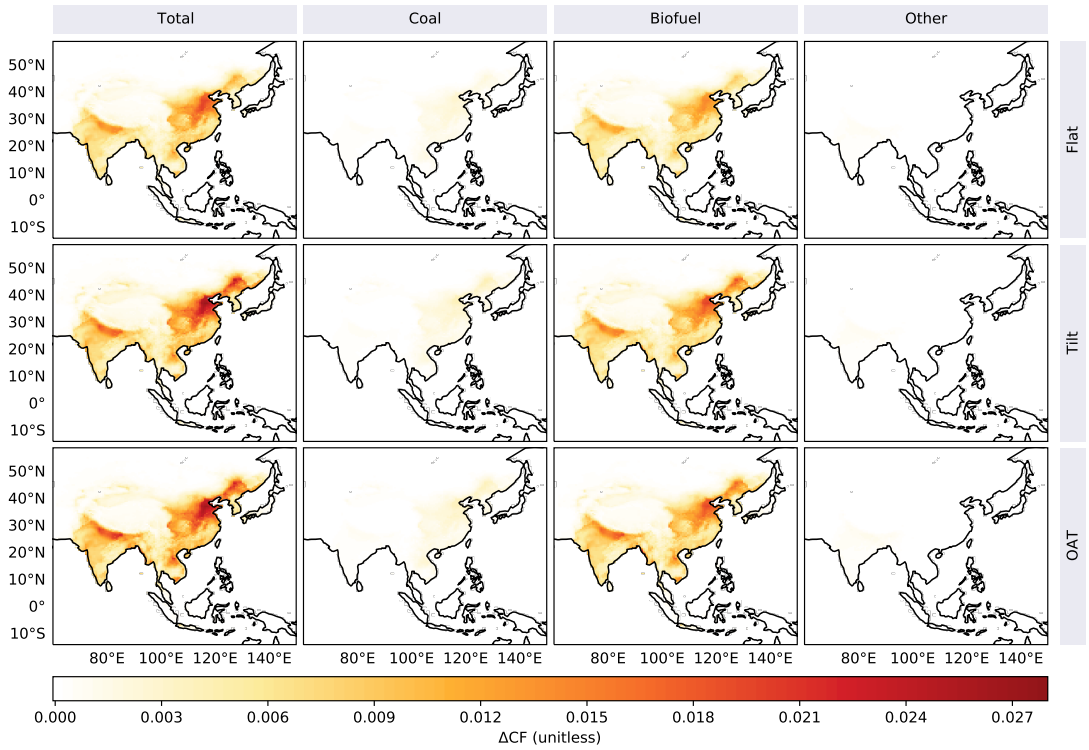


Figure 4.8: Mean spatial distribution of cleaning benefits over Asian domain from January 1 to February 29, 2008 as a result of completely reducing residential fuel emissions for flat, tilt, and one-axis tracking panels. Similar spatial patterns (not shown) are found for other extents (25%, 50%, and 75%) to which residential emissions are reduced.

beneficial for Eastern and Northeastern China, and the Indo-Gangetic Plain, respectively.

On the regional scale, Figure 4.10 shows that brightening benefits are comparable or slightly smaller than cleaning benefits over Eastern China and the Indo-Gangetic Plain, while brightening benefits are obviously smaller than cleaning benefits over Northeastern China. Flat panels always report the smallest benefits. Tilt panels report the highest benefits over Eastern and Northeastern China, followed by OAT panels, while the opposite is true over the Indo-Gangetic Plain.

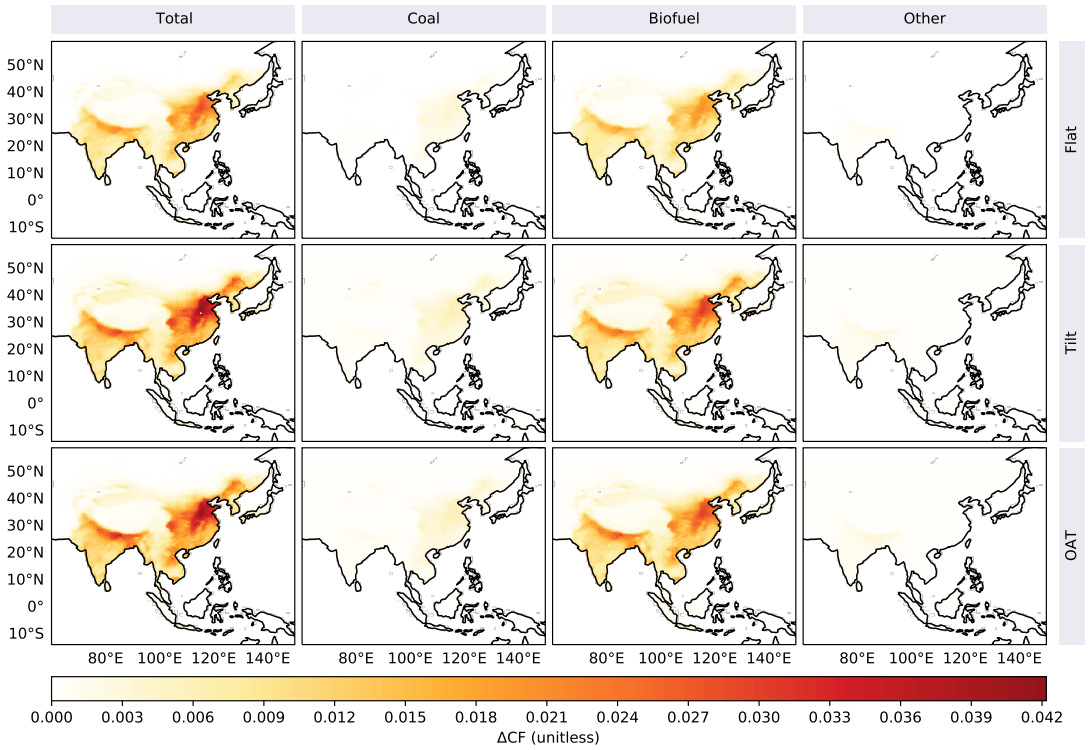


Figure 4.9: Mean spatial distribution of total benefits over Asian domain from January 1 to February 29, 2008 as a result of completely reducing residential fuel emissions for flat, tilt, and one-axis tracking panels. Similar spatial patterns (not shown) are found for other extents (25%, 50%, and 75%) to which residential emissions are reduced.

All benefits in all panels increase approximately linearly with the increase of residential emission reductions. The solid biofuel emission reductions contribute most benefits over the three regions, followed by coal and other fuel emission reductions over Eastern and Northeastern China, and the Indo-Gangetic Plain, respectively. Except for cleaning and total benefits under high emission reduction scenarios over Eastern China, the sum of the benefits from reducing individual residential fuel emissions approximately equal the benefits from reducing residential total emissions. The small data gaps in cleaning and total benefits under high emission reductions over Eastern China suggest the true logarithmic relationship between

residential emission reductions and corresponding benefits. Nonetheless, this true logarithmic relationship almost approximates to linear relationship in most cases.

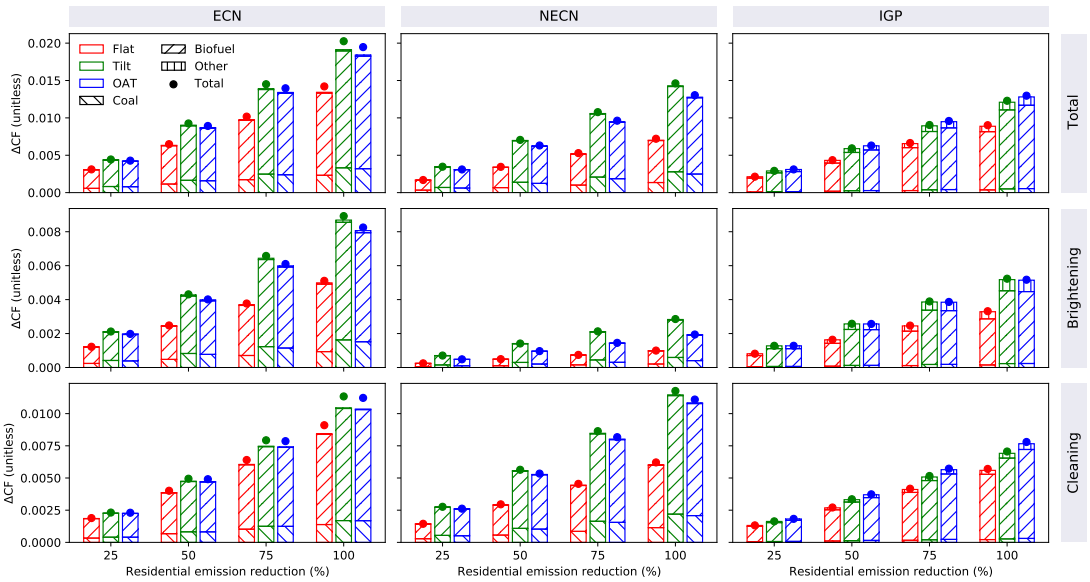


Figure 4.10: Absolute benefits to regional-scale PV efficiency of flat, tilt, and one-axis tracking panels from reducing residential fuel emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP).

Figure 4.11 shows that after scaling brightening, cleaning, and total benefits to PM dimming, soiling, and total impacts, respectively, these benefits become almost equal across the three panels, and they also present less spatial differences. This suggests that the more PV efficiency we loss, the more PV efficiency we can recover by reducing residential emissions. Completely removing residential emissions would lead to approximately 30% benefits over Eastern China and the Indo-Gangetic Plain, with brightening comparable to cleaning. Completely removing residential emissions would lead to approximately 30% benefits over Northeastern China, with brightening slightly smaller than cleaning.

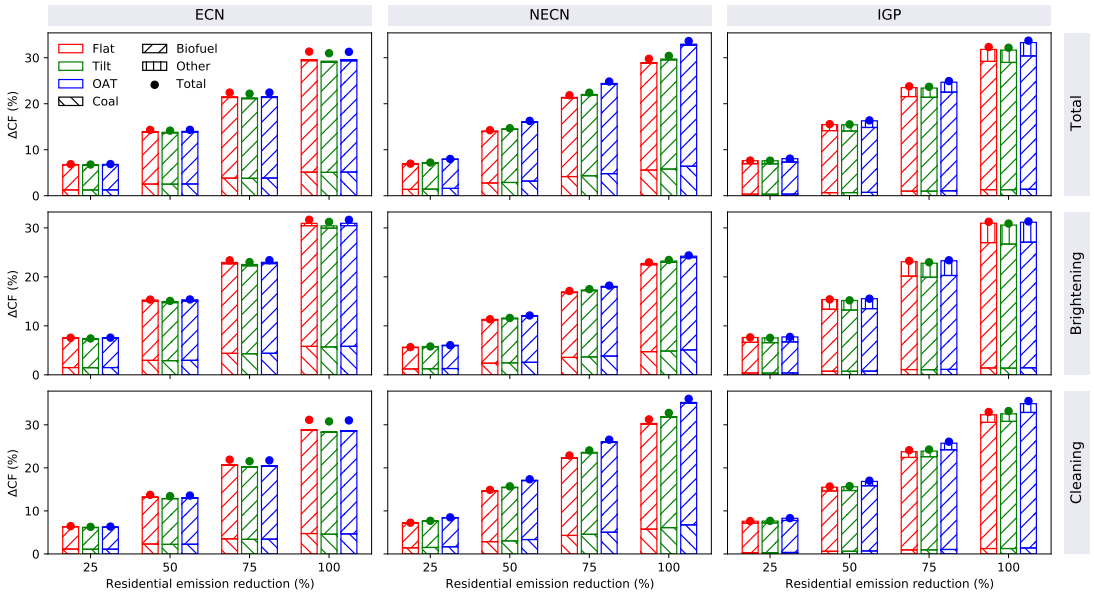


Figure 4.11: Percentage benefits (relative to the baseline case) to regional-scale PV efficiency of flat, tilt, and one-axis tracking panels from reducing residential fuel emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP).

4.5 Discussion and concluding remarks

To the best of my knowledge, this chapter is one of the earliest studies that quantifies the benefits of reducing residential fuel emissions to Asian wintertime air quality and PV power output. I identify three key regions (Eastern and Northeastern China and the Indo-Gangetic Plain) where reducing residential emissions leads to substantial benefits to air quality and PV power output. I find that these benefits change approximately linearly with the extent to which residential emissions are reduced, and that completely removing residential emissions results in 20–30% and 30–40% reductions in column-integrated AOD_{550nm} and ground-level $PM_{2.5}$ concentrations, respectively, and approximately 30% improvements in PV efficiency of the three widely used solar panel settings. I find that most of these

benefits result from reducing residential solid biofuel emissions. Reducing residential coal and other emissions are also important in Eastern and Northeastern China, and the Indo-Gangetic Plain, respectively. All of these findings provide valuable guidance to the formation of optimized, realistic emission mitigation strategies that produce multiple benefits.

To explain why reducing residential emissions leads to approximately linear benefits to air quality and PV power output, I investigate the emission structure of the residential sector. With $\text{CEDS}_{GBD-MAPS}$ monthly gridded emission data, Figure 4.12 shows that the emission structure of the residential sector is quite similar across the three regions, with carbon monoxide (CO) as the dominant-emitting air pollutant followed by non-methane volatile organic compounds (NMVOC) and organic carbon (OC). These large volume of carbon emissions, primarily owing to the low combustion and thermal efficiencies of residential devices and the general absence of any end-of-pipe controls, will form into carbon aerosols in the atmosphere. In particular, OC is emitted directly into the atmosphere to form what is called primary (POA) organic aerosol (OA). CO is not a direct precursor of OA, but it is feasible that GEOS-Chem calculates anthropogenic contribution to secondary OA (SOA) using CO emissions as a proxy, i.e. 6.9% scaled co-emission of SOA precursor from fossil fuel CO (Pai et al., 2020). Therefore, Figure 4.13 shows that reducing residential emissions will mainly cause decline in POA and SOA. Despite the many nonlinear physical and chemical processes from emission release to PM pollution formation, Figure 4.13 also discloses that after aggregating to temporal and regional mean statistics, reducing residential emissions consistently leads approximately linearly to reductions in levels of major PM chemical components with respect to both ground-level $\text{PM}_{2.5}$ and column-integrated AOD_{550nm}. The lumped ground-level $\text{PM}_{2.5}$ concentrations and column-integrated AOD_{550nm} then logically follow this pattern of approximately linear benefits from reducing residential emissions

as shown in Figures 4.4 and 4.5. As described earlier, ground-level $PM_{2.5}$ concentrations and column-integrated AOD_{550nm}, to some extent, can be regarded as the proxies for PM soiling and dimming impacts on PV efficiency. As such, I subsequently find the approximately linear benefits from residential emission reductions sequentially extend to brightening, cleaning, and thereby total benefits, as shown in Figures 4.10 and 4.11.

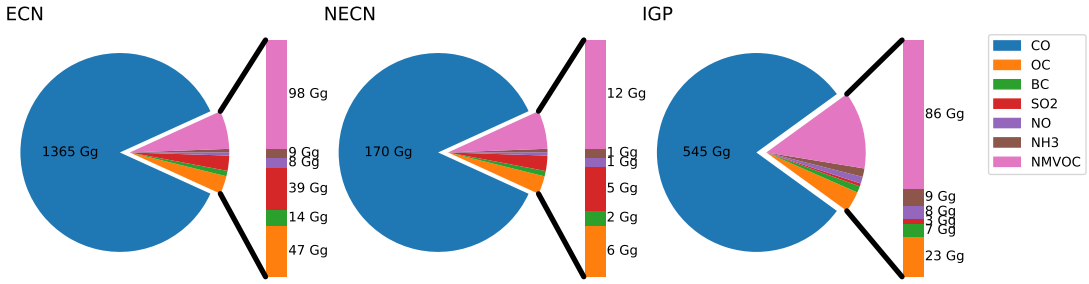


Figure 4.12: Total emissions of seven major air pollutants from the residential sector from January 1 to February 29, 2008 over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP).

I focus on the period of January 1 to February 29, 2008 because Chapter 3 have found that the consistent year-to-year benefits to PV power output from reducing residential emissions are mainly concentrated on wintertime. This way I also partly overcome the large computational costs. The comparison between Figures 4.12 and 4.14 shows that while the total residential emissions decreased from winter 2007 to winter 2016, the residential emission structure remains almost unchanged over the three regions this chapter identified and focused on. CO is still the dominant-emitting air pollutant. In this sense, while there will be variations in the analysis between years due to declined baseline emissions and varying meteorology, and variations in relative contributions from individual fuel types, the key finding of the approximately linear benefits to both air quality and PV power output from reducing residential emissions is likely to keep solid.

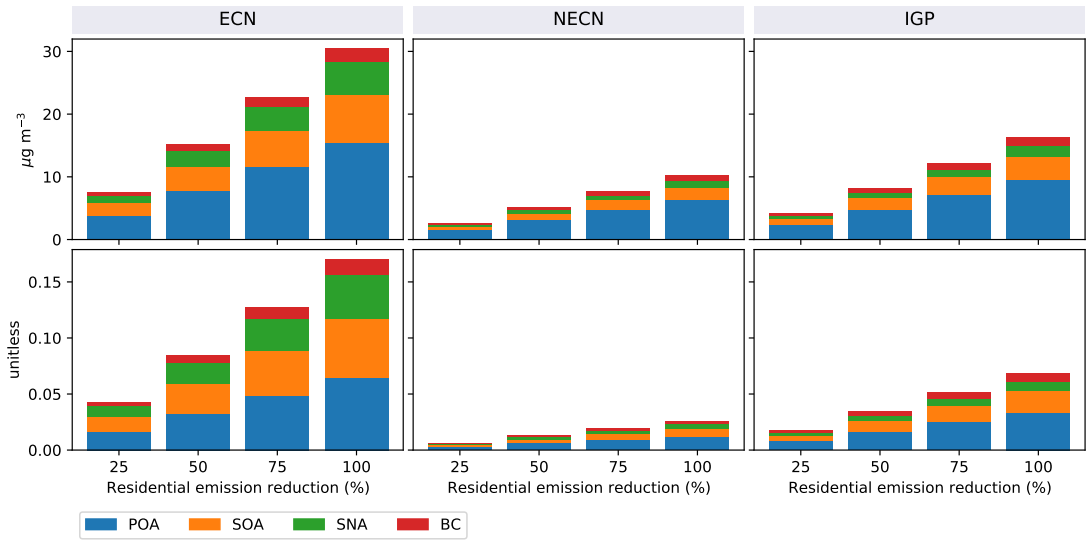


Figure 4.13: Absolute reductions in regional-scale levels of (upper) ground-level $PM_{2.5}$ and (bottom) column-integrated AOD_{550nm} major chemical components (POA: primary organic aerosol, SOA: secondary organic aerosol, SNA: sulfate - nitrate - ammonium, BC: black carbon) from reducing residential emissions over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP).

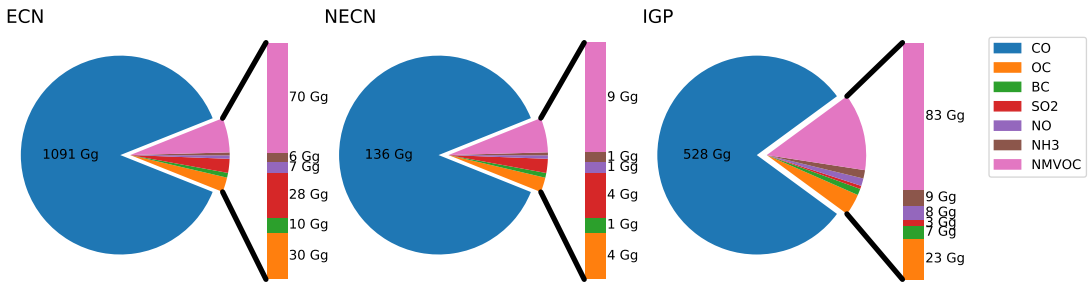


Figure 4.14: Total emissions of seven major air pollutants from the residential sector from January 1 to February 28, 2017 over Eastern China (ECN), Northeastern China (NECN), and the Indo-Gangetic Plain (IGP).

Chapter 5

A model framework to reduce bias in ground-level $PM_{2.5}$ concentrations inferred from satellite-retrieved AOD

This chapter is adapted from a published paper on which I am the lead author. I performed all data analysis, and wrote the first draft of the manuscript. Prof Paul Palmer supervised the study, and provided comments on subsequent manuscript revisions. Two anonymous reviewers provided additional comments and suggestions during peer-review.

Yao, F. and Palmer P.I., 2021. A model framework to reduce bias in ground-level $PM_{2.5}$ concentrations inferred from satellite-retrieved AOD. *Atmospheric Environment*, 248, p.118217, https://doi.org/10.1016/j.atmosenv.2021.118217.

5.1 Abstract

We present a new method to infer ground-level fine particulate matter ($PM_{2.5}$) from satellite remote sensing observations of aerosol optical depth (AOD). The conventional method generally uses a range of modelling approaches to determine $PM_{2.5}$:AOD relationships that are subsequently used to infer ground-level $PM_{2.5}$ concentrations from satellite-retrieved AOD. Here, we use a high-resolution atmospheric chemistry simulation to explore how changes in the vertical distribution of aerosol extinction coefficients affects the $PM_{2.5}$:AOD relationship and how we can use that information to improve the robustness of inferred estimates of ground-level $PM_{2.5}$ over eastern China. We define a metric, Γ_{PBL}^{AOD}, that describes the fraction of AOD that resides in the planetary boundary layer compared with the total columnar AOD. We determine physically-meaningful $PM_{2.5}$:AOD relationships using data for which $\Gamma_{PBL}^{AOD} \geq 50\%$, a criterion based on sensitivity analyses on data clusters that we identify using a hierarchical clustering method. We use statistical and machine learning methods to develop independent models that describe these $PM_{2.5}$:AOD relationships, and use a Monte Carlo approach to quantify the improvement after our selection of more physically relevant data records. Benefiting from the improved representativeness of AOD for ground-level $PM_{2.5}$, our method effectively reduces bias in inferred estimates of ground-level $PM_{2.5}$ by 10-15% (9-12%) for space-borne sensors passing over in the morning (afternoon). It also captures more variations in ground-level $PM_{2.5}$ by up to 8% (5%) for space-borne sensors passing over in the morning (afternoon), particularly over areas dominated by natural aerosols such as dust. Accordingly, our method improves the seasonal ground-level $PM_{2.5}$ maps, e.g. the bias of the autumn (winter) mean of ground-level $PM_{2.5}$ estimates over Qinghai and Gansu provinces (Shaaxi, Shanxi, and Henan provinces) reduces from -8% to -5% (11% to 6%).

5.2 Introduction

Fine particulate matter, particles with an aerodynamic diameter $\leq$ 2.5 μm ($PM_{2.5}$), have well-documented deleterious impacts on human health (Lelieveld et al., 2015; Burnett et al., 2018), and under certain environment conditions can result in cloud condensation nuclei that subsequently alter cloud optical properties that perturb Earth's radiative balance (Stephens, 2005; Seinfeld et al., 2016). Current knowledge of ground-level $PM_{2.5}$ and its chemical constituents is limited to *in situ* instruments that are typically located in regional clusters across the globe, with generally fewer calibrated sensors available in less developed countries. To address this measurement gap, the science community has developed progressively more sophisticated methods to relate satellite remote sensing observations of aerosol optical depth (AOD), an integrated measure of aerosol extinction of light as it passes through the atmosphere, to ground-level $PM_{2.5}$ concentrations.

Translating satellite-retrieved AOD into estimates of ground-level $PM_{2.5}$ is non-trivial. Previous studies have used a wide range of process-driven and data-driven models to determine $PM_{2.5}$:AOD relationships (Figure 5.3). Process-driven models, such the GEOS-Chem global 3-D model of atmospheric chemistry and transport, take into account emission distributions, atmospheric chemistry and transport to simulate self-consistent $PM_{2.5}$:AOD (Van Donkelaar et al., 2010, 2015; Boys et al., 2014; Hammer et al., 2020). However, sources of model error, e.g. vertical mixing of surface emissions and heterogeneous chemistry, will introduce uncertainties in the corresponding ground-level $PM_{2.5}$ estimates. The alternative is to develop data-driven $PM_{2.5}$:AOD relationships using statistical methods, e.g. the linear mixed effects model (Lee et al., 2011; Zhang et al., 2018), geographically weighted regression model (Hu et al., 2013; Ma et al., 2014; Guo et al., 2017; He and Huang, 2018*a*,*b*), and two-stage model (Hu et al., 2014; Yao et al., 2019), and machine learning, e.g. artificial neural networks (Gupta and Christopher, 2009;

Wu et al., 2012; Guo et al., 2013) and the random forest (RF) (Hu et al., 2017; Wei et al., 2019, 2020). While these methods show promise, they are sometimes difficult to interpret without a supporting in-depth investigation.

A major criticism of data-driven models is that they do not explicitly take into consideration that AOD and ground-level $PM_{2.5}$ are different measures of atmospheric aerosols. Ground-level $PM_{2.5}$ measurements are mass concentrations ($\mu g\ m^{-3}$) of fine particles measured under controlled relative humidity (RH) conditions: 35% in the US (U.S. Environmental Protection Agency, 1997), and 50% in Europe (European Committee for Standardization (CEN), 1998) and China (HJ 618-2011, 2011). In contrast, AOD (unitless) reflects a vertical integration of aerosol extinction coefficients (m^{-1}) at a specific wavelength (typically 550 nm). There are a range of factors that can influence AOD and/or ground-level $PM_{2.5}$ and consequently strengthen or weaken their relationship. For example, covariation of free tropospheric aerosols (e.g. desert dust) and ground-level $PM_{2.5}$ may result in a strong relationship between the two measurements without any physical connection.

In this study, we propose a model framework to infer ground-level $PM_{2.5}$ from satellite-retrieved AOD. Most importantly, the method identifies observed scenes for which aerosol extinction coefficients is mostly within the planetary boundary layer (PBL), where variations are mostly likely to reflect changes in ground-level $PM_{2.5}$. In the next section we describe the ground-level $PM_{2.5}$ and AOD measurements, the nested version of the GEOS-Chem global 3-D model of atmospheric chemistry and transport, and the statistical regression and machine learning models that we use to develop our $PM_{2.5}$:AOD relationships. In section 5.4, we report the results of our method and compare them with conventional methods. In section 5.5, we conclude the paper by reflecting on our results, outlining limitations to our approach and discussing future research directions.

5.3 Data and methods

5.3.1 Ground-level $PM_{2.5}$ measurements

We collect hourly ground-level $PM_{2.5}$ measurements for 2014 from the China National Environmental Monitoring Center (CNEMC, `http://www.cnemc.cn/`). The data are measured either by the tapered element oscillating microbalance method (TEOM) or the beta-attenuation method (HJ 618-2011, 2011). We remove values by following the first three steps of the *ad hoc* quality control protocol developed by Jiang et al. (2020). We limit our study area to eastern China, defined as 95–140°E and 15–55°N (Figure 5.9) due to limited ground-level $PM_{2.5}$ monitoring sites over western China. For our analysis we consider mean ground-level $PM_{2.5}$ measurements taken between 1000 and 1100 and between 1300 and 1400 local times, corresponding to the 1030 and 1330 local equatorial overpass times of Terra and Aqua satellites. These mean values are described on the GEOS-Chem model 0.25° latitude × 0.3125° longitude nested grid, described below.

5.3.2 MODIS MAIAC and AERONET AOD retrievals

We use data retrieved from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) instrument aboard the Terra and Aqua satellites. In particular, we use the MODIS Collection 6 Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD product available on a 1 km spatial resolution (Lyapustin et al., 2011, 2018) covering eastern China for 2014 downloaded from the Level-1 and Atmosphere Archive & Distribution System (`https://ladsweb.modaps.eosdis.nasa.gov/`). We use Terra and Aqua MAIAC AOD at 550 nm, corresponding to particle sizes of 0.1–2 μm and comparable to the size range of $PM_{2.5}$ (Kahn et al., 1998). To evaluate the MAIAC AOD data

over China, we compare values against AOD retrievals from the NASA AErosol RObotic NETwork (AERONET, version 3: `https://aeronet.gsfc.nasa.gov/`) as these ground-based data are generally recognized as a 'ground truth' for satellite observations of aerosol optical properties (Holben et al., 1998).

AERONET AOD data are available at three data quality levels: Level 1.0 (unscreened), Level 1.5 (cloud-screened and quality controlled), and Level 2.0 (quality-assured). Due to the small number of AERONET sites over China that provide continuous coverage, we use all the MAIAC and Level 2 AERONET AOD from 2000 to 2018 in China to conduct the evaluation. The MAIAC AOD are at 550 nm but AERONET do not provide AOD at 550 nm. Following (Eck et al., 1999; Martins et al., 2017), we interpolate AERONET AOD to 550 nm using quadratic fits on a log-log scale (equation 5.1).

$$lnAOD_{\lambda} = \beta_2(ln\lambda)^2 + \beta_1(ln\lambda) + \beta_0, \tag{5.1}$$

where λ is the wavelength typically including 380, 440, 500, 675, 870, and 1020 nm. AOD_{λ} is the AOD measured at wavelength λ. β_2, β_1, and β_0 are coefficients for corresponding variables.

To take into account of the differences in spatiotemporal resolutions of MAIAC and AERONET AOD, we calculate the mean of AERONET AOD in four temporal windows (i.e. 30 to 120 minutes with an interval of 30 minutes) centred at satellite overpass times to compare it with the mean of MAIAC AOD in 25 spatial windows (i.e. 5 to 125 km with an interval of 5 km) centred at each AERONET site. The calculation of the mean of AERONET and MAIAC AOD are inversely weighted by the temporal distance to satellite overpass times and spatial distance to each AERONET site, respectively. In particular, we calculate the Pearson correlation coefficient R and the mean percentage error $MPE = \frac{1}{N}\sum(\frac{\overline{AERONET} - \overline{MAIAC}}{\overline{AERONET}})$, where $\overline{AERONET}$ and $\overline{MAIAC}$ denote the mean of AERONET and MAIAC

AOD collected in a certain spatiotemporal window, respectively, and N is the sample size.

Figure 5.1 illustrates how these statistics change with different spatiotemporal windows. Terra and Aqua data with and without quality assurance (QA) flags (i.e. QA for AOD = '0000' denotes best data) filtering are analyzed separately to account for their potential difference. We find that a temporal window of 60 minutes retains a large sample size without compromising on the quality of the agreement. Similarly, we find that a spatial window of 25 km is a trade-off between data sample size and minimizing the MPE. Therefore, we determine the spatiotemporal window corresponding to 60 minutes and 25 km for matching MAIAC and AERONET AOD. This spatiotemporal window is also within the mesoscale aerosol homogeneity (specifically, for horizontal scales of 40-400 km and temporal scales of 2-48 h) found in Anderson et al. (2003).

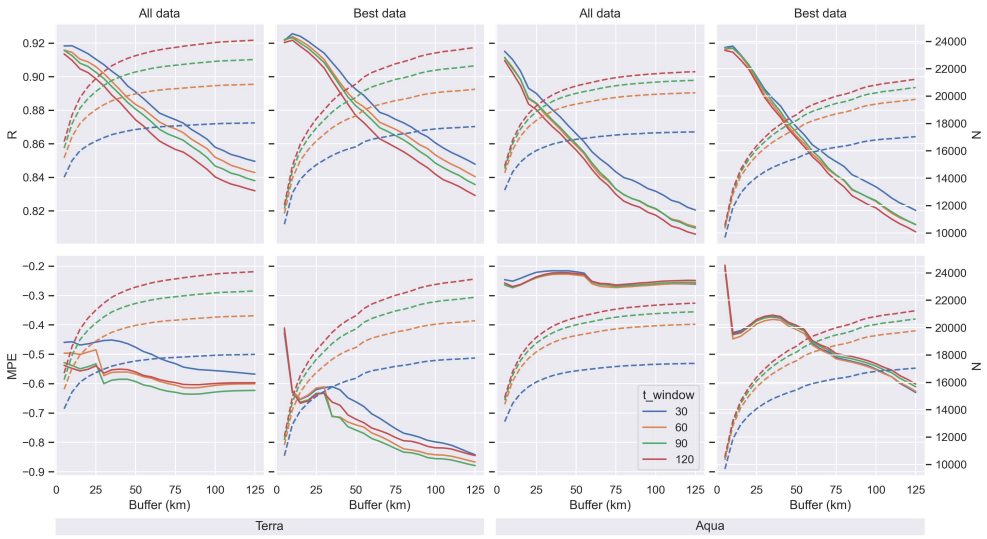


Figure 5.1: Spatiotemporal window determination for matching MAIAC and AERONET AOD from 2000 to 2018 in China. Solid lines denote the change of statistics and dashed lines denote the change of the sample size.

Figure 5.2 further shows the evaluation results in more details using the spatiotemporal window determined above. The Terra MODIS MAIAC AOD without QA

filtering have 19,008 match-ups with the AERONET AOD, with the R and the MPE equivalent to 0.91 and -0.48, respectively, while those by QA filtering have 16,709 match-ups, with the R and MPE equivalent to 0.91 and -0.6, respectively. This suggests that the accuracy of the Terra MODIS MAIAC AOD is insensitive to their QA flags over China. We find a similar result for Aqua. This allows us to use the complete MAIAC dataset. We regrid the 1 km Terra and Aqua MAIAC AOD data product to the coarser GEOS-Chem model 0.25° latitude × 0.3125° longitude nested grid.

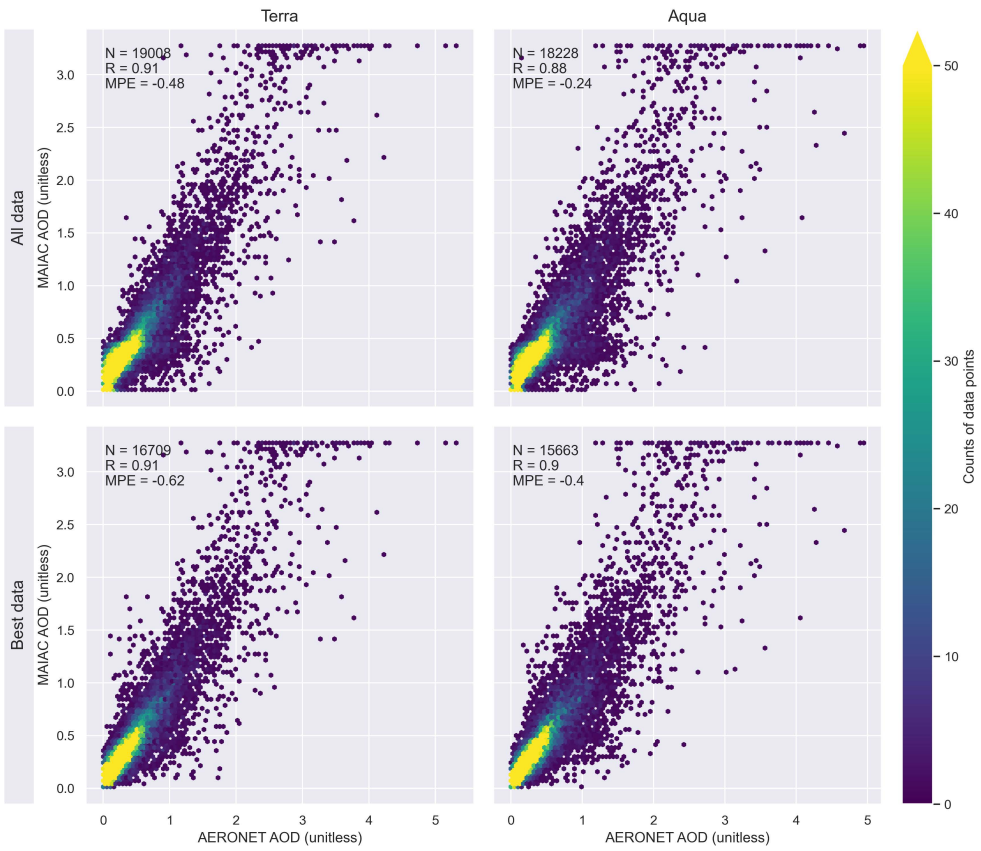


Figure 5.2: Density scatterplots of AERONET and (left) Terra and (right) Aqua MAIAC AOD from 2000 to 2018 in China (bottom) with and (top) without quality assurance flags filtering under the spatiotemporal window of 60 minutes and 25 km.

5.3.3 Framework

Figure 5.3 describes the computational approach we propose to infer ground-level $PM_{2.5}$ concentrations from MODIS MAIAC AOD retrievals. The approach is described by four broad steps: 1) data clustering to identify spatially proximate areas where ground-level $PM_{2.5}$ and AOD behave in a coherent manner; 2) data suitability to retain data where a certain fraction of the columnar AOD is found in the PBL, which we assume is well-mixed and consequently reflects variations of ground-level $PM_{2.5}$; 3) statistical and machine learning data-driven models of the relationship between ground-level $PM_{2.5}$ and AOD, which is subsequently used to (step 4) infer ground-level $PM_{2.5}$ concentrations from satellite observations of AOD. Below we further describe each step in our approach and clearly outline any assumptions we have made.

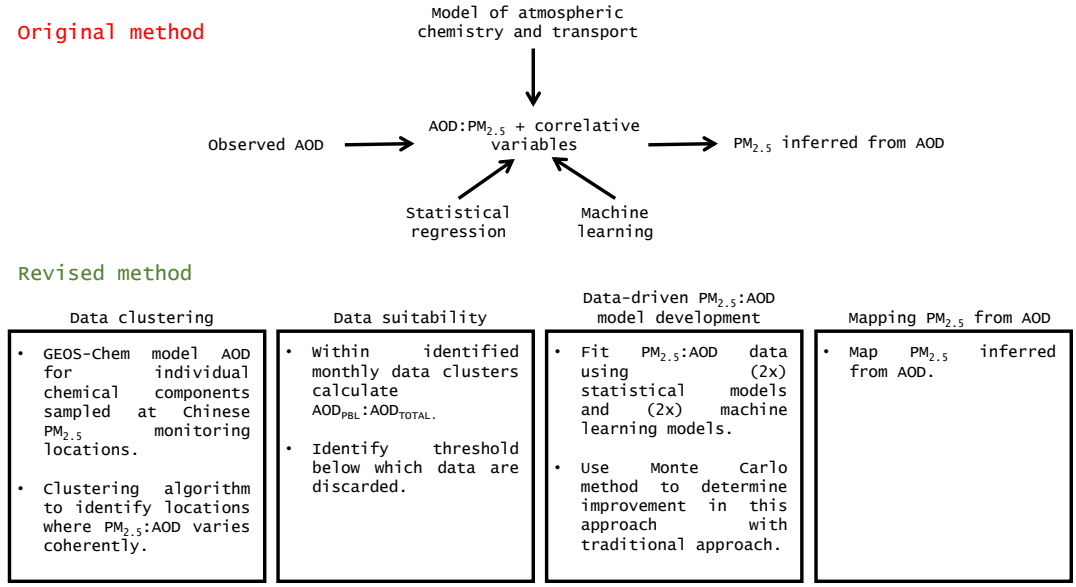


Figure 5.3: A brief description of the original method of inferring ground-level $PM_{2.5}$ concentrations from satellite observations of AOD, and the revised method that has developed in this study and most importantly takes into account the fraction of columnar AOD that resides in the planetary boundary layer.

5.3.4 GEOS-Chem Model of Atmospheric Chemistry and Transport

We use v12.5.0 of the GEOS-Chem global 3-D model of atmospheric chemistry and transport (Bey et al., 2001) to relate emissions and atmospheric chemistry to hourly ground-level $PM_{2.5}$ and 3-D fields of aerosol extinction coefficients and its individual chemical components over eastern China during 2014 on a 0.25° latitude × 0.3125° longitude nested grid.

We use the 2° latitude × 2.5° longitude global version of the GEOS-Chem model to provide lateral, time-dependent boundary conditions for the nested simulation (Wang et al., 2004). We define a 3-box "buffer zone" to incorporate spurious edge effects from the lateral boundary conditions for the nested simulation, and exclude data from this zone from further scientific analysis. We spin-up both models from initial conditions, provided by a longer GEOS-Chem model run, for three months prior to the start of our study year 2014 to minimize the impact of initial conditions. The atmospheric transport and chemistry time steps are 10 mins and 20 mins, respectively, for the global-scale model, and 5 mins and 10 mins, respectively, for the nested model (Philip et al., 2016).

We drive both models by the Goddard Earth Observing System-Forward Processing (GEOS-FP) meteorological analyses (Lucchesi, 2018), provided by the Global Modeling and Assimilation Office at the NASA Goddard Space Flight Center. We use 47 vertical terrain-following sigma levels that describe the atmosphere from the surface to 0.01 hPa, of which about 30 are typically below the dynamic tropopause. The global model uses the same meteorological data but as a spatial resolution of 2° latitude × 2.5° longitude.

The configuration for emission inventories follows that of previous chapters except that we substitute the global anthropogenic emissions with regional ones, which

includes the Multi-resolution Emission Inventory for China (Zheng et al., 2018), and the MIX Asian emission inventory for surrounding Asian regions beyond China (Li, Zhang, Kurokawa, Woo, He, Lu, Ohara, Song, Streets, Carmichael et al., 2017).

We use the NO_x-O_x-hydrocarbon-aerosol-bromine-chlorine-iodine mechanism with complex secondary organic aerosol (SOA) and semi-volatile primary organic aerosol (POA) applied from the surface to the tropopause. We apply the spatially and seasonally resolved ratios of organic matter to organic carbon to calculate organic aerosols from the simulated organic carbon (Philip et al., 2014). We follow the method detailed in section 1.2.4 to calculate $PM_{2.5}$ concentrations under the 50% RH, 298 K, and 1013.25 hPa condition and select the values of the lowest model layer to correspond to ground-level $PM_{2.5}$ measurements described as above. 3-D fields of aerosol extinction coefficients and its individual chemical components (sulfate – SO4, nitrate – NIT, ammonium – NH4, POA, SOA, black carbon – BC, sea salt – SS, and dust) are calculated at 550 nm using relative-humidity-dependent aerosol optical properties (Latimer and Martin, 2019) and dust optics (Ridley et al., 2012).

Due to the updates to selection of input emission inventories and chemical mechanisms, we are required to re-evaluate the performance of the GEOS-Chem model. Throughout this chapter, we sample at the time and location of ground-based and satellite observations. To evaluate the model, we compare ground-level $PM_{2.5}$ measurements with $PM_{2.5}$ values from the lowest model layer. We compare MODIS MAIAC AOD retrievals with the simulated AOD from all model layers. To describe these comparisons we continue to use the Pearson correlation coefficient R and the mean percentage error $MPE = \frac{1}{N}\sum(\frac{O-M}{O})$, where O and M denote observations and model values, respectively, and N denote the number of comparison points.

The temporal correlation between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values are 0.74 ± 0.12 (median ± median absolute deviation (MAD), similarly hereinafter) and 0.5 ± 0.18, respectively (Figure 5.4). The temporal MPE between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values are -0.01 ± 0.28 and -0.16 ± 0.3, respectively (Figure 5.5). The spatial correlation between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values are 0.6 ± 0.14 and 0.55 ± 0.18, respectively (Figure 5.6). The spatial MPE between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values are 0.11 ± 0.23 and -0.07 ± 0.3, respectively (Figure 5.7). According to these statistics, we believe that the GEOS-Chem generally reproduces the observed variations of AOD and ground-level $PM_{2.5}$, providing confidence in use of this model in our study.

5.3.5 Data clustering

The relationship between ground-level $PM_{2.5}$ and AOD varies spatially (Xiao et al., 2018; Yao et al., 2019), so we develop spatially adaptive models to infer ground-level $PM_{2.5}$ concentrations from satellite observations of AOD. We assume that ground-level $PM_{2.5}$ and AOD behave in a coherent manner over spatially proximate areas where individual chemical components of AOD vary similarly. We identify these areas using the unweighted pair group method with arithmetic mean (UPGMA) hierarchical clustering method. The UPGMA hierarchical clustering method includes three steps:

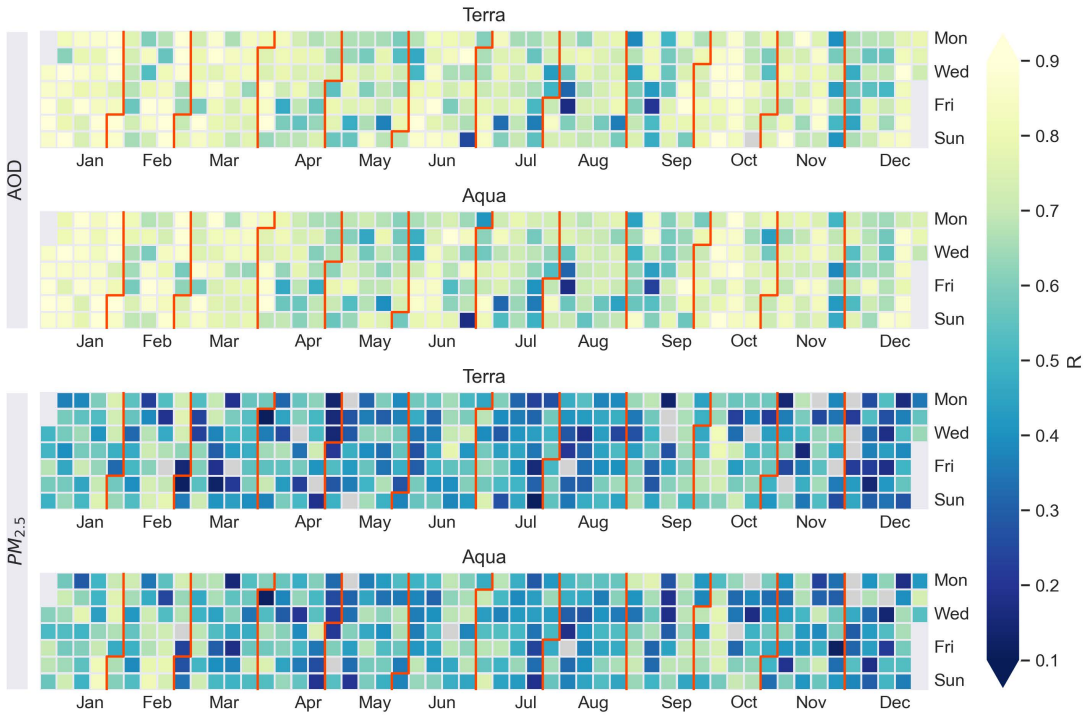


Figure 5.4: Temporal Pearson correlation coefficient (R) (top panel) between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and (bottom panel) between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values using spatially collocated data on a daily basis at the (rows 1 and 3) Terra and (rows 2 and 4) Aqua overpass times in 2014 over eastern China. Gray pixels denote statistically insignificant R values at $\alpha = 0.05$.

Step 1 Calculate the correlation distance between any two ground-level $PM_{2.5}$ monitoring locations using equation 5.2. The distance between two locations will be related to the extent to which we expect agreement of individual chemical components of AOD at these two locations. Repeating the calculation for all combinations of ground-level $PM_{2.5}$ monitoring locations returns a distance matrix:

$$D_{ij} = \frac{1}{N}\sum_{k=1}^{N}(1 - R_{ijk}), \tag{5.2}$$

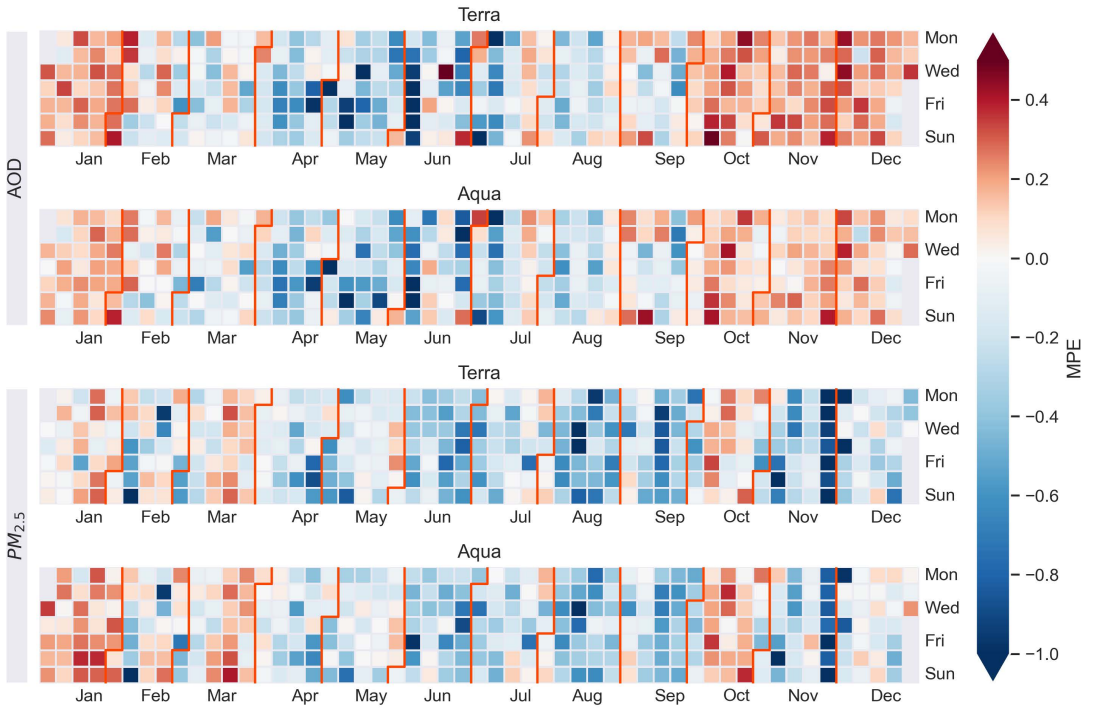


Figure 5.5: Temporal mean percentage error (top panel) between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and (bottom panel) between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values using spatially collocated data on a daily basis at the (rows 1 and 3) Terra and (rows 2 and 4) Aqua overpass times in 2014 over eastern China.

where R_{ijk} denotes the Pearson correlation between ground-level $PM_{2.5}$ monitoring location i and j for kth chemical component of AOD using a certain period (e.g. a month and a year in this study) of time series data. N is the total number of individual chemical components of AOD, which in this study is eight. The distance D will vary between 0 and 2.

Step 2 Starting from the initial distance matrix formed in Step 1, in which all ground-level $PM_{2.5}$ monitoring locations are thought as a single cluster, we first search the pair that has the minimum distance. Supposing they are A and B, we then merge them into a new cluster called $A \cup B$. Next, we calculate the

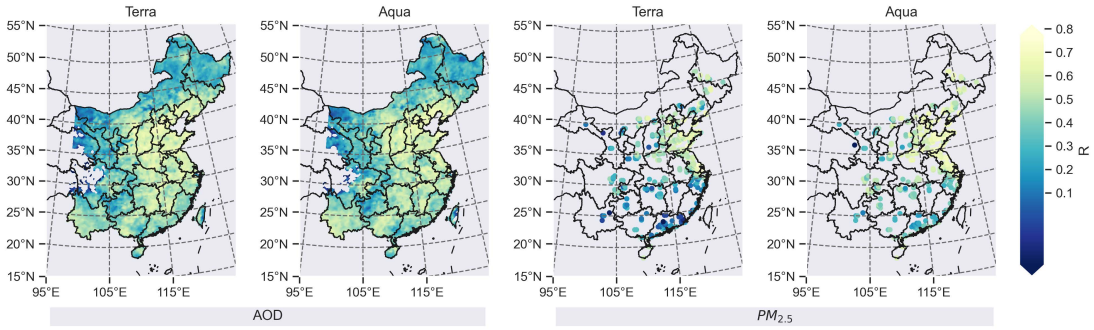


Figure 5.6: Spatial Pearson correlation coefficient (R) (left panel) between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and (right panel) between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values using daily collocated data on a spatial basis at the (columns 1 and 3) Terra and (columns 2 and 4) Aqua overpass times in 2014 over eastern China. Gray pixels denote statistically insignificant R values at $\alpha = 0.05$.

distances between $A \cup B$ and the remaining clusters by equation 5.3. Hence we reduce the distance matrix by one row and one column. We repeat this process until the distance matrix has been downgraded to a simple scalar, so that the original ground-level $PM_{2.5}$ monitoring locations are all merged and we obtain a dendrogram:

$$D_{A \cup B,X} = \frac{|A| \cdot D_{A,X} + |B| \cdot D_{B,X}}{|A| + |B|}, \tag{5.3}$$

where X denotes a remaining cluster, $D_{A,X}$ and $D_{B,X}$ denotes its distance from A and B. $|A|$ and $|B|$ denotes the total number of original ground-level $PM_{2.5}$ monitoring locations that cluster A and B hold. The updated distance is then with respect to initial distances rather than mathematical procedure, i.e. why the method is called unweighted.

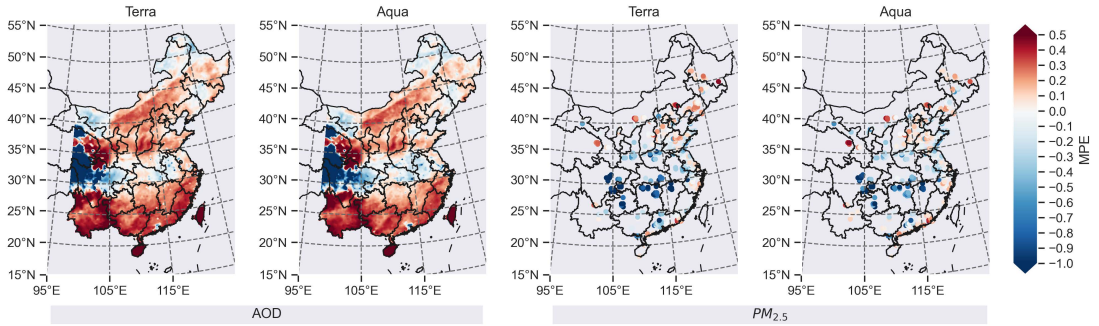


Figure 5.7: Spatial mean percentage error (left panel) between the MODIS MAIAC AOD retrievals and the GEOS-Chem model values and (right panel) between the ground-level $PM_{2.5}$ measurements and the GEOS-Chem model values using daily collocated data on a spatial basis at the (columns 1 and 3) Terra and (columns 2 and 4) Aqua overpass times in 2014 over eastern China.

Step 3 A clear relationship exists between the distance threshold chosen for the dendrogram and the number of clusters emerged. The value of the distance threshold is pivotal. The number of spatial clusters would be very small if the distance threshold is set too high. Consequently, their areas would be very large, hiding the spatial heterogeneity of the $PM_{2.5}$:AOD relationship. The number of spatial clusters would be very large if the distance threshold is set too low. As a result, their areas and the size of the model-fitting dataset would be very small, decreasing the reliability of the follow-on statistical and machine learning modelling. We test a series of values and finally choose 0.5 as the distance threshold to make a good balance between the number of spatial clusters and the size of the model-fitting dataset in each spatial cluster. In this sense, the mean similarity of variations of individual chemical components of AOD within the same cluster is no less than 0.5. We finally generate Thiessen polygons (Thiessen, 1911) from the ground-level $PM_{2.5}$ monitoring locations and assign the GEOS-Chem model 0.25° latitude × 0.3125° longitude nested grid cells within each Thiessen polygon to the same cluster of the corresponding ground-level $PM_{2.5}$ monitoring location in the center.

We conduct the data clustering on a monthly and an annual scale. We identify data for which we expect the strongest physical relationship between ground-level $PM_{2.5}$ and AOD by combining the monthly data clusters with simple correlation techniques (section 5.4.2). We develop robust $PM_{2.5}$:AOD models for geographical regions represented by the clustered ground-level $PM_{2.5}$ monitoring locations by combining the annual data clusters with sophisticated statistical and machine learning models (section 5.4.3).

5.3.6 Data suitability

Intuitively, we expect the strongest physical relationship between ground-level $PM_{2.5}$ and total columnar AOD when the majority of the contributing aerosol extinction coefficients resides in the PBL that is connected to aerosols at the surface. First, we define the fraction of total columnar AOD that resides in the PBL (Γ_{PBL}^{AOD}) using the 3-D fields of aerosol extinction coefficients simulated from the GEOS-Chem model. We then explore the robustness of the $PM_{2.5}$:AOD relationship and how it varies with Γ_{PBL}^{AOD}. We quantify robustness using the Pearson and Spearman rank correlation coefficients, acknowledging that the $PM_{2.5}$:AOD may vary monotonically without necessarily being linear. By examining how these correlations vary with Γ_{PBL}^{AOD} for each monthly data cluster we identify a threshold value for Γ_{PBL}^{AOD} that ensures a robust $PM_{2.5}$:AOD relationship that is determined by a large number of observations. By using this approach, we identify suitable data that provide the strongest constraints on our data-driven models that we describe below.

5.3.7 Data-driven model development

We use two statistical methods and two machine learning methods to model the $PM_{2.5}$:AOD relationship for each annual data cluster with a range of auxiliary meteorological predictors. The meteorological predictors include PBL height (PBLH), mean RH in the PBL (RH_PBL), surface skin temperature (TS), total precipitation (PRECTOT), 10-metre eastward wind (U10M), 10-metre northward wind (V10M), and sea level pressure (SLP). Mean values of these fields from 1000-1100 and 1300-1400 local times, corresponding to Terra and Aqua overpass times, are taken from the GEOS-FP meteorological data (Lucchesi, 2018).

We first construct a pooled ordinary least square (PooledOLS) model (equation 5.4):

$$\begin{aligned} {PM_{2.5}}_g^d = \quad & \beta_0 + \beta_{AOD} AOD_g^d + \beta_{PBLH} PBLH_g^d \\ & + \beta_{RH\_PBL} RH\_PBL_g^d + \beta_{TS} TS_g^d \\ & + \beta_{PRECTOT} PRECTOT_g^d + \beta_{U10M} U10M_g^d \\ & + \beta_{V10M} V10M_g^d + \beta_{SLP} SLP_g^d + \epsilon_g^d, \end{aligned} \tag{5.4}$$

where ${PM_{2.5}}_g^d$, AOD_g^d, $PBLH_g^d$, $RH\_PBL_g^d$, TS_g^d, $PRECTOT_g^d$, $U10M_g^d$, $V10M_g^d$, SLP_g^d represent the ground-level $PM_{2.5}$, MAIAC AOD, PBLH, RH_PBL, TS, PRECTOT, U10M, V10M, SLP, respectively, at grid g on day d during Terra or Aqua overpass time; β_{AOD}, β_{PBLH}, β_{RH_PBL}, β_{TS}, $\beta_{PRECTOT}$, β_{U10M}, β_{V10M}, β_{SLP} denote the corresponding coefficients; β_0 is the intercept; and ϵ_g^d is the error term at grid g on day d during Terra or Aqua overpass time.

We then consider a time-fixed-effects model (TFEM), equivalent to equation 5.4 but with an additional day-specific offset β_d, that allows us to capture temporal

heterogeneities in the $PM_{2.5}$:AOD relationship (Yao et al., 2018). We choose to fix time in the fixed-effects model because our data analysis suggests that aerosols are more variable in the time domain than in the space domain (Figures 5.8 versus 5.9).

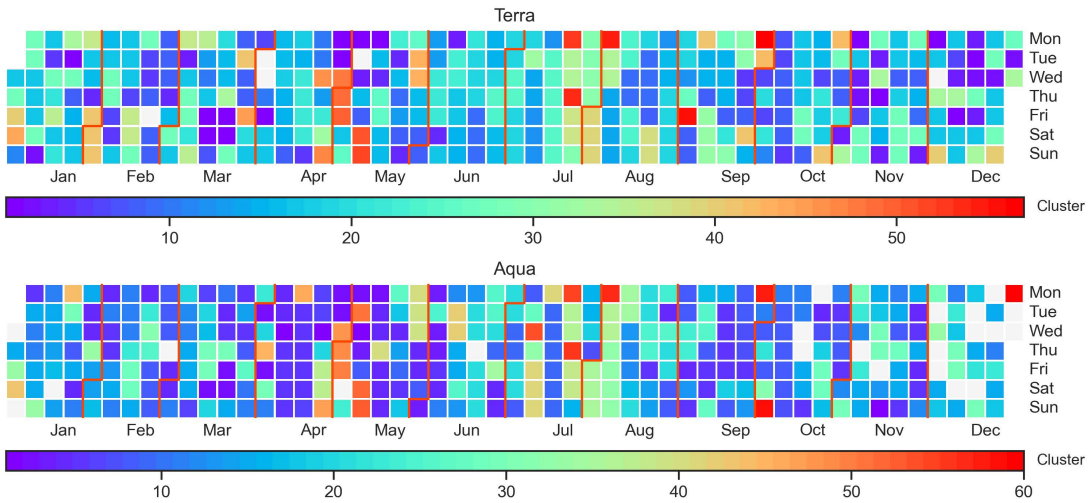


Figure 5.8: Temporal clusters determined by spatially collocated individual chemical components of AOD simulated from the GEOS-Chem model using the UPGMA hierarchical clustering method at the (top) Terra and (bottom) Aqua overpass times in 2014 over eastern China. That is we calculate the correlation distance between two day of years using their spatially collocated data in step 1 of the UPGMA hierarchical clustering method.

We use a RF model as an alternative approach to describe the relationship between ground-level $PM_{2.5}$ and AOD. RF is an ensemble of decision trees, with each tree constructed using bootstrap samples drawn from the original data (Breiman, 2001). For our study, the final prediction is the mean of predictions from all individual decision trees. Unlike statistical regression models, such as the PooledOLS model, RF typically involves fewer, less restrictive assumptions regarding independence of predictors and distributions of outcomes, with no *a priori* requirement of a formula between the predictors and the outcome variables.

We generate two RF models (RF1: equation 5.5 and RF2: equation 5.6) that correspond to the PooledOLS and TFEM regression models, respectively:

$$\begin{aligned} PM_{2.5}{}_g^d = \quad & f(AOD_g^d, PBLH_g^d, RH\_PBL_g^d, TS_g^d, \\ & PRECTOT_g^d, U10M_g^d, V10M_g^d, SLP_g^d) \qquad (5.5) \\ = \quad & f(AOD_g^d, PBLH_g^d, RH\_PBL_g^d, TS_g^d, \\ & PRECTOT_g^d, U10M_g^d, V10M_g^d, SLP_g^d, DOY_g^d), \qquad (5.6) \end{aligned}$$

where $f(\cdot)$ denotes a function representing the RF, DOY_g^d denotes the day of year at grid g on day d during Terra or Aqua overpass time, and the remaining variables are as previously defined. Each RF model is determined by using a 1,000 fully grown and unpruned decision trees.

We use both the full data (ignoring the data suitability step described above) and the suitable data (taking into consideration Γ_{PBL}^{AOD}) to train the statistical and machine learning models. We fit a linear regression between predicted and measured ground-level $PM_{2.5}$. We use the coefficient of determination (R^2) of the linear regression and the mean percentage error $MPE = \frac{1}{N}\sum(\frac{PM_{2.5obs} - PM_{2.5pred}}{PM_{2.5obs}})$, where $PM_{2.5obs}$ and $PM_{2.5pred}$ are observed and predicted $PM_{2.5}$, respectively, and N is the number of comparison points, to evaluate the goodness of model fitting (MF). Positive (negative) MPE values denote negative (positive) model bias.

5.3.8 Data-driven model validation and comparison

We conduct a sample-based 10-fold cross validation (CV) to evaluate model performance on yet unseen data. First, we split randomly and evenly the full/suitable data into 10 subsets. For each subset, we then calculate predictions

using the model trained from the remaining nine subsets. Next, we fit a linear regression between the CV predictions and observations and report the R^2 and the MPE. These CV statistics reflect the model performance on yet unseen data. The difference between MF and CV statistics can help identify whether a model under- or over-fits the data.

By comparing the MF and CV statistics of models trained by the full data and by the suitable data we can evaluate the comparative model improvement. We use a Monte Carlo approach to determine whether the potential model improvement is statistically significant. First, we randomly select a subset of the full data that matches the length of the suitable data and we repeat that process 1,000 times. Using each selected data sample, we build the statistical and machine learning models and perform the sample-based 10-fold CV. We subsequently obtain a series of MF and CV statistics, based on which we calculate the possibility (p value) of obtaining a model that performs no worse than our approach but has no skill of AOD vertical partitioning.

5.3.9 Ground-level $PM_{2.5}$ mapping

Finally, we use the statistical and machine learning models to predict ground-level $PM_{2.5}$ concentrations during Terra and Aqua overpass times on a daily basis across eastern China at the GEOS-Chem nested model horizontal resolution of 0.25° latitude $\times$ 0.3125° longitude. We use both the models trained by the full and the suitable data for ground-level $PM_{2.5}$ prediction. For the latter, we only predict at locations and times where Γ_{PBL}^{AOD} is no less than the determined threshold. We further compile daily ground-level $PM_{2.5}$ estimates to seasonal values, based on which we make another comparison between models trained by the full and the suitable data.

5.4 Results

5.4.1 Results of data clustering

Figure 5.9 shows the data clustering results based on one year of individual chemical components of AOD simulated from the GEOS-Chem model using the UPGMA hierarchical clustering method, described above. We determine a total of 13 spatial clusters with similar extent across China. Generally, we find little difference between results using model values taken from the 1030 and 1330 Terra and Aqua overpass times.

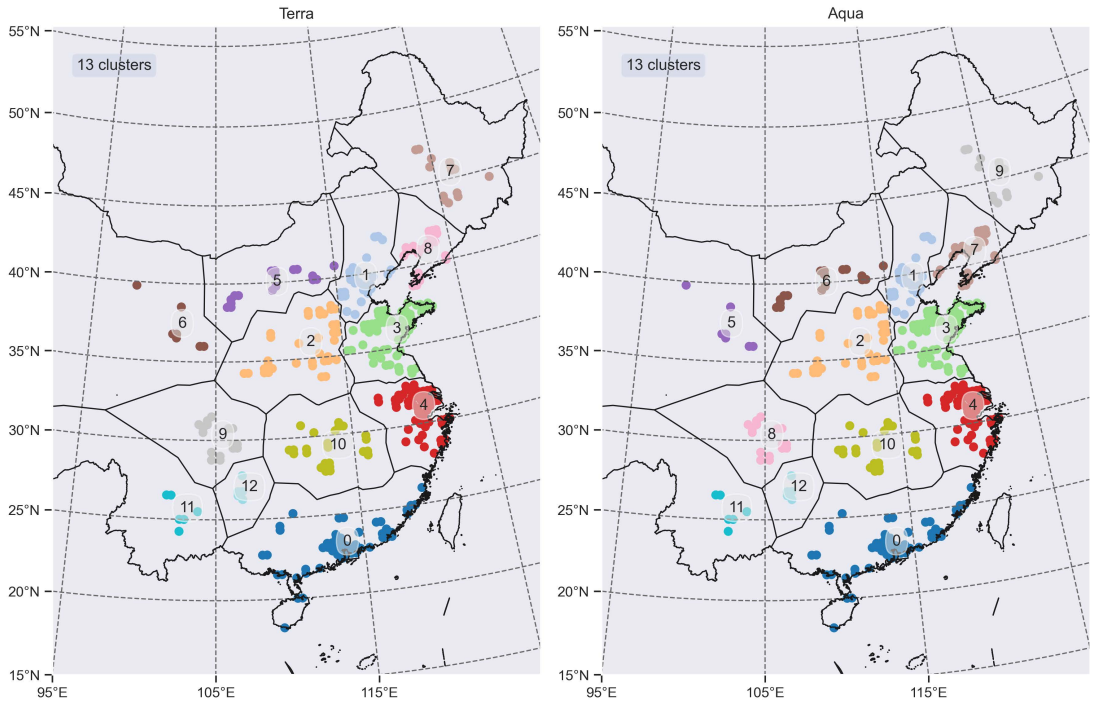


Figure 5.9: Spatial clusters of ground-level $PM_{2.5}$ monitoring locations over eastern China during 2014 for (left) Terra and (right) Aqua overpass times. Clusters are determined by one year of GEOS-Chem 3-D model fields of individual chemical components of AOD using the UPGMA hierarchical clustering method.

We find that the majority of these spatial clusters have boundaries similar to those

of urban agglomerations, e.g. Pearl River Delta urban agglomeration (cluster 0), Yangtze River Delta urban agglomeration (cluster 4), and the Yangtze River Mid-Reaches urban agglomeration (cluster 10). This, to some extent, suggests remarkable influence of anthropogenic activities on aerosols. We also find that that most clusters are dominated by secondary aerosols except cluster 6 for Terra and cluster 5 for Aqua overpass time, where dust aerosol is prevalent (Table 5.1). Moreover, these dust aerosols are frequently found in the free troposphere, e.g. the median ± median absolute deviation of the fraction of dust aerosols above the PBL at ground-level $PM_{2.5}$ monitoring locations in cluster 6 for Terra and cluster 5 for Aqua overpass time are 61 ± 21 % and 44 ± 21%, respectively.

Table 5.1: The fraction of individual chemical componenets to the annual AOD of 2014 in each cluster of eastern China derived from the GEOS-Chem model.

Terra									Aqua								
Cluster	SO_4	NIT	NH_4	POA	SOA	BC	DUST	SS	Cluster	SO_4	NIT	NH_4	POA	SOA	BC	DUST	SS
0	0.22	0.24	0.15	0.03	0.21	0.03	0.08	0.04	0	0.22	0.23	0.14	0.02	0.23	0.03	0.08	0.04
1	0.19	0.29	0.14	0.04	0.18	0.03	0.13	0.01	1	0.19	0.27	0.14	0.03	0.19	0.04	0.13	0.01
2	0.19	0.30	0.16	0.03	0.14	0.04	0.14	0.00	2	0.20	0.29	0.16	0.02	0.15	0.04	0.14	0.00
3	0.17	0.37	0.17	0.03	0.13	0.04	0.09	0.01	3	0.18	0.36	0.17	0.02	0.14	0.04	0.09	0.01
4	0.20	0.31	0.16	0.03	0.16	0.04	0.09	0.02	4	0.21	0.31	0.16	0.02	0.17	0.04	0.08	0.02
5	0.21	0.18	0.12	0.02	0.16	0.03	0.29	0.00	5	0.16	0.15	0.10	0.01	0.12	0.02	0.45	0.00
6	0.16	0.15	0.09	0.01	0.11	0.02	0.46	0.00	6	0.22	0.18	0.12	0.01	0.16	0.03	0.28	0.00
7	0.19	0.19	0.11	0.07	0.28	0.03	0.13	0.01	7	0.19	0.30	0.15	0.03	0.20	0.04	0.10	0.01
8	0.18	0.30	0.15	0.03	0.19	0.03	0.10	0.01	8	0.24	0.27	0.17	0.03	0.16	0.03	0.10	0.00
9	0.24	0.28	0.17	0.03	0.15	0.03	0.10	0.00	9	0.19	0.18	0.11	0.06	0.29	0.03	0.13	0.01
10	0.21	0.32	0.17	0.03	0.16	0.04	0.07	0.01	10	0.21	0.31	0.17	0.03	0.17	0.04	0.07	0.01
11	0.23	0.16	0.13	0.05	0.28	0.03	0.11	0.01	11	0.24	0.14	0.12	0.03	0.30	0.03	0.12	0.01
12	0.27	0.26	0.17	0.03	0.17	0.04	0.07	0.00	12	0.28	0.23	0.17	0.02	0.19	0.04	0.06	0.00

Figure 5.10 extends our data clustering analysis to monthly scales. Generally we find, irrespective of the local overpass times, more spatial clusters in warmer months (e.g. April-September) than during colder months (e.g. January-March, October-December), suggesting a wider spatial association of aerosols during colder months, e.g. from domestic heating.

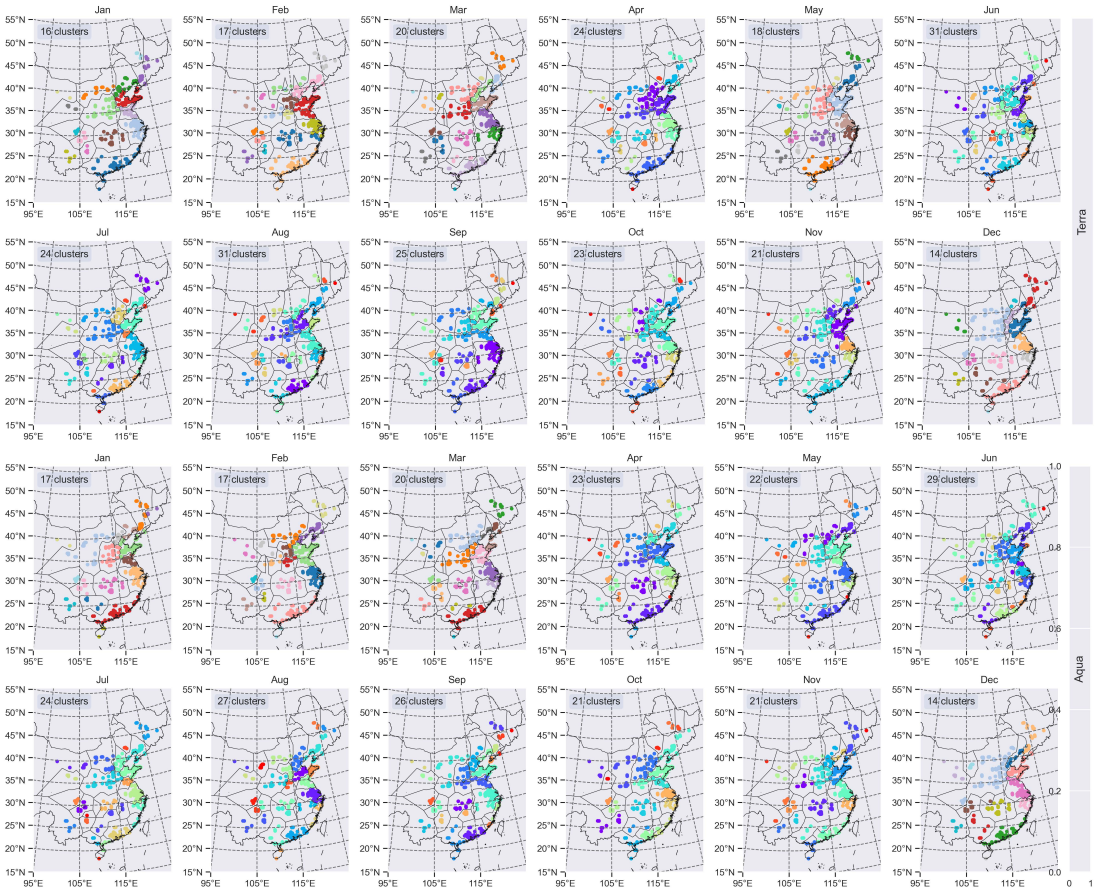


Figure 5.10: Spatial clusters of ground-level $PM_{2.5}$ monitoring locations over eastern China in each month of 2014 for (top panel) Terra and (bottom panel) Aqua overpass times. Clusters are determined by monthly GEOS-Chem 3-D model fields of individual chemical components of AOD using the UPGMA hierarchical clustering method. Number shown inset denote how many clusters determined for a specific month and satellite overpass time.

5.4.2 Results of data suitability

For each cluster in each calendar month shown in Figure 5.10 we analyze how the $PM_{2.5}$:AOD relationship changes with Γ_{PBL}^{AOD}. Figure 5.11 shows an example analysis. We find that the Pearson and Spearman rank correlation coefficients

between ground-level $PM_{2.5}$ and AOD generally increase with increasing Γ_{PBL}^{AOD}, as expected, despite a progressively decreasing number of observations that results in some fluctuations. There are variations in our analysis between clusters and months, but Figure 5.11 illustrates how Γ_{PBL}^{AOD} can be used effectively to identify a subset of data that ensures a physically-meaningful $PM_{2.5}$-AOD relationship. Based on our analysis over all clusters and months, we use a Γ_{PBL}^{AOD} value of 0.5 as a threshold, below which we discard the corresponding data.

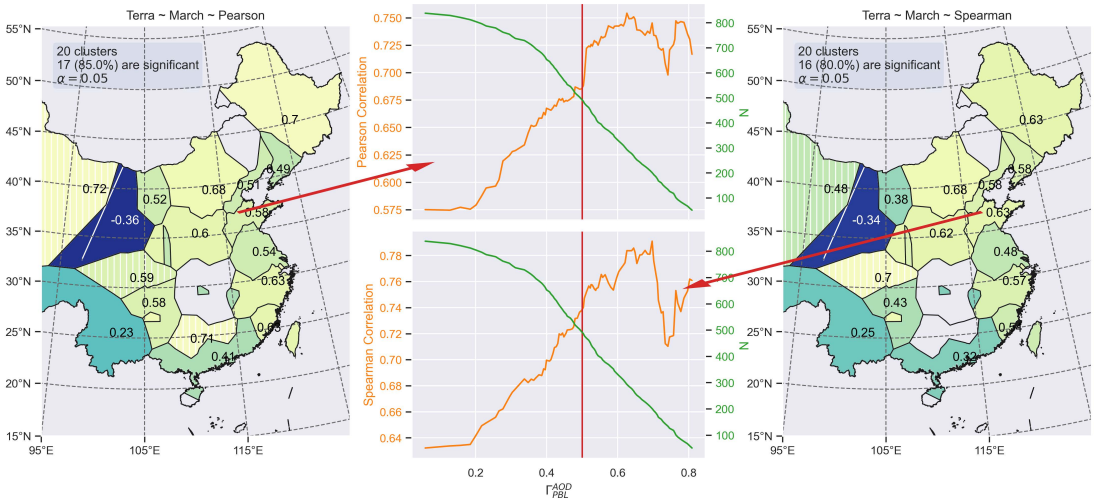


Figure 5.11: (Orange) Pearson and Spearman rank correlation coefficients between ground-level $PM_{2.5}$ and AOD, and the (green) corresponding number of observations, as a function of Γ_{PBL}^{AOD}, in a specific monthly cluster. We use a Γ_{PBL}^{AOD} value of 0.5 (red vertical lines) as a threshold, below which we discard the corresponding data.

5.4.3 Results of data-driven model development, validation, and comparison

Table 5.2 shows the overall MF and CV results of the PooledOLS, TFEM, RF1, and RF2 models. The difference between MF and CV statistics are smallest for the PooledOLS model and largest for the RF models. We interpret this as the PooledOLS model underfitting the data and the RF models overfitting the

data. The R^2 in the RF models are generally higher than those in the PooledOLS model and the TFEM model, suggesting the non-linearity in the $PM_{2.5}$:AOD relationship has been captured by the non-linear machine learning models. The models that account for day-to-day variations generally have a higher R^2 and a smaller absolute MPE. This is consistent with day-specific offsets in the TFEM model and the DOY predictor in the RF2 model having skill in reproducing temporal variations in the $PM_{2.5}$:AOD relationship. We generally find that models perform better for data collected during Aqua overpass time when the PBLH is close to its highest values (Chatfield et al., 2020). We also find that fewer data are discarded due to the $\Gamma_{PBL}^{AOD} \geq 0.5$ criterion at the Aqua overpass time ($\approx 16\%$) compared to the earlier Terra overpass time ($\approx 37\%$).

We find that the model trained by the suitable data generally has smaller absolute MPE values than the model using the full data, e.g. the model trained by the suitable data reduces the model bias using the full data by 10-15% and 9-12% for the Terra and Aqua overpass times, respectively, during cross validation. We find that the model trained by the suitable data generally has comparable or larger R^2 values than the model using the full data, e.g. the model trained by the suitable data improves the model R^2 using the full data by up to 8% and 5% for the Terra and Aqua overpass times, respectively, during cross validation. Moreover, we find the p values are all equivalent to zero, suggesting that the comparative model improvement is statistically significant.

For brevity, we focus on CV results in each cluster shown in Table 5.3. In terms of CV MPE, we find that most clusters obtain a statistically significant smaller absolute value. The number of these clusters are 11, 7, 10, and 10 in the PooledOLS, TFEM, RF1, and RF2 models, respectively, during the earlier Terra overpass time. Corresponding values for the Aqua overpass time are 11, 7, 11, and 9, respectively. In addition, the decrease of the absolute MPE values is spatially

Table 5.2: Overall model fitting and cross validation results of the PooledOLS, TFEM, RF1, and RF2 models in 2014 over eastern China. N, R^2, and MPE represent statistics of the model trained by the full data, while N$'$, $R^{2'}$, and MPE$'$ represent those of the model trained by the suitable data. ${R^2}_p$ and MPE_p represent the possibility of obtaining a model performance no worse than that of our approach by randomly selecting a subset of the full data that matches the length of the suitable table to train the model.

	Model	Terra								Aqua							
		N	N$'$	R^2	$R^{2'}$	${R^2}_p$	MPE	MPE$'$	MPE_p	N	N$'$	R^2	$R^{2'}$	${R^2}_p$	MPE	MPE$'$	MPE_p
MF	PooledOLS	57819	36692	0.37	0.39	0.0	-0.48	-0.41	0.0	55939	46961	0.43	0.45	0.0	-0.45	-0.41	0.0
	TFEM	57819	36692	0.63	0.67	0.0	-0.25	-0.21	0.0	55939	46961	0.69	0.71	0.0	-0.24	-0.21	0.0
	RF1	57819	36692	0.96	0.96	0.0	-0.12	-0.10	0.0	55939	46961	0.96	0.97	0.0	-0.11	-0.10	0.0
	RF2	57819	36692	0.97	0.96	0.0	-0.10	-0.09	0.0	55939	46961	0.97	0.97	0.0	-0.10	-0.09	0.0
CV	PooledOLS	57819	36692	0.36	0.39	0.0	-0.48	-0.41	0.0	55939	46961	0.43	0.45	0.0	-0.45	-0.41	0.0
	TFEM	57819	36692	0.58	0.58	0.0	-0.27	-0.24	0.0	55939	46961	0.64	0.66	0.0	-0.26	-0.23	0.0
	RF1	57819	36692	0.63	0.63	0.0	-0.32	-0.28	0.0	55939	46961	0.67	0.69	0.0	-0.31	-0.28	0.0
	RF2	57819	36692	0.68	0.66	0.0	-0.29	-0.26	0.0	55939	46961	0.73	0.73	0.0	-0.28	-0.25	0.0

homogeneous. For instance, we observe the biggest decrease in both clusters 3 and 4 dominated by anthropogenic aerosols and clusters 5 and 6 dominated by natural aerosols (Table 5.1). A few clusters show larger absolute MPE values but the increase are all limited to ≤ 0.03, a very small value. Different model performances across clusters under the same criterion, $\Gamma_{PBL}^{AOD} \geq 0.5$, reflects the importance of individual chemical components that comprise $PM_{2.5}$ and AOD (and thereby the data clustering step) to the development of robust $PM_{2.5}$:AOD models.

In terms of CV R^2, its behaviour is slightly different during Terra and Aqua overpass times. For the Aqua overpass time, we find that most clusters obtain a statistically significantly higher R^2. For the Terra overpass time, we admit that there is a smaller number of clusters showing statistically significant increased CV R^2. A few clusters present declined CV R^2 to a slightly large extent, suggesting that $\Gamma_{PBL}^{AOD} \geq 0.5$ might be too strict a criterion for the Terra overpass time from the perspective of capturing ground-level $PM_{2.5}$ concentrations. Nonetheless, we find that cluster 6 (5), corresponding to the Terra (Aqua) overpass time, where dust aerosol is prevalent (Table 5.1) generally shows the largest improvement in CV R^2 that is distinguishable from those in other clusters. This highlights

that using more AOD data blindly without considering their representativeness for ground-level $PM_{2.5}$ will not always lead to better estimates of ground-level $PM_{2.5}$ even if a sophistical machine learning model is used. Careful filtering of these data is required to build confidence and robustness of resulting estimates, and eventually scientific exploitation of satellite-derived ground-level $PM_{2.5}$ data products.

5.4.4 Ground-level $PM_{2.5}$ maps

For illustrative purposes, we show seasonal mean values inferred from RF2 using data from the Aqua overpass time. Figure 5.12 shows the ground-level $PM_{2.5}$ estimates inferred from the full data and from the subset of these data where $\Gamma_{PBL}^{AOD} \geq 0.5$. We find that the seasonal distributions of $PM_{2.5}$ inferred from the full and reduced AOD datasets are broadly similar. However, we find two areas characterized by large differences between these seasonal estimates.

The first is located around Qinghai and Gansu provinces (spatially corresponding to cluster 5) that is highly influenced by dust aerosols (Table 5.1). The larger values over these regions compared to the surrounding regions particularly in the autumn, inferred from the full data, are likely erroneous. This is because the spatial distribution of seasonal ground-level $PM_{2.5}$ should be relatively smooth without too many sudden changes. For the same regions, using the reduced data we find there is an obvious reduction of up to 20 $\mu g\ m^{-3}$, alleviating the extent to which the large anomaly occurs. Further, we calculate the MPE between these seasonal estimates and the seasonal observations compiled from daily ground-level $PM_{2.5}$ measurements. We find that the corresponding MPE over cluster 5 in the autumn goes from -8% to -5% when the seasonal estimates are inferred from the full to the reduced data. This indicates that the positive model bias occurs

Table 5.3: Cross validation results of the PooledOLS, TFEM, RF1, and RF2 models in each cluster of eastern China in 2014.

	Cluster	Terra								Aqua							
		N	N′	R^2	$R^{2'}$	$R^2{}_p$	MPE	MPE′	MPE_p	N	N′	R^2	$R^{2'}$	$R^2{}_p$	MPE	MPE′	MPE_p
PooledOLS	0	7857	5410	0.39	0.42	0.00	-0.37	-0.34	0.00	7776	6776	0.46	0.51	0.00	-0.30	-0.29	0.00
	1	6172	3458	0.34	0.33	0.73	-0.84	-0.77	0.00	5650	4716	0.46	0.47	0.07	-0.80	-0.77	0.02
	2	6312	4246	0.34	0.36	0.05	-0.46	-0.44	0.25	6039	5475	0.37	0.38	0.13	-0.50	-0.49	0.03
	3	10975	7202	0.30	0.34	0.00	-0.54	-0.42	0.00	10277	8388	0.35	0.38	0.00	-0.48	-0.40	0.00
	4	7672	5301	0.31	0.29	0.99	-0.31	-0.25	0.00	7243	6045	0.35	0.37	0.00	-0.29	-0.25	0.00
	5	4105	2020	0.24	0.24	0.42	-0.52	-0.40	0.00	1612	1313	0.31	0.35	0.01	-0.41	-0.35	0.00
	6	1603	756	0.16	0.22	0.00	-0.44	-0.37	0.00	3974	3145	0.30	0.35	0.00	-0.64	-0.54	0.00
	7	2278	1380	0.42	0.47	0.00	-0.44	-0.41	0.09	5376	3953	0.33	0.35	0.02	-0.49	-0.46	0.00
	8	5206	3006	0.24	0.29	0.00	-0.54	-0.50	0.00	1203	1097	0.47	0.51	0.00	-0.33	-0.28	0.00
	9	1189	667	0.40	0.38	0.76	-0.32	-0.27	0.02	2222	1757	0.58	0.65	0.00	-0.39	-0.36	0.02
	10	2920	1808	0.19	0.26	0.49	-0.36	-0.31	0.00	2861	2611	0.29	0.27	0.84	-0.32	-0.31	0.06
	11	1266	1233	0.28	0.30	0.00	-0.26	-0.25	0.00	1341	1331	0.28	0.28	0.10	-0.24	-0.24	0.76
	12	264	205	0.39	0.45	0.02	-0.37	-0.31	0.02	365	354	0.42	0.44	0.03	-0.36	-0.39	0.36
TFEM	0	7857	5410	0.63	0.63	0.36	-0.22	-0.22	0.02	7776	6776	0.63	0.65	0.00	-0.22	-0.20	0.00
	1	6172	3458	0.54	0.53	0.14	-0.48	-0.42	0.00	5650	4716	0.65	0.67	0.01	-0.40	-0.39	0.07
	2	6312	4246	0.51	0.53	0.00	-0.27	-0.27	0.33	6039	5475	0.56	0.59	0.00	-0.28	-0.26	0.00
	3	10975	7202	0.59	0.59	0.14	-0.26	-0.23	0.00	10277	8388	0.63	0.64	0.01	-0.24	-0.21	0.00
	4	7672	5301	0.60	0.59	0.44	-0.17	-0.14	0.00	7243	6045	0.67	0.69	0.00	-0.15	-0.13	0.00
	5	4105	2020	0.40	0.36	0.37	-0.38	-0.31	0.00	1612	1313	0.32	0.38	0.00	-0.32	-0.26	0.00
	6	1603	756	0.19	0.27	0.00	-0.33	-0.25	0.02	3974	3145	0.49	0.53	0.00	-0.45	-0.38	0.00
	7	2278	1380	0.63	0.49	1.00	-0.25	-0.23	0.11	5376	3953	0.59	0.68	0.00	-0.24	-0.22	0.00
	8	5206	3006	0.57	0.60	0.00	-0.26	-0.23	0.00	1203	1097	0.70	0.72	0.08	-0.15	-0.14	0.12
	9	1189	667	0.62	0.44	0.98	-0.17	-0.18	0.51	2222	1757	0.69	0.74	0.00	-0.25	-0.23	0.08
	10	2920	1808	0.57	0.48	0.99	-0.21	-0.18	0.00	2861	2611	0.60	0.60	0.06	-0.16	-0.17	0.43
	11	1266	1233	0.50	0.50	0.94	-0.13	-0.13	0.87	1341	1331	0.39	0.40	0.71	-0.14	-0.15	0.68
	12	264	205	0.44	0.34	0.47	-0.26	-0.27	0.50	365	354	0.56	0.58	0.13	-0.29	-0.25	0.24
RF1	0	7857	5410	0.67	0.68	0.00	-0.24	-0.22	0.00	7776	6776	0.65	0.67	0.01	-0.24	-0.22	0.00
	1	6172	3458	0.62	0.60	0.00	-0.53	-0.45	0.00	5650	4716	0.70	0.71	0.04	-0.49	-0.43	0.00
	2	6312	4246	0.55	0.57	0.00	-0.36	-0.35	0.33	6039	5475	0.62	0.62	0.01	-0.34	-0.31	0.00
	3	10975	7202	0.64	0.64	0.00	-0.30	-0.26	0.00	10277	8388	0.68	0.70	0.00	-0.26	-0.22	0.00
	4	7672	5301	0.64	0.63	0.09	-0.19	-0.16	0.00	7243	6045	0.68	0.71	0.00	-0.19	-0.17	0.00
	5	4105	2020	0.40	0.37	0.12	-0.44	-0.37	0.00	1612	1313	0.37	0.39	0.06	-0.39	-0.37	0.01
	6	1603	756	0.24	0.38	0.00	-0.40	-0.32	0.00	3974	3145	0.50	0.52	0.00	-0.56	-0.49	0.00
	7	2278	1380	0.68	0.63	0.54	-0.37	-0.35	0.00	5376	3953	0.59	0.62	0.00	-0.30	-0.26	0.00
	8	5206	3006	0.59	0.62	0.00	-0.32	-0.29	0.00	1203	1097	0.70	0.73	0.01	-0.24	-0.22	0.01
	9	1189	667	0.64	0.55	0.85	-0.23	-0.23	0.10	2222	1757	0.79	0.80	0.08	-0.35	-0.32	0.00
	10	2920	1808	0.56	0.43	1.00	-0.27	-0.25	0.01	2861	2611	0.61	0.59	0.87	-0.23	-0.22	0.04
	11	1266	1233	0.50	0.50	0.20	-0.21	-0.21	0.35	1341	1331	0.41	0.41	0.18	-0.22	-0.22	0.41
	12	264	205	0.52	0.52	0.20	-0.30	-0.27	0.03	365	354	0.61	0.63	0.13	-0.36	-0.37	0.25
RF2	0	7857	5410	0.72	0.72	0.12	-0.22	-0.21	0.00	7776	6776	0.70	0.70	0.32	-0.21	-0.21	0.00
	1	6172	3458	0.68	0.64	0.14	-0.47	-0.41	0.00	5650	4716	0.75	0.75	0.05	-0.44	-0.40	0.00
	2	6312	4246	0.58	0.59	0.00	-0.34	-0.33	0.33	6039	5475	0.68	0.67	0.13	-0.30	-0.28	0.00
	3	10975	7202	0.69	0.67	0.15	-0.26	-0.24	0.00	10277	8388	0.73	0.73	0.00	-0.23	-0.20	0.00
	4	7672	5301	0.73	0.69	0.91	-0.17	-0.15	0.00	7243	6045	0.77	0.77	0.04	-0.16	-0.14	0.00
	5	4105	2020	0.44	0.40	0.11	-0.42	-0.36	0.00	1612	1313	0.40	0.42	0.09	-0.38	-0.36	0.00
	6	1603	756	0.27	0.40	0.00	-0.39	-0.32	0.00	3974	3145	0.55	0.58	0.00	-0.51	-0.45	0.00
	7	2278	1380	0.70	0.65	0.56	-0.33	-0.31	0.00	5376	3953	0.69	0.68	0.03	-0.27	-0.25	0.00
	8	5206	3006	0.64	0.64	0.00	-0.29	-0.28	0.00	1203	1097	0.73	0.75	0.06	-0.21	-0.20	0.03
	9	1189	667	0.66	0.59	0.93	-0.22	-0.22	0.32	2222	1757	0.80	0.80	0.11	-0.32	-0.31	0.01
	10	2920	1808	0.64	0.51	1.00	-0.22	-0.21	0.03	2861	2611	0.70	0.68	0.66	-0.18	-0.18	0.07
	11	1266	1233	0.60	0.60	0.62	-0.16	-0.16	0.82	1341	1331	0.51	0.51	0.04	-0.17	-0.17	0.13
	12	264	205	0.57	0.55	0.35	-0.28	-0.26	0.03	365	354	0.65	0.68	0.05	-0.34	-0.34	0.30

frequently over cluster 5 in autumn months and the seasonal estimates inferred from the reduced data are much closer to observations.

The second is located around Shaanxi, Shanxi, and Henan provinces (spatially corresponding to cluster 2). The ground-level $PM_{2.5}$ pollution is severe in winter

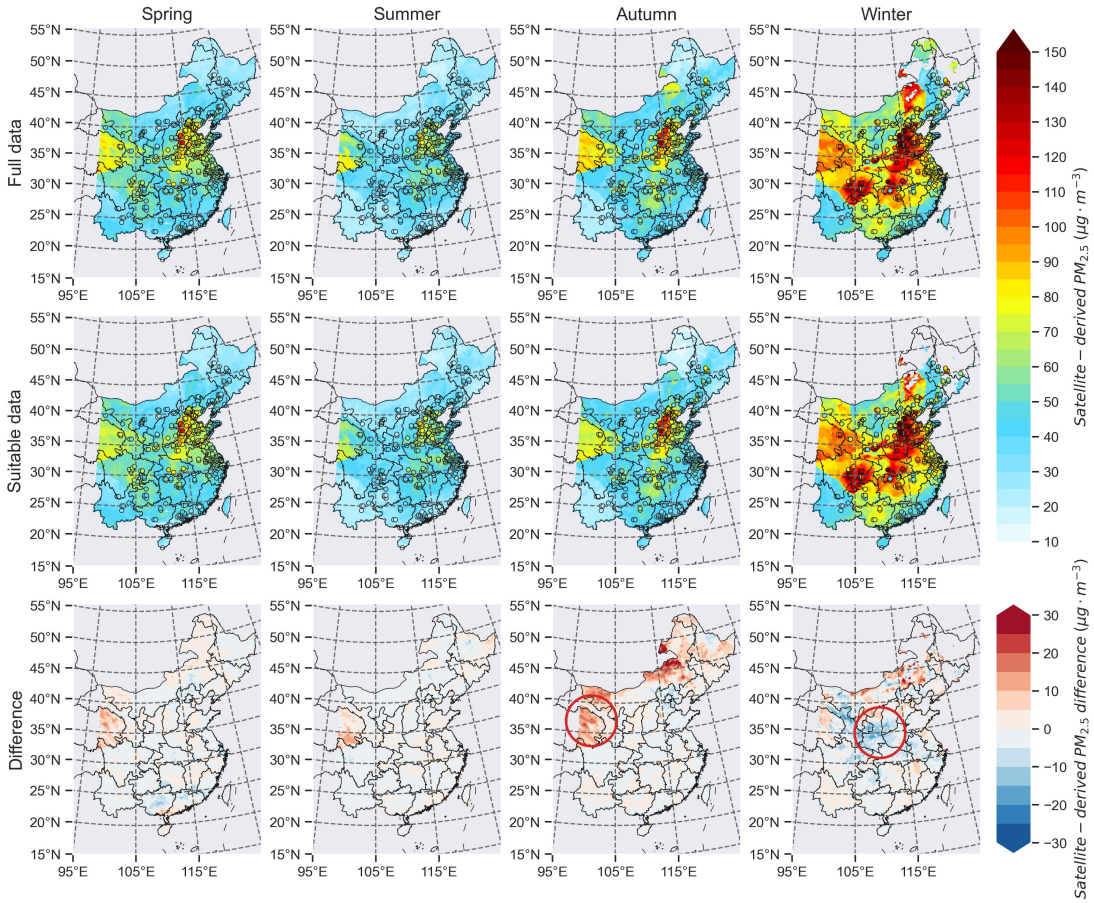


Figure 5.12: Seasonal ground-level $PM_{2.5}$ maps in 2014 over eastern China inferred from the RF2 using the data from the Aqua overpass time at 0.25° latitude × 0.3125° longitude coincide with observed ground-level $PM_{2.5}$ measurements. The first row presents results using the RF2 trained by the full data. The second row presents results using the RF2 trained by the reduced data where $\Gamma_{PBL}^{AOD} \geq 0.5$. The third row presents their differences. Two red circles in the last two subplots highlight areas where large differences are located.

months, causing negative model bias in most previous studies. We find that the winter mean of ground-level $PM_{2.5}$ estimates inferred from the reduced data is higher than those inferred from the full data over cluster 2. Further, we find that the MPE of the winter mean of ground-level $PM_{2.5}$ estimates inferred from the full and the reduced data against the observations are 11% and 6%, respectively. This

suggests that the negative model bias occurs frequently over cluster 2 in winter months and the seasonal estimates inferred from the reduced data are much closer to observations.

5.5 Discussion and concluding remarks

We have proposed a model framework to infer ground-level $PM_{2.5}$ from satellite observations of AOD. Most importantly, the method highlights the need to take into account the fraction of AOD in the PBL, where variations are most likely to reflect changes in ground-level $PM_{2.5}$. The experiment over eastern China during 2014 shows that the method can effectively reduce bias in inferred estimates of ground-level $PM_{2.5}$ and captures more variations in ground-level $PM_{2.5}$ particularly over areas dominated by natural aerosols. We attribute the comparative model improvement to the two benefits due to restricting the data analysis to the suitable data based on the $\Gamma_{PBL}^{AOD} \geq 0.5$ criterion. First, we improve the extent to which AOD is connected to ground-level $PM_{2.5}$. Second, for areas where natural aerosols like dust dominate and are frequently found in the free troposphere, we reduce the disparities in aerosol compositions between AOD and ground-level $PM_{2.5}$. These two aspects jointly improve the representativeness of using AOD as the proxy for ground-level $PM_{2.5}$.

As a result of our criterion we do not have a temporally and spatially continuous ground-level $PM_{2.5}$ field, with gaps left where AOD is not sufficiently representative of aerosol extinction in the PBL. A looser or stricter criterion would retain more or less data, respectively. Our ability to infer ground-level $PM_{2.5}$ depends on this criterion retaining sufficient data. It then becomes a trade-off between the quality and quantity of the AOD data being used to infer ground-level $PM_{2.5}$ concentrations. Alternatively, the resulting ground-level $PM_{2.5}$ estimates inferred

from satellite observations of AOD could be incorporated into a data assimilation framework where gaps are filled by a dynamic model of aerosols. In work (not shown) we also explore whether we can improve the model performance by using $\sum_{PBL}^{AOD} = AOD_{MAIAC} \times \Gamma_{PBL}^{AOD}$ as the main predictor to avoid data loss. However, we have not found any model improvement even on a national scale. Hence, by discarding the data for which $\Gamma_{PBL}^{AOD} < 0.5$ we likely remove the circumstance that the AOD is not sufficiently mixed in the PBL to reflect changes in ground-level $PM_{2.5}$.

A large quantity of research has developed statistical and machine learning models to infer ground-level $PM_{2.5}$ concentrations from satellite observations of AOD and more studies are increasingly emerging. To date, however, limited have investigated the suitability of the methodology. Through this study we demonstrate that caution should be exercised when relating AOD to ground-level $PM_{2.5}$ via data-driven models. Our model framework is sufficiently generic in that it can be applied to other periods and other geographical regions, after taking into account differences in the composition of AOD and ground-level $PM_{2.5}$ that will help define the process-driven and data-driven models.

Chapter 6

Conclusions and prospects

6.1 Thesis overview

Particulate matter (PM) in the atmosphere scatters and absorbs the solar radiation that would otherwise reach solar photovoltaic (PV) panels. PM deposited on solar PV panels further impedes the solar radiation being received by PV semiconductor materials. Consequently, PM represents a major barrier to maximizing the performance of solar PV technologies and thereby compromise our ability to generate clean, carbon-free energy. Atmospheric PM near the surface, particularly those with an aerodynamic diameter $\leq$ 2.5 μm ($PM_{2.5}$), also represents a major threat to human health. Our knowledge of PM pollution and PV power output, however, remains incomplete and uncertain. In this thesis, I aimed to enhance our understanding of PM pollution and PV power output with a focus on their links by using data and models.

I integrated the state-of-the-science GEOS-Chem chemical transport model with the PVLIB-Python which is a solar PV performance model. I designed and carried out a number of global and regional time-slice model experiments covering the

period of 2008–2017 to meet the research goals set in this thesis. To start with, I quantified the impacts of PM pollution on PV power output and distinguished the contribution of atmospheric and deposited PM. With the feasibility of controlling anthropogenic emissions in mind, I then quantified the benefits to PV power output from reducing every individual anthropogenic source sector emissions and thereby identified that the residential sector is where halving emissions provides the largest benefits over East and South Asia. I also quantified the benefits to PV power output by routine sweeping of PM deposited on solar panels. Next, I quantified the reductions in PM pollution and improvements in PV power output through a series of residential fuel emission reduction experiments over an Asian wintertime spatiotemporal domain.

Considering the importance of a reliable ground-level $PM_{2.5}$ data product for relevant epidemiological studies, I finally turned to improving current methods of estimating spatially resolved ground-level $PM_{2.5}$ concentrations from satellite-retrieved AOD. I developed a framework that uses the GEOS-Chem chemical transport model to identify physically relevant data to be modelled by statistical and machine learning models. I used rigorous Monte Carlo simulations to prove that the framework is better than conventional methods.

6.2 Summary of key findings

Below I give a summary of the major findings that address the main scientific questions outlined in section 1.1.

1. **What are the impacts of PM pollution on PV power output and how do we mitigate these impacts?**

   (a) The extent to which PM pollution damages PV power output can be

comparable to the maximum PV power output that can be achieved. The PV power output losses due to deposited PM far exceed those due to atmospheric PM.

(b) Deep cuts in residential emissions substantially benefit PV power output over East and South Asia. The industrial and energy sectors are the next two important places to target following the residential one over South Asia.

(c) Routine sweeping of PM deposited on solar panels is effective at improving PV power output. Even an annual sweeping routine could recover an approximately 40% PV power output in Central Asia and Southern South America.

(d) Eastern and Northeastern China, and the Indo-Gangetic Plain are the three key regions where reducing residential particularly solid biofuel emissions leads approximately linearly to reductions in PM pollution and improvements in PV power output.

2. **How do we produce a value-added ground-level $PM_{2.5}$ data product by using data and models?**

(a) Using chemical transport models to identify physically relevant data that are subsequently modelled by statistical and machine learning models is needed to improve the robustness of ground-level $PM_{2.5}$ estimates inferred from satellite-retrieved AOD.

(b) Rigorous Monte Carlo simulations can help assess the improved robustness of ground-level $PM_{2.5}$ estimates inferred from satellite-retrieved AOD. The improved robustness is mainly reflected by smaller biases that are statistically significant.

6.3 Significance and implications

My thesis, the first of its kind in many aspects, furthers our knowledge of PM pollution and PV power output and provides a holistic assessment of their links. By using a range of data and cutting-edge models, I have addressed a series of important scientific questions that are relevant to designing policies to improve air quality and PV power output.

First, the state-of-the-art GEOS-Chem chemical transport model has been integrated with the PVLIB-Python, a solar PV performance model. I used this integrated model to quantify the impacts of PM pollution on PV power output on a global and decadal scale. Thanks to the detailed mechanism of coupled gas-aerosol phase chemistry included in the GEOS-Chem model, I have not only reproduced spatial patterns of PM impacts on PV power output reported by previous studies, but also additionally identified coastal regions and Southern South American desert region where there are large deposited PM impacts.

Second, I have quantified the first estimates of benefits to PV power output from reducing every individual anthropogenic source sector emissions. These calculations help identify that the residential sector is where halving emissions provides the largest benefits to PV power output over East and South Asia. A 50% reduction in residential emissions would lead to an additional 7,355 GWh yr^{-1} (US\$625 million yr^{-1}) and 1,721 GWh yr^{-1} (US\$136 million yr^{-1}) produced in China and India, respectively.

Third, I have quantified the benefits of reducing residential fuel emissions to Asian wintertime air quality and PV power output. Despite the many nonlinear physical and chemical processes from emission release to PM pollution formation, I have found approximately linear benefits to Asian wintertime air quality and PV power output from reducing residential particularly solid biofuel emissions.

I attribute these approximately linear benefits to the large volume of residential carbon emissions that typically form into carbon aerosols in an approximately linear way in the atmosphere.

Fourth, to the best of my knowledge, I was the first to explicitly consider the representativeness of satellite-retrieved AOD for ground-level $PM_{2.5}$ concentrations when establishing statistical and machine learning models to convert the former to the latter. This is mainly achieved by using chemical transport models to identify physically relevant data to be modelled by statistical and machine learning models. Rigorous Monte Carlo simulations prove the advantage of this framework particularly in reducing bias of ground-level $PM_{2.5}$ estimates inferred from satellite-retrieved AOD.

Although there is increasing evidence suggesting that reducing residential emissions play important roles in reducing PM pollution and associated premature death, the monetary effects of such measures are difficult to robustly quantify. However, the transition toward clean fuels has foreseeable barriers of infrastructure development, such as the construction of a natural-gas pipeline network. Consequently, justifying the residential emission controls which are effective but expensive is relatively hard. Accordingly, residential sources have been largely overlooked in national environmental policies until very recently. For instance, China has not launched a work plan for clean fuel substitution in Beijing-Tianjin-Hebei and their surrounding areas until 2017 (National Development and Reform Commission of China, 2017). By identifying large, approximately linear, and more visible benefits to PV power output from reducing residential particularly solid biofuel emissions, my thesis further strengthens the significance of residential emission controls, and additionally provides a salient argument for accelerating the shift from non-clean to clean fuels within the residential sector.

Inferring ground-level $PM_{2.5}$ concentrations from satellite-retrieved AOD to serve

follow-on epidemiological studies has become a popular research niche. Nonetheless, bias embodied in inferred ground-level $PM_{2.5}$ estimates due to inherent differences of ground-level $PM_{2.5}$ and satellite-retrieved AOD could be sequentially extended to subsequent estimates of population exposure and premature death. My thesis proposes a model framework that considers the representativeness of satellite-retrieved AOD for ground-level $PM_{2.5}$ to reduce this kind of bias and thereby could benefit relevant epidemiological studies.

6.4 Suggestions for future research

My thesis has undoubtedly not studied all aspects of PM pollution and PV power output and their links. Here I summarize several avenues for future research on PM pollution and PV power output.

First, my whole work is centred around a single integrated model comprising the GEOS-Chem 3-D global model of atmospheric composition equipped with an online radiative transfer calculations and the PVLIB-Python model. Some of the limitations and sources of uncertainty of the integrated model have been discussed in section 2.5. Here I also want to mention that although the GEOS-Chem model has been widely evaluated in the literature and in the separate research chapters regarding specific configurations, many processes in the model remain incompletely parameterized and calibrated. For instance, the dry deposition process only has distinction between dust, sea salt, and others. This is definitely imperfect and may compromise the robustness of the findings presented in this thesis. Similar issues might also exist in the PVLIB-Python model. In light of these deficiencies, I underscore that all the findings presented in this thesis should be interpreted in the context of the models and associated configurations

I adopted. Continually advancing model developments may slightly modify but will undoubtedly refine the findings presented in this thesis.

Second, to reduce PM pollution and its impacts on PV power output, I mainly focus on strategies of reducing emissions from individual anthropogenic source sectors. However, different source sectors are likely to be simultaneously targeted to form emission control policies. Taking China's Air Pollution Prevention and Control Action Plan as an example, its control measures such as "Phase out outdated industrial capacities" target energy and industrial sectors simultaneously (Zhang et al., 2019). A complete and comprehensive evaluation of the effectiveness of these already implemented measures, together with the present evaluation work of the effectiveness of single source sector reductions, can provide critical information for developing better emission control policies.

Third, by focusing on period of 2008–2017, my thesis is mainly concerned with the past current. However, climate change likely complicates projections of future PV power output in several ways. PM-related emissions are expected to decline in the future, which will result in more solar radiation to be utilized by solar panels, boosting PV power output. By contrast, ambient temperature is expected to increase with more occurrence of heat waves, jeopardizing PV power output. Changes in frequency of warm, cloudy weather can also substantially alter PV power output. Accurately projecting future PV power output and attributing its changes to individual drivers can offer valuable implications for solar grid planning and operations.

Fourth, my thesis mainly focuses on PV power output, but PV power intermittency is a similarly important variable desiring attention. Ideally, a stable solar energy system continually producing abundant electricity is preferred. The impacts of anthropogenic emission reductions on PV power intermittency, and

future predictions of PV power intermittency in the context of climate change and carbon neutrality can be important topics of future research.

Fifth, my thesis only focuses on the impacts of PM pollution on solar PV panels that are comprised of crystalline silicon cells. While all types of solar PV panels would be subject to the availability of irradiance and thereby the level of PM pollution, further investigation is needed into the impacts of PM pollution on thin-film and other solar PV panels.

Sixth, my thesis simply focuses on the impacts of PM pollution on PV power output but not the opposite. Deploying renewable solar energy in power sector to displace conventional energy generation has air quality and climate benefits. The improvements in PV power output from anthropogenic emission controls will further enhance these benefits, leading to a virtuous cycle. It would be of great interest to study the whole chain of this virtuous cycle in the future.

Appendices

A: Dry deposition calculations in GEOS-Chem "Classic"

The dry deposition velocity (V_d) is viewed as the reciprocal of aerodynamic (R_a) and surface (R_s) resistances in series and gravitational settling velocity (V_g):

$$V_d = V_g + \frac{1}{R_a + R_s} \tag{A.1}$$

The gravitational settling velocity is calculated as:

$$V_g = \frac{\rho_p D_p^2 g C_c}{18\mu}, \tag{A.2}$$

where ρ_p is PM density, D_p is PM diameter, g is the acceleration of gravity, C_c is the slip correction coefficient, and μ is the viscosity coefficient of air. $\frac{\rho_p D_p^2 g}{18}$ can be readily derived and is regarded as a constant. μ as a function of temperature equals $1.8 \times 10^{-5}(\frac{T}{298})^{0.85}$, where T is in K. C_c is given by:

$$C_c = 1 + \frac{2\eta}{D_p}[1.257 + 0.4e^{-\frac{1.1D_p}{2\eta}}], \tag{A.3}$$

where η is the mean free path of air molecules and is calculated as a function of temperature, pressure, and air's kinematic viscosity (v). Calculation of values of C_c indicate that it is typically less than 2% for PM exceeding 10 μm and in this sense can be neglected. On the other hand, the drag force for a 0.1-μm-diameter PM is reduced by a factor of almost three as a result of this slip correction.

The aerodynamic resistance R_a is expressed as:

$$R_a = \frac{\ln \frac{z_R}{z_0} - \psi_H}{\kappa u_*}, \tag{A.4}$$

where z_R is the height at which the dry deposition velocity is evaluated, z_0 is the roughness length as a function of land use and season categories (see Table 3 in Zhang et al. (2001)), ψ_H is the stability function, κ is the Von Karman constant, and u_* is the friction velocity.

The surface resistance R_s is parameterized as:

$$R_s = \frac{1}{\epsilon_0 u_*(E_B + E_{IM} + E_{IN})R_1}, \tag{A.5}$$

where E_B, E_{IM}, and E_{IN} are collection efficiency from Brownian diffusion, impaction, and interception, respectively. R_1 is a correction factor representing the fraction of PM that sticks to the surface. $\epsilon_0 = 3.0$ is an empirical constant.

E_B relies on the Schmidt number, $Sc = \frac{v}{D}$, where D is the Brownian diffusivity, given by:

$$D = \frac{kTC_c}{3\pi\mu D_p}, \tag{A.6}$$

where k is Boltzmann's constant, T is the absolute temperature in K.

E_B is represented as:

$$E_B = Sc^{-\gamma}, \tag{A.7}$$

where γ lies between $\frac{1}{2}$ and $\frac{2}{3}$ with larger values for rougher surfaces (see Table 3 in Zhang et al. (2001)).

E_{IM} is governed by the Stokes number St which has the form:

$$E_{IM} = (\frac{St}{\alpha + St})^{\beta}, \tag{A.8}$$

where α depends on the land use category (see Table 3 in Zhang et al. (2001)) while $\beta = 2$. St is defined as follows:

$$\begin{aligned} St &= \frac{V_g u_*^2}{gv}, \text{ smooth surfaces or surfaces with bluff roughness elements} \\ &= \frac{V_g u_*}{gA}, \text{ vegetated surfaces} \end{aligned} \tag{A.9}$$

where A is the characteristic radius of collectors, the value of which depends on land use and season categories (see Table 3 in Zhang et al. (2001)).

E_{IN} depends on PM diameter and the characteristic radius of collectors:

$$E_{IN} = \frac{1}{2}(\frac{D_p}{A})^2 \tag{A.10}$$

Whether PM posses sufficient inertia to bounce off a surface depends on its Stokes number. Following Slinn (1982), R_1 is expressed as:

$$R_1 = e^{-St^{\frac{1}{2}}} \tag{A.11}$$

If the surface is wet, $St \to 0$, $R_1 = 1$, that is, PM is supposed to stick to a wet surface regardless of its size.

Like calculations in $PM_{2.5}$ and AOD, hygroscopic growth is required to be considered for hydrophilic PM to correct the dry radius r_d and thereby D_p to corresponding wet ones by using RH as the bridge (Gerber, 1985):

$$r_w = [\frac{C_1 r_d^{C_2}}{C_3 r_d^{C_4} - \ln RH} + r_d^3]^{\frac{1}{3}}, \tag{A.12}$$

where r_w is the corrected wet radius, and C_1, C_2, C_3, and C_4 are empirical constants (see Table 1 in Zhang et al. (2001)).

The calculation for R_s involves too many parameters to calibrate for different PM species with varying sizes under different underlying surfaces and meteorological conditions. With the acknowledgement of the difficulty of collecting detailed data for all circumstances, GEOS-Chem 'Classic' currently calculates R_s for sea salt, dust, and others, respectively (Pye et al., 2009). Sea salt is further differentiated between accumulation and coarse modes with distinct diameters and densities. Dust is further differentiated among those distributed in the first four, fifth, sixth, and seventh size bins with distinct diameters and densities. PM diameter and

density are assumed to be 0.5 microns and 1.5 g cm^{-3} (1500 kg m^{-3}) for others. Hygroscopic growth is accounted for sea salt only.

B: Estimating DNI and DHI from GHI using the Erbs model

The Erbs model is an empirical, simply fitted model developed from hourly average observations of GHI, DNI, and DHI. It uses the clearness index (k_t) and the diffuse fraction (k_d) to piecewise link the three terms.

The clearness index, k_t, is defined as:

$$k_t = \frac{GHI}{E_a \cos\theta_s}, \tag{B.1}$$

where θ_s is calculated on line by the PVLIB-Python solar position calculation function. E_a is the extraterrestrial irradiance on a plane normal to the Sun at the TOA and varies throughout the year because of the Earth's elliptical orbit. The varying Earth-Sun distance is empirically predictable during the year:

$$E_a = E_{sc} \times (\frac{R_{av}}{R})^2, \tag{B.2}$$

where $E_{sc} = 1367W\ m^{-2}$ is the solar constant, R_{av} is the mean Earth-Sun distance, R is the actual Earth-Sun distance depending on the day of the year. $(\frac{R_{av}}{R})^2$ can be calculated as:

$$(\frac{R_{av}}{R})^2 = \quad 1.00011 + 0.034221\cos B + 0.00128\sin B \\ +0.000719\cos(2B) + 0.000077\sin(2B), \tag{B.3}$$

where $B = 2\pi\frac{DOY}{265}$ radians with DOY denoting the day of the year.

$k_d = \frac{DHI}{GHI}$ can then be derived from k_t as:

$$\begin{aligned} k_d \quad &= 1 - 0.09k_t, \ (k_t \leq 0.22) \\ &= 0.9511 - 0.1604k_t + 4.388k_t^2 - 16.638k_t^3 + 12.336k_t^4, \ (0.22 < k_t \leq 0.8) \\ &= 0.165, \ (k_t > 0.8), \end{aligned} \tag{B.4}$$

which subsequently resolves $DHI = GHI \times k_d$, based on which and equation 1.37 DNI is determined as:

$$DNI = \frac{GHI - DHI}{\cos\theta_s} \tag{B.5}$$

Note that in the Erbs model, DNI will be set to zero for times when θ_s is greater than a pre-specified value (default to 87°) to avoid potential large errors. In this case, DHI will also be set equal to GHI to ensure that the closure relationship indicated by equation 1.37 remains valid.

C: Transposing irradiance to solar panels

E_b is expressed as:

$$E_b = DNI \times \cos(AOI), \tag{C.1}$$

where AOI is the angle of incidence:

$$AOI = \cos^{-1}[\cos\theta_s \cos\theta_T + \sin\theta_s \sin\theta_T \cos(\theta_A - \theta_{A,array})], \tag{C.2}$$

where θ_T and $\theta_{A,array}$ are the tilt and azimuth angles of the solar panels. For flat panels, $\theta_T = 0$ and $\theta_{A,array} = 180°$. For tilt panels, θ_T = latitude and $\theta_{A,array} = 180°$. For OAT panels, θ_T and $\theta_{A,array}$ are dynamically calculated to track the Sun in the east-west direction.

E_g is expressed as:

$$E_g = GHI \times albedo \times \frac{1 - \cos(1 - \theta_T)}{2}, \tag{C.3}$$

where *albedo* is the surface albedo stored in 'weather' provided by the MERRA-2 meteorology.

E_d generally contains: 1) the isotropic component which represents the uniform irradiance from the sky dome, 2) circumsolar diffuse component which represents the forward scattering of radiation concentrated in the area immediately surrounding the sun, and 3) horizon brightening component. Published models use

different semi-empirical approaches for estimating the combination of these components. In this thesis, I choose the Hay and Davies diffuse model (Loutzenhiser et al., 2007). It defines an anisotropy index, A_i:

$$A_i = \frac{DNI}{E_a}, \tag{C.4}$$

and a ratio of cosines of AOI to θ_s, R_b:

$$R_b = \frac{\cos(AOI)}{\cos\theta_s} \tag{C.5}$$

It then calculates E_d as:

$$E_d = DHI \times [A_i R_b + (1 - A_i)\frac{1 + \cos\theta_T}{2}] \tag{C.6}$$

D: Calculations of DC, AC, and CF in PVLIB-Python model

The maximum-power $P_{mp} = V_{mp} \times I_{mp}$ with V_{mp} and I_{mp} determined as follows:

$$\begin{aligned} V_{mp} &= V_{mp0} + C_2 N_s \delta \ln\frac{E_e}{E_0} + C_3 N_s (\delta \ln\frac{E_e}{E_0})^2 \\ &\quad + [B_{vmp0} + M_{bvmp}(1 - \frac{E_e}{E_0})](T_c - T_0) \\ I_{mp} &= I_{mp0}[C_0\frac{E_e}{E_0} + C_1(\frac{E_e}{E_0})^2][1 + \alpha_{Imp}(T_c - T_0)], \end{aligned} \tag{D.1}$$

where V_{mp} and I_{mp} are the voltage and current at the maximum-power point. δ is a function of T_c defined as:

$$\delta = \frac{n \times k(T_c + 273.15)}{q}, \tag{D.2}$$

where k and q are Boltzmann's constant and the elementary charge constant, respectively. n is the empirically determined "diode factor". All other parameters in equation D.1 are defined as:

V_{mpo} Maximum power voltage at reference condition

$C_0 - C_3$ Empirically determined coefficients relating V_{mp} and I_{mp} to E_e and T_c

N_s Number of cells in a module's cell string(s)

B_{vmp0} Maximum power voltage temperature coefficient at reference condition

M_{bvmp} Coefficient providing the irradiance dependence for the BetaVmp temperature coefficient at reference irradiance

T_0 reference temperature

I_{mpo} Maximum power current at reference condition

α_{imp} Maximum power current temperature coefficient at reference condition

AC power (P_{AC}) is then calculated as:

$$P_{AC} = [\frac{P_{AC0}}{A - B} - C(A - B)](P_{DC} - B) + C(P_{DC} - B)^2, \tag{D.3}$$

where

$P_{AC0} =$ Maximum AC power rating for inverter at reference condition

$P_{DC} =$ DC input power, assuming to be P_{mp}

$$A = P_{DC0}[1 + C_1(V_{DC} - V_{DC0})]$$
$$B = P_{s0}[1 + C_2(V_{DC} - V_{DC0})]$$
$$C = C_0[1 + C_3(V_{DC} - V_{DC0})],$$

where

V_{DC} = DC input voltage, assuming to be V_{mp}

P_{DC0} DC power input that results in P_{AC0} output at reference voltage V_{DC0}

V_{DC0} DC voltage at which the AC power rating is achieved with P_{DC0} power input

P_{s0} DC power required to start the inversion process

C_0 Parameter defining the curvature (parabolic) of the relationship between AC power and DC power at the reference operating condition

$C_1 - C_3$ Empirical coefficients allowing P_{DC0}, P_{s0}, and C_0 to vary linearly with DC voltage input, respectively.

The CF is finally determined as the ratio of P_{AC} to P_{AC0}, i.e. $CF = \frac{P_{AC}}{P_{AC0}}$.

References

Agency, E. P. (1996), 'Air quality criteria for particulate matter', *Report nos EPA/600/P-95/001aF-cF. 3v* .

Aguilera, R., Corringham, T., Gershunov, A. and Benmarhnia, T. (2021), 'Wildfire smoke impacts respiratory health more than fine particles from other sources: observational evidence from southern california', *Nature communications* **12**(1), 1–8.

Albrecht, B. A. (1989), 'Aerosols, cloud microphysics, and fractional cloudiness', *Science* **245**(4923), 1227–1230.

Amos, H. M., Jacob, D. J., Holmes, C., Fisher, J. A., Wang, Q., Yantosca, R. M., Corbitt, E. S., Galarneau, E., Rutter, A., Gustin, M. et al. (2012), 'Gas-particle partitioning of atmospheric hg (ii) and its effect on global mercury deposition', *Atmospheric Chemistry and Physics* **12**(1), 591–603.

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A. and Holmén, K. (2003), 'Mesoscale variations of tropospheric aerosols', *Journal of the Atmospheric Sciences* **60**(1), 119–136.

Andreae, M. O. (1995), Chapter 10 - climatic effects of changing atmospheric aerosol levels, *in* A. Henderson-Sellers, ed., 'Future climates of the world: a modelling perspective', Vol. 16 of *World Survey of Climatology*, Elsevier, pp. 347–398.
URL: *https://www.sciencedirect.com/science/article/pii/S0168632106800337*

Barber, J. L., Thomas, G. O., Kerstiens, G. and Jones, K. C. (2004), 'Current issues and uncertainties in the measurement and modelling of air–vegetation exchange and within-plant processing of pops', *Environmental pollution* **128**(1-2), 99–138.

Beckett, K. P., Freer-Smith, P. and Taylor, G. (1998), 'Urban woodlands: their role in reducing the effects of particulate pollution', *Environmental pollution* **99**(3), 347–360.

Bergin, M. H., Ghoroi, C., Dixit, D., Schauer, J. J. and Shindell, D. T. (2017), 'Large reductions in solar energy production due to dust and particulate air pollution', *Environmental Science & Technology Letters* **4**(8), 339–344.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J. and Schultz, M. G. (2001), 'Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation', *Journal of Geophysical Research: Atmospheres* **106**(D19), 23073–23095.

Bokoye, A. I., Royer, A., O'Neil, N., Cliche, P., Fedosejevs, G., Teillet, P. and McArthur, L. (2001), 'Characterization of atmospheric aerosols across canada from a ground-based sunphotometer network: Aerocan', *Atmosphere-Ocean* **39**(4), 429–456.

Bouwman, A., Lee, D., Asman, W., Dentener, F., Van Der Hoek, K. and Olivier, J. (1997), 'A global high-resolution emission inventory for ammonia', *Global biogeochemical cycles* **11**(4), 561–587.

Boyle, L., Flinchpaugh, H. and Hannigan, M. (2015), 'Natural soiling of photovoltaic cover plates and the impact on transmission', *Renewable Energy* **77**, 166–173.

Boys, B., Martin, R., Van Donkelaar, A., MacDonell, R., Hsu, N., Cooper, M., Yantosca, R., Lu, Z., Streets, D., Zhang, Q. et al. (2014), 'Fifteen-year global time series of satellite-derived fine particulate matter', *Environmental science & technology* **48**(19), 11109–11118.

Boyson, W. E., Galbraith, G. M., King, D. L. and Gonzalez, S. (2007), Performance model for grid-connected photovoltaic inverters., Technical report, Sandia National Laboratories.

Breiman, L. (2001), 'Random forests', *Machine learning* **45**(1), 5–32.

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope, C. A., Apte, J. S., Brauer, M., Cohen, A., Weichenthal, S. et al. (2018), 'Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter', *Proceedings of the National Academy of Sciences* **115**(38), 9592–9597.

Carn, S., Yang, K., Prata, A. and Krotkov, N. (2015), 'Extending the long-term record of volcanic so2 emissions with the ozone mapping and profiler suite nadir mapper', *Geophysical Research Letters* **42**(3), 925–932.

Central Electricity Authority (2021), 'Executive summary report', Available at `https://cea.nic.in/executive-summary-report/?lang=en`.

Chaichan, M. T. and Kazem, H. A. (2016), 'Experimental analysis of solar intensity on photovoltaic in hot and humid weather conditions', *International Journal of Scientific & Engineering Research* **7**(3), 91–96.

Chatfield, R. B., Sorek-Hamer, M., Esswein, R. F. and Lyapustin, A. (2020), 'Satellite mapping of pm 2.5 episodes in the wintertime san joaquin valley: a "static" model using column water vapor', *Atmospheric Chemistry and Physics* **20**(7), 4379–4397.

Che, H., Xia, X., Zhao, H., Dubovik, O., Holben, B. N., Goloub, P., Cuevas-Agulló, E., Estelles, V., Wang, Y., Zhu, J. et al. (2019), 'Spatial distribution of aerosol microphysical and optical properties and direct radiative effect from the china aerosol remote sensing network', *Atmospheric Chemistry and Physics* **19**(18), 11843–11864.

Che, H., Zhang, X., Chen, H., Damiri, B., Goloub, P., Li, Z., Zhang, X., Wei, Y., Zhou, H., Dong, F. et al. (2009), 'Instrument calibration and aerosol optical depth validation of the china aerosol remote sensing network', *Journal of Geophysical Research: Atmospheres* **114**(D3).

Che, H., Zhang, X.-Y., Xia, X., Goloub, P., Holben, B., Zhao, H., Wang, Y., Zhang, X.-C., Wang, H., Blarel, L. et al. (2015), 'Ground-based aerosol climatology of china: aerosol optical depths from the china aerosol remote sensing network (carsnet) 2002–2013', *Atmospheric Chemistry and Physics* **15**(13), 7619–7652.

Chen, S. A., Vishwanath, A., Sathe, S. and Kalyanaraman, S. (2016), 'Shedding light on the performance of solar panels: a data-driven view', *ACM SIGKDD Explorations Newsletter* **17**(2), 24–36.

Chenni, R., Makhlouf, M., Kerbache, T. and Bouzid, A. (2007), 'A detailed modeling method for photovoltaic cells', *Energy* **32**(9), 1724–1730.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A. and Nakajima, T. (2002), 'Tropospheric aerosol optical thickness from the gocart model and comparisons with satellite and sun photometer measurements', *Journal of the atmospheric sciences* **59**(3), 461–483.

Chowdhury, S., Dey, S., Guttikunda, S., Pillarisetti, A., Smith, K. R. and Di Girolamo, L. (2019), 'Indian annual ambient air quality standard is achievable by completely mitigating emissions from household sources', *Proceedings of the National Academy of Sciences* **116**(22), 10711–10716.

Ciencewicki, J. and Jaspers, I. (2007), 'Air pollution and respiratory viral infection', *Inhalation toxicology* **19**(14), 1135–1146.

Conibear, L., Butt, E. W., Knote, C., Arnold, S. R. and Spracklen, D. V. (2018), 'Residential energy use emissions dominate health impacts from exposure to ambient particulate matter in india', *Nature communications* **9**(1), 1–9.

Crawford, B., Hagan, D. H., Grossman, I., Cole, E., Holland, L., Heald, C. L. and Kroll, J. H. (2021), 'Mapping pollution exposure and chemistry during an extreme air quality event (the 2018 kīlauea eruption) using a low-cost sensor network', *Proceedings of the National Academy of Sciences* **118**(27).

Creutzig, F., Agoston, P., Goldschmidt, J. C., Luderer, G., Nemet, G. and Pietzcker, R. C. (2017), 'The underestimated potential of solar energy to mitigate climate change', *Nature Energy* **2**(9), 1–9.

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y. and Schwartz, J. (2016), 'Assessing pm2. 5 exposures with high spatiotemporal resolution across the continental united states', *Environmental science & technology* **50**(9), 4712–4721.

Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., Denn, F. M., Duprat, T., Fukuda, M., Grobe, H. et al. (2018), 'Baseline surface radiation network (bsrn): structure and data description (1992–2017)', *Earth System Science Data* **10**(3), 1491–1501.

Dutton, E. G., Reddy, P., Ryan, S. and DeLuisi, J. J. (1994), 'Features and effects of aerosol optical depth observed at mauna loa, hawaii: 1982–1992', *Journal of Geophysical Research: Atmospheres* **99**(D4), 8295–8306.

Eck, T. F., Holben, B., Reid, J., Dubovik, O., Smirnov, A., O'neill, N., Slutsker, I. and Kinne, S. (1999), 'Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols', *Journal of Geophysical Research: Atmospheres* **104**(D24), 31333–31349.

Ekins-Daukes, N. and Kay, M. (2019), 'Brighten the dark skies', *Nature Energy* **4**(8), 633–634.

Erbs, D., Klein, S. and Duffie, J. (1982), 'Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation', *Solar energy* **28**(4), 293–302.

European Committee for Standardization (CEN) (1998), *Air quality determination of the PM10 fraction of suspended particulate matter reference method and field test procedure to demonstrate reference equivalence of measurement methods (European Standard EN 12341)*.

Fairlie, T. D., Jacob, D. J. and Park, R. J. (2007), 'The impact of transpacific transport of mineral dust in the united states', *Atmospheric Environment* **41**(6), 1251–1266.

Feng, X., Lin, H., Fu, T.-M., Sulprizio, M. P., Zhuang, J., Jacob, D. J., Tian, H., Ma, Y., Zhang, L., Wang, X. et al. (2021), 'Wrf-gc (v2. 0): online two-way coupling of wrf (v3. 9.1. 1) and geos-chem (v12. 7.2) for modeling regional atmospheric chemistry–meteorology interactions', *Geoscientific Model Development* **14**(6), 3741–3768.

Feron, S., Cordero, R. R., Damiani, A. and Jackson, R. B. (2021), 'Climate change extremes and photovoltaic power output', *Nature Sustainability* **4**(3), 270–276.

Fountoukis, C. and Nenes, A. (2007), 'Isorropia ii: a computationally efficient thermodynamic equilibrium model for k^+-ca^{2+}-mg^{2+}-nh_4+-na^+-so_4^{2-};-no_3^-;-cl^-;-h_2o aerosols', *Atmospheric Chemistry and Physics* **7**(17), 4639–4659.
URL: *https://www.atmos-chem-phys.net/7/4639/2007/*

Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M. C., Fowler, D., Koren, I., Langford, B., Lohmann, U. et al. (2015), 'Particulate matter, air quality and climate: lessons learned and future needs', *Atmospheric chemistry and physics* **15**(14), 8217–8299.

Gao, M., Cao, J. and Seto, E. (2015), 'A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of pm2. 5 in xi'an, china', *Environmental pollution* **199**, 56–65.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R. et al. (2017), 'The modern-era retrospective analysis for research and applications, version 2 (merra-2)', *Journal of climate* **30**(14), 5419–5454.

Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Tong, D., Zheng, B., Peng, Y., Huang, X. et al. (2021), 'Tracking air pollution in china: Near real-time pm2. 5 retrievals from multiple data sources', *arXiv preprint arXiv:2103.06520* .

Gerber, H. E. (1985), Relative-humidity parameterization of the navy aerosol model (nam), Technical report, Naval Research Lab Washington DC.

Gil, V., Gaertner, M. A., Gutierrez, C. and Losada, T. (2019), 'Impact of climate change on solar irradiation and variability over the iberian peninsula using regional climate models', *International Journal of Climatology* **39**(3), 1733–1747.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R. et al. (2019), 'Advancements in the aerosol robotic network (aeronet) version 3 database–automated near-real-time quality control algorithm with improved cloud screening for sun photometer aerosol optical depth (aod) measurements', *Atmospheric Measurement Techniques* **12**(1), 169–209.

GlobalPetrolPrices.com (2021), 'Electricity prices for households, september 2020', Available at `https://www.globalpetrolprices.com/electricity\_prices/`.

Gong, S. (2003), 'A parameterization of sea-salt aerosol source function for sub- and super-micron particles', *Global biogeochemical cycles* **17**(4).

Guenther, A., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. and Wang, X. (2012), 'The model of emissions of gases and aerosols from nature version 2.1 (megan2. 1): an extended and updated framework for modeling biogenic emissions', *Geoscientific Model Development* **5**(6), 1471–1492.

Guo, J., Wu, Y., Zhang, X. and Li, X. (2013), 'Estimation of pm2. 5 over eastern china from modis aerosol optical depth using the back propagation neural network', *Huan jing ke xue= Huanjing kexue* **34**(3), 817–825.

Guo, Y., Tang, Q., Gong, D.-Y. and Zhang, Z. (2017), 'Estimating ground-level pm2. 5 concentrations in beijing using a satellite-based geographically and temporally weighted regression model', *Remote Sensing of Environment* **198**, 140–149.

Gupta, P. and Christopher, S. A. (2009), 'Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. a neural network approach', *Journal of Geophysical Research: Atmospheres* **114**(D20).

Hagolle, O., Dedieu, G., Mougenot, B., Debaecker, V., Duchemin, B. and Meygret, A. (2008), 'Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: Application to formosat-2 images', *Remote sensing of environment* **112**(4), 1689–1701.

Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., Levy, R. C., Garay, M. J., Kalashnikova, O. V., Kahn, R. A. et al. (2020), 'Global estimates and long-term trends of fine particulate matter concentrations (1998–2018)', *Environmental Science & Technology* **54**(13), 7879–7890.

Hansen, J. E. and Travis, L. D. (1974), 'Light scattering in planetary atmospheres', *Space science reviews* **16**(4), 527–610.

He, Q. and Huang, B. (2018*a*), 'Satellite-based high-resolution pm2. 5 estimation over the beijing-tianjin-hebei region of china using an improved geographically and temporally weighted regression model', *Environmental pollution* **236**, 1027–1037.

He, Q. and Huang, B. (2018*b*), 'Satellite-based mapping of daily high-resolution ground pm2. 5 in china via space-time regression modeling', *Remote sensing of environment* **206**, 72–83.

Heald, C., Ridley, D., Kroll, J., Barrett, S., Cady-Pereira, K., Alvarado, M. and Holmes, C. (2014), 'Contrasting the direct radiative effect and direct radiative forcing of aerosols', *Atmospheric Chemistry and Physics* **14**(11), 5513–5527.

Herman, J. and Celarier, E. (1997), 'Earth surface reflectivity climatology at 340–380 nm from toms data', *Journal of Geophysical Research: Atmospheres* **102**(D23), 28003–28011.

Hernandez, R. R., Easter, S. B., Murphy-Mariscal, M. L., Maestre, F. T., Tavassoli, M., Allen, E. B., Barrows, C. W., Belnap, J., Ochoa-Hueso, R., Ravi, S. et al. (2014), 'Environmental impacts of utility-scale solar energy', *Renewable and sustainable energy reviews* **29**, 766–779.

HJ 618-2011 (2011), *Determination of atmospheric articles PM10 and PM2.5 in ambient air by gravimetric method.*
URL: *http://english.mee.gov.cn/Resources/standards/Air_Environment/air_method/201111/*

Holben, B. N., Eck, T., Slutsker, I. a., Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y., Nakajima, T. et al. (1998), 'Aeronet—a federated instrument network and data archive for aerosol characterization', *Remote sensing of environment* **66**(1), 1–16.

Holmes, C., Prather, M. and Vinken, G. (2014), 'The climate impact of ship no x emissions: an improved estimate accounting for plume chemistry', *Atmospheric Chemistry and Physics* **14**(13), 6801–6812.

Holmgren, W. F., Hansen, C. W. and Mikofski, M. A. (2018), 'pvlib python: A python package for modeling solar energy systems', *Journal of Open Source Software* **3**(29), 884.

Hsu, N. C., Tsay, S.-C., King, M. D. and Herman, J. R. (2004), 'Aerosol properties over bright-reflecting source regions', *IEEE transactions on geoscience and remote sensing* **42**(3), 557–569.

Hsu, N. C., Tsay, S.-C., King, M. D. and Herman, J. R. (2006), 'Deep blue retrievals of asian aerosol properties during ace-asia', *IEEE transactions on geoscience and remote sensing* **44**(11), 3180–3195.

Hsu, N., Jeong, M.-J., Bettenhausen, C., Sayer, A., Hansell, R., Seftor, C., Huang, J. and Tsay, S.-C. (2013), 'Enhanced deep blue aerosol retrieval algorithm: The second generation', *Journal of Geophysical Research: Atmospheres* **118**(16), 9296–9315.

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J. and Liu, Y. (2017), 'Estimating pm2. 5 concentrations in the conterminous united states using the random forest approach', *Environmental science & technology* **51**(12), 6936–6944.

Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., Quattrochi, D. A., Sarnat, J. A. and Liu, Y. (2013), 'Estimating ground-level pm2. 5 concentrations in the southeastern us using geographically weighted regression', *Environmental Research* **121**, 1–10.

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., Quattrochi, D. A., Puttaswamy, S. J. et al. (2014), 'Estimating ground-level pm2. 5 concentrations in the southeastern united states using maiac aod retrievals and a two-stage model', *Remote Sensing of Environment* **140**, 220–232.

Hudman, R., Moore, N., Mebust, A., Martin, R., Russell, A., Valin, L. and Cohen, R. (2012), 'Steps towards a mechanistic model of global soil nitric oxide emissions: implementation and space based-constraints', *Atmospheric Chemistry and Physics* **12**(16), 7779–7795.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A. and Collins, W. D. (2008), 'Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models', *Journal of Geophysical Research: Atmospheres* **113**(D13).

IEA (2020), Renewables 2020, Technical report, International Energy Agency, Paris.
URL: *https://www.iea.org/reports/renewables-2020*

Ineichen, P., Perez, R., Seal, R., Maxwell, E. and Zalenka, A. (1992), 'Dynamic global-to-direct irradiance conversion models', *ASHRAE transactions* **98**(1), 354–369.

IRENA (2021), Renewable capacity statistics 2021, Technical report, International Renewable Energy Agency, Abu Dhabi.
URL: *https://www.irena.org/publications/2021/March/Renewable-Capacity-Statistics-2021*

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., Cofiño, A. S., Di Luca, A., Faria, S. H., Gorodetskaya, I. V. et al. (2020), 'An update of ipcc climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets', *Earth System Science Data* **12**(4), 2959–2970.

Jacob, D. J. (1999), *Introduction to atmospheric chemistry*, Princeton University Press.

Jaeglé, L., Quinn, P., Bates, T., Alexander, B. and Lin, J.-T. (2011), 'Global distribution of sea salt aerosols: new constraints from in situ and remote sensing observations', *Atmospheric Chemistry and Physics* **11**(7), 3137–3157.

Jerez, S., Tobin, I., Vautard, R., Montávez, J. P., López-Romero, J. M., Thais, F., Bartok, B., Christensen, O. B., Colette, A., Déqué, M. et al. (2015), 'The impact of climate change on photovoltaic power generation in europe', *Nature communications* **6**(1), 1–8.

Jia, H., Ma, X., Yu, F. and Quaas, J. (2021), 'Significant underestimation of radiative forcing by aerosol–cloud interactions derived from satellite-based methods', *Nature Communications* **12**(1), 1–11.

Jiang, Z., Jolleys, M. D., Fu, T.-M., Palmer, P. I., Ma, Y., Tian, H., Li, J. and Yang, X. (2020), 'Spatiotemporal and probability variations of surface pm2. 5 over china between 2013 and 2019 and the associated changes in health risks: An integrative observation and model analysis', *Science of the Total Environment* p. 137896.

Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R., Caudill, M., Rickard, J., Davis, M. et al. (2016), 'Community air sensor network (cairsense) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern united states', *Atmospheric Measurement Techniques* **9**(11), 5281–5292.

Johnson, M. (2010), 'A numerical scheme to calculate temperature and salinity dependent air-water transfer velocities for any gas', *Ocean Science* **6**(4), 913–932.

Kahn, R., Banerjee, P., McDonald, D. and Diner, D. J. (1998), 'Sensitivity of multiangle imaging to aerosol optical depth and to pure-particle size distribution and composition over ocean', *Journal of Geophysical Research: Atmospheres* **103**(D24), 32195–32213.

Kaufman, Y., Tanré, D., Remer, L. A., Vermote, E., Chu, A. and Holben, B. (1997), 'Operational remote sensing of tropospheric aerosol over land from eos moderate resolution imaging spectroradiometer', *Journal of Geophysical Research: Atmospheres* **102**(D14), 17051–17067.

Keller, C. A., Long, M. S., Yantosca, R. M., Da Silva, A., Pawson, S. and Jacob, D. J. (2014), 'Hemco v1. 0: a versatile, esmf-compliant component for calculating emissions in atmospheric models', *Geoscientific Model Development* **7**(4), 1409–1417.

King, D. L., Gonzalez, S., Galbraith, G. M. and Boyson, W. E. (2007), 'Performance model for grid-connected photovoltaic inverters', *Sandia National Laboratories SAND2007-5036* .

King, D. L., Kratochvil, J. A. and Boyson, W. E. (2004), *Photovoltaic array performance model*, United States. Department of Energy.

Koelemeijer, R., De Haan, J. and Stammes, P. (2003), 'A database of spectral surface reflectivity in the range 335–772 nm derived from 5.5 years of gome observations', *Journal of Geophysical Research: Atmospheres* **108**(D2).

Labordena, M., Neubauer, D., Folini, D., Patt, A. and Lilliestam, J. (2018), 'Blue skies over china: The effect of pollution-control on solar power generation and revenues', *PloS one* **13**(11), e0207028.

Lacis, A. (1995), 'Climate forcing, climate sensitivity, and climate response: A radiative modeling perspective on atmospheric aerosols.', *Aerosol forcing of climate* pp. 11–42.

Latimer, R. N. and Martin, R. V. (2019), 'Interpretation of measured aerosol mass scattering efficiency over north america using a chemical transport model', *Atmospheric Chemistry and Physics* **19**(4), 2635–2653.

Lee, H., Liu, Y., Coull, B., Schwartz, J. and Koutrakis, P. (2011), 'A novel calibration approach of modis aod data to predict pm 2.5 concentrations', *Atmospheric Chemistry and Physics* **11**(15), 7991–8002.

Lee, T. Y. and Kaufman, Y. J. (1986), 'Non-lambertian effects on remote sensing of surface reflectance and vegetation index', *IEEE transactions on geoscience and remote sensing* **GE-24**(5), 699–708.

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

Li, M., Zhang, Q., Kurokawa, J.-i., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R. et al. (2017), 'Mix: a mosaic asian anthropogenic emission inventory under the international collaboration framework of the mics-asia and htap', *Atmospheric Chemistry and Physics (Online)* **17**(2).

Li, X., Mauzerall, D. L. and Bergin, M. H. (2020), 'Global reduction of solar power generation efficiency due to aerosols and panel soiling', *Nature Sustainability* **3**(9), 720–727.

Li, X., Wagner, F., Peng, W., Yang, J. and Mauzerall, D. L. (2017), 'Reduction of solar photovoltaic resources due to air pollution in china', *Proceedings of the National Academy of Sciences* **114**(45), 11867–11872.

Li, Z., Xu, H., Li, K., Li, D., Xie, Y., Li, L., Zhang, Y., Gu, X., Zhao, W., Tian, Q. et al. (2018), 'Comprehensive study of optical, physical, chemical, and radiative properties of total columnar atmospheric aerosols over china: An overview of sun–sky radiometer observation network (sonet) measurements', *Bulletin of the American Meteorological Society* **99**(4), 739–755.

Liang, Q., Stolarski, R., Kawa, S., Nielsen, J., Douglass, A., Rodriguez, J., Blake, D., Atlas, E. L. and Ott, L. (2010), 'Finding the missing stratospheric br y: a global modeling study of chbr 3 and ch 2 br 2', *Atmospheric Chemistry and Physics* **10**(5), 2269–2286.

Liao, Q., Zhu, M., Wu, L., Pan, X., Tang, X. and Wang, Z. (2020), 'Deep learning for air quality forecasts: a review', *Current Pollution Reports* pp. 1–11.

Lin, H., Jacob, D. J., Lundgren, E. W., Sulprizio, M. P., Keller, C. A., Fritz, T. M., Eastham, S. D., Emmons, L. K., Campbell, P. C., Baker, B. et al. (2021), 'Harmonized emissions component (hemco) 3.0 as a versatile emissions component for atmospheric models: application in the geos-chem, nasa geos, wrf-gc, cesm2, noaa gefs-aerosol, and noaa ufs models', *Geoscientific Model Development Discussions* pp. 1–26.

Lin, J., Du, M., Chen, L., Feng, K., Liu, Y., Martin, R. V., Wang, J., Ni, R., Zhao, Y., Kong, H. et al. (2019), 'Carbon and health implications of trade restrictions', *Nature communications* **10**(1), 1–12.

Lin, J.-T. and McElroy, M. B. (2010), 'Impacts of boundary layer mixing on pollutant vertical profiles in the lower troposphere: Implications to satellite remote sensing', *Atmospheric Environment* **44**(14), 1726–1739.

Lin, S.-J. and Rood, R. B. (1996), 'Multidimensional flux-form semi-lagrangian transport schemes', *Monthly Weather Review* **124**(9), 2046–2070.

Liu, H., Jacob, D. J., Bey, I. and Yantosca, R. M. (2001), 'Constraints from 210pb and 7be on wet deposition and transport in a global three-dimensional chemical tracer model driven by assimilated meteorological fields', *Journal of Geophysical Research: Atmospheres* **106**(D11), 12109–12128.

Liu, J., Li, W. and Li, J. (2016), 'Quality screening for air quality monitoring data in china', *Environmental pollution* **216**, 720–723.

Liu, J., Mauzerall, D. L., Chen, Q., Zhang, Q., Song, Y., Peng, W., Klimont, Z., Qiu, X., Zhang, S., Hu, M. et al. (2016), 'Air pollutant emissions from chinese households: A major and underappreciated ambient pollution source', *Proceedings of the National Academy of Sciences* **113**, 7756–7761.

Liu, Y., Paciorek, C. J. and Koutrakis, P. (2009), 'Estimating regional spatial and temporal variability of pm2. 5 concentrations using satellite data, meteorology, and land use information', *Environmental health perspectives* **117**(6), 886–892.

Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J. and Koutrakis, P. (2005), 'Estimating ground-level pm2. 5 in the eastern united states using satellite remote sensing', *Environmental science & technology* **39**(9), 3269–3278.

Loutzenhiser, P., Manz, H., Felsmann, C., Strachan, P., Frank, T. and Maxwell, G. (2007), 'Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation', *Solar Energy* **81**(2), 254–267.

Lucchesi, R. (2018), 'File specification for geos fp', *GAMO Office Note No.4 (Version 1.2)* .
URL: *https://gmao.gsfc.nasa.gov/pubs/office_notes*

Lyapustin, A., Wang, Y., Korkin, S. and Huang, D. (2018), 'Modis collection 6 maiac algorithm', *Atmospheric Measurement Techniques* **11**(10), 5741–5765.

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R. and Reid, J. (2011), 'Multiangle implementation of atmospheric correction (maiac): 2. aerosol algorithm', *Journal of Geophysical Research: Atmospheres* **116**(D3).

Ma, Z., Hu, X., Huang, L., Bi, J. and Liu, Y. (2014), 'Estimating ground-level pm2. 5 in china using satellite remote sensing', *Environmental science & technology* **48**(13), 7436–7444.

Marais, E. A., Jacob, D. J., Jimenez, J. L., Campuzano-Jost, P., Day, D. A., Hu, W., Krechmer, J., Zhu, L., Kim, P. S., Miller, C. C. et al. (2016), 'Aqueous-phase mechanism for secondary organic aerosol formation from isoprene: application to the southeast united states and co-benefit of so 2 emission controls', *Atmospheric Chemistry and Physics* **16**(3), 1603–1618.

Martin, R. V., Brauer, M., van Donkelaar, A., Shaddick, G., Narain, U. and Dey, S. (2019), 'No one knows which city has the highest concentration of fine particulate matter', *Atmospheric Environment: X* **3**, 100040.

Martins, V. S., Lyapustin, A., de Carvalho, L. A., Barbosa, C. C. F. and Novo, E. M. L. d. M. (2017), 'Validation of high-resolution maiac aerosol product over south america', *Journal of Geophysical Research: Atmospheres* **122**(14), 7537–7559.

Maxwell, E. L. (1987), A quasi-physical model for converting hourly global horizontal to direct normal insolation, Technical report, Solar Energy Research Institute.

McDuffie, E. E., Martin, R. V., Spadaro, J. V., Burnett, R., Smith, S. J., O'Rourke, P., Hammer, M. S., van Donkelaar, A., Bindle, L., Shah, V. et al. (2021), 'Source sector and fuel contributions to ambient pm2. 5 and attributable mortality across multiple spatial scales', *Nature Communications* **12**(1), 1–12.

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M. and Martin, R. V. (2020), 'A global anthropogenic emission inventory of atmospheric pollutants from

sector-and fuel-specific sources (1970–2017): an application of the community emissions data system (ceds)', *Earth System Science Data* **12**(4), 3413–3442.

McLinden, C., Olsen, S., Hannegan, B., Wild, O., Prather, M. and Sundet, J. (2000), 'Stratospheric ozone in 3-d models: A simple chemistry and the cross-tropopause flux', *Journal of Geophysical Research: Atmospheres* **105**(D11), 14653–14665.

Meng, J., Martin, R. V., Li, C., van Donkelaar, A., Tzompa-Sosa, Z. A., Yue, X., Xu, J.-W., Weagle, C. L. and Burnett, R. T. (2019), 'Source contributions to ambient fine particulate matter for canada', *Environmental science & technology* **53**(17), 10269–10278.

Meral, M. E. and Dincer, F. (2011), 'A review of the factors affecting operation and efficiency of photovoltaic based electricity generation systems', *Renewable and Sustainable Energy Reviews* **15**(5), 2176–2184.

Miao, R., Chen, Q., Zheng, Y., Cheng, X., Sun, Y., Palmer, P. I., Shrivastava, M., Guo, J., Zhang, Q., Liu, Y. et al. (2020), 'Model bias in simulating major chemical components of pm 2.5 in china', *Atmospheric Chemistry and Physics* **20**(20), 12265–12284.

Millet, D. B., Guenther, A., Siegel, D. A., Nelson, N. B., Singh, H. B., de Gouw, J. A., Warneke, C., Williams, J., Eerdekens, G., Sinha, V. et al. (2010), 'Global atmospheric budget of acetaldehyde: 3-d model analysis and constraints from in-situ and satellite observations', *Atmospheric Chemistry and Physics* **10**(7), 3405–3425.

Ministry of New and Renewable Energy (2020), 'State-wise installed capacity of grid interactive solar power as on 31.12.2019', Available at `https://mnre.gov.in/`.

Murray, L. T., Jacob, D. J., Logan, J. A., Hudman, R. C. and Koshak, W. J. (2012), 'Optimized regional and interannual variability of lightning in a global chemical transport model constrained by lis/otd satellite data', *Journal of Geophysical Research: Atmospheres* **117**(D20).

National Development and Reform Commission of China (2017), 'Work plan for clean heating in winter in northern china (2017–2021)', Available at `http://www.gov.cn/xinwen/2017-12/20/content_5248855.htm` (2021/08/16).

National Energy Administration (2020), 'Statistics of photovoltaic power generation in 2019', Available at `http://www.nea.gov.cn/2020-02/28/c_138827923.htm` (2020/02/28).

National Energy Administration (2021), 'Reply to "the proportion of electricity generation from renewable energies in total electricity consumption of the

society"', Available at http://www.nea.gov.cn/2021-04/16/c_139884169.htm (2021/04/16).

Nightingale, P. D., Malin, G., Law, C. S., Watson, A. J., Liss, P. S., Liddicoat, M. I., Boutin, J. and Upstill-Goddard, R. C. (2000), 'In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers', *Global Biogeochemical Cycles* **14**(1), 373–387.

Ordóñez, C., Lamarque, J.-F., Tilmes, S., Kinnison, D. E., Atlas, E. L., Blake, D. R., Sousa Santos, G., Brasseur, G. and Saiz-Lopez, A. (2012), 'Bromine and iodine chemistry in a global chemistry-climate model: description and evaluation of very short-lived oceanic sources', *Atmospheric Chemistry and Physics* **12**(3), 1423–1447.

Pai, S. J., Heald, C. L., Pierce, J. R., Farina, S. C., Marais, E. A., Jimenez, J. L., Campuzano-Jost, P., Nault, B. A., Middlebrook, A. M., Coe, H. et al. (2020), 'An evaluation of global organic aerosol schemes using airborne observations', *Atmospheric Chemistry and Physics* **20**(5), 2637–2665.

Panagea, I. S., Tsanis, I. K. and Koutroulis, A. G. (2017), Climate change impact on photovoltaic energy output: the case of greece, *in* 'Climate Change and the Future of Sustainability', Apple Academic Press, pp. 85–106.

Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico, G., Mattis, I. et al. (2014), 'Earlinet: towards an advanced sustainable european aerosol lidar network', *Atmospheric Measurement Techniques* **7**(8), 2389–2409.

Park, R. J., Jacob, D. J., Field, B. D., Yantosca, R. M. and Chin, M. (2004), 'Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the united states: Implications for policy', *Journal of Geophysical Research: Atmospheres* **109**(D15).

Peters, I., Karthik, S., Liu, H., Buonassisi, T. and Nobre, A. (2018), 'Urban haze and photovoltaics', *Energy & Environmental Science* **11**(10), 3043–3054.

Pfenninger, S. and Staffell, I. (2016), 'Long-term patterns of european pv output using 30 years of validated hourly reanalysis and satellite data', *Energy* **114**, 1251–1265.

Philip, S., Martin, R., Pierce, J., Jimenez, J., Zhang, Q., Canagaratna, M., Spracklen, D., Nowlan, C., Lamsal, L., Cooper, M. et al. (2014), 'Spatially and seasonally resolved estimate of the ratio of organic mass to organic carbon', *Atmospheric Environment* **87**, 34–40.

Philip, S., Martin, R. V. and Keller, C. A. (2016), 'Sensitivity of chemistry-transport model simulations to the duration of chemical and transport operators: a case study with geos-chem v10-01', *Geoscientific Model Development* **9**(5), 1683–1695.

Philip, S., Martin, R. V., Snider, G., Weagle, C. L., van Donkelaar, A., Brauer, M., Henze, D. K., Klimont, Z., Venkataraman, C., Guttikunda, S. K. et al. (2017), 'Anthropogenic fugitive, combustion and industrial dust is a significant, underrepresented fine particulate matter source in global atmospheric models', *Environmental Research Letters* **12**(4), 044018.

Population of Asia (2019), '2019 demographics: densities, ratios, growth rate, clock, rate of men to women', Available at `https://www.populationof.net/` (2021/08/16).

Pye, H., Chan, A., Barkley, M. and Seinfeld, J. (2010), 'Global modeling of organic aerosol: the importance of reactive nitrogen (no x and no 3)', *Atmospheric Chemistry and Physics* **10**(22), 11261–11276.

Pye, H., Liao, H., Wu, S., Mickley, L. J., Jacob, D. J., Henze, D. K. and Seinfeld, J. (2009), 'Effect of changes in climate and emissions on future sulfate-nitrate-ammonium aerosol levels in the united states', *Journal of Geophysical Research: Atmospheres* **114**(D1).

Reda, I. and Andreas, A. (2004), 'Solar position algorithm for solar radiation applications', *Solar energy* **76**(5), 577–589.

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P. and Partyka, G. S. (2017), 'Land surface precipitation in merra-2', *Journal of Climate* **30**(5), 1643–1664.

Riddick, S., Dragosits, U., Blackall, T., Daunt, F., Wanless, S. and Sutton, M. (2012), 'The global distribution of ammonia emissions from seabird colonies', *Atmospheric Environment* **55**, 319–327.

Ridley, D., Heald, C. and Ford, B. (2012), 'North african dust export and deposition: A satellite and model perspective', *Journal of Geophysical Research: Atmospheres* **117**(D2).

SamlexAmerica (2021), 'Solar power system basics', Avaible at `https://samlexamerica.com/resources-support/faq/solarpowersystembasics/` (2021/07/26).

Sarver, T., Al-Qaraghuli, A. and Kazmerski, L. L. (2013), 'A comprehensive review of the impact of dust on the use of solar energy: History, investigations, results, literature, and mitigation approaches', *Renewable and sustainable energy Reviews* **22**, 698–733.

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., Feingold, G., Ghan, S., Guenther, A. B., Kahn, R. et al. (2016), 'Improving our fundamental understanding of the role of aerosol- cloud interactions in the climate system', *Proceedings of the National Academy of Sciences* **113**(21), 5781–5790.

Shen, G., Ru, M., Du, W., Zhu, X., Zhong, Q., Chen, Y., Shen, H., Yun, X., Meng, W., Liu, J. et al. (2019), 'Impacts of air pollutants from rural chinese households under the rapid residential energy transition', *Nature communications* **10**(1), 1–8.

Sherwen, T., Evans, M., Carpenter, L., Andrews, S., Lidster, R., Dix, B., Koenig, T., Sinreich, R., Ortega, I., Volkamer, R. et al. (2016), 'Iodine's impact on tropospheric oxidants: a global model study in geos-chem', *Atmospheric Chemistry and Physics* **16**(2), 1161–1186.

Simon, H., Reff, A., Wells, B., Xing, J. and Frank, N. (2015), 'Ozone trends across the united states over a period of decreasing nox and voc emissions', *Environmental science & technology* **49**(1), 186–195.

Skartveit, A., Olseth, J. A. and Tuft, M. E. (1998), 'An hourly diffuse fraction model with correction for variability and surface albedo', *Solar Energy* **63**(3), 173–183.

Slinn, W. (1982), 'Predictions for particle deposition to vegetative canopies', *Atmospheric Environment (1967)* **16**(7), 1785–1794.

Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S., Shelow, D., Hindin, D. A., Kilaru, V. J. and Preuss, P. W. (2013), 'The changing paradigm of air pollution monitoring', *Environmental science & technology* **47**(20), 11369–11377.

Stephens, G. L. (2005), 'Cloud feedbacks in the climate system: A critical review', *Journal of climate* **18**(2), 237–273.

Stettler, M., Eastham, S. and Barrett, S. (2011), 'Air quality and public health impacts of uk airports. part i: Emissions', *Atmospheric environment* **45**(31), 5415–5424.

Streets, D. G., Wu, Y. and Chin, M. (2006), 'Two-decadal aerosol trends as a likely explanation of the global dimming/brightening transition', *Geophysical Research Letters* **33**(15).

Sweerts, B., Pfenninger, S., Yang, S., Folini, D., Van der Zwaan, B. and Wild, M. (2019), 'Estimation of losses in solar energy production from air pollution in china since 1960 using surface radiation data', *Nature Energy* **4**(8), 657–663.

Tegen, I. and Lacis, A. A. (1996), 'Modeling of particle size distribution and its influence on the radiative properties of mineral dust aerosol', *Journal of Geophysical Research: Atmospheres* **101**(D14), 19237–19244.

Thiessen, A. H. (1911), 'Precipitation averages for large areas', *Monthly weather review* **39**(7), 1082–1089.

Turpin, B. J. and Lim, H.-J. (2001), 'Species contributions to pm2. 5 mass concentrations: Revisiting common assumptions for estimating organic mass', *Aerosol Science & Technology* **35**(1), 602–610.

Twomey, S. (1974), 'Pollution and the planetary albedo', *Atmospheric Environment (1967)* **8**(12), 1251–1256.

Twomey, S. A., Piepgrass, M. and Wolfe, T. (1984), 'An assessment of the impact of pollution on global cloud albedo', *Tellus B* **36**(5), 356–366.

Unger, N., Zheng, Y., Yue, X. and Harper, K. L. (2020), 'Mitigation of ozone damage to the world's land ecosystems by source sector', *Nature Climate Change* **10**(2), 134–137.

U.S. Environmental Protection Agency (1997), *Reference Method for the Determination of Fine Particulate Matter as PM2.5 in the Atmosphere. 40CFR50, Appendix L.*

Van Donkelaar, A., Martin, R. V., Brauer, M. and Boys, B. L. (2015), 'Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter', *Environmental health perspectives* **123**(2), 135–143.

Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C. and Villeneuve, P. J. (2010), 'Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application', *Environmental health perspectives* **118**(6), 847–855.

Van Donkelaar, A., Martin, R. V., Li, C. and Burnett, R. T. (2019), 'Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors', *Environmental science & technology* **53**(5), 2595–2611.

Vidyanandan, K. (2017), 'An overview of factors affecting the performance of solar pv systems', *Energy Scan* **27**(28), 216.

Vinken, G. C., Boersma, K. F., Jacob, D. J. and Meijer, E. W. (2011), 'Accounting for non-linear chemistry of ship plumes in the geos-chem global chemistry transport model', *Atmospheric Chemistry and Physics* **11**(22), 11707–11722.

Wallace, J. M. and Hobbs, P. V. (2006), *Atmospheric science: an introductory survey*, Vol. 92, Elsevier.

Wang, G., Zhang, R., Gomez, M. E., Yang, L., Zamora, M. L., Hu, M., Lin, Y., Peng, J., Guo, S., Meng, J. et al. (2016), 'Persistent sulfate formation from london fog to chinese haze', *Proceedings of the National Academy of Sciences* **113**(48), 13630–13635.

Wang, J. and Christopher, S. A. (2003), 'Intercomparison between satellite-derived aerosol optical thickness and pm2. 5 mass: Implications for air quality studies', *Geophysical research letters* **30**(21).

Wang, J., Li, J., Ye, J., Zhao, J., Wu, Y., Hu, J., Liu, D., Nie, D., Shen, F., Huang, X. et al. (2020), 'Fast sulfate formation from oxidation of so 2 by no 2 and hono observed in beijing haze', *Nature communications* **11**(1), 1–7.

Wang, J., Ye, J., Zhang, Q., Zhao, J., Wu, Y., Li, J., Liu, D., Li, W., Zhang, Y., Wu, C. et al. (2021), 'Aqueous production of secondary organic aerosol from fossil-fuel emissions in winter beijing haze', *Proceedings of the National Academy of Sciences* **118**(8).

Wang, Q., Jacob, D. J., Fisher, J. A., Mao, J., Leibensperger, E., Carouge, C., Sager, P. L., Kondo, Y., Jimenez, J., Cubison, M. et al. (2011), 'Sources of carbonaceous aerosols and deposited black carbon in the arctic in winter-spring: implications for radiative forcing', *Atmospheric Chemistry and Physics* **11**(23), 12453–12473.

Wang, Q., Jacob, D. J., Spackman, J. R., Perring, A. E., Schwarz, J. P., Moteki, N., Marais, E. A., Ge, C., Wang, J. and Barrett, S. R. (2014), 'Global budget and radiative forcing of black carbon aerosol: Constraints from pole-to-pole (hippo) observations across the pacific', *Journal of Geophysical Research: Atmospheres* **119**(1), 195–206.

Wang, X., Heald, C., Ridley, D., Schwarz, J., Spackman, J., Perring, A., Coe, H., Liu, D. and Clarke, A. (2014), 'Exploiting simultaneous observational constraints on mass and absorption to estimate the global direct radiative forcing of black carbon and brown carbon', *Atmospheric Chemistry and Physics* **14**(20), 10989–11010.

Wang, Y. X., McElroy, M. B., Jacob, D. J. and Yantosca, R. M. (2004), 'A nested grid formulation for chemical transport over asia: Applications to co', *Journal of Geophysical Research: Atmospheres* **109**(D22).

Weagle, C. L., Snider, G., Li, C., van Donkelaar, A., Philip, S., Bissonnette, P., Burke, J., Jackson, J., Latimer, R., Stone, E. et al. (2018), 'Global sources of fine particulate matter: interpretation of pm2. 5 chemical composition observed

by spartan using a global chemical transport model', *Environmental science & technology* **52**(20), 11670–11681.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L. and Cribb, M. (2019), 'Estimating 1-km-resolution pm2. 5 concentrations across china using the space-time random forest approach', *Remote Sensing of Environment* **231**, 111221.

Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., Guo, J., Peng, Y., Li, J., Lyapustin, A. et al. (2020), 'Improved 1 km resolution pm2. 5 estimates across china using enhanced space–time extremely randomized trees.', *Atmospheric Chemistry & Physics* **20**(6).

Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., Su, T. and Cribb, M. (2021), 'Reconstructing 1-km-resolution high-quality pm2. 5 data records from 2000 to 2018 in china: spatiotemporal variations and policy implications', *Remote Sensing of Environment* **252**, 112136.

Werf, G. R., Randerson, J. T., Giglio, L., Leeuwen, T. T. v., Chen, Y., Rogers, B. M., Mu, M., Van Marle, M. J., Morton, D. C., Collatz, G. J. et al. (2017), 'Global fire emissions estimates during 1997–2016', *Earth System Science Data* **9**(2), 697–720.

Wesely, M. (2007), 'Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models', *Atmospheric Environment* **41**, 52–63.

Wild, M. (2009), 'Global dimming and brightening: A review', *Journal of Geophysical Research: Atmospheres* **114**(D10).

Wu, S., Mickley, L. J., Jacob, D. J., Logan, J. A., Yantosca, R. M. and Rind, D. (2007), 'Why are there large differences between models in global budgets of tropospheric ozone?', *Journal of Geophysical Research: Atmospheres* **112**(D5).

Wu, Y., Guo, J., Zhang, X., Tian, X., Zhang, J., Wang, Y., Duan, J. and Li, X. (2012), 'Synergy of satellite and ground based observations in estimation of particulate matter in eastern china', *Science of the Total Environment* **433**, 20–30.

Xiao, Q., Chang, H. H., Geng, G. and Liu, Y. (2018), 'An ensemble machine-learning model to predict historical pm2. 5 concentrations in china from satellite data', *Environmental science & technology* **52**(22), 13260–13269.

Xiao, Q., Wang, Y., Chang, H. H., Meng, X., Geng, G., Lyapustin, A. and Liu, Y. (2017), 'Full-coverage high-resolution daily pm2. 5 estimation using maiac aod in the yangtze river delta of china', *Remote Sensing of Environment* **199**, 437–446.

Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B. and Bai, Y. (2015), 'Daily estimation of ground-level pm2. 5 concentrations over beijing using 3 km resolution modis aod', *Environmental science & technology* **49**(20), 12280–12288.

Xin, J., Wang, Y., Li, Z., Wang, P., Hao, W. M., Nordgren, B. L., Wang, S., Liu, G., Wang, L., Wen, T. et al. (2007), 'Aerosol optical depth (aod) and ångström exponent of aerosols observed by the chinese sun hazemeter network from august 2004 to september 2005', *Journal of Geophysical Research: Atmospheres* **112**(D5).

Xin, J., Wang, Y., Pan, Y., Ji, D., Liu, Z., Wen, T., Wang, Y., Li, X., Sun, Y., Sun, J. et al. (2015), 'The campaign on atmospheric aerosol research network of china: Care-china', *Bulletin of the American Meteorological Society* **96**(7), 1137–1155.

Xue, T., Geng, G., Han, Y., Wang, H., Li, J., Li, H.-t., Zhou, Y. and Zhu, T. (2021), 'Open fire exposure increases the risk of pregnancy loss in south asia', *Nature Communications* **12**(1), 1–10.

Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T. and Zhang, Q. (2019), 'Spatiotemporal continuous estimates of pm2. 5 concentrations in china, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations', *Environment international* **123**, 345–357.

Yao, F., Si, M., Li, W. and Wu, J. (2018), 'A multidimensional comparison between modis and viirs aod in estimating ground-level pm2. 5 concentrations over a heavily polluted region in china', *Science of the Total Environment* **618**, 819–828.

Yao, F., Wu, J., Li, W. and Peng, J. (2019), 'A spatially structured adaptive two-stage model for retrieving ground-level pm2. 5 concentrations from viirs aod in china', *ISPRS Journal of Photogrammetry and Remote Sensing* **151**, 263–276.

Yun, X., Shen, G., Shen, H., Meng, W., Chen, Y., Xu, H., Ren, Y., Zhong, Q., Du, W., Ma, J. et al. (2020), 'Residential solid fuel emissions contribute significantly to air pollution and associated health impacts in china', *Science advances* **6**(44), eaba7621.

Zender, C. S., Bian, H. and Newman, D. (2003), 'Mineral dust entrainment and deposition (dead) model: Description and 1990s dust climatology', *Journal of Geophysical Research: Atmospheres* **108**(D14).

Zhang, L., Gong, S., Padro, J. and Barrie, L. (2001), 'A size-segregated particle dry deposition scheme for an atmospheric aerosol module', *Atmospheric environment* **35**(3), 549–560.

Zhang, L., Kok, J. F., Henze, D. K., Li, Q. and Zhao, C. (2013), 'Improving simulations of fine dust surface concentrations over the western united states by optimizing the particle size distribution', *Geophysical Research Letters* **40**(12), 3270–3275.

Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W. et al. (2019), 'Drivers of improved pm2. 5 air quality in china from 2013 to 2017', *Proceedings of the National Academy of Sciences* **116**(49), 24463–24469.

Zhang, T., Zhu, Z., Gong, W., Zhu, Z., Sun, K., Wang, L., Huang, Y., Mao, F., Shen, H., Li, Z. et al. (2018), 'Estimation of ultrahigh resolution pm2. 5 concentrations in urban areas using 160 m gaofen-1 aod retrievals', *Remote Sensing of Environment* **216**, 91–104.

Zhang, Y., Li, Z., Bai, K., Wei, Y., Xie, Y., Zhang, Y., Ou, Y., Cohen, J., Zhang, Y., Peng, Z. et al. (2021), 'Satellite remote sensing of atmospheric particulate matter mass concentration: Advances, challenges, and perspectives', *Fundamental Research* **1**(3), 240–258.

Zhang, Y., Li, Z., Zhang, Y., Hou, W., Xu, H., Chen, C. and Ma, Y. (2014), 'High temporal resolution aerosol retrieval using geostationary ocean color imager: application and initial validation', *Journal of Applied Remote Sensing* **8**(1), 083612.

Zhao, B., Zheng, H., Wang, S., Smith, K. R., Lu, X., Aunan, K., Gu, Y., Wang, Y., Ding, D., Xing, J. et al. (2018), 'Change in household fuels dominates the decrease in pm2. 5 exposure and premature mortality in china in 2005–2015', *Proceedings of the National Academy of Sciences* **115**(49), 12401–12406.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J. et al. (2018), 'Trends in china's anthropogenic emissions since 2010 as the consequence of clean air actions', *Atmospheric Chemistry and Physics* **18**(19), 14095–14111.