

# The role of small-scale cloud, aerosol, and radiation processes for Earth's climate

Thesis by  
Clare E. Singer

In Partial Fulfillment of the Requirements for the  
Degree of  
Ph.D. in Environmental Science and Engineering

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2024  
Defended July 20, 2023

© 2024

Clare E. Singer

ORCID: 0000-0002-1708-0997

All rights reserved

## ACKNOWLEDGEMENTS

There are many people I need to thank for both supporting me in my pursuit of the Ph.D. and for distracting me enough along the way to make sure I had a good time.

To my partner, my biggest supporter, and my soon-to-be husband, Dan. Thank you for waking up at weird hours to call me on the other side of the world when I was away doing science. Thank you for your huge sacrifice of taking on a remote graduate study to live with me in Pasadena and moving with me to New York. Thank you for all the home-cooked dinners and lunches that made my office mates jealous.

To my parents, your support goes back to the very beginning. I had the immense privilege to learn the joys and perils of academic science from a young age. Thank you for always bringing me along to conferences, and taking off time to explore the cool places we were able to travel as a result. Thank you for fiercely protecting my love of learning and never pushing me more than I pushed myself as a child, which was probably already too much. Mom, thank you for leading by example and showing me that there's no better time than right now to have a big adventure.

To my friends, in Pasadena and far beyond. Thank you for your support and immense curiosity. Thank you for letting me teach you about the weather on every one of our hikes. Sherry, thank you for always asking what I'm studying so you can give my elevator pitch to your family. Thank you to my friends at Caltech who have cheered me on and cheered me up throughout the past five years. Lily and Sara, I wouldn't have made it this far, this happily, without the two of you.

To my research group. The CliMA House has been an incredibly special place to do a Ph.D., and mostly because of the wonderful people who sit inside it everyday. Thank you for teaching me everything I know about climate and coding. Thank you for celebrating every birthday with cake and every Friday with a BBQ. Thank you for indulging my passions for baking and organizing. Thanks especially to Zhaoyi Shen and Anna Jaruga for teaching me so much and being amazing role models.

To my advisor, Tapio Schneider. Thank you for your support, guidance, and trust. I am so glad I ended up at Caltech working with you, and that you accepted me into the group despite an unconventional start. Often along the way I felt lost or confused, but you never doubted I could figure my own way out. Your willingness to let me succeed has brought me so much confidence, thank you.

And finally, a bizarre thank you to the COVID-19 pandemic lock-downs. Thank you for upending life as we knew it, but providing a small silver lining of a chance to unplug from the daily grind and reconnect in new ways with friends and family. Thank you for providing the impetus for Dan to move to California and for giving us the freedom to spend three months in Steamboat skiing with friends and family. Thank you for inspiring two road trips from Chicago to D.C. and extending my visits home from days to weeks. And thank you for pushing me to start reading novels again and sending snail mail to loved ones on the East Coast.

## ABSTRACT

What makes clouds ethereal and beautiful also makes them complex and challenging to understand and to model. The important (thermo)dynamical processes of clouds occur at scales from microns (cloud-aerosol interactions), to meters (turbulence), to thousands of kilometers (synoptic weather patterns), and every scale in between. In this thesis, I explore several facets of how clouds interact with, respond to, and shape Earth's climate. I focus on small-scale processes, using high-resolution models and theory, to understand phenomena that can have large-scale impacts.

In the first three chapters of this thesis, I explore the idea of stratocumulus-cumulus transitions. Chapters 1 and 2 develop and demonstrate a conceptual model of a cloud-topped atmospheric boundary layer, which is rooted in mixed-layer theory. This model is able to concisely explain both the spatial stratocumulus-cumulus transition observed in the historical period, as well as a transition that has only been hypothesized by models, which may occur in the future as the direct effect of extreme concentrations of atmospheric CO<sub>2</sub>, or which may have occurred in the past. I use this conceptual model to show the importance of sea surface temperature variations for driving the climatological transition, and on sea surface warming as a positive feedback for the CO<sub>2</sub>-induced transition. Chapter 3 extends this work to understand the global response to CO<sub>2</sub>-induced stratocumulus-cumulus transitions and the role for spatial teleconnections by embedding this conceptual model of the boundary layer into a global climate model (GCM). In the GCM we see both a fast adjustment in low cloud cover to CO<sub>2</sub>, as well as a slower surface temperature-mediated feedback. Under CO<sub>2</sub> quadrupling, the stratocumulus cloud regions shrink in extent as the cloud-top longwave cooling is inhibited by CO<sub>2</sub> and surface temperatures also increase.

The final two chapters diverge from the previous theme to present two studies using very high-resolution models to explore how clouds interact with i) aerosols and ii) radiation. In Chapter 4, using a particle-based cloud microphysics model, I find that aerosol hygroscopicity, determined by the chemical composition of the particles, can alter stratocumulus cloud macrophysical properties, like liquid water path by up to 25% (in the regime of small aerosol sizes). I compare these results to a more standard moment-based microphysics model and find that this model is overly sensitive to aerosol hygroscopicity in the regime of small aerosol sizes, but realistically represents the negative sensitivity for large aerosol sizes. Finally, in

Chapter 5, I use a Monte Carlo 3D radiative transfer solver to estimate the global albedo bias introduced in models which make the standard assumption that photon fluxes in the horizontal are zero (the so-called Independent Column Approximation). I extrapolate globally from a set of resolved tropical cloud fields, using a learned empirical relation between top-of-atmosphere flux bias and cloud water path. I conclude that in a global model that resolves clouds at small-enough spatial scales, the tropical-mean, annual-mean bias may be on the order of  $3 \text{ W m}^{-2}$ .

## PUBLISHED CONTENT AND CONTRIBUTIONS

- [1] C. E. Singer et al. “Top-of-Atmosphere Albedo Bias from Neglecting Three-Dimensional Cloud Radiative Effects”. In: *Journal of the Atmospheric Sciences* 78.12 (2021), pp. 4053–4069. DOI: 10.1175/JAS-D-21-0032.1.

C.E.S. participated in the design of the project and led the radiative transfer modeling, analysis of the simulation data, and writing and editing of the manuscript.

*Used for Chapter 5.*

Published content used with permission from the American Meteorological Society.

## TABLE OF CONTENTS

Acknowledgements . . . . .	iii
Abstract . . . . .	v
Published Content and Contributions . . . . .	vii
Table of Contents . . . . .	vii
Chapter 0 : Introduction . . . . .	1
0.1 Background . . . . .	1
0.2 Thesis Outline . . . . .	3
Chapter I : Stratocumulus-cumulus transition explained by bulk boundary layer theory . . . . .	9
1.1 Introduction . . . . .	10
1.2 Bulk boundary layer model description . . . . .	12
1.3 Observational data for model comparison and evaluation . . . . .	18
1.4 Results . . . . .	19
1.5 Discussion . . . . .	26
1.6 Conclusions . . . . .	27
1.7 Appendix A: Subsetting daily reanalysis data . . . . .	29
Chapter II : CO <sub>2</sub> -driven stratocumulus cloud breakup in a bulk boundary layer model . . . . .	35
2.1 Introduction . . . . .	36
2.2 Bulk boundary layer model with interactive SSTs . . . . .	37
2.3 Ensemble Kalman inversion for parameter calibration . . . . .	40
2.4 Results . . . . .	43
2.5 Conclusions . . . . .	49
Chapter III : CO <sub>2</sub> -induced stratocumulus cloud breakup in a global climate model . . . . .	55
3.1 Introduction . . . . .	55
3.2 Methodology . . . . .	58
3.3 Preliminary results: Response to CO <sub>2</sub> quadrupling . . . . .	63
3.4 Future work . . . . .	67
Chapter IV : Stratocumulus cloud sensitivity to aerosol hygroscopicity . . . . .	71
4.1 Introduction . . . . .	71
4.2 Methods . . . . .	75
4.3 Results from simulations with particle-based microphysics . . . . .	79
4.4 Comparison with 2-moment bulk scheme . . . . .	86
4.5 Conclusions . . . . .	90
4.6 Appendix A: Supplemental figures . . . . .	91
Chapter V : Top-of-atmosphere albedo bias from neglecting three-dimensional cloud radiative effects . . . . .	98
5.1 Introduction . . . . .	99

5.2	Methods . . . . .	101
5.3	Radiative flux bias dependence on zenith angle . . . . .	106
5.4	Seasonal cycle of radiative flux bias . . . . .	113
5.5	Implications for Climate Models . . . . .	115
5.6	Summary and conclusions . . . . .	119
5.7	Appendix A: LES model setup . . . . .	121
5.8	Appendix B: Radiative transfer details . . . . .	124
5.9	Appendix C: Cloud property proxy for flux bias . . . . .	126
Appendix A : Gender disparities in GPS qualifying exam outcomes . . . . .		134
A.1	Remarks from the author . . . . .	134
A.2	Part 1: Are there robust statistical differences in exam outcome between men and women? . . . . .	135
A.3	Part 2: Why are there differences in exam outcome between men and women? . . . . .	143

## *Chapter 0*

### INTRODUCTION

The overarching theme of the works presented in this thesis will be examining the role that clouds play in Earth's climate system. To this end, the chapters that follow will focus on two aspects: how the background meteorological and aerosol conditions determine the properties of the clouds that form, and how clouds in turn affect the climate system they exist within via radiative feedbacks. Specifically, I will present results using a range of models, from highly idealized conceptual frameworks, to very high-resolution small-scale simulations, to full complexity climate models.

#### **0.1 Background**

Contrary to the popular nickname for Earth as the “blue marble,” the Earth often appears mostly white from space due to the clouds. Clouds cover 70% of Earth's surface on average, reflecting around 15% of all incoming shortwave radiation back to space (Cesana et al., 2019). In total, the Earth reflects 29% of the incoming radiation (i.e. the albedo is 0.29), and the clouds account for about half of this, thus exerting a strong cooling effect (Stephens et al., 2015). Clouds regulate the Earth's temperature, both through this shortwave reflection, but also through their greenhouse effect by trapping outgoing longwave radiation (OLR). Clouds generally fall into three categories, illustrated in Figure 0.1:

1. Low clouds (stratocumulus and cumulus), which have a strong cooling effect due to their optical thickness and weak warming effect due to their warm cloud-top temperatures.
2. High clouds (cirrus), which typically are optically thin at visible wavelengths and thus have a weak cooling effect, but are still quite opaque in the infrared and thus still have a strong warming effect since they emit at similarly cold temperatures.
3. Deep clouds (cumulonimbus), which also have a strong cooling due to their large albedo, but also a strong warming due to their cold cloud-tops that emit at very cold temperatures and reduce OLR.

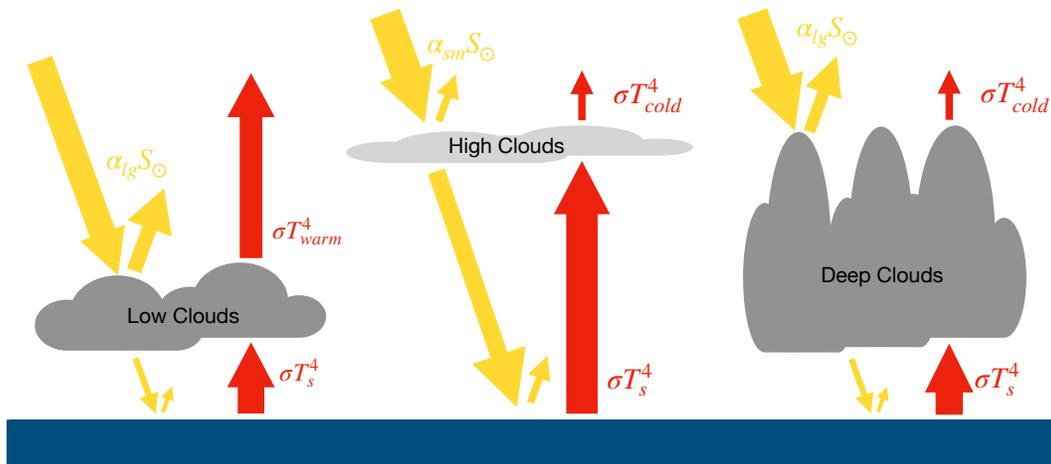


Figure 0.1: Schematic illustration of shortwave and longwave cloud radiative effects from low, high, or deep clouds. Low clouds have a large albedo  $\alpha_{lg}$  and weak greenhouse effect  $T_{warm} \sim T_s$ , which results in a net cooling. High clouds have a small albedo  $\alpha_{sm}$  and strong greenhouse effect  $T_{cold} \ll T_s$ , which results in net warming. Deep clouds combine both, because they have a large albedo and strong greenhouse effect, and thus result in near zero net radiative effect.

Chapters 1 through 4 of this thesis will focus on low clouds, particularly subtropical marine stratocumulus clouds. Chapter 5 is more expansive, including both shallow and deep clouds.

Low clouds, or specifically subtropical marine boundary layer (MBL) clouds, are a main focus in this thesis primarily because of their outsized role in the climate system (Stevens and Brenguier, 2009), but also because of the difficulties we have simulating them which lead to large errors in current generation global climate models (GCMs) (Bony and Dufresne, 2005). MBL clouds fall into two main categories: stratocumulus (Sc) and cumulus (Cu). Sc clouds occur over eastern subtropical ocean basins (e.g. off the coast here near Pasadena) where there is strong subsidence and relatively cool sea surface temperatures (SST). The Sc-topped boundary layer is shallow (approx. 1 km) and well-mixed, due to an upside-down overturning circulation driven by longwave cloud-top radiative cooling. The cloud layer itself is usually quite geometrically thin (approx. 200 m), but optically thick, with nearly 100% cloud cover. Cu clouds in contrast occur further west where the SSTs are warmer and the subsidence is weaker. They penetrate deeper into the atmosphere (approx. 2 km) with the cloud layer decoupled from the sea surface. Cu clouds are scattered, with typical cloud cover of 10–20%, but the albedo of each cloud is still fairly high. Observations along transects from the east to west across

subtropical ocean basins show a climatological transition from Sc to Cu that is termed the “stratocumulus-cumulus transition” or SCT (Bretherton, McCoy, et al., 2019) and this has been replicated in numerical experiments as well (Roode et al., 2016; Tan et al., 2016).

In addition to the background meteorology, the formation of a cloud is predicated on the existence of aerosol particles that can serve as cloud condensation nuclei (CCN). The efficacy of a CCN depends on its size and composition, which can be parameterized by the hygroscopicity using  $\kappa$ -Köhler theory (Petters and Kreidenweis, 2007). Aerosol-cloud interactions (ACI) contribute substantially to the overall uncertainty of climate prediction, partly because of the difficulty both simulating and measuring the important processes of condensation and coalescence that take place at the microscale (Myhre et al., 2013; Fan et al., 2016; Gettelman et al., 2019).

Finally, the coupling goes both ways: not only are the properties of the clouds that form strongly controlled by the atmospheric conditions within which they occur, but the climate system is also strongly controlled by clouds. Clouds control the climate state through their radiative effects, e.g. how they scatter shortwave and absorb longwave radiation. This controls the heating and cooling within the atmosphere and of the Earth’s surface. The precise amount of heating and cooling from clouds — their radiative effects — depends on the detailed structure of the clouds. In climate models, and higher resolution simulations, we make approximations to the radiative transfer to make it computationally tractable. The most important of these is called the independent column approximation, or ICA, wherein we assume that horizontal fluxes of photons between atmospheric columns are negligible (Cahalan et al., 1994; Schäfer et al., 2016; Hogan and Bozzo, 2018; Hogan, Fielding, et al., 2019). Without the ICA, the distributed computation in a global model becomes impossible, because all vertical columns must now exchange information at a rate that makes it impossible to scale the computation across multiple processing units. However, this approximation necessarily neglects three-dimensional effects of clouds which can be important, especially for clouds with larger aspect ratios.

## **0.2 Thesis Outline**

The works presented in this thesis use a variety of tools and methods ranging from idealized, conceptual models, to very high-resolution limited-domain models, to coarse-resolution global models. The first three chapters explore the theme of stratocumulus-cumulus transitions (SCT). Chapter 1, develops a conceptual model

of a cloud-topped atmospheric boundary layer. This model builds on the foundation of mixed-layer models for stratocumulus clouds (e.g., Lilly, 1968; Bretherton and Wyant, 1997; Stevens, 2006), but extends the framework to explicitly include decoupling of the cloud-topped boundary layer, breakup of the stratocumulus deck, and existence of cumulus clouds (Bretherton and Wyant, 1997; Schneider et al., 2019). In Chapter 1, I show how this model applies to the climatological SCT observed across the North East Pacific transect between California and Hawaii. I use the model to quantify the importance of different meteorological factors in driving this transition.

In Chapter 2, I extend the bulk boundary layer model by coupling it to a slab ocean with interactive surface temperatures and an idealized representation of radiative transfer. This extended model is therefore able to capture important feedbacks between cloud-thinning and surface warming that is ignored when sea surface temperatures (SSTs) are fixed (Tan et al., 2016). With this setup, I demonstrate the direct effect of  $\text{CO}_2$  on stratocumulus clouds by running experiments with increasing  $\text{CO}_2$ , analogous to the large-eddy simulations (LES) from Schneider et al., 2019. I show that this conceptual model also has a critical concentration of  $\text{CO}_2$  that leads to stratocumulus cloud breakup and exhibits hysteresis behavior, with the clouds not reforming until  $\text{CO}_2$  concentrations are lowered much past the tipping point.

Chapter 3 takes another step forward, delving into one of the key limitations of the work in Chapter 2, which is the idealized representation of how a single location with climatological stratocumulus cloud cover is coupled with the rest of the tropics. In this chapter, I apply the key idea of decoupling induced cloud breakup from the bulk boundary layer model to the Community Earth System Model (CESM). CESM has a very complex representation of boundary layer turbulence and clouds, called the Cloud Layers Unified By Binormals (CLUBB) scheme, which I replace with a simple diagnostic cloud fraction based on the degree of decoupling of the boundary layer, as calculated in the theory from Chapters 1 and 2. I explore how including feedbacks between local cloud cover and large-scale changes in circulation and thermodynamics modifies the prediction of a stratocumulus tipping point with  $\text{CO}_2$  predicted in Chapter 2 and Schneider et al., 2019.

The last two chapters diverge from this main theme to explore two detailed aspects of clouds and their interactions with aerosols and radiation. In these chapters I use two different high-resolution large eddy models, coupled to first, a high-resolution microphysics scheme, and second to a high-resolution radiative transfer scheme.

In Chapter 4, I use a Lagrangian cloud model with particle-based microphysics to explore how aerosol properties influence the macroscopic cloud. This relatively new type of microphysics scheme contrasts to traditional bulk (moment) representations of aerosols, cloud droplets, and rain drops — where the distributions of these particles are characterized by a finite number of (usually 1 or 2) moments of the distribution — by instead tracking a random statistical sample of these particles explicitly throughout the domain (Shima et al., 2009; Arabas et al., 2015; Dziekan et al., 2019; Grabowski et al., 2019). These particle-based schemes often go by the name superdroplet methods, where the particles that are tracked are called “superdroplets,” and can be used in a range of dynamical models from 0D boxes, to 1D parcels, to fully 3D LES. Chapter 4 examines the role of aerosol hygroscopicity on cloud formation and in which regimes hygroscopicity may be an important factor to consider. In particular, these superdroplet schemes are well-suited for studying the role of aerosol composition because they can introduce additional properties to the particles (e.g., aerosol chemical composition, ice crystal density or shape) with little additional computational cost. Studies of aerosol-cloud-interactions historically have only looked at hygroscopicity effects in parcel models where the microphysical effects do not feedback on the cloud dynamics (Reutter et al., 2009; Chen et al., 2016; Pöhlker et al., 2021).

Chapter 5 ties together themes from throughout the thesis, asking the question: “how do the smallest scales influence the largest?”. I look at the 3D radiative effects of clouds, and present an estimate of the global albedo bias introduced by neglecting these effects in models. I use LES to simulate clouds in various dynamical regimes and then uses a 3D radiative transfer model to track the trajectories of individual photons as they scatter through the cloudy domain (Mayer and Kylling, 2005; Emde et al., 2016). By doing so, I am able to calculate the bias introduced by typical radiative transfer schemes that neglect these 3D effects (Pincus et al., 2003; Shonk and Hogan, 2008), and extrapolate from these few LES cases up to a global estimate (Barker, Kato, et al., 2012; Barker, Cole, Li, Yi, et al., 2015; Barker, Cole, Li, and Salzen, 2016).

Finally, this thesis concludes with an appendix discussion on the topic of the “climate” of the geoscience community at Caltech (not the physical climate of Earth). I present data collected during my Ph.D. on the Caltech GPS Division’s Qualifying Exam and results I found of disparity in exam outcome by gender.

## References

- [1] G. Cesana et al. “The Cumulus And Stratocumulus CloudSat-CALIPSO Dataset (CASCCAD)”. In: *Earth System Science Data* 11.4 (2019), pp. 1745–1764. DOI: 10.5194/essd-11-1745-2019.
- [2] G. L. Stephens et al. “The albedo of Earth”. In: *Reviews of Geophysics* 53.1 (2015), pp. 141–163. DOI: 10.1002/2014RG000449.
- [3] B. Stevens and J.-I. Brenguier. “Cloud-controlling Factors: Low Clouds”. In: *Clouds Perturbed Clim. Syst. Their Relatsh. to Energy Balanc. Atmos. Dyn. Precip.* 1802. 2009.
- [4] S. Bony and J.-L. Dufresne. “Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models”. In: *Geophys. Res. Lett.* 32.20 (2005), p. L20806. DOI: 10.1029/2005GL023851.
- [5] C. S. Bretherton, I. L. McCoy, et al. “Cloud, Aerosol, and Boundary Layer Structure across the Northeast Pacific Stratocumulus–Cumulus Transition as Observed during CSET”. In: *Mon. Weather Rev.* 147.6 (2019), pp. 2083–2103. DOI: 10.1175/MWR-D-18-0281.1.
- [6] S. R. de Roode et al. “Large-Eddy Simulations of EUCLIPSE–GASS Lagrangian Stratocumulus-to-Cumulus Transitions: Mean State, Turbulence, and Decoupling”. In: *J. Atmos. Sci.* 73.6 (2016), pp. 2485–2508. DOI: 10.1175/JAS-D-15-0215.1.
- [7] Z. Tan et al. “Large-eddy simulation of subtropical cloud-topped boundary layers: 1. A forcing framework with closed surface energy balance”. In: *J. Adv. Model. Earth Syst.* 8.4 (2016), pp. 1565–1585. DOI: 10.1002/2016MS000655.
- [8] M. D. Petters and S. M. Kreidenweis. “A single parameter representation of hygroscopic growth and cloud condensation nucleus activity”. In: *Atmos. Chem. Phys.* 7.8 (2007), pp. 1961–1971. DOI: 10.5194/acp-7-1961-2007.
- [9] G. Myhre et al. “Anthropogenic and Natural Radiative Forcing”. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by T. Stocker et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2013. Chap. 8, pp. 659–740.
- [10] J. Fan et al. “Review of Aerosol–Cloud Interactions: Mechanisms, Significance, and Challenges”. In: *Journal of the Atmospheric Sciences* 73.11 (2016), pp. 4221–4252. DOI: 10.1175/JAS-D-16-0037.1.
- [11] A. Gettelman et al. “Cloud Microphysics Across Scales for Weather and Climate”. In: *Current Trends in the Representation of Physical Processes in Weather and Climate Models*. Ed. by D. A. Randall et al. Singapore: Springer Singapore, 2019, pp. 71–94. DOI: 10.1007/978-981-13-3396-5\_4.

- [12] R. F. Cahalan et al. “Independent Pixel and Monte Carlo Estimates of Stratocumulus Albedo”. In: *Journal of the Atmospheric Sciences* 51.24 (1994), pp. 3776–3790. DOI: 10.1175/1520-0469(1994)051<3776:IPAMCE>2.0.CO;2.
- [13] S. A. K. Schäfer et al. “Representing 3-D cloud radiation effects in two-stream schemes: 1. Longwave considerations and effective cloud edge length”. In: *Journal of Geophysical Research: Atmospheres* 121.14 (2016), pp. 8567–8582. DOI: 10.1002/2016JD024876.
- [14] R. J. Hogan and A. Bozzo. “A Flexible and Efficient Radiation Scheme for the ECMWF Model”. In: *Journal of Advances in Modeling Earth Systems* 10 (8 2018). <https://doi.org/10.1029/2018MS001364>, pp. 1990–2008. DOI: <https://doi.org/10.1029/2018MS001364>.
- [15] R. J. Hogan, M. D. Fielding, et al. “Entrapment: An Important Mechanism to Explain the Shortwave 3D Radiative Effect of Clouds”. In: *Journal of the Atmospheric Sciences* 2019 (1 2019), pp. 48–66. DOI: 10.1175/JAS-D-18-0366.1.
- [16] D. K. Lilly. “Models of cloud-topped mixed layers under a strong inversion”. In: *Quart. J. Roy. Meteor. Soc.* 94 (1968), pp. 292–309. DOI: 10.1002/qj.49709440106.
- [17] C. S. Bretherton and M. C. Wyant. “Moisture Transport, Lower-Tropospheric Stability, and Decoupling of Cloud-Topped Boundary Layers”. In: *J. Atmos. Sci.* 54 (1997), pp. 148–167. DOI: 10.1175/1520-0469(1997)054<0148:MTL TSA>2.0.CO;2.
- [18] B. Stevens. “Bulk boundary-layer concepts for simplified models of tropical dynamics”. In: *Theor. Comput. Fluid Dyn.* 20.5-6 (2006), pp. 279–304. DOI: 10.1007/s00162-006-0032-z.
- [19] T. Schneider et al. “Possible climate transitions from breakup of stratocumulus decks under greenhouse warming”. In: *Nat. Geosci.* 12.3 (2019), pp. 163–167. DOI: 10.1038/s41561-019-0310-1.
- [20] S.-I. Shima et al. “The super-droplet method for the numerical simulation of clouds and precipitation: a particle-based and probabilistic microphysics model coupled with a non-hydrostatic model”. In: *Q. J. R. Meteorol. Soc.* 135.642 (2009), pp. 1307–1320. DOI: 10.1002/qj.441.
- [21] S. Arabas et al. “libcloudph++ 1.0: a single-moment bulk, double-moment bulk, and particle-based warm-rain microphysics library in C++”. In: *Geosci. Model Dev.* 8 (2015), pp. 1677–1707. DOI: 10.5194/gmd-8-1677-2015.
- [22] P. Dziekan et al. “University of Warsaw Lagrangian Cloud Model (UWLCM) 1.0: a modern large-eddy simulation tool for warm cloud modeling with Lagrangian microphysics”. In: *Geosci. Model Dev.* 12 (2019), pp. 2587–2606. DOI: 10.5194/gmd-12-2587-2019.

- [23] W. W. Grabowski et al. “Modeling of Cloud Microphysics: Can We Do Better?” In: *Bulletin of the American Meteorological Society* 100.4 (2019), pp. 655–672. DOI: 10.1175/BAMS-D-18-0005.1.
- [24] P. Reutter et al. “Aerosol-and updraft-limited regimes of cloud droplet formation: influence of particle number, size and hygroscopicity on the activation of cloud condensation nuclei (CCN)”. In: *Atmos. Chem. Phys.* 9 (2009), pp. 7067–7080. DOI: 10.5194/acp-9-7067-2009.
- [25] J. Chen et al. “New understanding and quantification of the regime dependence of aerosol-cloud interaction for studying aerosol indirect effects”. In: *Geophys. Res. Lett.* 43.4 (2016), pp. 1780–1787. DOI: 10.1002/2016GL067683.
- [26] M. L. Pöhlker et al. “Aitken mode particles as CCN in aerosol- and updraft-sensitive regimes of cloud droplet formation”. In: *Atmos. Chem. Phys.* 21.15 (2021), pp. 11723–11740. DOI: 10.5194/acp-21-11723-2021.
- [27] B. Mayer and A. Kylling. “Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use”. In: *Atmospheric Chemistry and Physics* 5.7 (2005), pp. 1855–1877. DOI: 10.5194/acp-5-1855-2005.
- [28] C. Emde et al. “The libRadtran software package for radiative transfer calculations (version 2.0.1)”. In: *Geoscientific Model Development* 9.5 (2016), pp. 1647–1672. DOI: 10.5194/gmd-9-1647-2016.
- [29] R. Pincus et al. “A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields”. In: *Journal of Geophysical Research: Atmospheres* 108.13 (2003), pp. 1–5. DOI: 10.1029/2002jd003322.
- [30] J. K. Shonk and R. J. Hogan. “Tripleclouds: An efficient method for representing horizontal cloud inhomogeneity in 1D Radiation schemes by using three regions at each height”. In: *Journal of Climate* 21.11 (2008), pp. 2352–2370. DOI: 10.1175/2007JCLI1940.1.
- [31] H. W. Barker, S. Kato, et al. “Computation of Solar Radiative Fluxes by 1D and 3D Methods Using Cloudy Atmospheres Inferred from A-train Satellite Data”. In: *Surveys in Geophysics* 33.3-4 (2012), pp. 657–676. DOI: 10.1007/s10712-011-9164-9.
- [32] H. W. Barker, J. N. S. Cole, J. Li, B. Yi, et al. “Estimation of Errors in Two-Stream Approximations of the Solar Radiative Transfer Equation for Cloudy-Sky Conditions”. In: *Journal of the Atmospheric Sciences* 72.11 (2015), pp. 4053–4074. DOI: 10.1175/JAS-D-15-0033.1.
- [33] H. W. Barker, J. N. S. Cole, J. Li, and K. von Salzen. “A parametrization of 3-D subgrid-scale clouds for conventional GCMs: Assessment using A-Train satellite data and solar radiative transfer characteristics”. In: *Journal of Advances in Modeling Earth Systems* 8.2 (2016), pp. 566–597. DOI: 10.1002/2015MS000601.

*Chapter 1***STRATOCUMULUS-CUMULUS TRANSITION EXPLAINED BY  
BULK BOUNDARY LAYER THEORY**

- [1] C. E. Singer and T. Schneider. “Stratocumulus-cumulus transition explained by bulk boundary layer theory”. In: *Journal of Climate* (2023). In revision.

[This chapter is temporarily embargoed.]

*Chapter 2***CO<sub>2</sub>-DRIVEN STRATOCUMULUS CLOUD BREAKUP IN A  
BULK BOUNDARY LAYER MODEL**

- [1] C. E. Singer and T. Schneider. “CO<sub>2</sub>-driven stratocumulus cloud breakup in a bulk boundary layer model”. In: *Journal of Climate* (2023). In revision.

[This chapter is temporarily embargoed.]

*Chapter 3***CO<sub>2</sub>-INDUCED STRATOCUMULUS CLOUD BREAKUP IN A  
GLOBAL CLIMATE MODEL**

[This chapter is temporarily embargoed.]

*Chapter 4***STRATOCUMULUS CLOUD SENSITIVITY TO AEROSOL  
HYGROSCOPICITY**

[This chapter is temporarily embargoed.]

## TOP-OF-ATMOSPHERE ALBEDO BIAS FROM NEGLECTING THREE-DIMENSIONAL CLOUD RADIATIVE EFFECTS

- [1] C. E. Singer et al. “Top-of-Atmosphere Albedo Bias from Neglecting Three-Dimensional Cloud Radiative Effects”. In: *Journal of the Atmospheric Sciences* 78.12 (2021), pp. 4053–4069. doi: 10.1175/JAS-D-21-0032.1.

© American Meteorological Society. Used with permission.

**Abstract.** Clouds cover on average nearly 70% of Earth’s surface and regulate the global albedo. The magnitude of the shortwave reflection by clouds depends on their location, optical properties, and three-dimensional (3D) structure. Due to computational limitations, Earth system models are unable to perform 3D radiative transfer calculations. Instead they make assumptions, including the independent column approximation (ICA), that neglect effects of 3D cloud morphology on albedo. We show how the resulting radiative flux bias (ICA-3D) depends on cloud morphology and solar zenith angle. We use high-resolution (20–100 m horizontal resolution) large-eddy simulations to produce realistic 3D cloud fields covering three dominant regimes of low-latitude clouds: shallow cumulus, marine stratocumulus, and deep convective cumulonimbus. A Monte Carlo code is used to run 3D and ICA broadband radiative transfer calculations; we calculate the top-of-atmosphere (TOA) reflected flux and surface irradiance biases as functions of solar zenith angle for these three cloud regimes. Finally, we use satellite observations of cloud water path (CWP) climatology, and the robust correlation between CWP and TOA flux bias in our LES sample, to roughly estimate the impact of neglecting 3D cloud radiative effects on a global scale. We find that the flux bias is largest at small zenith angles and for deeper clouds, while the albedo bias is most prominent for large zenith angles. In the tropics, the annual-mean shortwave radiative flux bias is estimated to be  $3.1 \pm 1.6 \text{ W m}^{-2}$ , reaching as much as  $6.5 \text{ W m}^{-2}$  locally.

**Significance Statement.** Clouds cool the Earth by reflecting sunlight back to space. The amount of reflection is determined by their location, details of their 3D structure, and the droplets or ice crystals they are composed of. Global models

cannot simulate the 3D structure of clouds because computational power is limited, so they approximate that clouds only scatter sunlight in a 1D vertical column. In this study, we use local models to directly simulate how clouds scatter sunlight in 3D and compare with a 1D approximation. We find the largest bias for overhead sun and for deeper clouds. Using satellite observations of bulk cloud properties, we estimate the tropical annual-mean bias introduced by the 1D approximation to be  $3.1 \pm 1.6 \text{ W m}^{-2}$ .

## 5.1 Introduction

Earth's average albedo is roughly 29%, with clouds accounting for about half of the solar radiative energy fluxes reflected back to space (Stephens et al., 2015). Accurately simulating clouds is crucial for modeling Earth's albedo. However, Earth system models (ESMs) struggle to accurately represent the albedo's magnitude, spatial patterns, and seasonal variability (Bender et al., 2006; Voigt et al., 2013; Engström et al., 2015). Simulating clouds is difficult for several reasons, but one major factor is their wide range of spatial scales. Clouds have complex three-dimensional (3D) morphologies created by turbulent motions at length scales down to tens of meters or smaller. However, the typical resolution of an ESM is around only 10–100 km in the horizontal and 100–200 m in the vertical in the lower troposphere (Schneider et al., 2017). This discrepancy means that clouds are not explicitly resolved in ESMs. Instead, they are represented by parameterizations and, for purposes of radiative transfer (RT) calculations, are approximated as broken plane-parallel structures within grid cells (Marshak and Davis, 2005).

The plane-parallel approximation (PPA) leads to important biases in RT calculations (R. Cahalan and Wiscombe, 1992). Over the past 20 years, RT solvers have made significant progress in reducing some of these biases, either by making use of semi-empirical deterministic parameterizations of cloud heterogeneity (J. K. Shonk and Hogan, 2008) or through stochastic sampling of plane-parallel cloudy columns based on assumed distributions and characteristics of cloud structural properties (Pincus et al., 2003). These approximate solvers are likely to become even more accurate in the future, as dynamical parameterizations provide increasingly detailed cloud statistics (e.g., Cohen et al., 2020). Moreover, the PPA bias may be reduced in ESMs by using embedded cloud-resolving models (Cole, Barker, Randall, et al., 2005; Kooperman et al., 2016), albeit at great additional computational expense, in an approach known as cloud superparameterization (Khairoutdinov and Randall, 2001).

This progress has led to a renewed interest in another source of bias that was, until recently, overshadowed by biases due to the PPA: the treatment of horizontal radiative fluxes in ESMs (R. F. Cahalan et al., 1994; Schäfer et al., 2016; Hogan, Fielding, et al., 2019). ESMs make the independent column approximation (ICA) when performing RT calculations. This approximation neglects horizontal radiative fluxes, decoupling the RT calculation between atmospheric columns to make the problem computationally tractable. 3D radiative transfer will remain too expensive to run in ESMs in the foreseeable future, making the ICA a necessary simplification (Hogan and Bozzo, 2018). For this reason, it is important to quantify and document biases due to the ICA.

In this context, the effect of cloud structure on horizontal radiative transfer has gained attention, enabled by advances in computation that make 3D RT feasible at high spectral resolution (Mayer and Kylling, 2005; Emde et al., 2016; Villefranque et al., 2019; Gristey et al., 2019; Veerman et al., 2020). The structural differences between ICA and a full 3D RT calculation have been documented before (Marshak, Davis, Wiscombe, and Titov, 1995; O’Hirok and Gautier, 1998; O’Hirok and Gautier, 2005; Barker, Stephens, et al., 2003; Barker, Kato, et al., 2012), and many alternatives to ICA have been proposed to minimize their mismatch (e.g., Marshak, Davis, Wiscombe, and R. Cahalan, 1995; Várnai and Davies, 1999; Frame et al., 2009; Hogan and J. K. P. Shonk, 2013; Wissmeier et al., 2013; Okata et al., 2017; Oreopoulos and Barker, 1999; Klinger and Mayer, 2016; Klinger and Mayer, 2020; Hogan, Fielding, et al., 2019).

Nevertheless, most studies have been focused on theoretical cases, small spatial and temporal domains, or improving satellite retrieval algorithms. Some notable exceptions are Cole, Barker, O’Hirok, et al. (2005), who calculate the ICA bias from two-dimensional (2D) RT in a superparameterized cloud resolving model at 4 km horizontal resolution; and Barker, Cole, Li, Yi, et al. (2015) and Barker, Cole, Li, and Salzen (2016), who calculate the ICA bias using 2D cloud fields retrieved from CloudSat and CALIPSO.

Here we discuss the magnitude of the bias that results from neglecting the 3D cloud radiative effects by making the ICA. We use large-eddy simulations (LES) to generate 3D cloud fields representing three canonical cloud regimes: shallow cumulus convection, stratocumulus, and deep convection. These cloud regimes are representative of the clouds typically found in the tropics. Previous studies that quantify 3D cloud radiative effects globally have used 2D cloud fields retrieved from

satellites or superparameterized models, or inferred 3D fields using some stochastic generator (O’Hirok and Gautier, 1998; O’Hirok and Gautier, 2005; Barker, Cole, Li, Yi, et al., 2015; Barker, Cole, Li, and Salzen, 2016). These approaches can better represent the spatial distribution of cloud types but are restricted to the coarse resolution of satellite footprints or rely on assumptions to generate 3D fields. We instead use high-fidelity models to generate realistic 3D cloud fields at very high resolution, sacrificing some ability to generalize beyond the tropics from our limited number of LES cases. All of these methods present different challenges, either in generalizing to global scales, or in representing the details of small scales, but the simplifications are necessary because 3D cloud retrievals from satellite are not yet available. However, recent progress in stereoscopic observations is bringing us closer to having global high-resolution observations of 3D cloud structure (e.g., Romps and Öktem, 2018; Castro et al., 2020).

We calculate the bias between the true reflected flux and the flux approximated by ICA using a Monte Carlo RT code. The shortwave radiative flux bias is shown to vary with solar zenith angle and cloud type. Because the solar zenith angle varies with the diurnal and seasonal cycle, we quantify the effect of the 3D bias on these timescales. Finally, using global satellite observations of cloud climatology, we estimate the spatiotemporal bias that would result in global models that resolve clouds but still make the ICA. As stated earlier, most ESMs make the ICA and use some cloud heterogeneity parameterization to reduce the PPA bias, so the bias associated with only the ICA is an underestimate of the total bias. Because of the diversity of assumptions made by global models to account for phenomena such as cloud overlap, and the fundamental resolution dependence of cloud heterogeneity emulators, in this study we focus on the bias resulting from RT using only the ICA on fully resolved 3D cloud structures from LES.

## 5.2 Methods

### Large-eddy simulations of clouds

We generate three-dimensional cloud fields from high-resolution LES using the anelastic solver PyCLES (Pressel, Kaul, et al., 2015; Pressel, Mishra, et al., 2017). The LES are run in three dynamical regimes to simulate shallow cumulus (ShCu), stratocumulus (Sc), and deep-convective cumulonimbus clouds (Cb). Figure 5.1 shows volume renderings of each cloud regime alongside profiles of cloud fraction; key properties of the different cloud regime simulations can be found in Table 5.1 with more details in appendix A. In general, LES are capable of reproduc-

ing observed cloud characteristics (e.g., cloud cover, liquid water path, cloud depth) including aspects of 3D structures (Griewank et al., 2020).

ShCu clouds are convective clouds with typical cloud cover of 10–20% and cloud top height (CTH) around 2 km. They occur frequently over tropical oceans, on average covering 20% but up to 40% of their surface (Cesana et al., 2019). In this study, ShCu are represented by two LES case studies, BOMEX and RICO, which represent non-precipitating and precipitating shallow convection over tropical oceans, respectively (Siebesma et al., 2003; vanZanten et al., 2011). Sc clouds are shallower, with CTH only around 1 km. They have near 100% cloud cover and typically blanket subtropical oceans off the west coast of continents (Cesana et al., 2019). Sc are represented by the DYCOMS-II RF01 LES case of a Sc deck off the coast of California (Stevens et al., 2005). Cb clouds are deep convective thunderstorm clouds that occur frequently over mid-latitude continents in summer and in the tropics, e.g., in the intertropical convergence zone (ITCZ). Their CTH can reach up to 15 km or higher, they often contain ice, and anvils at the top contribute to a cloud cover around 30%. Cb clouds are represented in this paper by the TRMM-LBA LES case, based on measurements of convection over land in the Amazon (Grabowski et al., 2006).

An ensemble of snapshots is used to estimate the mean and variance of the bias for each cloud type. The snapshots are chosen to be at least one convective turnover time apart (1 hour for BOMEX and RICO, 30 minutes for DYCOMS-II RF01, and 90 minutes for TRMM-LBA). For ShCu and Sc, we take snapshots evenly spaced in time starting once the simulation has reached a statistically quasi-steady state, after an initial spin-up period. For the Cb case we take snapshots from an initial-condition ensemble at several time points representative of transient and fully-developed deep convection at 4, 5.5, and 7 hours into the simulation (10:00, 11:30, and 13:00 local time). We also analyze the effect of convective aggregation in Cb (Jeevanjee and Romps, 2013; Wing et al., 2017; Patrizio and Randall, 2019) by analyzing snapshots from an initial-condition ensemble run over a larger domain  $((40 \text{ km})^2$ , compared to the original  $(20 \text{ km})^2$ ). In both cases, we use only the snapshots at 13:00 local time of fully-developed deep convection, characterized by stable liquid and ice water paths, for the cloud-type specific calculations. The rest of the snapshots are used in our estimate of the tropical shortwave flux bias. We choose ensemble sizes that capture the natural variability of morphology in each LES case: 20 for ShCu (10 each of BOMEX and RICO) and 5 for Sc; for Cb we take 15 snapshots from each time

LES Case Name	Cloud Type	Domain Size (km <sup>3</sup> )	Resolution (m <sup>3</sup> )	Cloud cover	CWP (g m <sup>-2</sup> )	CTH (km)	Cloud Phase
BOMEX	ShCu	6.4 × 6.4 × 3	20 × 20 × 20	0.22 ± 0.03	44.6 ± 7.8	1.72 ± 0.08	Liquid
RICO	ShCu	12.8 × 12.8 × 6	40 × 40 × 40	0.25 ± 0.01	90 ± 20	2.28 ± 0.18	Liquid
DYCOMS II RF01	Sc	3.36 × 3.36 × 1.5	35 × 35 × 5	0.996 ± 0.002	53.8 ± 0.7	0.911 ± 0.004	Liquid
TRMM-LBA, <i>t</i> = 4 hr	Cb	20 × 20 × 22	100 × 100 × 50	0.35 ± 0.01	280 ± 20	6.4 ± 0.5	Liquid/Ice
TRMM-LBA, <i>t</i> = 5.5 hr	Cb	20 × 20 × 22	100 × 100 × 50	0.33 ± 0.02	380 ± 60	9.4 ± 1.5	Liquid/Ice
TRMM-LBA, <i>t</i> = 7 hr	Cb	20 × 20 × 22	100 × 100 × 50	0.32 ± 0.06	290 ± 80	10 ± 2	Liquid/Ice
TRMM-LBA agg.	Cb	40 × 40 × 22	100 × 100 × 50	0.30 ± 0.02	360 ± 80	12.2 ± 0.6	Liquid/Ice

Table 5.1: LES case properties: name, type of cloud (shallow cumulus, ShCu; stratocumulus, Sc; deep convective cumulonimbus, Cb), domain size, resolution, in-cloud cloud water path (CWP), cloud top height (CTH), and thermodynamic phase. Shown are ensemble means and standard deviations not accounting for spatial variance within a single ensemble member.

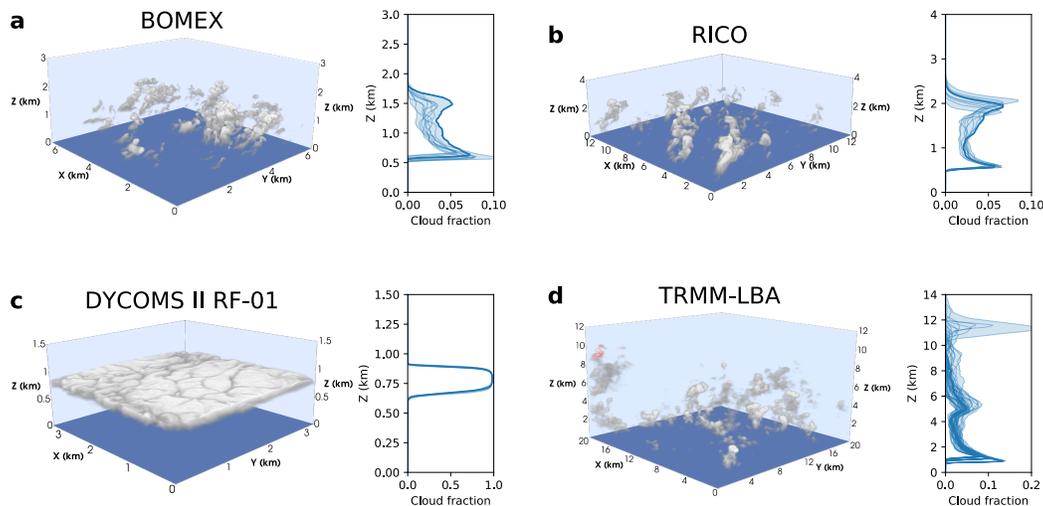


Figure 5.1: Snapshots of LES clouds, showing liquid water specific humidity (gray to white, low to high) and ice water specific humidity (red to white, low to high). Subplots to the right show vertical profiles of cloud fraction for each case. The thick line shows the profile for the specific snapshot in the 3D rendering, the thin lines show all other snapshots, and the shading shows the range. (a) and (b) Shallow convective clouds. (c) Stratocumulus clouds. (d) Deep convective clouds. Note that the domain sizes vary between the cases.

point (45 in total) from the  $(20 \text{ km})^2$  TRMM-LBA simulations and 5 snapshots of fully-developed, more aggregated deep convection from the  $(40 \text{ km})^2$  TRMM-LBA aggregated (agg.) simulations. The smaller ensemble is determined to sufficiently capture the dynamical variability for the larger domain.

The increase in convective aggregation for the larger domain Cb simulations can be seen in typical measures such as the variance of the column relative humidity or total precipitable water (Wing et al., 2017) (see appendix A, Fig. A1). The domain-mean cloud cover, cloud top height, and cloud water path from the two sets of Cb simulations are similar, indicating that the difference in radiative flux bias is being driven by a change in the aggregation or domain size. Larger domains may lead to even more aggregation (Patrizio and Randall, 2019); however, synoptic noise may become important and disrupt the self-aggregation of convection on large scales in reality (Bretherton, 2015). The Sc and ShCu results are unchanged for larger domain sizes (not shown), but we do see an expected reduction in variance across the ShCu ensemble ( $N_{\text{LES}} = 10$ ) due to the larger dynamical variability captured in each snapshot of the larger domain.

### **Radiative transfer computations**

The RT calculations were done using the libRadtran software package with the MYSTIC Monte Carlo solver (Mayer and Kylling, 2005; Mayer, 2009; Emde et al., 2016). Details of the set-up can be found in appendix B. The MYSTIC solver requires 3D fields of liquid and ice water content and particle effective radius as input. We use MYSTIC to do the full 3D RT and ICA calculations. The LES uses simple microphysics schemes that do not explicitly compute the effective radius. To compute the effective radius, we follow the parameterization from Ackerman et al. (2009) and Blossey et al. (2013) for liquid and Wyser (1998) for ice (appendix B). For the RT calculation, MYSTIC finds the scattering phase function from pre-computed lookup tables. In the case of liquid droplets, which are assumed spherical, the full Mie phase function is used. For the case of ice clouds, a parameterization of the habit-dependent scattering must be used. We use the hey parameterization with “general habit mixture” (Yang et al., 2013; Emde et al., 2016). The results are insensitive to the choice of ice parameterization (Fig. B1) because the reflected flux signal is dominated by the liquid droplets for the clouds we simulated.

### **Observations of cloud climatology**

The LES cloud fields allow for precise calculation of the 3D cloud radiative effect on small domains. To estimate the global impact of the 3D cloud radiative effect, we use the results from LES along with satellite observations of cloud climatology and surface albedo to scale up from these few cases to a global picture. We find that in-cloud cloud water path (CWP), defined as the domain-mean cloud water path divided by cloud cover, is a simple but robust predictor of the flux bias (will be shown in Section 5). We use the International Satellite Cloud Climatology Project (ISCCP) D2 dataset of CWP (W. B. Rossow et al., 1999; W. Rossow and Duenas, 2004; Marchand et al., 2010; C. Stubenrauch et al., 2012; C. J. Stubenrauch et al., 2013). The ISCCP D2 cloud product is a monthly climatological mean with spatial resolution of  $1^\circ \times 1^\circ$  constructed from measurements during the period 1984–2007. These data are collected by a suite of weather satellites that are combined into a 3-hourly global gridded product at the D1 level and are averaged, including a mean diurnal cycle, into the D2 product we use.

We also account for the observed surface albedo that varies seasonally and spatially and affects the flux bias. We use observations of surface albedo from the Global Energy and Water Exchanges Project’s surface radiation budget product version 3.0, which is aggregated to a monthly mean climatology for the period 1984–2007 and

gridded to  $1^\circ \times 1^\circ$ .

### 5.3 Radiative flux bias dependence on zenith angle

#### Top-of-atmosphere

The top-of-atmosphere (TOA) radiative flux bias is measured (in  $\text{W m}^{-2}$ ) as the difference in reflected irradiance between the ICA and 3D RT calculations averaged over the full domain. A positive bias means that, under the ICA, clouds reflect more energy back to space than in reality (i.e., 3D), implying that the Earth's surface is artificially dimmed (cooled) in a model that uses the ICA. The albedo bias ( $\Delta\alpha$ ) is computed as the flux bias ( $\Delta F = F_{\text{ICA}} - F_{\text{3D}}$ ) divided by the total incoming solar flux ( $F_{\text{in}}$ ),

$$\Delta\alpha = \frac{\Delta F}{F_{\text{in}}} \times 100\%. \quad (5.1)$$

Fig. 5.2 shows the flux and albedo biases (ICA–3D) for the five cases of ShCu, Sc, and Cb clouds. The solid lines show the ensemble mean bias, and the shading denotes one standard deviation ( $\sigma$ ). The combined variance ( $\sigma^2$ ) is computed as,

$$\sigma^2 = \frac{1}{N_{\text{LES}}} \sum_{i=1}^{N_{\text{LES}}} \left[ \left( \sigma_{i,\text{ICA}}^2 + \sigma_{i,\text{3D}}^2 \right) + (\Delta F_i - \langle \Delta F \rangle)^2 \right] \quad (5.2)$$

where  $N_{\text{LES}}$  is the number of ensemble members,  $\sigma_{i,\text{ICA}}$  and  $\sigma_{i,\text{3D}}$  are the standard deviations from the MYSTIC solver photon tracing,  $\Delta F_i$  is the TOA flux bias of each ensemble member, and  $\langle \cdot \rangle$  denotes a mean over the LES ensemble. This variance includes both the statistical noise from the Monte Carlo RT and the dynamical variability of the cloud field (which are assumed to be uncorrelated). The Monte Carlo noise is proportional to  $1/\sqrt{n}$  where  $n = 10^4$  is the number of photons used for the RT simulation, and is in fact  $\sim 0.7\%$  for these calculations. The variance between cloud scenes is much larger than the Monte Carlo error, by more than an order of magnitude.

Sc show negligible deviation between ICA and 3D reflected fluxes. For convective clouds (ShCu and Cb), the bias from the ICA is positive, except for ShCu at very large solar zenith angles. At large zenith angles, ShCu show a large negative flux and albedo bias for ICA. ShCu scatter far fewer photons than Cb due to the low cloud cover and their smaller optical thickness, corresponding to small vertical extent. Cb exhibit the largest reflected irradiance and also the largest bias between the ICA and 3D RT calculations. While the mean flux bias is similar, the structure of the bias with zenith angle is markedly different for the two domain sizes (Fig. 5.2). For the small-domain simulations with a lesser degree of aggregation, the bias is

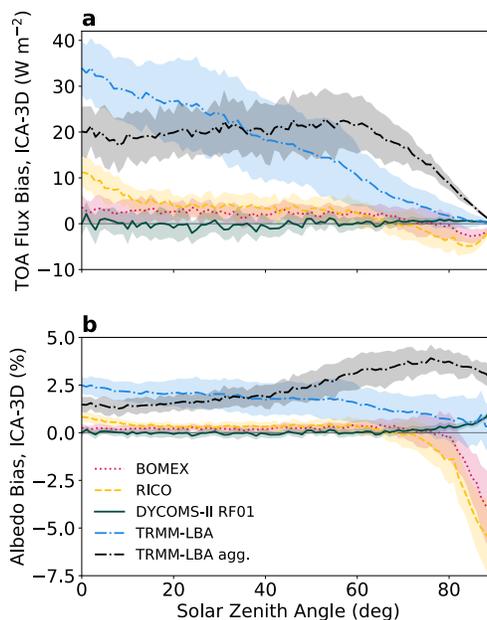


Figure 5.2: Bias (ICA-3D) in (a) TOA reflected flux and (b) albedo as a function of zenith angle for ShCu (BOMEX and RICO), Sc (DYCOMS-II RF01), and Cb (TRMM-LBA and TRMM-LBA agg.). For each cloud type, average fluxes (with shaded  $1\sigma$  error bars) are computed over the individual snapshots. Positive bias means the ICA approximation is reflecting more incoming flux than in the 3D RT calculation.

approximately linear with zenith angle (as seen by Barker, Cole, Li, Yi, et al. (2015) and Barker, Cole, Li, and Salzen (2016)). For the more aggregated case, the flux bias is nearly uniform up until a solar zenith angle of  $60^\circ$  and then decreases rapidly towards zero; this translates to an albedo bias that peaks at large zenith angles (around  $70^\circ$ ).

The convective clouds show much more variation than the stratiform clouds between snapshots due to the variability in cloud cover even in a statistically steady state. The less aggregated Cb clouds have the largest variability, which is expected since the domain size is small relative to the scale of the clouds, i.e., in each snapshot we capture only approximately one deep convective cloud, compared to many small cumulus clouds; therefore, we are effectively averaging over fewer realizations even though we take our ensemble size to be larger. Similarly, for the more aggregated Cb clouds, since we use a four times larger domain, a smaller ensemble ( $N_{LES} = 5$  compared to 15) is large enough to capture the variability.

In the ICA, the horizontal photon fluxes between neighboring columns are ignored.

For the Sc clouds that uniformly cover the whole domain (Fig. 5.1c), this assumption has little effect: the flux bias is near zero for all zenith angles. However, for cumulus clouds, making the ICA has two effects that are described in detail by Hogan, Fielding, et al. (2019).

1. The long-recognized “cloud-side illumination” effect in 3D radiative transfer. This describes how horizontally propagating photons can encounter the side of a cloud and can be scattered by it, rather than being restricted to hit the top of a cloud in the ICA. Side-illumination happens when photons travel across columns at slant angles, brightening the cloud sides and enhancing cloud reflectance; it also creates larger shadows, or larger effective cloud cover. This effect acts to enhance reflectance in 3D, and thus would appear as a negative ICA flux bias in our terminology.
2. The newer “entrapment” effect that Hogan, Fielding, et al. (2019) presented. This mechanism is similar to the “upward trapping” mechanism discussed by Várnai and Davies (1999). It describes how in 3D a scattered photon may be intercepted by another cloud, or the same cloud, in a different column higher in the domain and scattered back down to the surface. In the ICA by contrast, when a photon travels through clear-sky and is scattered by a cloud, it will necessarily travel back through the same column of clear-sky to the TOA. The entrapment mechanism acts to decrease cloud reflectance in 3D, i.e., it creates a positive flux bias.

The calculated 3D effects we show in Fig. 5.2 are a combination of these competing mechanisms. At large solar zenith angles, cloud shadowing, by which clouds can shade each other, clear-sky regions, and the surface when photons are coming in at slant angles, can be important for surface irradiance and surface fluxes (Frame et al., 2009; Veerman et al., 2020).

For small zenith angles, when the sun is overhead, the convective clouds (ShCu and Cb) produce a positive flux bias because entrapment is dominant over cloud-side illumination. For large zenith angles, the flux and albedo bias from ShCu is negative because cloud-side illumination becomes the dominant effect. In the mean, the solar zenith angle at which the flux bias becomes negative is around  $70^\circ$ , but for the individual ensemble members this ranges from around  $45^\circ$  to  $75^\circ$ . This has been seen before for ShCu by Barker, Cole, Li, Yi, et al. (2015) and Barker, Cole, Li, and Salzen (2016) and Hogan, Fielding, et al. (2019). For Cb clouds,

however, even at large zenith angles, the flux and albedo biases remain positive, indicating that the entrainment mechanism continues to dominate over cloud-side illumination. This is not the case for every scene in the Cb ensemble, but it is true in the mean, in agreement with the results from Hogan, Fielding, et al. (2019). This difference between ShCu and Cb is related to the aspect ratio of the clouds; the cloud-side illumination mechanism can only become dominant if the aspect ratio is small (clouds are not too deep). Furthermore, in the case of the more aggregated Cb clouds, a greater degree of aggregation decreases the surface area to volume ratio of the clouds, or what Schäfer et al. (2016) call the length of cloud edge, or cloud perimeter. A smaller cloud perimeter will decrease the cloud side illumination as well as the entrainment efficiency of the cloud (Hogan, Fielding, et al., 2019). The uncertainty in flux bias due to the degree of aggregation of deep convection is much larger than the spread across the LES ensemble and represents a structural uncertainty, which is more challenging to quantify.

These 3D cloud effects can be understood from Fig. 5.1, which shows illustrations of the clouds from the four LES cases. The scattered cumulus in the BOMEX and RICO cases are shallow and spaced apart, which allows for cloud-side illumination at large zenith angles to dominate over the entrainment mechanism. The DYCOMS-II RF01 stratocumulus clouds are quite horizontally homogeneous over this small domain, therefore, ICA biases are small. As discussed in Hogan, Fielding, et al. (2019), when in-cloud heterogeneity is larger, e.g., for open-celled marine stratocumulus, the entrainment effect is larger. Finally, for the deep TRMM-LBA clouds, the entrainment mechanism remains dominant even for large zenith angles because the clouds at higher levels can intercept and trap outgoing photons that are able to escape to TOA in the ICA.

In addition to the LES ensembles described previously, we run one additional set of tests to quantify the dependence of the flux bias calculations on the LES resolution (Fig. 5.3). We take the original LES simulations and systematically coarse-grain the 3D fields to lower resolution. Doing so ensures that we do not change the dynamics of the clouds so that we can test the effect of resolution on only the radiative transfer. We are not able to bridge the gap all the way to ESM scales (10–100 km horizontal resolutions) due to computational limits on running the LES, but we show results across a range of horizontal scales. When coarse-graining, we keep the vertical resolution fixed to better represent the very large aspect ratio grid boxes found in ESMs compared to the relatively isotropic grid boxes in LES. The

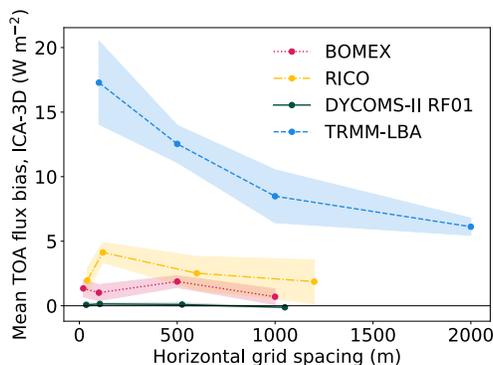


Figure 5.3: Mean TOA reflected flux bias across all solar zenith angles computed for different resolutions of the same cloud fields. The horizontal axis shows the horizontal resolution; the vertical resolution is kept fixed. The four cases of ShCu, Sc, and Cb are shown in the same colors as Fig. 5.2. For each case, three snapshots from the original ensemble are used and the spread is shown by the shading.

mean TOA flux bias is nearly constant across resolutions for the shallow clouds (Sc and ShCu). For Cb, the mean TOA flux bias decreases with larger grid spacing, as expected, from around  $17 \text{ W m}^{-2}$  at the original resolution and down to  $6 \text{ W m}^{-2}$  for 2 km horizontal resolution. Since the bias does not asymptote as we move towards smaller horizontal grid spacing, we expect that if the LES were run at even higher resolutions, we would find an even larger larger bias between the ICA and 3D. We conclude that our estimated bias is a lower bound in this regard.

## Surface

Using the same radiative transfer calculations, we also quantify the bias in downwelling surface irradiance. This bias as a function of solar zenith angle is shown for each cloud type in Fig. 5.4. The total bias (Fig. 5.4a) is the sum of a direct component and diffuse component, shown in Fig. 5.4b and c, respectively. These components largely offset each other, as has been found in previous studies (Gristey et al., 2019). The direct surface irradiance bias is always positive and the diffuse always negative, resulting from side-illumination (shadowing) and entrapment, respectively. Note that these two mechanisms have opposite effects in the surface irradiance bias and the TOA reflected flux bias.

For the total surface irradiance bias (Fig. 5.4), we observe a similar pattern to the TOA bias (Fig. 5.2), except with the opposite sign. This has been noted before (Barker, Cole, Li, Yi, et al., 2015), and is to be expected given that when the ICA

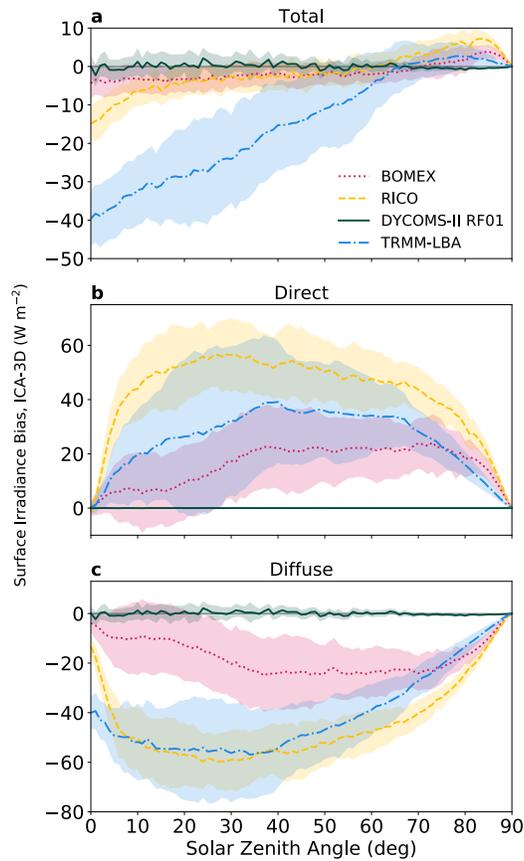


Figure 5.4: Surface irradiance bias (ICA-3D) as function of zenith angle for ShCu (BOMEX and RICO), Sc (DYCOMS-II RF01), and Cb (TRMM-LBA). The total surface irradiance bias (a) is split into the direct (b) and diffuse (c) components, which largely compensate each other, especially at larger zenith angles. For each cloud type, average fluxes (with shaded  $1\sigma$  error bars) are computed over the individual snapshots. Positive bias means the ICA approximation has more downwelling radiation at the surface than the 3D calculation.

produces anomalous extra TOA reflectance, it simultaneously decreases the surface irradiance with respect to the 3D calculation. We can quantify this by considering in a simple way how the TOA reflected flux and surface irradiance depend on the incoming flux, surface albedo ( $\alpha_s$ ), cloud albedo ( $\alpha_c$ ), and cloud cover ( $f_c$ ). The total scene albedo stems from scattering by the clouds and scattering by the surface. Considering up to two scattering events, we can write

$$\alpha = f_c \alpha_c + (1 - f_c) \alpha_s + 2 f_c (1 - f_c) (1 - \alpha_c) \alpha_s.$$

The first term comes from reflection directly from the clouds, the second from reflection directly from the surface, and the third from reflection of diffuse radiation

from the surface. The albedo bias is therefore

$$\Delta\alpha = f_c\Delta\alpha_c(1 - 2\alpha_s(1 - f_c)), \quad (5.3)$$

where  $\Delta\alpha = \alpha_{\text{ICA}} - \alpha_{\text{3D}}$ . From Eq. 5.3 we see that the albedo bias will decrease with surface albedo because when the surface accounts for a larger fraction of the total albedo the cloud bias is less pronounced. For the downwelling surface irradiance ( $I$ ), we can do the same and consider up to two scattering events,

$$I = [(1 - f_c) + f_c(1 - \alpha_c) + f_c(1 - f_c)\alpha_c\alpha_s] F_{\text{in}},$$

where the first term comes from direct irradiance, the second from forward scattering through the cloud, and the third from multiple scattering first off the surface and then back down off the cloud. To first order the surface irradiance does not depend on the surface albedo, but including higher order terms we see that the surface irradiance increases with surface albedo. The surface irradiance bias ( $\Delta I = I_{\text{ICA}} - I_{\text{3D}}$ ) can be written as

$$\Delta I = -f_c\Delta\alpha_c(1 - \alpha_s(1 - f_c))F_{\text{in}}. \quad (5.4)$$

Finally, with these approximations, and noting that  $\Delta F = \Delta\alpha F_{\text{in}}$ , we can relate the surface irradiance bias to the TOA reflected flux bias by

$$\Delta I = -\Delta F \left( \frac{1 - \alpha_s(1 - f_c)}{1 - 2\alpha_s(1 - f_c)} \right). \quad (5.5)$$

Fig. 5.5 shows the calculated surface irradiance bias compared to the predicted bias calculated from the TOA bias via Eq. 5.5. The relation between surface irradiance bias and TOA reflected flux bias is robust and predictable, with an  $r^2 = 0.92$ . Eq. 5.5 considers up to two-scattering events, but considering only one scattering event, which simplifies the equation to  $\Delta I = -\Delta F$ , does not significantly change the quality of the prediction ( $r^2 = 0.91$  in this case).

While the surface irradiance bias is critically important for surface fluxes, which may have dynamical feedbacks on the clouds, or for impacts on vegetation and carbon uptake (Veerman et al., 2020), the relevant quantity for the overall climate system energetics is the TOA energy balance. The remainder of this paper is devoted to analyzing the TOA reflected flux bias and the relevance of 3D cloud radiative effects for climate.

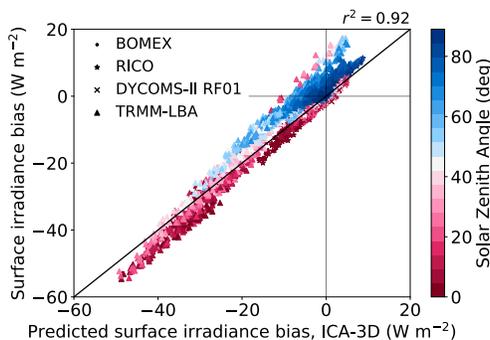


Figure 5.5: Bias (ICA-3D) in surface irradiance as predicted by bias in TOA reflected flux via Eq. 5.5 compared to the computed surface irradiance bias. Cloud type and zenith angle are indicated by marker shape and color, respectively. The 1:1 line is shown for reference.

#### 5.4 Seasonal cycle of radiative flux bias

The solar zenith angle varies over the course of the day from sunrise to sunset, and therefore the dependence of the bias on zenith angle manifests itself as a diurnally varying bias. But the zenith angle also varies on seasonal timescales for different latitudes.

To assess the climate impact of the ICA bias, we consider the flux and albedo bias for each cloud type as a function of day of year and latitude. This calculation is done by assuming that the LES-generated cloud field is present at any given latitude circle on any given day of the year. This exercise is done without a claim to be realistic, but to demonstrate the impact each cloud type might have on Earth given the spatiotemporal variations of solar zenith angle. For any location and time, including a diurnal cycle, the solar zenith angle is calculated and the flux bias is estimated based on the results presented in Fig. 5.2. The flux and albedo biases are computed hourly and averaged to show the daily-mean bias.

Fig. 5.6 shows the annual mean and seasonal cycle of TOA flux and albedo biases for each cloud type. To estimate the uncertainties of the annual-mean bias, we calculate the LES ensemble spread as follows. For each hour in the year and each latitude, the solar zenith angle is calculated, and we interpolate between integer zenith angles in the flux bias calculations to find the mean flux bias. This is done individually for each LES cloud scene in the ensemble. The ensemble mean for each latitude and day of the year is shown (colored contour plots in Fig. 5.6), as well as the annual mean of the ensemble (black lines on Fig. 5.6). The spread across the ensemble in

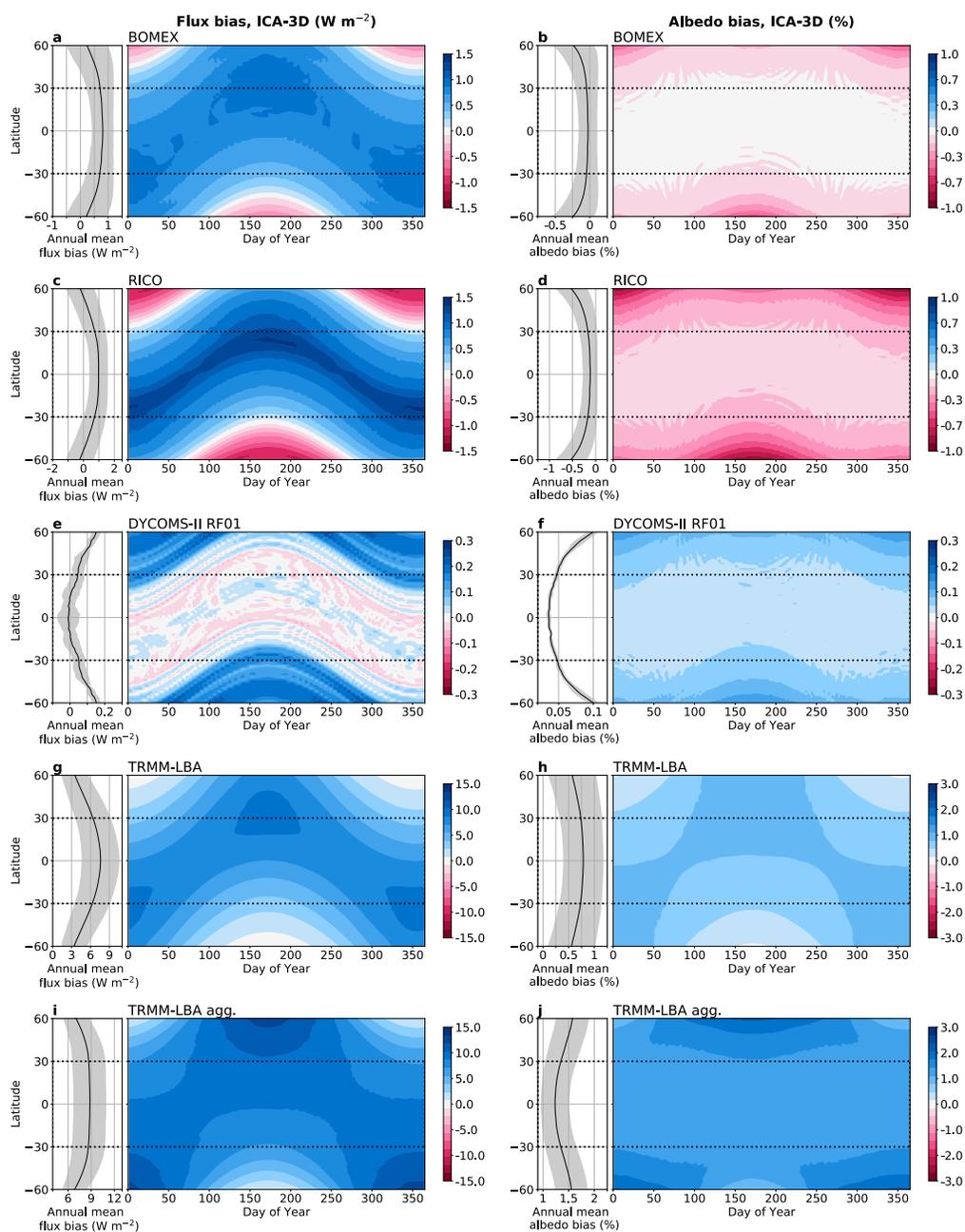


Figure 5.6: Daily mean bias (ICA-3D) as a function of latitude and day of year assuming the globe is covered by (a-d) ShCu (BOMEX and RICO), (e-f) Sc (DYCOMS-II RF01), (g-h) Cb (TRMM-LBA), and (i-j) more aggregated Cb (TRMM-LBA agg.). Left column shows flux bias, and right columns shows albedo bias. Note the color scales vary between LES cases. Inset panels on the left show annual average biases with shaded error bars that denote the spread across the LES ensembles as described in the text. Only latitudes from 60°S to 60°N are shown because the simulated clouds are not representative of the high-latitude regions.

the annual mean is shown as one standard deviation (gray shading on Fig. 5.6).

Both ShCu cases show similar patterns of flux bias with latitude and time (Fig. 5.6a and c). As seen in Fig. 5.2, these cases both have a small positive bias for small solar zenith angles, transitioning to a small negative bias for larger zenith angles, which is manifest here as a positive bias at low latitudes, transitioning to a negative bias only in mid-latitude winters. The albedo bias for both ShCu cases is near zero in the tropics and becomes more negative at higher latitudes. Sc show a very small flux (and albedo) bias for all solar zenith angles due to their high cloud cover and horizontal homogeneity, but they do exhibit a small positive flux bias ( $\sim 0.2 \text{ W m}^{-2}$ ) during winter in mid-latitudes (Fig. 5.6e). For Cb, the flux bias is comparatively large and always positive (Fig. 5.2). In the less aggregated state, the flux bias is nearly linear in zenith angle, which gives rise to a bias pattern that roughly mimics the insolation pattern with latitude and day of year (Fig. 5.6g). In the more aggregated state, the flux bias is roughly constant in the tropics and overall larger than in the less aggregated case (Fig. 5.6i). The albedo bias for Cb is largest and positive during summer, though seasonal variations are less pronounced for aggregated convection (Fig. 5.6h and j). In particular, deep convective clouds are frequently found in the ITCZ, which migrates with the insolation maxima and therefore results in a tropical TOA reflected flux bias that peaks in each hemisphere during their respective summers and is smallest during the shoulder seasons (Fig. 5.6g and i).

## 5.5 Implications for Climate Models

To make an assessment of the effect that the 3D radiative transfer through cloud fields has on climate simulated with ESMs, we must account for the climatological occurrence of different cloud types in space and time. A simple parameter that can account for much of the flux bias variability in our calculations is in-cloud cloud water path (CWP), defined as domain-mean cloud water path divided by cloud cover. By regressing the flux bias against CWP for integer solar zenith angles between  $0^\circ$  and  $90^\circ$ , constraining the regression lines to pass through the origin because there is no flux bias in clear-sky conditions ( $\text{CWP} = 0$ ), we observe a robust positive correlation between CWP and flux bias (Fig. 5.7). The best fit line and confidence intervals are estimated with Gaussian process regression; we use a dot product kernel, with the intercept constrained to zero. We apply regularization by specifying the “nugget” (the values added to the diagonal of the correlation matrix) as the empirically calculated variance scaled by a constant factor. The variance is calculated as the sample variance in a  $100 \text{ g m}^{-2}$  CWP interval around each point.

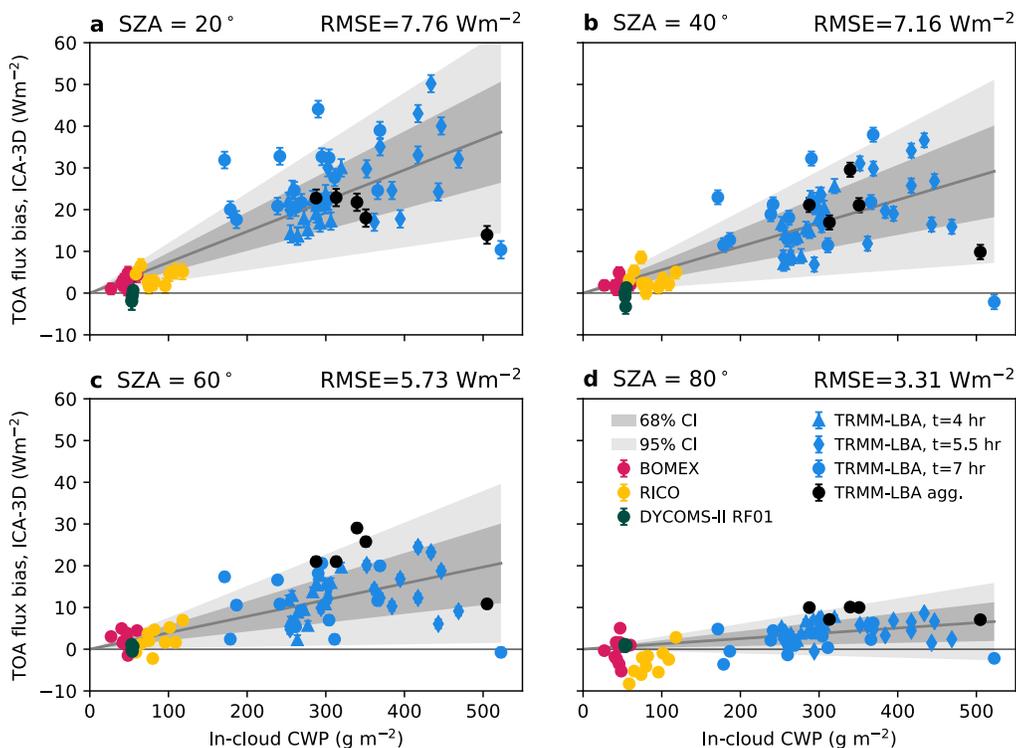


Figure 5.7: Scatter plot of in-cloud cloud water path (CWP) from LES domain against flux bias at zenith angles (a)  $20^\circ$ , (b)  $40^\circ$ , (c)  $60^\circ$ , and (d)  $80^\circ$ . LES ensemble members are plotted with the same color convention as in Fig. 5.2. The grey lines show the regression fit constrained to go through the origin. The gray shaded areas show the 68% and 95% confidence intervals. The rms error of the regression is indicated at the top.

The positive correlation between CWP and flux bias, though not perfect, allows us to approximate TOA flux biases using CWP on the global scale. We choose CWP as our proxy for flux bias because it is robustly observed by satellite and, among the other cloud properties we explored (e.g., cloud top height), the best predictor for flux bias (Fig. C1). Despite the fact that the radiative flux bias certainly depends on more than just CWP, we use it here as a first approximation to model the flux bias.

Using this relationship between CWP and flux bias for a series of zenith angles, we can use the observed climatological CWPs from ISCCP to infer the resulting flux bias that would be associated with using the ICA for RT calculations in place of 3D RT. The monthly temporal resolution is not inherently an issue for this analysis given that we use a linear relationship between CWP and flux bias.

Additionally, we may account for the variations in surface albedo. In the RT

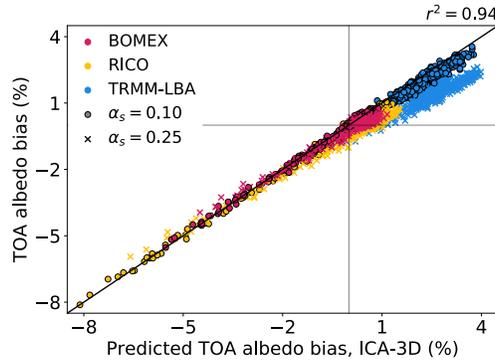


Figure 5.8: Predicted albedo bias from Eq. 5.6 compared to the computed albedo bias for convective cloud scenes (BOMEX, RICO, and TRMM-LBA) with different surface albedos. The predicted surface albedo is calculated from the simulations using a surface albedo of  $\alpha_O = 0.06$  corresponding to an ocean surface. Each point represents the albedo bias at integer solar zenith angle from  $0^\circ$  to  $90^\circ$  for 5 ensemble members of each LES case. The colors denote the different cases and the symbol shapes denote the surface albedo. The 1:1 line is shown for reference.

calculations previously shown, we assume a constant surface albedo of  $\alpha_O = 0.06$ , corresponding to an ocean surface. The surface albedo affects the computed TOA and surface flux biases as shown in Eqs. 5.3 and 5.4. The albedo bias, written in Eq. 5.3, scales with the factor  $(1 - 2\alpha_s(1 - f_c))$ . We can therefore correct for the effect of the surface albedo by multiplying our computed flux or albedo bias by the ratio of the surface absorptions:

$$\Delta\alpha|_{\alpha_s} = \left( \frac{1 - 2\alpha_s(1 - f_c)}{1 - 2\alpha_O(1 - f_c)} \right) \Delta\alpha|_{\alpha_O}. \quad (5.6)$$

We run additional RT calculations for the convective cloud cases (BOMEX, RICO, and TRMM-LBA) and vary the specified surface albedo from  $\alpha_s = 0.1$  to  $0.25$  (spanning the range for sub-polar land surfaces). Fig. 5.8 shows the correlation between the predicted TOA albedo bias using Eq. 5.6 and the explicitly calculated TOA albedo bias. The prediction based on Eq. 5.6 is robust with an  $r^2$  value of 0.94. For stratocumulus clouds, since the cloud cover is nearly 1, the albedo bias depends very little on the surface albedo and is not shown. We use Eq. 5.6 globally to account for variations in the observed surface albedo without the need to run additional RT calculations and interpolate between discrete values.

To construct the annual-mean flux bias map shown in Fig. 5.9, we first calculated the solar zenith angle for each location on Earth and each hour of the year. Then,

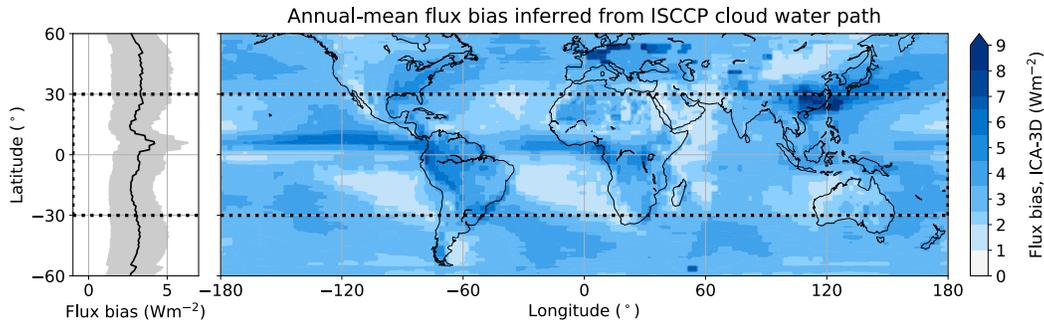


Figure 5.9: Map of annual-mean flux bias inferred from ISCCP in-cloud CWP. Left panel shows the zonally averaged flux bias in the black line and  $1\sigma$  error bars in the grey shading that are derived from the linear regression in Fig. 5.7. Like in Fig. 5.6, we show only  $60^\circ\text{S}$  to  $60^\circ\text{N}$  because the clouds we have modeled are not representative of high-latitudes.

we obtained the flux bias given the observed CWP from the linear regression at the given zenith angle (Fig. 5.7). Finally, we made a correction using Eq. 5.6, based on the ratio of the observed surface absorption to the assumed ocean surface absorption used in the MYSTIC RT calculations. The resulting flux bias is an estimate of the bias that would be present in an ESM that is able to resolve the relevant dynamical scales of clouds, but makes the ICA during radiative transfer. This bias is smaller than the bias present in current ESMs, which also contains the biases due to PPA and cloud parameterizations, given their very coarse horizontal resolution (Cole, Barker, Randall, et al., 2005).

We focus on the tropics ( $30^\circ\text{S}$  to  $30^\circ\text{N}$ , dotted box on Fig. 5.9), where our estimation of flux bias based on the LES cases is most robust and relevant; for higher-latitudes, we do not capture all the relevant cloud regimes with our sample of LES clouds, and so our flux bias estimate needs to be interpreted with caution. Shown in the left inset plot is the zonal-mean flux bias. The shading represents  $1\sigma$  error from the regression of flux bias on CWP shown in Fig. 5.7 (as opposed to spatial or temporal variability).

The largest bias occurs in the ITCZ region and the storm track regions, especially over eastern Asia where the climatological CWP is maximal (Fig. 5.9). It corresponds to locations where the tallest clouds on Earth exist and where the mean zenith angle is smallest. The region of maximum bias migrates seasonally following the location of the ITCZ (and maximum insolation). Seasonal variations in cloud cover and cloud type are also manifest in the seasonal cycle of the 3D flux bias. In the annual

mean, the zonal-mean tropical flux bias is estimated to be  $3.1 \pm 1.6 \text{ W m}^{-2}$ , and the maximum local flux bias in the annual mean is around  $6.5 \text{ W m}^{-2}$  (99th percentile). The annual-mean, zonal-mean tropical albedo bias is  $0.7 \pm 0.4\%$  and is locally as large as  $1.5\%$  (99th percentile).

Our results are of the same order as those reported in Cole, Barker, O’Hirok, et al. (2005) and Barker, Cole, Li, Yi, et al. (2015) and Barker, Cole, Li, and Salzen (2016). Cole, Barker, O’Hirok, et al. (2005) also found the largest flux bias occurring over the ITCZ region, with a maximum bias of  $5 \text{ W m}^{-2}$  and tropical zonal-average bias of  $1.5 \text{ W m}^{-2}$  during the boreal winter. The larger value reported here is likely due to the fact that Fig. 5.9 averages over the shoulder seasons and the regression is based on higher-resolution cloud scenes, as quantified in Figures 5.3 and 5.6.

## 5.6 Summary and conclusions

In this paper we estimated the TOA flux and albedo biases that result from neglecting 3D radiative transfer through cloudy atmospheres. Although TOA short-wave radiative flux biases in current ESMs are predominantly due to deficiencies of subgrid-scale dynamical parameterizations that generate cloud cover biases, as convection parameterizations improve and model resolution increases, the relative contribution of 3D radiative effects to the total model error will increase. We have quantified the radiative flux and albedo bias that results from making the ICA by using a 3D Monte Carlo radiative transfer scheme applied to LES-generated 3D cloud fields. The flux and albedo biases were assessed across different cloud regimes and solar zenith angles. We took our findings from four canonical LES cases and applied them to observed climatological cloud occurrence to infer the spatially- and temporally-resolved flux and albedo biases.

We find that the largest flux bias comes from deep convective clouds at small solar zenith angles. The albedo bias is large and negative for shallow cumulus clouds at large solar zenith angles. These results quantitatively agree with previous studies using LES clouds to assess 3D effects (Hogan, Fielding, et al., 2019). There is room for future work considering a larger ensemble of cloud morphologies, which could be generated again by LES or alternatively could be retrieved from satellite observations. Our inferred global flux bias is based on only four tropical/subtropical LES cases and therefore does not represent the full diversity of extratropical cloud morphologies. This methodology cannot fully capture the effects of mid-latitude storms, for instance, which is why we do not emphasize our results outside of the

tropics.

We use the observed positive correlation between CWP and TOA flux bias from our LES ensemble to estimate the global spatiotemporal bias from neglecting 3D radiative transfer in a high-resolution ESM. We choose a simple linear model to map from satellite observations of climatological CWP to TOA flux bias. The deviations in our regression fit suggest that there is potential for a more robust mapping from cloud properties to radiative flux bias. Future work is necessary to explore this path towards a parameterization of 3D radiative effects in ESMs.

The large flux bias for Cb clouds at small solar zenith angles translates into a seasonal bias that peaks just off the equator in the summer hemisphere, tracking the position of the ITCZ. We estimate the annual-mean tropical-mean flux bias to be  $3.1 \pm 1.6 \text{ W m}^{-2}$ . The flux bias computed here is small compared to the TOA shortwave flux errors typical for CMIP5 and CMIP6 models, which are on the order of  $10 \text{ W m}^{-2}$  in the mean (Zhao et al., 2018; Hourdin et al., 2020) and can reach  $50 \text{ W m}^{-2}$  in stratocumulus regions (Brient et al., 2019). However, the 3D bias is still comparable to the signal of anthropogenic greenhouse gas emissions for the coming decades, which is on the order of  $2.5\text{--}3.1 \text{ W m}^{-2}$  (Myhre et al., 2013). These results highlight the importance of considering the 3D radiative fluxes through clouds for Earth's radiation budget and Earth system modeling.

**Acknowledgements.** C.E.S. acknowledges support from NSF Graduate Research Fellowship under Grant No. DGE-1745301. I.L. is supported by a fellowship from the Resnick Sustainability Institute at Caltech. This research was additionally supported by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We thank three anonymous reviewers for helpful comments on an earlier version of this paper.

**Data availability statement.** All code or data used in this paper are freely available online. The LES were run using the PyCLES code (<https://climate-dynamics.org/software/#pycles>). The radiative transfer computations were done using the libRadtran code (<http://www.libradtran.org>). Post-processed LES 3D fields used as input files for libRadtran computations are available in Singer et al. (2021). The ISCCP data were downloaded from <https://climserv.ipsl>.

polytechnique.fr/gewexca/ and the GEWEX albedo measurements were downloaded from <https://eosweb.larc.nasa.gov>.

### 5.7 Appendix A: LES model setup

LES are performed using the anelastic fluid solver PyCLES (Pressel, Kaul, et al., 2015). Subgrid-scale fluxes are treated implicitly by the WENO scheme used in the numerical discretization of the equations (Pressel, Mishra, et al., 2017).

For each case, the characteristic timescale of convection is evaluated and taken to be representative of the dynamical decorrelation time  $\tau$ . Snapshots are taken at least one dynamical decorrelation time apart, so that the cloud samples can be treated as independent in a statistical analysis of the flux biases. The decorrelation timescale is calculated as

$$\tau = \frac{z_{bl}}{w^*} + \frac{d_c}{\bar{w}_u}, \quad (5.7)$$

where  $z_{bl}$  is the mixed-layer height,  $w^* = \left( z_{bl} \overline{w'b'} \Big|_s \right)^{1/3}$  is the Deardoff convective velocity,  $d_c$  is the cloud depth, and  $\bar{w}_u$  is the mean updraft velocity within the cloud.

#### Shallow cumulus (ShCu) convection, BOMEX

The BOMEX LES case study is described in Siebesma et al. (2003). Surface boundary conditions,  $\overline{w'q'_l} \Big|_s$  and  $\overline{w'\theta'_l} \Big|_s$  are prescribed, resulting in sensible and latent heat fluxes of about 10 and 130 W m<sup>-2</sup>, respectively. The atmospheric column is forced by clear-sky longwave radiative cooling, neglecting radiative cloud effects. A prescribed subsidence profile induces mean vertical advection of all fields, and specific humidity is further forced by large-scale horizontal advective drying in the lower 500 m. The liquid-water specific humidity is diagnosed through a saturation adjustment procedure. For BOMEX, the characteristic timescale of convection is  $\tau \approx 40$  min, where  $z_{bl} = 500$  m,  $w^* = 0.66$  m s<sup>-1</sup>,  $d_c = 1300$  m, and  $\bar{w}_u = 0.85$  m s<sup>-1</sup>, and snapshots are taken every 1 hour. The domain size is set to 6.4 km in the horizontal and 3 km in the vertical. Results are reported for an isotropic resolution of  $\Delta x_i = 20$  m.

#### Shallow cumulus (ShCu) convection, RICO

The RICO LES case study is described in vanZanten et al. (2011). The surface sensible and latent heat fluxes are modeled using bulk aerodynamic formulae with drag coefficients as specified in vanZanten et al. (2011), resulting in fluxes of around 6 and 145 W m<sup>-2</sup>, respectively. The atmospheric column is forced by

prescribed profiles for subsidence and large-scale heat and moisture forcings that are a combination of radiative and advective forcings. The two-moment cloud microphysics scheme from Seifert and Beheng (2006) is used with cloud droplet concentration set to  $N_d = 70 \text{ cm}^{-3}$ . For RICO, the characteristic timescale of convection is  $\tau \approx 50 \text{ min}$ , where  $z_{bl} \approx 500 \text{ m}$ ,  $w^* \approx 0.62 \text{ m s}^{-1}$ ,  $d_c = 2500 \text{ m}$ , and  $\bar{w}_u \approx 1.2 \text{ m s}^{-1}$ , and snapshots are taken every 1 hour. The domain size is set to 12.8 km in the horizontal and 6 km in the vertical. Results are reported for an isotropic resolution of  $\Delta x_i = 40 \text{ m}$ .

### **Stratocumulus-topped marine boundary layer (Sc), DYCOMS-II RF01**

The simulation setup for DYCOMS-II RF01 follows the configuration of Stevens et al. (2005). The initial state consists of a well-mixed layer topped by a strong inversion in temperature and specific humidity, with  $\Delta\theta_l = 8.5 \text{ K}$  and  $\Delta q_t = -7.5 \text{ g kg}^{-1}$ . Surface latent and sensible heat fluxes are prescribed as 115 and  $15 \text{ W m}^{-2}$ , respectively. In addition, the humidity profile induces radiative cooling above cloud-top and warming at cloud-base. As in BOMEX, the liquid-water specific humidity is diagnosed through a saturation adjustment procedure. For the stratocumulus clouds, without strong updrafts and a thin cloud layer, the characteristic convective timescale is taken to be just the first term of Eq. (5.7), which evaluates to  $\tau \approx 20 \text{ min}$ , with  $z_{bl} = 850 \text{ m}$  and  $w^* = 0.8 \text{ m s}^{-1}$ . Snapshots taken every 30 minutes are used in the analysis. The domain size is set to 3.36 km in the horizontal and 1.5 km in the vertical. Results are reported for a resolution of  $\Delta z = 5 \text{ m}$  in the vertical and  $\Delta x = \Delta y = 35 \text{ m}$  in the horizontal.

### **Deep convection (Cb), TRMM-LBA**

Deep convective clouds are generated using the TRMM-LBA configuration detailed in Grabowski et al. (2006), based on observations of the diurnal cycle of convection in the Amazon during the rainy season. The diurnal cycle is forced by the surface fluxes, which are prescribed as a function of time. The magnitude of the fluxes maximizes 5.25 hours after dawn, with a peak latent and sensible heat fluxes of 554 and  $270 \text{ W m}^{-2}$ , respectively. The radiative cooling profile is also prescribed as a function of time. We use the one-moment microphysics scheme based on Kaul et al. (2015) with modifications described in Shen et al. (2020). Since this case study is not configured to reach a steady state, the simulation is run up to  $t = 7$  hours. Deep convection is considered to be fully developed after 5 hours, when the liquid-water and ice-water paths stabilize (Grabowski et al., 2006). The ensemble

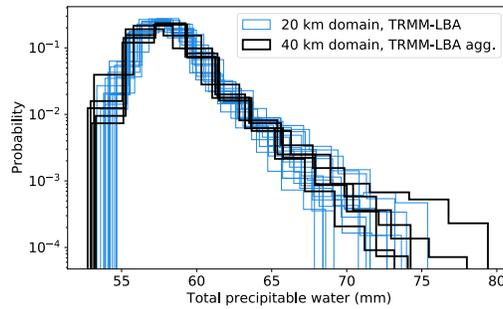


Figure 5.10: Normalized histogram of total precipitable water from the TRMM-LBA simulations in a 20 km domain vs. 40 km domain, which we use as a less and more-aggregated case of deep convection. The variance across the ensemble, shown by the width of the histogram, is representative of the degree of convective aggregation.

of cloud snapshots is formed by sampling after  $t = 4, 5.5,$  and 7 hours from a set of simulations with different initial conditions. For the idealized case (Figs. 5.2 and 5.6) only the 15 snapshots from  $t = 7$  hours are used. The characteristic convective timescale is given by just the second term of Eq. (5.7),  $\tau = \int_0^{z_{ct}} w_u^{-1} dz \approx 80$  min, where  $z_{ct}$  and  $w_u$  are the cloud-top height and updraft vertical velocity averaged over the last two hours, respectively. The random perturbations used in the initialization ensure that all cloud snapshots in the ensemble are uncorrelated. The domain size is set to 20 km in the horizontal and 22 km in the vertical. Results are reported for a resolution of  $\Delta z = 50$  m in the vertical and  $\Delta x = \Delta y = 100$  m in the horizontal.

For the large-domain simulations, we double the domain size to 40 km in the horizontal and run a smaller ensemble of  $N_{LES} = 5$  simulations. The mean cloud cover, cloud top heights, and cloud water path in the large and small domain ensembles are comparable at 0.30 and 0.32, 12.2 and 10.0 km, and 360 and 290  $\text{g m}^{-2}$ , respectively. The large-domain simulations show a higher degree of aggregation as measured by the variance in total precipitable water, 4.3  $\text{mm}^2$ , compared to 3.7  $\text{mm}^2$  in the original 20 km domain. Fig. A1 shows histograms of the total precipitable water for each of the TRMM-LBA simulations at 7 hours ( $N_{LES} = 15$  for the 20 km domain, and  $N_{LES} = 5$  for the 40 km domain). The wider histograms for the large-domain simulations illustrate the larger variance in this field, which is indicative of a higher degree of convective aggregation.

## 5.8 Appendix B: Radiative transfer details

We use the libRadtran MYSTIC Monte Carlo solver for the 3D and ICA radiative transfer calculations with no variation reduction techniques applied. The ICA is done using the `mc_ipa` setting which horizontally averages the fluxes from the full 3D calculation. The RT is done using  $n = 10^4$  photons using the `atlas_plus_modtran` solar spectrum. The atmospheric molecular absorption is done using the `kato2` correlated- $k$  parameterization (Kato et al., 1999; Mayer and Kylling, 2005). The atmospheric profile (pressure, temperature, density, and specific humidity) defined for the radiative transfer comes from the LES output and aerosols are neglected in these calculations. Because we only consider the flux bias and not absolute values, the LES are not embedded in a clear-sky atmosphere with a profile extension to a fixed height, although this is typical for radiative transfer calculations. The surface is treated as a Lambertian scatterer and the albedo was set to  $\alpha_s = 0.06$  for all RT calculations. The observed surface albedo is accounted for through the approximation described in Section 5. Both the LES and RT assume doubly-periodic horizontal boundary conditions and the spatial resolution, which can be found in Table 5.1 for each case, is the same.

The MYSTIC solver from libRadtran requires 3D fields of liquid and ice water content and particle effective radius as input. The LES uses bulk microphysics schemes and does not explicitly compute the effective radius. For liquid-only clouds, the parameterization from Ackerman et al. (2009) and Blossey et al. (2013) with assumed droplet number of  $N_d = 10^8 \text{ m}^{-3}$  is used. The full Mie scattering phase function is taken from the libRadtran lookup tables. Because the lookup tables are only valid for droplets with radius greater than  $1 \mu\text{m}$ , smaller calculated effective radii were rounded to this minimum value.

For ice clouds, the parameterization from Wyser (1998) is used. The `hey` parameterization from Yang et al. (2013) and Emde et al. (2016) with habit type set to `ghm` (general habit mixture) is used. The `hey` parameterization uses the complete scattering phase function as calculated from single scattering models for ice crystals in (Yang et al., 2013), rather than employing an approximation like Henyey-Greenstein phase function, which has been shown to be another source of error in RT (Barker, Cole, Li, Yi, et al., 2015). The results are not dependent on the exact choice for ice crystal shape or roughness (Fig. B1). Note that the `hey` ice parameterization is only valid for radii less than  $90 \mu\text{m}$ , and larger calculated effective radii were rounded to this maximum value.

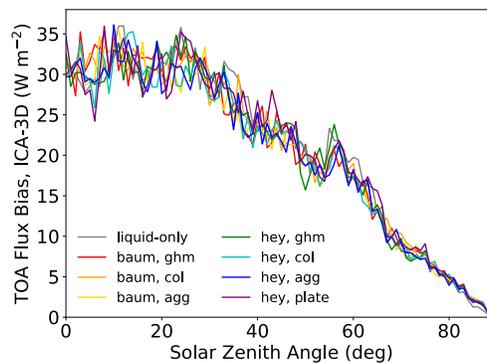


Figure 5.11: TOA reflected bias across zenith angles for different ice parameterizations in one TRMM-LBA cloud snapshot.

Deep convective clouds, reaching upwards of 10 km, nearly always contain ice crystals in addition to liquid water. Optical properties of ice crystals depend on their size, shape (or habit), and surface smoothness. Two different parameterizations, with three and four habit choices, respectively, were tested. The differences between these parameterization variants are negligible; they are much smaller than the variability stemming from the cloud dynamics (statistical spread between snapshots) and also much smaller than the magnitude of the 3D effects (Fig. B1).

The `hey` parameterization with general habit mixture (`ghm`) is used in the main text (Yang et al., 2013; Emde et al., 2016). This parameterization is valid for a spectral range from  $0.2 - 5\mu\text{m}$ , and for ice effective radii from  $5 - 90\mu\text{m}$ . `hey` assumes smooth crystals and allows for four choices of habit: `ghm`, solid column (`col`), rough aggregate (`agg`), and `plate`.

The other parameterization tested was `baum_v36` (Heymsfield et al., 2013; Yang et al., 2013; Baum et al., 2014). This parameterization is valid over a wider spectral range ( $0.2 - 99\mu\text{m}$ ), but a narrower effective radius range ( $5 - 60\mu\text{m}$ ). Particles with effective radius outside of the accepted range were rounded to the maximum allowed value. The `baum_v36` parameterization assumes severely roughened particles. It allows for three choices of habit: `ghm`, solid column (`col`), and rough-aggregate (`agg`).

These seven variants are compared in Fig. B1 for one cloud snapshot from the TRMM-LBA case and they show very similar results. Also shown in Fig. B1 is a RT calculation done on the same cloud field, but only including the liquid droplets and ignoring the ice particles. We use the full Mie scattering phase function without

any parameterization for the liquid portion of the cloud in all cases. The difference between the liquid-only and liquid and ice simulated absolute fluxes can be up to 20% depending on the parameterization used (not shown), but the flux bias (ICA - 3D) is very similar for the liquid-only and all ice parameterizations.

### 5.9 Appendix C: Cloud property proxy for flux bias

We explored several different cloud properties to use as a proxy for the flux bias. Our limited study concluded that the in-cloud cloud water path (CWP) was the best proxy because it shows a strong positive, linear correlation with flux bias. Other cloud scene properties we examined included cloud top height (CTH), cloud cover (cc), and the geometric mean of covered area and uncovered area,  $\sqrt{cc(1 - cc)}$ . The linear regression fits are shown in Fig. C1. The rms error for CWP is the smallest. Although CTH (or cloud depth) are also reasonable proxies, they are more difficult to measure from satellite, and therefore we use CWP in this study. An important extension to this work would be to allow for multiple cloud properties and a more complex model than a linear fit to describe the flux bias. However, with our limited data from only four LES cases in this present study, we do not feel justified to use a more complex model.

### References

- [1] G. L. Stephens et al. “The albedo of Earth”. In: *Reviews of Geophysics* 53.1 (2015), pp. 141–163. doi: 10.1002/2014RG000449.
- [2] F. A.-M. Bender et al. “22 views of the global albedo—comparison between 20 GCMs and two satellites”. In: *Tellus A: Dynamic Meteorology and Oceanography* 58.3 (2006), pp. 320–330. doi: 10.1111/j.1600-0870.2006.00181.x.
- [3] A. Voigt et al. “The observed hemispheric symmetry in reflected shortwave irradiance”. In: *J. Climate* 26 (2013), pp. 468–477. doi: <https://doi.org/10.1175/JCLI-D-12-00132.1>.
- [4] A. Engström et al. “The nonlinear relationship between albedo and cloud fraction on near-global, monthly mean scale in observations and in the CMIP5 model ensemble”. In: *Geophysical Research Letters* 42.21 (2015), pp. 9571–9578. doi: <https://doi.org/10.1002/2015GL066275>.
- [5] T. Schneider et al. “Climate goals and computing the future of clouds”. In: *Nat. Climate Change opinion & comment* 7.1 (2017), pp. 3–5. doi: 10.1038/nclimate3190.

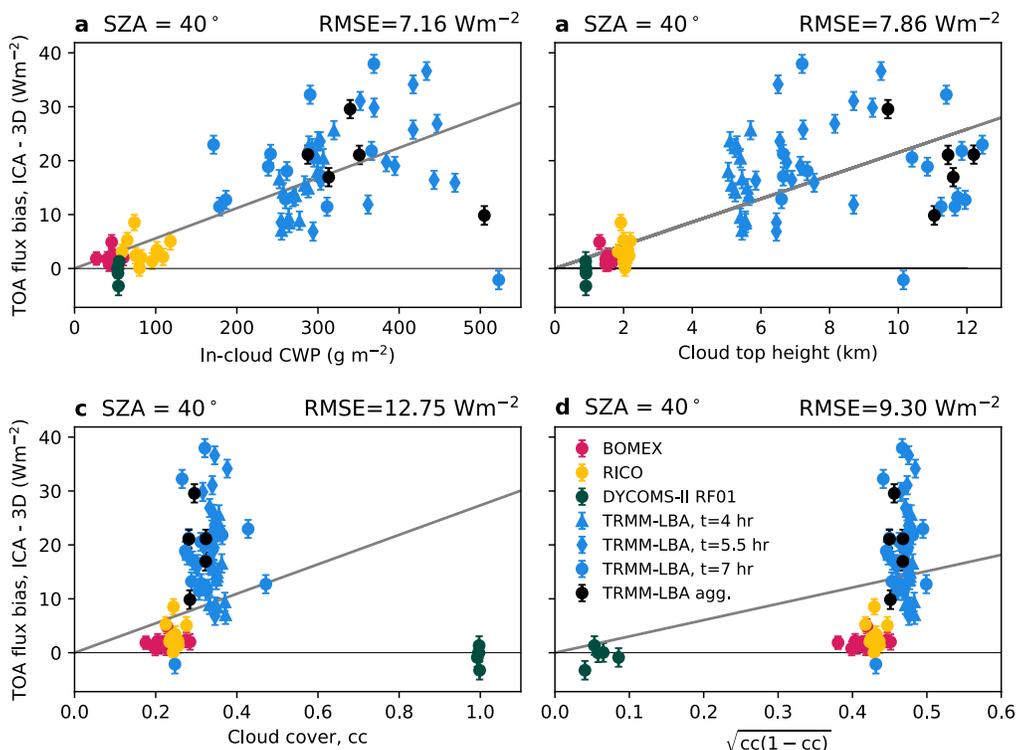


Figure 5.12: Regression of flux bias on several different cloud properties (a) in-cloud cloud water path (CWP), (b) cloud top height, (c) cloud cover, and (d)  $\sqrt{cc(1-cc)}$ . All panels show the flux bias for a zenith angle of  $40^\circ$ . (a) is the same as Fig. 5.7b. The rms error is shown for each panel and is smallest for the in-cloud CWP case (a).

- [6] A. Marshak and A. Davis, eds. *3D radiative transfer in cloudy atmospheres*. Physics of Earth and Space Environments. Berlin/Heidelberg: Springer-Verlag, 2005. DOI: 10.1007/3-540-28519-9.
- [7] R. Cahalan and W. Wiscombe. “Proceedings of the Second Atmospheric Radiation Measurement (ARM) Science Team Meeting”. In: *Plane-Parallel Albedo Bias*. December. Denver, Colorado, 1992, p. 35.
- [8] J. K. Shonk and R. J. Hogan. “Tripleclouds: An efficient method for representing horizontal cloud inhomogeneity in 1D Radiation schemes by using three regions at each height”. In: *Journal of Climate* 21.11 (2008), pp. 2352–2370. DOI: 10.1175/2007JCLI1940.1.
- [9] R. Pincus et al. “A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields”. In: *Journal of Geophysical Research: Atmospheres* 108.13 (2003), pp. 1–5. DOI: 10.1029/2002jd003322.
- [10] Y. Cohen et al. “Unified Entrainment and Detrainment Closures for Extended Eddy-Diffusivity Mass-Flux Schemes”. In: *Journal of Advances in*

- Modeling Earth Systems* 12.9 (2020), e2020MS002162. DOI: 10.1029/2020MS002162.
- [11] J. N. S. Cole, H. W. Barker, D. A. Randall, et al. “Global consequences of interactions between clouds and radiation at scales unresolved by global climate models”. In: *Geophysical Research Letters* 32.6 (2005). DOI: 10.1029/2004GL020945.
- [12] G. J. Kooperman et al. “Robust effects of cloud superparameterization on simulated daily rainfall intensity statistics across multiple versions of the Community Earth System Model”. In: *Journal of Advances in Modeling Earth Systems* 8.1 (2016), pp. 140–165. DOI: 10.1002/2015MS000574.
- [13] M. F. Khairoutdinov and D. A. Randall. “A cloud resolving model as a cloud parameterization in the NCAR Community Climate System Model: Preliminary results”. In: *Geophysical Research Letters* 28.18 (2001), pp. 3617–3620. DOI: 10.1029/2001GL013552.
- [14] R. F. Cahalan et al. “Independent Pixel and Monte Carlo Estimates of Stratocumulus Albedo”. In: *Journal of the Atmospheric Sciences* 51.24 (1994), pp. 3776–3790. DOI: 10.1175/1520-0469(1994)051<3776:IPAMCE>2.0.CO;2.
- [15] S. A. K. Schäfer et al. “Representing 3-D cloud radiation effects in two-stream schemes: 1. Longwave considerations and effective cloud edge length”. In: *Journal of Geophysical Research: Atmospheres* 121.14 (2016), pp. 8567–8582. DOI: 10.1002/2016JD024876.
- [16] R. J. Hogan, M. D. Fielding, et al. “Entrapment: An Important Mechanism to Explain the Shortwave 3D Radiative Effect of Clouds”. In: *Journal of the Atmospheric Sciences* 2019 (1 2019), pp. 48–66. DOI: 10.1175/JAS-D-18-0366.1.
- [17] R. J. Hogan and A. Bozzo. “A Flexible and Efficient Radiation Scheme for the ECMWF Model”. In: *Journal of Advances in Modeling Earth Systems* 10 (8 2018). <https://doi.org/10.1029/2018MS001364>, pp. 1990–2008. DOI: <https://doi.org/10.1029/2018MS001364>.
- [18] B. Mayer and A. Kylling. “Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use”. In: *Atmospheric Chemistry and Physics* 5.7 (2005), pp. 1855–1877. DOI: 10.5194/acp-5-1855-2005.
- [19] C. Emde et al. “The libRadtran software package for radiative transfer calculations (version 2.0.1)”. In: *Geoscientific Model Development* 9.5 (2016), pp. 1647–1672. DOI: 10.5194/gmd-9-1647-2016.
- [20] N. Villefranque et al. “A Path-Tracing Monte Carlo Library for 3-D Radiative Transfer in Highly Resolved Cloudy Atmospheres”. In: *Journal of Advances in Modeling Earth Systems* 11.8 (2019), pp. 2449–2473. DOI: 10.1029/2018MS001602.

- [21] J. J. Gristey et al. “Surface solar irradiance in continental shallow cumulus fields: Observations and large eddy simulation”. In: *Journal of the Atmospheric Sciences* (2019), pp. 19–0261. DOI: 10.1175/JAS-D-19-0261.1.
- [22] M. A. Veerman et al. “Three-Dimensional Radiative Effects By Shallow Cumulus Clouds on Dynamic Heterogeneities Over a Vegetated Surface”. In: *Journal of Advances in Modeling Earth Systems* 12.7 (2020). DOI: 10.1029/2019MS001990.
- [23] A. Marshak, A. Davis, W. Wiscombe, and G. Titov. “The verisimilitude of the independent pixel approximation used in cloud remote sensing”. In: *Remote Sensing of Environment* 52 (1995), pp. 71–78. DOI: 10.1016/0034-4257(95)00016-T.
- [24] W. O’Hirok and C. Gautier. “A Three-Dimensional Radiative Transfer Model to Investigate the Solar Radiation within a Cloudy Atmosphere. Part I: Spatial Effects”. In: *Journal of the Atmospheric Sciences* 55.12 (1998), pp. 2162–2179. DOI: 10.1175/1520-0469(1998)055<2162:ATDRTM>2.0.CO;2.
- [25] W. O’Hirok and C. Gautier. “The impact of model resolution on differences between independent column approximation and Monte Carlo estimates of shortwave surface irradiance and atmospheric heating rate”. In: *Journal of the Atmospheric Sciences* 62.8 (2005), pp. 2939–2951. DOI: 10.1175/JAS3519.1.
- [26] H. W. Barker, G. L. Stephens, et al. “Assessing 1D Atmospheric Solar Radiative Transfer Models: Interpretation and Handling of Unresolved Clouds”. In: *Journal of Climate* 16.16 (2003), pp. 2676–2699. DOI: 10.1175/1520-0442(2003)016<2676:ADASRT>2.0.CO;2.
- [27] H. W. Barker, S. Kato, et al. “Computation of Solar Radiative Fluxes by 1D and 3D Methods Using Cloudy Atmospheres Inferred from A-train Satellite Data”. In: *Surveys in Geophysics* 33.3-4 (2012), pp. 657–676. DOI: 10.1007/s10712-011-9164-9.
- [28] A. Marshak, A. Davis, W. Wiscombe, and R. Cahalan. “Radiative smoothing in fractal clouds”. In: *Journal of Geophysical Research* 100.D12 (1995), p. 26247. DOI: 10.1029/95JD02895.
- [29] T. Várnai and R. Davies. “Effects of cloud heterogeneities on shortwave radiation: Comparison of cloud-top variability and internal heterogeneity”. In: *Journal of the Atmospheric Sciences* 56.24 (1999), pp. 4206–4224. DOI: 10.1175/1520-0469(1999)056<4206:EOCHOS>2.0.CO;2.
- [30] J. W. Frame et al. “An application of the tilted independent pixel approximation to cumulonimbus environments”. In: *Atmospheric Research* 91 (2009), pp. 127–136. DOI: 10.1016/j.atmosres.2008.05.005.

- [31] R. J. Hogan and J. K. P. Shonk. “Incorporating the Effects of 3D Radiative Transfer in the Presence of Clouds into Two-Stream Multilayer Radiation Schemes”. In: *Journal of the Atmospheric Sciences* 70.2 (2013), pp. 708–724. DOI: 10.1175/JAS-D-12-041.1.
- [32] U. Wissmeier et al. “paNTICA: A fast 3D radiative transfer scheme to calculate surface solar irradiance for NWP and LES models”. In: *Journal of Applied Meteorology and Climatology* 52.8 (2013), pp. 1698–1715. DOI: 10.1175/JAMC-D-12-0227.1.
- [33] M. Okata et al. “A study on radiative transfer effects in 3-D cloudy atmosphere using satellite data”. In: *Journal of Geophysical Research: Atmospheres* 122.1 (2017), pp. 443–468. DOI: 10.1002/2016JD025441.
- [34] L. Oreopoulos and H. W. Barker. “Accounting for subgrid-scale cloud variability in a multi-layer 1d solar radiative transfer algorithm”. In: *Quarterly Journal of the Royal Meteorological Society* 125.553 (1999), pp. 301–330. DOI: 10.1002/qj.49712555316.
- [35] C. Klinger and B. Mayer. “The Neighboring Column Approximation (NCA) – A fast approach for the calculation of 3D thermal heating rates in cloud resolving models”. In: *J. Quant. Spectrosc. Radiat. Transf.* 168 (2016), pp. 17–28. DOI: 10.1016/j.jqsrt.2015.08.020.
- [36] C. Klinger and B. Mayer. “Neighboring Column Approximation—An Improved 3D Thermal Radiative Transfer Approximation for Non-Rectangular Grids”. In: *J. Adv. Model. Earth Syst.* 12.1 (2020). DOI: 10.1029/2019MS001843.
- [37] J. N. S. Cole, H. W. Barker, W. O’Hirok, et al. “Atmospheric radiative transfer through global arrays of 2D clouds”. In: *Geophysical Research Letters* 32.19 (2005). DOI: 10.1029/2005GL023329.
- [38] H. W. Barker, J. N. S. Cole, J. Li, B. Yi, et al. “Estimation of Errors in Two-Stream Approximations of the Solar Radiative Transfer Equation for Cloudy-Sky Conditions”. In: *Journal of the Atmospheric Sciences* 72.11 (2015), pp. 4053–4074. DOI: 10.1175/JAS-D-15-0033.1.
- [39] H. W. Barker, J. N. S. Cole, J. Li, and K. von Salzen. “A parametrization of 3-D subgrid-scale clouds for conventional GCMs: Assessment using A-Train satellite data and solar radiative transfer characteristics”. In: *Journal of Advances in Modeling Earth Systems* 8.2 (2016), pp. 566–597. DOI: 10.1002/2015MS000601.
- [40] D. M. Romps and R. Öktem. “Observing clouds in 4d with multiview stereophotogrammetry”. In: *Bulletin of the American Meteorological Society* 99.12 (2018), pp. 2575–2586. DOI: 10.1175/BAMS-D-18-0029.1.
- [41] E. Castro et al. “Determination of Cloud-top Height through Three-dimensional Cloud Reconstruction using DIWATA-1 Data”. In: *Scientific Reports* 10.1 (2020), pp. 1–13. DOI: 10.1038/s41598-020-64274-z.

- [42] K. G. Pressel, C. M. Kaul, et al. “Large-eddy simulation in an anelastic framework with closed water and entropy balances”. In: *J. Adv. Model. Earth Syst.* 7.3 (2015), pp. 1425–1456. doi: 10.1002/2015MS000496.
- [43] K. G. Pressel, S. Mishra, et al. “Numerics and subgrid-scale modeling in large eddy simulations of stratocumulus clouds”. In: *Journal of Advances in Modeling Earth Systems* 9.2 (2017), pp. 1342–1365. doi: 10.1002/2016MS000778.
- [44] P. J. Griewank et al. “Size dependence in chord characteristics from simulated and observed continental shallow cumulus”. In: *Atmospheric Chemistry and Physics* 20.17 (2020), pp. 10211–10230. doi: 10.5194/acp-20-10211-2020.
- [45] G. Cesana et al. “The Cumulus And Stratocumulus CloudSat-CALIPSO Dataset (CASCCAD)”. In: *Earth System Science Data* 11.4 (2019), pp. 1745–1764. doi: 10.5194/essd-11-1745-2019.
- [46] A. P. Siebesma et al. “A large eddy simulation intercomparison study of shallow cumulus convection”. In: *Journal of the Atmospheric Sciences* 60.10 (2003), pp. 1201–1219. doi: 10.1175/1520-0469(2003)60<1201:ALESIS>2.0.CO;2.
- [47] M. C. vanZanten et al. “Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO”. In: *Journal of Advances in Modeling Earth Systems* 3.2 (2011). doi: 10.1029/2011MS000056.
- [48] B. Stevens et al. “Evaluation of large-eddy simulations via observations of nocturnal marine stratocumulus”. In: *Monthly Weather Review* 133.6 (2005), pp. 1443–1462. doi: 10.1175/MWR2930.1.
- [49] W. W. Grabowski et al. “Daytime convective development over land: A model intercomparison based on LBA observations”. In: *Quarterly Journal of the Royal Meteorological Society* 132.615 (2006), pp. 317–344. doi: 10.1256/qj.04.147.
- [50] N. Jeevanjee and D. M. Romps. “Convective self-aggregation, cold pools, and domain size”. In: *Geophysical Research Letters* 40.5 (2013), pp. 994–998. doi: 10.1002/grl.50204.
- [51] A. A. Wing et al. “Convective Self-Aggregation in Numerical Simulations: A Review”. In: *Surveys in Geophysics* 38.6 (2017), pp. 1173–1197. doi: 10.1007/s10712-017-9408-4.
- [52] C. R. Patrizio and D. A. Randall. “Sensitivity of Convective Self-Aggregation to Domain Size”. In: *Journal of Advances in Modeling Earth Systems* 11.7 (2019), pp. 1995–2019. doi: 10.1029/2019MS001672.
- [53] C. S. Bretherton. “Insights into low-latitude cloud feedbacks from high-resolution models”. In: *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 373.2054 (2015). doi: 10.1098/rsta.2014.0415.

- [54] B. Mayer. “Radiative transfer in the cloudy atmosphere”. In: *EPJ Web of Conferences* 1 (2009), pp. 75–99. DOI: 10.1140/epjconf/e2009-00912-1.
- [55] A. S. Ackerman et al. “Large-Eddy Simulations of a Drizzling, Stratocumulus-Topped Marine Boundary Layer”. In: *Monthly Weather Review* 137.3 (2009), pp. 1083–1110. DOI: 10.1175/2008MWR2582.1.
- [56] P. N. Blossey et al. “Marine low cloud sensitivity to an idealized climate change: The CGILS LES intercomparison”. In: *J. Adv. Model. Earth Syst.* 5.2 (2013), pp. 234–258. DOI: 10.1002/jame.20025.
- [57] K. Wyser. “The effective radius in ice clouds”. In: *Journal of Climate* 11.7 (1998), pp. 1793–1802. DOI: 10.1175/1520-0442(1998)011<1793:TERIIC>2.0.CO;2.
- [58] P. Yang et al. “Spectrally consistent scattering, absorption, and polarization properties of atmospheric ice crystals at wavelengths from 0.2 to 100  $\mu\text{m}$ ”. In: *Journal of the Atmospheric Sciences* 70.1 (2013), pp. 330–347. DOI: 10.1175/JAS-D-12-039.1.
- [59] W. B. Rossow et al. “Advances in understanding clouds from ISCCP”. In: *Bulletin of the American Meteorological Society* 80.11 (1999), pp. 2261–2287. DOI: 10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2.
- [60] W. Rossow and E. Duenas. “The International Satellite Cloud Climatology Project (ISCCP) web site: An online resource for research”. In: *Bulletin of the American Meteorological Society* 85.2 (2004), pp. 167–176. DOI: 10.1175/BAMS-85-2-167.
- [61] R. Marchand et al. “A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS”. In: *Journal of Geophysical Research* 115.D16 (2010). DOI: 10.1029/2009JD013422.
- [62] C. Stubenrauch et al. *Assessment of global cloud data sets from satellites: A project of the world climate research programme Global Energy and Water Cycle Experiment (GEWEX) Radiation Panel*. Tech. rep. 23. World Climate Research Programme, 2012.
- [63] C. J. Stubenrauch et al. “Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel”. In: *Bulletin of the American Meteorological Society* 94.7 (2013), pp. 1031–1049. DOI: 10.1175/BAMS-D-12-00117.1.
- [64] M. Zhao et al. “The GFDL Global Atmosphere and Land Model AM4.0/LM4.0: 2. Model Description, Sensitivity Studies, and Tuning Strategies”. In: *Journal of Advances in Modeling Earth Systems* 10.3 (2018), pp. 735–769. DOI: 10.1002/2017MS001209.

- [65] F. Hourdin et al. “LMDZ6A: the atmospheric component of the IPSL climate model with improved and better tuned physics”. In: *Journal of Advances in Modeling Earth Systems* (2020). DOI: 10.1029/2019MS001892.
- [66] F. Brient et al. “Evaluating Marine Stratocumulus Clouds in the CNRM-CM6-1 Model Using Short-Term Hindcasts”. In: *J. Adv. Model. Earth Syst.* 11.1 (2019), pp. 127–148. DOI: 10.1029/2018MS001461.
- [67] G. Myhre et al. “Anthropogenic and Natural Radiative Forcing”. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by T. Stocker et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2013. Chap. 8, pp. 659–740.
- [68] C. Singer et al. *Data for “Top-of-atmosphere albedo bias from neglecting three-dimensional radiative transfer through clouds”*. 2021. DOI: 10.22002/D1.1637.
- [69] A. Seifert and K. D. Beheng. “A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description”. In: *Meteorology and Atmospheric Physics* 92.1-2 (2006), pp. 45–66. DOI: 10.1007/s00703-005-0112-4.
- [70] C. M. Kaul et al. “Sensitivities in Large-Eddy Simulations of Mixed-Phase Arctic Stratocumulus Clouds Using a Simple Microphysics Approach”. In: *Monthly Weather Review* 143.11 (2015), pp. 4393–4421. DOI: 10.1175/MWR-D-14-00319.1.
- [71] Z. Shen et al. “Statistically Steady State Large-Eddy Simulations Forced by an Idealized GCM: 1. Forcing Framework and Simulation Characteristics”. In: *Journal of Advances in Modeling Earth Systems* 12.2 (2020). DOI: 10.1029/2019MS001814.
- [72] S. Kato et al. “The  $k$ -distribution method and correlated- $k$  approximation for a shortwave radiative transfer model”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 62.1 (1999), pp. 109–121. DOI: [https://doi.org/10.1016/S0022-4073\(98\)00075-2](https://doi.org/10.1016/S0022-4073(98)00075-2).
- [73] A. J. Heymsfield et al. “Ice cloud particle size distributions and pressure-dependent terminal velocities from in situ observations at temperatures from 0° to -86°C”. In: *Journal of the Atmospheric Sciences* 70.12 (2013), pp. 4123–4154. DOI: 10.1175/JAS-D-12-0124.1.
- [74] B. A. Baum et al. “Ice cloud single-scattering property models with the full phase matrix at wavelengths from 0.2 to 100 $\mu$ m”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 146 (2014), pp. 123–139. DOI: 10.1016/j.jqsrt.2014.02.029.

*Appendix A***GENDER DISPARITIES IN GPS QUALIFYING EXAM  
OUTCOMES****A.1 Remarks from the author**

The following chapter details information on outcomes of the Caltech GPS qualifying exam that have been collected by myself and a collection of graduate student colleagues. The chapter is broken into two sections:

1. The first section is based on a report I wrote for Heather Knutson's Ge/Ay 117 course in Winter 2020. The analysis was updated in Spring 2021. Data collection has been transferred to the GPS Division.
2. The survey, presented later in this chapter, originated as a response to the statistical analysis of Part 1. The survey targeted the same cohort of students (and recent alumni) who made up the years of analysis in the original report.

Myself and my collaborators on this work are indebted to the generation of graduate students who came before us, who preserved these ideas through a whisper network, and motivated the initial statistical study.

The following is the first ever compilation of GPS qualifying exam data. The results are not surprising to those engaged with the issue, but are deeply troubling nonetheless. I hope that this report inspires a recognition of the importance of collecting meta-data about the exams. Without this multi-year dataset it is impossible to discern any trends from small annual samples with lots of variability. However, the numbers do add up over time and with only six (and now seven, or more) years, these data are able to demonstrate stark discrepancies in exam outcomes. It is necessary for the Division to keep a thorough record of exams and use that data to help inform changes to the Qualifying Exams themselves, mentoring of graduate students, and even admissions criteria.

This chapter is presented as a record of the efforts made by myself and my colleagues during our time as Ph.D. students at Caltech with the hopes that it will preserve our knowledge gained and the progress we have made through tireless advocacy.

## **A.2 Part 1: Are there robust statistical differences in exam outcome between men and women?**

### **Summary**

This project uses a statistical model to understand GPS qualifying exam outcomes. I use a data set of exam outcomes from the past six years to inform my model. I test the hypothesis that demographic information about the students can be used to predict the outcome of the qualifying exam.

The previous literature addressing topics of bias in exam scores and outcomes is very rich. A thorough review has not been conducted, but there are so many more extensions and additions to this analysis that can, and should, be inspired by previous work.

### **What is the exam?**

In the Caltech Geological and Planetary Science Division (GPS), all graduate students take a qualifying exam in September at the end of their first year of study. The exam format consists of a written report (10 pages) and an oral defense (3 hours) of the two proposition projects to a committee of historically five, and now four, faculty members, including the two proposition advisors.

### **Data**

These data have been compiled through interviews with individuals, scraped from online CVs, and corroborated by friends and colleagues. These data only extend back to 2014 due to limited time and resources for collecting them. Data for additional individuals exists dating back 10 years, but only data from “complete” years (where every individual in a cohort has been accounted for) has been included in this analysis to limit sampling bias.

The anonymized dataset consists of information for each of the 119 students who have taken the GPS qualifying exam since 2014. The information collected includes: (1) year exam was taken, (2) option within GPS, (3) (binary) gender of student, (4) are they a native or non-native English speaker, (5) did they have a masters before coming to Caltech, and (6) the outcome of the exam.

Figure A.1 shows the number of students who passed and failed the qualifying exam in each of the six years. Figure A.2 shows the total number of students who passed and failed the exam by option, gender, or English language status.

Additional information about whether a student who failed retook the exam and the

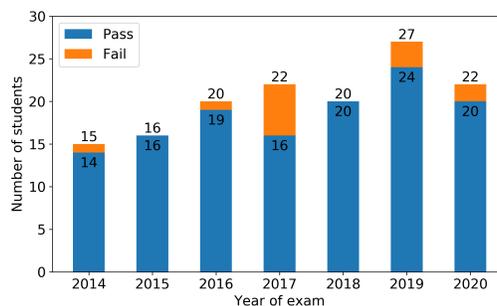


Figure A.1: Number of students per year who passed or failed the exam. Labels are total number of students and number who passed for each year.

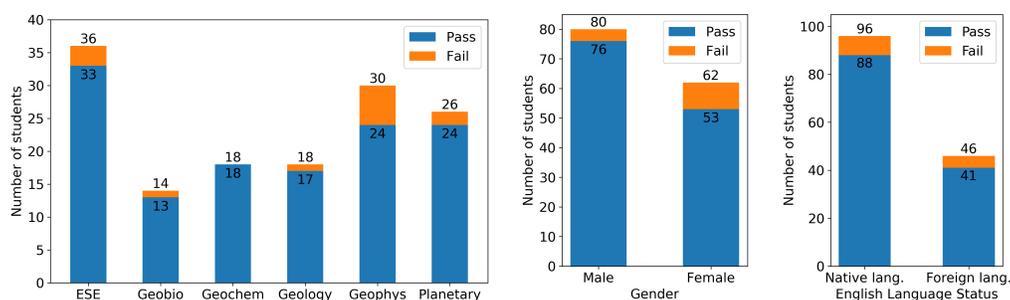


Figure A.2: Number of students per option (left), gender (center), or English language status (right) who passed or failed the exam. Labels are total number of students and number who passed for each category.

outcome of that retake has also been tracked, but is not included in the analysis. For some individuals, there is data on whether they left the PhD program after passing their exam but before completing the degree. Additional variables that are of interest to the author based in previous literature include the student's status as a underrepresented minority or first-generation college student, bachelor's GPA, GRE score, and the faculty on the examining committee. These data have yet to be collected.

For each additional variable included in the model, more individuals are needed to constrain the uncertainties on the power of that predictor, which poses a challenge to this analysis. For the present study only the students' option, gender, and English language status are included as predictors. These factors were determined as being potentially predictive via a preliminary analysis of these data. Figure A.3 shows the fraction of students per category who passed or failed the exam. It is very obvious that being in the Geophysics option is a relevant predictor for exam success. Given

that women and foreign speakers of English fail at higher rates than men and native speakers (respectively), it is possible that these are also important predictors. While having a masters degree appears to be less important (and also possibly correlated with English language status given that a large number of international students come with prior masters degrees).

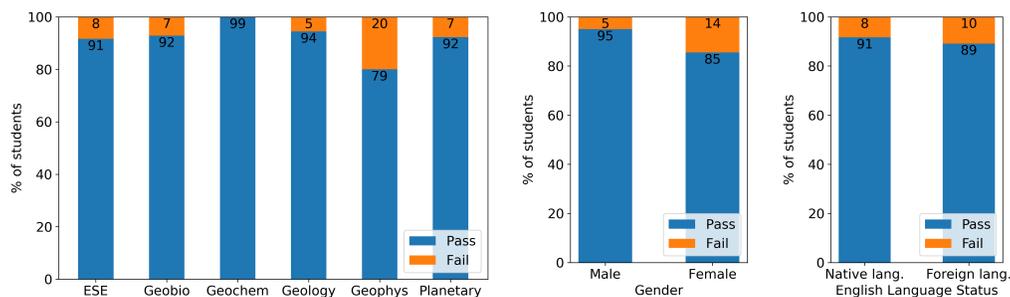


Figure A.3: Percentage of students per option (left), gender (center), or English language status (right) who passed or failed the exam. Labels are percentage of students who passed or failed for each category.

The goal of this study is to determine whether these patterns observed in the data are statistically significant, and if so, how much being in one of these groups statistically helps or hurts your probability of passing your qualifying exam.

### Statistical Methods

Based on the preliminary analysis shown above, I consider the dependence of exam outcome on three variables for my model: (1) option, (2) gender, and (3) English language status. The statistical analysis is done in the following steps.

1. Because these variables are categorical, I first create dummy variables to use in the model. For each variable, there must be  $n - 1$  binary dummy variables where  $n$  is the number of possible categories. This is to avoid the “dummy variable trap.” So the total number of variables will be  $N = 1 + \sum_{i=1}^3 (n_i - 1)$ . When one variable for each category is thrown away I am implicitly creating a “base state” against which all perturbations are compared against.
2. The exam outcome is a binary variable (pass or fail). For this reason I am using a logistic model

$$\sigma(t) = \frac{1}{1 + e^{-t}}. \quad (\text{A.1})$$

Here,  $t$  is known as a “latent variable” because it is not observed. In this case, I propose using

$$t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N + \beta_{N+1} x_1 x_2 + \beta_{N+2} x_1 x_3 + \dots, \quad (\text{A.2})$$

a linear combination of the explanatory variables ( $\vec{x} = \{x_1 \dots x_N\}$ ), including interaction terms. These interaction terms quantify the importance of intersectional identities. These can be included in extensions of the present analysis, but were omitted due to time and data constraints.

For simplicity, I reduce this model to

$$t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5, \quad (\text{A.3})$$

where  $\beta_0$  is the constant term representing the base case,  $\beta_1$  represents the effects of gender,  $\beta_2$  represents the effects of English language status, and  $\beta_3 - \beta_5$  represent the effect of option. It is possible/likely that some of these  $\beta_i$  terms are zero within uncertainty.

3. I use a Bayesian approach to calculate these coefficients  $\beta_i$  using these data. I model the process of the qualifying exams as a Bernoulli process, where the demographic factors will predict a certain probability of passing ( $p_k = \sigma(f(\vec{x}_k))$ ). The result, if a person actually passes or not, follows a binomial distribution. Therefore, the likelihood can be calculated as

$$\mathcal{L} = \prod_{k=1}^M p_k^{y_k} (1 - p_k)^{1-y_k}. \quad (\text{A.4})$$

4. Finally, I use a standard Metropolis-Hastings Markov Chain Monte Carlo (MCMC) to fit  $\beta_i$ . With the results of the MCMC, I calculate the probability of passing  $p = \sigma(\vec{x})$  and the odds  $e^{f(\vec{x})}$ . Furthermore, I calculate the odds ratio. For example, the odds ratio for exam outcome for men vs. women, given that the base state of  $x_1 = 0$  is woman,

$$OR = \frac{e^{f(\vec{x}|x_1=1)}}{e^{f(\vec{x}|x_1=0)}} = e^{\beta_1} \quad (\text{A.5})$$

quantifies how many times more likely it is to pass given that you are a man vs. woman.

## Results

The MCMC was initialized by first doing optimal estimation of the parameters. This led to very rapid convergence and the burn-in period was negligible. The MCMC was run for 50,000 steps. The initial parameter values were

$$\vec{\beta}_{guess} = [2.55, 1.14, -0.14, -0.57, -1.96, -0.01]$$

and the parameters converged to

$$\vec{\beta}_{final} = [2.70, 1.23, -0.13, -0.37, -2.11, 0.22].$$

Figure A.4 shows a corner plot of the six coefficients fit in the model  $\beta_0 - \beta_5$ . The probability of passing for the base state is

$$p_{base} = \frac{1}{1 + e^{-\beta_0}} = \frac{1}{1 + e^{-2.70}} = 0.937.$$

**The  $\beta_i$  coefficients that are statistically different from zero (at the  $1\sigma$  level) are  $\beta_1$  and  $\beta_4$  which correspond to the gender effect ( $x_1 = 0$  for women and  $x_1 = 1$  for men) and geophysics effect.** Because  $\beta_1$  is positive, this means that it is more likely for men to pass the exam compared to women, and  $\beta_4$  being negative means it is more likely for someone in geophysics to fail the exam compared to someone in ESE (the base state).

Given that women and students who speak English as a foreign language were both failing the exam at twice the rate of their peers (Figure A.3), why is gender but not English language status a statistically significant predictor? This must be because English language status is correlated with some other variable. This becomes evident when looking at Figure A.5: gender is not highly correlated with option, but English language status is.

While students who speak English as a foreign language fail at higher rates, it turns out this is because most of these students are in the Geophysics or ESE options, while very few of these students are in Geobiology, Geochemistry, or Geology. But, Geophysics has by far the highest failure rate (20%) and Geochemistry and Geology have a 0% failure rate (in the past six years).

On the other hand, for gender, there is no strong correlation between options. All options except Geophysics have nearly equal proportions of male and female students. But by looking at this breakdown by gender and option it is clear that

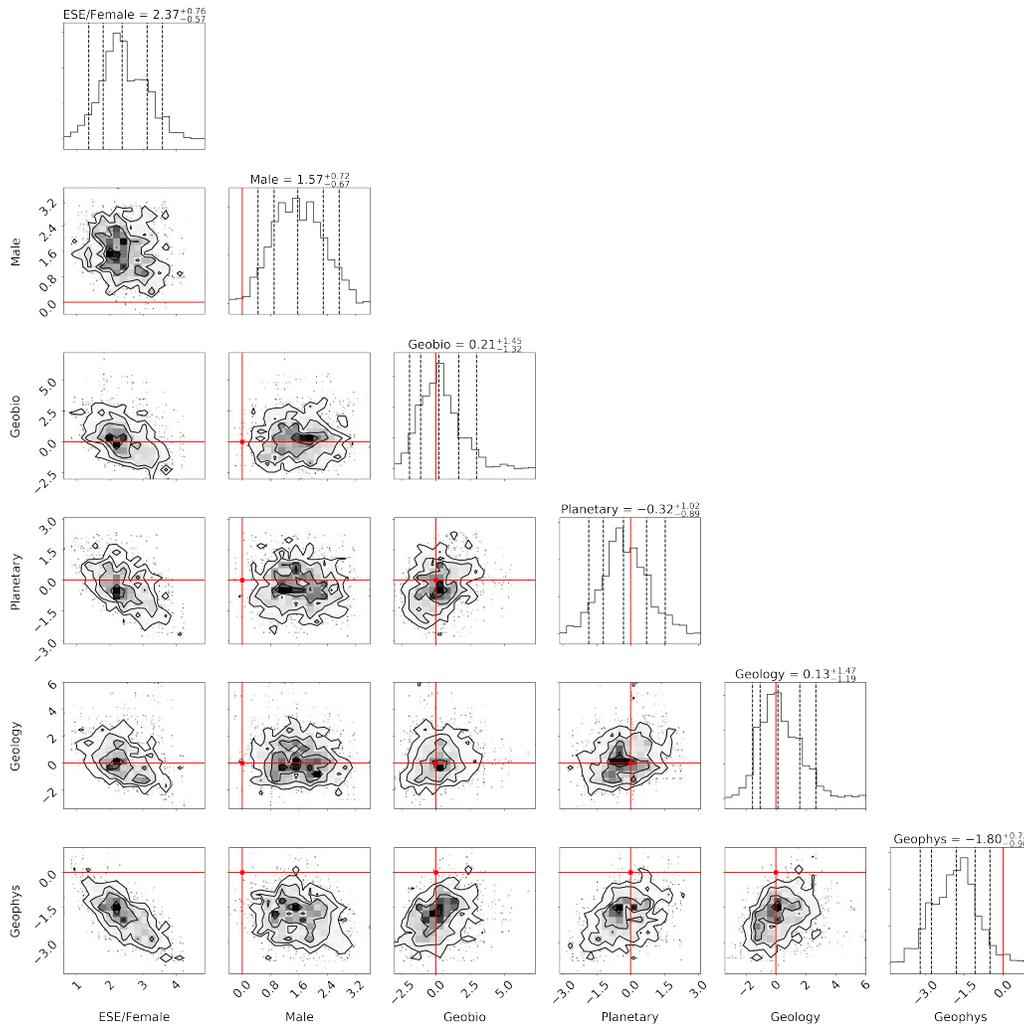


Figure A.4: Corner plot of  $\beta_i$  values from MCMC analysis. The solid lines show  $\beta_i = 0$  and the dashed lines show the mean and  $\pm 1\sigma$  error bars for each  $\beta_i$ . (Note: The null hypothesis does not correspond to  $\beta_0 = 0$ , so this is not plotted.)

women in Geophysics are failing at a significantly higher rate than a) women in other options and b) men in Geophysics.

To further study this effect, we can examine an abbreviated model with predictor variables that are not statistically different from zero excluded. The new model is

$$\sigma(\vec{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_4 x_4)}}. \quad (\text{A.6})$$

The parameter best fit results with the new model are shown in Figure A.6 using MCMC with only 10,000 steps.

The shorter MCMC run can be justified by calculating the autocorrelation length for

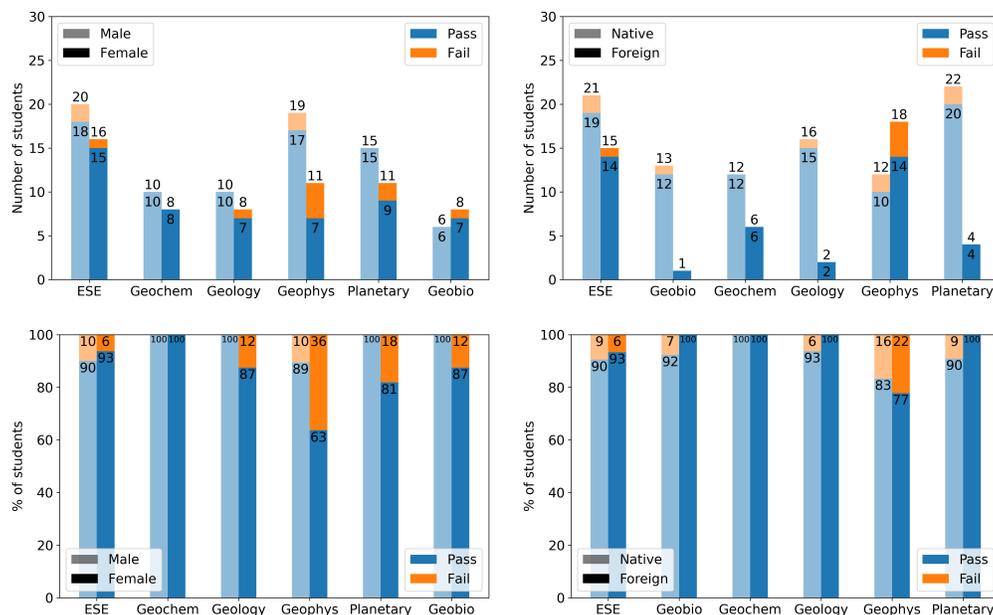


Figure A.5: Number of students per option per gender (top left) or English language status (top right) who passed or failed the exam. Labels are total number of students and number who passed for each category. Bottom panels show percentage of students who passed or failed the exam in each category (left for gender and right for English language status).

the chain. The autocorrelation length was found to be  $ACL \sim 300$ , so,  $N = 10000$  steps obeys the rule-of-thumb that  $ACL/N \geq 10$ . I obtain slightly different  $\beta_i$  values from this abbreviated model, but they agree well within the  $1\sigma$  error bars. This is shown in Table A.1.

	full model	abbreviated model
$\beta_0$	$2.37^{+0.76}_{-0.57}$	$2.19^{+0.41}_{-0.37}$
$\beta_1$	$1.57^{+0.72}_{-0.67}$	$1.39^{+0.68}_{-0.59}$
$\beta_5$	$-1.80^{+0.72}_{-0.96}$	$-1.49^{+0.54}_{-0.60}$

Table A.1: Table showing  $\beta_i$  values calculated via MCMC using the full model with all 5 predictors versus the abbreviated model with only two predictors. The results agree.

Based on this model and the  $\beta_i$  parameters fit with MCMC, I can calculate the predicted probability of passing the GPS qualifying exam for the different demographic groups. In particular I am interested in the difference between male and female and the difference between Geophysics and ESE (the base case). These mean

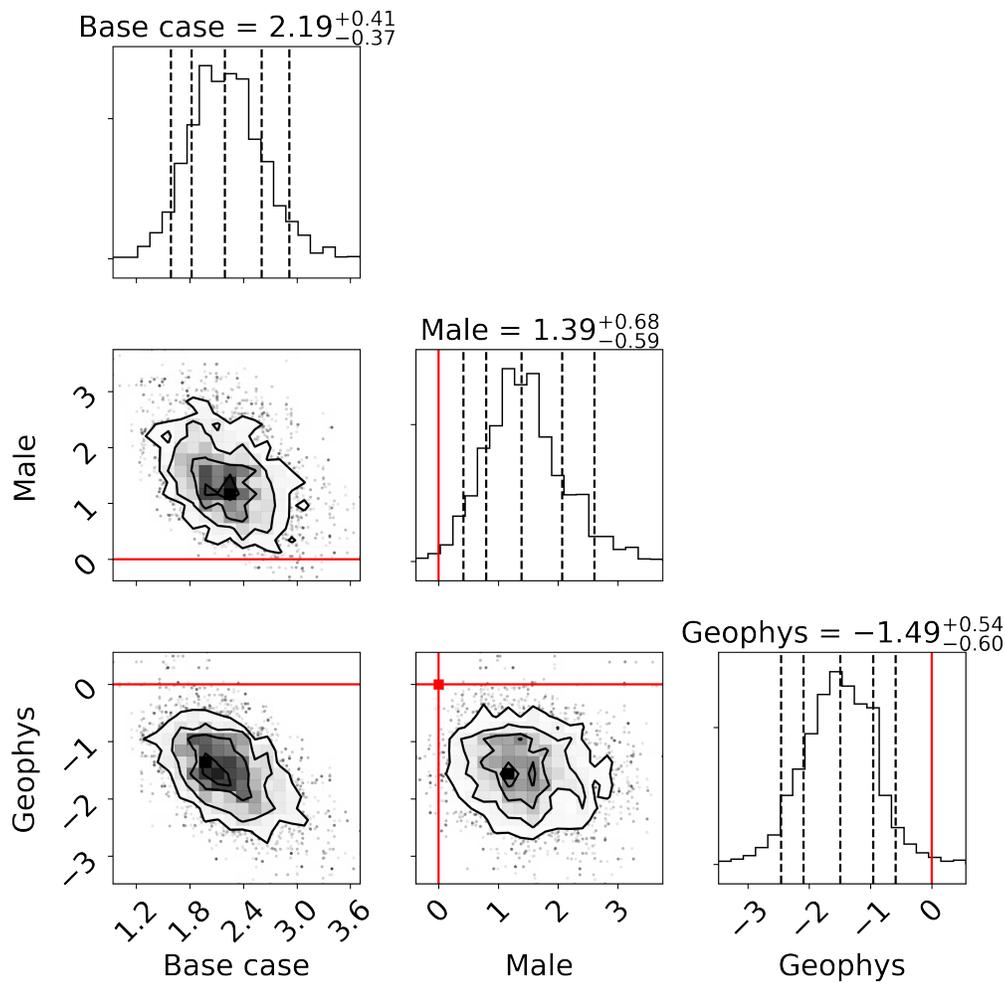


Figure A.6: Corner plot of  $\beta_i$  values from MCMC analysis. The solid lines show  $\beta_i = 0$  and the dashed lines show the mean and  $\pm 1\sigma$  error bars for each  $\beta_i$ .

predicted values and one standard deviation are show in Table A.2. These values were calculated as

$$\delta\sigma = \sqrt{\left(\frac{\partial\sigma}{\partial\beta_0}\delta\beta_0\right)^2 + \left(\frac{\partial\sigma}{\partial\beta_1}\delta\beta_{1x_1}\right)^2 + \left(\frac{\partial\sigma}{\partial\beta_4}\delta\beta_{4x_4}\right)^2} \quad (\text{A.7})$$

The error bars on these passing probabilities are large, but the trends still clearly stand out. To more confidently demonstrate these discrepancies across gender and option categories, more years of data must be added to this analysis to increase the sample size beyond  $N = 119$  students.

	Women	Men
ESE	90 ± 4%	97 ± 2%
Geophysics	67 ± 15%	89 ± 9%

Table A.2: Probability of passing the GPS Qualifying Exam based on the logistic model with MCMC fit best parameters.

### **Preliminary conclusions**

These preliminary findings indicated that more years of data are needed, but also more data on the student experience, beyond pass/fail rates, is needed to gain understanding of why students are failing, why the failure rates are so disproportionate across options and across genders, and what changes could be implemented to remedy these problems. To follow up, myself along with two other students, staff from the Title IX office, and in coordination with the GPS Core Committee (the group of faculty responsible for administering the exams) designed and conducted a survey of current students and recent alumni about their experiences surrounding the qualifying exams. The details of the survey are discussed in the following sections.

### **A.3 Part 2: Why are there differences in exam outcome between men and women?**

*Co-authors, listed alphabetically: Lilian A. Dove, Sara E. Murphy, Clare E. Singer*

#### **The survey**

The survey was conducted in Summer 2021. The survey was written by Lily Dove, Sara Murphy, and Clare Singer, with help from Allie McIntosh (Equity & Title IX Office), and after approval from Mike Brown, Ed Stolper, and Paul Wennberg (GPS/ESE Core Committee<sup>1</sup> members). The survey was administered to all current students who had taken the qualifying exam and to as many recent alumni as we could contact. The survey results were collected and anonymized by the Caltech Title IX office. The authors only ever had access to anonymized, aggregated, or redacted results from the survey.

<sup>1</sup>The Core Committee is the faculty committee responsible for administering the qualifying exams each year. Every examining committee consists of at least one member of the Core Committee, often two. The administration of the ESE exams is done by a separate ESE Core Committee, which operates in parallel to the GPS Core Committee. Typically faculty on the Core Committee change every year. One result of this analysis and the survey was to expand the scope of the Core Committee slightly. They now claim some responsibility for the design of the exam, not just the implementation. Furthermore, the chair of the Core Committee now has a three year, rather than one year, term so that they may provide continuity and oversee changes on slightly longer timescales.

## Survey questions

### 1. Before quals:

- a) How supported did you feel while preparing for your exam? (1-5)
- b) How confident were you leading up to your exam? (1-5)
- c) How informed were you about the format of the exam? (1-5)
- d) How prepared did you feel for your exam? (1-5)
- e) When did you start your first (second) proposition? (Mult. choice)
- f) How often did you meet with your first (second) advisor about your proposition during your first year? (Mult. choice)
- g) Was this frequent enough to meet your advising needs? (Yes/No)
- h) What was your experience leading up to quals? (Open-ended)  
*E.g. Who gave you the most support? What was the messaging you received from your advisor about the exam? Did you have adequate information about what to expect and from whom did you get that information?*

### 2. During the exam:

- a) Was the format of the exam as it was described to you? (1-5)
- b) Do you feel that the qualifying exam allowed you to show your achievements and growth as a researcher over the course of the first year? (1-5)
- c) Were the questions asked during your exam relevant to your research? (Yes/No)
- d) Were the questions asked in a respectful manner? (Yes/No)
- e) What was your experience on the day of quals? (Open-ended)  
*E.g. What did you feel the exam was testing? Were your committee members attentive, engaged, respectful during the exam? Did the exam challenge you in a positive way? – Please be as specific as you would like. Your answers to all questions will be anonymized and/or aggregated before shown to the faculty.*

### 3. After the exam:

- a) Did the qualifying exam make you aware of your weaknesses as a researcher in a constructive way? (1-5)

- b) How helpful was the feedback from your committee members? (1-5)
- c) What 2-3 words best describe your feelings immediately after finishing the exam?
- d) What 2-3 words best describe your feelings 1 week after finishing the exam?
- e) Did you talk to your committee members after the exam? (Yes/No)
- f) When did you talk to your committee members? (Mult. choice)
- g) If you didn't talk with your committee members after the exam, why not? (Open-ended)
- h) What was your experience after quals? (Open-ended)  
*E.g. From whom did you get feedback? How useful was that feedback? Did you feel confident in your abilities as a scientist after the exam? Do you feel like you know why you passed/failed? Did the feedback you received help you identify your strengths and weaknesses? Did it help you figure out how to improve on those weaknesses?*

4. General questions:

- a) In an ideal world, how would you restructure or change quals? (Open-ended)
- b) Are there events or activities during the first year, prior to the qualifying exam, that would improve exam preparation? (Open-ended)

5. Demographic questions:

- a) When did you take quals? (Mult. choice)
- b) What was the outcome of your exam? (Pass/Fail/Other)
- c) Racial identity (Mult. choice)
- d) Gender identity (Mult. choice)
- e) Do you identify as LGBTQIA+? (Yes/No)
- f) Do you identify as a first generation student? (Yes/No)
- g) Do you identify as someone with a disability? (Yes/No)
- h) Are you a native English speaker? (Yes/No)

**Email sent to students about survey**

“Dear G2+ GPS students,

Consistent feedback from students across the GPS Division demonstrates that the Qualifying Exam is an area that requires further reflection and discussion within the division. With the assistance of the GPS/ESE Qualifying Exam Core Committees and the Equity and Title IX Office, we have constructed a survey to assess the current state of the qualifying exam. The goal of this form is to collect student experiences from before, during, and after their qualifying exam, to help develop a clear picture of where students feel supported, where there can be improvement, and where the exam is meeting and/or falling short of its intended goals. Please take the survey (link) before June 15.

While we have a number of 1-5 questions on the form, we would like to emphasize the long-answer questions. Please take the time to fill them out, if you feel comfortable. The purpose of this form is to collect stories and details, and this is the information that will be most helpful. Below, you can see the long-form questions so you can think about them ahead of time.

All questions on this survey are optional, and your responses will never be shared in a way that may be identifying. The responses to this form will only be directly accessible to Deputy Equity and Title IX Coordinator Allie McIntosh (allie@caltech.edu) and Julie Lee who will further anonymize the stories/information before sharing them with the faculty.

As an incentive, 4 survey respondents will receive \$25 gift cards to their choice of Urban Plates or Vroman’s Bookstore. You must enter your email in the secondary google form after you complete the survey to be entered into the lottery.

If you have any questions or concerns, please reach out to us.

Best,

Sara Murphy, Clare Singer, Lily Dove (ESE G3s)

GPS Qualifying Exam Core Committee (Chair: Mike Brown)

ESE Qualifying Exam Core Committee (Chair: Paul Wennberg)”

## Overview of survey results

### Who took the survey?

The survey was completed by 71 current and former Caltech GPS students, which represented approximately a 65% completion rate (compared to an average of 110 total students in the Division). The demographics of students taking the survey was commensurate with the demographics of the overall student body. Figure A.7 shows histograms of survey-taker demographics including English language status, ability status, first-generation student status, year of exam, racial identity, and exam outcome. For all questions, demographic and otherwise, survey participants were allowed to opt out or not answer; these responses are labeled as “NoResponse” throughout the results.

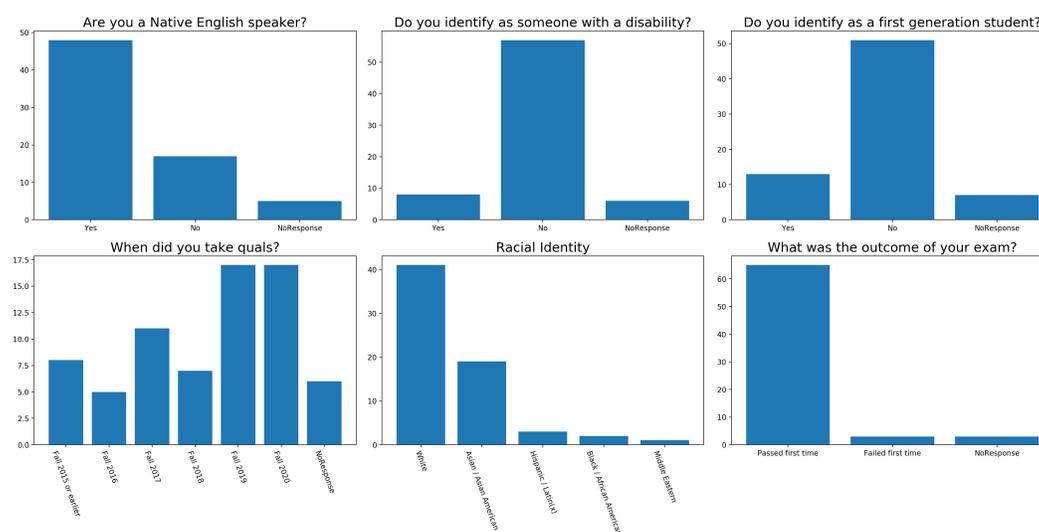


Figure A.7: Demographic information on survey participants for categories that could not be dis-aggregated due to concerns of anonymity: a) Native English speaker, b) ability status, c) first-generation student status, d) year of exam, e) racial identity, and f) exam outcome.

### General impressions

First, we can investigate the aggregated results from the survey to understand overall student impressions of the exam. Figure A.8 shows responses to the eight numerical survey questions asking about student experience *before*, *during*, and *after* the exam. Overall, students felt most positively about their knowledge of the exam format and most negatively about their experience receiving feedback on the exam.

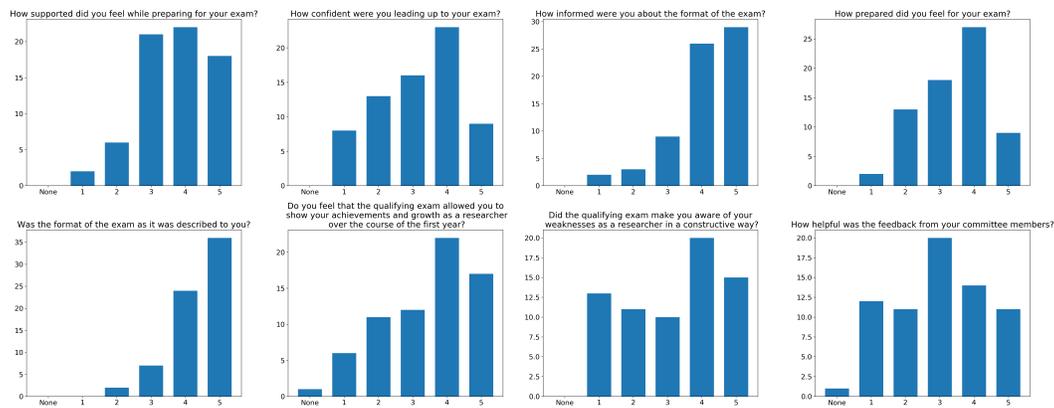


Figure A.8: Histograms of responses to questions with numerical answers (1–5) asking students how a) supported, b) confident, c) informed, d) prepared they felt before the exam, e) if the exam format was in fact as expected, f) whether the exam format allowed them to demonstrate their achievements, g) if the exam provided constructive criticisms, and h) how helpful the feedback was overall from the committee. Students felt most negatively about the feedback that was provided (g,h).

To gather more qualitative data on student experience during the 3 hour exam, we asked students to describe in 2-3 words how they felt when they finished. The word cloud in Figure A.9 shows these responses. Words are sized by the number of times they were written by students in answer to the survey. The most common feelings were relieved, exhausted, disappointed, tired, happy, confused, relaxed, excited, depressed, overwhelmed, demoralized, and so on.

The survey also asked open-ended questions to which students could respond with long-format answers. Students were asked about their experience during each phase of the first year (before, during, or after the exam). To summarize some of these responses, the answers were coded as “overall positive,” “neutral/mixed,” or “overall negative.” These results are shown in Figure A.10. Overall, the majority of students felt mixed or neutral before the exam, equally split during the exam, and majority negative after the exam. The time period when the most students felt overall positive was on the day of the exam itself, and the least students felt this way after the exam.

From the long-format responses, a theme emerged of disrespect by the committee towards the students. Students described situations of “hostile” questions, arguments between committee members, and being laughed at during the exam. Overall, an astounding number of students mentioned disrespectful behavior in their responses, including:



- 12% of respondents mentioned committee members falling asleep during the exam,
- 12% of respondents said they were asked questions in a rude/disrespectful manner, and
- 7% of respondents mentioned committee members arriving late or leaving early.

Finally, long-format answers were coded with respect to how students described their interactions with different groups of people. Figure A.11 shows how students characterized their interactions with their faculty “advisors” compared to other student or postdoc “peers” in their experience leading up to the exam. When responses mentioned interactions with either of these groups, the response was ranked as either positive or negative. Overall, many more students wrote about interactions with their faculty advisors compared to interactions with peers. About 20% more interactions with faculty were ranked as negative than positive, compared to about 50% more of cited interactions with students or postdocs being ranked positive than negative. Additionally, many responses described the power imbalance between students and faculty and how difficulties with communication or other interactions often arose from this inherent, but often ignored, power imbalance.

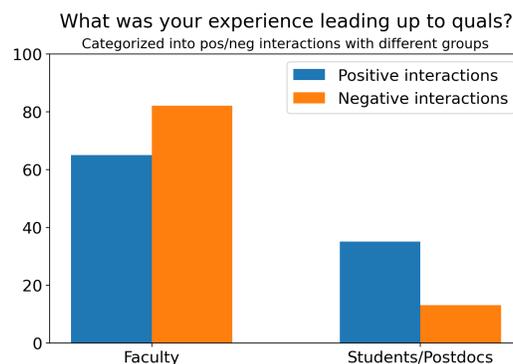


Figure A.11: Percentage of responses that described positive or negative interactions with faculty or with other students and postdocs. Responses could indicate both positive and negative interactions with either or both groups, so the total does not sum to 100%. Overall, students described interacting more often with faculty than other students/postdocs and these interactions were described as both positive and negative at fairly comparable rates, but with more negative interactions described. Fewer descriptions of interactions with students/postdocs existed, but these were mostly described as positive.

## Gender discrepancies

For all of the questions asked, we can disaggregate the responses by the students' gender identity. For other identity characteristics this was not always possible while still retaining anonymity due to the small sample size. Similarly for gender only a gender binary (men and women) is presented in these results because the number of nonbinary respondents was too small to present anonymously. Therefore, this analysis only focuses on gender discrepancies in exam experience. However, the same questions asked here can and should be asked pertaining to other identity characteristics like race, first-generation status, and more.

We focus here on the numerical questions about exam experience shown in Figure A.8. Figure A.12 shows the same results, but split by gender. The results here are striking and show that across the whole timeline — before, during, and after the exam — women are feeling less supported, less prepared, and receiving worse feedback than their male peers.

1. For all questions except b (confidence) and d (preparedness), the mode answer from women was less than the mode answer from men.
2. For all questions more women responded with numerical values 1 or 2 than men, and fewer responded with 4 and 5.
3. For questions c to e (information, preparedness, format knowledge), only women responded with the lowest value.

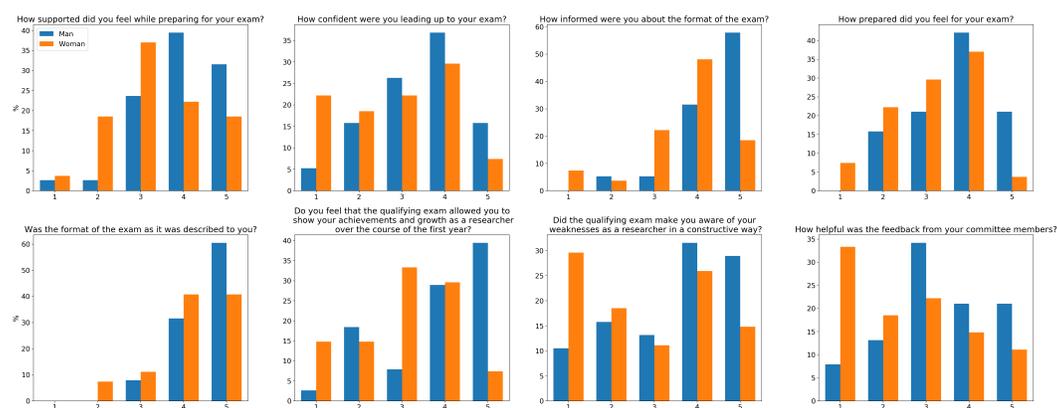


Figure A.12: Histogram of scores for each of the eight questions with numerical 1–5 answers broken down by gender of respondent, normalized by total number of respondents from each category.

To help summarize these results, Figure A.13 shows the average score on each of the eight questions separately for men and women, with  $0.5\sigma$  error bars. For all eight questions, the average score from women was less than from men, though of course with considerable spread (as evident from the full histograms in Figure A.12).

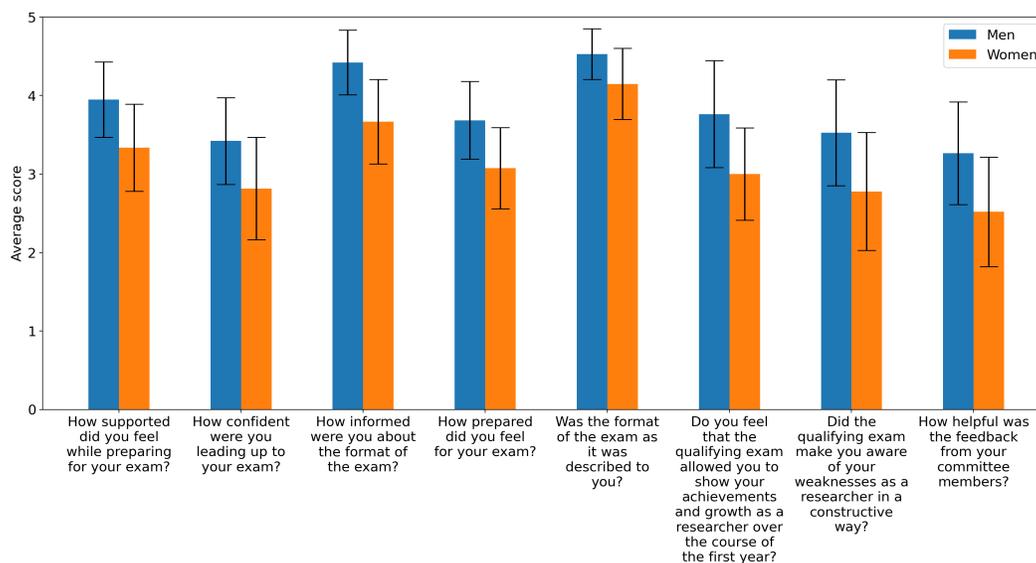


Figure A.13: Average score for each of the eight questions with numerical 1–5 answers aggregated by gender of respondent. The first four questions ask about experiences before the exam, the second three about the experience during the exam, and the last question about the feedback received after the exam. For all eight questions the average response from women was lower than from men. The error bars show  $0.5\sigma$  to represent the spread in answers amongst respondents.

### Recommendations to the Division

Based on the survey results, we recommended a number of improvements the Division could make to the examination itself and the process surrounding first year advising. These recommendations were discussed with the Core Committee in Summer 2022.

1. Have committee members write down 3 exam successes and 3 areas for improvement immediately after the exam in order to provide more actionable and specific feedback.
2. Empower committee chairs to call out un-attentive, distracted, or sleeping members during the exam.

3. Advertise Caltech Accessibility Student Services and consider making the exam shorter or splitting into two sessions in order to provide students with disabilities an equitable exam-taking experience.
4. Openly discuss the possibility of failing and what happens with regard to financial support, visa status, and academics in order to reduce undue stress on first year students.
5. Write and publish a statement describing the faculty goals for the qualifying exam and what it is testing and how success is measured.
6. Require mandatory yearly faculty mentorship training and scheduled check-ins between advisors and first years.
7. Create a universal course for first years that covers presentation skills, writing an abstract, and building mentorship circles.
8. Organize communication between G1s and G2+s to learn about quals organized by option representatives.
9. Offer oral finals instead of written finals in most GPS graduate classes.
10. Offer spring term sessions or a class where students can practice talking about their research in a low-stakes environment.
11. Change the required paper to a poster that can be engaged with by the full division and give student practice speaking about their work.
12. Consider more flexible scheduling (spread exams over a semester).
13. Maintain a database of the qualifying exam results by gender identity, race, and other protected characteristics.

### **Recent progress**

Continued conversations with the qualifying exam committee have extended until the writing of this thesis. Some changes already implemented in the exam process are documented here. The authors recognize the efforts of Professor Mike Brown in working to improve the qualifying exam since 2022, some of which is documented here: [https://www.gps.caltech.edu/documents/4829/Core\\_report\\_2022.pdf](https://www.gps.caltech.edu/documents/4829/Core_report_2022.pdf)

1. 1-credit lunchtime seminar course for first year students that covers aspects ranging from how to find an advisor, to how to conduct fieldwork safely and ethically, to how to get involved in outreach, to where on campus to find resources from scientific writing to gender harassment support.
2. Mandatory acknowledgement from students that their proposition advisors have signed-off on their project abstracts when they are first submitted in May.
3. Mid-summer (July) progress report to ensure that students and advisors are communicating appropriately about expectations.
4. Annotated rubric provided to students before the exam that clearly explains what each requirement means.
5. Examining committee meets one week prior to exam to decide upon a list of topics to cover based on the submitted abstracts.
6. A record of the past topic/abstract pairs will be made publicly available for students to read while preparing for their own exam.
7. Reduction of committee size from five to four members to reduce faculty burn out during exam season and increase attentiveness from each faculty during the examinations.
8. Detailed written feedback letter sent to students after the exam.<sup>2</sup>

---

<sup>2</sup>This does not yet apply to the ESE option.