Verification of cloud production in the Community Atmosphere Model: A comparison of two data assimilation techniques

Bethany Sutherland

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2018

Reading Committee:

Chris Bretherton, Chair

Ka-Kit Tung

Program Authorized to Offer Degree:
Department of Applied Mathematics

University of Washington

Abstract

Verification of cloud production in the
Community Atmosphere Model:
A comparison of two data assimilation techniques

Bethany Sutherland

Chair of the Supervisory Committee:
Professor Chris Bretherton
Atmospheric Sciences and Applied Mathematics

Clouds play an important role in regulating our climate, and it is vital that we are able to accurately simulate them in global climate models. Data assimilation can be used to force the model towards simulating a specific historical event, and then the clouds simulated can be compared to those observed during the event. Two data assimilation methods, Newtonian relaxation and Kalman filtering, are used to force the Community Atmosphere Model version 6 towards the state of the atmosphere that was observed in July of 2013. The use of the two methods produced significantly different results. The results produced using Newtonian relaxation indicate the Community Atmosphere Model simulates clouds reasonable well. The results produced using Kalman filtering support the conclusion that there may have been an error in the way the model was set up for that run.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AMSR	**A**dvanced **M**icrowave **S**canning **R**adiometer
CAM	**C**ommunity **A**tmosphere **M**odel
CERES	**C**louds and the **E**arth's **R**adiant **E**nergy **S**ystem
DART	**D**ata **A**ssimilation **R**esearch **T**estbed
FDDA	**F**our-**D**imensional **D**ata **A**ssimilation
GCM	**G**lobal **C**limate **M**odel
LWP	**L**iquid **W**ater **P**ath
NASA	**N**ational **A**eronautics and **S**pace **A**dministration
NCAR	**N**ational **C**enter for **A**tmospheric **R**esearch
NOAA	**N**ational **O**ceanic and **A**tmospheric **A**dministration
OLR	**O**utgoing **L**ongwave **R**adiation
Q	**S**pecific **H**umidity
RH	**R**elative **H**umidity
RSW	**R**eflected **S**hort **W**ave
SOCRATES	**S**outhern **O**cean **C**louds **R**adiation **A**erosol **T**ransport **E**xperimental **S**tudy
T	**T**emperature
TOA	**T**op **O**f **A**tmosphere

ACKNOWLEDGMENTS

I would like to express my utmost appreciation to Professor Chris Bretherton for advising me this year and for teaching me so much about our atmosphere, conducting research, and performing data analysis. These lessons will be invaluable to me as I now begin my scientific career.

I would also like to thank Professor Ka-Kit Tung for sitting on my committee. I want to express my gratitude to Andrew Gettelman and Kevin Raeder for providing me with the CAM results and to Chris Jones who kindly provided me his MATLAB codes to use as a starting place. Chris's codes proved extremely helpful and I learned a great deal about programming while trying to use and modify them. Thank you to the National Science Foundation for providing funding for this work through NSF grant AGS-1660604.

Last but certainly not least, I would like to thanks my wonderful husband for his endless support and encouragement in all of my endeavors.

DEDICATION

to my father,
for inspiring my love of science
and for teaching me to think deeply and critically.

"don't guess"

Chapter 1

INTRODUCTION

Clouds play a pivotal role in the regulation of our climate. Depending on their composition distribution, and elevation, clouds can have a cooling effect, a warming effect or neither [6]. In order to make well founded predictions of the way the climate is likely to evolve in the future we must be able to include clouds and cloud feedbacks in our climate models in a realistic and physically consistent manner. The goal of this project is to verify version 6 of the Community Atmosphere Model's (CAM6) ability to to simulate clouds. This is done by using data assimilation techniques to force certain aspects of the model towards a historical event, and then comparing the clouds simulated with those observed during that time. One motivation for this analysis is to determine if CAM6 is a good candidate for use in conjunction with analysis of data obtained during the Southern Ocean Clouds Radiation Aerosol Transport Experimental Study (SOCRATES), a field campaign conducted in early 2018.

1.1 The role of clouds in regulating Earth's energy balance

The earth climate system receives energy from the sun in the form of radiation. Clouds affect the Earth's energy budget by reflecting incoming radiation, absorbing and re-emitting outgoing longwave (infrared) radiation (like an atmospheric blanket), and redistributing radiant energy around the globe. Figure 1.1 shows many of the ways radiation reacts with our atmosphere and the role clouds play in the system. Small changes in the distribution of clouds can have significant impacts on the global climate; being able to accurately simulate clouds in our global climate models is vital to understanding how the climate will evolve.

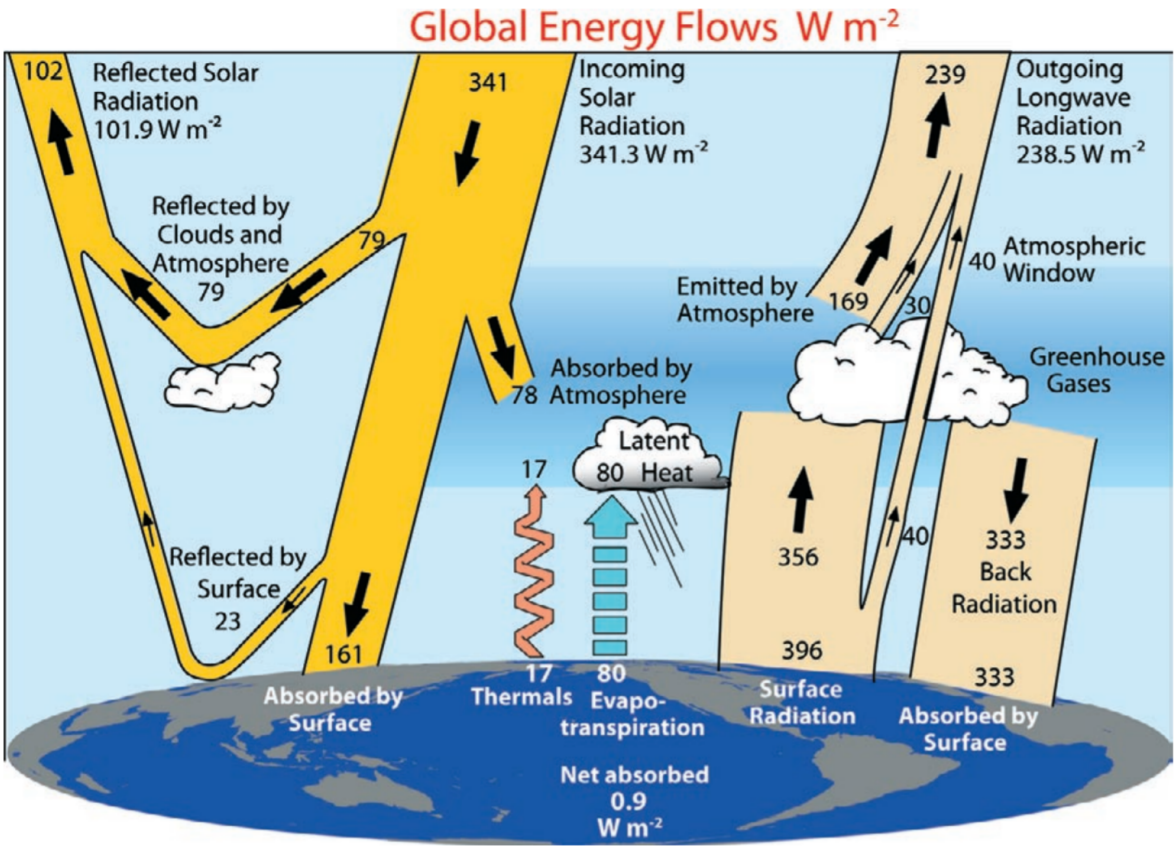

Figure 1.1: Figure reproducted from Trenberth et al. 2009. "The global annual mean Earth's energy budget for the Mar 2000 to May 2004 period (Wm^{-2}). The broad arrows indicate the schematic flow of energy in proportion to their importance." [14]

1.1.1 Cloud cooling effect

Low altitude clouds are generally optically thick, meaning they intercept radiation effectively. Low clouds also have a high albedo which means they reflect a high amount of the incoming solar radiation. By reflecting incoming radiation back out to space, clouds lower the amount of energy warming the climate. Clouds which are low in the atmosphere radiate energy at nearly the same temperature as the earth, so they do not have a significant effect on the emitted infrared radiation. Thus, low clouds contribute to cooling the system by increas-

ing the albedo and lowering the amount of absorbed solar radiation without significantly changing the emitted longwave radiation [1, 6].

1.1.2 Cloud warming effect

Clouds high in the atmosphere tend to have a low albedo, and do not intercept much of the incoming solar radiation. However they do absorb the outgoing longwave radiation. They then re-emit the energy both upwards towards space and back towards the ground, like an insulating blanket layer. The height of these clouds mean that they are emitting at a lower temperature than the earth and cause the amount of energy leaving the system to be reduced. Overall these clouds tend to contribute to warming because they reduce the outgoing longwave radiation, but do not absorb enough solar radiation to offset the longwave reduction.[1, 6].

1.2 Modeling Clouds

The way that clouds contribute to the overall energy budget depends on the size, composition, and distribution of clouds globally. Currently it has been shown that clouds have a cooling effect overall[1]. However the way that the balance between cooling and warming will shift due to anthropogenic changes is not well understood. The way cloud responds to warming will change the top of atmosphere (TOA) radiative flux is by a wide margin the largest source of uncertainty in GCMs' simulated response to a CO_2 forcing [4]. A necessary step towards credibly simulating cloud responses to anthropogenic forcings is being able to accurately simulate clouds in our current climate. One of the main challenges to modeling clouds is that they rely on small-scale physical processes which cannot be represented on the course grids utilized in GCMs [4]. Most of this small scale physics must therefore be accounted for using parameterizations which can be imprecise. Lack of cloud observations for model verification, and lack of understanding of cloud processes also contribute to the difficulty[1]. [15] showed that when comparing a GCM with cloud feedback (where the distribution of cloud cover is predicted by the model) to the same model with prescribed clouds, the climate sensitivity

Type of cloud	indicated by
low cloud	increased RSW, no change in OLR
high clouds	may increase RSW, decreased OLR

Table 1.1: Summary of the relationship between TOA radiation and clouds

to a doubling of CO_2 will be drastically different in the two cases. The surface temperatures of the cloud feedback case were found to increase 1.25 times that of the prescribed case. In a similar study [5] found that the cloud feedback might be even more pronounced.

1.3 Using TOA Radiation to understand clouds

Most of the data we have about the state of our atmosphere is obtained via remote sensing techniques. The use of systems of satellites, such as NASA's Clouds and the Earth's Radiant Energy System (CERES), enable us to gain frequent and periodic observations across the globe. Without satellites the amount of data we would have about the atmosphere would be very sparse. Therefore, we must utilize the data which can be remotely obtained for model verification. One of the things which can be easily measured using satellites is the TOA radiative fluxes in different wavelength bands.

We can determine cloud presence using TOA radiative fluxes. Because clouds usually have a higher albedo than the earth under it, the presence of clouds (except very thin ones) will increase the amount of reflected short wave (RSW) radiation. So generally places where the observed RSW at the TOA is particularly high indicate the presence of thick, reflective clouds in those locations. The earth and clouds emit long wave radiation according to their temperature. Clouds which are high in the atmosphere will emit at temperatures much lower than the earth is emitting at. So the presence of a high cloud will reduce the amount of outgoing longwave radiation (OLR) reaching the TOA, even if the cloud is thin. Therefore RSW can be used as a proxy for the presence of clouds and when combined with the OLR the general altitude of the clouds can be determined. [1].

Chapter 2

DATA ASSIMILATION METHODS

Data assimilation is a hybrid method of modeling which incorporates data measurements of the system with the governing equations. Both measurements and simulations are significantly hampered by uncertainty and the presence of noise and therefore neither is fully trustworthy. However by using the experimental measurements to help inform the model, and vice versa, the predictive ability of the model can be greatly improved [12]. Data assimilation methods have long been used with great success to incorporate additional data observations in numerical weather prediction[9].

Validating cloud production in GCMs is challenging because most climate studies are conducted with the model being integrated over periods of 30 years or longer. However observations are not available for most parameters associated with clouds for that length of time. A method for model validation over shorter time periods was needed. Jeuken et al. 1996 first proposed using the data assimilation technique of Newtonian relaxation as a method for validation of GCM's physical parameterization schemes [9]. Data assimilation can be used to constrain certain aspects of the model, the temperature and winds for example, towards a historical state. This then allows us to answer the following question: if those specific aspects of the model match the observed state, will the model simulate clouds similar to what was observed? [9].

Two different type of data assimilation, Newtonian relaxation and Kalman filtering, were used on Community Atmosphere Model version 6 (CAM6). CAM6 is the newest version of the atmospheric component of the Community Earth System Model (CESM). CESM was developed by the National Center for Atmospheric Research to be a community tool for investigating Earth system interactions [8]. The newest version has not been publicly

released as of printing, but the model source code and the documentation for CESM version 1.0 are freely available at http://www.cesm.ucar.edu/models/cesm1.0/.

Newtonian relaxation and Kalman filtering were both used for the period of July 2013 to determine how well CAM6 is able to simulate clouds. The run of CAM6 where Newtonian relaxation was used will hereafter be referred to as CAM-N, and the version where Kalman filtering was used will be referred to as CAM-DART.

2.1 Newtonian Relaxation - Nudging

Newtonian relaxation, also known as "nudging", is a method of data assimilation which relaxes the model towards the observed data by the addition of a non-physical relaxation term to the model equations.

$$\frac{\partial X}{\partial t} = F_m(X) + G(X_{obs} - X) \tag{2.1}$$

where X represents the model variable being nudged. F_m is the change in X caused by the dynamical and physical processes of the model which govern the evolution of X, this is how the X would evolve if the model were run freely. The nudging is done by the addition of the second term, where $(X_{obs} - X)$ is the difference between the model predicted value and the observed data and G is the relaxation coefficient[9] which has units of inverse time.

Choosing an appropriate G is important when doing Newtonian relaxation. Choosing too large of a G will cause the relaxation term to dominate in 2.1. However a G that is too small will mean that the inclusion of the observations will have little impact on the results[9]. [7] determined that the optimal value of the relaxation coefficient should depend on the variable being nudged, the observational accuracy, and the magnitude of the model forcing. However taking all this into account would lead to using G which varies in both time and space. This amount of complexity is usually not done in most applications[9]. [11] suggested that for consideration with clouds, a six hour relaxation time is reasonable because it is longer than the lifetime of an individual cloud, but is short enough that temperature perturbation are unlikely to influence the large scale circulation.

For CAM-N the temperature and wind fields were nudged towards the Modern-Era Retrospective analysis for Research and Applications (MERRA) dataset fields. The MERRA dataset is a reanalysis using the the Goddard Earth Observing System Data Assimilation System, Version 5 (GEOS-5) and is run by NASA's Global Modeling and Assimilation Office[3]. Andrew Gettelman of University Corporation for Atmospheric Research (UCAR) conducted the CAM-N run and chose a relaxation time of 48 hours because it displayed the best climate reproduction for clouds and radiative fluxes. Nudging the temperature and winds is advantageous, because the fields being constrained are a bit removed from those which are directly used to resolve the clouds.

2.2 Kalman Filtering - DART

Ensemble filters are a different method of data assimilation where a group of forecasts (ensemble) are used every time data is assimilated. The ensemble of forecasts acts as a random draw from the probability distribution of the models state which has already incorporated all previous observations [2].

Following [12, 13] the Kalman filter can be derived as follows. The full dynamics to advance the system from time t_k to time t_{k+1} without approximation is given by

$$x_{k+1} = f(x_k) + q_{k+1} \qquad \text{(truth)} \tag{2.2}$$

and the model approximation of the system is given by

$$x_{0_{k+1}} = f(x_{0_k}). \qquad \text{(model approx.)} \tag{2.3}$$

q_{k+1} is a Gaussian white noise sequence which accounts for missing physics that are not accounted for in the model. The analysis, x_{0_k}, is the best estimation of the state of the system at time t_k and is represented by the blue asterisks in Figure 2.1. Using (2.2) and (2.3) together and Taylor expanding $f(x_k)$ around x_{0_k} the error between the truth and the

forecast can be expressed as

$$x_{k+1} - x_{0_{k+1}} = f(x_k) - f(x_{0_k}) + q_{k+1} \tag{2.4}$$

$$= (x_k - x_{0_k})f'(x_{0_k}) + \frac{1}{2}(x_k - x_{0_k})^2 f''(x_{0_k}) + \cdots + q_{k+1}. \tag{2.5}$$

The expectation or the expected value of a random variable is a weighted average with weighting according the the probability that X assumes the value, and is defined as

$$E[X] = \sum_{x:p(x)>0} xp(x) \tag{2.6}$$

where $p(x)$ is the probability mass function. The error variance between the correct solution and the predicted one is then found by squaring (2.4), neglecting higher order terms, and taking the expectation

$$E[(x_{k+1} - x_{0_{k+1}})^2] = E[(x_k - x_{0_k})^2](f'(x_{0_k}))^2 + E[q_{k+1}^2]. \tag{2.7}$$

The error variances P_{k+1} and P_k at times t_{k+1} and t_k respectively can be defined as

$$P_{k+1} = E[(x_{k+1} - x_{0_{k+1}})^2] \tag{2.8}$$

$$P_k = E[(x_k - x_{0_k})^2] \tag{2.9}$$

P_{k+1} now contains the dynamics errors, and P_k accounts for the errors in the estimation of the initial state. Using this notation, Equation 2.7 can be rewritten as

$$P_{k+1} = P_k(f'(x_{0_k}))^2 + E[q_{k+1}^2]. \tag{2.10}$$

Using this measure of the error variance, along with an observation, x_{obs} and it's the error variance, R, the data assimilated prediction of the state of the system, $\bar{x}_{k+1}$, can be calculated using

$$\bar{x}_{k+1} = x_{0_{k+1}} + \frac{P_{k+1}}{P_{k+1} + R}(x_{obs} - x_{0_{k+1}}) \tag{2.11}$$

The adjustment to the model approximation, $x_{0_{k+1}}$, caused by the second term in 2.11 is called the innovation. Note that in the case where there is no observational error ($R = 0$),

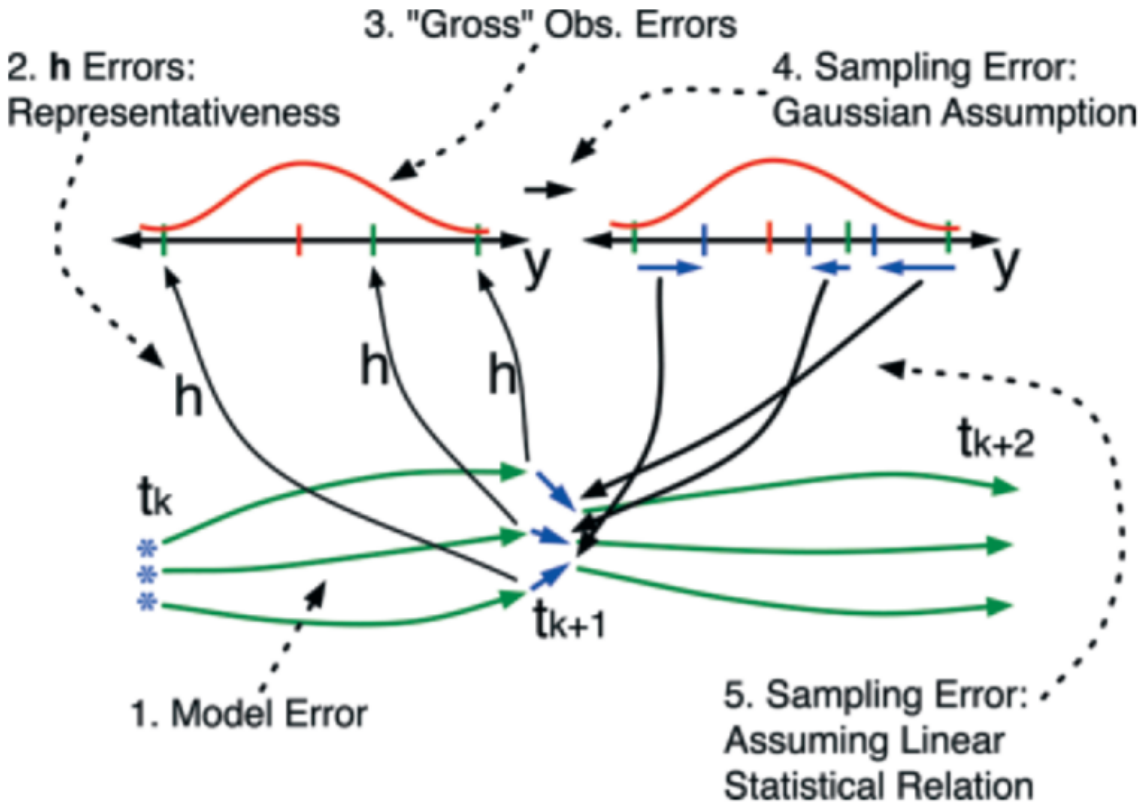

Figure 2.1: Figure reproduced from [2]. The steps in DART Kalman filtering: (1) Each model state at time t_k (blue asterisks) is advanced to time t_{k+1} by CAM (green vectors). (2) Forward observation operator, h, applied to each state to get estimates of observations from model (green ticks on upper axes) (3) The true observed value (red tick) and the observation likelihood (red curve, the probability that the instrument would have observed what it did if y were the true value) are combined with the ensemble estimate to obtain (4) an updated ensemble estimate (blue ticks) and increments (blue vector below top right axis) 5) The increments to the observation ensemble are regressed onto each state vector component independently to update the model states (blue vectors on end of green vectors) (6) The model is then used to advance the updated state estimates to time t_{k+2} (7) Repeat [2]

the prediction from 2.11 will take on exactly the value of the observation. See [12] for the extension of this to vector form.

The other method of data assimilation for CAM6, was done using the Data Assimilation Research Testbed (DART), a facility developed by the National Center for Atmospheric Research (NCAR) and were produced and provided by Kevin Raeder of NCAR. DART uses Kalman filtering on an ensemble of model forecasts to produce a prediction of the possible states of the system. One advantage to this is the the spread in each of the ensemble members gives an indication of the uncertainty of results, but an idea of what they system thinks is the most likely outcome can be found by taking an average of the ensemble. Figure 2.1 shows a simplified cartoon of the way that Kalman filtering is applied in DART using a 3 member ensemble example. For more on this assimilation techniques see [2]. DART assimilates a variety of observations, including radiosondes, surface meteorological observations, and profiles of refractive index from GPS satellite occulations. One disadvantage of the default DART implementation is that the majority of the observations being assimilated are sparse over the ocean.

Chapter 3

METHODS

The data assimilated model results were produced by Andrew Gettlman and Kevin Raeder of NCAR. The top of atmosphere radiation fluxes were compared with version 3A of CERES Daily Synoptic Radiative Fluxes and Clouds (SYN1deg-Day) product on a 1°grid [10]. The CERES product data is publicly available at http://ceres.larc.nasa.gov. Liquid water path comparisons were made using passive microwave satellite retrievals over the ocean from the Advanced Microwave Scanning Radiometer on JAXA's GCOM-W1 spacecraft (AMSR-2) which is sponsored by the NASA Stand Alone Mission of Opportunity Program [10]. The data from this satellite archived and produced by Remote Sensing Systems (RSS) and can be found at http://www.remss.com.

The method of data analysis used here follows that used in [10].

CAM6 outputs the net longwave flux at top of model (OLR) and the net solar flux at top of model. The RSW was calculated by subtracting the net solar flux at TOA from the solar insulation. The monthly mean solar insulation was used as a daily value was not available. It is believed that this does not contribute significantly to the results. Particularly since the effects would be averaged out for the monthly averaged results. Additionally this should not affect the differences seen in the two different DA runs, since both were corrected in the same manner.

CAM-N outputs every 6 hours on a grid with 192 latitude points, 288 longitude points, and 32 vertical levels for the entire month of July 2013. CAM-DART outputs 40 individual runs on the same grid and at the same times as CAM-N but only for the period of July 19th 2013 through July 30th 2013. The daily average for each CAM6 variable on the grid was calculated for comparison with the daily CERES data. As the CERES data was on a

1° grid it was more coarse than the CAM6 output in the latitudinal direction, but finer than the CAM6 output in the longitudinal direction. Therefore the coarsest grid for each direction with 180 latitude points and 288 longitude points was used, this will be called the analysis grid. Each of the CAM6 and CERES data sets were individually regridded to the analysis grid by averaging the data from all cells that fall within each analysis grid cell. The ensemble average of all 40 CAM-DART was used for this purpose.

The AMSR-2 data for the liquid water path is on a 0.25° grid and therefore is much finer than the CAM6 outputs in both directions. Therefore the AMSR-2 data was regridded to match the original CAM6 grid and the two can be compared without regridding the CAM6 data as well. The observed LWP in each grid cell is compared with the CAM6 LWP within that grid cell which occurred at the nearest time to when the observation was taken. Cells where the average time differed from its neighboring cells by more than 6 hours are excluded to account for not wanting to average data from two adjacent passes of the satellite.

The difference between the model predicted value and the observed value for each variable was found by subtracting the observed value from the model value for each day. The bias of the model relative to CERES daily observations for TOA OLR and RSW is calculated by taking the area weighted global average of the difference, ignoring missing data. The root mean square error (RMSE) is the square root of the global area weighted average of the square of the difference.

Monthly averages are also calculated ignoring any missing data.

Chapter 4

RESULTS

The difference between the TOA radiation values between the two versions of CAM6 and CERES is shown in Figure 4.1. The CAM-N results are on the left and CAM-DART is on the right. The OLR differences are on the top row and the RSW differences are on the bottom row. CAM-N could be considered the best, in that it has a lower root mean squared error for both the OLR and RSW. These results indicate that CAM6 does resolve clouds reasonable well and that data assimilation using Newtonian relaxation of the temperature and wind fields, with a 48 hour timescale might be the best method of constraining the

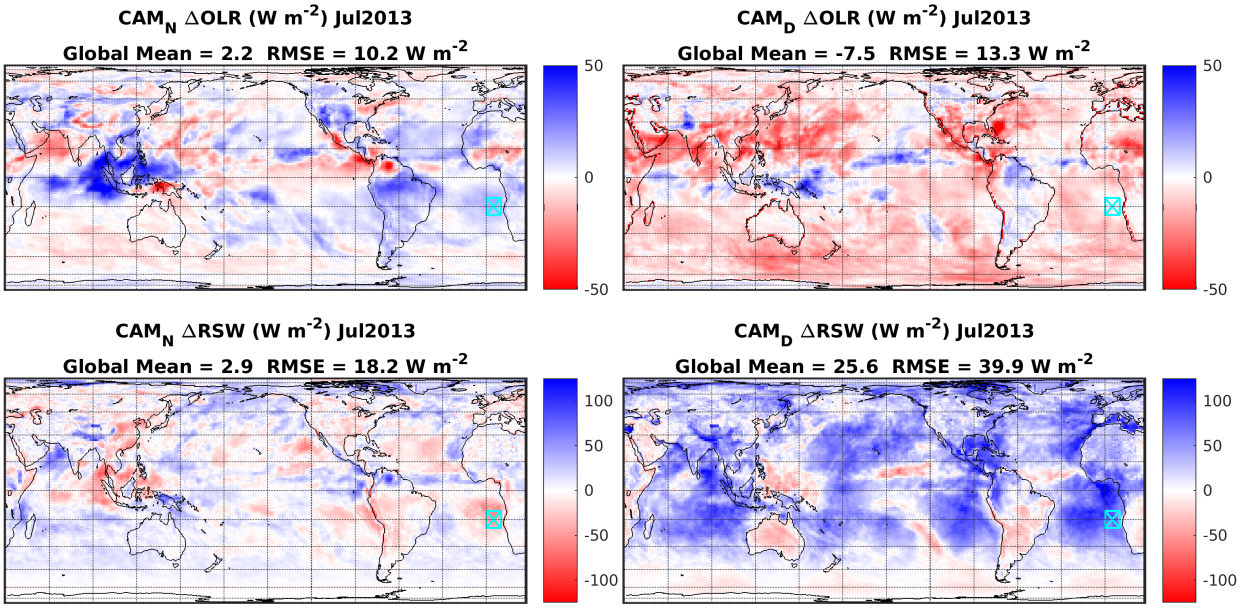

Figure 4.1: The difference between to TOA radiation model results and CERES observed data for July 2013. Averaged over the entire month for CAM-N (left), and averaged over the period between July 19th and July 30th for CAM-DART(right).

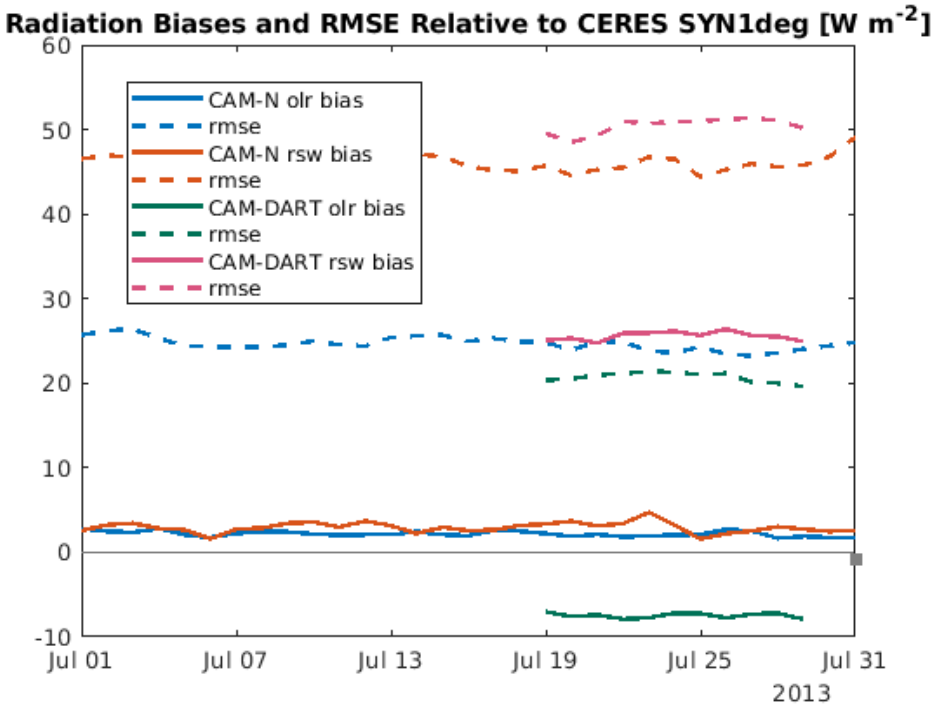

Figure 4.2: Average biases (solid lines) and root mean squared error (dashed lines) for OLR (blue hues) and RSW(red hues) for both versions of CAM for each day in July 2013.

model for comparison with observed data.

Figure 4.2 shows the average global RMSE and bias for each day in the period. The consistency of these results indicates that the biases are consistent and robust. This means that they are systematic, and not likely due to a particular time sampled.

The difference between the AMSR-2 LWP data and modeled LWP are plotted on a logarithmic scale in Figure 4.3. The global mean LWP measured by AMSR-2 is significantly higher than both models, despite the low bias in RSW. A pattern of lower LWP is present in both models near the equator and particularly in the maritime continent between India and Australia. However there are large regions of too high LWP in CAM-DART which are not present in CAM-N.

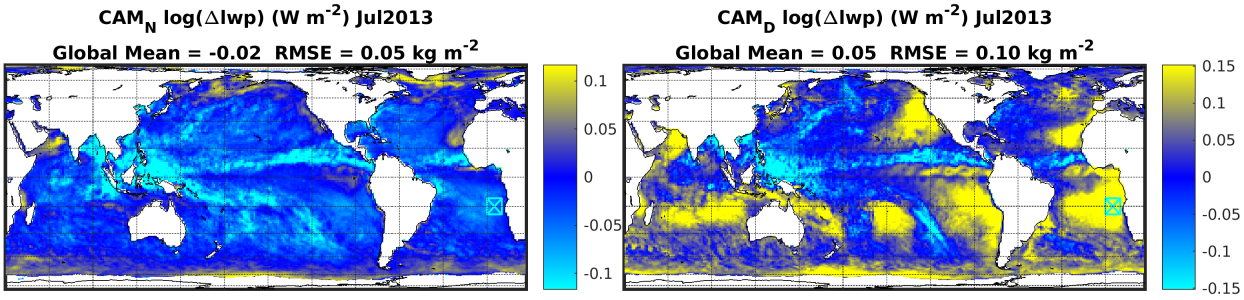

Figure 4.3: The difference between CAM liquid water path results and AMSR observed data for July 2013. The log of the differences is shown to make the small changes more apparent.

4.1 Interpreting the results

An understanding of cloud production can be reached by using information from the OLR, RSW and LWP together. For example, the results for CAM-N can be interpreted as follows. In the bottom left panel of Figure 4.1 the small positive bias in the ΔRSW indicate that CAM-N is slightly overproducing clouds in general. Looking at the cyan boxed region off the western coast of Africa specifically, negative ΔRSW indicates an underproduction of clouds. We can learn more about the types of clouds not being produced by looking at the OLR (top left panel of Figure 4.1 and LWP (left panel of Figure 4.3). The positive ΔOLR indicates that that there is too little high clouds in this region, while the negative ΔLWP as shown in Figure 4.3 suggests that low clouds are being under-produced as well.

In contrast, the area in the Pacific ITCZ, and Central America in particular, show the model is producing too much RSW, and therefore too much clouds. The negative ΔOLR indicates that there is too much high clouds. The negative ΔLWP shows that even though it is overproducing clouds in the region in general, it is still slightly under-producing the low clouds here.

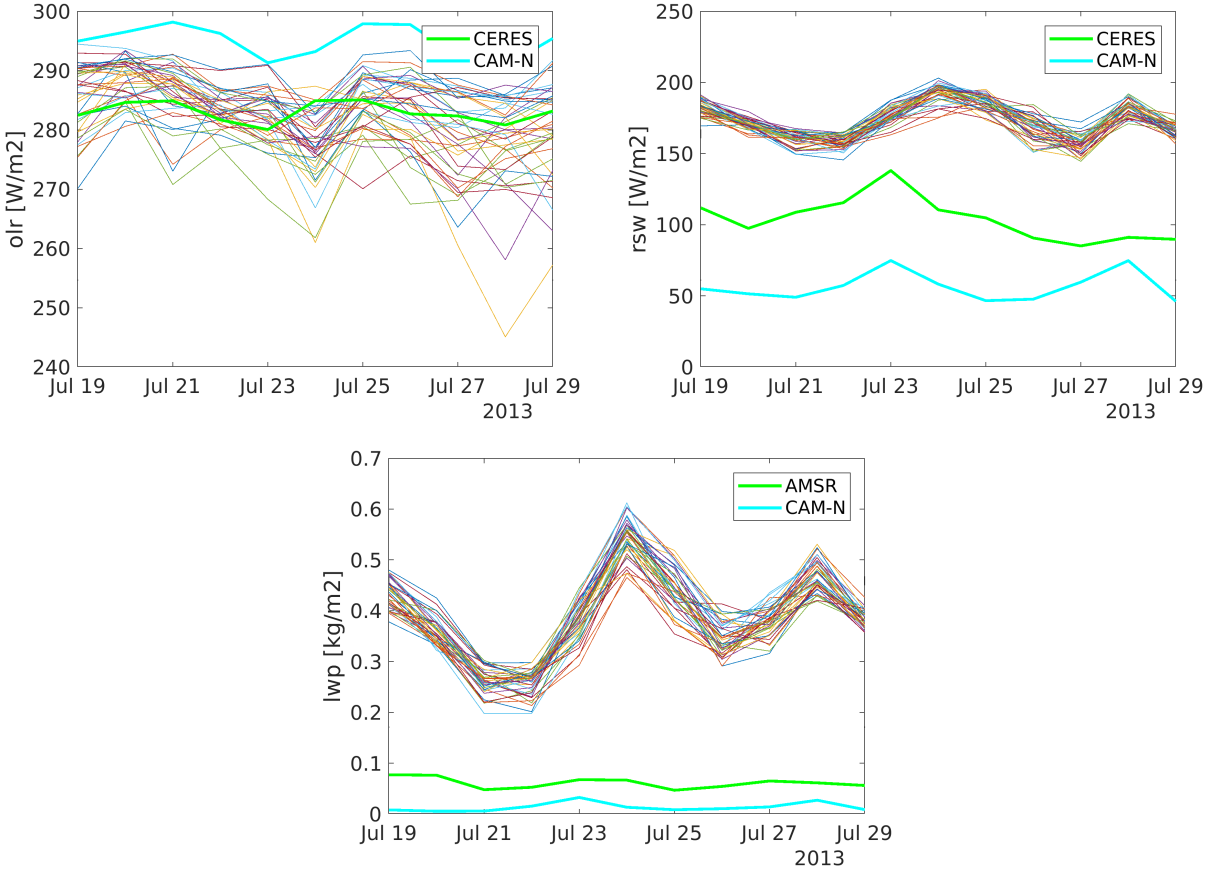

Figure 4.4: Individual CAM-DART runs and CAM-N averaged over the boxed region for each day in July

4.2 Regional analysis

One surprising result is the large difference between the two tests of CAM. The differences are evident both in the TOA radiation values as seen in Figure 4.1 as well as in the liquid water paths as seen in Figure 4.3. The nudged version performed significantly better than the DART version. Since DART uses an ensemble of 40 runs, it is significantly more computationally expensive, and the preliminary results indicate the extra expense does not pay off. In order to try to understand the differences between the two results, and why the different methods of assimilating data into the same model would produce such different outcomes,

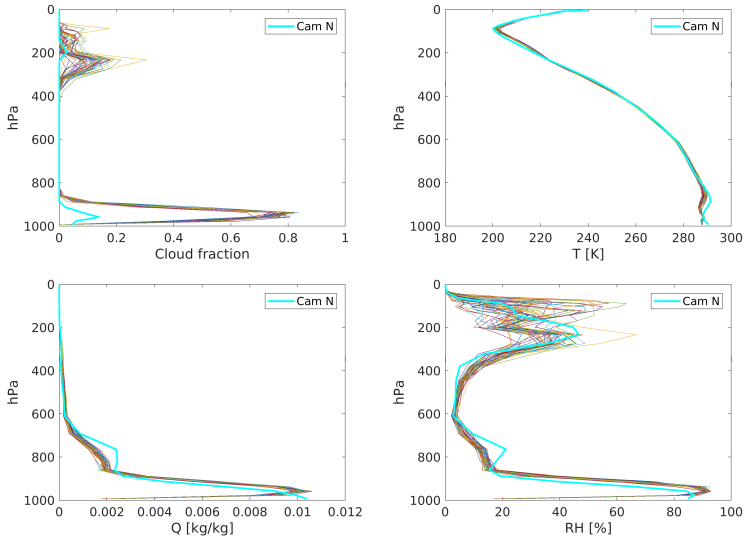

Figure 4.5: Vertical profile of cloud fraction (top left), temperature (top right), specific humidity (bottom left) and relative humidity (bottom right) for individual CAM-DART runs and CAM-N averaged over the boxed region for July 24th 2013.

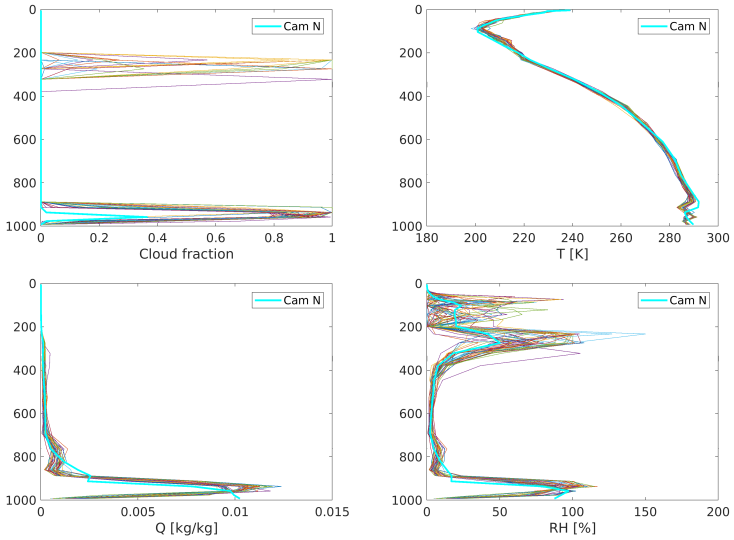

Figure 4.6: Vertical profile of cloud fraction (top left), temperature (top right), specific humidity (bottom left) and relative humidity (bottom right) for individual CAM-DART runs and CAM-N for a single grid point located at -15.5497° latitude and 5° longitude on July 24th 2013 at noon.

a regional analysis was performed on a subset of the data for a region off of the western coast of Africa. The region used was the area between -10° and -20° latitude and 0° and 10° longitude (shown in cyan box in Figures 4.1 and 4.3). In this region both CAM-N and CAM-DART show clear deviations from the observed data, and they deviate in opposite directions. In this analysis we examined each of the 40 CAM-DART runs individually in comparison with the CAM-N results.

In Figure 4.4 the observed OLR is within the scope of the individual DART runs (top left) and in this case CAM-N appears to be the biased one. The RSW (top right), and LWP (bottom left) show significant and opposite biases for both versions of the model. This indicates much of the excess clouds produced by CAM-DART are low clouds which would have similar OLR to the surface, but would effectively reflect shortwave radiation from the sun.

Additionally to understand the vertical structure of the differences between the two versions, several of the 3D variables are plotted with height. The average for the entire cyan box region at each height is shown in Figure 4.5, and Figure 4.6 shows the vertical profile for a single point at -15.5° latitude and 5° longitude. The top left of these figures confirms that the cloud fraction for each CAM-DART run is significantly larger than what is seen in the CAM-N results. The temperature profiles are in fairly good agreement (top right), although the DART temperatures do tend to skew slightly cold in the lower cloud region and slightly warm in the high cloud region. From the bottom we can see that there are large differences in the humidity between the two, which is likely the cause of the much higher amount of clouds in CAM-DART than in CAM-N. Of particular interest is the minimum in humidity near the surface of the CAM-DART runs which does not show up in CAM-N. Given that the region being analyzed is over the ocean, this minimum in humidity at the surface seems to be an unphysical phenomena and the lack of such a phenomena in the CAM-N run means that this may be an indicator of something going wrong in CAM-DART.

To try to determine why this drop in humidity near the surfaces is occurring in DART the innovations for each time step were calculated. This was done by looking at the average of all

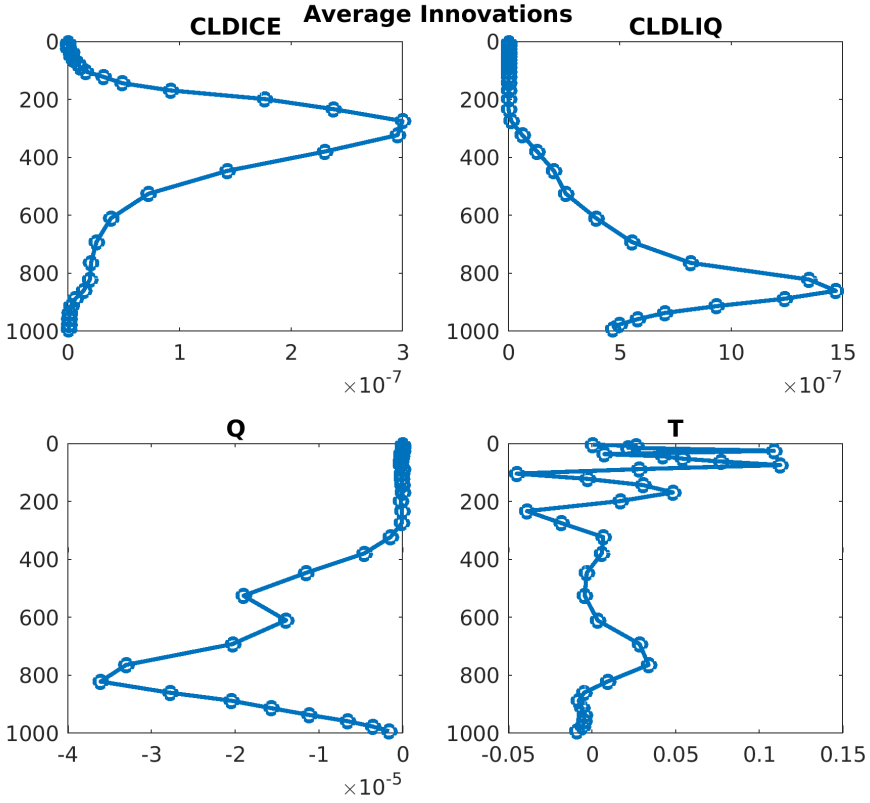

Figure 4.7: Global average innovations for all runs of CAM-DART for entire time period.

the CAM-DART runs and subtracting the pre-assimilation value from the post-assimilation value at each time. The innovation at each level averaged over the globe and over the time period is shown in Figure 4.7. The bottom left panel shows that the affect of the data assimilation on the specific humidity approaches zero at the surface. Figure 4.8 shows the affect of the data assimilation over time, and confirms that the data assimilation is not having a significant impact on the surface specific humidity.

4.3 Discussion

Through the use of data assimilation we have been able to confirm that CAM6 simulates clouds fairly successfully and is a good candidate for comparison with the SOCRATES data.

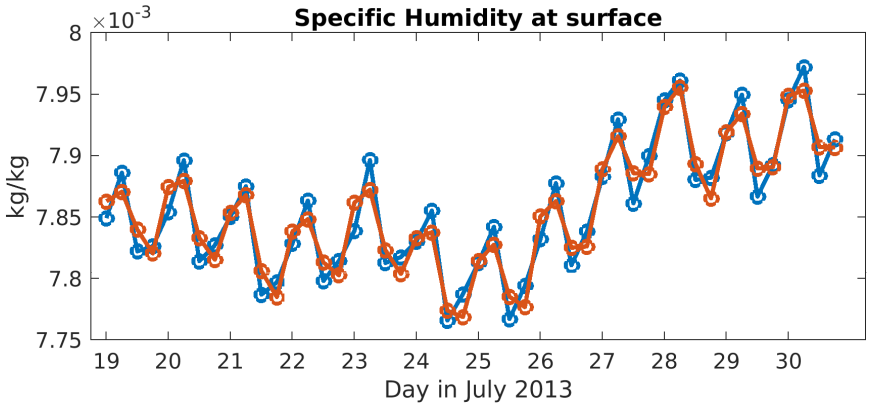

Figure 4.8: Global average innovations for all runs of CAM-DART for entire time period.

A significant difference was noted when different form of data assimilation were used. The source of the large discrepancy between the two data assimilation methods is still uncertain. However we were able to show the presence of a unphysically low humidity at the surface of CAM-DART is not caused by the data assimilation. We therefore conclude that something may have been done incorrectly in the set up of the DART run. We plan to continue working with colleagues at NCAR/UCAR to hopefully fix it and then make a better comparison.

BIBLIOGRAPHY

[1] National Aeronautics and Space Administration. The importance of understanding clouds. *NASA Facts*.

[2] Jeffrey Anderson, Tim Hoar, Kevin Raeder, Hui Liu, Nancy Collins, Ryan Torn, and Avelino Avellano. The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, 90(9):1283–1296, 2009.

[3] S. W. Berrick, D. Ostrenga, and S. Shen. Modern Era Retrospective-analysis for Research and Applications (MERRA) Data and Services at the GES DISC. *AGU Fall Meeting Abstracts*, pages A13A–0225, December 2008.

[4] Zelinka Mark D. Hartmann Dennis L. Ceppi Paulo, Brient Florent. Cloud feedback mechanisms and their representation in global climate models. *WIREs Clim Change*, 2017.

[5] J. Hansen, A. Lacis, D. Rind, G. Russell, P. Stone, I. Fung, R. Ruedy, and J. Lerner. *Climate Sensitivity: Analysis of Feedback Mechanisms*, pages 130–163. American Geophysical Union (AGU), 1984.

[6] Dennis L. Hartmann. *Global Physical Climatology (Second Edition)*. Elsevier, Boston, USA, 2016.

[7] James E. Hoke and Richard A. Anthes. The initialization of numerical models by a dynamic-initialization technique. *Monthly Weather Review*, 104(12):1551–1556, 1976.

[8] James W. Hurrell, M. M. Holland, P. R. Gent, S. Ghan, Jennifer E. Kay, P. J. Kushner, J.-F. Lamarque, W. G. Large, D. Lawrence, K. Lindsay, W. H. Lipscomb, M. C. Long,

N. Mahowald, D. R. Marsh, R. B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W. D. Collins, J. J. Hack, J. Kiehl, and S. Marshall. The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360, 2013.

[9] A. B. M. Jeuken, P. C. Siegmund, L. C. Heijboer, J. Feichter, and L. Bengtsson. On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation. *Journal of Geophysical Research: Atmospheres*, 101(D12):16939–16950, 1996.

[10] C. S. Bretherton R. Sun J. Han H. Guo Jones, C. R. and M. Zhao. Hindcast cloud and radiation diagnostics for ncep gfs and gfdl am3. submitted 11/2016, rejected, resubmitted 4/2018, 2018.

[11] Gabriel J. Kooperman, Michael S. Pritchard, Steven J. Ghan, Minghuai Wang, Richard C. J. Somerville, and Lynn M. Russell. Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the community atmosphere model 5. *Journal of Geophysical Research: Atmospheres*, 117(D23):n/a–n/a, 2012. D23204.

[12] J. Nathan Kutz. *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data*. Oxford University Press, Inc., New York, NY, USA, 2013.

[13] Robert N. Miller, Michael Ghil, and Franois Gauthiez. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51(8):1037–1056, 1994.

[14] Kevin E. Trenberth, John T. Fasullo, and Jeffrey Kiehl. Earth's global energy budget. *Bulletin of the American Meteorological Society*, 90(3):311–324, 2009.

[15] R. T. Wetherald and S. Manabe. Cloud feedback processes in a general circulation model. *Journal of the Atmospheric Sciences*, 45(8):1397–1416, 1988.