Final Technical Report – CMDV (CM)4, University of Washington contribution
Report number: DOE-UWASH-16433-1
Institution: Department of Atmospheric Sciences, University of Washington, Seattle
Institutional PI: Christopher S. Bretherton

Abstract

The original goal of this project was to diagnose and improve CLUBB, the turbulence and cloud fraction parameterization used in DOE's E3SM model. For this purpose, we planned to use large-eddy simulation (LES) and ARM observations from the NE Pacific Ocean and the SGP site, in coordination with ARM's LASSO program. It was determined that for these cloud regimes, many of the turbulent 'moment closures' which underlie CLUBB's mathematical formulation are inconsistent with LES, which is an appropriate benchmark for testing this. The research assistant found that these closures could be more accurately formulated using machine learning using the LES as a training dataset. However the resulting parameterization proved to quickly drift away from physical plausibility. A novel machine-learning based boundary layer parameterization called MARBLE based on matching the time evolution of a parameterized cloud-topped boundary layer to reanalysis was then developed and published.

Subject Categories

54 (Environmental Sciences)
97 (Mathematics and computing)

Keywords

Cloudy boundary layers; machine learning; neural net; climate modeling; atmospheric boundary layer parameterization

Summary of major results

Our plan at the start of $(CM)^4$ was to diagnose and improve CLUBB, the turbulence and cloud fraction parameterization used in E3SM. For this purpose, we planned to use large-eddy simulation (LES) and ARM observations from the NE Pacific Ocean during the MAGIC program and from the SGP site during RACORO and ongoing shallow convection episodes, in coordination with LASSO.

Graduate research assistant Jeremy McGibbon, the primary contributor working with PI Bretherton on the UW component of the $(CM)^4$ project, started work on this. To predict atmospheric boundary layer and cloud evolution, CLUBB prognoses 11 turbulent moments in each grid cell (grid-means of temperature, humidity and horizontal wind components, 6 variances and covariances of temperature, humidity and vertical velocity, and the third power of vertical velocity). The prognostic equations for these equations contain numerous other high-order turbulence moments that must be related to these 11 variables via human-designed 'closure' assumptions. Jeremy quickly found that many of CLUBB's closure assumptions do not hold very well in LES. He lost confidence that incremental improvement of CLUBB was a good road to progress. This led to the first thrust of our research, CHOMP, which used machine learning to derive more accurate closure assumptions using LES as a training dataset. However, despite considerable effort, implementing these closure assumptions did not yield a version of CLUBB that produced physical plausible results for more than a couple of hours of simulation. This work is described in Section 1.

A year ago, we abandoned CHOMP and began work on a completely different machine learning boundary layer parameterization, MARBLE, which is trained exclusively on a high-resolution global reanalysis, ERA5. This parameterization has shown very promising results in a single-column setting, discussed in Section 2 and described in a paper to appear shortly in *Geophys. Res. Lett.* Usama Anber of BNL, part of Andy Vogelmann's group has been implementing it in the WRF model to see if it also produces good results when run as part of a regional weather/climate model with 50 km grid resolution. This is perhaps the most seminal achievement of our $(CM)^4$ work at the University of Washington.

Lastly, Jeremy, while working on this project, developed a collaboration with Joy Monteiro at U. Stockholm to develop a Python-based framework for hierarchical climate modeling called CliMT/Sympl. This work was published in GMD. For brevity, it will not be discussed in this report since it is tangential to the overall thrust of $(CM)^4$ even though it has attracted some outside interest from the climate model development community.

The $(CM)^4$ funding helped support Jeremy's PhD dissertation, completed in July 2019. The dissertation includes a comprehensive discussion of the CHOMP work, CliMT/Sympl, and MARBLE and is attached along with the two published paper as a resource/reference to all of these efforts. It describes a successful application of MARBLE to the SGP region that is not included in the published papers and gives technical details of the equations, machine learning schemes, technical implementation, etc.

1. CHOMP/CLUBB

CHOMP (Closure of Higher Orders by Machine-learning Parameterization) uses the same set of 11 prognostic higher-order turbulence closure equations as E3SM's boundary layer scheme CLUBB. CHOMP uses machine learning (ML) based on a suite of large-eddy simulations of cloud-topped boundary layers to 'close' unknown terms in these equation in terms of the prognosed moments. This is in contrast to CLUBB, which closes these terms with several assumptions involving the shape of the joint probability distribution between temperature, humidity and vertical velocity, as well as the correlations between joint moments of vertical velocity, buoyancy and vertical pressure gradient.

As described in Jeremy's thesis, CHOMP's approach more accurately represents the unknown terms in the equations compared to CLUBB. The primary testing and training dataset was an extensive suite of LES outputs for the MAGIC campaign, simulating about 1200 hours of marine stratocumulus and shallow cumulus in diverse weather conditions (McGibbon and Bretherton 2017). The LES was found to match the ship-observed cloud and boundary layer conditions for this suite of cases quite well, with little systematic bias. We have also tried training our ML scheme on RACORO and LASSO data for SGP shallow convection cases. The few days of model outputs proved too short to be a robust training dataset.

After much trial and error, we developed a customized polynomial regression approach to learning the needed turbulent moments from the LES that has the accuracy of a random forest but with outputs that are smooth functions of the inputs. This is a highly desirable feature for a parameterization, especially since we needed to take vertical derivatives of some moments. Fig. 1 compares error statistics from this scheme, a random forest, and using the CLUBB closures for the 12 moments that must be diagnosed within CLUBB. These errors are based on LES-diagnosed values over a range of initial profiles for one of the MAGIC cruises. The upper panel shows mean bias and the bottom panel shows the correlation coefficient with the LES results. In the upper panel, values near zero are desirable, and the two machine learning schemes clearly achieve this much better than CLUBB. In the lower panel, a high R^2 (fraction of variance explained in the LES-diagnosed moment) is best; again, the two machine learning schemes outperform CLUBB on almost all moments. Thus, we were hopeful that CHOMP would yield a more accurate version of CLUBB in a prognostic setting as well.

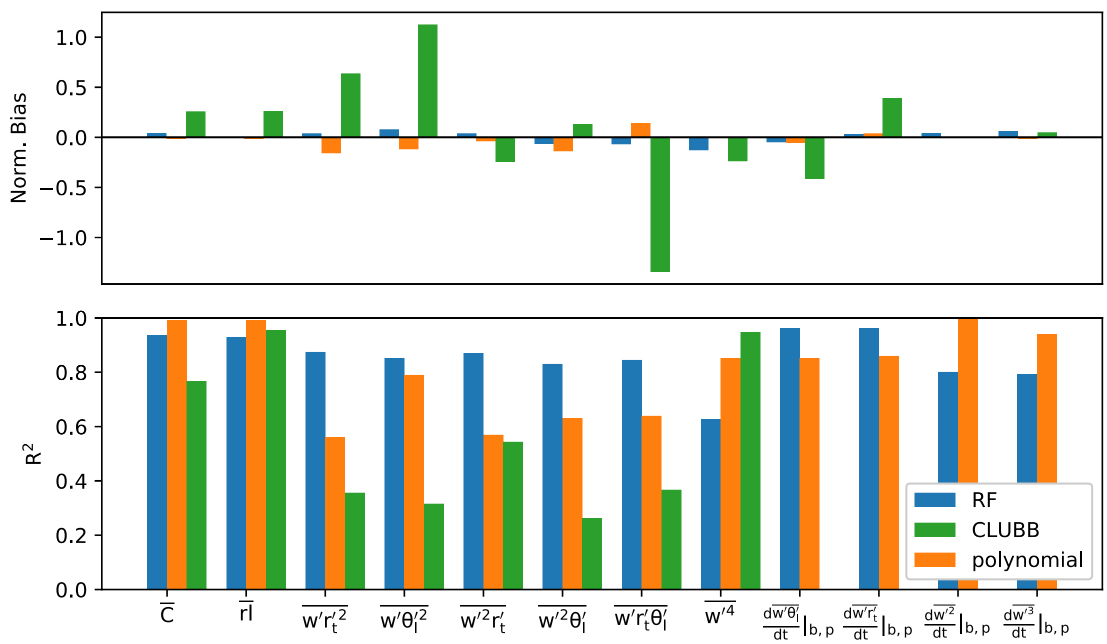

Fig. 1: Bias and squared correlation coefficient for CLUBB, random forest, and polynomial fit over Leg 15A. Evaluation is performed on values normalized such that their mean is zero and variance is 1 over the MAGIC training legs. Equal weighting is given to each model level.

CHOMP runs stably for periods of an hour or more when initialized starting with LES estimates of all its prognostic moments, but some of the prognostic moments that slowly drift away from their LES reference values, as illustrated in Fig. 1. Higher-order moments drift far worse, as shown in Figs. 3.8 and 3.9 of Jeremy's thesis. Despite a year of effort, we were unable to get CHOMP to converge to a quasi-steady equilibrium state given typical NE Pacific boundary layer thermodynamic profiles and forcings, so we decided to abandon this approach.

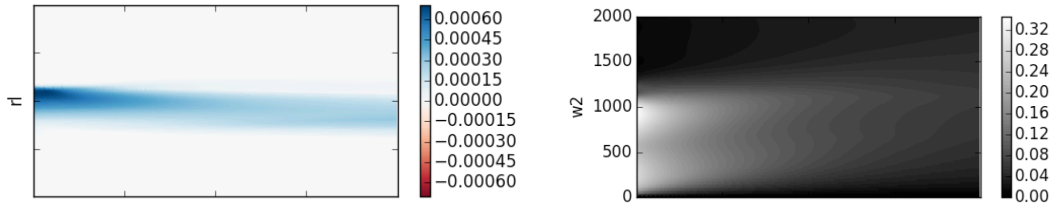

Fig. 2: CHOMP simulations of liquid water content (rl, in kg/kg) and vertical velocity variance (w2, in m^2/s^2) a stratocumulus layer from MAGIC Leg 15A over a 4000 s period.
Ideally, the simulations should stay very close to their initial profile (at left edge of the plots), but both quantities systematically drift away from this state toward a thinner cloud with much less vertical velocity variance.

With E3SM single-column model simulations of RACORO, CLUBB's turbulence lengthscale fluctuates alarmingly between time steps, a symptom that CLUBB also has problems settling into a reasonable state. CLUBB commonly hits artificial limiters (e. g. on this lengthscale) that prevent it from giving unphysical results despite what its underlying equations try to do. We hypothesize that these limiters (not included in CHOMP) may be vital to CLUBB's stability. We also suspect CHOMP might have worked better with a more extensive LES training dataset.

2. MARBLE

An artificial neural network is trained to reproduce thermodynamic tendencies and boundary layer properties from 8 years of ERA5 HIRES reanalysis data over the summertime Northeast Pacific stratocumulus to trade cumulus transition region (Fig. 3). The network (Fig. 4) is trained prognostically using 7-day forecasts rather than using diagnosed instantaneous tendencies alone. The resulting model, Machine Assisted Reanalysis Boundary Layer Emulation (MARBLE), skillfully reproduces the boundary layer structure and cloud properties of the reanalysis data in 7-day single-column prognostic simulations over withheld testing periods. Radiative heating profiles are well-simulated (Fig. 5), and the mean climatology and variability of the stratocumulus to cumulus transition are accurately reproduced (Fig. 6). MARBLE more closely tracks the reanalysis than does a comparable configuration of the underlying forecast model.

MARBLE learns not only to reproduce the effects of the physical parameterizations in the underlying weather forecast model but also to add the systematic impacts of analysis increments from data assimilation on the reanalysis. This serves as a natural form of model bias correction. In this study, MARBLE was trained and applied to a particular region and season. We have also successfully trained a version of MARBLE for summertime boundary layer parameterization over the southern U. S. Great Plains, and it is being tested as the lower tropospheric diabatic process parameterization for regional boundary-forced three-dimensional simulations with a mesoscale model.

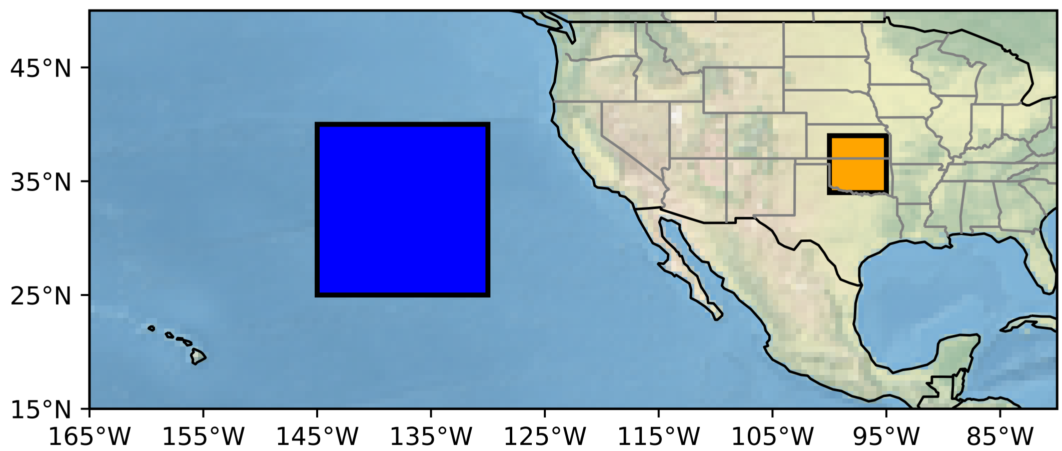

Fig. 3: NE Pacific and SGP study regions used for MARBLE training and evaluation

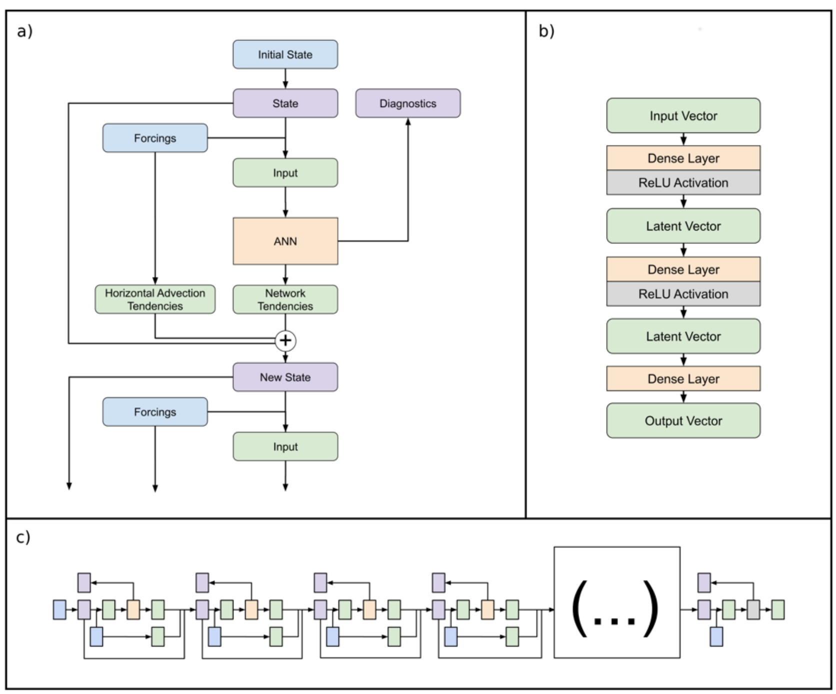

Fig. 4: Artificial neural network (ANN) structure and training approach, showing a) how components are connected in the first timestep, b) the internal structure of the ANN component, and c) how those components are chained together to train over multiple timesteps. Orange represents operations with trainable weights, grey represents static operations without trainable weights, purple represents output vectors whose values are included in the optimized loss function, green represents internal vectors used by the model, and blue represents data collected from ERA5 used as input to the model. At each timestep in c) the ANN uses the same weights. Network tendencies include both clear-sky radiative heating and residual tendencies.

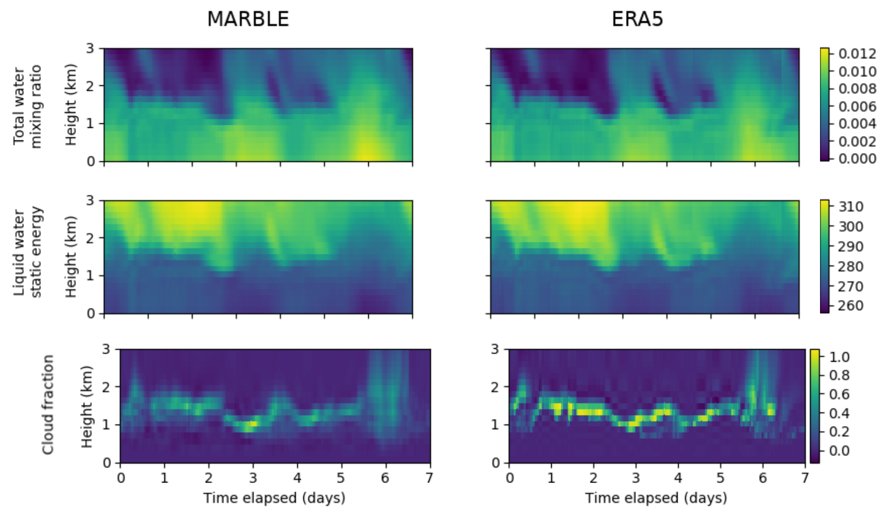

Fig. 5: Randomly selected seven-day single-column prognostic simulation showing time-height sections of humidity, liquid static energy (like potential temperature) and cloud fraction from MARBLE using withheld testing data, compared to ERA5. Initialized 0700 UTC June 11, 2016 at 30 N, 133.5 W.

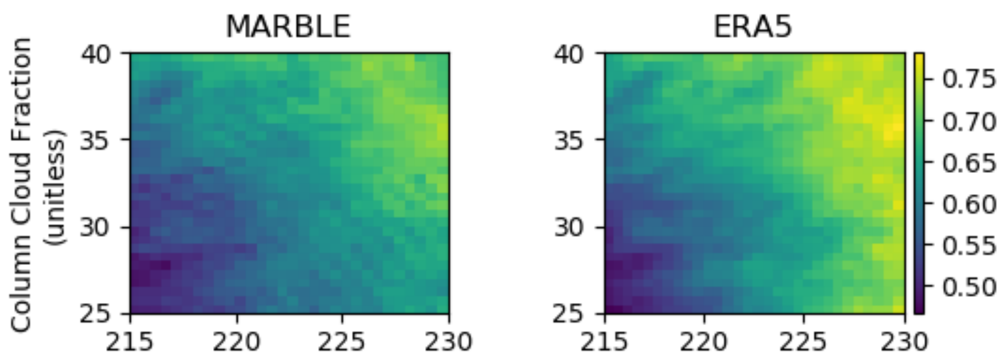

Fig. 6: Horizontally-resolved time-mean values of cloud fraction averaged across the last 3.5 days of 7-day MARBLE simulations for the Northeast Pacific in the testing period, JJA 2016.

In conclusion, we did not follow our intended research plan. We nevertheless produced a novel machine learning parameterization that is a promising prototype for future global boundary layer schemes, and we showed that LES are a useful training dataset for machine learning of turbulent statistics in atmospheric boundary layers, even though our resulting scheme proved an intractable modification to CLUBB.

Journal Articles Published/In Press:

Monteiro, J. M., McGibbon, J., and Caballero, R., 2018: sympl (v. 0.4.0) and climt (v. 0.15.3) – towards a flexible framework for building model hierarchies in Python, Geosci. Model Dev., 11, 3781-3794, https://doi.org/10.5194/gmd-11-3781-2018.

McGibbon, J., and Bretherton, C. S., 2019: Machine-Assisted Reanalysis Boundary Layer Emulation (MARBLE). Single-column emulation of reanalysis of the Northeast Pacific marine boundary layer. *Geophys. Res. Lett.*, **46**, https://doi.org/10.1029/2019MS001647.

Conference Posters and Presentations

McGibbon, J. J., and C. S. Bretherton: *Machine learning of high-order turbulence closures*. AMS 2017 Annual Meeting, Artificial Intelligence Symposium, Seattle WA, Jan. 2017.

McGibbon, J. J., and C. S. Bretherton: *Keep it Sympl: Representing Clouds and Turbulence with CHOMP*. SciPy conference, Austin, TX, Jul. 2017.

McGibbon, J. J., and C. S. Bretherton: *Defining Higher-Order Turbulent Moment Closures with a Random Forest, artificial neural network, and polynomial regression*. AGU Fall Meeting, New Orleans, Dec. 2017.

Bretherton, C.S. and J. J. McGibbon, *Machine learning for moist physics parameterizations in weather and climate models*. Invited oral presentation at the First Workshop on Leveraging Artificial Intelligence (AI) in the Exploitation of Satellite Earth Observations and Numerical Weather Prediction, College Park, MD, April 23-25, 2019.

Ph D Thesis

McGibbon, J. J., 2019: *Improving prognostic moist turbulence parameterization with machine learning and software design*. Ph. D. Thesis, Department of Atmospheric Sciences, University of Washington, Seattle, WA, July 2019.
https://www.dropbox.com/s/c7rgr1xnzkwwubp/McGibbon_PhD_Dissertation_201907.pdf?dl=0

New software:

sympl (System for Modelling Planets): A general Python-based framework for earth system modeling. Components contain all the information about the kinds of inputs they expect and outputs that they provide. Components can be used interchangeably, even when they rely on different units or array configurations. sympl provides basic functions and objects which could be used by any type of Earth system model. Available at:

https://github.com/mcgibbon/sympl

climt (Climate Modelling and diagnostics Toolkit):

An Earth system modelling toolkit that contains scientific components built over the sympl base objects. Components can be written in any language accessible from Python, and Fortran/C libraries are accessed via Cython. climt aims to provide different user APIs which trade-off simplicity of use against flexibility of model building, thus appealing to a wide audience. Model building, configuration and execution is through a Python script (or Jupyter Notebook), enabling researchers to build an end-to-end Python based pipeline along with popular Python based data analysis tools. Available at:

https://github.com/CliMT/climt

Personnel:

Jeremy J. McGibbon, graduate research assistant

Matthew E. Wyant, research scientist

Christopher S. Bretherton, Professor and PI

Reference

McGibbon, J., and Bretherton, C. S., 2017: Skill of ship-following large-eddy simulations in reproducing MAGIC observations across the Northeast Pacific stratocumulus to cumulus transition region. *J. Adv. Model. Earth Syst.*, **9**, 810–831, doi:10.1002/2017MS000924.