Using EUREC[4]A/ATOMIC Field Campaign Data to Improve Trade-Wind Regimes in the Community Atmosphere Model

Skyler Graap[1] and Colin M. Zarzycki[1]

[1]Department of Meteorology and Atmospheric Science, Pennsylvania State University, University Park, PA 16802

Correspondence: Colin M. Zarzycki (czarzycki@psu.edu)

Abstract. Improving the prediction of clouds in shallow cumulus regimes via turbulence parameterization in the planetary boundary layer (PBL) will likely increase the global skill of global climate models (GCMs) because this cloud regime is common over tropical oceans where low cloud fraction has a large impact on Earth's radiative budget. This study attempts to improve the prediction of PBL structure in tropical trade-wind regimes in the Community Atmosphere Model (CAM) by updat-
5 ing its formulation of momentum flux in CLUBB (Cloud Layers Unified by Binormals), which currently does not by default allow for upgradient momentum fluxes. Hindcast CAM output from custom CLUBB configurations which permit counter-gradient momentum fluxes are compared to in-situ observations from weather balloons collected during the ElUcidating the RolE of Cloud–Circulation Coupling in ClimAte and Atlantic Tradewind Ocean–Atmosphere Mesoscale Interaction Campaign (EUREC[4]A/ATOMIC) field campaign in the Tropical Atlantic in early 2020. Comparing a version with CAM-CLUBB with
10 a prognostic treatment of momentum fluxes results in vertical profiles that better match previously published LES results. Countergradient fluxes are frequently simulated between 950 hPa and 850 hPa over the EUREC[4]A/ATOMIC period in CAM-CLUBB. Further modification to the PBL parameterization by implementing a more generalized calculation of the turbulent length scale reduces model bias and RMSE relative to sounding data. Benefits are also seen in the diurnal cycle, although more systematic model errors persist. A cursory budget analysis suggests the buoyant production of momentum fluxes, both
15 above and below the jet maximum, significantly contributes to the frequency and depth of countergradient vertical momentum fluxes in the study region. This paper provides evidence that higher-order turbulence parameterizations may offer pathways for improving the simulation of trade-wind regimes in global models.

1 Introduction

The increase in atmospheric temperatures caused by anthropogenic greenhouse forcing will inevitably lead to changes in the
20 properties of the land surface and the structures of the atmosphere and ocean. These changes can act to either enhance or diminish the effect of the original forcing and are thus known as positive or negative feedbacks, respectively. Among the feedback mechanisms captured in global climate models (GCMs), those relating to changes in cloud profiles represent the largest source of uncertainty in the simulated climate response to increased greenhouse gas concentrations (Ceppi et al., 2017).

Low clouds reflect a significant portion of incoming shortwave radiation but emit longwave radiation at a rate comparable
25 to the surface given the similarity in temperature. This leads to what is called the 'low cloud radiative feedback' whereby

an increase in low cloud cover has a net cooling effect on the surface by preventing solar warming, while still allowing for radiational cooling. Near-surface cumulus and stratocumulus clouds are among the most important clouds for this feedback given that they have a sufficient optical depth to prevent sunlight from reaching the surface, can exist at low latitudes that experience high insolation, and can cover large surface areas.

30   Changes in low cloud fractions in the tropics have been described by Ceppi et al. (2017) as one of the three main components of the global cloud feedback in GCMs. The Hadley cell is a conceptual model of the atmospheric circulation comprised of rising motion near the equator and sinking motion in the subtropics leading to easterly winds (known as the trade winds) at the surface and westerly winds aloft in the tropics. Within this cell regions exist where different large-scale patterns of clouds, known as cloud regimes, tend to arise repeatedly. One of these is the tropical trade-wind cumulus regime, characterized by the formation

35 of many small separate cumulus clouds as a result of shallow convection in the boundary layer over tropical oceans (Ruppert, 2016). Poleward of this cloud regime, in a region known as the subtropical stratocumulus to trade cumulus transition (STCT), there is a gradual transition as the shallow cumulus clouds feed into an overlying stratocumulus layer (Stevens et al., 2002). Poleward of this, the stratocumulus layer breaks up. A large portion of stratocumulus clouds found over subtropical oceans are associated with the transitional regime and thus the STCT has a large impact on the overall climate system cloud-radiative

40 feedback (Stevens et al., 2002; Trenberth et al., 2001). Improvements in GCM prediction of boundary layer structure in the tropical trade-wind regime could improve not only the representation of cloud cover changes locally, but also the prediction of downstream cloud cover change in the STCT where the shallow cumulus clouds feed into a broader stratocumulus layer.

     The structure of the PBL is determined in large part by turbulent vertical fluxes which work to redistribute quantities like heat, moisture, and momentum. This turbulence occurs at scales much smaller than typical grid spacing of GCMs and must be

45 parameterized. The vertical flux of horizontal momentum (henceforth simply "vertical momentum flux") can be thought of as the horizontally averaged covariance between the horizontal wind and the vertical wind ($\overline{u_h'w'}$) where u_h is either the zonal (u) or meridional (v) component of the wind. In most GCMs, the time tendency of $\overline{u_h'w'}$ is parameterized with diagnostic eddy diffusivity (commonly referred to as "K Theory" (Berkowicz and Prahm, 1979; Stensrud, 2007)). This turbulence closure defines $\overline{u_h'w'}$ as the product of the existing vertical gradient in horizontal momentum and a coefficient denoted as K. Such a

50 closure can only act to move existing horizontal momentum to an altitude with less momentum (downgradient flux). Recently, it has been shown in large eddy simulations (LES) that momentum fluxes moving in the opposite direction – upgradient fluxes working to move momentum to altitudes with greater horizontal momentum, also referred to as countergradient fluxes – can occur in tropical shallow convection (Larson et al., 2019; Dixit et al., 2020; Helfer et al., 2021). In order for GCMs to capture these upgradient fluxes, they must prognose $\overline{u_h'w'}$. Such a parameterization includes many different source and sink terms in

55 its calculation of $\overline{u_h'w'}$ time tendency, with each term being related to a physical process.

     Larson et al. (2019) (henceforth L19) attempted to model $\overline{u_h'w'}$ in marine shallow cumulus layers in a single-column model using data from several field campaigns (including the Barbados Oceanographic and Meteorological Experiment (BOMEX), which took place over the tropical North Atlantic (Holland and Rasmusson, 1973)). Their model utilizes the higher-order Cloud Layer Unified by Binormals (CLUBB) parameterization and is run in both a mode that only allows downgradient diffusion and

60 a mode that prognoses $\overline{u_h'w'}$. They found that the prognostic momentum configuration was better able to recreate the structure

of wind profiles described by an LES run based on the field campaign. LES simulations are integrated at a much higher spatial resolution than operational models and can serve as a spatiotemporally-continuous 'bridge' to point observations that are limited in space and time. The vertical profile of momentum during BOMEX featured a characteristic easterly jet near the top of the boundary layer and the prognostic momentum run was able to recreate the 3-layer structure of $\overline{u_h'w'}$ described by the LES where there is downgradient $\overline{u_h'w'}$ from the surface to near the jet maximum, upgradient flux in the few hundred meters above this jet maximum, and weak $\overline{u_h'w'}$ above this layer.

Similarly, Dixit et al. (2020) (henceforth D20) found upgradient $\overline{u_h'w'}$ in the cloud layer of a tropical shallow convection regime in their investigation of vertical momentum transport using multi-day large eddy models with data from the BOMEX and RICO (Rain in Shallow Cumulus Over the Ocean (Rauber et al., 2007)) field campaigns, both of which took place in the western tropical North Atlantic. Their analysis reveals that these upgradient fluxes are driven by non-hydrostatic pressure gradients and horizontal circulations generated by convection. The effects of these mesoscale dynamics can therefore not be represented by downgradient diffusion alone.

Helfer et al. (2021) also noted upgradient momentum fluxes in their LES simulations run for the tropical North Atlantic in a time period corresponding to the NARVAL (Next-generation Aircraft Remote-sensing for VALidation studies) flight campaign in December 2013 (Vial et al., 2019). They demonstrated that these upgradient fluxes could not be captured by pure K theory based on their calculated profiles of what the coefficient K would have to be as derived by dividing $\overline{u_h'w'}$ by the existing vertical gradient in horizontal momentum ($\frac{dU}{dz}$), sometimes referred to as 'effective diffusivity' (Bryan et al., 2017; Nardi et al., 2022). These profiles showed that negative K would be required (i.e. upgradient fluxes are occurring) for both u and v, in certain layers of a vertical structure similar to that found in L19, particularly in the winter. These profiles were calculated for the innermost grid of their LES hindcasts which consisted of multiple nested domains and were ultimately forced by reanalysis data.

This study seeks to build on the findings of L19 by using data from a more recent and intensive study that took place in generally the same region as BOMEX and RICO (the joint ElUcidating the RolE of Cloud–Circulation Coupling in ClimAte and Atlantic Tradewind Ocean–Atmosphere Mesoscale Interaction Campaign (EUREC[4]A/ATOMIC) field campaign) to evaluate how prognosing, rather than diagnosing, $\overline{u_h'w'}$ affects a three-dimensional GCM's performance in predicting boundary layer structure in tropical trade-wind regimes. Here we focus on the Community Atmosphere Model (CAM), a component of the Community Earth System Model (CESM). Several experimental versions of CAM are created, each of which implements CLUBB and includes a prognostic eddy diffusivity that uses a Reynolds averaging closure. The difference between separate prognostic momentum runs lies in how the vertical turbulent length scale is estimated. Output from these versions of CAM, as well as from the default unmodified version, are compared to state variable data from 1,546 weather balloon soundings collected during the six-week EUREC[4]A/ATOMIC field campaign.

2 Data and Methods

All of the observational data used in this study to evaluate model predictions come from the EUREC[4]A/ATOMIC mass data collection field campaign. EUREC[4]A/ATOMIC was conducted over the tropical North Atlantic Ocean just east of Barbados in January and February 2020 (Stevens et al., 2021). Boundary layer measurements collected for this field campaign are of higher resolution and quality than previous field campaigns in the same region like BOMEX and RICO (Savazzi et al., 2022). While recent, these data are beginning to be exploited to evaluate model performance in this region. For example, Savazzi et al. (2022) used the weather balloon sounding, dropsonde, and lidar data from EUREC[4]A/ATOMIC to characterize the wind profile structure of the boundary layer and to evaluate the performance of the Integrate Forecast System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) along with the related ERA5 reanalysis data in the prediction of boundary layer wind profiles during EUREC[4]A/ATOMIC. Some of the techniques employed by Savazzi et al. (2022) to evaluate the performance of IFS using this data set are used here to evaluate the performance of CAM.

2.1 EUREC[4]A/ATOMIC Sounding Data

During the EUREC[4]A/ATOMIC campaign, radiosondes attached to weather balloons were launched from four ships and the Barbados Cloud Observatory (BCO) over the course of 43 consecutive days from 8 January to 19 February 2020. For most of this period, soundings were attempted every four hours from all five stations, but not all stations reported every day (see Figure 1 in Stephan et al. (2020) for a complete time series of all balloon launches). Most balloon launches recorded data during both the ascent of the balloon and the descent of the radiosonde with a parachute after the balloon burst, however only data from the ascents are used here because the descent data are likely less reliable given the rapid fall speed. The four ships were moving during the field campaign, but at all times, all ships were located somewhere between 6 and 16 ° N and between 50 and 60 ° W (see Fig. 2 in Stephan et al. (2020) for a complete time series of ship locations). All stations launched Vaisala RS41-SGP radiosondes and recorded horizontal wind (u and v components), temperature (T), relative humidity, and pressure at even intervals of 10 meters altitude starting at 30 or 40 meters above the surface until balloon burst, up to a maximum altitude of 31 km. Additionally, 47 radiosondes of Meteomodem type M10 were launched from one of the ships (the L'Atalante) without parachutes (Stephan et al., 2020). These soundings also reported data every 10 meters.

2.2 CAM Configurations

The version of CAM studied here is CAM version 6 (Bogenschutz et al., 2018; Gettelman et al., 2019) using the spectral element (SE) dynamical core (Lauritzen et al., 2018). We apply an unstructured cubed-sphere grid with nominal 1° (111km, also referred to as CAM-SE's ne30np4 grid) horizontal grid spacing and 58 vertical levels with finer grid spacing in the atmospheric boundary layer. The height of the lowest model level is approximately 22m and the model top is approximately 40 km. The most significant parameterization change in CAM6 from predecessor versions is the addition of CLUBB as a unified turbulence scheme to replace otherwise separate boundary layer, shallow-convection, and macrophysics parameterizations (Bogenschutz et al., 2013; Danabasoglu et al., 2020). CLUBB is a high-order closure that represents moist turbulence with

a simple multivariate probability density function to describe sub-grid variations in potential temperature (θ), water vapor
125   mixing ratio (Q), and vertical velocity (w) (Golaz et al., 2002; Larson, 2022). CLUBB is discretized in the vertical by centered
differencing or else upwind differencing on a staggered grid and implements a semi-implicit time stepper where the time
stepping method is simple backward Euler (Larson, 2022). State variables solved for in the dynamical core of CAM include
air temperature (T), Q, u, v, and surface pressure (p_s). Since CAM is a hydrostatic model, the vertical pressure velocity (ω)
is diagnosed from the continuity equation. Other quantities, such as turbulence outputs, are solved for in the model's subgrid
130   parameterization suite.

     In this study, CAM is initialized twice daily (00Z and 12Z) with the 0.25° ERA5 (Hersbach et al., 2020) reanalysis data using
the Betacast software package, first described in Zarzycki and Jablonowski (2015). To initialize the model, the ERA5 state field
is mapped to the CAM grid using high-order remap operators, with the hydrostatic correction of Trenberth et al. (1993) applied
to balance the model state against CAM's lower-resolution orography. The model was run with prescribed ocean and ice fields
135   using observations from NOAA's Optimum Interpolation (OI) dataset (Reynolds et al., 2002) and are fixed for the duration
of the hindcasts. The model's land state was generated by using three-hourly surface forcing derived from ERA5 to drive an
offline version of the Community Land Model (CLM) for the 12 months before the EUREC⁴A/ATOMIC period. Subsequent
land initializations leverage the 12-hour land surface forecast from the previous cycle as in Zarzycki and Jablonowski (2015).
This creates a surface state consistent with atmospheric observations during the period prior to the simulation, although it is
140   worth noting that we anticipate impacts from the land surface model are negligible given the domain of interest and duration of
the hindcasts. The model is then integrated for 72 hours in different configurations providing output every 30 minutes for each
day of the EUREC⁴A/ATOMIC Core Period (8 January - 19 February 2020). In order that CAM output from runs initialized
0, 1, and 2 days prior are available for all days during the field campaign in addition to approximately a week following it,
CAM is initialized for the three days leading up to the campaign and then every day during it (from 00Z 5 January 2020 to
145   12Z 25 February 2020), resulting in 104 initializations for each configuration discussed below. All simulations were completed
using the Cheyenne supercomputer, maintained at Computational Information Systems Lab and funded by National Science
Foundation (CISL, 2019).

2.2.1   Diagnostic Versus Prognostic Configurations

The unaltered version of CAM described above (henceforth known as x001 or "the default run") is the run against which the
150   other configurations of CAM are compared. This model version corresponds to the configuration of CAM that was used in the
CESM version 2 release (Danabasoglu et al., 2020) that was used to generate the simulation submitted to the Coupled Model
Intercomparison Project version 6 (CMIP6). In this configuration, $\overline{u'_h w'}$ are calculated using a diagnostic eddy diffusivity
approximation:

$$\overline{u'w'} = -K_m \frac{\partial \overline{u}}{\partial z} \tag{1}$$

155 $$\overline{v'w'} = -K_m \frac{\partial \overline{v}}{\partial z} \tag{2}$$

where K_m is a tunable transfer coefficient (Golaz et al., 2002). Here, $\overline{u'_h w'}$ is simply a function of the vertical shear of the resolved horizontal wind. The turbulent transfer coefficient is defined to be positive, and thus, such a diagnosis is incapable of producing $\overline{u'_h w'}$ that acts to move momentum 'up' the existing gradient.

An experimental CAM configuration (x101) is created by replacing the eddy diffusivity closure by using a higher order

160 closure described by Eq. 3 to prognose $\overline{u'_h w'}$. This closure, which calculates the time tendency of $\overline{u'_h w'}$ by considering several source and sink terms, can be considered an incomplete third order closure since $\overline{w'^3}$ is prognosed by CLUBB (Larson, 2022; Larson et al., 2019; Nardi et al., 2022). We stress that, aside from this change, all other components of x001 and x101 are identical. Unless otherwise specified, all model settings and configurations are the default used in CAM6 for the CESM2 release.

$$\frac{\partial \overline{u'_h w'}}{\partial t} = \underbrace{-\overline{w}\frac{\partial \overline{u'_h w'}}{\partial z}}_{1} - \underbrace{\frac{1}{\rho}\frac{\partial \rho\overline{w'^2 u'_h}}{\partial z}}_{2} - \underbrace{(1-C_{uu,shear})\overline{w'^2}\frac{\partial \overline{u_h}}{\partial z}}_{3} - \underbrace{(1-C_7)\overline{u'_h w'}\frac{\partial \overline{w}}{\partial z}}_{4} + \underbrace{(1-C_7)\frac{g}{\theta_{vs}}\overline{u'_h \theta'_v}}_{5} - \underbrace{\frac{C_6}{\tau}\overline{u'_h w'}}_{6} - \underbrace{\epsilon_{u_h w}}_{7}$$

165 $$\tag{3}$$

Here, ρ is the air density, g is gravity, θ_v is virtual potential temperature, τ is the eddy turnover time scale, and $C_{uu,shear}$ is an empirical constant with a default value of 0.3. C_6 and C_7 are also tunable constants, although they are left as 4 and 0.5, respectively, for all simulations here. The terms here describe how $\overline{u'_h w'}$ can either be generated or dissipated through 1) advection by the mean vertical wind, 2) turbulent advection by perturbations in the vertical wind, 3) turbulent production by

170 updrafts and downdrafts, 4) turbulent production from pre-existing $\overline{u'_h w'}$ existing in a vertical gradient in the mean vertical wind, 5) buoyant production, 6) a 'return-to-isotropy' adjustment that has the magnitude of $\overline{u'_h w'}$ decay over time, and 7) a residual dissipation term (Nardi et al., 2022). The derivation of this equation is described in Appendix A alongside additional turbulence closures for the remaining unsolved terms.

In configuration x101, the eddy turnover time scale, τ, which describes the rate of decay in the 'return-to-isotropy' term, is

175 calculated as the vertical turbulent length scale (L) divided by the square root of turbulent kinetic energy (TKE or $\overline{e}$) as defined in Eq. 25 of Golaz et al. (2002):

$$\tau = \frac{L}{\sqrt{\overline{e}}} \tag{4}$$

and TKE is calculated from variances of each wind component (each predicted by CLUBB):

$$\overline{e} = \frac{1}{2}(\overline{u'^2} + \overline{v'^2} + \overline{w'^2}) \tag{5}$$

180   This turbulent length scale is described by the mean of the upward and downward distances a parcel could travel before its change in potential energy from buoyancy equals the total turbulent kinetic energy that it started with (Golaz et al. (2002) Eqs. 36, 37, 38). This formulation of τ depends only on turbulent kinetic energy (TKE) and atmospheric stability. In x101, L is calculated as described above and τ is diagnosed from that value of L and TKE. The same is the case in x001.

2.3   Prognostic Configurations with Experimental Vertical Turbulent Length Scale Estimates

185   To explore the impact of the shape of the turbulence profile (i.e., the shape of either L or τ profiles), we explore an alternative treatment of τ described in Guo et al. (2021). Here, τ can be calculated using a set of 'building blocks' describing the dissipation of turbulent eddies:

$$\frac{1}{\tau} = \underbrace{C_{\tau,bkgnd} \frac{1}{\alpha}}_{1} + \underbrace{C_{\tau,sfc} \frac{u^*}{\kappa} \frac{1}{(z - z_{sfc} + d)}}_{2} + \underbrace{C_{\tau,shear} \sqrt{\left(\frac{\partial \overline{u}}{\partial z}\right)^2 + \left(\frac{\partial \overline{v}}{\partial z}\right)^2}}_{3} + \underbrace{C_{\tau,N2} \sqrt{N^2}}_{4} \qquad (6)$$

In this equation (the sum of Eqs. 19 and 20 in Guo et al. (2021)), α is a constant (1000 s), u^* is the friction velocity,
190   κ is the Von Karman constant, N is the Brunt-Väisälä frequency, d is a small displacement height, and $C_{\tau,bkgnd}$, $C_{\tau,sfc}$, $C_{\tau,shear}$, and $C_{\tau,N2}$ are all empirical constants. This equation considers 1) a background dissipation rate, 2) dissipation due to frictional effects near the surface, 3) dissipation due to vertical wind shear, and 4) dissipation in a stable atmosphere (set to 0 in buoyantly unstable and neutral layers). Each term here includes a different tunable coefficient (i.e., the C_τ terms). We perform four additional CAM configurations (x201, x202, x203, and x204) which prognose $\overline{u_h' w'}$ (like x101) but use Eq. 6 to calculate
195   τ. We note that τ does appear in other prognostic CLUBB equations (e.g., turbulent fluxes of scalars) and therefore impacts additional prognostic quantities in the PBL beyond just $\overline{u_h' w'}$ (Larson, 2022).

Each of these runs differs only in their values of those four tunable coefficients, which are determined here by a simple Nelder-Mead optimization (Nelder and Mead, 1965). Specifically, a set of very short (48-hour) hindcasts is run, optimizing various tunable parameters in CLUBB to minimize the difference in the predicted wind field after 2 days when compared
200   against ERA5 reanalysis at the same time. Optimization is completed relative to global ERA5 reanalysis data rather than the local EUREC[4]A/ATOMIC data to ensure a reasonable global simulation. Four different sets of parameters are generated by applying four different starting simplexes with the Nelder-Mead algorithm. The algorithm was not run to a formal convergence, with each optimization lasting approximately 50 iterations. We stress that this should be thought of as providing different sets of parameters that generate an atmospheric state plausibly consistent with ERA5 reanalysis. In other words, each of these four
205   configurations should be considered different components of a small sensitivity analysis as opposed to rigorous attempts to calibrate model performance, particularly at global scales. The exact values of the parameters for Eqs. 3 and 6 used in these different runs are described by the first five terms in Table 1. The other four parameters listed in this table are subtle tuners on the equations for $\overline{w'^2}$ and the $C_{\tau,N2}$ term.

The relationship between L and τ described in Eq. 4 is applied, although L is now diagnosed from turbulent kinetic energy
210   and τ as:

Model Ver.	$C_{\tau,bkgrnd}$	$C_{\tau,sfc}$	$C_{\tau,shear}$	$C_{\tau,N2}$	$C_{uu,shear}$	$C_{uu,buoy}$	$C_{\tau,N2,clr}$	$C_{\tau,N2,wp2}$	$C_{\tau,N2,xp2}$
x201	0.42	0.05	0.37	0.06	0.09	0.12	1.00	0.10	0.05
x202	0.41	0.04	0.34	0.07	0.01	0.36	0.84	0.19	0.13
x203	0.64	0.04	0.46	0.10	0.07	0.12	1.30	0.02	0.15
x204	0.45	0.04	0.20	0.10	0.05	0.30	0.90	0.20	0.15

Table 1. Specific values of the parameters that differ between prognostic configurations with experimental vertical turbulent length scale estimates. $C_{\tau,bkgrnd}$ modifies the background τ, $C_{\tau,sfc}$, $C_{\tau,shear}$, $C_{\tau,N2}$ modify the strength of influence of the surface, vertical wind shear, and stability on τ respectively, and $C_{uu,shear}$ modifies the turbulent production of $\overline{u'_h w'}$ by updrafts and downdrafts. The next four parameters are not included in the equations mentioned thus far, but $C_{uu,buoy}$ serves as a tuning in the CLUBB equation for $\overline{w'^2}$ and $C_{\tau,N2,clr}$, $C_{\tau,N2,wp2}$, and $C_{\tau,N2,xp2}$ all serve as subtle tunings on $C_{\tau,N2}$ as described in Guo et al. (2021).

$$L = \tau * \sqrt{e} \qquad (7)$$

That is, τ is computed first and L is diagnosed using this in combination with TKE (Larson, 2022). Henceforth, these configurations (x201, x202, x203, and x204) which use Eq. 6 to calculate τ (and thus L) will be referred to as the 'experimental length scale runs'.

2.4 Comparison to Observational Soundings

2.4.1 Interpolation of CAM Output

In order to directly compare model output to observational data, model estimates of state variables are calculated for every point reported for every sounding. This is done for every model configuration where 1-day lead time predictions are used (24-48 hours after model initialization) to reduce forecast error and better constrain the simulations based on the initial conditions. Similar results are found when 2-day leads are considered instead (not shown). The profiles are found by taking data from only the model column nearest a sounding and linearly interpolating the vertical profiles of T, Q, u, and v. The 'nearest' model column is calculated as that with the smallest great circle distance from the latitude and longitude reported by a balloon at 1 km geopotential height, or if no data were reported for this level, the next lowest altitude for which coordinates are reported. 1 km geopotential height is chosen as the reference point for each sounding because this study mainly focuses on the lowest 2.5 km of the atmosphere. Soundings that do not report any data for altitudes above 1 km are not considered in this study.

Each sounding profile is compared to a purely vertical profile in the model output, but this is reasonable since ascent rates were rapid enough and horizontal wind speeds were slow enough that balloons tended to drift only around 10 km horizontally in the lowest 5 km altitude (the layer of focus), while the nearest model columns are separated by approximately 100 km. Similarly, each observational profile is compared only to model output from the single timestep that is nearest in time to when the sounding reached 1 km geopotential height. This is reasonable since typical balloon ascent rates were 3 to 5 m/s (or

about 1 km in 3 to 5 minutes) and model timesteps are 30 minutes apart. Once a model timestep and column are chosen for a particular sounding, the interpolated vertical profile used in the direct comparison is generated. Since CAM6 uses a hybrid sigma-pressure vertical coordinate, the heights at which CAM data are output can vary between columns and time steps. These reporting altitudes are found for each column and timestep that were chosen to correspond to an observational sounding in each

235 model run. The vertical grid spacing of CAM is around 50 meters near the surface, 250 meters at 2 km altitude, and 500 meters at 5 km altitude. This is much coarser than observations, which report every 10 meters. At each of these 10-meter levels, state variable values meant to represent model output are calculated by taking the linear vertical distance-weighted average of those values reported at the two nearest model levels. Those observational points that lie below the lowest model level simply take the value of that lowest level. There is no analog to this at high altitudes since model output is reported for higher altitudes

240 than all soundings. For each interpolated model prediction that corresponds to a point in the observations, a bias is calculated for each state variable predicted. This is done by simply subtracting the value measured by the observation from that value predicted by the model.

2.4.2 Statistical Profile Calculations

Mean, median, and 25th/75th percentile profiles for state variables in observations are all estimated for the whole domain

245 space and time by calculating those metrics at each 10-meter altitude level over all soundings during the campaign. These statistical profiles are also created for the output of each model configuration and lead time by performing the calculations on the corresponding state variable model output that has been interpolated to the observations' 10-meter grid spacing.

These profiles are calculated both for all times of day (by including all soundings) and for particular times of day, by only including soundings whose launch times fit within particular hours of the day. Specifically, eight sets of time-of-day-specific

250 profiles are created, each of which only takes into account those data that were collected by balloons launched during particular non-overlapping 3-hour increments, beginning with 00:00-03:00 UTC (02:00-05:00 local time).

Mean profiles are also created that estimate the vertical profiles in model bias and root mean squared error (RMSE). Bias profiles are simply created by averaging the aforementioned biases calculated at each point, while RMSE profiles are created by, for each altitude, taking the square root of the sum of the squares of each bias from every sounding at that altitude.

255 ## 3 Results of the Addition of Prognostic Momentum

3.1 Momentum Profiles

We first investigate the impact on simulated profiles by replacing parameterization of $\overline{u_h' w'}$ by eddy diffusivity with the prognostic equation (Eq. 3). It can be seen in Fig. 1a that the default version of CAM (x001, red dotted line) tends to overestimate the magnitude of the easterly winds at most altitudes below 2.5 km and places the easterly jet maximum at a higher altitude

260 when compared to EUREC[4]A/ATOMIC observations (solid black line). Given the limited number of soundings that report

below 40 m, mean profiles observational profiles below 40 m are not representative of the domain and have been removed from all plots in this study. The same goes for the corresponding plots of errors calculated from those observations.

When adding the prognostic $\overline{u_h'w'}$ formulation in configuration x101 (dashed green line), the jet maximum becomes stronger in magnitude by around 0.5 m/s but narrower in depth meaning that the vertical gradient of u becomes steeper in the region of

265 the jet. Above this layer, the strong easterly wind bias in x001 is reduced in x101. Although the easterly bias is increased by up to 0.25 m/s in x101 at altitudes below the jet maximum, the maximum bias in u is actually around 0.2 m/s smaller in x101 than x001, and RMSEs are reduced by up to 0.3 m/s at altitudes between 1 and 2 km in x101 (see Figure 2). Biases and RMSEs can become large below 200 m because very few model levels are present in this layer where real-world conditions can vary significantly with height. Model predictions at these altitudes are highly sensitive to the surface layer formulation, which is not

270 the focus of this study. We also note that winds are generally too strong throughout the lowest 2.5km. Ignoring the Coriolis force, a turbulence parameterization only rearranges the wind profile in the vertical. This may also imply that the surface is not inducing enough drag on the lowest model level, although we leave this evaluation for future work.

Figure 1b shows the profiles of $\overline{u'w'}$. No observational profiles exist for turbulence covariances since only state variables are measured by the radiosondes in EUREC[4]A/ATOMIC. While some aircraft observations of such fluxes were collected as part

275 of the field campaign (Brilouet et al., 2021), these flights covered a small time window of the campaign and observations were generally taken along horizontal surfaces. Below the jet maximum, both x001 and x101 show similar $\overline{u_h'w'}$. However, $\overline{u_h'w'}$ differ greatly above the altitude of the jet maximum (above approximately 800m). Both profiles feature negative $\overline{u'w'}$ at these altitudes, but the magnitude 'overshoot' (i.e., the magnitude of negative $\overline{u'w'}$ values before returning towards 0 with height) is much greater for x001.

280 These $\overline{u'w'}$ profiles are qualitatively very similar to analogous results described in Fig. 8 of L19. They compared results from a prognostic $\overline{u_h'w'}$ idealized single-column model and an LES running the BOMEX test case. The implementation of prognostic $\overline{u_h'w'}$ made the easterly jet more narrow and reduced the magnitude of negative $\overline{u'w'}$ above the jet maximum, which resulted in better agreement with their LES runs. This better matches the LES simulation L19 applies as a physically-based reference. This along with observations in our study being qualitatively similar to the LES-derived profiles in L19,

285 suggests the addition of prognostic $\overline{u_h'w'}$ improves the realism of how the jet is simulated in x101. The behaviors seen in highly constrained single-column simulations and idealized LES runs can be reproduced in short-term initialized real-world hindcasts when compared against field observations, demonstrating potential utility in applying such a hierarchical analysis for model development applications.

Magnitudes of the northerly winds are enhanced by up to 0.5 m/s below the height of the jet and reduced by up to 0.6 m/s

290 above it in x101 compared to x001, leading to a larger vertical wind shear (see Figure 1c). In x101, $\overline{v'w'}$ is also about half as negative at altitudes between 300 m and 2 km. Differences in wind component structure between x001 and x101 are related to differences in $\overline{u_h'w'}$ profile structure. Although the overall biases in v are similar between x001 and x101, the differences in profile structure are very similar to those differences in the v component profiles described by Figure 8 in L19. L19 found that their model predictions of both v and $\overline{v'w'}$ profiles better matched LES when prognostic $\overline{u_h'w'}$ were included in their model.
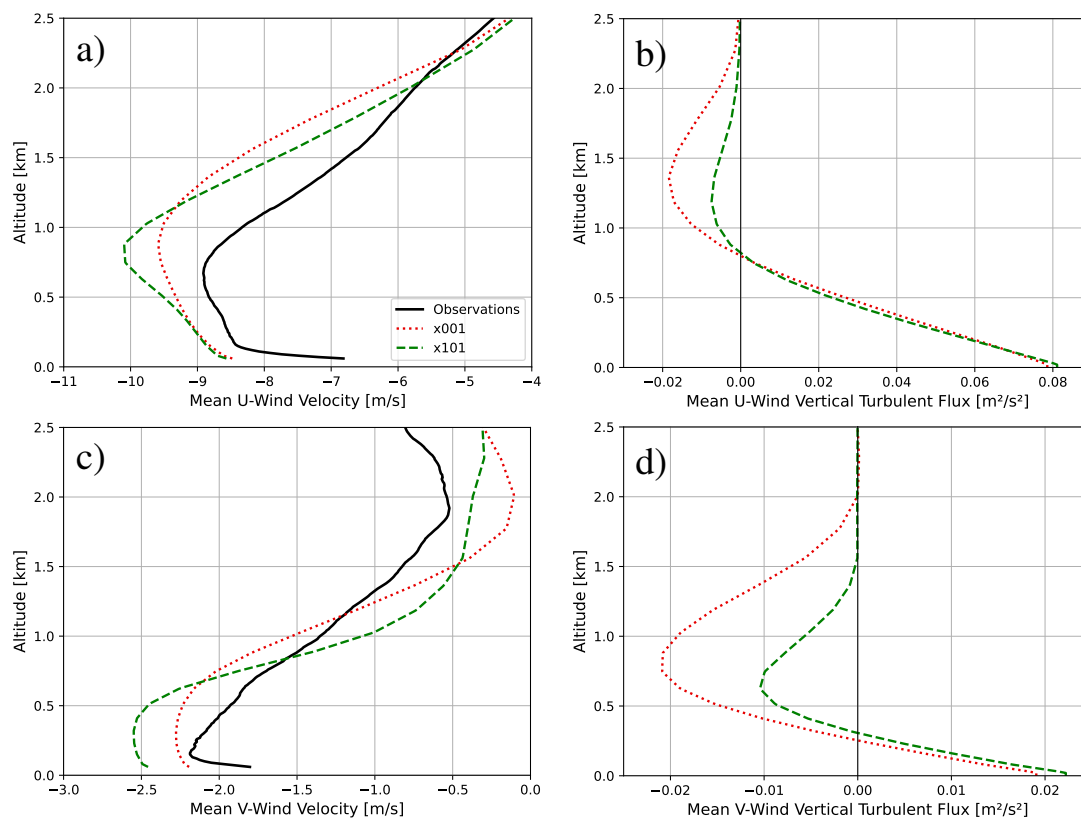
Figure 1. Domain mean vertical profiles from observations, CAM x001, and CAM x101 for horizontal wind components (u and v, panels a. and c.) and vertical turbulent fluxes of horizontal momentum ($\overline{v'_h w'}$, panels b. and d.).

The x101 simulation is also in better agreement, qualitatively, than x001, with the winds and $\overline{u'_h w'}$ profiles for BOMEX found in Fig. 2 of D20. The smaller negative $\overline{u'w'}$ in the layer above the jet is closer to the nearly zero $\overline{u'w'}$ in this layer in D20. Similarly, the less negative $\overline{v'w'}$ in the layer around the jet maximum in x101 is more similar to the relatively weak $\overline{v'w'}$ in that layer in D20, although v winds appear overall much weaker in BOMEX (around 1 m/s maximum) than in our study (around 2 m/s maximum). These qualitative similarities to D20 in profiles predicted by x101 make sense given that D20 also noted the existence of countergradient fluxes in their simulations. These fluxes can be captured by the x101 simulation but not by x001 because of the addition of prognostic $\overline{u'_h w'}$ calculations.

Changes in the u and v wind profiles may be linked via the characteristic turning of winds with height in the atmospheric boundary layer known as the Ekman spiral (Ekman, 1905; Clarke, 1970). This large-scale turning ultimately results from the surface imparting a frictional force on the air immediately above it. This frictional force acts in the direction opposing the motion of the air and, in steady-state, must be balanced by the Coriolis force. This causes the wind to be rotated near the surface relative to its direction aloft. In the Northern Hemisphere, this wind is observed to veer (i.e., rotate clockwise) with height until it reaches geostrophy. The depth over which this rotation occurs, known as the Ekman Layer, depends on vertical turbulent

momentum transport which depends on the air's eddy viscosity. A higher viscosity makes it so the momentum imparted by the surface is dissipated over a smaller depth and hence the Ekman Layer is more shallow and features a steeper change in the wind profile.

Figure 2 displays mean profiles of both u and v, and horizontal wind speed ($|\overline{\boldsymbol{U_h}}|$) for both CAM configurations, along with corresponding vertical profiles of the biases and root mean squared errors associated with these variables. For both the bias (middle row) and RMSE (lower row), values closer to zero are desirable and reflect better agreement with the sounding data. It can be seen that although x101 has a stronger jet maximum than x001, it has a reduced maximum easterly bias when compared to x001 since its jet placement matches observations better. It can also be seen that the reduced easterly bias in x101 corresponds with a reduced overall $|\overline{\boldsymbol{U_h}}|$ bias. There is a noticeable decrease in the RMSEs of u and $|\overline{\boldsymbol{U_h}}|$ of about 0.3 m/s moving from x001 to x101 near the jet maximum. The remainder of the RMSE profiles are quite similar between x001 and x101, which implies that other model biases are important drivers in solution error rather than $\overline{u_h'w'}$.

Upgradient fluxes are not apparent in any mean momentum profile in Fig. 1 as vertical wind shear ($\frac{\partial u_h}{\partial z}$) sign changes occur at nearly the same altitudes where $\overline{u_h'w'}$ sign changes occur in both x001 and x101 (although not exactly because of linear interpolations working on model levels of inconsistent heights). Upgradient fluxes are, however, present in individual profiles. One way to describe where upgradient fluxes are occurring is by calculating an "effective eddy diffusivity" (K_{eff}) and finding where it is negative. This quantity backs out what the transfer coefficient K_m described in Eqs. 1 and 2 would have to be in order to predict the given $\overline{u_h'w'}$ profile from the vertical wind shear. Eq. 8 describes this calculation essentially as a rearrangement of Eqs. 1 and 2. The coefficient K_m is always positive in a model that diagnoses momentum flux (and thus $\overline{u_h'w'}$ always works downgradient). Here, a negative value of K_{eff} indicates that upgradient fluxes are occurring.

$$K_{eff} = -\frac{\overline{u_h'w'}}{\frac{\partial \overline{u_h}}{\partial z}} \tag{8}$$

Figure 3 describes all model levels below 600 hPa on each of the 1,546 recreated soundings (before linear interpolation is applied) where negative values of K_{eff} are found for u for both x001 and x101 as black points. Some points in x001 are found to have negative K_{eff}, but these arise because CAM outputs u_h and $\overline{u'w'}$ at different points within its timestep. This can lead to u in low shear environments being updated by other subroutines such that small changes induce a sign flip in $\frac{\partial \overline{u}}{\partial z}$ which results in K_{eff} being erroneously calculated as negative. In order to exclude such occurrences, points where K_{eff} is found to be negative, but the absolute value of $\frac{\partial \overline{u}}{\partial z}$ is smaller than 0.07 m/s per km (i.e., essentially unsheared layers), are shown in orange. This threshold was chosen to be larger than the largest value of $\frac{\partial \overline{u}}{\partial z}$ found for any point with negative K_{eff} in x001 since this model configuration is incapable of generating true upgradient fluxes within the CLUBB subroutine. Most points with negative K_{eff} in x101 are above this threshold and remain black in the corresponding panel. It is evident that x101 does indeed produce countergradient fluxes that are not apparent in the x001 simulations.

Most upgradient $\overline{u'w'}$ predicted by x101 fall in a layer between 950 hPa and 850 hPa, which roughly corresponds to 600 to 1400 m above the ocean surface. This is a similar range of altitudes as where L19 found upgradient fluxes when running the BOMEX test case with both LES and single-column models (their Fig. 1), which was between approximately 770 and 1070 m
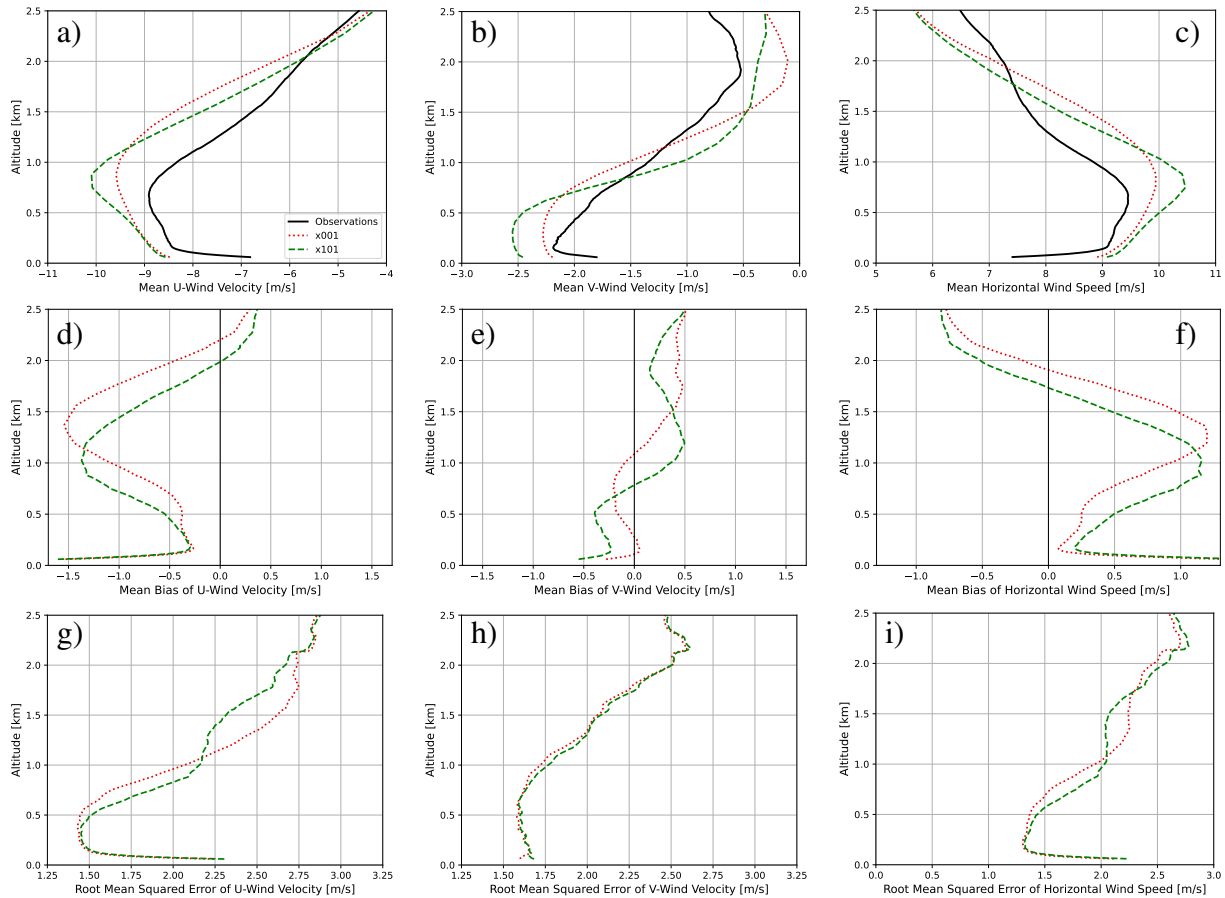
Figure 2. Vertical profiles of means, mean errors (biases), and root mean squared errors of various CAM configurations for horizontal wind components (u and v) and overall horizontal wind magnitude ($|\boldsymbol{U_h}|$) (observations are also included in the mean profiles).

altitude. These altitudes are also similar to where Helfer et al. (2021) calculated negative K_{eff} between approximately 600 and 1700 m altitude for their large eddy model hindcast using data from the NARVAL campaign (their Fig. 10). This demonstrates that a high-order turbulence scheme can reproduce these countergradient fluxes in global ESM simulations and that they occur when the atmospheric state is initialized with real-world conditions. From these results and the $\overline{u_h'w'}$ profile structures of

345 past LES, we speculate that the zonal jet is more physically realistically represented when prognostic $\overline{u_h'w'}$ is applied in lieu of traditional eddy diffusivity by comparing short-term initialized hindcasts using a climate model compared to intensive field campaign data. Confidence is added to this hypothesis by qualitatively similar findings in recent work investigating LES simulations with atmospheric forcing consistent with EUREC[4]A/ATOMIC field campaign conditions. This underscores the utility of applying initialized hindcasts to help bridge the gap that has traditionally existed between process-oriented analyses

350 (e.g., single-column models, LES, observations) and long-term (e.g., multi-decadal) climate simulations.
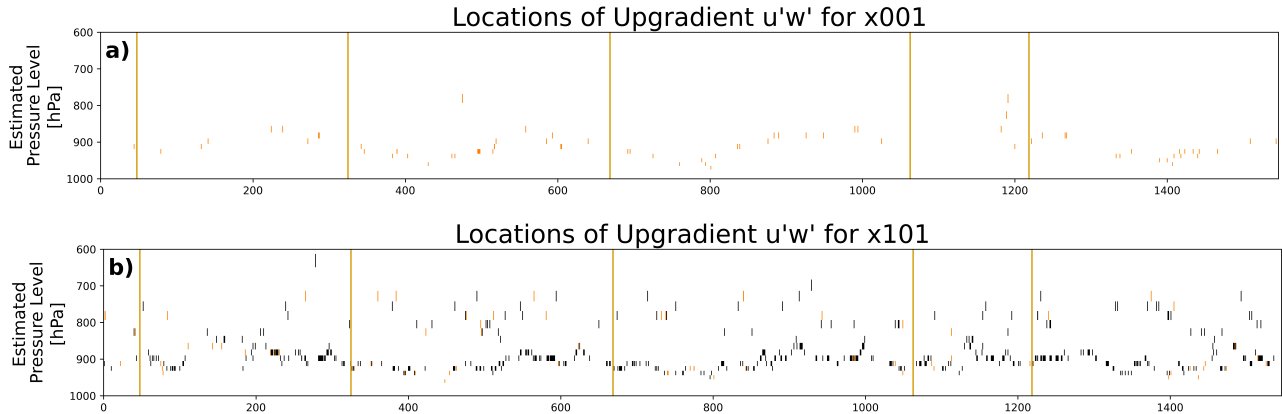
Figure 3. Diagrams displaying where effective eddy diffusivity (K_{eff}) is negative (and thus where upgradient fluxes are occurring) in x001 and x101 output for the vertical flux of zonal momentum ($\overline{u'w'}$). Black and light red dots indicate where upgradient fluxes are calculated to be occurring, but light red dots indicate where negative K_{eff} was also calculated with a very small value for the vertical gradient of zonal momentum ($\frac{\partial u}{\partial z} < 0.075 \frac{m}{s}$ per km) (and thus where the upgradient flux calculation is likely spurious). The vertical axis is a rough estimate of the pressure level of the model output and the horizontal axis is the index of each re-created sounding in the original data. The vertical yellow lines separate the soundings based on which observatory or "mission" they are from. Within each mission, the soundings are in chronological order. From left to right, the six "missions" are those balloons launched from: L'Atalante with Meteomodem radiosondes, L'Atalante with Vaisala radiosondes, the Barbados Cloud Observatory, Meteor, Maria S. Merian, and the Ronald H. Brown.

3.2 Thermodynamic Profiles

While we only change equations related to $\overline{u'_h w'}$ in x101, it is worth considering how these changes may feed back onto the atmospheric state and therefore modulate thermodynamic profiles (T and Q) and their fluxes. It is revealed in Fig. 4 how the predictions of thermodynamic quantities also change when the prognostic $\overline{u'_h w'}$ formulation is introduced. Figure 4a displays profiles of θ rather than T itself to highlight the stability of layers. The default run features a sizeable cold bias for all altitudes below 2.5 km, a cold bias that is only slightly changed (on the order of a tenth of a Kelvin) in x101. In observations, the domain mean Q profile features a "dry nose" around 1 km and a "moist nose" around 1.7 km while both model configurations predict a smoother decrease in moisture with height, meaning they both have moist biases around 1 km and dry biases around 1.7 km altitude. Both configurations also have a dry bias in the lowest 500 meters. Although the directions of these biases are consistent between model configurations, their magnitudes do change on the order of a few tenths of g/kg. The dry bias below 500 m is roughly cut in half from about 0.4 g/kg in x001 to 0.2 k/kg in x101 while the dry bias centered around 1.7 km is degraded in x101 by around 0.1 g/kg.

These differences in thermodynamic profiles are not as large as the differences in the momentum profiles but do exist. This would seem to contradict the findings in L19 where there was no noticeable difference found in the thermodynamic profiles predicted by the prognostic versus the diagnostic $\overline{u'_h w'}$ configurations of the single-column model. The structures of
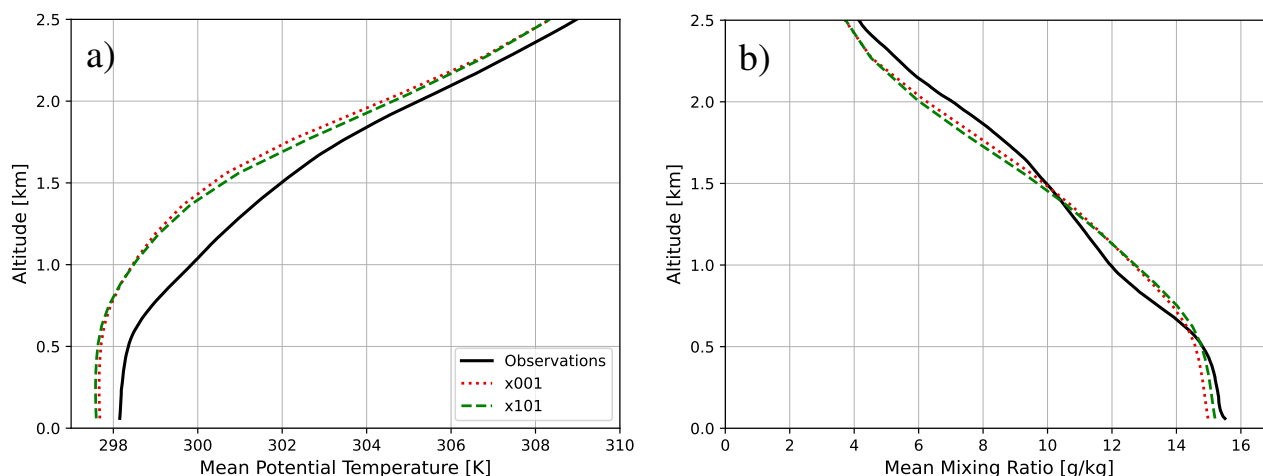
Figure 4. Domain mean vertical profiles from observations, CAM x001, and CAM x101 for potential temperature (θ) and water vapor mixing ratio (Q).

the thermodynamic profiles from the LES in L19 are very similar to those from observations in this study, and those profiles from the single-column model in L19 have similar shapes to the CAM output in this study. We hypothesize that the differences in thermodynamic profiles between x001 and x101 indicate there is additional two-way feedback between $\overline{u'_h w'}$ and scalar fluxes in CAM due to the hindcast framework (i.e., $\overline{u'_h w'}$ changes the atmospheric state, which is further modified and advected

370 by the dynamical core, which then is passed back to the physical parameterizations, including CLUBB, etc.). This feedback would not occur in the single-column model in L19 (which applies a large-scale nudging to specify the mean state fields that are used by the subgrid turbulence scheme).

4 Results of the Experimental Vertical Turbulent Length Scale Formulation

4.1 Dynamic and Thermodynamic Profiles

375 Additional bias reduction in the hindcasts can be achieved with the application of the experimental estimate of L in simulations (x201, x202, x203, and x204) in both the dynamic and thermodynamic profiles (recall that these runs apply the prognostic momentum configuration as in x101 but add an experimental modification to how L is calculated). This is shown in Figs. 5 and 6, which are similar to Figs. 1 and 4 except they include the additional CAM configurations with the experimental formulation for L using the coefficient values described in Table 1.

380 Like x101, the experimental length scale runs have an easterly jet that is more narrow. Unlike x101, however, (which has an enhanced easterly wind bias at the jet maximum), some of them feature reduced easterly bias relative to x001 in this layer. Profiles of v in the experimental length scale runs also tend to qualitatively match observations better than x101. These runs also produce θ profiles with cold biases a few more tenths of Kelvins smaller than both x101 and x001, and Q profiles that
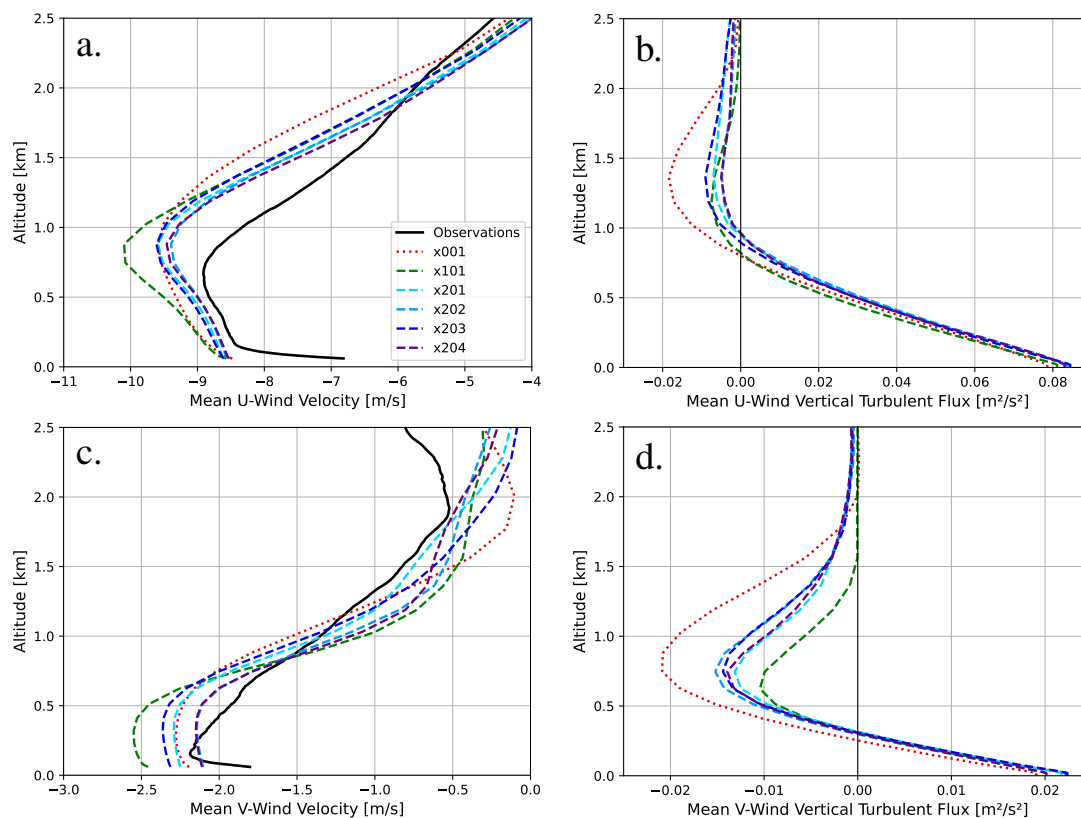
Figure 5. As in Figure 1 but including all CAM configurations

match observations more closely than both x101 and x001 at most altitudes. The dry bias in the lowest 500 meters is nearly
385 eliminated in the experimental length scale runs.

How bias profiles for wind variables depend on the time of day is described for all model configurations in Fig. 7 (based
on sounding launch time). On plots corresponding to individual components, red colors indicate where CAM tends to predict
values that are too negative (more easterly or more northerly) than in reality, while blue colors indicate where wind components
are too positive. On the plots of $\overline{|U_h|}$, the violet colors indicate where CAM tends to overpredict the strength of the wind, while
390 green is where it tends to underpredict. The jet layer easterly (negative) bias in the default run is present at all times of day, but
strongest in the daytime. A smaller magnitude westerly (positive) bias seems to exist between 2 and 5 km in x001: present at
most times of day, except the afternoon when it is small or slightly reversed. Much like the wind magnitudes themselves, biases
in v are generally smaller than those of u, but generally, x001 features a background southerly (positive) bias that is largest at
night and away from the surface. Bias in $\overline{|U_h|}$ appears dominated by biases in u, with winds being too strong in the jet layer,
395 especially in the daytime, and too weak above this, especially at night.

Bias reduction in u when adding the prognostic $\overline{u_h'w'}$ equation can be seen here at almost all times of day when moving
from x001 to x101, particularly between about 1 and 2 km altitude where the maximum magnitude bias changes from around
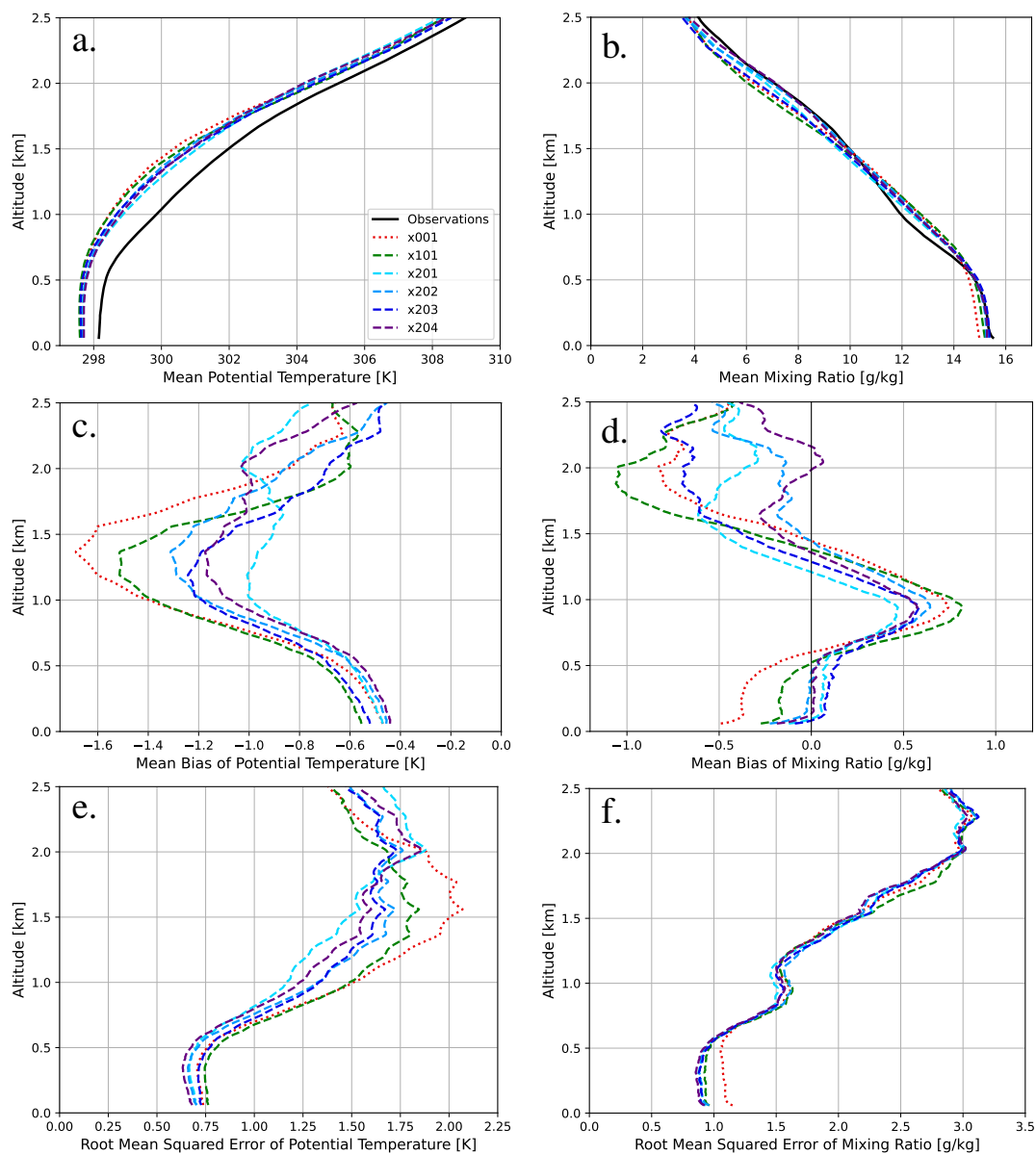
Figure 6. Vertical Profiles of domain mean (top row), mean bias (middle row), and root mean squared error (bottom row) of potential temperature (θ) (left column) and water vapor mixing ratio (Q) (right column) for all CAM configurations.

-1.5 m/s to around -1.2 m/s. This bias is further reduced by adding the experimental length scale estimate in the next 4 rows, especially in the lowest 2 km where the maximum magnitude bias becomes around only -1.0 m/s. For v wind, biases appear

400 mostly the same in all configurations with perhaps the background nocturnal southerly bias being made a few tenths of a m/s worse in some of the experimental length scale runs. Biases in $\overline{|U_h|}$ are similarly reduced moving from x001 to x101 and further reduced moving from x101 to the experimental length scale runs, likely owing to the dominance of u biases.

Errors in state fields throughout the rest of the troposphere (above 2.5km or so) are largely unaffected by the differences between CAM configurations (not shown). Consistent biases in the background tropospheric likely arise from errors in model

405 initialization and from discretization in the dynamical core. Such errors will propagate into boundary layer prediction no matter the skill of the turbulence parameterization, particularly when one considers the free atmosphere as an upper boundary condition to the system. This along with errors arising from the surface layer formulation are likely why, despite improvements seen in bias magnitude for boundary layer winds, the general pattern of bias sign with regards to altitude and time of day remains quite similar for all configurations.

410 Diurnal cycles of mean biases for these three momentum variables between 200 and 2 km are described in Figs. 8. This range of altitudes is chosen to focus on errors in the boundary layer and to exclude errors in the surface layer and the free troposphere. Errors tend to saturate around 2 km in all model configurations, becoming constant with height (e.g., Figs. 2g-i). There is a clear pattern in observations where the winds tend to be weakest in the early afternoon and strongest in the early morning hours. This mean diurnal cycle is captured in each model configuration, but the magnitude of the easterly wind component is

415 always overpredicted. All 3 panels have a range of 3 m/s on their vertical axes. A minor mean reduction in the strong easterly jet bias of around 0.1 m/s can be seen moving from x001 to x101 in panel a). Much greater mean bias reductions in the range of 0.2 to 0.4 m/s can then be seen moving from x101 to the experimental prognostic runs. By comparison, biases in v are all very small, making the mean bias patterns for $\overline{|U_h|}$ essentially the same as those in u (except a more negative u is a larger $\overline{|U_h|}$ here).

420 Figure 9 describes where negative values of K_{eff} are found for u for the experimental length scale runs alongside x101 in the same manner as Fig. 3. Upgradient fluxes are more common and tend to occur in deeper layers in the experimental length scale runs compared to x101, although they are still common in the layer from 950 to 850 hPa, they now often extend higher to near 750 hPa (or roughly 2500 m). We emphasize that these more frequent predictions of upgradient fluxes are not necessarily more accurate, but do demonstrate a likely connection between the prediction of upgradient fluxes and modifications to various

425 terms in the vertical momentum flux budget aimed at improving the shape of the near-surface u and v profiles.

Again, we emphasize that these experimental length scale runs should be generally treated as sensitivity analyses. In other words, we explore how more generalized treatments of eddy turnover timescales could impact simulated state profiles when added to a prognostic momentum setup in the study region. Given how these configurations are better at reducing biases in thermodynamic fields than x101, it may be useful to study the effects of individual tuning parameters further in future work.
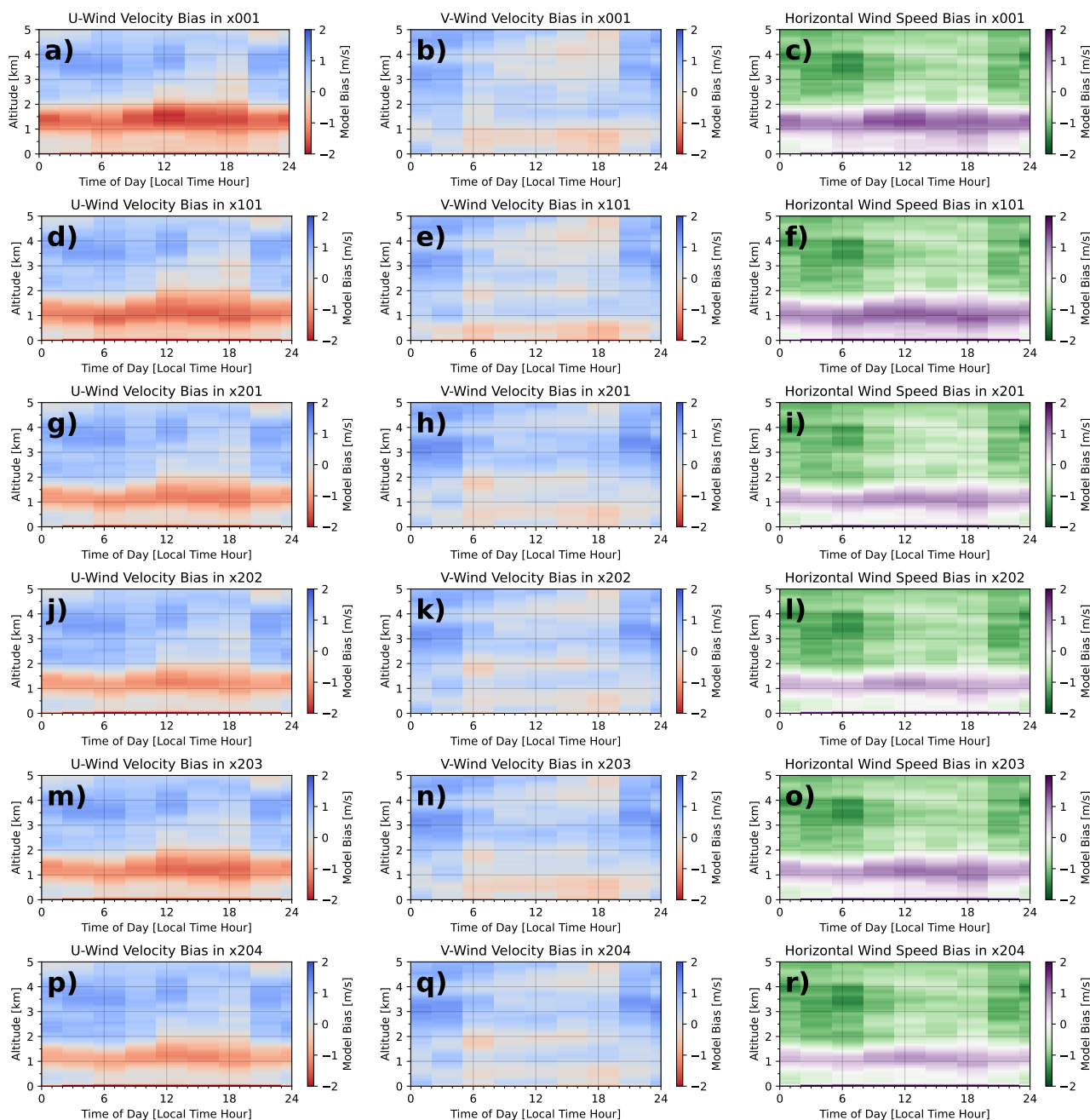
Figure 7. Plots of biases in mean zonal wind speeds (u) (left column), meridional wind speeds (v) (middle column), and horizontal wind magnitudes ($\overline{|U_h|}$) (right column) as a function of time of day and altitude predicted by runs x001 (a,b,c), x101 (d,e,f), x201 (g,h,i), x202 (j,k,l), x203 (m,n,o), and x204 (p,q,r). Biases are averaged every 10 meters of altitude and in eight 3-hour blocks based off of sounding launch times.
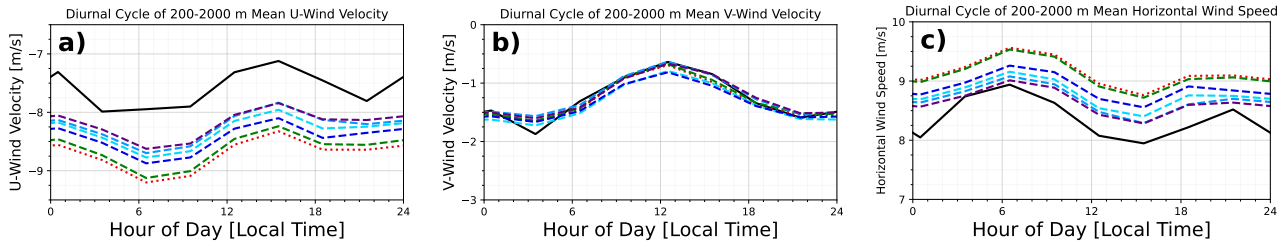
Figure 8. Mean biases in mean zonal wind speeds (u) (a), meridional wind speeds (v) (b), and horizontal wind magnitudes ($\overline{|U_h|}$) (c) predicted by each CAM configuration averaged between 150 m and 750 m altitude. Biases are averaged in eight 3-hour blocks based off of sounding launch times.

4.2 Mean Biases and Root Mean Squared Errors

To quantify the performance of these configurations in simulating EUREC^{4}A/ATOMIC observations, Figure 10 displays the mean biases between altitudes of 200 m and 2 km for each CAM configuration in several state variables. Biases are first calculated for each sounding profile and then the mean is taken over all soundings at each specific altitude (every 10 meters). The blue and red shadings indicate how these biases have changed from the default run (x001). Red colors indicate that the absolute magnitude of the mean bias has increased and blue colors indicate that this magnitude has decreased. The color scale here runs from a 100% decrease in bias magnitude in the darkest blue (complete bias elimination) to a 100% increase in the darkest red (doubling of the bias).

Starting on the left, the column for x001 is completely white because each value serves as the reference bias for the corresponding variable. When the prognostic $\overline{u'_h w'}$ is added, mean biases are reduced on the order of 5 to 10% for most variables in x101. The exceptions to this are v and Q, which see very slight increases in mean bias. The coloring here is not particularly meaningful for these two variables, however, given how small the corresponding mean biases are in x001 to begin with (a minuscule absolute change in these biases appears as a significant relative change). The fact that the $\overline{|U_h|}$ bias is reduced also implies that the u bias reduction is a more important contributor than the v bias degradation. Moving now to the right four columns which represent the prognostic runs when the experimental length scale estimate is added, the blue shading becomes darker, indicating a further reduction of mean bias in most variables. The greatest improvements are seen in u and $\overline{|U_h|}$, as was seen in the profiles with the better depiction of the jet.

Biases cannot paint a full picture, since they do not account for errors that have no mean tendency. Figure 11, is identical to Fig. 10 except it describes root mean squared errors (RMSEs) rather than biases (and has a much more sensitive color scale that runs from a 15% decrease to a 15% increase). Predictions of u are indeed improved when measured by aggregate RMSE reduction (albeit by a few percent) when a prognostic $\overline{u'_h w'}$ calculation is implemented. Somewhat more significant RMSE reductions for u are seen in the next four columns with the inclusion of various versions of the experimental estimate for L. RMSEs are slightly increased in x101 for v and $\overline{|U_h|}$ (by about one percent), but this is not seen in the experimental length scale runs where each version sees a decrease in RMSE for all momentum variables. Although mean u bias between 200 m
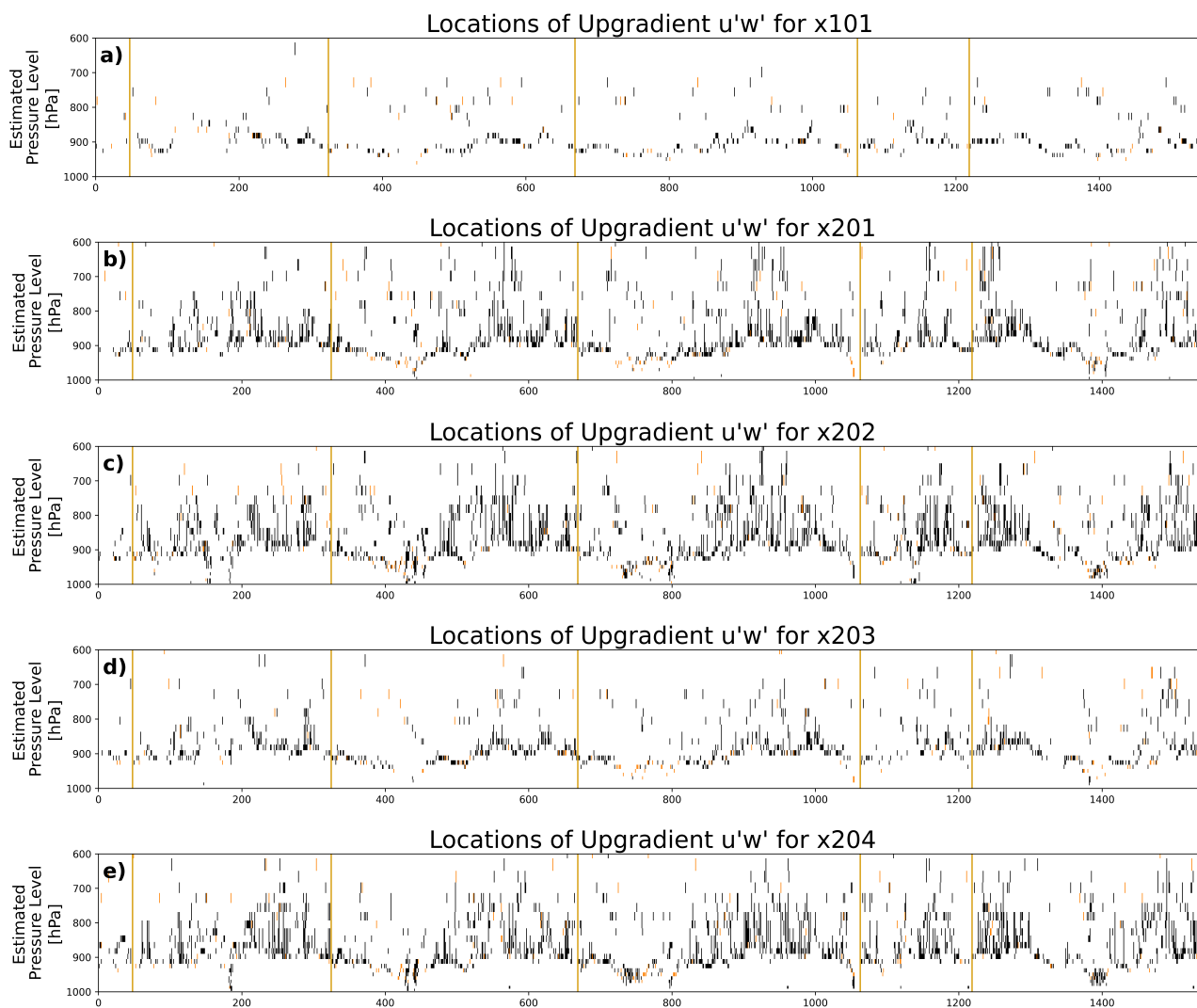
Figure 9. As in Figure 3, except for CAM output from the experimental length scale runs and x101 (panel a) in this figure is identical to panel b) in Figure 3).

and 2 km is reduced in x101 relative to x001, recall that the improvement in the structure of the wind profile seen when moving
455 from x001 to x101 is accompanied by an increase in the strength of the easterly jet, which itself has an easterly bias in x001
(see Fig. 2). The worsened u biases at certain altitudes in x101 likely counteract any improvements in layer-mean RMSEs
that may come from a more accurate wind profile structure. These increased jet layer biases were not seen, however, in the
experimental length scale runs. As such, the improvements that come from improved wind profiles can be seen as reduced
RMSEs for momentum variables in x201, x202, x203, and x204.

460   Improvements in thermodynamic fields are also visible as reductions in RMSEs. This is particularly interesting for x101
relative to x001 since the code used to calculate the turbulent fluxes of scalars (i.e., T and Q) was the same in these runs. Such
improvements again suggest downstream effects of better resolving momentum profile structure via feedback with mean state
fields: a phenomenon not seen in single-column models. Even here, the experimental length scale runs feature improvements
over the default prognostic run, providing further evidence for both the ability of the experimental length scale formulation
465 to improve boundary layer structure and for the significance of this two-way dynamic-thermodynamic feedback. The largest
RMSE reductions seen here for any variable are those in T and the closely related θ, which range between 10 and 20 percent
for the experimental length scale runs. These apparent temperature improvements are likely dominated by the reduction of
the cold bias seen at almost all altitudes when moving from x001 to x101 and then to the experimental length scale runs. A
correspondence of those altitudes with the greatest cold bias reduction to those altitudes with the greatest RMSE reduction can
470 be seen in Fig. 6.

4.3   Horizontal Momentum Budgets

Given the improvement in wind profile predictions relative to observations moving from x101 to the experimental length scale
runs, it is worth comparing how the individual terms that contribute to the time tendency of $\overline{u_h' w'}$ in Eq. 3 differ between them.
If the state variable predictions of a given configuration better match observations, it is conceivable that the corresponding
475 modeled momentum budget profiles that helped make these predictions are themselves better descriptions of physical reality.
In other words, studying these budget terms may provide physical insight into why one configuration's predictions may be
better than those of another. Figure 12 describes vertical profiles of the $\overline{u_h' w'}$ budget terms described in Eq. 3 for both x101
and an example experimental length scale run which scored well in the RMSE evaluation section (x204).

    The "mean advection" term (1) corresponds to the advection of existing $\overline{u_h' w'}$ by the mean vertical wind, while the "turbulent
480 advection" term (2) represents that advection by turbulent perturbations in w. Term 3 is the turbulent production of $\overline{u_h' w'}$ by
variance in w acting in a vertical u or v gradient, while term 4 is that turbulent production by pre-existing $\overline{u_h' w'}$ acting in a
vertical w gradient.

    The "buoyant production" term (5) describes the net change in $\overline{u_h' w'}$ from covariance between parcels of particular values of
buoyancy and with horizontal momentum. "Return to isotropy" (term 6) refers to the effective dissipation of $\overline{u_h' w'}$ determined
485 by τ, and term 7 is the "residual" dissipation term. The "time tendency" is the sum of all other terms (the left-hand side of
the equation), but here, this term is typically orders of magnitude smaller than any of the individual budget terms because of
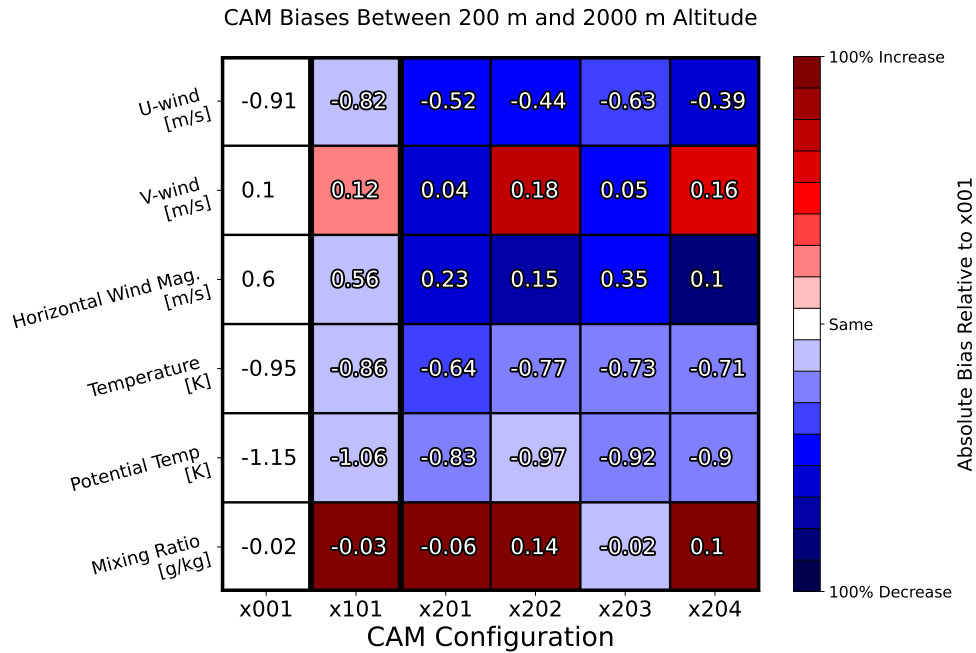
Figure 10. Chart describing absolute errors (biases) of CAM predictions between 200m and 2km altitude relative to sounding data for all model configurations and state variables. All levels are equally weighted. Numbers in each cell describe the actual bias for the corresponding variable and configuration. Colors describe how these errors compare to that of the same variable in the default configuration (x001). Blue colors indicate that the error has a smaller absolute value and red colors indicate that the error has a larger absolute value compared to x001. The colors are scaled such that the darkest blue would be a complete bias eradication and the darkest red would be a doubling of the reference bias in x001. All levels are equally weighted.

how many terms nearly balance each other. In order to make the overall time tendency apparent on the same x-axis scale, it is multiplied by 10 in Fig. 12.

490 One of the most notable differences in these plots is the strong reduction of turbulent production (by $\overline{w'^2}$) in the lowest 1 km in x204 compared to x101 for both the zonal and meridional components. This is accompanied by a similar reduction in the compensating "return-to-isotropy" term, whose magnitude is related to the magnitude of the net $\overline{u_h'w'}$ produced. Another notable difference is the changing of the sign of the buoyancy production term from weakly negative in x101 to notable positive below 700 m, and negative above in x204, particularly in the zonal momentum budget. We hypothesize that this may be related to increased stratification in the θ profile in x204 making vertical transport or air parcels due to buoyancy more difficult in 495 the lowest 700 m. In that case, improvement in the thermodynamic profile in x204 could be leading to changes in atmospheric stability, which in turn lead to changes in buoyant production of $\overline{u_h'w'}$ which then feeds back to changes in the dynamic profiles. Since downgradient diffusion corresponds to a simple balance between turbulent production and return-to-isotropy, the fact that the buoyancy term is so large in x204 could explain the enhanced upgradient fluxes in Fig. 9.
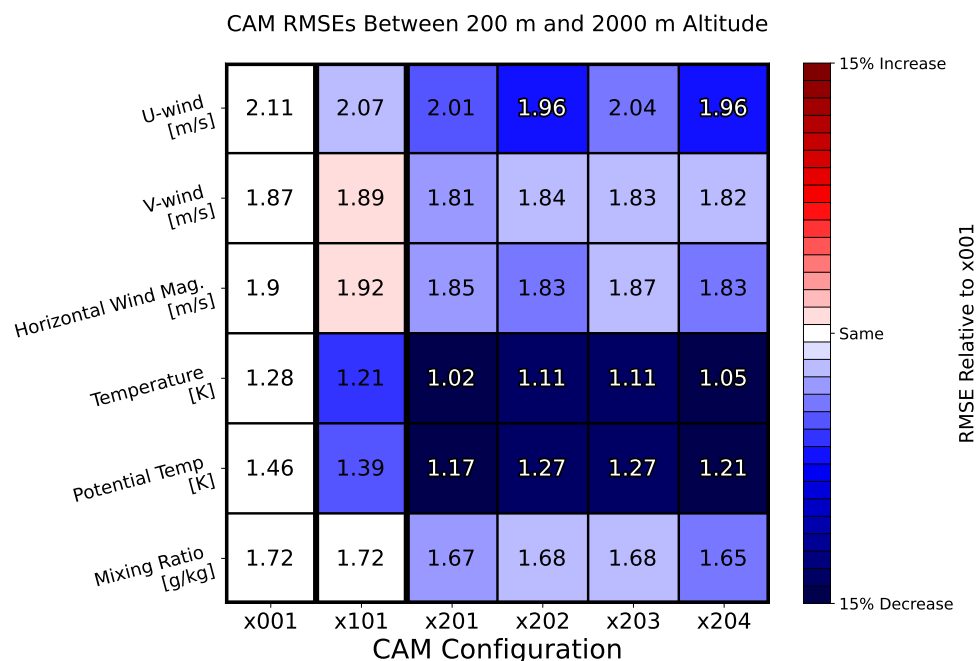
Figure 11. As in Figure 10 but for root mean squared errors. Notice that the color bar has changed to have extrema of +/- 15%.

While relatively qualitative in nature, the evaluation of initialized model simulations against observed state profiles, with
subsequent analysis of turbulence budget terms that either improve or degrade said profiles could provide useful pathways
for parameterization tuning and physical interpretation in future work. The lack of direct observations of turbulent quantities
in this study limits the depth of analysis that can be done here. Estimating similar budgets from LES could prove useful
in understanding whether these changes within the $\overline{u_h' w'}$ budget that lead to more skillful vertical provides are truly due to
improvements in physical processes. This is a target for future work.

5 Discussion

We use 1-day-lead hindcasts produced by a general circulation model (CAM6) to evaluate its prediction of planetary boundary
layer structure in a tropical maritime trade-wind regime. CAM is run in various configurations which vary in how turbulent
momentum fluxes are calculated. One configuration (x001) diagnoses these $\overline{u_h' w'}$ by implementing traditional downgradient
diffusion while all other configurations prognose $\overline{u_h' w'}$ using the unified turbulence scheme, CLUBB. One prognostic con-
figuration uses the default calculation for a vertical turbulent length scale estimate included in CLUBB, while another four
configurations use a more generalized equation to derive L from the eddy diffusivity timescale, each with a different set of
empirically determined coefficients. Predictions from each configuration are evaluated through comparisons to high-quality,
high-resolution, real-world data from 1,546 weather balloons launched during the EUREC[4]A/ATOMIC field campaign.
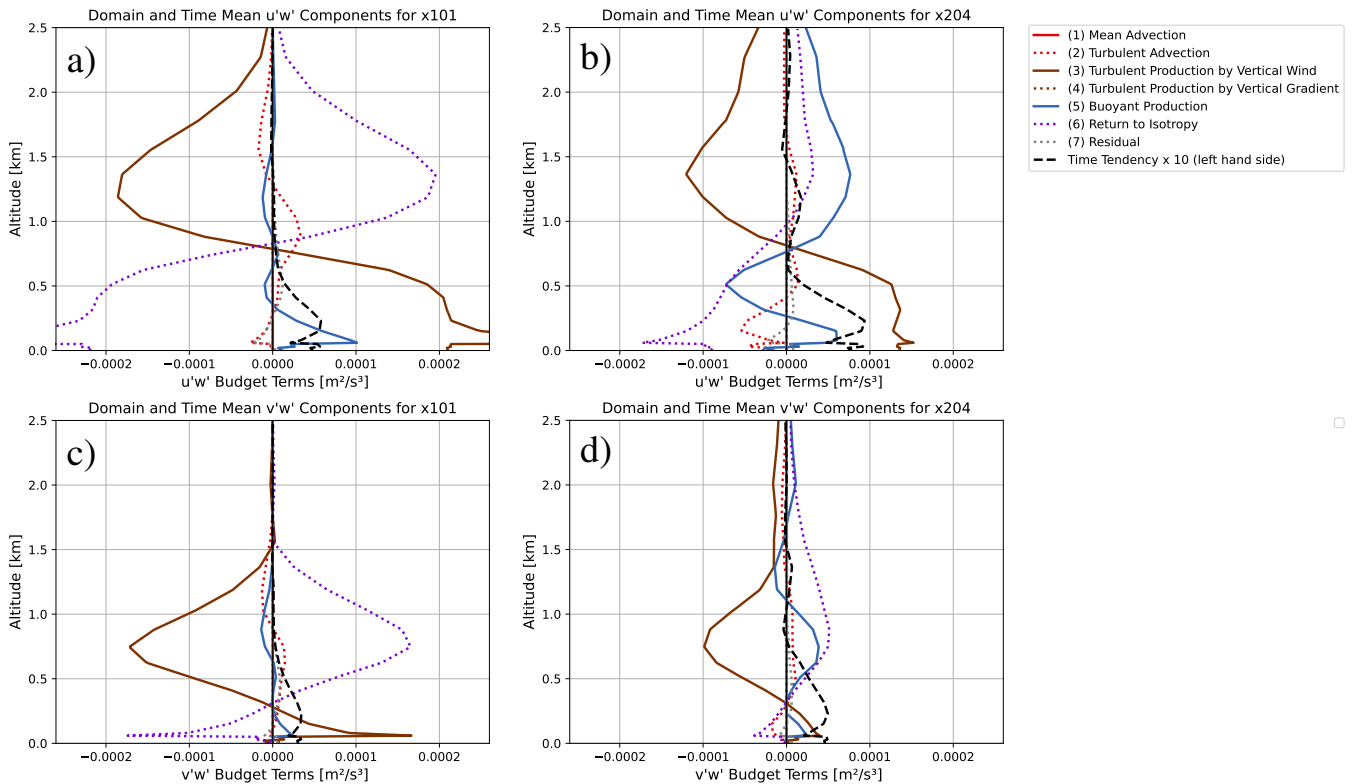
Figure 12. Domain-mean, time-mean, vertical profiles for the components affecting the time tendency of $\overline{u'w'}$ (top row) and $\overline{v'w'}$ (bottom row) for x101 (left column) and x204 (right column) as described in Eq. 3.

Default CAM6 is found to be too diffusive over the EUREC^{4}A/ATOMIC domain when run with standard eddy diffusivity. That is, when compared to observations, it predicts a jet that is too broad in terms of altitude and vertical gradients of u and v that are too weak. The introduction of prognostic $\overline{u_h'w'}$ reduces these biases by predicting a narrower jet, albeit one that is still too strong in terms of maximum velocity. This is a qualitative improvement in terms of how well the structure of this jet matches both observations from EUREC^{4}A/ATOMIC and results from LES in both L19 and D20. This suggests higher-order momentum flux formulations, particularly those that permit countergradient fluxes, may be able to improve the representation of lower troposphere structure in trade-wind regimes, perhaps in conjunction with improvements to the surface layer formulation.

Further improvements in the prediction of boundary layer wind profiles are observed (as measured by root mean squared error in the relevant layer) when the experimental formulation of the turbulent length scale L as first described in Guo et al. (2021) is included, and the relevant parameters are quasi-optimized. This suggests a more flexible, regime-specific strategy for estimating L in GCMs can provide further improvement in the vertical structure of $\overline{u_h'w'}$ and subsequent wind profiles in these regions. These results do not point to any particular set of parameters leading to the best predictions but rather demonstrate that model predictions of boundary layer structure are sensitive to and can be improved via the tuning of these coefficients.

This study is a targeted regional investigation and as such, the improvements seen here cannot necessarily be generalized to the global climate system without further exploration. Reductions of errors in any particular run here do not necessarily imply

530 that that run would generate better predictions globally. A parameterization that improves the structure of the boundary layer in a steady-state shallow cumulus regime over a relatively homogeneous calm ocean might also make predictions worse in regions with more orography and heterogeneous dynamical forcing. Model grid spacing is still on the order of $1°$ in mountainous regions where topography can vary vertically by kilometers and thus these regions have the same requirement for subgrid parameterizations. How $\overline{u_h' w'}$ in the boundary layer responds to this roughness in boundary layer structure still must be captured

535 by the same parameterization used by the model over the flat ocean surface. Therefore, one suggestion arising from this work is to more closely tie model development experiments to a variety of field campaign datasets and regions.

Although forecasts may improve when $\overline{u_h' w'}$ is prognosed rather than diagnosed, potential trade-offs exist in terms of computational cost and complexity. In the case of the CLUBB code specifically, the total computational cost of CLUBB increases by a few percent when adding prognostic $\overline{u_h' w'}$ if scalar fluxes have already been prognosed. This is only because many of the

540 computations used to calculate scalar fluxes can be reused (Larson et al., 2019). Nonetheless, this is an increase in computational cost and one that would be larger in a model where a high-order closure is not already being implemented. Even besides the issue of computational cost, the inclusion of equations with more terms used to prognose $\overline{u_h' w'}$ increases the complexity of the model, thereby increasing the risk of introducing artifacts and increasing the difficulty of understanding model behavior (Mihailović et al., 2014).

545 The use of short-term initialized hindcasts here can serve to bridge a hierarchy gap between using long-term climate integrations and using single-column models or LES as tools for improving GCMs. This can be done since the large-scale environment is realistic in these hindcasts while significant model biases still appear within 1-2 days after initialization as can be seen here with CAM's prediction of wind speeds and jet height. Unlike in single-column models, here the simulated atmosphere can vary spatially and the subgrid parameterizations in question are allowed to impact the large-scale flow. This is unlike the 'one-way'

550 transfer of information generally conferred by nudged configurations. Since the model is initialized with an observed state, observational profiles can be directly compared to the model simulation in a deterministic sense, rather than being averaged and compared to climatology in a more traditional assessment. Initialized simulations are also cheap compared to traditional climatology comparisons, with the six different sets of experiments here costing less than a single multi-year tuning run typically used by climate modeling centers.

555 Improvements in boundary layer structure are likely limited by propagation of errors from the near-surface layer and from the background troposphere generated from the model's dynamical core and initialization. This issue arises from the nature of a global model and is not present when working with a single-column model or LES where the background forcing is prescribed as in Larson et al. (2019). Direct comparisons of findings here to the findings of past studies are thus inherently limited because of this innate difference between the types of models implemented. Future work should test how sensitive the improvements

560 demonstrated in this study are to the surface layer formulation and to the structure of existing background errors that remain unaffected by changes in turbulence parameterization.

In order to improve predictions globally, modelers should identify other regions with strong biases that are thought to result from boundary layer parameterization. Analyses similar to this can prove fruitful for either noting similar errors or determining parameterizations where responses may differ with respect to varying atmospheric regimes. Additional field campaigns report-
565 ing detailed observations in these regions alongside LES tailored to those regions would greatly benefit future studies seeking to improve turbulence parameterizations in GCMs.

Code and data availability. EUREC[4]A/ATOMIC soundings and derived quantities data used for this project were acquired from https://doi.org/10.25326/137 and are described in Stephan et al. (2020). The version of the Community Atmosphere Model run here was cesm2.2.0 and is available at https://github.com/ESCOMP/CAM. ERA5 data used to initialize the hindcasts was downloaded from the Copernicus
570 Climate Data Store (CDS), available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. The Betacast software used for hindcast configuration and initialization is available at https://github.com/zarzycki/betacast and is described in Zarzycki and Jablonowski (2015). The version of Betacast used in this manuscript is archived at https://doi.org/10.5281/zenodo.8184863. The data generated for this project (cesm_x*.tar) are available via Penn State's Data Commons at https://doi.org/10.26208/DCSY-HY63. A checkout of the model code (cesm_EUREC4A_sourcetree.tar.gz), case directories for the various configurations (EUREC4A_cases.tar.gz), and the scripts (gmd-prog-
575 upwp-scripts.tar.gz) used to analyze the data and reproduce the results of this manuscript are available via Zenodo at https://doi.org/10.5281/zenodo.8184358.

Appendix A: Prognostic momementum derivation and closures

Starting from Eq. 3.3 in Larson (2022):

$$\frac{\partial \overline{u_h'w'}}{\partial t} = \underbrace{-\overline{w}\frac{\partial \overline{u_h'w'}}{\partial z}}_{mean-adv} \underbrace{-\frac{1}{\rho}\frac{\partial \rho\overline{w'^2u_h'}}{\partial z}}_{turb-adv} \underbrace{-\overline{w'^2}\frac{\partial \overline{u_h}}{\partial z}}_{turb-prod} \underbrace{-\overline{u_h'w'}\frac{\partial \overline{w}}{\partial z}}_{accum} + \underbrace{\frac{g}{\theta_{vs}}\overline{u_h'\theta_v'}}_{buoy-prod} \underbrace{-\frac{1}{\rho}\overline{(u_h'\frac{\partial p'}{\partial z} + w'\frac{\partial p'}{\partial x_h})}}_{pressure} \underbrace{-\epsilon_{u_hw}}_{dissip} \tag{A1}$$

580 where ρ is density of air, g is gravity, θ_v is virtual potential temperature, and θ_{vs} is a dry base-state potential temperature value. Substituting in Eq. 3.30 from Larson (2022):

$$\underbrace{-\frac{1}{\rho}\overline{(u_h'\frac{\partial p'}{\partial z} + w'\frac{\partial p'}{\partial x_h})}}_{pressure} \approx \underbrace{-\frac{C_6}{\tau}\overline{u_h'w'}}_{pr1} + \underbrace{C_7\overline{u_h'w'}\frac{\partial \overline{w}}{\partial z}}_{pr2} \underbrace{-C_7\frac{g}{\theta_{vs}}\overline{u_h'\theta_v'}}_{pr3} + \underbrace{C_{uu,shear}\overline{w'^2}\frac{\partial \overline{u_h}}{\partial z}}_{pr4} \tag{A2}$$

where C_6, C_7, and $C_{uu,shear}$ are all empirical coefficients. Note that $C_{uu,shear}$ is equivalent to C_{7upwp} from Nardi et al. (2022). Eq. A1 Becomes

585

$$\underbrace{\frac{\partial \overline{u_h' w'}}{\partial t}}_{} = \underbrace{-\overline{w}\frac{\partial \overline{u_h' w'}}{\partial z}}_{mean-adv} \underbrace{-\frac{1}{\rho}\frac{\partial \rho \overline{w'^2 u_h'}}{\partial z}}_{turb-adv} \underbrace{-\overline{w'^2}\frac{\partial \overline{u_h}}{\partial z}}_{turb-prod} \underbrace{-\overline{u_h' w'}\frac{\partial \overline{w}}{\partial z}}_{accum} + \underbrace{\frac{g}{\theta_{vs}}\overline{u_h' \theta_v'}}_{buoy-prod}$$

$$\underbrace{-\frac{C_6}{\tau}\overline{u_h' w'}}_{pr1} + \underbrace{C_7\overline{u_h' w'}\frac{\partial \overline{w}}{\partial z}}_{pr2} - \underbrace{C_7\frac{g}{\theta_{vs}}\overline{u_h' \theta_v'}}_{pr3} + \underbrace{C_{uu,shear}\overline{w'^2}\frac{\partial \overline{u_h}}{\partial z}}_{pr4} - \underbrace{\epsilon_{u_h w}}_{dissip} \quad (A3)$$

Rearranging terms with common expressions:

590

$$\underbrace{\frac{\partial \overline{u_h' w'}}{\partial t}}_{} = \underbrace{-\overline{w}\frac{\partial \overline{u_h' w'}}{\partial z}}_{mean-adv} \underbrace{-\frac{1}{\rho}\frac{\partial \rho \overline{w'^2 u_h'}}{\partial z}}_{turb-adv} \underbrace{-\overline{w'^2}\frac{\partial \overline{u_h}}{\partial z}}_{turb-prod} + \underbrace{C_{uu,shear}\overline{w'^2}\frac{\partial \overline{u_h}}{\partial z}}_{pr4} \underbrace{-\overline{u_h' w'}\frac{\partial \overline{w}}{\partial z}}_{accum}$$

$$+ \underbrace{C_7\overline{u_h' w'}\frac{\partial \overline{w}}{\partial z}}_{pr2} + \underbrace{\frac{g}{\theta_{vs}}\overline{u_h' \theta_v'}}_{buoy-prod} - \underbrace{C_7\frac{g}{\theta_{vs}}\overline{u_h' \theta_v'}}_{pr3} - \underbrace{\frac{C_6}{\tau}\overline{u_h' w'}}_{pr1} - \underbrace{\epsilon_{u_h w}}_{dissip} \quad (A4)$$

Combining like terms gives us:

$$\underbrace{\frac{\partial \overline{u_h' w'}}{\partial t}}_{} = \underbrace{-\overline{w}\frac{\partial \overline{u_h' w'}}{\partial z}}_{mean-adv} \underbrace{-\frac{1}{\rho}\frac{\partial \rho \overline{w'^2 u_h'}}{\partial z}}_{turb-adv} \underbrace{-(1-C_{uu,shear})\overline{w'^2}\frac{\partial \overline{u_h}}{\partial z}}_{turb-prod} \underbrace{-(1-C_7)\overline{u_h' w'}\frac{\partial \overline{w}}{\partial z}}_{accum}$$

595

$$+ \underbrace{(1-C_7)\frac{g}{\theta_{vs}}\overline{u_h' \theta_v'}}_{buoy-prod} - \underbrace{\frac{C_6}{\tau}\overline{u_h' w'}}_{pr1} - \underbrace{\epsilon_{u_h w}}_{dissip} \quad (A5)$$

Further, the closure used for $\rho \overline{w'^2 u_h'}$ is closed with Eq. 5 in Larson et al. (2019):

$$\frac{\partial}{\partial z}(\rho \overline{w'^2 u_h'}) \approx \frac{\partial}{\partial z}(a_1 \frac{\overline{w'^3}}{\overline{w'^2}}\overline{u_h' w'}) \quad (A6)$$

the closure for $\overline{u_h' \theta_v'}$ is as in Eq. 33 in Golaz et al. (2002):

$$\overline{u_h' \theta_v'} = \overline{u_h' \theta_l'} + \frac{1-\epsilon_0}{\epsilon_0}\theta_0 \overline{u_h' q_t'} + (\frac{L_v}{c_p}(\frac{p_0}{p})^{\frac{R_d}{c_p}} - \frac{1}{\epsilon_0}\theta_0)\overline{u_h' q_l'} \quad (A7)$$

600   where θ_l is the liquid water potential temperature, R_d and R_v are the gas constants for dry air and water respectively, $\epsilon_0 = \frac{R_d}{R_v}$, C_p is the heat capacity of dry air at constant pressure, L_v is the latent heat of vaporization of water, P_0 is a reference pressure, θ_0 is a reference potential temperature, r_l is liquid specific water content, and r_t is total specific water content. $\overline{u_h' \theta_l'}$ and $\overline{u_h' q_l'}$ are in turn closed with Eq. 9 in Larson et al. (2019):

$$\overline{u_h'\psi'} = (\frac{\tau}{C_{u\psi}^{pi}})(-\overline{u_h'w'}\frac{\partial\overline{\psi}}{\partial z} - (1 - C_{u\psi}^{ps})\overline{w'\psi'}\frac{\partial\overline{u_h}}{\partial z}) \tag{A8}$$

The variable ψ here represents either θ_l or r_t and the constants $C_{u\psi}^{pi}$ and $C_{u\psi}^{ps}$ are set equal to C_6 (2) and 0 respectively. Finally, the closure used for $\epsilon_{u_h w}$ is setting it to 0 as in Eq. 3.31 in Larson (2022):

$$\underbrace{-\epsilon_{u_h w}}_{dissip} \approx 0 \tag{A9}$$

Author contributions. SG: literature review, data organization and cleaning, data analysis, figure generation, results interpretation, writing. CZ: conceptualization, simulation generation, proofreading and formatting, project administration, supervision, and funding acquisition.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. SG and CZ would like to thank Ying Pan, Jerry Y. Harrington, Kyle Nardi, and Vince Larson who provided comments on a draft of this manuscript. SG would also like to thank Allen Mewhinney and Marley Majetic for their suggestions, encouragement, and help with coding.

This research is jointly funded as part of a Climate Process Team (CPT) under Grant AGS-1916689 from the National Science Foundation (NSF) and Grant NA19OAR4310363 from the National Oceanic and Atmospheric Administration (NOAA). The authors acknowledge computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the NSF. Additional data analysis for this research was performed on the Pennsylvania State University's Institute for Computational and Data Sciences' Roar supercomputer.

References

620 Berkowicz, R. and Prahm, L. P.: Generalization of K Theory for Turbulent Diffusion. Part I: Spectral Turbulent Diffusivity Concept, Journal of Applied Meteorology and Climatology, 18, 266–272, https://doi.org/10.1175/1520-0450(1979)018<0266:GOTFTD>2.0.CO;2, 1979.

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., and Schanen, D. P.: Higher-Order Turbulence Closure and Its Impact on Climate Simulations in the Community Atmosphere Model, J. Clim., 26, 9655–9676, https://doi.org/10.1175/JCLI-D-13-00075.1, 2013.

625 Bogenschutz, P. A., Gettelman, A., Hannay, C., Larson, V. E., Neale, R. B., Craig, C., and Chen, C.-C.: The path to CAM6: coupled simulations with CAM5.4 and CAM5.5, Geoscientific Model Development, 11, 235–255, https://doi.org/10.5194/gmd-11-235-2018, 2018.

Brilouet, P.-E., Lothon, M., Etienne, J.-C., Richard, P., Bony, S., Lernoult, J., Bellec, H., Vergez, G., Perrin, T., Delanoë, J., Jiang, T., Pouvesle, F., Lainard, C., Cluzeau, M., Guiraud, L., Medina, P., and Charoy, T.: The EUREC^{4}A turbulence dataset derived from the SAFIRE ATR 42 aircraft, Earth System Science Data, 13, 3379–3398, https://doi.org/10.5194/essd-13-3379-2021, 2021.

630 Bryan, G. H., Worsnop, R. P., Lundquist, J. K., and Zhang, J. A.: A Simple Method for Simulating Wind Profiles in the Boundary Layer of Tropical Cyclones, Boundary-Layer Meteorology, 162, 475–502, https://doi.org/10.1007/s10546-016-0207-0, 2017.

Ceppi, P., Brient, F., Zelinka, M. D., and Hartmann, D. L.: Cloud feedback mechanisms and their representation in global climate models, WIREs Climate Change, 8, e465, https://doi.org/https://doi.org/10.1002/wcc.465, 2017.

CISL: Cheyenne: HPE/SGI ICE XA System (University Community Computing), Boulder, CO: National Center for Atmospheric Research, 635 https://doi.org/10.5065/D6RX99HX, 2019.

Clarke, R. H.: Observational studies in the atmospheric boundary layer, Q. J. R. Meteorolog. Soc., 96, 91–114, https://doi.org/10.1002/qj.49709640709, 1970.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, 640 W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, e2019MS001 916, https://doi.org/https://doi.org/10.1029/2019MS001916, e2019MS001916 2019MS001916, 2020.

645 Dixit, V., Nuijens, L., and Helfer, K. C.: Counter-Gradient Momentum Transport Through Subtropical Shallow Convection in ICON-LEM Simulations, Journal of Advances in Modeling Earth Systems, 13, 2020.

Ekman, V. W.: On the influence of the Earth's rotation on ocean-currents, 1905.

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., Lamarque, J.-F., Fasullo, J. T., Bailey, D. A., Lawrence, D. M., and Mills, M. J.: High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2), Geophysical 650 Research Letters, 46, 8329–8337, https://doi.org/10.1029/2019GL083978, 2019.

Golaz, J.-C., Larson, V. E., and Cotton, W. R.: A PDF-Based Model for Boundary Layer Clouds. Part I: Method and Model Description, Journal of the Atmospheric Sciences, 59, 3540 – 3551, https://doi.org/10.1175/1520-0469(2002)059<3540:APBMFB>2.0.CO;2, 2002.

Guo, Z., Griffin, B. M., Domke, S., and Larson, V. E.: A Parameterization of Turbulent Dissipation and Pressure Damping Time Scales in Stably Stratified Inversions, and its Effects on Low Clouds in Global Simulations, Journal of Advances in Modeling Earth Systems, 13, 655 e2020MS002 278, https://doi.org/10.1029/2020MS002278, 2021.

Helfer, K. C., Nuijens, L., and Dixit, V. V.: The role of shallow convection in the momentum budget of the trades from large-eddy-simulation hindcasts, Q. J. R. Meteorolog. Soc., 147, 2490–2505, https://doi.org/10.1002/qj.4035, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holland, J. Z. and Rasmusson, E. M.: Measurements of the Atmospheric Mass, Energy, and Momentum Budgets Over a 500-Kilometer Square of Tropical Ocean, Mon. Weather Rev., 101, 44–55, https://doi.org/10.1175/1520-0493(1973)101<0044:MOTAME>2.3.CO;2, 1973.

Larson, V. E.: CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere, arXiv preprint arXiv:1711.03675, 2022.

Larson, V. E., Domke, S., and Griffin, B. M.: Momentum Transport in Shallow Cumulus Clouds and Its Parameterization by Higher-Order Closure, Journal of Advances in Modeling Earth Systems, 11, 3419–3442, https://doi.org/https://doi.org/10.1029/2019MS001743, 2019.

Lauritzen, P. H., Nair, R. D., Herrington, A. R., Callaghan, P., Goldhaber, S., Dennis, J. M., Bacmeister, J. T., Eaton, B. E., Zarzycki, C. M., Taylor, M. A., Ullrich, P. A., Dubos, T., Gettelman, A., Neale, R. B., Dobbins, B., Reed, K. A., Hannay, C., Medeiros, B., Benedict, J. J., and Tribbia, J. J.: NCAR Release of CAM-SE in CESM2.0: A Reformulation of the Spectral Element Dynamical Core in Dry-Mass Vertical Coordinates With Comprehensive Treatment of Condensates and Energy, Journal of Advances in Modeling Earth Systems, 10, 1537–1570, https://doi.org/10.1029/2017MS001257, 2018.

Mihailović, D. T., Mimić, G., and Arsenić, I.: Climate Predictions: The Chaos and Complexity in Climate Models, Adv. Meteorol., 2014, https://doi.org/10.1155/2014/878249, 2014.

Nardi, K. M., Zarzycki, C. M., Larson, V. E., and Bryan, G. H.: Assessing the Sensitivity of the Tropical Cyclone Boundary Layer to the Parameterization of Momentum Flux in the Community Earth System Model, Monthly Weather Review, 150, 883 – 906, https://doi.org/10.1175/MWR-D-21-0186.1, 2022.

Nelder, J. A. and Mead, R.: A simplex method for function minimization, The computer journal, 7, 308–313, 1965.

Rauber, R. M., Stevens, B., Ochs, H. T., Knight, C., Albrecht, B. A., Blyth, A. M., Fairall, C. W., Jensen, J. B., Lasher-Trapp, S. G., Mayol-Bracero, O. L., Vali, G., Anderson, J. R., Baker, B. A., Bandy, A. R., Burnet, E., Brenguier, J.-L., Brewer, W. A., Brown, P. R. A., Chuang, R., Cotton, W. R., Di Girolamo, L., Geerts, B., Gerber, H., Göke, S., Gomes, L., Heikes, B. G., Hudson, J. G., Kollias, P., Lawson, R. R., Krueger, S. K., Lenschow, D. H., Nuijens, L., O'Sullivan, D. W., Rilling, R. A., Rogers, D. C., Siebesma, A. P., Snodgrass, E., Stith, J. L., Thornton, D. C., Tucker, S., Twohy, C. H., and Zuidema, P.: Rain in Shallow Cumulus Over the Ocean: The RICO Campaign, Bull. Am. Meteorol. Soc., 88, 1912–1928, https://doi.org/10.1175/BAMS-88-12-1912, 2007.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W.: An Improved In Situ and Satellite SST Analysis for Climate, Journal of Climate, 15, 1609–1625, https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2, 2002.

Ruppert, J. H.: Diurnal timescale feedbacks in the tropical cumulus regime, J. Adv. Model. Earth Syst., 8, 1483–1500, https://doi.org/10.1002/2016MS000713, 2016.

Savazzi, A. C. M., Nuijens, L., Sandu, I., George, G., and Bechtold, P.: The representation of the trade winds in ECMWF forecasts and reanalyses during EUREC^{4}A, Atmospheric Chemistry and Physics, 22, 13 049–13 066, https://doi.org/10.5194/acp-22-13049-2022, 2022.

Stensrud, D. J.: Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models, Cambridge University Press, https://doi.org/10.1017/CBO9780511812590, 2007.

695 Stephan, C. C., Schnitt, S., Schulz, H., Bellenger, H., de Szoeke, S. P., Acquistapace, C., Baier, K., Dauhut, T., Laxenaire, R., Morfa-Avalos, Y., Person, R., Meléndez, E. Q., Bagheri, G., Böck, T., Daley, A., Güttler, J., Helfer, K. C., Los, S. A., Neuberger, A., Röttenbacher, J., Raeke, A., Ringel, M., Ritschel, M., Sadoulet, P., Schirmacher, I., Stolla, M. K., Wright, E. E., Charpentier, B., Doerenbecher, A., Wilson, R. K., Jansen, F., Kinne, S., Reverdin, G., Speich, S., Bony, S., and Stevens, B.: Ship- and island-based atmospheric soundings from the 2020 EUREC4A field campaign, Earth System Science Data, 2020.

700 Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., Karstensen, J., Quinn, P. K., Speich, S., Acquistapace, C., Aemisegger, F., Albright, A. L., Bellenger, H., Bodenschatz, E., Caesar, K.-A., Chewitt-Lucas, R., de Boer, G., Delanoë, J., Denby, L., Ewald, F., Fildier, B., Forde, M., George, G., Gross, S., Hagen, M., Hausold, A., Heywood, K. J., Hirsch, L., Jacob, M., Jansen, F., Kinne, S., Klocke, D., Kölling, T., Konow, H., Lothon, M., Mohr, W., Naumann, A. K., Nuijens, L., Olivier, L., Pincus, R., Pöhlker, M., Reverdin, G., Roberts, G., Schnitt, S., Schulz, H., Siebesma, A. P., Stephan, C. C., Sullivan, P., Touzé-Peiffer, L., Vial, J., Vogel, R., Zuidema, P., Alexander, 705 N., Alves, L., Arixi, S., Asmath, H., Bagheri, G., Baier, K., Bailey, A., Baranowski, D., Baron, A., Barrau, S., Barrett, P. A., Batier, F., Behrendt, A., Bendinger, A., Beucher, F., Bigorre, S., Blades, E., Blossey, P., Bock, O., Böing, S., Bosser, P., Bourras, D., Bouruet-Aubertot, P., Bower, K., Branellec, P., Branger, H., Brennek, M., Brewer, A., Brilouet, P.-E., Brügmann, B., Buehler, S. A., Burke, E., Burton, R., Calmer, R., Canonici, J.-C., Carton, X., Cato Jr., G., Charles, J. A., Chazette, P., Chen, Y., Chilinski, M. T., Choularton, T., Chuang, P., Clarke, S., Coe, H., Cornet, C., Coutris, P., Couvreux, F., Crewell, S., Cronin, T., Cui, Z., Cuypers, Y., Daley, A., Damerell, 710 G. M., Dauhut, T., Deneke, H., Desbios, J.-P., Dörner, S., Donner, S., Douet, V., Drushka, K., Dütsch, M., Ehrlich, A., Emanuel, K., Emmanouilidis, A., Etienne, J.-C., Etienne-Leblanc, S., Faure, G., Feingold, G., Ferrero, L., Fix, A., Flamant, C., Flatau, P. J., Foltz, G. R., Forster, L., Furtuna, I., Gadian, A., Galewsky, J., Gallagher, M., Gallimore, P., Gaston, C., Gentemann, C., Geyskens, N., Giez, A., Gollop, J., Gouirand, I., Gourbeyre, C., de Graaf, D., de Groot, G. E., Grosz, R., Güttler, J., Gutleben, M., Hall, K., Harris, G., Helfer, K. C., Henze, D., Herbert, C., Holanda, B., Ibanez-Landeta, A., Intrieri, J., Iyer, S., Julien, F., Kalesse, H., Kazil, J., Kellman, A., Kidane, 715 A. T., Kirchner, U., Klingebiel, M., Körner, M., Kremper, L. A., Kretzschmar, J., Krüger, O., Kumala, W., Kurz, A., L'Hégaret, P., Labaste, M., Lachlan-Cope, T., Laing, A., Landschützer, P., Lang, T., Lange, D., Lange, I., Laplace, C., Lavik, G., Laxenaire, R., Le Bihan, C., Leandro, M., Lefevre, N., Lena, M., Lenschow, D., Li, Q., Lloyd, G., Los, S., Losi, N., Lovell, O., Luneau, C., Makuch, P., Malinowski, S., Manta, G., Marinou, E., Marsden, N., Masson, S., Maury, N., Mayer, B., Mayers-Als, M., Mazel, C., McGeary, W., McWilliams, J. C., Mech, M., Mehlmann, M., Meroni, A. N., Mieslinger, T., Minikin, A., Minnett, P., Möller, G., Morfa Avalos, Y., Muller, C., Musat, I., 720 Napoli, A., Neuberger, A., Noisel, C., Noone, D., Nordsiek, F., Nowak, J. L., Oswald, L., Parker, D. J., Peck, C., Person, R., Philippi, M., Plueddemann, A., Pöhlker, C., Pörtge, V., Pöschl, U., Pologne, L., Posyniak, M., Prange, M., Quiñones Meléndez, E., Radtke, J., Ramage, K., Reimann, J., Renault, L., Reus, K., Reyes, A., Ribbe, J., Ringel, M., Ritschel, M., Rocha, C. B., Rochetin, N., Röttenbacher, J., Rollo, C., Royer, H., Sadoulet, P., Saffin, L., Sandiford, S., Sandu, I., Schäfer, M., Schemann, V., Schirmacher, I., Schlenczek, O., Schmidt, J., Schröder, M., Schwarzenboeck, A., Sealy, A., Senff, C. J., Serikov, I., Shohan, S., Siddle, E., Smirnov, A., Späth, F., Spooner, B., Stolla, 725 M. K., Szkółka, W., de Szoeke, S. P., Tarot, S., Tetoni, E., Thompson, E., Thomson, J., Tomassini, L., Totems, J., Ubele, A. A., Villiger, L., von Arx, J., Wagner, T., Walther, A., Webber, B., Wendisch, M., Whitehall, S., Wiltshire, A., Wing, A. A., Wirth, M., Wiskandt, J., Wolf, K., Worbes, L., Wright, E., Wulfmeyer, V., Young, S., Zhang, C., Zhang, D., Ziemen, F., Zinner, T., and Zöger, M.: EUREC[4]A, Earth System Science Data, 13, 4067–4119, https://doi.org/10.5194/essd-13-4067-2021, 2021.

Stevens, D. E., Ackerman, A. S., and Bretherton, C. S.: Effects of Domain Size and Numerical Resolution on the Simulation of Shallow Cumulus Convection, Journal of the Atmospheric Sciences, 59, 3285 – 3301, https://doi.org/10.1175/1520-0469(2002)059<3285:EODSAN>2.0.CO;2, 2002.

Trenberth, E., Berry, C., and Buja, E.: Vertical Interpolation and Truncation of Model-coordinate Data, https://doi.org/10.5065/D6HX19NH, 1993.

Trenberth, K. E., Caron, J. M., and Stepaniak, D. P.: The atmospheric energy budget and implications for surface fluxes and ocean heat transports, Climate Dynamics, 17, 259–276, https://doi.org/10.1007/PL00007927, 2001.

Vial, J., Vogel, R., Bony, S., Stevens, B., Winker, D. M., Cai, X., Hohenegger, C., Naumann, A. K., and Brogniez, H.: A New Look at the Daily Cycle of Trade Wind Cumuli, J. Adv. Model. Earth Syst., 11, 3148–3166, https://doi.org/10.1029/2019MS001746, 2019.

Zarzycki, C. M. and Jablonowski, C.: Experimental Tropical Cyclone Forecasts Using a Variable-Resolution Global Model, Monthly Weather Review, 143, 4012–4037, https://doi.org/10.1175/MWR-D-15-0159.1, 2015.