Supporting Information for "A machine learning parameterization of clouds in a coarse-resolution climate model for unbiased radiation"

Brian Henn[1], Yakelyn R. Jauregui[2], Spencer K. Clark[1,3], Noah D. Brenowitz[4], Jeremy McGibbon[1], Oliver Watt-Meyer[1], Andrew G. Pauling[5], Christopher S. Bretherton[1]

[1]Allen Institute for Artificial Intelligence, Seattle, WA

[2]University of Washington, Seattle, WA

[3]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ

[4]NVIDIA Corporation, Santa Clara, CA

[5]University of Otago, Dunedin, NZ

Contents of this file

Introduction

We describe the validation of the Python port of the Fortran RRTMG radiation scheme (S1), and the vertically-resolved shortwave and longwave cloud radiative effect heating rates (S2).

Text S1: Diagnostic radiation scheme validation

We show that the Python port of the Fortran RRTMG radiation scheme validates against Fortran outputs for the same model state variables and ancillary radiation inputs. First, we show that the clear-sky (i.e., neglecting the effects of clouds) radiative flux diagnostic outputs at the surface and the top of atmosphere are very nearly the same between the Python and Fortran codes (Fig. S1). For both shortwave and longwave, the differences between them are typically less than 0.1 W/m^2, with global-mean biases an order of magnitude smaller than that. Some discrepancies are seen in shortwave at the edge of the solar day, and in longwave possibly related to handling of aerosol effects, but both are small.

For total-sky diagnostic fluxes, the differences are shown in Fig. S2. The stochastic nature of MCICA in RRTMG in Fortran was ported to Python, but not in a reproducible way, such that individual instances of the scheme will produce difference realizations in each code. Therefore, we see that in cloudy regions, there are grid-cell level differences in shortwave and longwave radiation, which appear spatially similar to white noise. The stochastic differences for a given cloudy grid cell may be 20 W/m^2 or more. However, the global area-weighted mean differences are very small, much less than 1 W/m^2. To further illustrate that the differences between the Python and Fortran radiative flux diagnostics are due to their stochastic nature, we show time-averaged results in Fig. S3. We expect that if the Python port correctly reproduced the stochasticity of the MCICA scheme, which draws an independent sample of the overlap profile for each grid cell and radiation timestep, then the differences should disappear with sufficient time averaging, and this is what we observe. The global area-weighted means of the bias between Python and

Fortran over time is on the order of 0.01 W/m^2. Thus, we conclude that the Python port of the Fortran RRTMG scheme is sufficiently robust to capture the effects of clouds on radiative fluxes at the resolution we desire.

Text S2: Evaluation of vertically-resolved radiative fluxes

In addition to the two-dimensional maps of radiative fluxes at the surface and top of atmosphere, we are interested in how the coarsened-fine grid and ML clouds reproduce the vertically-resolved shortwave and longwave fluxes through the atmosphere. This is a harder target for the ML cloud field to match, as here we evaluate the heating rate at each point vertically, rather than its column-integrated effect.

As a diagnostic for vertically-resolved radiative flux, we examine the cloud radiative effect (CRE) heating rates due to shortwave and longwave radiation, i.e., the total-sky minus the clear-sky radiative heating rate. (The difference between the coarsened-fine and coarse model clear-sky heating rates is not large.)

We compare the CRE heating rates as directly coarsened from the fine-grid reference simulation with those from the coarse-grid model with different cloud configurations. These cloud configurations include the coarse nudged model's own cloud fields, the prescribed coarsened-fine cloud, and the raw and thresholded ML cloud.

For shortwave CRE in the fine-grid model there is broadly heating in the upper atmosphere and cooling underneath, particularly in the summer (Northern) hemisphere (Fig. S4a). The error of the shortwave CRE heating rates in the coarse-nudged baseline is substantial; there is too little heating aloft and too little cooling below, with typical error magnitudes in the coarse nudged baseline of about 70% of the heating rate magnitudes in the fine-grid model (Fig. S4b). The coarsened-fine clouds, in contrast, produce short-

wave CRE heating rates with error magnitudes that are only about 20% of the typical coarsened-fine heating rate magnitudes (Fig S4c). The ML cloud fields are able to capture about half of this improvement over the coarse nudged baseline, with little difference between the thresholded and raw ML cloud fields in terms of CRE errors.

For longwave CRE, the coarsened-fine heating rates broadly show cooling aloft and heating near the surface and underneath areas of tropical deep convection (Fig. S5a). The coarsened-fine clouds do produce longwave heating rates with smaller error magnitudes than the baseline coarse nudged clouds (Fig. S5b, c), but the improvement is less complete than with shortwave heating rates. The ML clouds have approximately similar error magnitudes to the coarse nudged baseline, though the bias pattern is different. In particular the ML clouds produce excessive longwave heating near the tropical tropopause and near the surface in the polar regions (Fig. S5d, e). For the ML cloud parameterization to be used to generate vertically-resolved CRE heating rates online and produce improved forecasts, i.e., to replace the existing cloud parameterization for this purpose, improved ML skill in these regions may be needed.

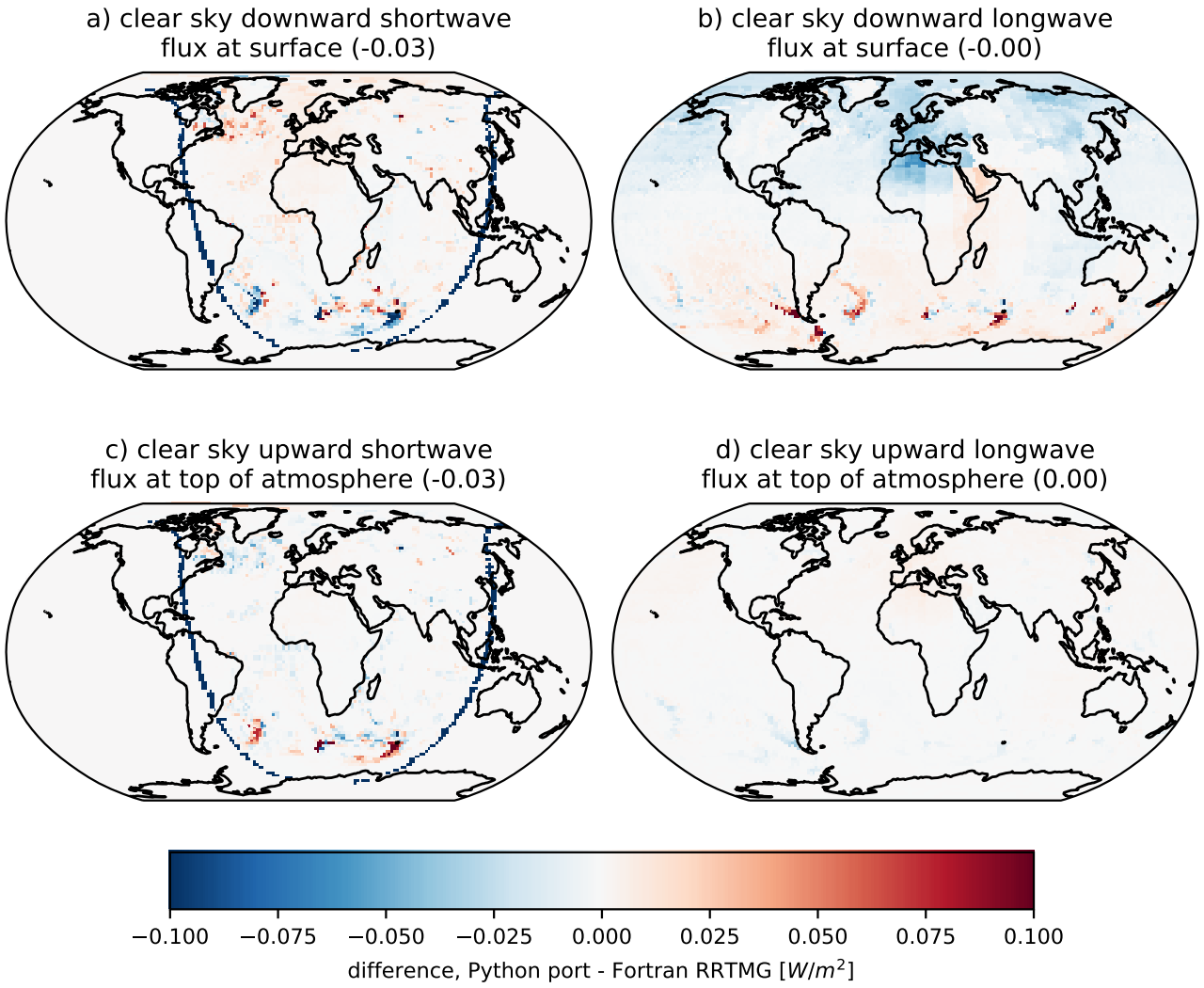

Figure S1. Clear-sky radiation diagnostic differences between the Python port of Fortran RRTMG and the original code, for a single snapshot time. Global area-weighted mean values are shown in parentheses.

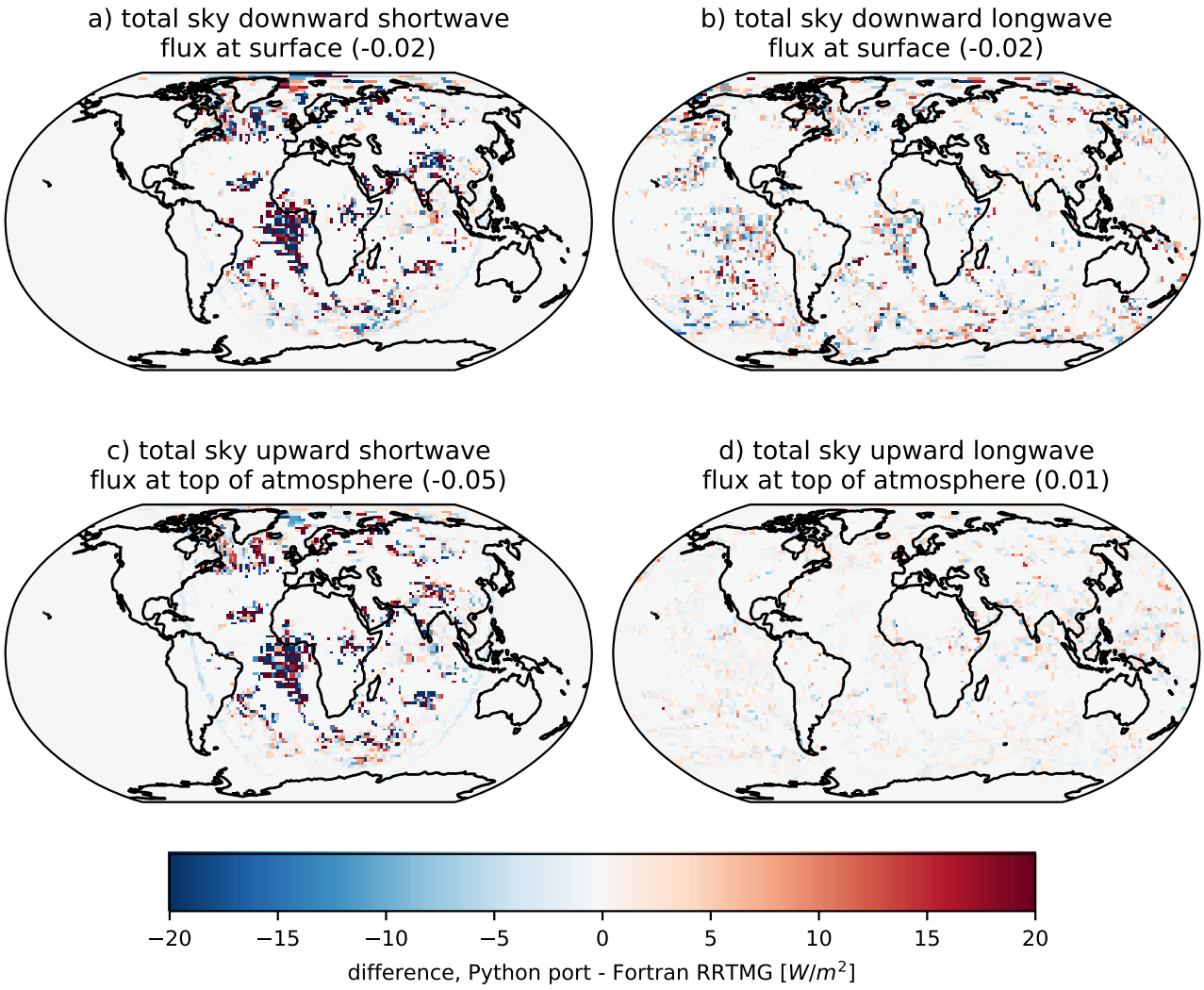

Figure S2. Total-sky radiation diagnostic differences between the Python port of Fortran RRTMG and the original code, for the same single snapshot time. Global area-weighted mean values are shown in parentheses.

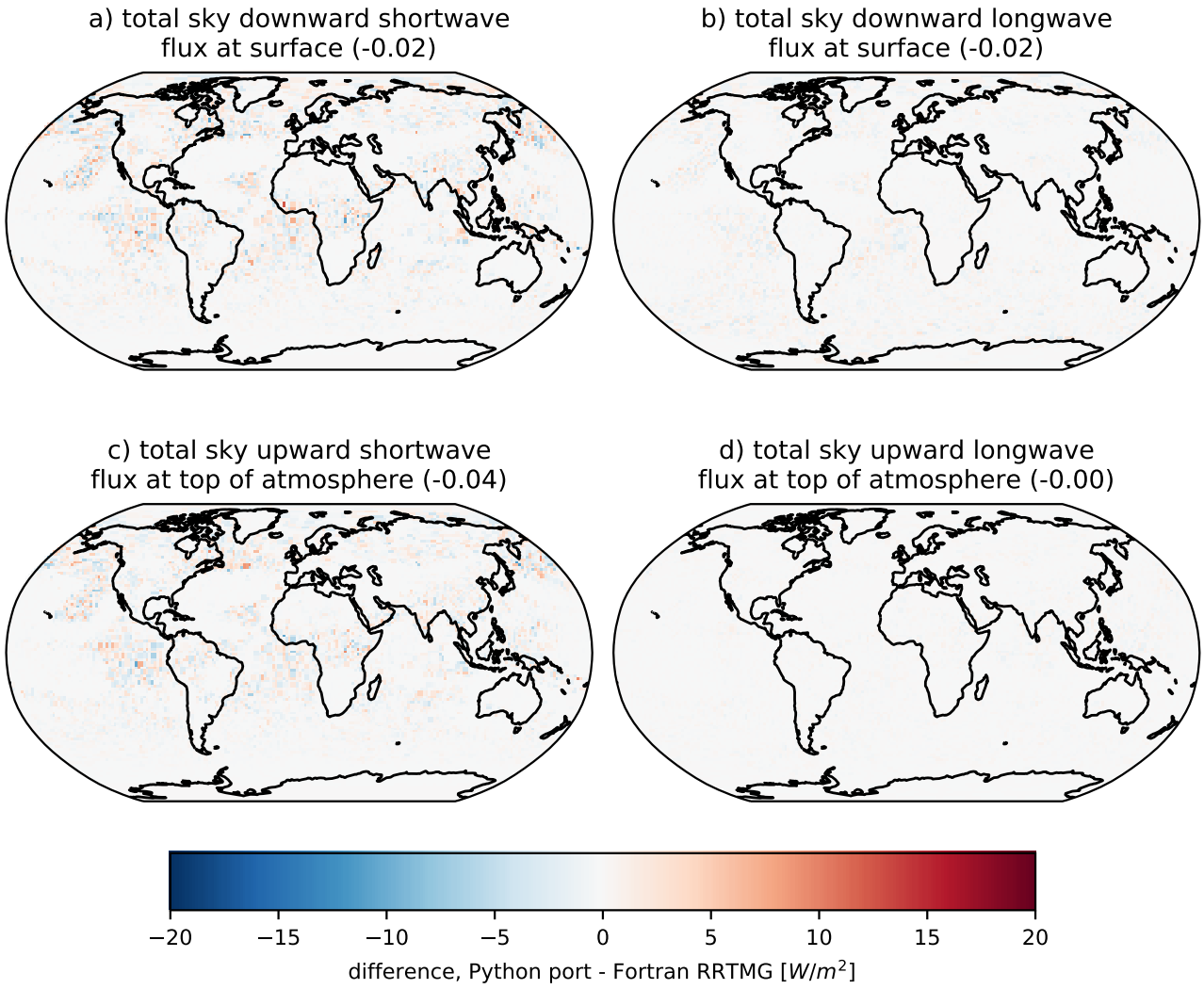

Figure S3. Total-sky radiation diagnostic differences between the Python port of Fortran RRTMG and the original code, averaged from hourly outputs over three days (72 snapshots). Global area-weighted mean values are shown in parentheses.

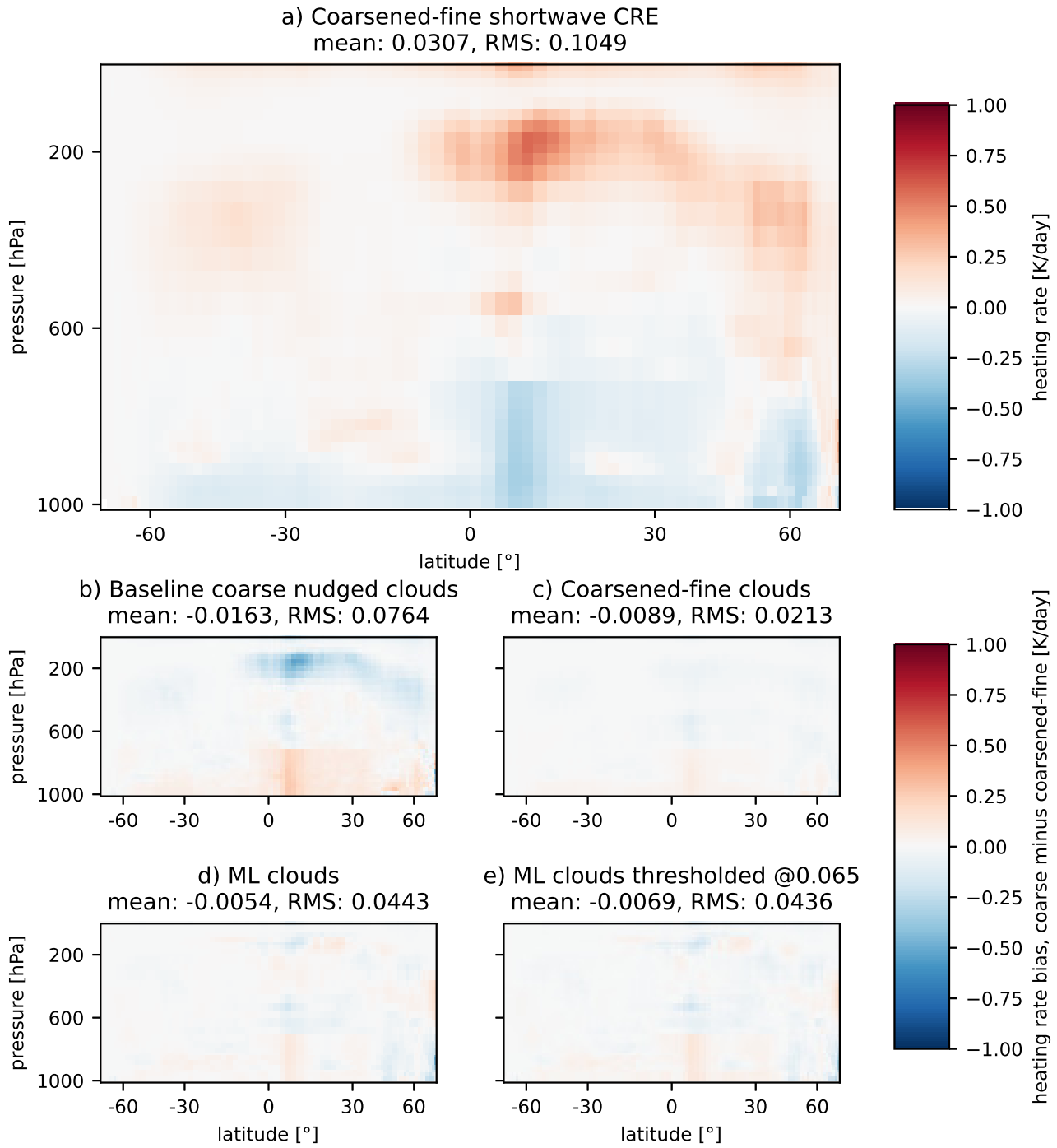

Figure S4. Vertically-resolved cloud radiative effect shortwave heating rates, shown as zonal- and time-averages. The time averaging is over every hour in the three days of the ML validation period. The mean and root-mean-squared metrics are mass and latitude-weighted.

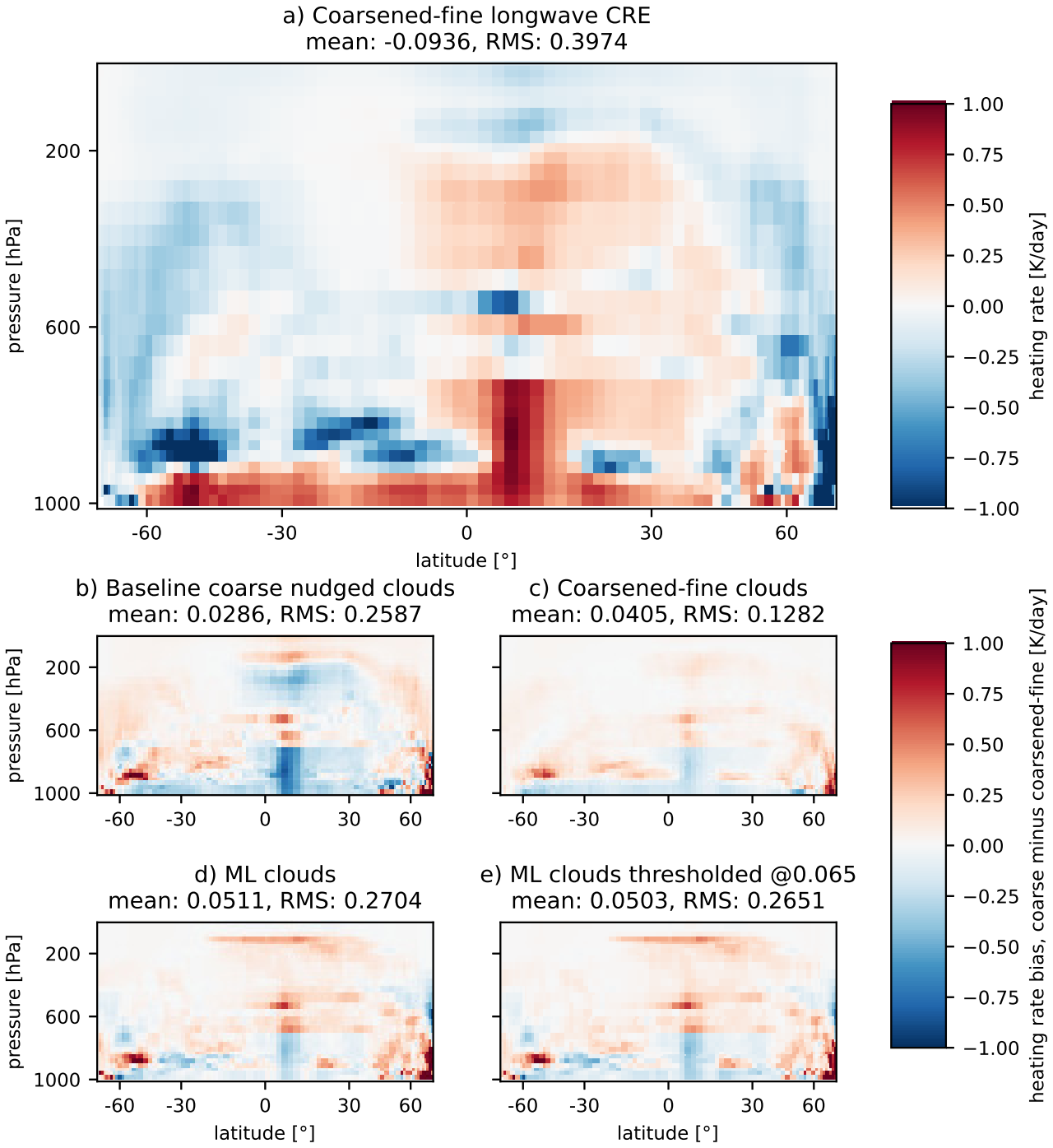

Figure S5. As in previous figure, but for longwave heating rates.