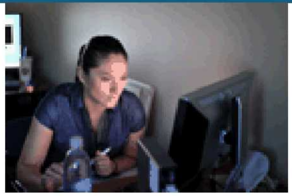
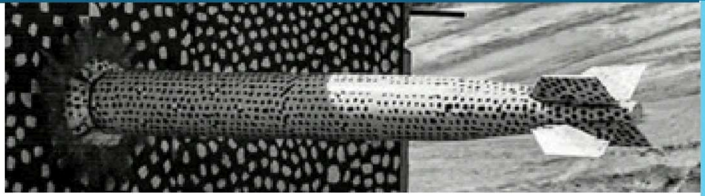
SAND2019-10549PE

Reducing the cost of radiative transfer in multi-scale atmosphere models





Benjamin R. Hillman

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Two paths toward a cloud-resolving E3SM

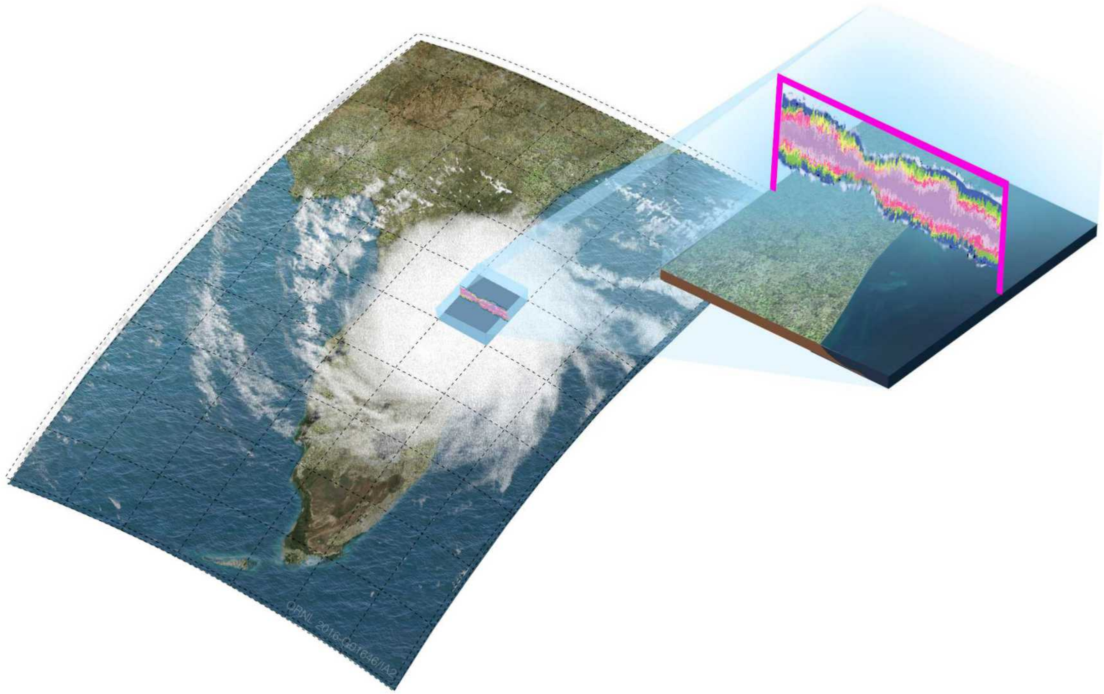
E3SM with the Multi-scale Modeling Framework (E3SM-MMF)

- Explicitly resolve the large-scale and cloud-scale dynamics *separately*
- Embedded cloud-resolving model within each physics column
- Capture *some aspects* of cloud-resolving simulation, at lower computational cost (climate-scale)

Simplified Cloud-Resolving E3SM Atmosphere Model (SCREAM)

- Push E3SM horizontal grid to cloud-resolving resolutions
- Simplified physics: P3 micro, Simplified Higher-Order Closure (SHOC), no deep convection
- *NOT A CLIMATE MODEL!*

Multi-scale Modeling Framework



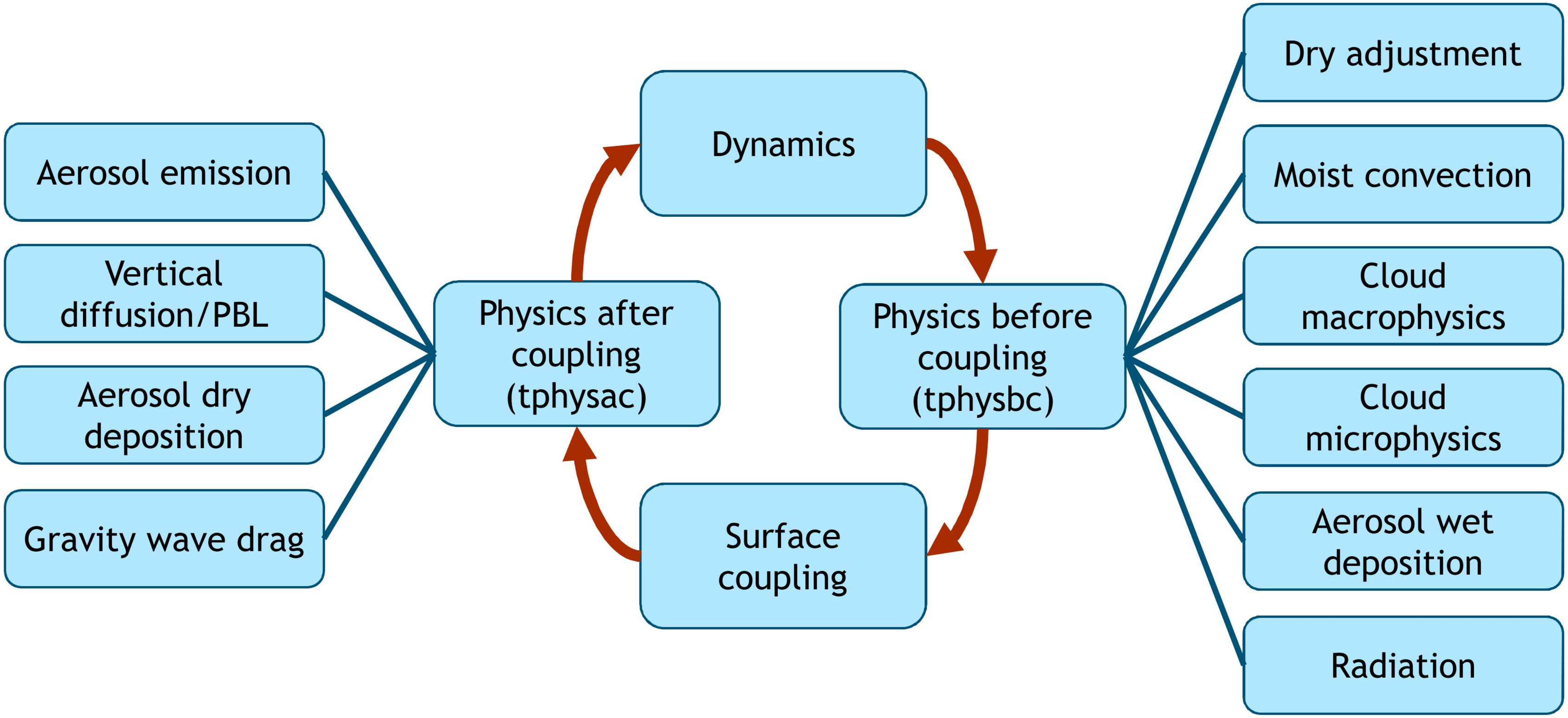
Traditional parameterizations introduce *structural uncertainty* in current global climate models

Pushing to higher resolutions allows us to drop more physical parameterizations, but this is *expensive*

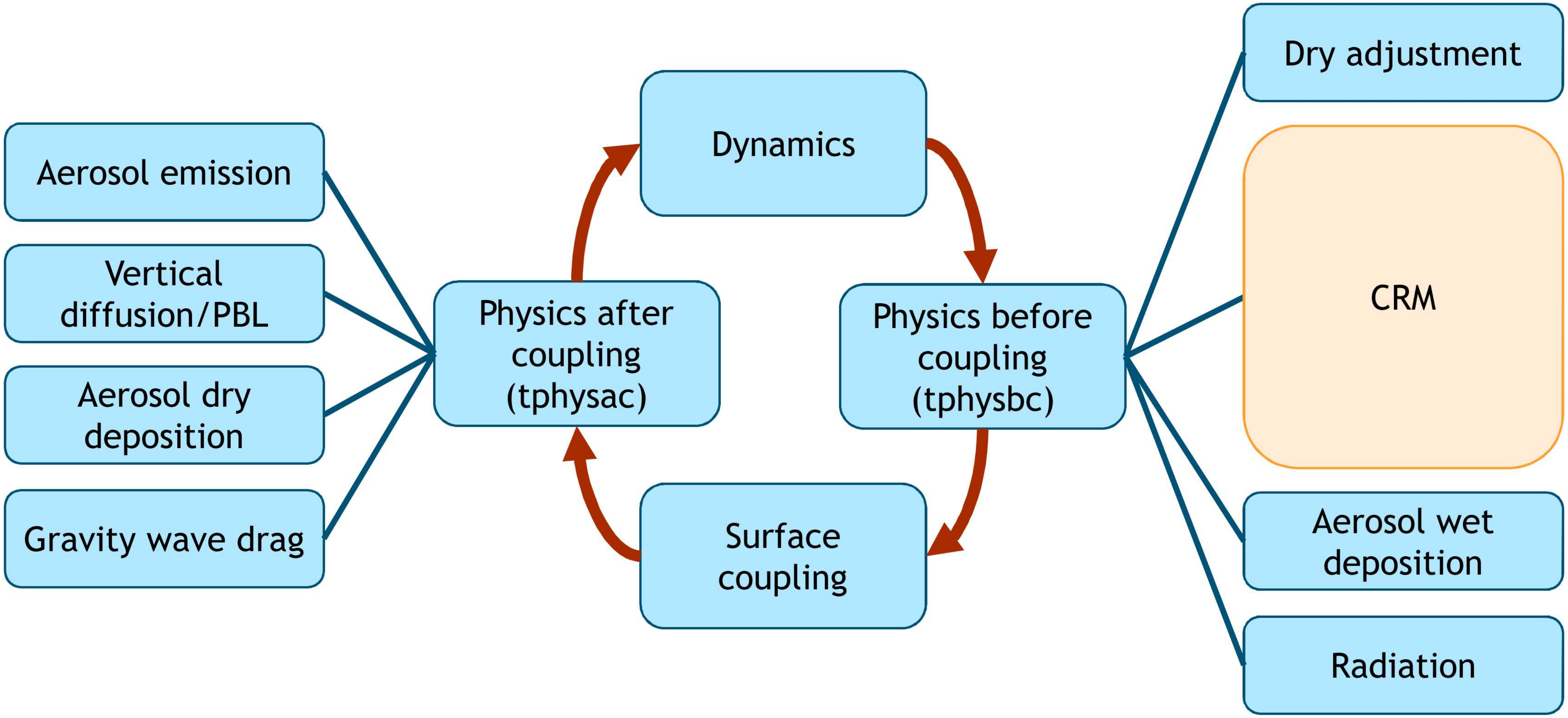
Replace cloud and convective parameterizations in a traditional global climate model with embedded cloud resolving models in each column

Exascale computers + MMF will make it possible to perform traditional climate simulations with *some aspects* of cloud resolving simulations

E3SM Atmosphere Model schematic



E3SM-MMF Atmosphere Model schematic



E3SM-MMF model specifics

“Host” GCM: E3SM Atmosphere Model

- HOMME spectral element dynamical core
- v1 physics, minus convection and cloud macro and micro physics

Embedded CRM: System for Atmospheric Modeling

- Currently using single-moment microphysics scheme
- Prescribed aerosol

Target throughput: 5 simulated years per day

Computational speed-ups

GPU port of CRM code (Matt Norman)

Mean State Acceleration (Chris Jones)

Reduced radiation resolution/frequency (Ben Hillman and Walter Hannah)

GPU port of CRM code

Entirety of the time-stepping loop within the CRM ported using OpenACC directives-based approach

CRM code refactored to include *ncol* dimension from global model to expose more parallelism; do multiple CRMs at once

15-16x speed-up on summit (two P9s vs six Voltas per node)

Benchmark on Summit for Gordon Bell submission used 4,600 nodes of Summit, achieved 2.5% peak double precision flop/s, throughput of 1.8 SYPD (with Mean State Acceleration and reduced radiation)

Mean State Acceleration (MSA)

Reduce number of timesteps required for CRM integration by artificially introducing a "mean-state" tendency

Push the CRM faster towards mean state

Rationale: turbulent eddies spin-up fast relative to evolution of mean state

Jones, Bretherton, and Pritchard (*JAMES*, 2015, https://doi.org/10.1002/2015MS000488)

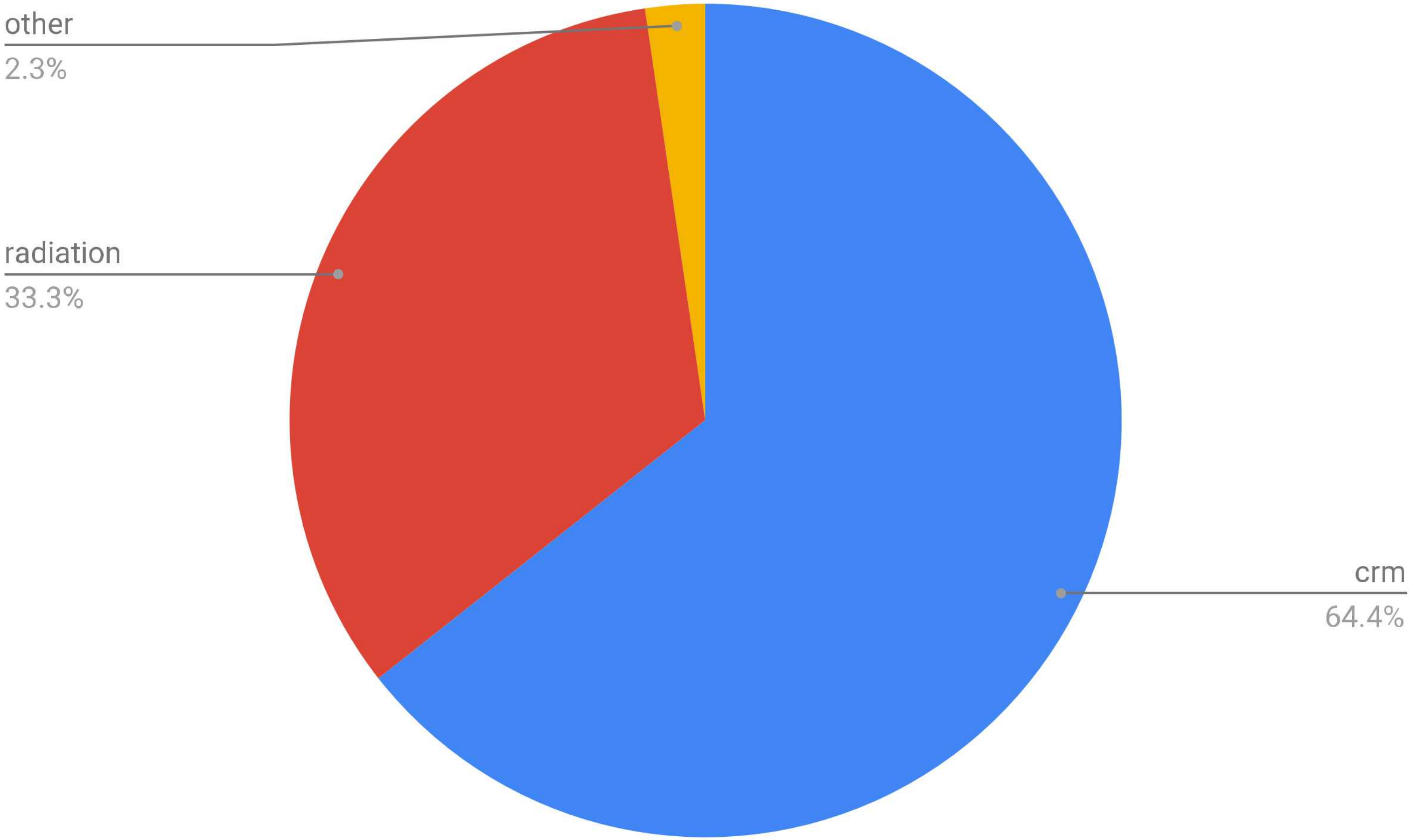
Currently an acceleration factor of about 4 appears to be stable

Radiation cost

Radiation is expensive!

Reduce computational cost of radiation by reducing frequency and number of columns

Balance between efficiency and accuracy



Relative cost of physics packages on Intel Sandy Bridge (chama)

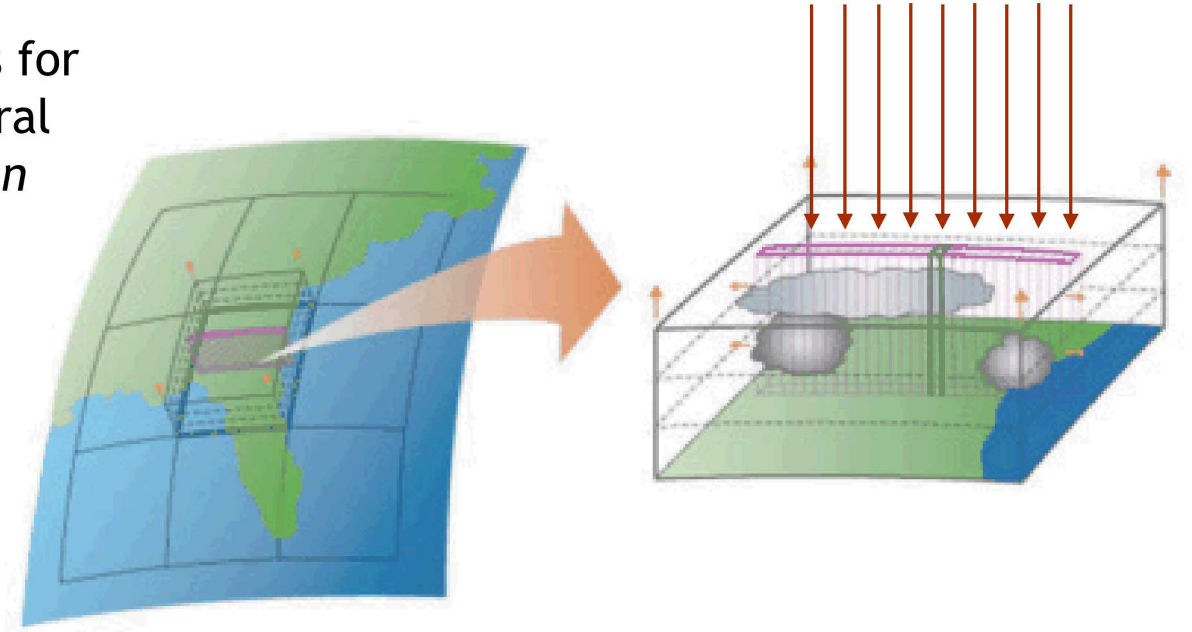
Why is radiation so expensive

Fluxes and heating rates: integrated quantities over many spectral intervals

Lots of columns, lots of bands…lots of calculations

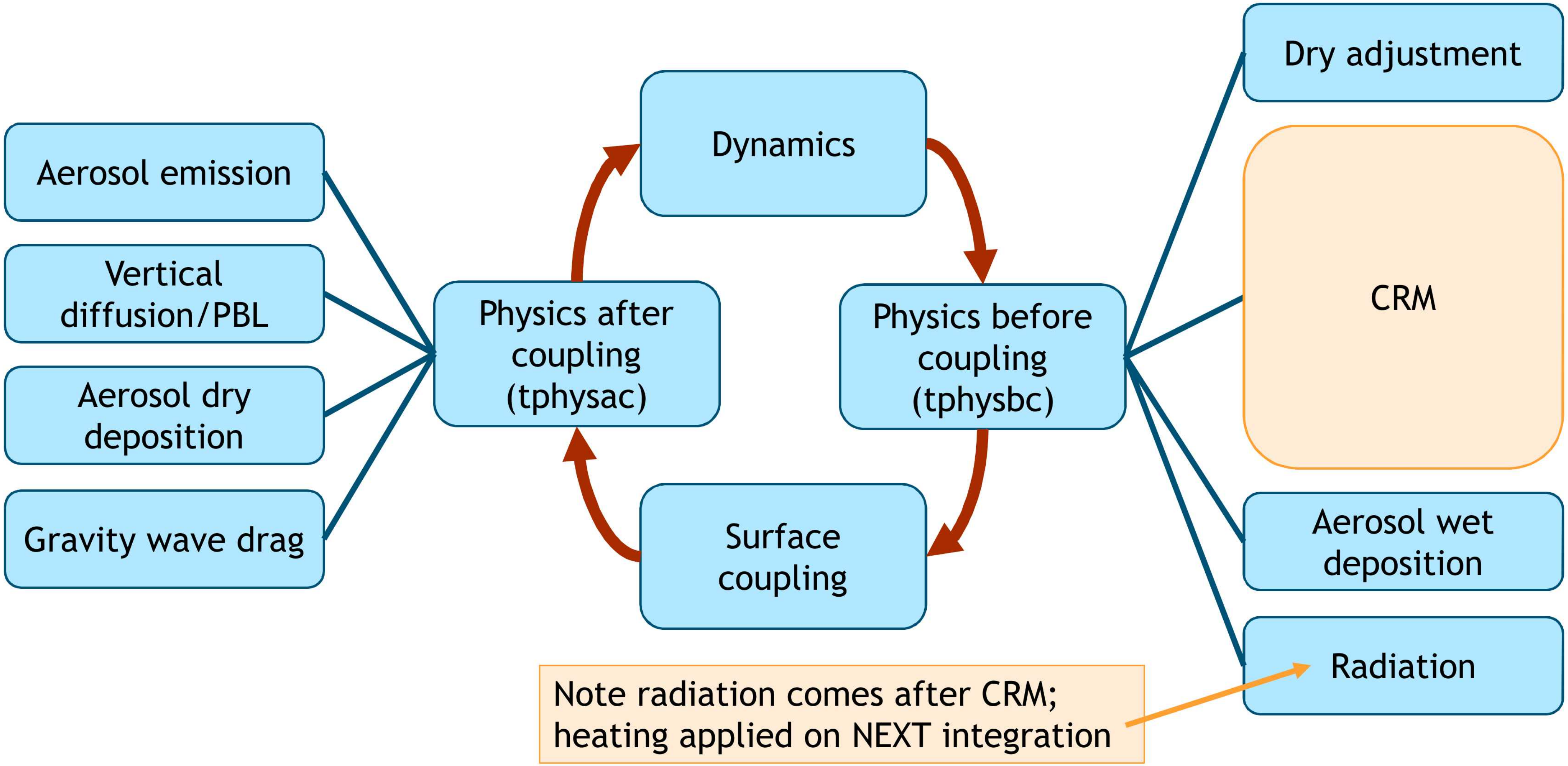
Exacerbated in models using the MMF: need to calculate fluxes and heating rates on *each CRM column*

Need to calculate fluxes for a large number of spectral intervals for *each column*



Now need to calculate fluxes for a large number of spectral intervals for each *CRM column* within each GCM column!

E3SM-MMF Atmosphere Model schematic



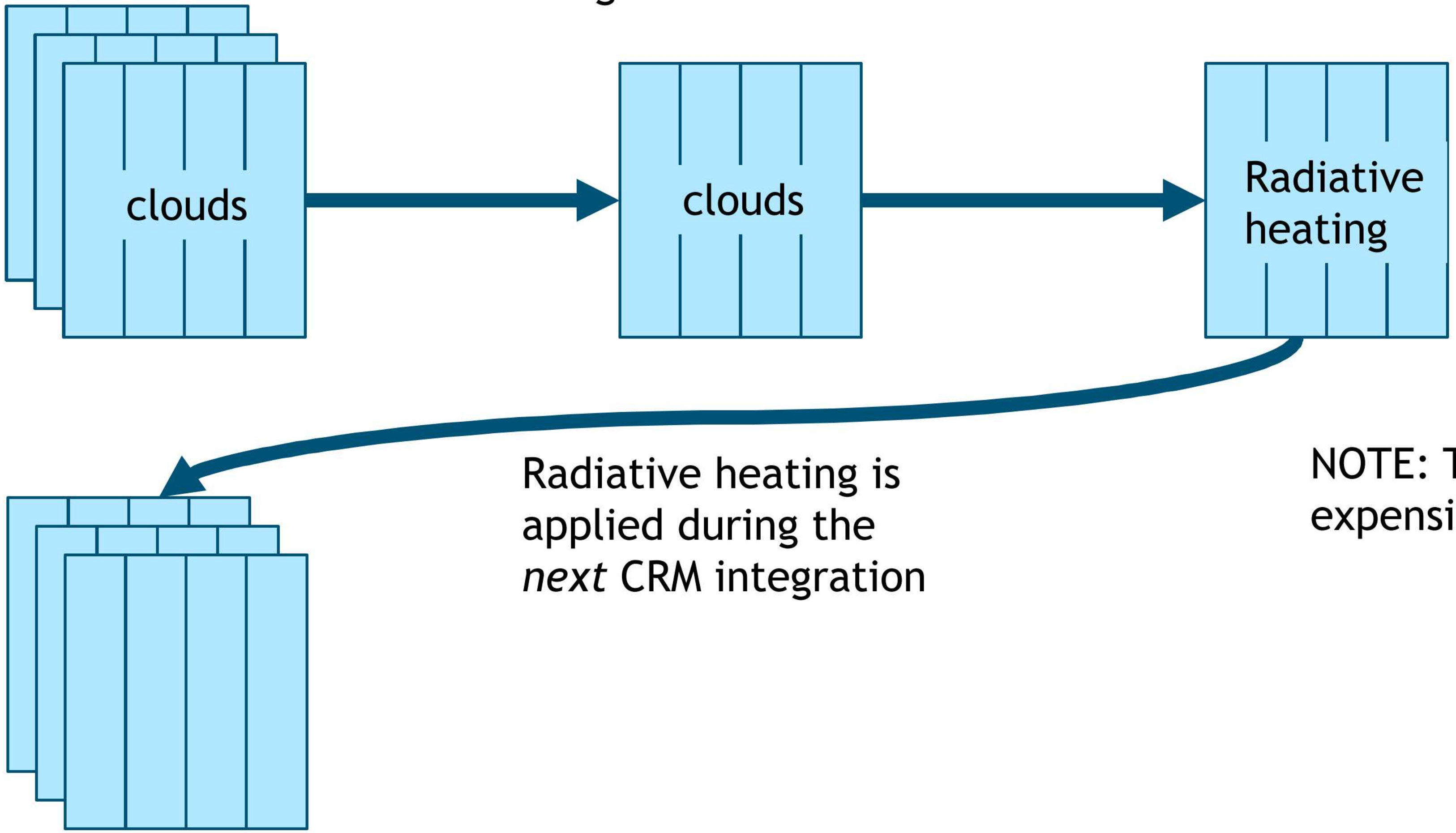
Radiative coupling in E3SM-MMF

Cloud properties within CRM at t_crm = 1, … , dt_gcm

Cloud properties *time-averaged* within the CRM, then passed back to global model

Radiative heating calculated from *time-averaged* profiles, but at each CRM column

t_gcm = n

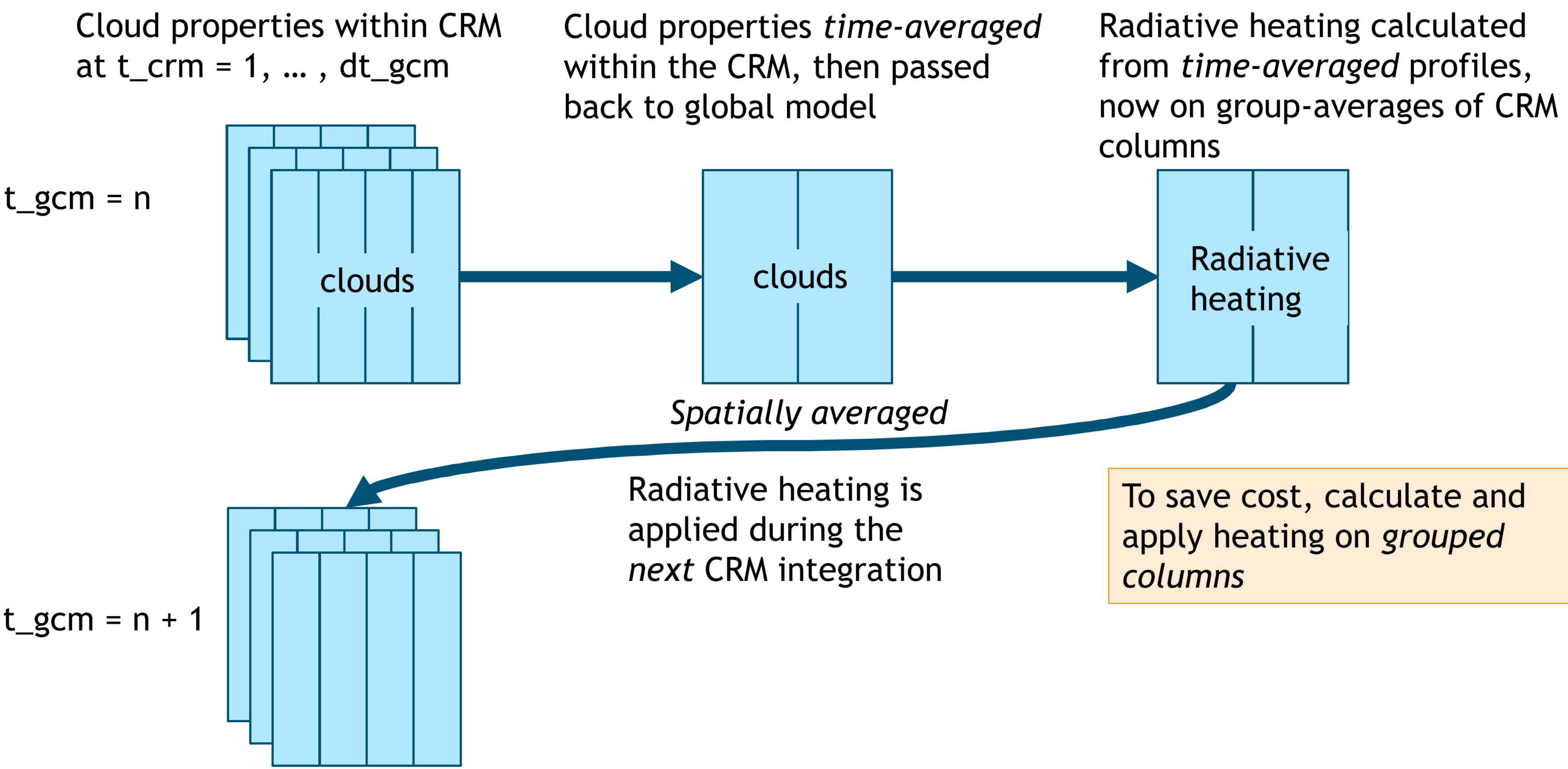


Radiative heating is applied during the *next* CRM integration

NOTE: This step is expensive!

t_gcm = n + 1

Radiative coupling in E3SM-MMF (reduced radiation)



To save cost, calculate and apply heating on *grouped columns*

Minimizing biases resulting from using less columns

Using spatially-averaged cloud properties removes cloud-scale features

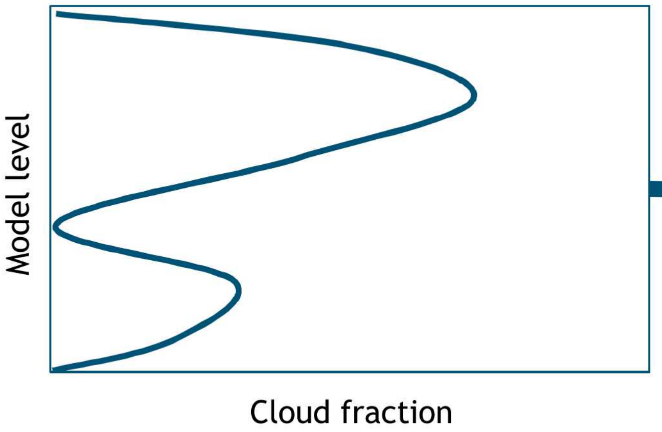
Individual columns no longer just “clear” or “cloudy” but have *partial cloud fraction C: [0, 1]*

This is accounted for in the GCM through subcolumn sampling in the “MCICA” approach: reconstruct psuedo-cloudy/clear elements from domain averages with stochastic subcolumn sampling

Enabling for the CRM adds no cost (already being done internally but with trivial inputs)

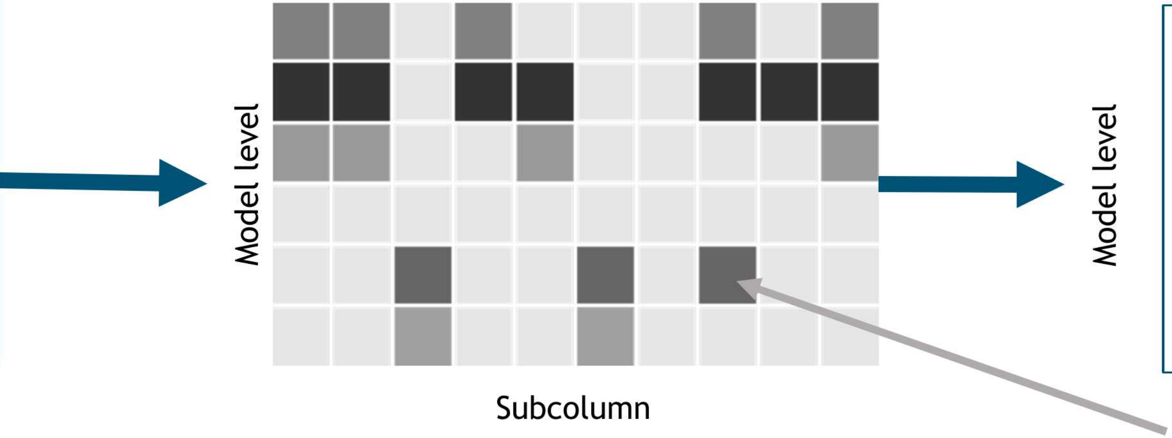
Subcolumn sampling

Start with column-mean profiles of cloud fraction (and condensate amount



(For GCM, this is a single gridbox average; for MMF, this is a single time-averaged CRM column)

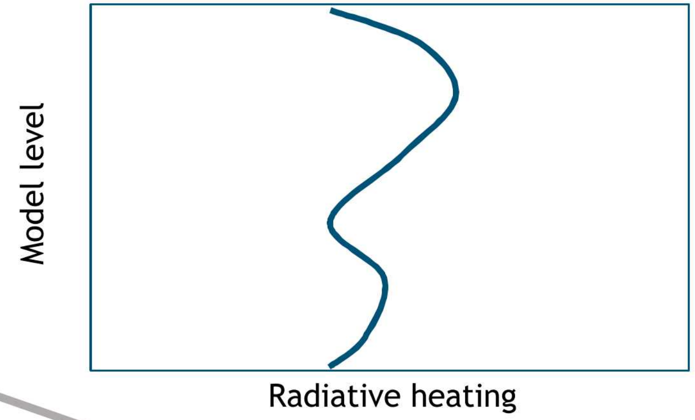
Stochastic subcolumn sampling with an overlap assumption to generate psuedo-resolved clear/cloudy elements



Colors: condensate amount

Note for CRM, this is subsampling *within a single CRM column*

A *single subcolumn* chosen for *each spectral band*; calculate band-resolved fluxes, return broadband heating



Experiment setup

Small, rapid test configurations (ne4 resolution, 1 month duration)

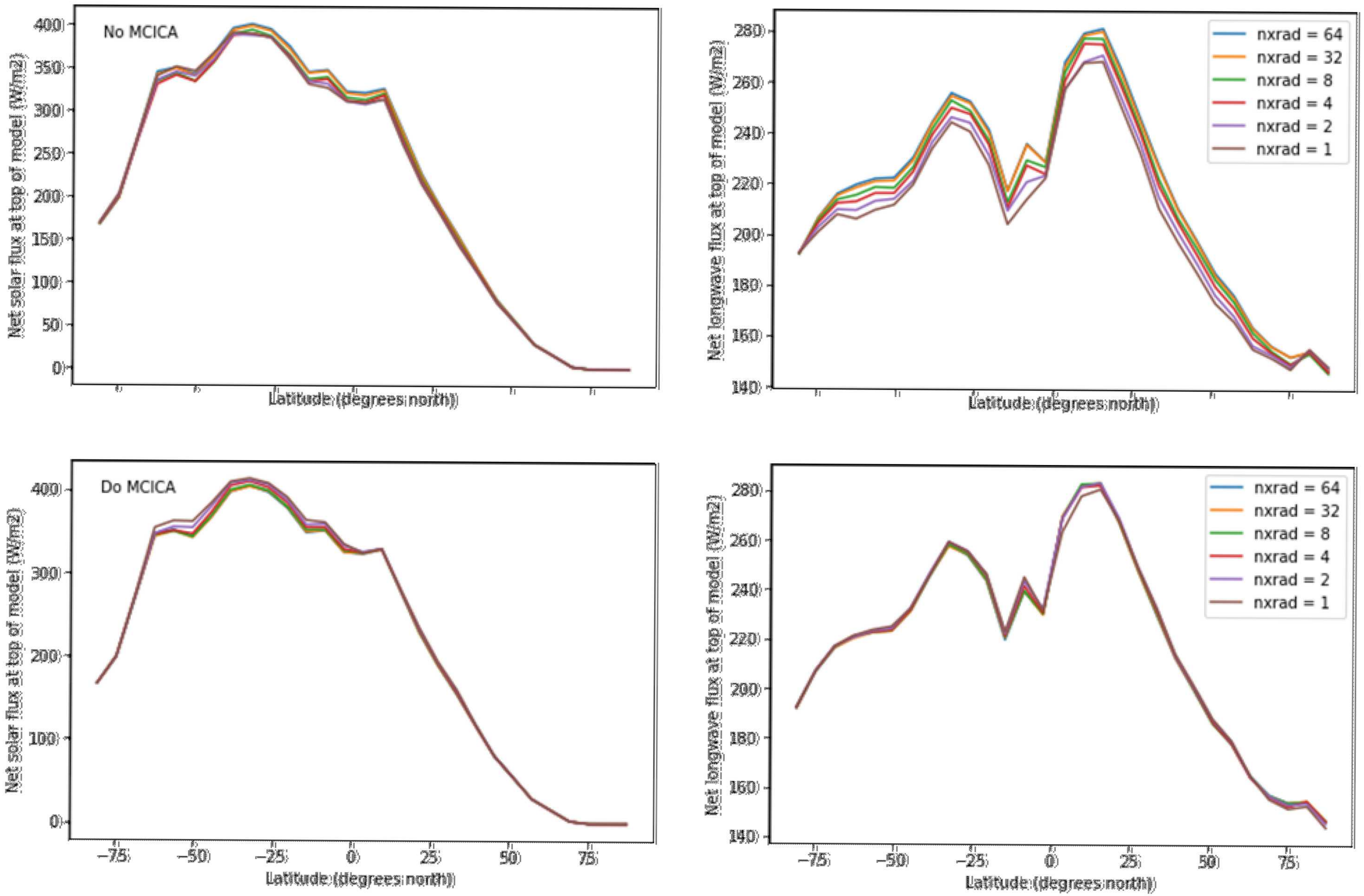
- Not a realistic resolution for climate, but sufficient for informing us about the affects of these changes since they largely affect the CRM itself, not the host model

CRM uses a 2D domain with 64 columns (oriented north-south) with a grid spacing of 1 km

Different cases using 1, 2, 4, 8, 32, and 64 columns for the radiation (i.e., group sizes of 64, 32, 16, 8, 2, and 1)

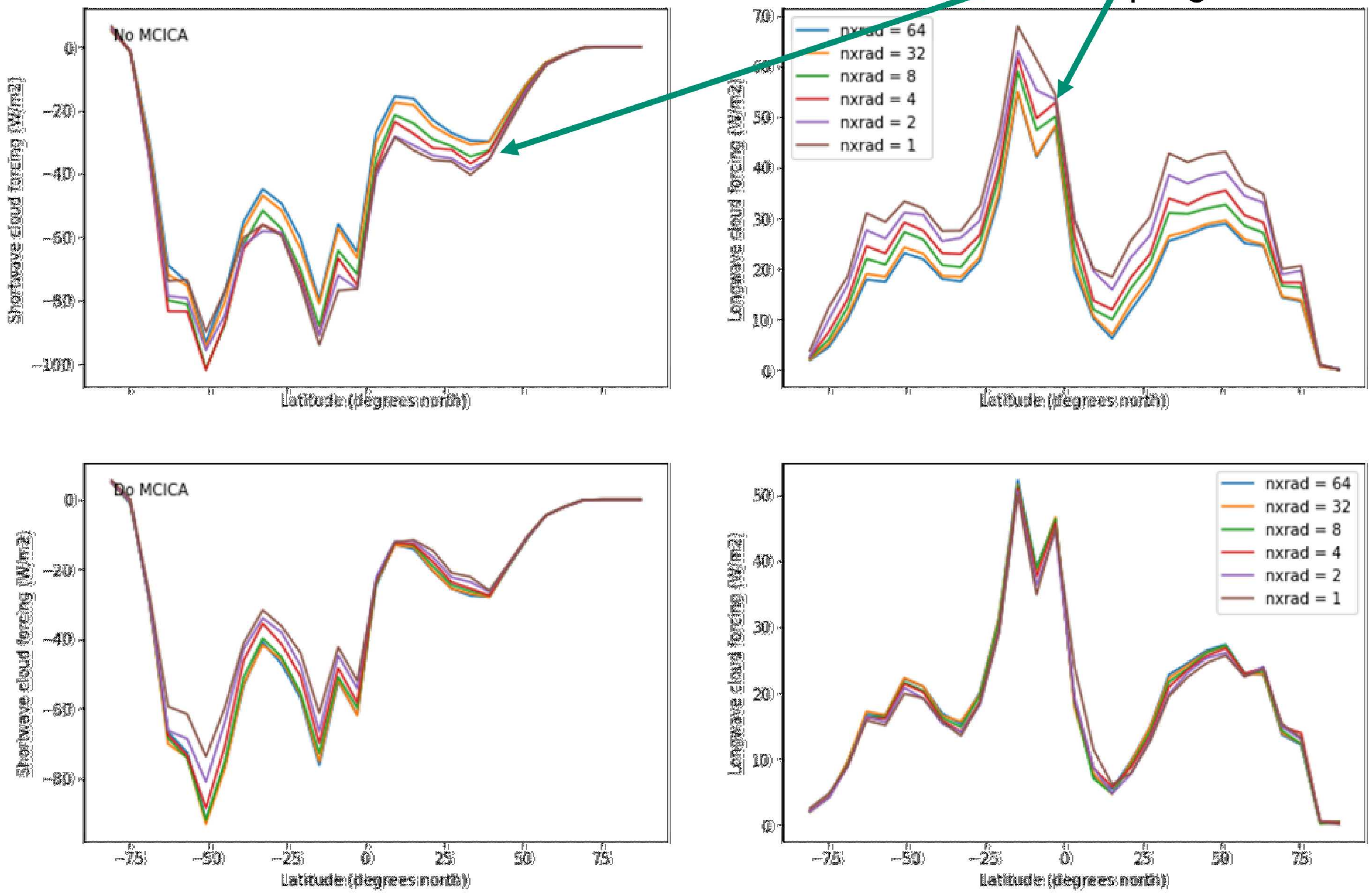
Identical tests with and without subcolumn sampling for overlap

Errors in TOA fluxes

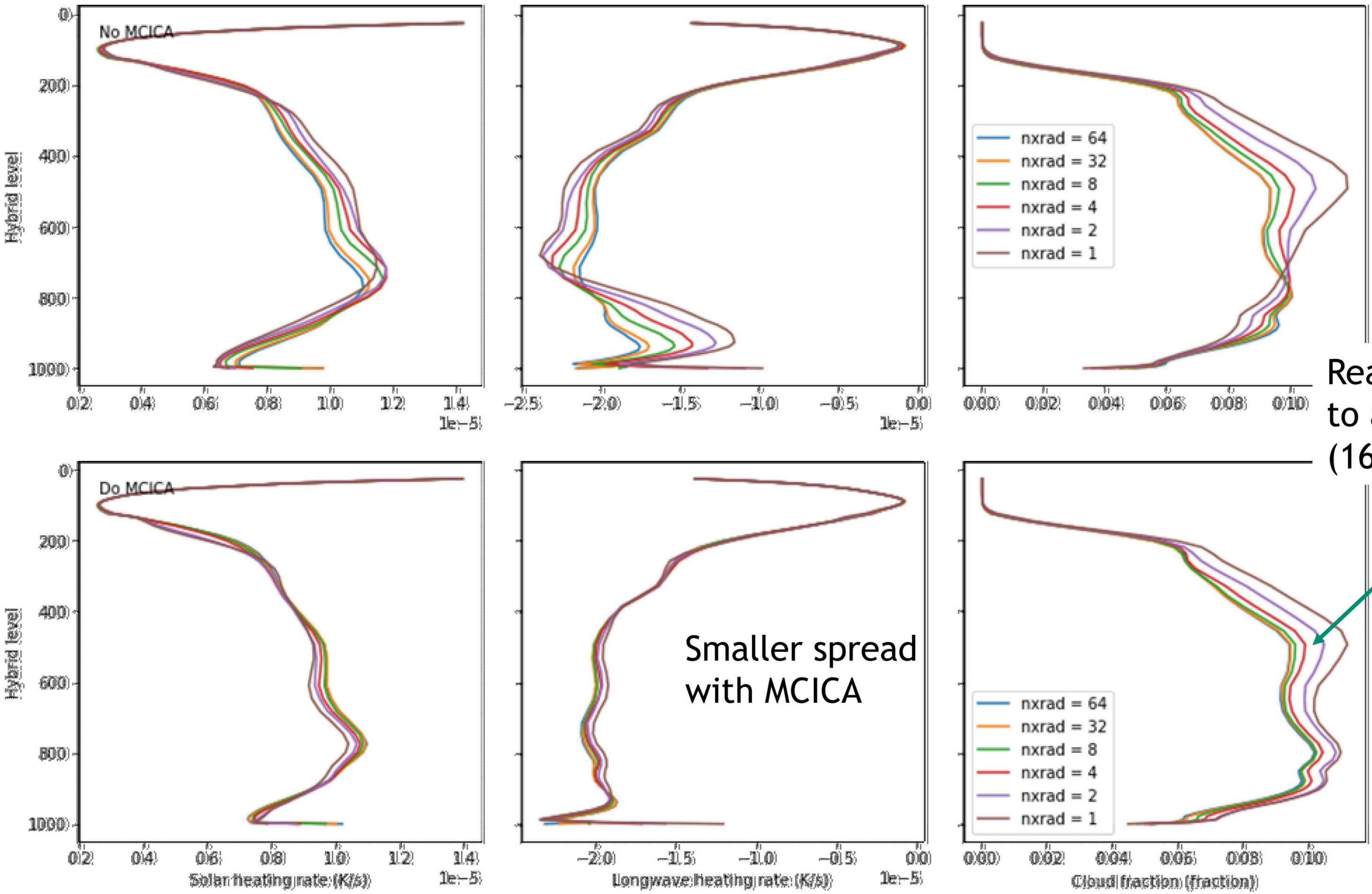


Errors in cloud radiative effects

Large sensitivity with no MCICA subcolumn sampling

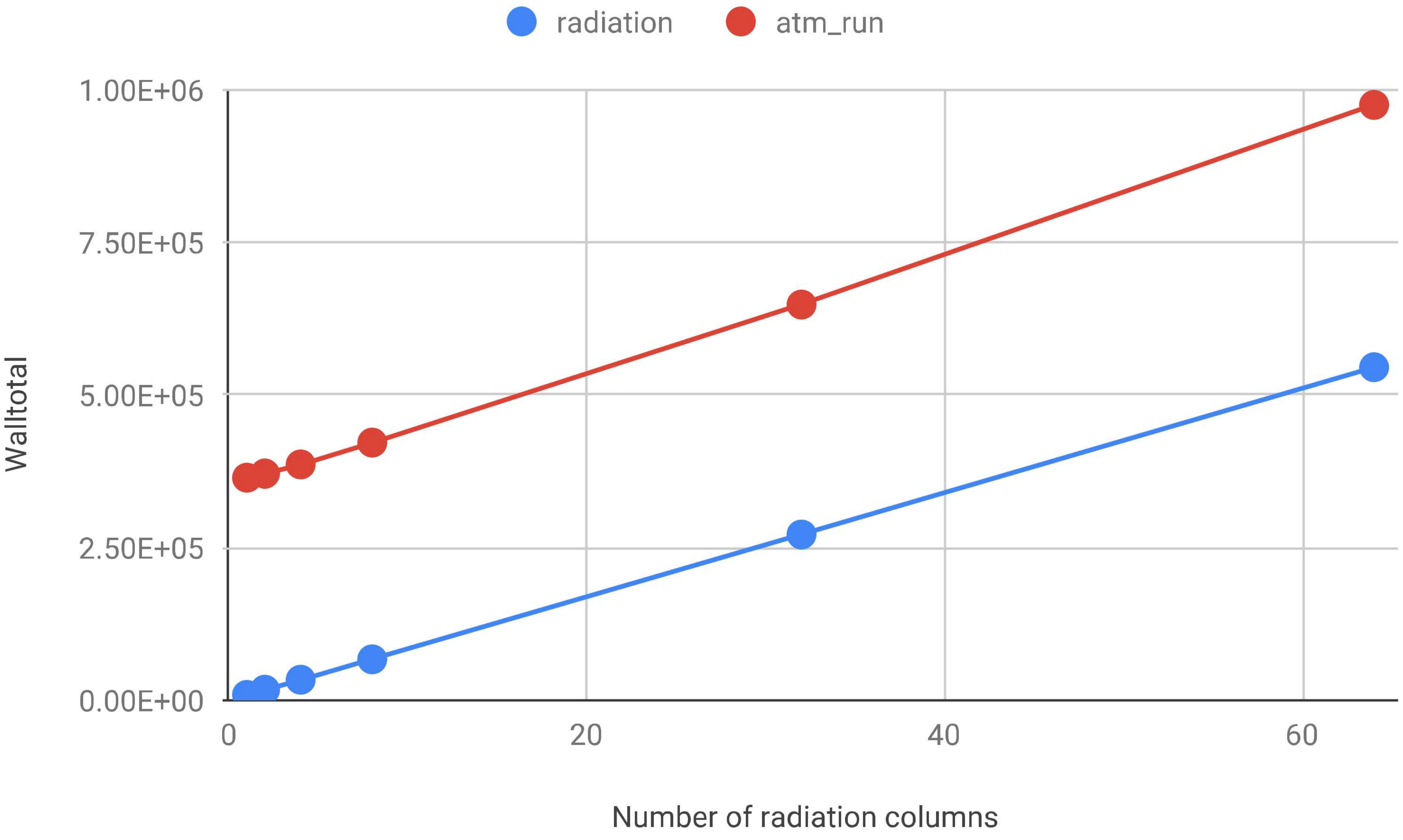


Errors in heating and cloud profiles



Reasonable down to about 4 columns (16x speed-up)

Speed-up from reduced radiation



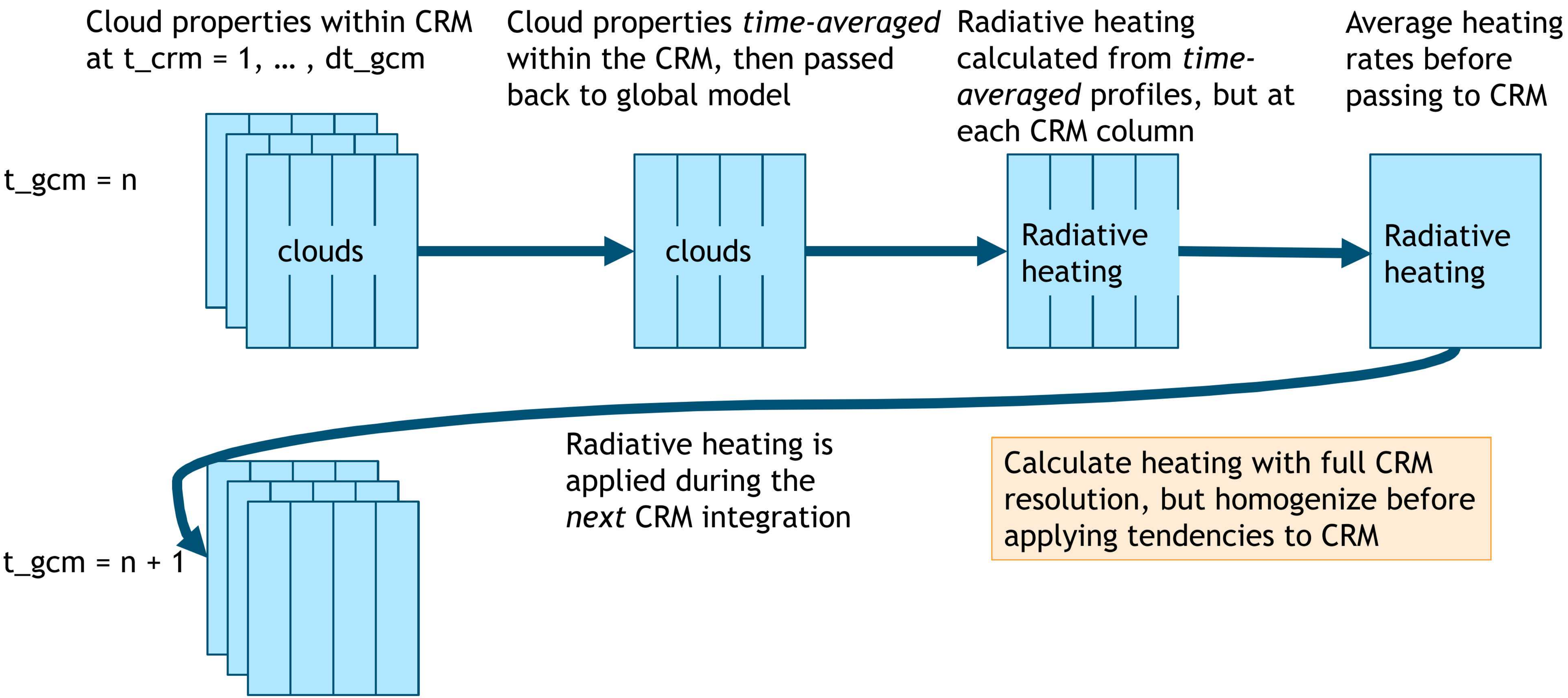
Can we just use domain averages?

It would be simple and cheap to just reduce down to *one* radiation column per CRM, and let the GCM handle radiation (even have nice ways of representing the variability)

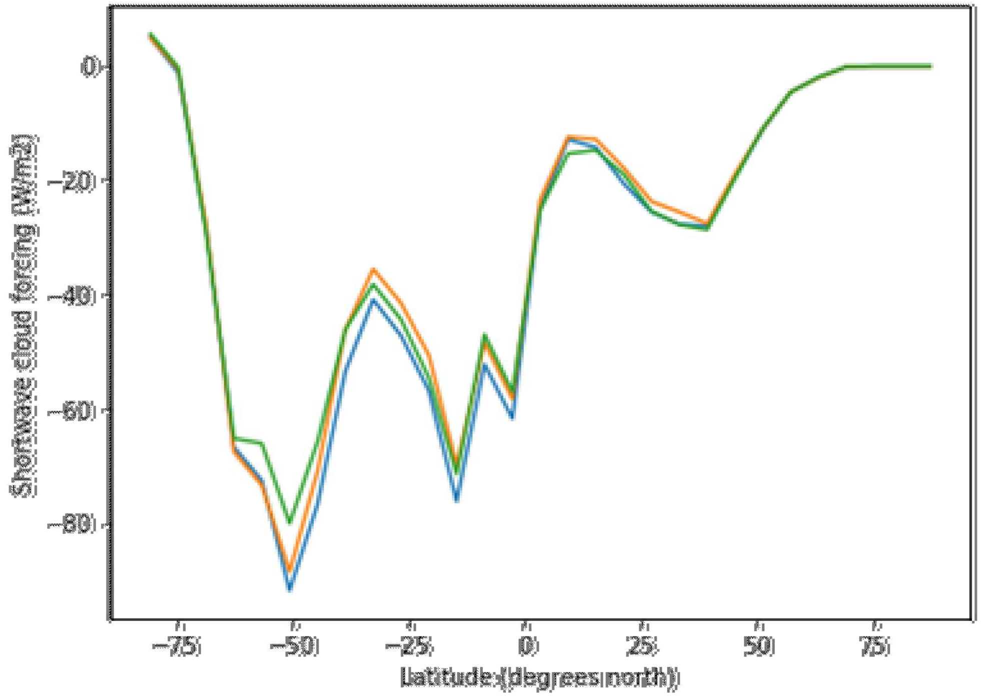
To test, calculate heating rates on full CRM resolution, but homogenize before applying them

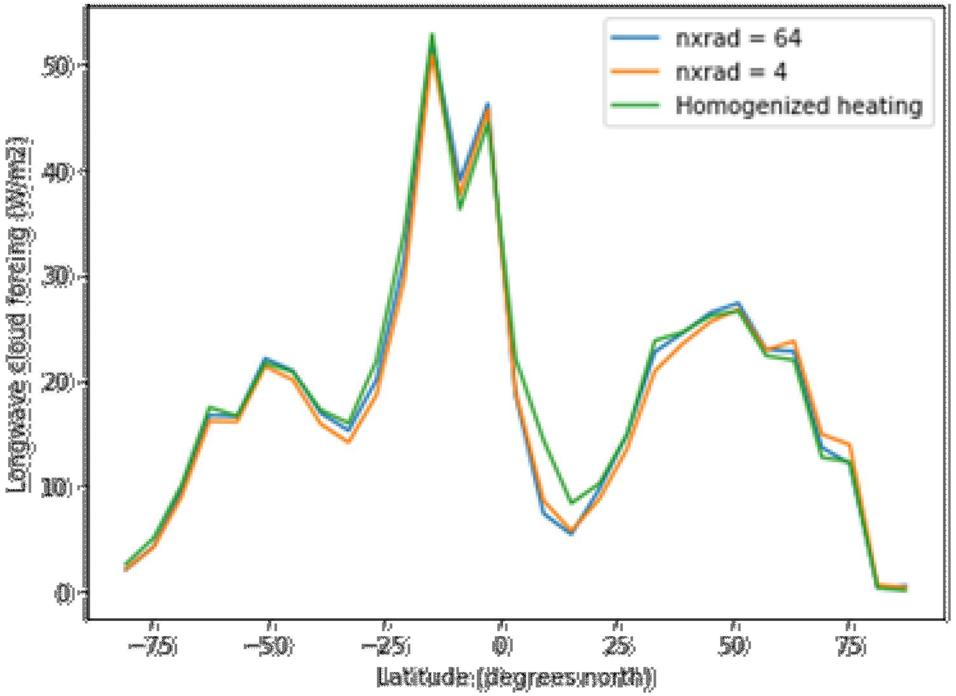
Tests the question: if we had a perfect subcolumn sampling scheme, could we get back the same solution when reducing down to one radiation column?

Homogenized heating test

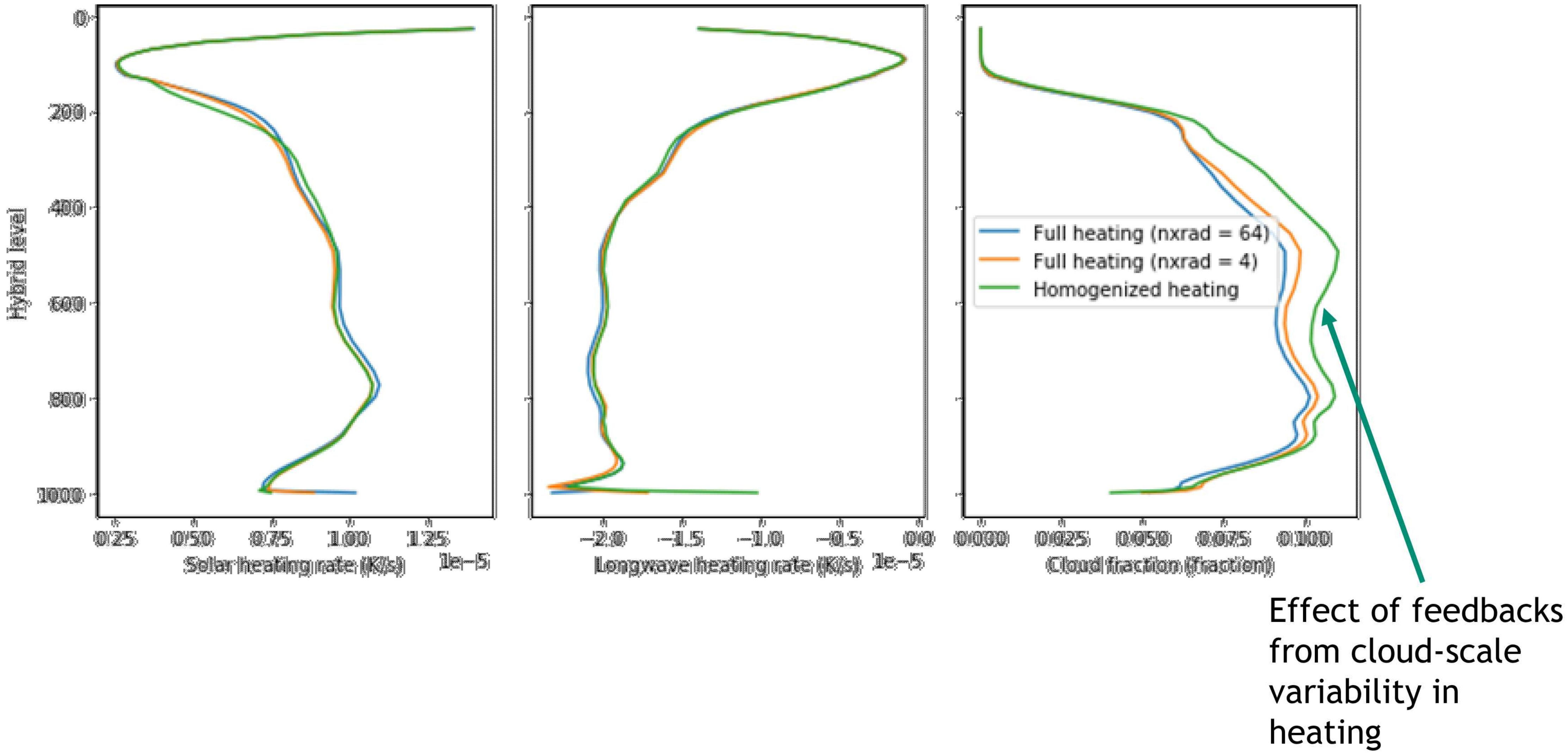


Homogenized heating test





Homogenized heating test



Summary of reducing spatial resolution

Reduced radiation alone has large impact on solution

Using subcolumn sampling reduced that impact

Can get significant speedup, with tolerable differences from baseline

Cannot reduce this all the way down to a single radiation column: lose impact of small-scale variability in heating on cloud evolution/micro-circulations

Reducing temporal frequency

How often do we need to update the heating?

Previous MMF implementations called radiation at every physics timestep

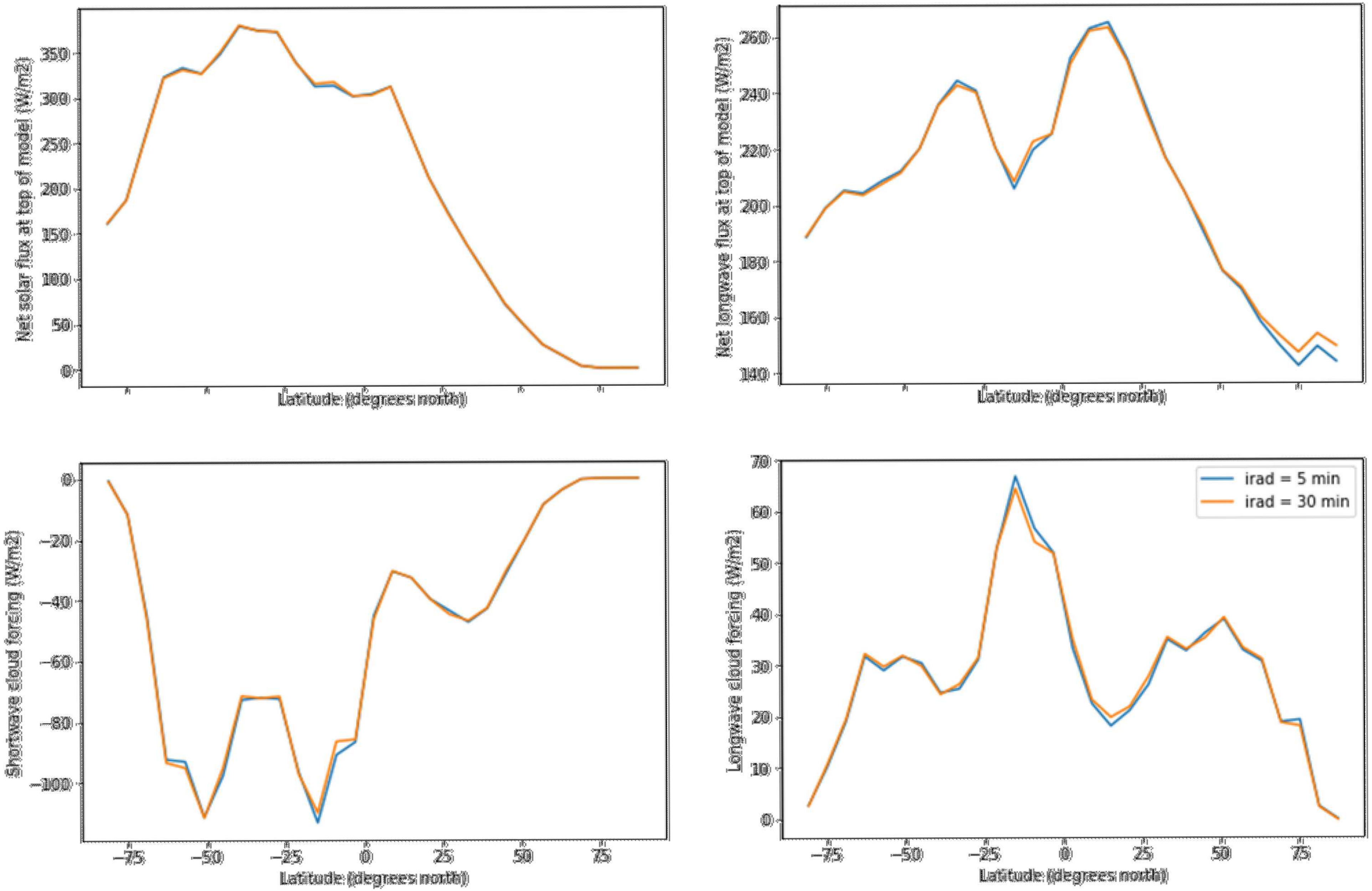
In E3SM, we usually only update the radiative heating once every 30 minutes

Can we get away with updating radiation less frequently in E3SM-MMF?

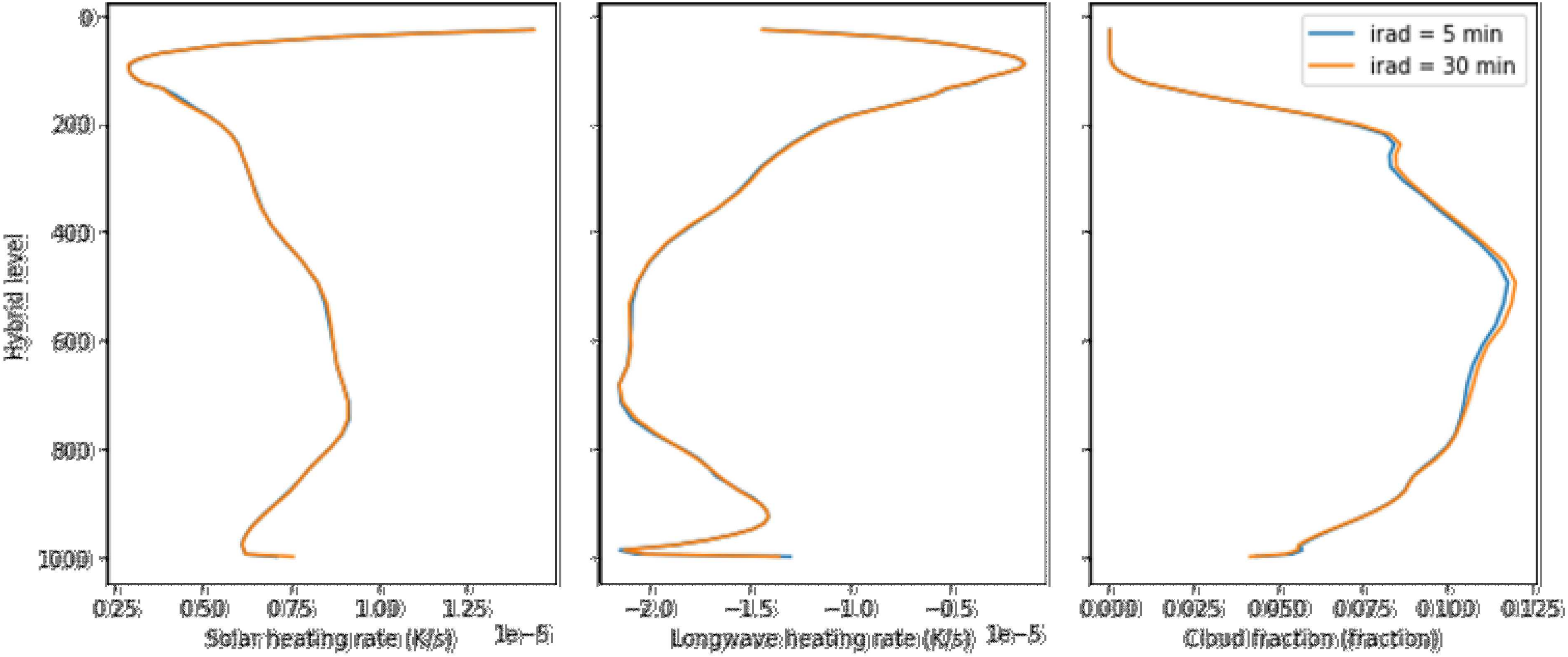
Experiment:

- Baseline: radiative heating updated every 5 minutes (every physics timestep)
- Test: radiative heating updated only every 30 minutes (every sixth timestep)
- Note that heating is still applied every CRM timestep

Biases due to reduced temporal frequency



Bias due to reduced temporal frequency



NOTE: different physics timestep
and CRM dx than previous tests

Summary of techniques for reducing cost

Reduced spatial resolution can provide significant speed-up, at cost of manageable errors in heating

Reduced temporal resolution provides additional speed-up, with little impact on simulation (although this needs further exploration)

Will this become uneccessary on the GPU?

Thank you!

Extra slides

Monte Carlo Independent Column Approximation

$$< F^{ICA} > = (1 - A_c) \sum_{k}^{K} w(\lambda_k) S(\lambda_k) F^{clr}(\lambda_k) + A_c \sum_{k}^{K} w(\lambda_k) S(\lambda_k) \sum_{j}^{J} p(s_j) F(s_j, \lambda_k)$$

Do this for every column

Sum over bands (order 10^2)

Sum over possible cloud states (subcolumns)

$$< F^{McICA} > \approx (1 - A_c) \sum_{k}^{K} w(\lambda_k) S(\lambda_k) F^{clr}(\lambda_k) + A_c \sum_{k}^{K} w(\lambda_k) S(\lambda_k) F(s_{rnd}, \lambda_k)$$

Grab a random subcolumn cloud state for each spectral interval, rather than calculate flux for each spectral interval for each cloud state