To exascale and beyond – The Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM), a performance portable global atmosphere model for cloud-resolving scales

A. S. Donahue[1], P. M. Caldwell[1], L. Bertagna[2], H. Beydoun[1], P. A. Bogenschutz[1], A. M. Bradley[2], T. C. Clevenger[2], J. Foucar[2], C. Golaz[1], O. Guba[2], W. Hannah[1], B. R. Hillman[2], J. N. Johnson[4], N. Keen[3], W. Lin[5], B. Singh[6], M. A. Taylor[2], J. Tian[1,6], C. R. Terai[1], P. A. Ullrich[1], X. Yuan[7]Y. Zhang[1]

[1]Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA, 94550, USA
[2]Sandia National Laboratories, Albuquerque, NM, USA
[3]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[4]Cohere Consulting LLC, Seattle, WA, USA
[5]Brookhaven National Laboratory, Upton, NY, USA
[6]Pacific Northwest National Laboratory, Richland, WA, USA
[7]Argonne National Lab, Lemont, IL

Key Points:

- Describes the C++/Kokkos implementation of the Simple Cloud–Resolving E3SM Atmosphere Model (SCREAMv1)
- SCREAMv1 leverages GPUs to surpass one simulated year per compute day at global 3km resolution
- High resolution improves some meso-scale features and the diurnal cycle but large-scale biases require improvement across all four seasons.

Corresponding author: Aaron S. Donahue, donahue5@llnl.gov

Abstract

The new generation of heterogeneous CPU/GPU computer systems offer much greater computational performance but are not yet widely used for climate modeling. One reason for this is that traditional climate models were written before GPUs were available and would require an extensive overhaul to run on these new machines. In addition, even conventional "high–resolution" simulations don't provide enough parallel work to keep GPUs busy, so the benefits of such overhaul would be limited for the types of simulations climate scientists are accustomed to. The vision of the Simple Cloud-Resolving Energy Exascale Earth System (E3SM) Atmosphere Model (SCREAM) project is to create a global atmospheric model with the architecture to efficiently use GPUs and horizontal resolution sufficient to fully take advantage of GPU parallelism. After 5 yrs of model development, SCREAM is finally ready for use. In this paper, we describe the design of this new code, its performance on both CPU and heterogeneous machines, and its ability to simulate real-world climate via a set of four 40 day simulations covering all 4 seasons of the year.

Plain Language Summary

This paper describes the design and development of a 3km version of the Energy Exascale Earth System Model (E3SM) atmosphere model, which has been fully rewritten in C++ using the Kokkos library for performance portability. This newly rewritten model is able to take advantage of the state–of–the–science high performance computing systems which use graphical processor units (GPUs) to mitigate much of the computational expense which typically plagues high–resolution global modeling. Taking advantage of this high–performance we are able to run four seasons of simulations at 3km global resolution. We discuss the biases, including the diurnal cycle, by comparing model results with satellite and ARM ground-based site data.

1 Introduction

Accurate prediction of regional climate change – particularly of extreme events – is essential for avoiding the worst impacts of climate change. Numerical simulations are our main tool for providing quantitative predictions. While regional models can add fine-scale detail, global models are needed for future prediction because all areas of the planet contribute to changes in the general circulation, which in turn affects climate change everywhere. Weather forecasts longer than a few days also require global models because the impact of a weather event propagates outwards to affect far-flung regions. The need to capture processes spanning the range of scales from micron-scale water droplet interactions to global circulations makes weather and climate modeling among the world's most computationally challenging applications.

Climate models deal with this incredible range of scales by dividing the world into discrete grid cells. Processes larger than the grid scale are resolved explicitly and smaller processes are parameterized based on information available at the grid scale. The first general circulation models (GCMs) had $\sim 5^\circ$ horizontal grid spacing and < 10 vertical levels (Hamilton, 2020; Manabe et al., 1965; D. Johnson & Arakawa, 1996). Modern GCMs divide the atmosphere into approximately 60 vertical levels and typically use $\sim 1^\circ$ horizontal grid spacing (Ingram & Bushell, 2021). At this scale, convective motions critical to the vertical transport of heat, moisture, and humidity are unresolved; efforts to parameterize these effects have met with limited success despite 50 yrs of efforts (Randall et al., 2003; Klein et al., 2013). The effects of topography on precipitation and winds are also sorely lacking at these scales, though some of these effects can be parameterized. Process interactions on scales unresolved by modern GCMs are also critical to capture but too numerous and complex to fully parameterize. Taken together,

insufficient resolution is widely understood to be a major source of predictive error in GCM predictions (Sanderson et al., 2008; Sherwood et al., 2014).

In light of this, a new breed of global models is being developed. Known variously as global storm-resolving models (GSRMs), km- or k-scale models, or global cloud-resolving models, these models attempt to explicitly resolve deep convection by employing horizontal grid spacing finer than 5 km. The first GSRM was the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) (Tomita & Satoh, 2004; Satoh et al., 2008), for which the world's fastest supercomputer at the time – Japan's Earth Simulator – was built. Developing GSRMs has become quite popular in recent years, with 12 models participating in the second DYnamics of the Atmospheric general circulation Modeled On Nonhydrostatic Domains (DYAMOND-PhaseII) GSRM intercomparison project in 2021 (`https://www.esiwace.eu/the-project/past-phases/dyamond-initiative/services-dyamond-winter`). These models have been found to greatly improve spatial distribution, intensity, and diurnal characteristics of precipitation (Caldwell et al., 2021) and show great promise for the representation of clouds (Tomita et al., 2005). They are also able to represent important large-scale weather events like tropical cyclones (JUDT et al., 2021), derechos (Liu et al., 2023), and mesoscale convective systems (Feng et al., 2018) which simply aren't resolved in conventional GCMs. They still struggle with boundary-layer clouds, which is unsurprising since boundary-layer eddies are smaller than the GSRM grid scale. Convection is only partially and crudely resolved at km scales. Microphysics is a critical area for these new models because it will always be parameterized in weather and climate models. Despite these imperfections, GSRMs capture many more scales and processes from first principles compared to GCMs and empirically perform better than GCMs in many metrics. Because this is such a new area of research, further rapid improvements are expected.

GSRMs have a major shortcoming, though: on present-day computer hardware, they are too slow and expensive to run long enough to capture the evolution of climate, or to run with multiple ensemble members as needed for modern weather prediction. Stevens et al. (2019) report that DYAMOND1 models averaged about 6 simulated days per computational wall day (SDPD), which makes the multi-century runs needed to predict climate change impossible. On the positive side, a lot can be learned from pairs of year-long idealized runs. Climate feedbacks in response to greenhouse gases can be estimated with relative accuracy from pairs of 1 year simulations with sea surface temperature (SST) prescribed at current values and uniformly raised by 4 K (Cess et al., 1989; Ringer et al., 2014; Qin et al., 2022). Effective radiative forcing (ERF) due to greenhouse gas increases can be assessed from short paired experiments (Hansen et al., 2005), and anthropogenic aerosol ERF can also be computed from paired simulations so long as both runs are nudged to follow the same series of weather events (Kooperman et al., 2012). Unfortunately, short simulations are not sufficient for assessing statistical significance of changes, particularly related to regional behavior and extremes. This is particularly problematic because planning for local impacts is increasingly the focus of climate science, and that planning requires robust statistics. It is ironic that very fine resolution is exactly what's needed to accurately simulate extremes and to capture local effects, yet it is this very resolution that precludes simulations long enough to adequately sample extreme events. Developing strategies to combine resolution and computational speed is a grand challenge for climate science in the 2020s. A partial solution to this conundrum is to explore the impact of expected climate-change trends on particular historical events (Shepherd et al., 2018). These pseudo-global warming storyline simulations are useful for giving us a sense of how climate change will modify storms, but don't tell us how the frequency of those storms will change.

Breakthroughs in high-performance computing (HPC) have the potential to solve this grand challenge. Several modeling centers are building exascale computers now – computers that are capable of completing an exaflop worth of calculations per second.

A central feature of these machines is the use of graphics processing units (GPUs), which accelerate code by performing huge numbers of calculations simultaneously. Higher resolution means more columns to process in parallel, so GSRMs are perfect for these new machines. Unfortunately, existing codes require major modification to run efficiently on GPUs because loop order, variable dimension order, and algorithmic strategies which were best on conventional CPU chips are sub-optimal on GPUs. Even worse, each GPU vendor supports a different set of programming models. Because no single programming strategy works on all major new HPC architectures, several strategies for performance portability have been introduced (as described in 2.4).

The Energy Exascale Earth System Model (E3SM) has a stated goal of building a model which runs efficiently on its world-leading exascale computers. This was accomplished over the last 4.5 years by a team of ∼6 full-time employee equivalents using templated C++ and the externally-developed Kokkos library (Trott et al., 2022) to produce code which runs efficiently on a wide variety of architectures. We started by extending version 1 of the Fortran-based E3SM Atmosphere Model to run credibly – but not efficiently – at 3.25 km resolution. The resulting model, known as version 0 of the Simple Cloud-Resolving E3SM Atmosphere Model (SCREAMv0 or simply v0) is described in (Caldwell et al., 2021) (hereafter C21). SCREAMv0 was used as a guide for rewriting the entire model in C++. We call the resulting code the E3SM Atmosphere Model in C++ (EAMxx). The GSRM configuration of EAMxx is called SCREAMv1 (or v1). Building a slow prototype first gave us a template for C++ porting and allowed us to do science with the model while we were still working on our performant version.

Our C++/Kokkos approach has proven effective. SCREAMv1 is running efficiently on several of the world's fastest computers and was awarded the inaugural 2023 Gordon Bell Prize for Climate Modeling (M. Taylor et al., 2023; *ACM Gordon Bell Prize in Climate Modeling*, 2023). It has also been critical to ensuring we can run on all of the computers available to us. Unlike other modeling centers, E3SM gets most of its computing allocation from DOE Leadership Computing Facilities, which choose architectures to satisfy a wide range of DOE applications. Unlike Fortran, C++ is better suited for a single code base that can adapt to a variety of configurable options, via template metaprogramming. A complete rewrite also allowed us to better structure and compartmentalize the code, which allowed us to implement a more robust and rich unit testing framework. The rich C++ language and its extensive standard library also allows for a higher level code, hiding some tedious implementation details from the code, leaving a simpler and leaner code. Finally, freeing ourselves from decades of legacy code also allowed us to remove older functionality that is no longer necessary of supported that are no longer supported.

Because we tested that Fortran and C++ implementations of each process were bit-for-bit, v0 and v1 behave very similarly. As described in Sect. 2, however, process coupling, aerosol implementation, and some tuning choices are different in v1 (see Sect. 2 for details). For efficiency, we rely on C21 to describe SCREAM model climate and focus here on differences in behavior between model versions and on aspects of model climate that were not analyzed in the previous paper. In particular, this paper includes analysis of the seasonal cycle, which was not possible using the shorter simulation in C21. This paper also makes extensive use of ground-based observational data for model validation, e.g., Atmospheric Radiation Measurement (ARM) data by US Department of Energy (DOE). ARM provides multi-year and even multi-decadal process-focused high-resolution observational datasets at strategic geographic points worldwide. Because GSRM grid resolution closely matches the spatial extent of ARM data, GSRM-ARM evaluation provides a great opportunity for high-resolution model improvement.

This paper provides an introduction to the EAMxx code and explains our software-engineering strategy in section 2 and highlights our computational performance in section 4. Our first SCREAMv1 simulation campaign and associated evaluation data are

described in section 2.5, followed by evaluation of weather and climate skill in Sect. 5. Conclusions follow in Sect. 6. In evaluating SCREAMv1, it is important to remember that the model is young and we have focused so far primarily on speed rather than skill. As we pivot now towards climate improvement, our efforts to develop a robust code validation and testing framework will be invaluable.

2 Model Description

2.1 Model Processes

SCREAMv0 and SCREAMv1 have nearly an identical set of process representations. For brevity we list them here and refer the reader to C21 for more details. Fluid dynamics is handled by the non-hydrostatic High Order Method Modeling Environment (HOMME-NH) (Dennis et al., 2005, 2012; M. A. Taylor et al., 2020; Bertagna et al., 2020). Clouds and turbulence are parameterized using the Simplified Higher Order Closure (SHOC) scheme (Bogenschutz & Krueger, 2013), while microphysics is based on the Predicted Particle Properties (P3) scheme of Morrison and Milbrandt (2015). For simplicity, P3 is run using a single ice category. Gas optical properties and radiative fluxes are computed using the RTE+RRTMGP radiative transfer package (Pincus et al., 2019) which has been rewritten in C++ using the YAKL performance portability library (Norman et al., 2022) for improved performance.

There are a few notable differences in the SCREAMv1 configuration versus SCREAMv0. While SCREAMv0 used a relative humidity-based ice cloud fraction scheme, SCREAMv1 sets ice cloud fraction based on cell-average ice mass mixing ratio q_i, where ice cloud fraction is set to 1 for $q_i > 10^{-14}$ kg kg^{-1} and to 0 elsewhere. Assuming the whole cell is filled with cloud whenever there's even a small amount of ice present is needed in order for P3 ice processes to be active for falling snow. Ice clouds aren't visible until ice mass becomes much larger, however, so cloud evaluation typically uses a much larger cutoff. We follow C21 in setting total cloud fraction to 1 whenever cloud ice or liquid mixing ratio is greater than 10^{-5} kg kg^{-1}. Otherwise cloud fraction is set to zero. This is done as a post-processing step.

Aerosol effects in SCREAMv1 are prescribed based on 5-year monthly climatologies from an E3SMv2 AMIP-type simulation (Golaz et al., 2022) with 1° spatial resolution. This low-resolution climatology is interpolated to GSRM resolution during run time. Aerosol optical properties (single scattering albedo, asymmetry parameter, and shortwave/longwave optical depths) are prescribed directly. Aerosols affect microphysics through the cloud-condensation nuclei (CCN) amount, which is read in from the E3SMv2 climatology. Where clouds are present, the liquid droplet concentration n_c at the beginning of microphysics is replaced by the CCN concentration wherever CCN$> n_c$. Elsewhere in the code, n_c is handled following the typical two-moment P3 specifications described in Morrison and Milbrandt (2015). Thus CCN stands in as a crude droplet activation scheme, making sure n_c doesn't drop below climatological values while allowing for cloud processing of drop number. This implementation is very similar to MACv2-SP as described in Stevens et al Stevens et al. (2017), but we get aerosol-relevant quantities from a model simulation rather than an idealized plume calculation and allow for a quasi prognostic n_c. We refer to this scheme as Simple Prescribed Aerosols (SPA). Ice number is yet to be included in SPA and is handled as in Morrison and Milbrandt 2015 Morrison and Milbrandt (2015).

Several other minor changes were made to correct deficiencies in SCREAMv0. Excessive disaggregated shallow "popcorn" convection in the deep tropics was found to be related to SCREAMv0's tendency to convert cloud water to rain too quickly. We reduced these problems in SCREAMv1 by disabling the enhancement of collision/coalescence due to an assumed gamma distribution for cloud water described in C21. SCREAMv0 pro-

duced conspicuous low level clouds characteristic of stratocumulus where trade cumulus are often found. To help partially ameliorate this, a tuning parameter related to the parameterization of the third moment of vertical velocity was adjusted to prevent the skewness of vertical velocity from readily going negative in unstable conditions. Specifically, this is the c_{diag} term which was increased from 5 to 7 (Bogenschutz & Krueger, 2013). In addition, to reduce turbulence oscillations within the boundary layer the turbulence length scale of SHOC was revised from having separate formulations for the subcloud layer and cloud layer to having one continuous formulation. The new formulation can be found in the appendix of Chinita et al. (2023). The impact of these changes is discussed in Sect. 5. Before performing simulations, we confirmed that global-average shortwave and longwave global imbalances were within 0.5 W m^{-2} of CERES-SYN for several 2 day simulations.

2.2 Process Coupling

Rewriting the model in C++ gives us the opportunity to fix shortcomings with the previous E3SM atmosphere model (hereafter EAMF90) which have become apparent over the years. This rewrite also allows us to take advantage of the flexibility provided by C++. Each process in EAMxx is an instance of the generic atmosphere process class. Processes include parameterizations (like microphysics) but also fluid dynamics solvers and simple adjustments like nudging towards observations. Modularity makes adding, removing, or changing the order of processes easy, enables the development of tools for use across all processes, and simplifies the work needed to create new processes. The default process order used for all EAMxx simulations discussed in this paper is

1. import fluxes from surface components
2. HOMME-NH
3. SHOC
4. SPA
5. P3
6. RRTMGP
7. export fluxes for surface components

This ordering follows best practices according to Donahue and Caldwell (2018). Unlike many models (including EAMF90) which split the atmosphere into a function call for processes before surface coupling and another function for after-surface processes, the EAMxx code is executed all at once, bracketed by communication from/to surface models. This greatly improves model readability and makes component-level parallelism easier. Splitting the atmosphere model into multiple chunks is often required in other models in order to implicitly couple surface fluxes with atmospheric turbulence while also passing the surface schemes the most balanced version of the atmosphere state (which typically comes after radiation). Neither EAMxx nor EAMF90 employ this sort of tight coupling with surface fluxes. Because the place where surface components are called in EAMF90 (before dynamics) is different from where their effects are applied (during SHOC), this shift in process order between EAMxx and EAMF90 is not expected to significantly affect the model solution.

A particular design goal for EAMxx was the ability to easily switch the timestepping strategy for the whole model. Because some processes (like liquid condensation) are diagnostic rather than prognostic and other processes (like dynamics) need to use sophisticated timestepping schemes of their own which don't make sense for other cruder processes (like microphysics), we chose to have each EAMxx process return the state after that process acts rather than the time tendencies due to that process. If needed (e.g., for parallel splitting, or higher-order process coupling), tendencies can easily be regenerated from the difference in state from after versus before a given process has acted.

The atmosphere process initialization and execution is handled by the *atmosphere driver* (AD). Additionally, the AD is responsible for the creation and handling of other infrastructure components, such as grids, I/O streams, scratch memory, and persistent variables. The latter are variables that need to be shared between processes, and are represented in EAMxx by `Field` objects. In addition to storing the actual raw data (both on host and device), fields also store a variety of metadata pointers, used to identify, manipulate, and track the field within the atmosphere timestep. Every atmosphere process must explicitly declare which fields it needs as inputs as well as which fields it computes as outputs. This allows the AD to provide only what is needed, avoiding unexpected updates to other variables. It also allows us to build a directed acyclic graph (DAG) representing how each variable is updated within the atmosphere timestep (see Figure 1). The AD uses this DAG to compile a list of variables that are needed for bit-for-bit model restarts. This ensures that the memory footprint of EAMxx restart files is kept at a minimum. In contrast, the hard-coded list of restart variables in EAMF90 must be constantly maintained to avoid errors or bloating of the restart file size.

2.3 Testing infrastructure

Modern software practices emphasize robust unit testing and continuous integration to ensure that new features do not break current code, to improve confidence in the validity of the model, and to enhance our ability to debug the code if (meaning, when) a buggy piece of code makes its way in the main branch. A drawback of the EAMF90 design is that individual processes rely on a complex web of other modules and tools, making it difficult to isolate a single process for unit testing without a) hard–coded suppression of all other processes or b) building the entire E3SM model. This design made unit testing EAMF90 prohibitively difficult. In EAMxx, unit testing is made easier by extensively implementing the separation of concerns principle. In particular, each process is separated into its internal implementation and an interface layer. The former is intended to be completely self contained, though within EAMxx all internal process representations are broken into a number of smaller functions (to aid unit testing). Information to and from these functions are passed solely via arguments (typically Kokkos views and simple data structures) or (in the case of physical constants or user-specified settings) by data members of in-scope classes. The interface layer connects EAMxx-specific constructs (related to the AD) with the more generic constructs used by the actual process implementation. This separation of concerns allows us to individually test each atmosphere process, as well as the internal functions of each package. In SCREAMv1, numerous unit tests have been written to test individual functions inside the physics parametrization packages, as well as their interfaces. The driver and infrastructure data structures (including I/O) are also tested independently, meaning without the need to run a particular configuration of the model. A testing infrastructure developed within the project allows us to run tests for a variety of MPI ranks and OpenMP thread counts (where appropriate). A continuous integration system automatically runs tests on a variety of HPC architectures, both on a nightly basis, as well as before the integration of any code update. In the latter case, to speed up the integration of new features, we run a subset of the tests, while in the former, we run an extensive set of tests (including in-depth memory checks), to ensure the integrity of the code base.

2.4 Computational strategy

The huge computational cost of a GSRM requires efficient use of the largest HPC machines available. The last 10-15 years have seen an increase in diversity of chip architectures used in the largest HPC centers. On top of conventional CPUs, various other architectures have emerged, including many-core CPUs and, more importantly, GPUs. The latter have become a predominant choice for the largest supercomputers procured by the US Department of Energy.

Even though GPUs appear to be the dominant architecture for exascale in the coming years, performance portability (that is, the ability of a code base to run efficiently on different HPC architectures) is important in order to minimize disruption from possible architecture changes in the future. Furthermore, maintaining an efficient implementation on conventional CPUs is important because it makes debugging easier during development, and because CPUs are more efficient at low resolutions where opportunities for parallelism are limited.

When measuring performance, common metrics such as FLOPS or memory bandwidth are often used, which compare the application efficiency with the maximum possible efficiency on a given platform. However, the most common metric in the realm of climate sciences is Simulated Years Per Day (SYPD) or Simulated Days Per Day (SDPD). The latter is the metric we use for the evaluation of the SCREAMv1 model.

There are currently three main approaches to achieve performance portability in scientific applications:

- **Compiler directives**. The code is decorated with compiler directives, which provide hints to the compiler on how to optimize and/or parallelize the code. This is the approach used by OpenMP (*OpenMP Application Programming Interface Version 5.0*, 2018) or OpenACC (*The OpenACC Application Programming Interface Version 3.0*, 2019).
- **General purpose libraries**. Architecture-specific choices are delegated to a third party library (TPL). The application then uses the library interfaces as a programming model, obtaining a single code base that run on all the architectures supported by the TPL. Examples of this approach include Kokkos (Trott et al., 2022), YAKL (Norman et al., 2022), and Raja (Beckingsale et al., 2019).
- **Domain Specific Languages**. The developer writes code in a high level language, which is then translated into standard source code (such as C/C++/Fortran) by an intermediate compiler, which takes care of making optimization to the code targeted to the particular architecture where it will be executed. Examples of this approach in the climate/weather domain include GridTools (Afanasyev et al., 2021), PSyclone (*PSyclone User Guide*, n.d.), and CLAW (Clement et al., 2018), the latter sharing some similarity with the compiler directives approach.

Each of the above approaches comes with benefits and drawbacks. Choosing the best approach requires consideration of several factors, including compiler support, ease of development, support for future architectures, stability and maturity of the approach, prospects for long-term support, and amount of staff time available. In EAMxx, we adopted the second approach, using the Kokkos library for on-node parallelism. Kokkos is a very mature library that has already been successfully used in a variety of large projects (most notably, Trilinos (Trilinos Project Team, 2020 (acccessed May 22, 2020)) and other apps depending on it), demonstrating that it is reliable, effective, and robust. The main features of Kokkos are constructs to handle on-node parallelism and device-friendly multidimensional arrays. Kokkos exploits template metaprogramming to map abstract types to different concrete implementations on each architecture. For multidimensional arrays, it abstracts concepts such as the data type (e.g., `double` or `int`), the memory space where the data is located (e.g., CPU memory or GPU memory), and the layout (e.g., whether the rightmost index is the one that strides fastest or slowest). For parallelism, it abstracts concepts such as the parallel operation (e.g., a map, reduction, or scan), the threading device where the operation is executed (e.g., CPU or GPU), and the execution policy, that is, how the iterations are distributed among available threads.

In this work, we heavily leveraged the expertise developed during the conversion of the HOMME dynamical core to C++, which was also done using Kokkos, as described in Bertagna et al. (2019) and Bertagna et al. (2020). In those efforts,the authors estab-

lished some solid guidelines that became the cornerstone of EAMxx, which we can summarize in the following three most important aspects.

- **Parallelism**. To take advantage of the massive amount of parallelism on GPUs, the implementation needs to be able to expose all possible layers of parallelism in the algorithm and data structures. For EAMxx, this entailed using *hierarchical parallelism* to allow a team of threads to share intermediate work.
- **Vectorization**. In order to remain performant on CPU architectures, efficient use of vector instructions is necessary, as vectorization remains one of the most effective ways to increase FLOPS per energy used. In EAMxx, this required the development of lightweight data structures, to enhance the ability of the compiler to vectorize code.
- **Memory**. Codes arising from the discretization of PDEs tend to be more memory bound than compute bound. Hence, minimizing memory movement and memory footprint is important to keep the CPU and GPU busy. In EAMxx, this was done by using shared/reusable workspace data structures to compute intermediate quantities, by minimizing the size of persistent variables, and by increasing access time for each thread.

Among the above, hierarchical parallelism was the most important aspect, and therefore it is helpful to illustrate with a simple example the main concept, as well as how it can be implemented using Kokkos. Suppose we want to normalize the rows of an M-by-N dense matrix X. In plain C++, this task could be written as

```
for (int m=0; m<M; ++m) {
  double normSq = 0;
  for (int n=0; n<N; ++n) normSq += X(m,n)*X(m,n);
  for (int n=0; n<N; ++n) X(m,n) /= sqrt(normSq);
}
```

assuming X has the proper overload of `operator()`. Note that the `for` loop which computes `normSq` for a given row must complete before any entry in that row can be normalized. Without using hierarchical parallelism, if one had T threads available (with T potentially large), they could choose to either parallelize only the outer loop, or divide the code into two sets of nested loops, the first computing the norm of all rows, and the second doing the normalization, at the price of requiring more storage (as well as an extra kernel launch on GPU). Hierarchical parallelism aims to keep the original loop struc-

ture while parallelizing outer as well as inner loops. With Kokkos, we can use a `TeamPolicy` to allow groups of threads to cooperate and share intermediate results:

```
using namespace Kokkos;
using Policy = TeamPolicy<ExecutionSpace>;
using TeamMember = typename Policy::member_type;
auto f = KOKKOS_LAMBDA(const TeamMember& team) {
  int m = team.league_rank();
  double normSq = 0;
  auto compute_norm = [&](int n,double& sum){
    sum += X(m,n)*X(m,n);
  };
  parallel_reduce(TeamVectorRange(team,N), compute_norm, normSq);
  team.team_barrier();
  auto normalize = [&](int n}{
    X(m,n) /= sqrt(normSq);
  };
  parallel_for(TeamVectorRange(team,N), normalize);
};
parallel_for(Policy(M,N),f);
```

where X here is stored in a `Kokkos::View` (to allow accessing it from device). Each outer iteration is assigned to a team of threads, so that different teams handle different rows of X, and can therefore operate completely independently. Threads within the same team cooperate to compute the vector norm, which is used in the second loop to normalize the vector. A team synchronization barrier between the loops ensures the norm calculation has been completed before any thread of the team attempts to use it in the second loop. This barrier, as the name suggests, is only local to the team, and does not affect the work of other teams. The way Kokkos parallelizes each loop depends on the underlying threading model (e.g., CUDA vs OpenMP), and on the available resources.

2.5 Experimental Setup

All simulations in this paper were run on a cubed-sphere grid with 1024 x 1024 spectral elements on each face and a 4x4 grid of Gauss-Lobatto-Legendre nodes defining each element. This results in an equatorial horizontal grid spacing of ∼3.25km. The physics parameterizations are solved on a uniform finite-volume 2x2 grid for each individual spectral element, following Hannah et al. (2021). The vertical structure of the grid uses a terrain following hybrid coordinate system with a model top of 40km and 128 levels with varying layer thickness. See section 2 of C21 for more details on the model grid. All simulations ran with a top-level timestep of 100 seconds. Coupling with surface components and most atmospheric processes run at that timestep. Dynamics is CFL limited, and is thus substepped within the atmosphere timestep at a frequency of 12 dynamics steps per atmosphere timestep for a dynamics timestep of 8.33s. Conversely, radiation does not need to be run as frequently as the other physics parameterizations and is thus supercycled so that the radiation forcing is updated every three atmosphere timesteps for a radiation timestep of 300s. All simulations are run with prescribed ocean SSTs and sea-ice extent, while sea-ice thermodynamics and land behavior is interactive. The land model and thermodynamic sea-ice are both as in SCREAMv0 and EAMxx is coupled to the rest of E3SM with the same interfaces as the other components. Custom routines were written to translate between the C++ data structures of EAMxx and the Fortran interface of E3SM's coupler/driver (Craig et al., 2012)

Model validation is conducted using 40-day simulations for each of the four seasons. These simulations have start dates of January 20, 2021 (DYAMOND2 configuration), April 1, 2013, August 1, 2016 (DYAMOND1 configuration) and October 1, 2013.

For the remainder of the paper we will refer to these as the Jan, Apr, Aug and Oct simulations, respectively.

3 Observational Data for Model Validation

3.1 Satellite Data

Global validation is based on a number of satellite and reanalysis data products. Global precipitation comparison is based on the NASA Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (GPM-IMERG) dataset (Huffman & coauthors, 2019). Top-of-atmosphere radiative fluxes and cloud forcing are from the CERES Synoptic (SYN1deg) product (Doelling et al., 2013). All other variables are based on the European Center for Medium-Range Weather Forecasting's ERA5 reanalysis product (Hersbach et al., 2020).

3.2 ARM Data

Ground-based US/DOE ARM observations are used to assess the performance of SCREAMv1 in simulating surface variables and cloud fraction. To compare with the four SCREAMv1 seasonal runs, the observed composite diurnal cycle is constructed based on the ARM Best Estimate (ARMBE, Xie et al., 2010) data, which include measurements from long-term fixed observatories such as South Great Plains (SGP, a mid-latitude land site; data from 1996 to 2020 used in this paper), Eastern North Atlantic (ENA, a mid-latitude marine site; data used from 2014 to 2020), North Slope of Alaska (NSA, an Arctic site; data used from 1998 to 2020) and Tropical Western Pacific (TWP) C1 at Manus (1996 to 2011) and C3 at Darwin (2002 to 2011); and from the ARM mobile facility during the field campaign of Green Ocean Amazon (GoAmazon; from 2014 to 2015). SCREAM data whose cell centers are within a circle of 10 km radius are averaged for comparison against the multi-year average observational data for each hour of the diurnal cycle in each season. Increasing radius size to 20 km has negligible impact on results. Surface heat flux measurements are not available at all the TWP sites and in Jan at GoAmazon. Due to output issues, 2-meter vapor mixing ratio is not available from the SCREAMv1 Oct run. Cloud fraction at ARM sites are based on Radar, Lidar, and Ceilometer combined retrievals called Active Remote Sensing of Clouds (ARSCL, Clothiaux et al., 2000; Kollias et al., 2007) and sampled every 3-hourly here to match with the frequency of SCREAMv1 output of 3D variables. The cloud fraction variable in SCREAMv1 output was corrupted, so as in Caldwell et al. (2021), we assume cloud fraction is 0 or 100% wherever the sum of liquid and ice water content is smaller or greater than 10^{-5} kg/kg (respectively). These offline cloud-fraction calculations are performed on 3-hourly instantaneous snapshots. As SCREAMv1 only simulates one season each, we have considered interannual variability from observations, especially shown in the comparison on surface variables.

4 Computational performance

We distinguish between *production* and *performance* simulation configurations. The 40 day seasonal simulations described above are considered to be production simulations. Production-run job submissions typically ran 20 simulated days at a time and always occupied 1024 nodes of the OLCF Summit system, giving 1024 vertically extruded spectral elements and 4096 physics columns to each GPU. The critical differences between production and performance simulations is that the latter have no output, are run for a variety of node counts on a number of machines, and are much shorter (typically 12 to 48 simulated hours).

4.1 Computer systems

Production and performance simulations were run on the Summit supercomputer at the Oak Ridge Leadership Computing Facility (OLCF) at Oak Ridge National Laboratory (ORNL). Summit has 4608 compute nodes. Each node has two IBM Power 9 (AC 922) CPUs with 512GB of DDR4 memory and six NVIDIA Tesla V100 GPUs with 16GB of HBM2 memory per GPU. Summit's internode network is Mellanox EDR 100G InfiniBand. NVIDIA NVLINK connects the GPUs within a node. The software environment for our model runs includes GNU GCC 9.1.0, NVIDIA Cuda 11.0.3, and IBM Spectrum MPI 10.4.0.3.

Performance results were also obtained on the OLCF Frontier system and the NERSC Perlmutter systems. Frontier has 9408 nodes. Each node has one 64-core AMD "Optimized 3rd Gen" EPYC CPU and 4 AMD Instinct MI250X GPUs. Because each GPU has two Graphics Compute Dies (GCD), each node effectively has eight GPUs. Each node has 512GB of DDR4 CPU memory and 64GB of HBM2E memory per GPU, and a peak performance of 53 FP64 TFLOPS per GPU. The system interconnect consists of 4 HPE Slingshot 11 NICs per node, providing 100 GB/s network bandwidth, while the CPU-GPU and GPU-GPU interconnect is AMD Infinity Fabric.

The Perlmutter GPU partition has 1536 nodes. Each node has one 64-core AMD EPYC "Milan" 7763 CPU and four NVIDIA Ampere A100 GPUs, with 256GB of DDR4 CPU memory and 40GB of HBM memory per GPU, and a peak performance of 9.7 FP64 TFLOPS per GPU. The system interconnect has four HPE Slingshot 11 NICs per node.

The Perlmutter CPU partition has 3072 nodes. Each node has two 64-core AMD EPYC "Milan" 7763 CPUs, with 512GB of DDR4 memory, and a peak performance of 5 FP64 TFLOPS per node. The system interconnect has one HPE Slingshot 11 NIC per node.

On Summit, the job runner divides a node into resource sets. In all of our simulations on Summit, we use one GPU, one MPI rank, and seven cores per resource set. The land and thermodynamic sea ice models, the model coupler, and I/O run on the CPU; the atmosphere model computations run on the GPU. The land model threads effectively on the CPU; thus, we provided the land model with 14 threads per resource set and, thus, two hyperthreads per CPU core. The atmosphere model supports MPI buffers on either the GPU or CPU. We sometimes experienced communication slowdown and large communication imbalances when using GPU buffers and, thus, used CPU buffers in all Summit runs reported in this section. We estimate the performance penalty of using host buffers instead of GPU buffers on Summit is approximately 10% for the dynamical core and, thus, 5 to 7% overall. On Perlmutter and Frontier, we were able to use GPU buffers in all runs.

The model configuration is similar on Frontier, just adapted to the particular numbers relevant to a node: eight MPI ranks per node, with seven cores per MPI rank, providing seven threads per rank for the land model. Similarly, on Perlmutter's GPU partition, each node runs with four MPI ranks and eight cores and threads per rank. On the Perlmutter CPU partition, the simulations are configured to use one MPI rank per core, for 128 ranks per node. It is worth noting that the choice to use Kokkos in EAMxx facilitated rapid deployment on Frontier and Perlmutter.

4.2 Results

Figure 2 shows the results of a strong-scaling study using the performance configuration, and Table 1 lists the data for each of the curves. We have collected data from all systems described in Sec. 4.1. The Frontier and Summit results for the full model were reported in M. Taylor et al. (2023), and here we add additional sub-component results

Machine	Nodes	Model	Atm.	Dycore	DIRK	CAAR	Physics
Frontier	512	58.3	64.8	82.3	288.8	306.0	337.6
	1024	103.7	118.3	149.9	562.0	596.0	620.5
	2048	175.6	197.6	252.1	1012.0	1069.2	1043.6
	4096	283.2	332.8	424.0	1760.2	1886.0	1693.5
	8192	419.5	514.5	645.3	3173.7	3349.6	2911.1
Summit	1024	70.0	78.2	100.8	401.4	465.9	382.3
	2048	109.4	132.9	169.4	744.1	870.9	679.3
	3072	135.6	168.2	211.2	1057.1	1234.3	911.5
	4096	156.4	205.0	257.3	1325.4	1565.1	1107.5
	4608	159.3	216.1	266.4	1606.7	1882.8	1272.7
Perlmutter-GPU	384	39.1	44.5	52.3	175.9	181.1	360.0
	512	51.0	58.7	68.8	239.8	240.1	466.3
	1024	95.1	110.0	129.4	457.9	471.0	886.0
	1536	132.0	160.5	190.2	677.0	697.3	1233.8
Perlmutter-CPU	1536	22.6	25.0	34.5	109.2	179.3	97.0
	2048	35.3	38.4	52.3	122.2	252.4	156.4

Table 1. Data, in SDPD, corresponding to the curves in Figure 2.

as well as new CPU and GPU Perlmutter results. In the figure, the x-axis is the number of nodes used in the model run. The y-axis is the model or subcomponent throughput in SDPD. The y-axis range is different in (a) and (b). The maximum time among all ranks is used when computing the SDPD. "Model" is the full model and uses the top-level model run-loop timer. It is the wallclock time a job submission takes excluding model initialization. To compute node-hours per simulation, one can use this throughput times days to simulate, plus a constant model initialization time. Subcomponents include "Atmosphere", which uses the top-level atmosphere model timer; "Dycore", for the dynamical core; "Physics", for the combined column physics parameterizations; "Dycore::DIRK", for the vertical solver in the Diagonally Implicit Runge-Kutta IMEX scheme; and "Dycore::CAAR", for "compute and apply right hand side", the primary kernels in the discretization's horizontal scheme. In (a), only "Model" is fully accurate; the subcomponent timers time regions that contain communication but have no MPI barrier at start or finish and, thus, are timed only approximately. The black undecorated line is the reference for perfect scaling.

The Summit and Frontier curves are smooth over the range of node counts because a large number of submissions were run, and the time data set corresponding to the highest-throughput run is plotted. Far fewer submissions were made on the Perlmutter GPU and CPU systems. As a result, some of the curves are jagged due to system variability, particularly the DIRK and CAAR timers at 384 and 512 nodes on the GPU partition and the DIRK timer at 2048 nodes on the CPU partition.

The timers in (a) include MPI communication, while those in (b) do not. Thus, in (b), the subcomponents scale much better than in (a). In (b), the small loss in GPU scaling is due entirely to loss in parallel efficiency, which in turn is a strong function of the work volume per GPU kernel launch. In (a), the loss in scaling is due almost entirely to MPI communication. The fact that GPU scaling in (b) is much better than in (a) emphasizes that the amount of work we pack onto each GPU is much less important for performance than the volume and frequency of MPI communication we do (at least for GSRMs). The dynamical core's scaling almost entirely determines the atmosphere model's scaling. The GPU curves in Figure 2a show greater falloff from perfect scaling with increas-

ing node count than the results from Perlmutter-CPU. This is because the on-node compute time is much smaller than for CPU-only machines, making the non-compute time (i.e. MPI communication) more impactful. In short, if computations are really fast, then communication determines scaling behavior.

The speedup of GPU results relative to CPU in Figure 2 and Table 1 supports our decision to rewrite the model to allow using GPU architectures. On Perlmutter, when using 1536 nodes, we can compare performance when using GPU vs CPU: on average, using GPUs the code is 6x faster than when using CPUs, with physics running 12x faster on GPUs. This is a particularly useful comparison because Perlmutter's CPUs and GPUs are both state of the art for 2023; Bertagna et al. (2020) (hereafter SC20) had previously shown GPUs to be 14.1x faster for dycore calculations on 1024 Summit nodes, but Summit GPUs are of relatively higher quality than their CPUs. Note however, that GPU nodes are much more expensive and consume more energy than a CPU node. Also, we spent more time optimizing model performance for GPUs. These caveats notwithstanding, the real-world impact of our rewrite is that our SCREAMv1 production simulations ran more than 10x faster than SCREAMv0 runs ($\sim$ 55 SDPD on 1024 Summit nodes versus 4–5 SDPD on 1536 nodes of the older Intel Knight's Landing (KNL) Cori machine). This acceleration is due as much to accessing bigger and faster machines as GPU acceleration, but both of these conditions were enabled by our C++ rewrite.

Figure 2 also demonstrates that EAMxx runs well across a wide variety of architectures. Bertagna et al. (2019) (hereafter GMD19) expands on this finding by comparing performance of the EAMxx dycore on seven different computer architectures: two NVIDIA GPUs, one IBM CPU, and four Intel CPUs. In all these cases, performance is excellent, with particularly high performance per unit of energy on the Intel KNL and all GPU systems.

Another important point about our C++ rewrite is that it is as good as – and often better than – the original F90 model on CPUs, despite the fact that the original F90 model had been optimized over many years of use. For brevity, we refer readers to GMD19 and SC20 for proof of this point. In particular, GMD19 Figures 9 and 10 compare performance of the original F90 and new C++ dynamical cores at 0.25° resolution on NERSC's Cori-KNL, Cori-HSW Intel Haswell (HSW), and Edison Intel Ivy Bridge (IB) supercomputers. On the IB and HSW architectures, the C++ model is slightly faster than the F90 model. On the KNL architecture, the C++ model is substantially faster than the F90 model. Figure 6 in SC20 adds to these results with a comparison of the C++ and F90 models on the IBM Power9 CPUs of the Summit supercomputer; here again, the C++ model is slightly faster than the F90 one. These results also hold for the full atmosphere model (dynamical core and physics) on newer CPUs, where M. Taylor et al. (2023) shows the C++ model is slightly faster than the F90 one on Perlmutter AMD EYPCs. A crucial part of achieving better than parity with the F90 model is using packs for explicit vectorization on wide-vector CPUs such as the KNL architecture (as explained in Sect. 2.4). Thus on the CPU the new C++ model performs at least as well as the legacy, well-optimized F90 model.

Figure 3 illustrates the approximate proportion of time spent in certain subcomponents in three cases of interest: performance model runs on 1024 (left bar) and 4096 (right bar) nodes of Summit, and a production model run on 1024 nodes of Summit (middle bar). The production simulations used for the results presented in Sec. 5 relied on seven 20-day model runs plus one set of fragmented model runs due to machine issues. The seven 20-day runs had top-level model run-loop throughputs in simulated days per wallclock day (SDPD) of, in increasing order, 51.8, 54.3, 54.7, 56.8, 57.8, 58.9, and 60.7 SDPD. The production model run used in the figure has the median total run time of the seven available 20-day model runs (56.8 SDPD). The y-axis is normalized wallclock time per unit of simulated time. The production run is the reference and, thus, has a total proportion of 1. The subcomponents are the dynamical core, the physics param-

eterizations, the rest of the atmosphere model other than these (mainly output writes), and the model excluding the atmosphere model (primarily the land model and component coupler). Thus, for example, in the production run, the dynamical core takes a little over 57% of the total model wallclock time. The 1024-node performance run is almost 20% faster than the production run because it does not write output. However, the dynamical core and physics blocks are about the same between the two 1024-node runs, as we expect. Because dynamics consumes most of the simulation time while physics is only ∼15% of the total cost at 1024 nodes, significant performance improvement will require changes to the dynamical core. Since the dynamical core is the most mature and optimized part of the code, further substantial improvements are unlikely. Writing output in this version of the model was very expensive and has been a target for recent improvements (M. Taylor et al., 2023). Switching to 4096 nodes decreases physics and dynamics costs roughly proportionally (as documented in Figure 2(a)) but the non-atmosphere portion of run time approximately does not decrease. This is a coupler scaling problem which should be addressed in the future.

5 Model Climate

The inability to perform multi-decadal simulations poses challenges for evaluating climate skill in GSRMs. Climatological biases can be hidden both by model drift over time and by natural interannual variability. Model spin-up can also appear more biased then longer equilibriated simulations. Previous analyses of short hindcasts by climate models, however, have shown that many long-term climate biases – particularly those related to clouds – show up within a couple days of model initialization (Xie et al., 2012). These studies convince us that our 40 day simulations, as described in section 2.5, will provide a useful sense of climate model skill. In addition, we will show that many model biases and successes show up consistently across all 4 seasons and echo the conclusions of C21. This reassures us that our conclusions are statistically robust. Where possible, we also include estimates of interannual variability from the observations we compare the model against in order to provide a sense of whether our results reflect real model differences or sampling uncertainty. Finally, we include time series of model behavior across each simulation to reassure readers that our conclusions aren't corrupted by model drift.

5.1 Global Analysis

In this section, we provide a brief summary of large-scale model climate in order to orient readers to model behavior. Subsequent sections dive into the physical behavior in particular climate regimes which contribute to these global features.

Figure 4 shows time series of globally-averaged variables for all four seasons. Since these are free-running simulations, temporal correlation between SCREAM and ERA5 is not expected beyond what is induced by prescribed SST and insolation. None of the four variables drift substantially relative to ERA5 results (which have trends due to the seasonal cycle and perhaps large-scale conditions). Note as well that model biases tend to be consistent across all seasons. This indicates that the results SCREAMv1 provides on 40 day timescales are likely to also hold on longer timescales. The left-hand column also includes output from our previous SCREAMv0 DYAMOND2 simulation. Both v0 and v1 have surface wind speeds ∼1 m s^{-1} too fast and similarly large precipitation rates, but their near-surface temperature and water vapor paths differ, with SCREAMv1 not necessarily closer to ERA5. In particular, SCREAMv1 is almost 1 K too warm; SCREAMv0 started similarly warm but matches ERA5 well by the end of the simulation. Both the surface warm bias and wind-speed bias are explored further in Section 5.5.

A remarkable feature of these SCREAMv1 runs is that despite the fact that we made no attempt to tune v1 other than reducing the popcorn convection found in v0, the top of atmosphere (TOA) net-radiative fluxes in the global mean are within 1.2 Wm^{-2} (Fig-

ure 5). Overall, the top panel in Figure 5 shows that SCREAMv1 captures the north-south seasonal shift of peak TOA energy input as well as the latitudes of transition from TOA radiation being a net source to a net sink of energy. The bottom panel of Figure 5 shows that despite the general agreement, SCREAMv1 does tend to have a negative bias in the TOA net radiative flux poleward of 50°latitude in the spring and summer hemisphere and a positive bias over 5S to 30S.

If we separate the net radiative flux into its shortwave and longwave components, they also agree to within 3 Wm^{-2} in the global mean. This result is demonstrated in Figures 6 and 7, which again shows that global averages mask large compensating regional biases, particularly in the shortwave. Comparison of Figure 6 versus 7 reveals that shortwave biases are the principle cause of both excessive TOA emission in summertime high latitudes and of overly-strong absorption at low latitudes noted in Figure 5. Longwave radiation actually damps the tropical bias (as expected due to competing shortwave and longwave effects in high clouds); from this compensation and the location of local maxima we conclude that SCREAM's deep convective clouds are anemic. The tendency for shortwave biases to be largest in the local summertime is at least partly a response to larger TOA insolation in the summertime hemisphere. Figure 8 shows the contribution from clouds to shortwave radiation. Lack of stratocumulus decks to the west of subtropical continents is stark and explains the positive insolation biases in these regions. Anemic stratocumulus is a common problem even in high-resolution climate models. Figure 9 and 11 from Caldwell et al. (2021) provided a more optimistic view of stratocumulus from an earlier model version; worse results here suggest we have more work to do. Shortwave cloud impact is largest in the summer-hemisphere midlatitudes, though deep clouds in the Inter-Tropical Convergence Zone (ITCZ) also reflect a lot of radiation.

Large temporal variability makes assessing precipitation from 40 day simulations more difficult. Figure 9 shows that the v1 model captures the maximum in zonal-mean precipitation rate and the seasonal migration of the ITCZ. In general, precipitation in SCREAMv1 tends to organize in an overly-narrow zonal strip and leads to precipitation being aggregated in a smaller area than satellite GPM IMERG estimates (Figure 9). The ITCZ also appears sharper, with stronger precipitation rates, consistent with what other models exhibit when the deep convective scheme is turned off (Wedi et al., 2020). On the other hand, the South Pacific Convergence Zone (SPCZ) and the Tropical West Pacific are areas where precipitation is underestimated, particularly in the boreal summer and October seasons. Finally, the midlatitude storms in SCREAMv1 generally tend to produce more precipitation than IMERG estimates, but there are no systematic shifts that are detectable in these 40 day simulations. Precipitation in v0 and v1 are quite similar both in global-mean rate and in terms of regions of large bias (not shown for v0). In SCREAMv0, it was noted that the rain water path exceeded the mean cloud liquid water path. SCREAMv1 does not suffer from the same issue (not shown).

Finally, we assess the three-dimensional thermodynamic and dynamic structure in SCREAMv1 across the four seasons. The zonal-mean vertical structure of temperature in Figure 10 shows that warm bias is largest in the lower and upper troposphere. We also see in Figure 11 that dry bias in vertically-integrated water vapor comes primarily from a lack of moisture between 850 and 600 hPa over the tropics and subtropics. Below that drying signal, we also see that the boundary layer in SCREAMv1, particularly in the southern hemisphere, actually tends to be too humid. This is indicative of insufficient mixing between the boundary layer and lower free-troposphere and a lack of mid-level cloud detrainment. The upper troposphere, on the other hand, shows slight hints of excessive humidity in the model, which indicates that upper level detrainment of humidity is adequate in SCREAMv1. Further discussion about convection in the tropics is provided in Sect. 5.2.

Figure 12 shows that the overall structure, location, and strength of the zonal winds are reasonable for all 4 seasons, though there are some consistent biases. First, the mid-latitude zonal jets tend to be shifted equatorward compared to ERA5. Second, upper-tropospheric equatorial winds have too much easterly momentum, particularly in Aug. In most instances, the sign of model biases is more or less height invariant (though magnitude increases proportionally with increased windspeed aloft). Interestingly, the excessive wind speeds found at the surface (e.g. Figure 4) do not extend into the interior of the atmosphere.

5.2 Tropics

5.2.1 Tropical Convection

A key advantage of GSRMs is the ability to explicitly simulate deep convection, which is important for understanding the vertical distribution of heat and momentum in the tropical atmosphere. Despite the need for parameterized convection on smaller scales, we expect SCREAMv1's ability to resolve larger convective circulations will yield a realistic simulation of the tropical climate. Therefore the following discussion will assess how well SCREAMv1 can reproduce the observed distribution of tropical convection.

Figure 13a shows profiles of ice and liquid water content in the tropical western Pacific (-10°to 10°latitude, 160°to 180°east longitude) from SCREAMv1 and CloudSat observations (2B-CWC-RO) for the 40-day period starting Oct 1, 2013. We acknowledge that satellite estimates of physical quantities can potentially carry large biases and uncertainty due to the limitations of the instrument. Thus, it is preferred to use satellite simulator output when comparing to satellite data to simulate these limitations and make for a fair comparison (Bodas-Salcedo et al., 2011). Unfortunately, the satellite simulator implementation in SCREAMv1 was not available for the simulations discussed here. Therefore, we will limit our discussion to a qualitative comparison between model and observations in recognition of the imperfections in the satellite data.

The profile of ice cloud water in SCREAMv1 compares favorably with CloudSat, although the peak sits at a slightly higher altitude (Figure 13a). SCREAMv1 also produces ice below 5km where there are no ice clouds in CloudSat data, which suggests that the melting of ice may be delayed in the model for condensate falling below the melting level. It should be noted that the neither P3 microphysics nor CloudSat make distinctions between ice and snow, thus ice is taken here to include all frozen hydrometeors.

The profile of liquid cloud water is generally underestimated in SCREAMv1, but there is an especially notable lack of mid-level clouds, which are responsible for the trimodal distribution observed in the tropics (R. H. Johnson et al., 1999). This deficiency could be related to the mid-level dry bias seen in Figure 11. Note that saturation of the CloudSat radar means it underestimates low and mid-level clouds (Schulte et al., 2023), so lower liquid cloud amount in the model suggests a systematic issue with liquid clouds.

Figure 13b shows the probability density function of total condensed water path compared to MODIS level 2 data (MYDO6L2) as a reference. SCREAMv1 underestimates the prevalence of mid-thickness clouds with water path of 20-600 g/m^2 and overestimates the prevalence of thick clouds with water path greater than 600 g/m^2. This is a common bias in conventional general circulation models (Kay et al., 2012; Klein et al., 2013; Medeiros et al., 2023), and there is some evidence that it may be common to GSRMs as well (Kodama et al., 2012). Instrumental uncertainty prevents us from drawing conclusions regarding the prevalence of thin clouds. Currently, it is unclear how to address this in SCREAMv1.

5.2.2 Tropical ARM Sites

Evaluating SCREAM at individual ARM sites gives us access to a treasure trove of high-quality observations. In this section we explore the composite diurnal cycle of several variables for several locations. Local features such as buildup of convective available potential energy, solar heating, geography, and land-sea contrast can cause certain behaviors to manifest at particular times of day. Reproducing the composite daily timing of observed meteorology has historically been very difficult for models, and therefore is often used to gauge the realism of a model's convective representation (Collier & Bowman, 2004; Pritchard & Somerville, 2009; Xie et al., 2019). Given that SCREAM's major advantage is the ability to explicitly resolve convective storms, it is reasonable to expect the model to perform well in this context. To test this hypothesis, we explore the diurnal rhythm in 3 tropical sites whose location is included in Figure 8; comparing site locations against the seasonal cloud radiative effect climatology they are superimposed on is useful for understanding the behavior described below.

SCREAM does an excellent job of reproducing the diurnal and seasonal cycles of cloud fraction at Darwin (Figure 14). In particular, SCREAM successfully captures the build-up of clouds during Darwin's wet season in Austral summer (January) and the generally dry conditions found during the rest of the year. Note as well the realistic portrayal of deepening shallow cumulus clouds during daylight hours. Unsurprisingly, good simulation of clouds at Darwin is accompanied by realistic simulation of thermodynamic variables though – as in the global analysis – 2m temperature and wind speed are too high and liquid water path and column integrated water vapor tend to be low (Figure 15). The diurnal cycle of near-surface water vapor mixing ratio at Darwin is also quite good, though its diurnal-mean value is a bit too high. Wet-season (Jan) precipitation does capture the observed peak at 16:00 local time, but predicts little precipitation at other times, while observed precipitation is above 15 mm/day at all times of day.

In the West Pacific Warm Pool, deep convection is more prevalent and small shifts in the position of the ITCZ can result in huge model biases. Off the north coast of Indonesia, Manus has substantial convective clouds year-round (Figure 16). SCREAM does successfully maintain deep clouds year-round in this region, but (as noted in previous sections) high cloud fraction is too large and mid-level cloud is missing. SCREAM also predicts a layer of low clouds through most of the day which are not observed. Further evidence of convective dysfunction is found in the diurnal cycle of precipitation, which peaks around sunrise in SCREAM but in the early afternoon in observations (Figure 17c). Because SCREAM has too much high cloud and not enough mid-level clouds, its LWP is far lower than observed (Figure 17d) and its ice water path is too strong (not shown). The diurnal cycle of near-surface temperature, moisture, and winds look completely wrong at Manus, but this is at least partially due to the fact that Manus is an island and a portion of the simulation data contributing to this result samples points over the ocean. Land surfaces warm much more than ocean during the day.

GoAmazon is interesting because it is an inland case, but with evaporative fluxes from the surrounding rain forest providing fluxes a bit like the ocean. Observations on the right-hand side of Figure 18 show low clouds in the morning gradually growing into mid-level and then deep convection in the afternoon, as noted previously in Tian et al. (2021, 2022). As found for other regions, SCREAM does a nice job of simulating deepening boundary-layer clouds in the morning but skips the mid-level congestus step, transitioning directly to deep convection. There are some hints that SCREAM may be transitioning from boundary-layer to deep convective clouds too fast (particularly for Jan) though interestingly the timing of peak precipitation is generally correct (Figure 19). Note however that observed precipitation has similar peak intensity for all seasons except Aug, while SCREAM magnitude has more seasonal variation and is generally too weak. SCREAM also shows an unexpected peak in the early morning in Apr. Because both model and observations sample land points in GoAmazon, the diurnal cycle of near-surface temper-

ature and other variables is more consistently reproduced. There is no significant difference in the 10-meter wind speed during daytime. Surface latent and sensible heat fluxes (LHF and SHF, respectively) in the tropics are only available at the GoAmazon site. LHF is slightly overpredicted while SHF is far too large around midday. Interestingly, SHF has much stronger seasonal cycle than observed. Also, LHF is too large at night while SHF actually goes negative (not shown). Understanding these behaviors via surface water and energy budgets is future work.

5.3 Midlatitudes

5.3.1 Synoptic Weather

Mid-latitude weather is often associated with synoptic scale frontal boundaries resulting from the meandering path of the jet stream. Assessing how well these systems are represented in the simulations presented here is challenging due to the divergence of simulated and observed weather patterns that complicates the comparison of individual storms. Thus, we assess the position and intensity of zonal jets and their variability under the assumption that any issues in the representation of mid-latitude storms will be imprinted in the zonal mean jets.

Figure 20 shows the 3-hourly variance (shading) and temporal mean (contours) of zonal mean zonal wind for the last 30 days of each simulation alongside the corresponding data from ERA5. In all cases the latitudinal position and altitude of westerly wind maxima are consistent with ERA5, however the winter hemisphere maxima are slightly stronger in intensity. There are several notable differences in the magnitude of the variance zonal mean zonal wind. For instance, the variance is smaller than ERA5 in the Apr simulation, while it is larger than ERA5 in the Oct simulation. The Aug simulation shows particularly weak variance relative to ERA5 south of the southern hemisphere jet. The differences in the patterns make direct comparison somewhat difficult, which would likely be remedied by a longer record of simulation data. If any of these differences prove to be robust in longer simulations then this may indicate a fundamental problem in the generation of eddy kinetic energy or the role of convective heating in baroclinic instability. Given the importance of expected changes to jets and storm tracks in a future warming scenario (Barnes & Polvani, 2013) this topic may demand closer inspection in future work.

Atmospheric rivers are another synoptic-scale feature of mid-latitude weather that can expose sensitivities in the simulated general circulation. Figure 21 shows the spatial density of atmospheric rivers (ARs) over each simulation period from SCREAMv1 and ERA5. ARs are defined following the tracking criteria in Ullrich et al. (2021), in which smoothed Laplacian of the integrated vapor transport is used to check for values below a prescribed threshold. Although we do not expect an exact match in AR density, given that these simulations are conducted at weather time scales, there is considerable qualitative agreement between the ARs produced by SCREAM and those from ERA5. To quantify some of the differences between AR objects, histograms of AR width and lengths are computed and shown in Figure 22, computed using the principal component method of Inda-Díaz et al. (2021). Under this metric, ERA5 ARs are statistically the same length as SCREAMv1 ARs, but are about 5% wider than SCREAMv1 ARs. This difference is attributed to the resolution at which ARs are detected. Given the lack of any clear systematic difference in AR characteristics in SCREAMv1 we conclude that the model is adequately reproducing the salient properties of these events.

5.3.2 Continental US Precipitation

The central continental US is notable for the regular occurrence of strong organized convective systems during the spring and summer months that can produce various types

of severe weather hazards (Schumacher & Johnson, 2006; Hu et al., 2021; Stevenson & Schumacher, 2014). These systems can also contribute to nocturnal precipitation over certain regions. These systems have proven difficult to represent in models unless the resolution is sufficient to explicitly resolve convection (Feng et al., 2018; Lin et al., 2022).

Figure 23 shows a comparison of precipitation for April 5, 2013, both as a map of precipitation and 500mb geopotential height at 00Z (a-b) and a Hovmoller of precipitation rate (c-d) averaged meridionally over the red box indicated in the upper panels. Since this is five days since initialization, weather patterns have diverged, but we can still make a qualitative comparison to assess the realism of SCREAMv1. Nonetheless, there are several similar features in both model and observations, such as widespread precipitation in the Pacific Northwest, a moderately organized convective system over the central US, and a precipitating frontal system off the East coast. The mesoscale feature in the Central US appears to be associated with a synoptic wave feature indicated by the geopotential height field. The Hovmoller shows that this system maintains a coherent propagation across the entire region. All of these features are consistent with the system in observations, providing some confidence that the model can produce realistic convective system in this region.

Figure 24 shows the composite diurnal cycle for select regions representing the high mountains, high plains, mid-plains, and low plains with IMERG (Figure 24a) and SCREAM (Figure 24b). The composite IMERG data shows progressively delayed peaks across the central United states that are well documented, especially during Boreal summer (Carbone et al., 2002; Carbone & Tuttle, 2008). The SCREAM model reasonably reproduces the delay of peak precipitation from the high mountain region to the high plains, and slightly so for the transition to the mid-plains, but misses the dominant nocturnal peak over the low-plains that is often attributable to elevated mesoscale convective systems. We speculate that this deficiency is due to accumulated biases in the large-scale environment as the simulation progresses, rather than a deficiency in the model's ability to simulate convective systems at this scale. The strong late-afternoon diurnal peak over the Southeast region of the United State is also well simulated.

5.3.3 Midlatitude ARM Sites

We now turn our attention to ARM's two long-term mid-latitude sites: the land site at SGP and the marine site at ENA. The location of these sites is also included in Figure 8.

Similar to the tropics, SCREAMv1 does an excellent job in simulating the diurnal cycle of the surface-forced shallow cumulus observed from late morning through afternoon (Zhang & Klein, 2013; Zhang et al., 2017) at SGP, though the timing of modeled clouds slightly lags the observed evolution (Figure 25). In Apr SCREAMv1 seems to also be catching the primary nighttime peak both in precipitation (Figure 26) and clouds associated with propagating MCS as well as the secondary peak in the late afternoon or early evening which is associated with locally forced deep convection (Zhang & Klein, 2010). Precipitation is less well captured in other seasons. Note that SGP is in the "low plains" region in Figure 24, so we've already seen that SCREAM misses deeper convection during summer because propagating convection doesn't reach that far from the Rocky Mountains. SCREAM does slightly better at capturing midlevel clouds at SGP than at tropical sites, but there is still a notable lack of clouds around 3 km in all seasons for most times of day. Unlike the tropics, SCREAM generally has *too little* high cloud in this region, suggesting that there is some nuance to the conditions that lead to excessive high clouds in the model. Though SCREAM's cloud profiles in each season aren't a great match with observations, SCREAM does largely capture the general diurnal and seasonal cycle at SGP. One glaring exception, however, is Oct. SCREAM cloudiness maximizes in this month when in reality there should be very little cloud at this time.

SCREAM's column-integrated water vapor is within the bounds of observed interannual variations (Figure 26). SHF and LHF generally are as well, but - as for GoAmazon - nighttime SHF is too negative and nighttime LHF is once again too positive. Near-surface temperature and wind speed tend to be too high, especially at night. These deficiencies are partially due to excessive turbulent mixing in stable nighttime boundary layers (see Sect. 5.5).

SCREAM does a very good job of reproducing both the seasonal and diurnal cycle of cloudiness at ENA (Figure 27), though diurnal variations in low clouds in Aug are too large. This is perhaps explained by a a large spike in precipitation between 4 and 8 AM local time (the first diurnal peak in cloudiness) for that season (Figure 28). It could be that Aug statistics are getting corrupted by a single large storm; 30 days is certainly a short averaging period for evaluating midlatitude precipitation. Other seasons also have peak precipitation larger than the 2σ bounds on observed interannual variations, suggesting that precipitation efficiency in this boundary-layer cloud regime might be overpredicted. If this was the case, then diurnal variations in Aug cloud could be explained by dissipation due to excessive rain out. Similar to SGP, SCREAM does very well with low clouds and underestimates high clouds. Like everywhere else, midlevel clouds are anemic. As we've seen elsewhere, near-surface wind speeds are generally too high, though the worst season for boundary-layer clouds - Aug - actually captures wind speed quite well.

5.4 Polar Regions

5.4.1 Cloud Phase Analysis

Cloud phase strongly influences cloud radiative forcing. Even under extremely cold conditions, mixed-phase clouds are often observed in high latitudes over both the Southern and Northern polar regions. Model skill in simulating the phase partitioning can have strong implications on simulating the strength of polar amplification in response to greenhouse-gas induced warming. Below, the GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP) Chepfer et al. (2010); Cesana and Chepfer (2013) is used as a reference to evaluate the SCREAM-simulated cloud fraction and the phase partitioning.

Typically CALIPSO-GOCCP is used to evaluate a model's COSP simulator output. As instrument simulators were not available for this SCREAM simulation, model cloud phase partitioning is constructed offline based on the relative ratio of liquid versus ice cloud condensate. A grid cell is considered 100% cloudy if total condensate amount exceeds 1e-5 kg/kg. Otherwise, it is cloud free. Among cloudy cells, if liquid condensate amount is at least 10 times greater than ice condensate, it is treated as liquid cloud. Similarly if ice condensate amount is 10 times greater or larger, it is labeled ice cloud. If neither phase dominates, it is considered mixed-phase cloud. Note that CALIPSO-GOCCP does not have a mixed phase mode. The comparison with CALIPSO-GOCCP uses the aggregated fraction for each cloud category over 1 degree x 1 degree area over the polar cap. Because our estimate of cloud phase and the CALIPSO-GOCCP definition differ, comparisons should be taken as approximate.

Figures 29 and 30 compare the zonal-vertical distribution of the simulated clouds against CALIPSO-GOCCP cloud climatology for identical time periods. Instead of a mixed-phase mode for CALIPSO-GOCCP, we include a composite over all the scenes where the classification algorithm is unable to determine whether cloud is liquid or ice. The fraction of time when there are clouds but phase can't be determined is significant, particularly near the surface. Thus we also include a comparison of total cloud fraction in the right-hand panel.

Over both polar regions, the presence of mixed-phase clouds can be viewed in terms of the 'mixed-phase' clouds at model native grids, and the coexistence of liquid and ice

phase clouds at coarsened grids. The model captures the vertical structure and general seasonality of cloudiness (not shown). Boundary-layer bound low clouds dominate, though at subpolar latitudes, the amount of middle-level and high-level clouds are also significant. Maximum cloudiness occurs in the respective summer to fall seasons, when liquid clouds extend further poleward. The discrepancies between model and observed clouds are substantial, even after accounting for the uncertainty in the CALIPSO-GOCCP cloud phase classification. Modeled liquid-phase clouds near both poles are more confined to lower levels compared to CALIPSO-GOCCP. Ice cloud fractions above the boundary layer are clearly higher in the model than seen in the CALIPSO-GOCCP data. SCREAM also has stronger latitudinal gradients in cloudiness than observed.

5.4.2 Polar ARM Sites

ARM has one long-term monitoring site in the Arctic: NSA. See Figure 8 for location. As shown in Figure 31, SCREAMv1 captures the low-level stratiform clouds extending to the surface in both Aug and Oct, though the low-cloud layer tends to be too thick in Aug and is largely missing in Apr. Mid-level clouds, e.g. between 3 and 6 km, are underestimated in all seasons except Aug, and in particular have an interesting and unrealistic minimum in mid-level clouds around 6 am for all seasons except Jan. In Figure 32, surface air temperature, surface mixing ratio, and Jan precipitation rate are all modestly but consistently underestimated. LHF is grossly overpredicted in Oct and is missing the observed diurnal variability in Aug (not shown). Note that these are the seasons of minimum sea ice extent at NSA and our strategy of averaging SCREAM data with centers inside a circle of radius 10 km is likely to include open ocean data that is missing from NSA point observations. This probably also explains why SHF is too weak in Aug and too large in Oct (not shown).

5.5 Process specific biases

The above analysis of the four seasonal SCREAMv1 simulations exposed significant biases in surface winds, surface temperature and in the cloud forcing. Investigation of the causes of these biases identified three configuration choices to be the most likely cause. A series of shorter, days-long simulations were conducted for each season with proposed solutions adopted. In all cases we were able to improve model accuracy. Due to a lack of resources we were unable to extend these three solutions to the full seasonal cycle. However, these fixes will be adopted in subsequent SCREAMv1 simulation campaigns. In this section we discuss the impact of each solution individually.

5.5.1 Surface Winds Bias

As seen in Figure 4 there is a significant overestimation of near surface winds observed for all four seasonal simulations. This bias is also observed in the SCREAMv0 simulations. The wind speed is consistently too high in the global average. Further analysis, shown in Figure 33, points to regions with topography as having the greatest bias. These results motivated us to rethink our omission of a Turbulent Mountain Stress (TMS) scheme in the SCREAMv0 and SCREAMv1 physics suites. The original hypothesis was that a 3km horizontal grid scale would be sufficiently fine enough for the model to naturally represent the impact of topography on wind speed. To test this hypothesis, a set of five–day simulations were conducted for each of the four seasons with a newly implemented TMS scheme. As shown in Figure 33 the inclusion of TMS decreased the wind speed over land, especially near topography, improving the overall bias in the global simulations. These results point to the need for some parameterized mountain stress, even at the 3km grid scale.

5.5.2 Warm Temperature Bias

The top panel of Figure 34 shows that warm bias occurs primarily over land (as expected when SSTs are prescribed). Differences can be larger than 5 K over midlatitude land regions. Figure 10 shows that warm bias is largest in the lower and upper troposphere. This bias turns out to be at least partially due to earlier attempts to prevent runaway cold snaps by maintaining SHOC mixing of warm air from above even in cases of strong thermal stability. Further analysis of the cold snaps after these simulations revealed that the problem was actually due to the dycore creating new local minima in temperature. The addition of TMS caused this problem to go away, allowing us to return SHOC to more realistic behavior in stable boundary layer cases. The impact of the default SHOC diffusivities on near-surface temperature is illustrated in the bottom panel of Figure 34. A set of six–day simulations were conducted with the original configuration of SHOC. As seen in Figure 34 the greatest impact is over land, and in the correct direction to improve the warm temperature bias. Over ocean the results are more muted, which is not surprising given the prescribed SSTs.

5.5.3 Cloud Biases

Recall in Figure 6 that SCREAMv1 tends to over–predict the shortwave cloud forcing (SWCF) in the tropics and subtropics and under–predict SWCF in the high latitiudes. A similar result is shown in Figure 7 with respects to longwave cloud forcing (LWCF). In order to address these biases we focused on two tuning parameters that impact clouds, $lambda\_high$ in SHOC and max_total_ni in P3. The former helps reduce mixing in areas of high vertical stability. Thus, higher $lambda\_high$ values will reduce the mixing and help reduce dry air entrainment at stratocumulus top. This should in turn produce more abundant subtropical stratocumulus. The max_total_ni parameter imposes a maximum total ice number concentration. Increasing this parameter will allow for higher concentration in ice clouds, impacting cloud forcing at high-latitudes.

A six–day simulation was conducted for each season with $lambda\_high$ increased to 0.08 from 0.04 and with max_total_ni increased to $740{\cdot}10^3$ from $500{\cdot}10^3$. Figure 35 reveals an overall improvement in the short– and long–wave cloud forcing biases. In the tropics we observe increased solar reflection and more longwave trapping. Shortwave changes are especially prevalent along Western coastal regions due to increased stratocumulus.

In the high–latitudes, where the increased ice number concentration threshold is expected to make the greatest impact we see the reverse pattern with less solar reflection and decreased longwave trapping, as desired.

6 Conclusions

It is clear that the future of high performance computing is in heterogeneous CPU/GPU systems. In order to stay at the cutting edge of science global climate models need to adapt to these architectures. EAMxx is a complete rewrite of the Energy Exascale Earth System Model's atmosphere component. EAMxx saw great success in the adoption of a performance portable library such as Kokkos, making it possible to not only deploy EAMxx on two of the fastest DOE systems, Summit and Frontier, relatively quickly, but to take full advantage of the GPU architectures of these machines and accomplish greater than one simulated-year-per-day throughput on a 3km global grid. Surprisingly, it only took the equivalent of 6 full–time employees approximately 4.5 years to accomplish this. This was facilitated by the early decision to establish a robust unit testing framework that allowed the team to develop multiple features concurrently and be confident that model integrity was maintained. Furthermore, a bottom up redesign of the code allowed the team to abandon support for complicated and often unused legacy code. Although

an imperfect metric, a comparison of total lines of Fortran code in EAM to lines of C++ code in EAMxx shows a reduction by a full order of magnitude. The result is the first global km–scale atmosphere model to run at exascale.

In tandem with the software development of EAMxx, the team worked on the validation of the SCREAM physics package using the existing F90 EAM codebase, see C21 (Caldwell et al., 2021) for more details. Once the more performant EAMxx model was ready we were able to extend the original 3km analysis of C21. Global and regional-scale analysis across all four seasons revealed that SCREAMv1 captures the global net top of atmosphere imbalance to within 1.2 W m^{-2} and represents the mean characteristics of the midlatitude jets and the atmospheric rivers that are responsible for much of the poleward vapor flux. Many of the differences between SCREAM and observational products that were shown in C21 are shown here to span all four seasons, including excessively warm near-surface temperatures, overly-strong surface winds, a over-predicted global mean precipitation rate, and a mid-tropospheric dry bias. Despite seasonal variations, TOA radiative fluxes display a consistent story of too much absorbed shortwave in the tropics and subtropics that is balanced by too much reflected shortwave at high latitudes and too much outgoing longwave radiation. We have shown that a couple of these biases can be ameliorated with additional physics or changes in the tunings, including the strong near-surface wind speeds, strong outgoing longwave radiation, weak reflection of the stratocumulus, and warm near-surface temperatures over land. However, certain biases persist. In the Tropics there is a general tendency of the precipitation patterns to organize into narrow strips compared to satellite estimates. Additionally there is a dry bias in the lower troposphere above the boundary layer, which coincides with a lack of mid-level clouds and deep convective cloud tops, consistent with previous findings (Turbeville et al., 2022). Outside of the Tropics, biases also exist in the representation of cloud height and phase over both polar region, with liquid-phase clouds being more excessive and more confined to lower levels while having more ice-phase clouds aloft. What further changes are needed to the physics suite to address these biases is the subject of ongoing study.

Higher resolution simulations lend themselves well to comparison with ground-based observation stations. ARM's simultaneous co-located surface measurements and vertical profiling of cloud fraction provide a unique perspective for understanding the convective behaviors in responding to large scale environments and surface forcings. When compared with ARM data, SCREAMv1 showed good performance in capturing the diurnal development of boundary layer clouds across all 4 seasons. SCREAMv1 was found to be deficient in mid-level clouds, especially in the tropics. This is accompanied by a relatively warmer and drier mid-troposphere. SCREAMv1 tends to overestimate tropical high clouds. Both of these features might be attributed to an overly-quick transition from shallow to deep convection without the gradual intermediate stage of congestus clouds. Some common biases in SCREAMv1 shared across ARM sites in different seasons and different cloud regimes are: 1) significantly overestimated 10-meter wind speeds; 2) overestimated downward shortwave radiation (except NSA); 3) significantly underestimated column integrated liquid water path (except NSA); 4) underestimated column integrated precipitable water vapor (except Jan/Oct at SGP).

SCREAMv1 is exciting because it allows us to perform global storm-resolving simulations much faster than previous efforts by taking advantage of the GPUs on the world's fastest supercomputers. We have also shown that despite approaching convection–permitting spatial scales, there is still work to be done addressing persistent biases in our simulations. This is unsurprising because the runs shown are the first ever done with SCREAMv1, in contrast to the decades of development that have gone into conventional global climate models. In this context, the model skill shown here is a testament to the power of resolving the scales of motion between 10 and 100 km. Future work will continue to investigate these biases as well as leverage EAMxx performance to investigate longer simulations.

Open Research Section

All ARM observational data is publicly available and can be found at https://portal.nersc.gov/project/capt/SCREAM
All simulation data from the SCREAMv1 simulations can be accessed from NERSC via Globus (http://globus.org/) from the E3SM_SCREAMv1_4SeasonsRuns2024 Collection.

The version of the SCREAM codebase used for all full length simulations in this manuscript is available via Zenodo at the following DOI: 10.5281/zenodo.10724808

The code adjustments made to examine the impact of changing certain processes in the model discussed in section 5.5 is available via Zenodo at the following DOI's: https://doi.org/10.5281/zenodo.1072485 https://doi.org/10.5281/zenodo.10724859, and https://doi.org/10.5281/zenodo.10724864. Performance data, run scripts, and code for Table 1 and Figs. 2 and 3 are available at `https://github.com/E3SM-Project/perf-data` and are archived at `https://zenodo.org/records/10714334`. Within this repository, see `screamv1-frontier-feb2023/readme.txt` and `screamv1-summit-oct2022/readme.txt` for details.

The GPM-IMERG dataset used in this study is Version 06B Level 3 daily 0.1 degree x 0.1 degree final-run gauge-calibrated data as described in Huffman and coauthors (2019) and accessible via `https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_06/summary`. The version of CERES SYN1deg data used here is Edition 4.1 with a release date of August 22, 2019 and was accessed from: `https://ceres-tool.larc.nasa.gov/ord-tool/jsp/SYN1degEd41Selection.jsp`. This study also used ERA5 hourly data on both pressure levels (Hersbach et al., 2023a) and on single levels (Hersbach et al., 2023b) for evaluation, which can be downloaded from the Copernicus Climate Change Service (2023).

MODIS satellite data is available publicly at https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MYD06_L2#data-availability.

CALIPSO–GOCCP data was obtained from https://climserv.ipsl.polytechnique.fr/cfmip-obs/Calipso_goccp.html.

Acknowledgments

This research was supported as part of the Energy Exascale Earth System Model (E3SM) project (https://e3sm.org/), funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

Data were used from the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Climate Research Facilities at Southern Great Plains (SGP), Eastern North Atlantic (ENA), North Slope of Alaska (NSA), Tropical Western Pacific (TWP), and the field campaign Green Ocean Amazon (GoAmazon).

Yunyan Zhang was supported by the US DOE Atmospheric Systems Research (ASR) program Tying in High Resolution E3SM with ARM Data (THREAD) project. Jingjing Tian was supported by the US DOE Early Career Research Project awarded to Y. Zhang.

This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of two U.S. Department of Energy organizations (Office of Science and the National Nuclear Security Administration) responsible for the planning and preparation of a capable exascale ecosystem, including software, applications, hardware, advanced system engineering, and early testbed platforms, in support of the nation's exascale computing imperative.

This paper was prepared under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. IM Release number LLNL-JRNL-859924-DRAFT.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government

This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

An award of computer time was also provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program, performed using resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RLO1830.

The Work at BNL was supported by the Energy Exascale Earth System Model (E3SM) project funded by Department of Energy under contract DE-SC0012704.

Argonne National Laboratory's work was supported by the U.S. Department of Energy, Assistant Secretary for Environmental Management, Office of Science, under contract DE-AC02-06CH11357.

References

ACM Gordon Bell Prize in Climate Modeling. (2023). Association for Computing Machinery. Retrieved from `https://awards.acm.org/bell-climate` (Accessed: 2023-11-28)

Afanasyev, A., Bianco, M., Mosimann, L., Osuna, C., Thaler, F., Vogt, H., ... Schulthess, T. C. (2021). Gridtools: A framework for portable weather and climate applications. *SoftwareX*, *15*, 100707. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2352711021000522` doi: https://doi.org/10.1016/j.softx.2021.100707

Barnes, E. A., & Polvani, L. (2013). Response of the midlatitude jets, and of their variability, to increased greenhouse gases in the cmip5 models. *Journal of Climate*, *26*(18), 7117 - 7135. Retrieved from `https://journals.ametsoc.org/view/journals/clim/26/18/jcli-d-12-00536.1.xml` doi: https://doi.org/10.1175/JCLI-D-12-00536.1

Beckingsale, D. A., Burmark, J., Hornung, R., Jones, H., Killian, W., Kunen, A. J., ... Scogland, T. R. (2019). Raja: Portable performance for large-scale scientific applications. In *2019 ieee/acm international workshop on performance, portability and productivity in hpc (p3hpc)* (p. 71-81). doi: 10.1109/P3HPC49587.2019.00012

Bertagna, L., Deakin, M., Guba, O., Sunderland, D., Bradley, A. M., Tezaur, I. K., ... Salinger, A. G. (2019). Hommexx 1.0: a performance-portable atmospheric dynamical core for the energy exascale earth system model. *Geoscientific Model Development*, *12*(4), 1423–1441. Retrieved from https://gmd.copernicus.org/articles/12/1423/2019/ doi: 10.5194/gmd-12-1423-2019

Bertagna, L., Guba, O., Taylor, M. A., Foucar, J. G., Larkin, J., Bradley, A. M., ... Salinger, A. G. (2020). A performance-portable nonhydrostatic atmospheric dycore for the energy exascale earth system model running at cloud-resolving resolutions. In *Sc20: International conference for high performance computing, networking, storage and analysis* (p. 1-14). doi: 10.1109/SC41405.2020.00096

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., ... John, V. O. (2011). Cosp: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023 - 1043. Retrieved from https://journals.ametsoc.org/view/journals/bams/92/8/2011bams2856_1.xml doi: https://doi.org/10.1175/2011BAMS2856.1

Bogenschutz, P. A., & Krueger, S. K. (2013). A simplified pdf parameterization of subgrid-scale clouds and turbulence for cloud-resolving models. *Journal of Advances in Modeling Earth Systems*, *5*(2), 195-211. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jame.20018 doi: https://doi.org/10.1002/jame.20018

Caldwell, P. M., Terai, C. R., Hillman, B., Keen, N. D., Bogenschutz, P., Lin, W., ... others (2021). Convection-permitting simulations with the e3sm global atmosphere model. *Journal of Advances in Modeling Earth Systems*, *13*(11), e2021MS002544.

Carbone, R. E., & Tuttle, J. D. (2008). Rainfall occurrence in the u.s. warm season: The diurnal cycle. *Journal of Climate*, *21*(16), 4132 - 4146. Retrieved from https://journals.ametsoc.org/view/journals/clim/21/16/2008jcli2275.1.xml doi: https://doi.org/10.1175/2008JCLI2275.1

Carbone, R. E., Tuttle, J. D., Ahijevych, D., & Trier, S. B. (2002). *Inferences of predictability associated with warm season precipitation episodes.* (Vol. 59).

Cesana, G., & Chepfer, H. (2013). Evaluation of the cloud thermodynamic phase in a climate model using calipso-goccp. *Journal of Geophysical Research: Atmosphere*, *118*, 7922 - 7937. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/10.1002/jgrd.50376 doi: https:doi:10.1002/jgrd.50376

Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Ghan, S. J., Kiehl, J. T., ... Yagai, I. (1989). Interpretation of cloud-climate feedback as produced by 14 atmospheric general circulation models. *Science*, *245*(4917), 513-516. Retrieved from https://www.science.org/doi/abs/10.1126/science.245.4917.513 doi: 10.1126/science.245.4917.513

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J. L., Minnis, P., ... Zeng, S. (2010). The gcm oriented calipso cloud product (calipso-goccp). *Journal of Geophysical Research: Atmosphere*, *115*, D00H16. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2009JD012251 doi: https://doi.org/10.1029/2009JD012251

Chinita, M. J., Witte, M., Kurowski, M. J., Teixeira, J., Suselj, K., Matheou, G., & Bogenschutz, P. (2023). Improving the representation of shallow cumulus convection with the simplified-higher-order-closure–mass-flux (shoc+mf v1.0) approach. *Geoscientific Model Development*, *16*(7), 1909-1924. doi: 10.5194/gmd-16-1909-2023

Clement, V., Ferrachat, S., Fuhrer, O., Lapillonne, X., Osuna, C. E., Pincus, R., ... Sawyer, W. (2018). The CLAW DSL: Abstractions for performance portable weather and climate models. In *Proceedings of the platform for advanced scientific computing conference, pasc '18.* ACM. Retrieved from http://doi.acm.org/10.1145/3218176.3218226 doi:

10.1145/3218176.3218226

Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T., Miller, M. A., & Martner, B. E. (2000). Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the arm cart sites. *Journal of Applied Meteorology and Climatology*, *39*(5), 645–665.

Collier, J. C., & Bowman, K. P. (2004). Diurnal cycle of tropical precipitation in a general circulation model. *Journal of Geophysical Research: Atmospheres*, *109*(D17). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD004818 doi: https://doi.org/10.1029/2004JD004818

Craig, A. P., Vertenstein, M., & Jacob, R. (2012). A new flexible coupler for earth system modeling developed for CCSM4 and CESM1. *ijhpca*, *26*, 31–42. doi: 10.1177/1094342011428141

Dennis, J., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P., Mirin, A. A., ... Worley, P. H. (2012). CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. *Int. J. High Perform. Comput. Appl.*, *26*, 74–89.

Dennis, J., Fournier, A., Spotz, W. F., St-Cyr, A., Taylor, M. A., Thomas, S. J., & Tufo, H. (2005). High-resolution mesh convergence properties and parallel efficiency of a spectral element atmospheric dynamical core. *Int. J. High Perform. Comput. Appl.*, *19*(3), 225–235.

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., ... Sun, M. (2013). Geostationary enhanced temporal interpolation for ceres flux products. *Journal of Atmospheric and Oceanic Technology*, *30*(6), 1072 - 1090. Retrieved from https://journals.ametsoc.org/view/journals/atot/30/6/jtech-d-12-00136_1.xml doi: https://doi.org/10.1175/JTECH-D-12-00136.1

Donahue, A. S., & Caldwell, P. M. (2018). Impact of physics parameterization ordering in a global atmosphere model. *Journal of Advances in Modeling Earth Systems*, *10*(2), 481-499. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017MS001067 doi: https://doi.org/10.1002/2017MS001067

Feng, Z., Leung, L. R., Houze Jr., R. A., Hagos, S., Hardin, J., Yang, Q., ... Fan, J. (2018). Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection-permitting simulations over the united states. *Journal of Advances in Modeling Earth Systems*, *10*(7), 1470-1494. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001305 doi: https://doi.org/10.1029/2018MS001305

Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., ... Bader, D. C. (2022). The doe e3sm model version 2: Overview of the physical model and initial model evaluation. *Journal of Advances in Modeling Earth Systems*, *14*(12), e2022MS003156. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003156 (e2022MS003156 2022MS003156) doi: https://doi.org/10.1029/2022MS003156

Hamilton, K. (2020). At the dawn of global climate modeling: the strange case of the leith atmosphere model. *History of Geo- and Space Sciences*, *11*(1), 93–103. Retrieved from https://hgss.copernicus.org/articles/11/93/2020/ doi: 10.5194/hgss-11-93-2020

Hannah, W. M., Bradley, A. M., Guba, O., Tang, Q., Golaz, J.-C., & Wolfe, J. (2021). Separating Physics and Dynamics Grids for Improved Computational Efficiency in Spectral Element Earth System Models. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2020MS002419. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002419 doi: https://doi.org/10.1029/2020MS002419

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., ... Zhang, S. (2005). Efficacy of climate forcings. *Journal of Geophysical Research: Atmospheres*, *110*(D18). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD005776 doi: https://doi.org/10.1029/2005JD005776

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., ... Thépaut, J.-N. (2023a). *Era5 hourly data on pressure levels from 1940 to present.* (Accessed on 06-02-2023) doi: 10.24381/cds.bd0915c6

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., ... Thépaut, J.-N. (2023b). *Era5 hourly data on single levels from 1940 to present.* (Accessed on 06-02-2023) doi: 10.24381/cds.bd0915c6

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Hu, H., Feng, Z., & Leung, L.-Y. R. (2021). Linking flood frequency with mesoscale convective systems in the us. *Geophysical Research Letters*, *48*(9), e2021GL092546. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL092546 (e2021GL092546 2021GL092546) doi: https://doi.org/10.1029/2021GL092546

Huffman, G. J., & coauthors. (2019). *Nasa global precipitation measurement (gpm) integrated multi-satellite retrievals for gpm (imerg)* (NASA Algorithm Theoretical Basis Doc. No. version 06). Greenbelt MD, USA: NASA.

Inda-Díaz, H. A., O'Brien, T. A., Zhou, Y., & Collins, W. D. (2021). Constraining and characterizing the size of atmospheric rivers: A perspective independent from the detection algorithm. *Journal of Geophysical Research: Atmospheres*, *126*(16), e2020JD033746.

Ingram, W., & Bushell, A. C. (2021). Sensitivity of climate feedbacks to vertical resolution in a general circulation model. *Geophysical Research Letters*, *48*(12), e2020GL092268.

Johnson, D., & Arakawa, A. (1996). On the scientific contributions and insight of professor yale mintz. *Journal of Climate*, *9*(12), 3211 - 3224. Retrieved from https://journals.ametsoc.org/view/journals/clim/9/12/1520-0442_1996_009_3211_otscai_2_0_co_2.xml doi: https://doi.org/10.1175/1520-0442(1996)009<3211:OTSCAI>2.0.CO;2

Johnson, R. H., Rickenbach, T. M., Rutledge, S. A., Ciesielski, P. E., & Schubert, W. H. (1999). Trimodal characteristics of tropical convection. *Journal of Climate*, *12*(8), 2397 - 2418. Retrieved from https://journals.ametsoc.org/view/journals/clim/12/8/1520-0442_1999_012_2397_tcotc_2.0.co_2.xml doi: https://doi.org/10.1175/1520-0442(1999)012<2397:TCOTC>2.0.CO;2

JUDT, F., KLOCKE, D., RIOS-BERRIOS, R., VANNIERE, B., ZIEMEN, F., AUGER, L., ... ZHOU, L. (2021). Tropical cyclones in global storm-resolving models. *Journal of the Meteorological Society of Japan. Ser. II*, *99*(3), 579-602. doi: 10.2151/jmsj.2021-029

Kay, J. E., Hillman, B. R., Klein, S. A., Zhang, Y., Medeiros, B., Pincus, R., ... Ackerman, T. P. (2012). Exposing global cloud biases in the community atmosphere model (cam) using satellite observations and their corresponding instrument simulators. *Journal of Climate*, *25*(15), 5190 - 5207. Retrieved from https://journals.ametsoc.org/view/journals/clim/25/15/jcli-d-11-00469.1.xml doi: https://doi.org/10.1175/JCLI-D-11-00469.1

Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., & Gleckler, P. J. (2013). Are climate model simulations of clouds improving? an evaluation using the isccp simulator. *Journal of Geophysical Research: Atmospheres*, *118*(3), 1329-1342. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50141 doi: https://doi.org/10.1002/

jgrd.50141

Kodama, C., Noda, A. T., & Satoh, M. (2012). An assessment of the cloud signals simulated by nicam using isccp, calipso, and cloudsat satellite simulators. *Journal of Geophysical Research: Atmospheres*, *117*(D12). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD017317 doi: https://doi.org/10.1029/2011JD017317

Kollias, P., Miller, M. A., Luke, E. P., Johnson, K. L., Clothiaux, E. E., Moran, K. P., ... Albrecht, B. A. (2007). The atmospheric radiation measurement program cloud profiling radars: Second-generation sampling strategies, processing, and cloud data products. *Journal of Atmospheric and Oceanic Technology*, *24*(7), 1199–1214.

Kooperman, G. J., Pritchard, M. S., Ghan, S. J., Wang, M., Somerville, R. C. J., & Russell, L. M. (2012). Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the community atmosphere model 5. *Journal of Geophysical Research: Atmospheres*, *117*(D23). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012JD018588 doi: https://doi.org/10.1029/2012JD018588

Lin, G., Jones, C. R., Leung, L. R., Feng, Z., & Ovchinnikov, M. (2022). Mesoscale convective systems in a superparameterized e3sm simulation at high resolution. *Journal of Advances in Modeling Earth Systems*, *14*(1), e2021MS002660. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002660 (e2021MS002660 2021MS002660) doi: https://doi.org/10.1029/2021MS002660

Liu, W., Ullrich, P. A., Li, J., Zarzycki, C., Caldwell, P. M., Leung, L. R., & Qian, Y. (2023). The june 2012 north american derecho: A testbed for evaluating regional and global climate modeling systems at cloud-resolving scales. *Journal of Advances in Modeling Earth Systems*, *15*(4), e2022MS003358. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003358 (e2022MS003358 2022MS003358) doi: https://doi.org/10.1029/2022MS003358

Manabe, S., Smagorinsky, J., & Strickler, R. F. (1965). Simulated climatology of a general circulation model with a hydrologic cycle. *Monthly Weather Review*, *93*(12), 769 - 798. Retrieved from https://journals.ametsoc.org/view/journals/mwre/93/12/1520-0493_1965_093_0769_scoagc_2_3_co_2.xml doi: https://doi.org/10.1175/1520-0493(1965)093<0769:SCOAGC>2.3.CO;2

Medeiros, B., Shaw, J., Kay, J. E., & Davis, I. (2023). Assessing clouds using satellite observations through three generations of global atmosphere models. *Earth and Space Science*, *10*(7), e2023EA002918. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023EA002918 (e2023EA002918 2023EA002918) doi: https://doi.org/10.1029/2023EA002918

Morrison, H., & Milbrandt, J. A. (2015). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. part i: Scheme description and idealized tests. *Journal of the Atmospheric Sciences*, *72*(1), 287 - 311. Retrieved from https://journals.ametsoc.org/view/journals/atsc/72/1/jas-d-14-0065.1.xml doi: https://doi.org/10.1175/JAS-D-14-0065.1

Norman, M., Lyngaas, I., Bagusetty, A., & Berrill, M. (2022, Dec 08). Portable c++ code that can look and feel like fortran code with yet another kernel launcher (yakl). *International Journal of Parallel Programming*. Retrieved from https://doi.org/10.1007/s10766-022-00739-0 doi: 10.1007/s10766-022-00739-0

The OpenACC application programming interface version 3.0. (2019). Retrieved from https://www.openacc.org/sites/default/files/inline-images/Specification/OpenACC.3.0.pdf

OpenMP application programming interface version 5.0. (2018). Retrieved from https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5.0.pdf

Pincus, R., Mlawer, E. J., & Delamere, J. S. (2019). Balancing accuracy, efficiency, and flexibility in radiation calculations for dynamical models. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3074-3089. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001621 doi: https://doi.org/10.1029/2019MS001621

Pritchard, M. S., & Somerville, R. C. J. (2009). Assessing the diurnal cycle of precipitation in a multi-scale climate model. *Journal of Advances in Modeling Earth Systems*, *1*(4). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.3894/JAMES.2009.1.12 doi: https://doi.org/10.3894/JAMES.2009.1.12

Psyclone user guide. (n.d.). Retrieved from https://psyclone.readthedocs.io/en/stable/index.html (Accessed: 2023-02-13)

Qin, Y., Zelinka, M. D., & Klein, S. A. (2022). On the correspondence between atmosphere-only and coupled simulations for radiative feedbacks and forcing from co2. *Journal of Geophysical Research: Atmospheres*, *127*(3), e2021JD035460. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JD035460 (e2021JD035460 2021JD035460) doi: https://doi.org/10.1029/2021JD035460

Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547 - 1564. Retrieved from https://journals.ametsoc.org/view/journals/bams/84/11/bams-84-11-1547.xml doi: https://doi.org/10.1175/BAMS-84-11-1547

Ringer, M. A., Andrews, T., & Webb, M. J. (2014). Global-mean radiative feedbacks and forcing in atmosphere-only and coupled atmosphere-ocean climate change experiments. *Geophysical Research Letters*, *41*(11), 4035-4042. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL060347 doi: https://doi.org/10.1002/2014GL060347

Sanderson, B., Piani, C., Ingram, W., Stone, D., & Allen, M. (2008). Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics gcm simulations. *Clim. Dyn.*, *30*, 175-199.

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., & Iga, S. (2008). Nonhydrostatic icosahedral atmospheric model (nicam) for global cloud resolving simulations. *Journal of Computational Physics*, *227*(7), 3486-3514. Retrieved from https://www.sciencedirect.com/science/article/pii/S0021999107000654 (Predicting weather, climate and extreme events) doi: https://doi.org/10.1016/j.jcp.2007.02.006

Schulte, R. M., Lebsock, M. D., & Haynes, J. M. (2023). What cloudsat cannot see: liquid water content profiles inferred from modis and caliop observations. *Atmospheric Measurement Techniques*, *16*(14), 3531–3546. Retrieved from https://amt.copernicus.org/articles/16/3531/2023/ doi: 10.5194/amt-16-3531-2023

Schumacher, R. S., & Johnson, R. H. (2006). Characteristics of u.s. extreme rain events during 1999–2003. *Weather and Forecasting*, *21*(1), 69 - 85. Retrieved from https://journals.ametsoc.org/view/journals/wefo/21/1/waf900_1.xml doi: https://doi.org/10.1175/WAF900.1

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., ... others (2018). Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic change*, *151*, 555–571.

Sherwood, S., Bony, S., & Dufresne, J. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, *505*, 37-42.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., ... Mauritsen, T. (2017). Macv2-sp: a parameterization of anthropogenic aerosol optical properties and an associated twomey effect for use in cmip6. *Geoscientific Model Development*, *10*(1), 433–452. Retrieved from `https://gmd.copernicus.org/articles/10/433/2017/` doi: 10.5194/gmd-10-433-2017

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., ... others (2019). Dyamond: the dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, *6*(1), 1–17.

Stevenson, S. N., & Schumacher, R. S. (2014). A 10-year survey of extreme rainfall events in the central and eastern united states using gridded multisensor precipitation analyses. *Monthly Weather Review*, *142*(9), 3147 - 3162. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/142/9/mwr-d-13-00345.1.xml` doi: https://doi.org/10.1175/MWR-D-13-00345.1

Taylor, M., Caldwell, P. M., Bertagna, L., Clevenger, C., Donahue, A., Foucar, J., ... Wu, D. (2023). The simple cloud-resolving e3sm atmosphere model running on the frontier exascale system. In *Proceedings of the international conference for high performance computing, networking, storage and analysis.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3581784.3627044` doi: 10.1145/3581784.3627044

Taylor, M. A., Guba, O., Steyer, A., Ullrich, P. A., Hall, D. M., & Eldrid, C. (2020). An energy consistent discretization of the nonhydrostatic equations in primitive variables. *J. Adv. Model Earth Syst.*, *12*(1).

Tian, Y., Zhang, Y., & Klein, S. A. (2022). What determines the number and the timing of pulses in afternoon precipitation in the green ocean amazon (goamazon) observations? *Geophysical Research Letters*, *49*(2), e2021GL096075.

Tian, Y., Zhang, Y., Klein, S. A., & Schumacher, C. (2021). Interpreting the diurnal cycle of clouds and precipitation in the arm goamazon observations: Shallow to deep convection transition. *Journal of Geophysical Research: Atmospheres*, *126*(5), e2020JD033766.

Tomita, H., Miura, H., Iga, S., Nasuno, T., & Satoh, M. (2005). A global cloud-resolving simulation: Preliminary results from an aqua planet experiment. *Geophysical Research Letters*, *32*(8). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL022459` doi: https://doi.org/10.1029/2005GL022459

Tomita, H., & Satoh, M. (2004, jun). A new dynamical framework of nonhydrostatic global model using the icosahedral grid. *Fluid Dynamics Research*, *34*(6), 357. Retrieved from `https://dx.doi.org/10.1016/j.fluiddyn.2004.03.003` doi: 10.1016/j.fluiddyn.2004.03.003

Trilinos Project Team, T. (2020 (acccessed May 22, 2020)). The Trilinos Project Website [Computer software manual]. Retrieved from `https://trilinos.github.io`

Trott, C. R., Lebrun-Grandié, D., Arndt, D., Ciesko, J., Dang, V., Ellingwood, N., ... Wilke, J. (2022). Kokkos 3: Programming model extensions for the exascale era. *IEEE Transactions on Parallel and Distributed Systems*, *33*(4), 805-817. doi: 10.1109/TPDS.2021.3097283

Turbeville, S. M., Nugent, J. M., Ackerman, T. P., Bretherton, C. S., & Blossey, P. N. (2022). Tropical cirrus in global storm-resolving models: 2. cirrus life cycle and top-of-atmosphere radiative fluxes. *Earth and Space Science*, *9*(2), e2021EA001978. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EA001978` (e2021EA001978 2021EA001978) doi: https://doi.org/10.1029/2021EA001978

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). TempestExtremes v2.1: a community framework

for feature detection, tracking and analysis in large datasets. *Geoscientific Model Development Discussions*, 1–37. Retrieved from `https://github.com/ClimateGlobalChange/tempestextremes`

Wedi, N. P., Polichtchouk, I., Dueben, P., Anantharaj, V. G., Bauer, P., Boussetta, S., ... Vitart, F. (2020). A baseline for global weather and climate simulations at 1 km resolution. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002192. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002192` (e2020MS002192 10.1029/2020MS002192) doi: https://doi.org/10.1029/2020MS002192

Xie, S., Ma, H.-Y., Boyle, J. S., Klein, S. A., & Zhang, Y. (2012). On the correspondence between short- and long-time-scale systematic errors in cam4/cam5 for the year of tropical convection. *Journal of Climate*, *25*(22), 7937 - 7955. Retrieved from `https://journals.ametsoc.org/view/journals/clim/25/22/jcli-d-12-00134.1.xml` doi: https://doi.org/10.1175/JCLI-D-12-00134.1

Xie, S., McCoy, R. B., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Clothiaux, E. E., ... others (2010). Arm climate modeling best estimate data: A new data product for climate studies. *Bulletin of the American Meteorological Society*, *91*(1), 13–21.

Xie, S., Wang, Y.-C., Lin, W., Ma, H.-Y., Tang, Q., Tang, S., ... Zhang, M. (2019). Improved diurnal cycle of precipitation in e3sm with a revised convective triggering function. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2290-2310. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001702` doi: https://doi.org/10.1029/2019MS001702

Zhang, Y., & Klein, S. A. (2010). Mechanisms affecting the transition from shallow to deep convection over land: Inferences from observations of the diurnal cycle collected at the arm southern great plains site. *Journal of the Atmospheric Sciences*, *67*(9), 2943–2959.

Zhang, Y., & Klein, S. A. (2013). Factors controlling the vertical extent of fair-weather shallow cumulus clouds over land: Investigation of diurnal-cycle observations collected at the arm southern great plains site. *Journal of the Atmospheric Sciences*, *70*(4), 1297–1315.

Zhang, Y., Klein, S. A., Fan, J., Chandra, A. S., Kollias, P., Xie, S., & Tang, S. (2017). Large-eddy simulation of shallow cumulus over land: A composite case based on arm long-term observations at its southern great plains site. *Journal of the Atmospheric Sciences*, *74*(10), 3229–3251.

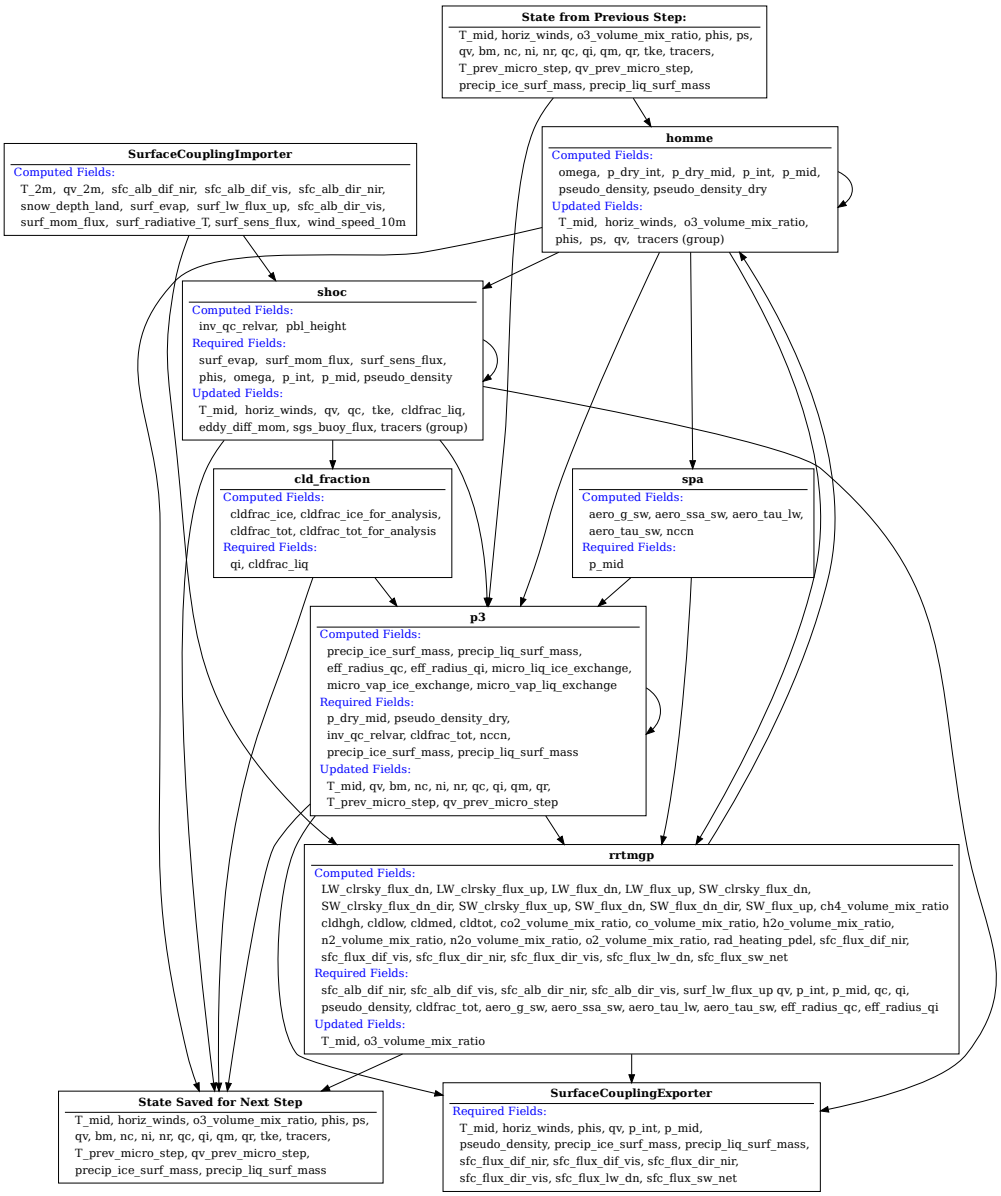


Figure 1. Directed acyclic graph (DAG) showing EAMxx processes and the variables passed between them.

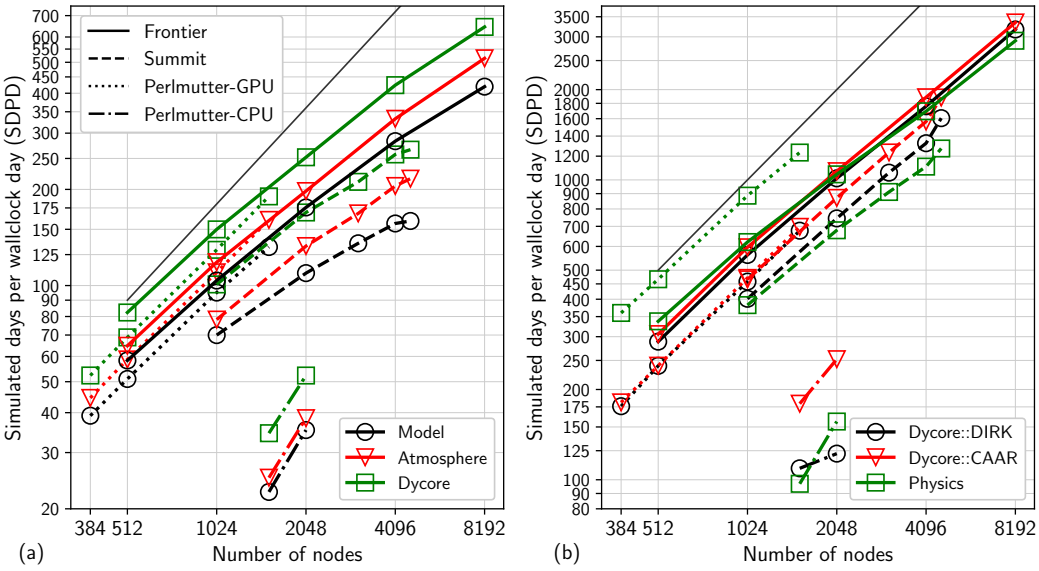


Figure 2. Strong scaling of the model. (a) Total model ("Model"), atmosphere model ("Atmosphere"), and dynamical core ("Dycore") throughput in simulated days per wallclock day as a function of number of Frontier nodes (solid lines), Perlmutter GPU nodes (dotted lines), Summit nodes (dashed lines), and Perlmutter CPU nodes (dash-dot lines). The black undecorated line is the perfect-scaling reference. (b) Physics parameterizations ("Physics") and vertical IMEX solver ("Dycore::DIRK") throughput. These subcomponents have no MPI communication.

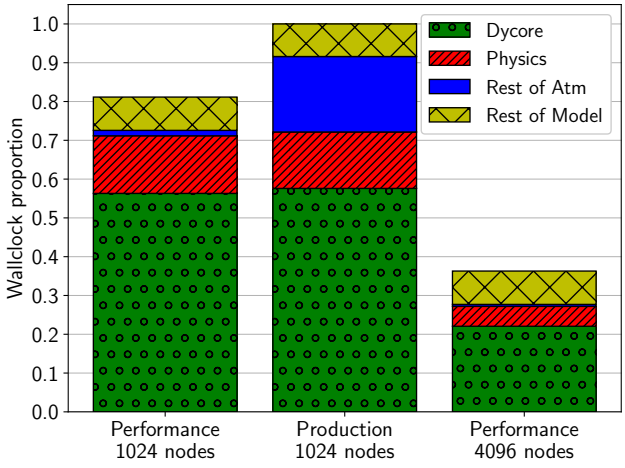


Figure 3. Proportion of wallclock time spent in certain subcomponents, referenced to a production run. The production run is the reference and, thus, has total wallclock time set to 1. All data in this figure are for simulations performed on Summit.

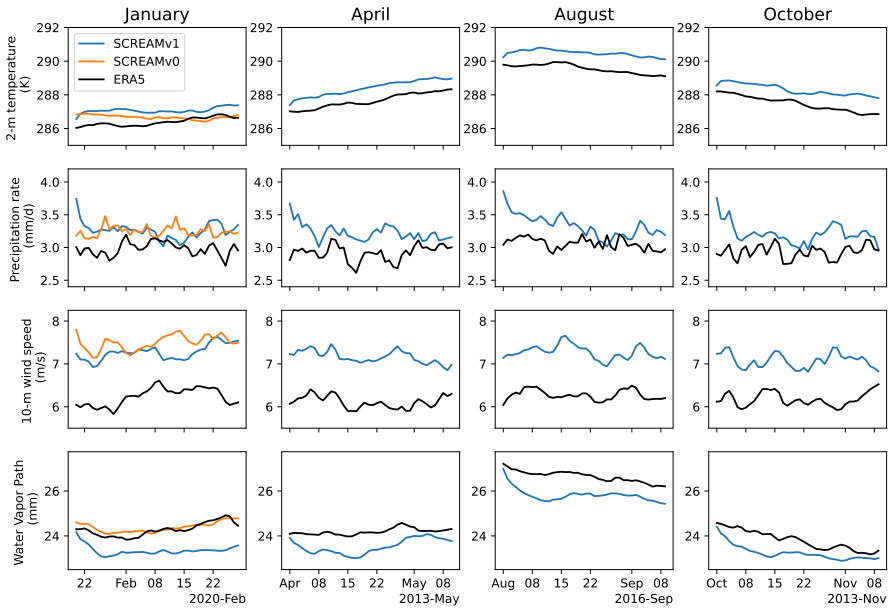


Figure 4. Time series of global-mean variables (listed on left side of each row) from SCREAMv1 and ERA5 for each season (listed at bottom of each column). For Jan, results from SCREAMv0 are also shown.

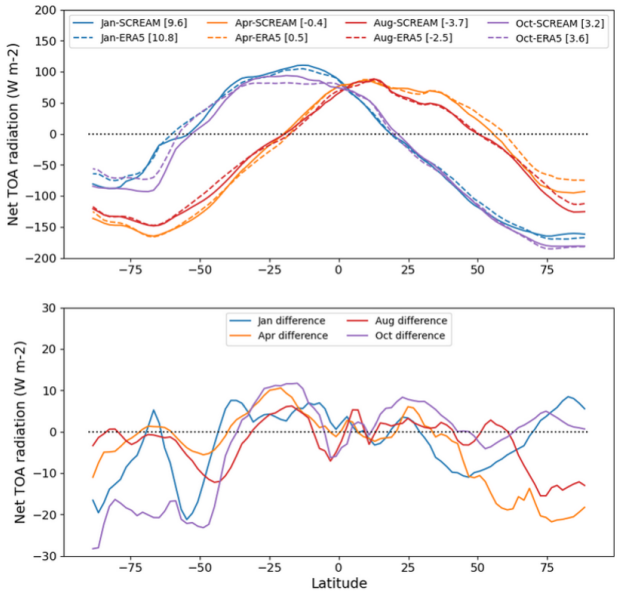


Figure 5. (top) Zonal mean net top-of-atmosphere radiative fluxes (positive downwards) in SCREAM (solid lines) and from CERES-SYN (dashed lines) from the last 30 days of the 40 day Jan (blue), Apr (orange), Aug (red), and Oct (purple) simulations. (bottom) The difference between SCREAMv1 and CERES-SYN.

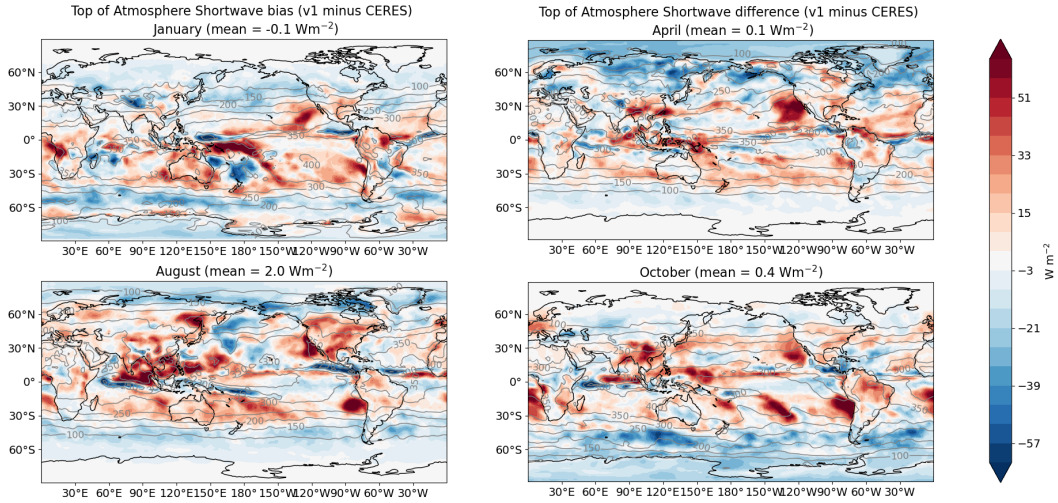


Figure 6. Time-mean differences in the net shortwave radiation over the last 30 days of the simulation between SCREAMv1 and CERES-SYN daily-mean radiative flux estimates in the Jan (a), Apr (b), Aug (c), and Oct (d) simulations. Line contours indicate mean downward fluxes in SCREAMv1, whereas colors indicate differences with respect to CERES, where positive values indicate more absorbed shortwave radiation in the model. Global mean differences are indicated in the title of each simulation.

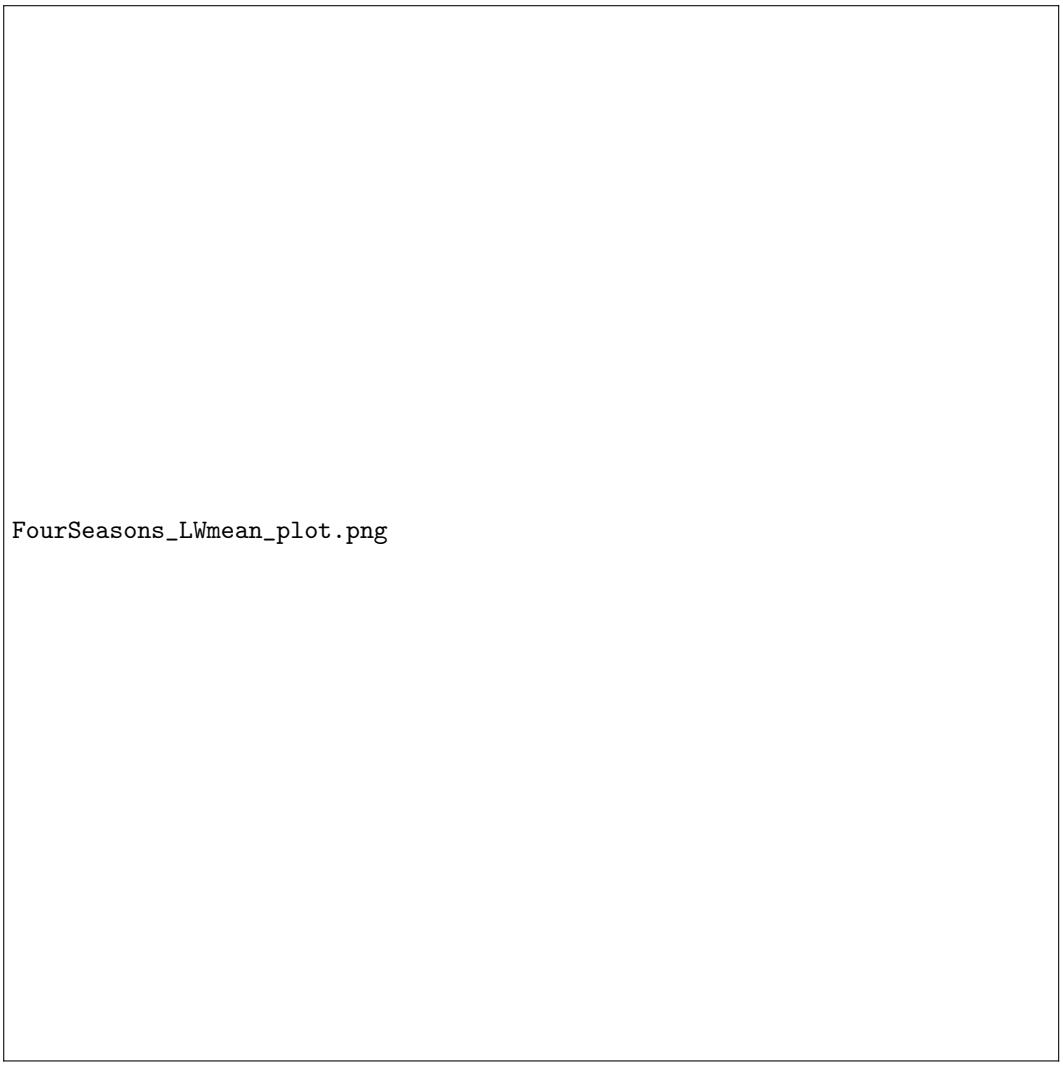


Figure 7. Time-mean differences in the net longwave radiation over the last 30 days of the simulation between SCREAMv1 and CERES-SYN daily-mean radiative flux estimates in the Jan, Apr, Aug, and Oct simulations. Contours indicate mean upward fluxes in SCREAMv1, whereas colors indicate differences with respect to CERES, where positive values indicate more outgoing longwave radiation in the model. Global mean differences are indicated in the title of each simulation.

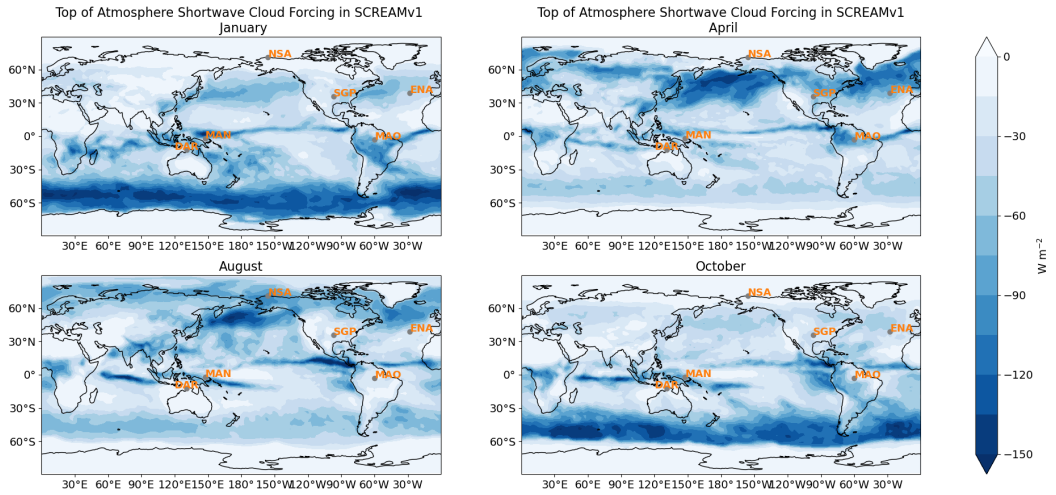


Figure 8. Temporally averaged shortwave cloud forcing in SCREAMv1 from the last 30-days of the Jan, Apr, Aug, and Oct simulations. The location of the ARM site locations (analyzed in Sect. 5.2.2, 5.3.3, and 5.4.1) are noted by gray circles labeled with the following shorthand: NSA=North Slope of Alaska, SGP=Southern Great Plains, ENA=Eastern North Atlantic, MAO=Manacapuru, Nau=Nauru, MAN=Manus, and DAR=Darwin.

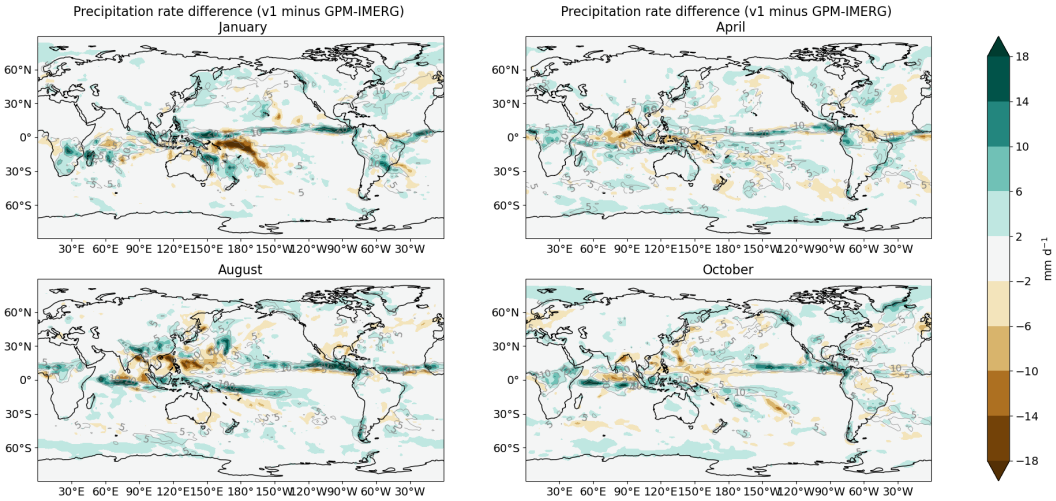


Figure 9. Time-mean differences in the precipitation rate in the last 30 days of the 40-day SCREAMv1 simulations, compared to GPM-IMERG estimates. Contours indicate the time-mean precipitation rates in SCREAMv1.

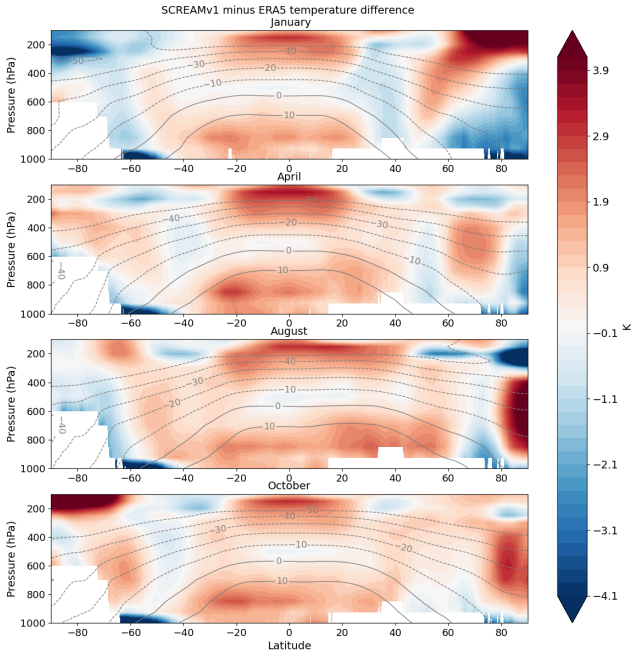


Figure 10. Zonal cross-sections of temperature differences in SCREAMv1, relative to ERA5 estimates. Time-averaged differences from the last 30 days of the simulation are shown. Line contours indicate zonal mean values from ERA5 in degree Celcius.

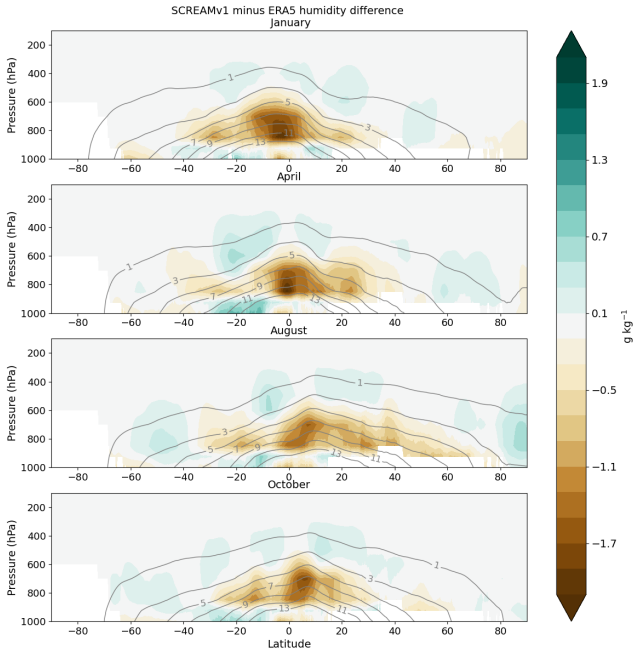


Figure 11. Zonal cross-sections of water vapor specific humidity differences in SCREAMv1, relative to ERA5 estimates. Time-averaged differences from the last 30 days of the simulation are shown. Line contours indicate zonal mean values from ERA5 in g kg^{-1}.

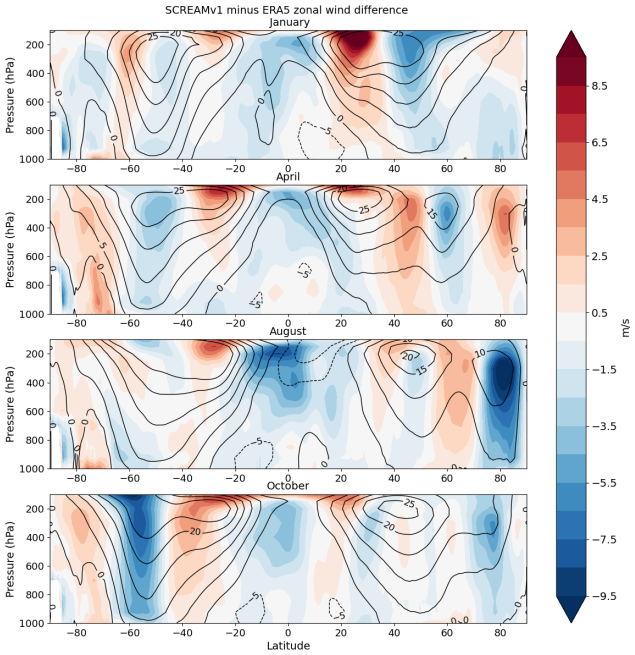


Figure 12. Colors indicate the zonal cross-sections of zonal wind differences in SCREAMv1, relative to ERA5 estimates, across the four seasons. Contour lines indicate the zonal-mean winds from ERA5. Both color and line contours show temporally averaged zonal winds from the last 30 days of the simulation.

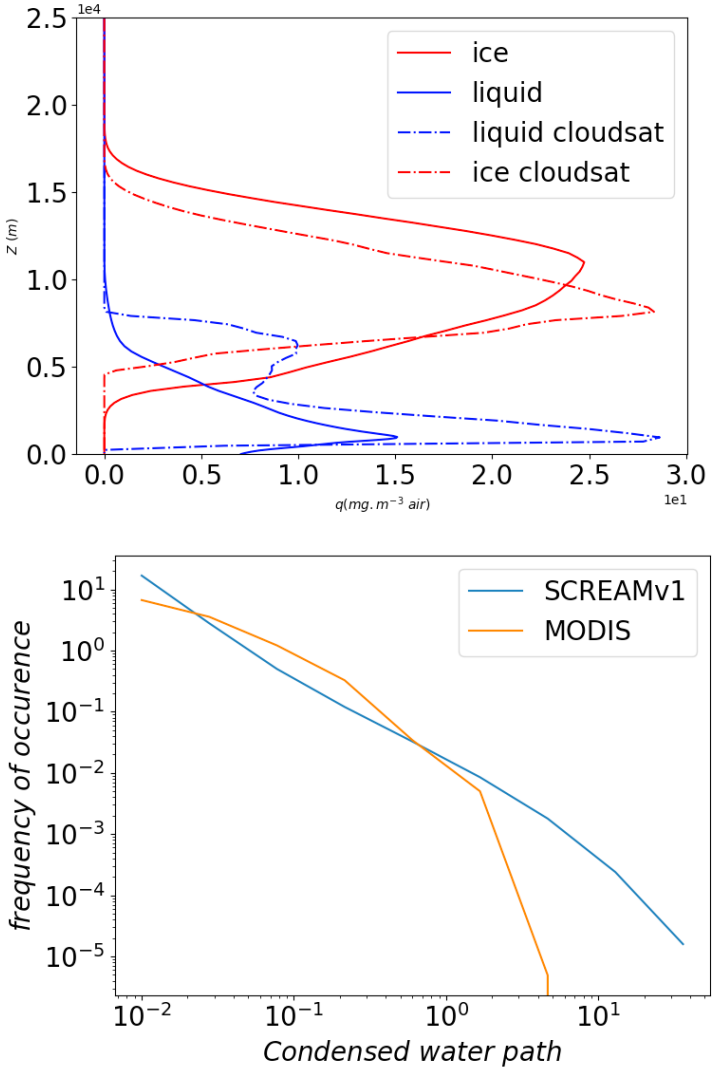


Figure 13. a) Liquid and ice cloud water content for SCREAMv1 (continuous lines) and CloudSat (dashed lines) b) Probability density function of total condensed water path for SCREAMv1 and MODIS. Both plots use data from the tropical Western Pacific for the 40 days starting Oct 1, 2013.

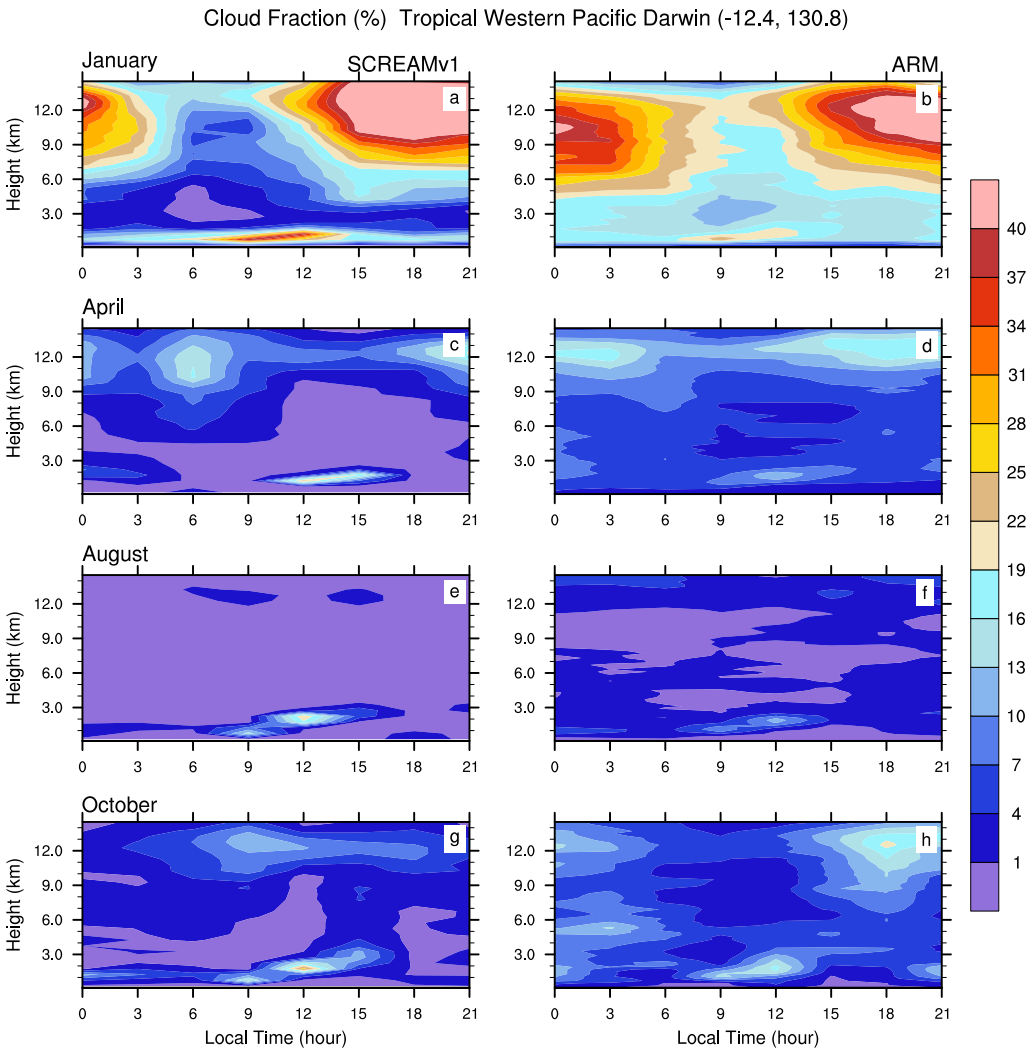


Figure 14. Composite diurnal cycle of cloud fraction (in percent) changing with height from SCREAMv1 (left column) and ARM observation (right column) at the ARM Tropical Western Pacific C3 (TWPC3) site at Darwin (12.4S, 130.8E) in January (top row), April (2nd row), August (3rd row) and October (bottom row).

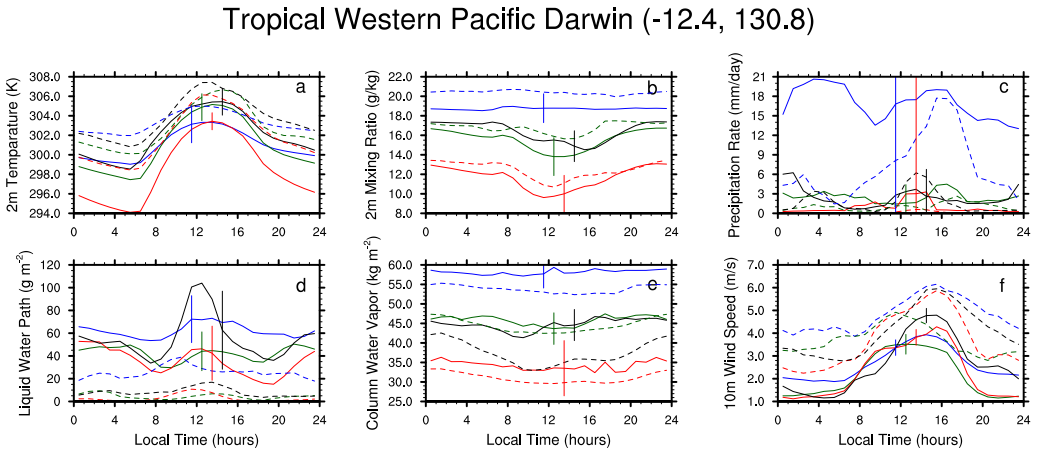


Figure 15. Composite diurnal cycle of surface variables by SCREAMv1 (dashed lines) and ARM observation (solid lines) at the ARM Tropical Western Pacific C3 (TWPC3) site at Darwin (12.4S, 130.8E) in January (blue), April (dark green), August (red) and October (black): a) 2-meter air temperature; b) 2-meter water vapor mixing ratio; c) precipitation rate; d) column integrated liquid water path; e) column integrated precipitable water vapor; f) 10-meter wind speed. The length of the vertical lines stands for two standard deviations (one on each side of the mean diurnal cycle values at certain local time around noon), which denote the interannual variability of the seasonal average diurnal cycles based on multi-year ARM observations.

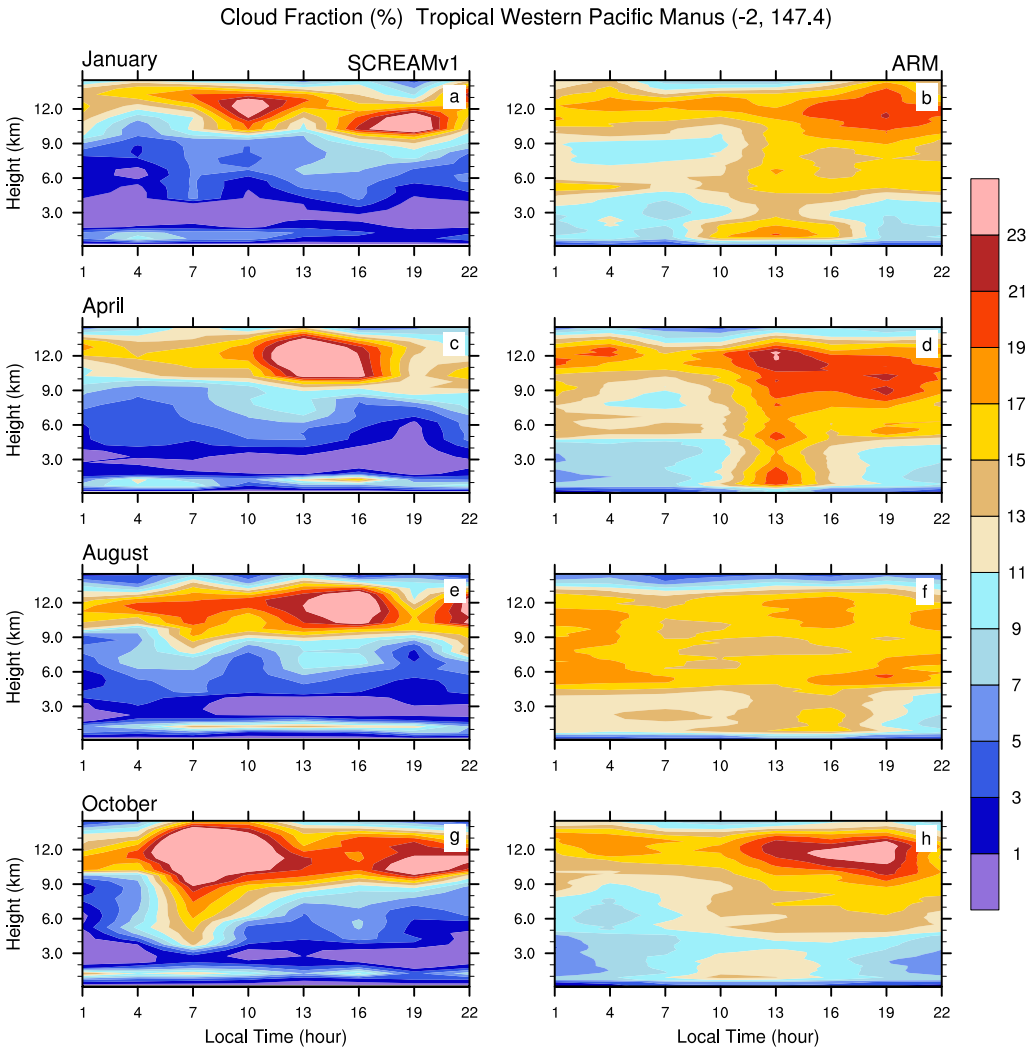


Figure 16. Same as in Figure 14, but for but for the ARM Tropical Western Pacific C1 (TWPC1) site at Manus (2S, 147.4E).

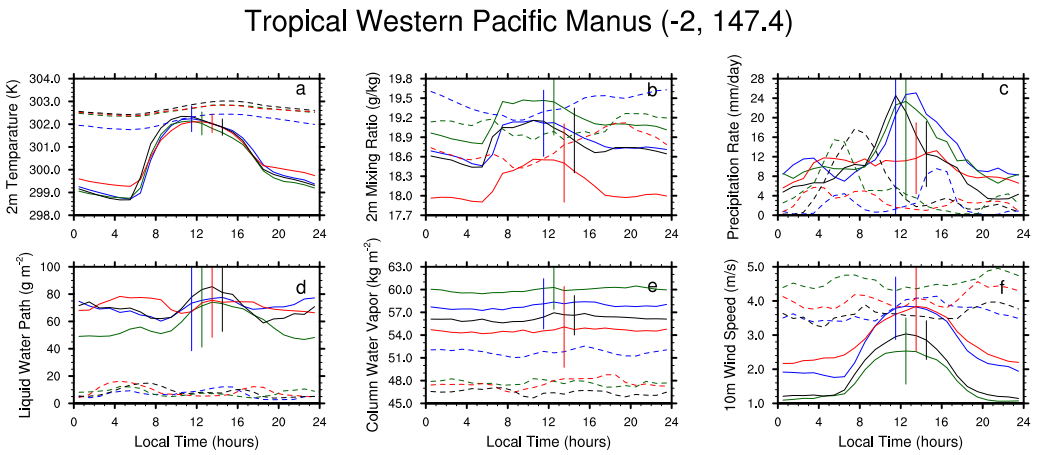


Figure 17. Same as in Figure 15, but for but for the ARM Tropical Western Pacific C1 (TWPC1) site at Manus (2S, 147.4E).

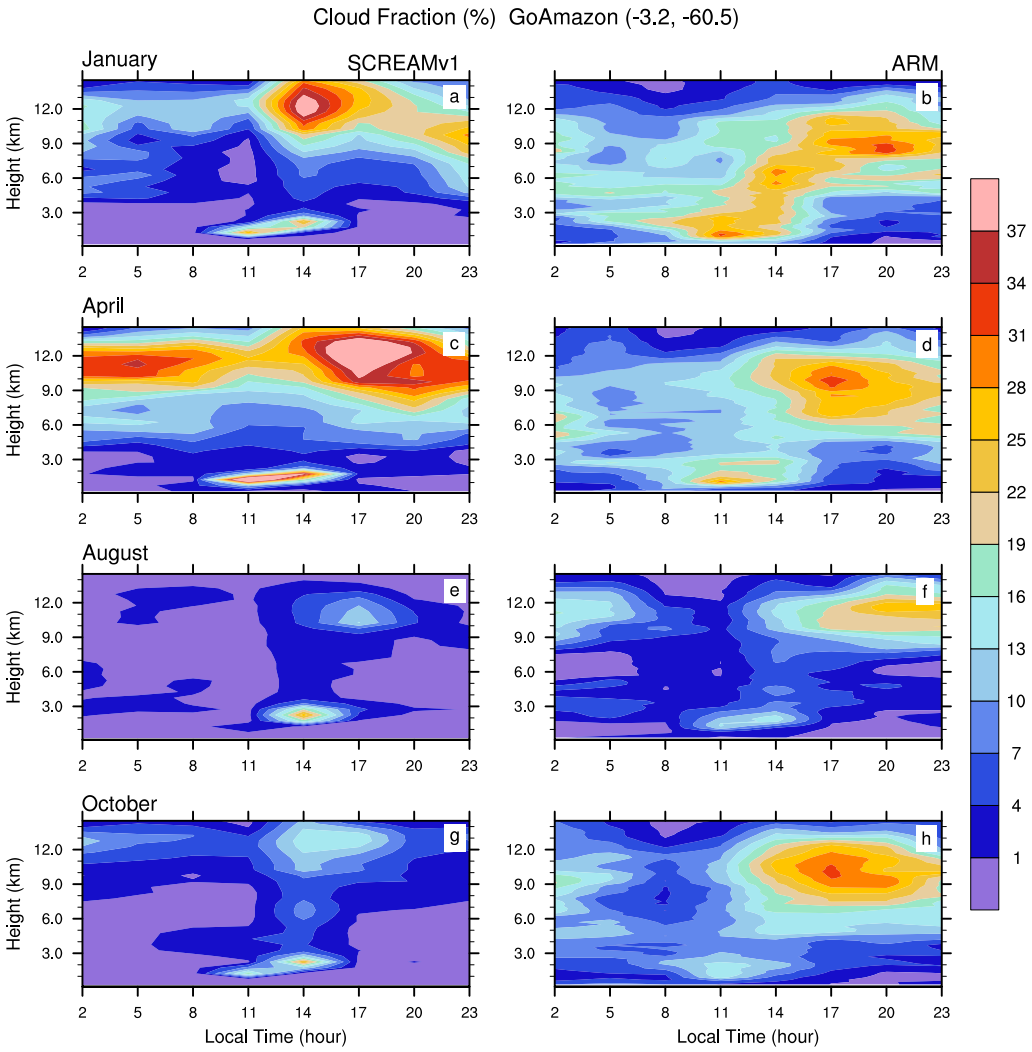


Figure 18. Same as in Figure 14, but for an ARM tropical land site during GoAmazon field campaign (MAO) at (3.2S, 60.5W).

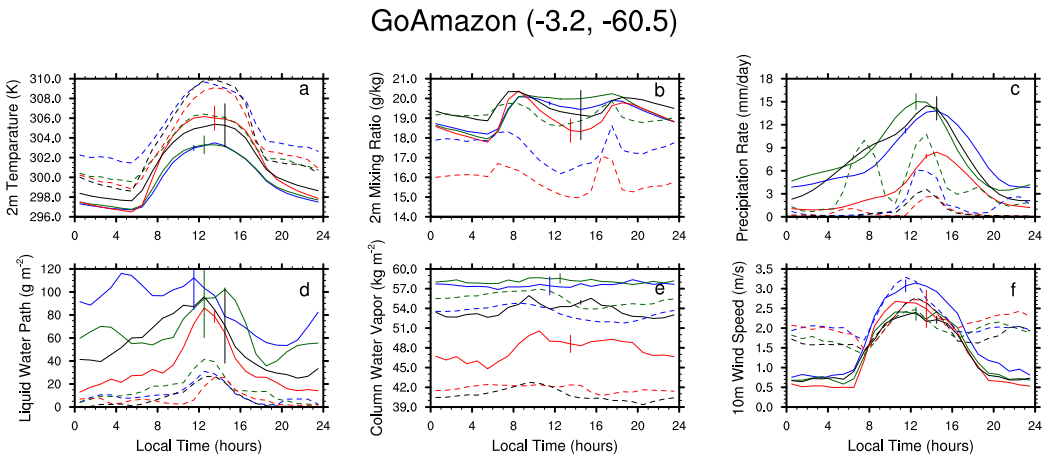


Figure 19. Same as in Figure 15, but for an ARM tropical land site during GoAmazon field campaign (MAO) at (3.2S, 60.5W).

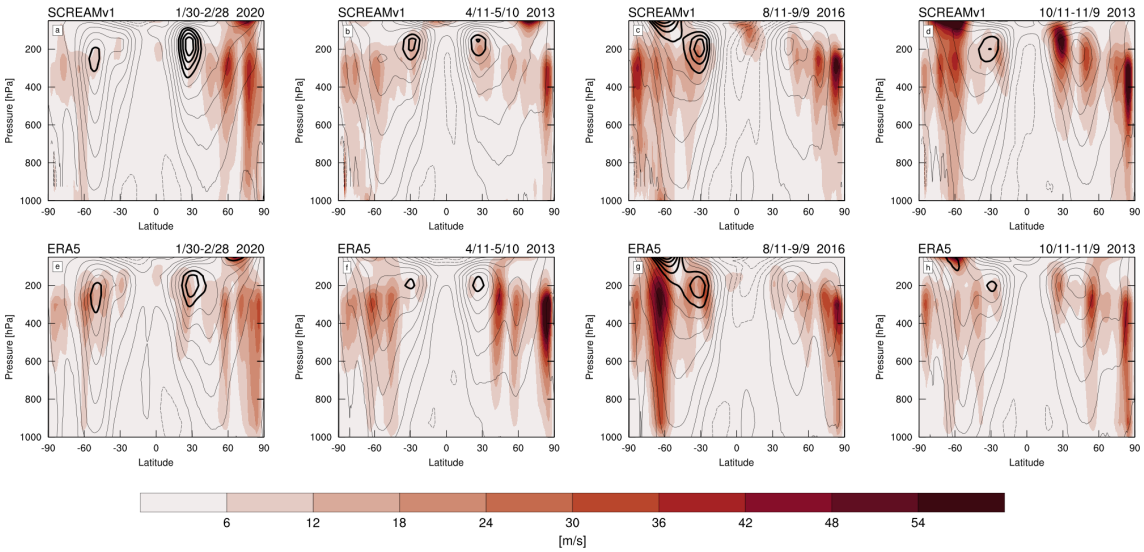


Figure 20. 3-hourly variance (shading) and time mean (contours) zonal mean zonal wind for SCREAM (a-d) and ERA5 (e-). Bold contours indicate values greater than 30 m/s and the contour interval is 5 m/s.

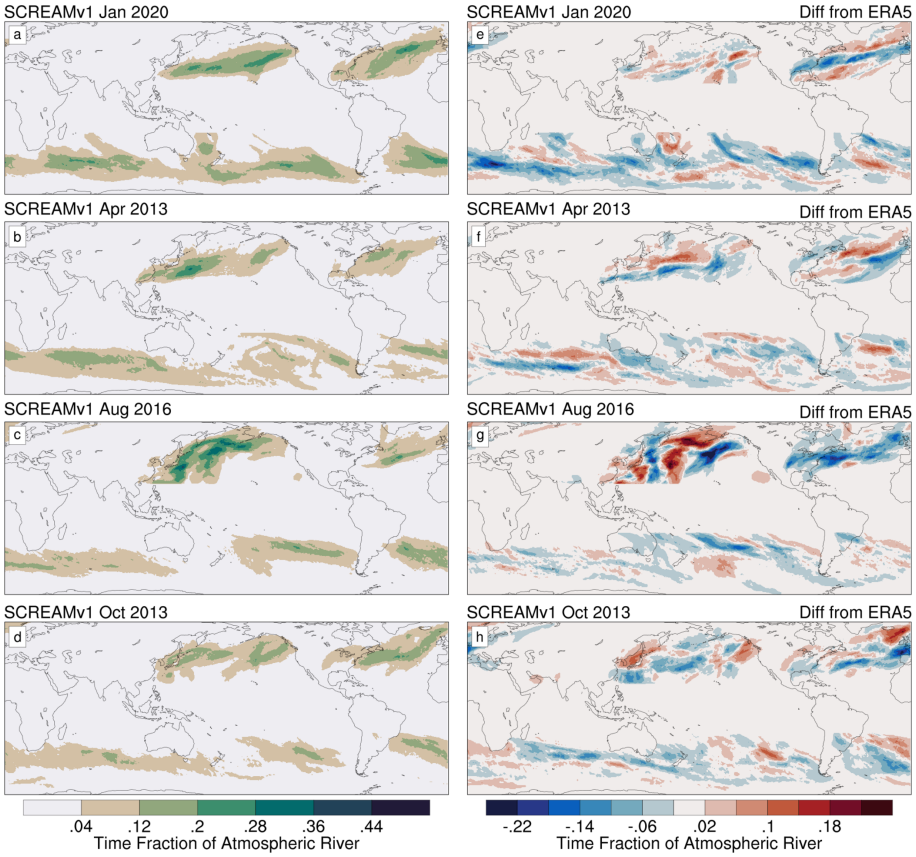


Figure 21. Atmospheric river density in SCREAMv1 as a fraction of time for each season (a-d) and the difference from ERA5 (e-h).

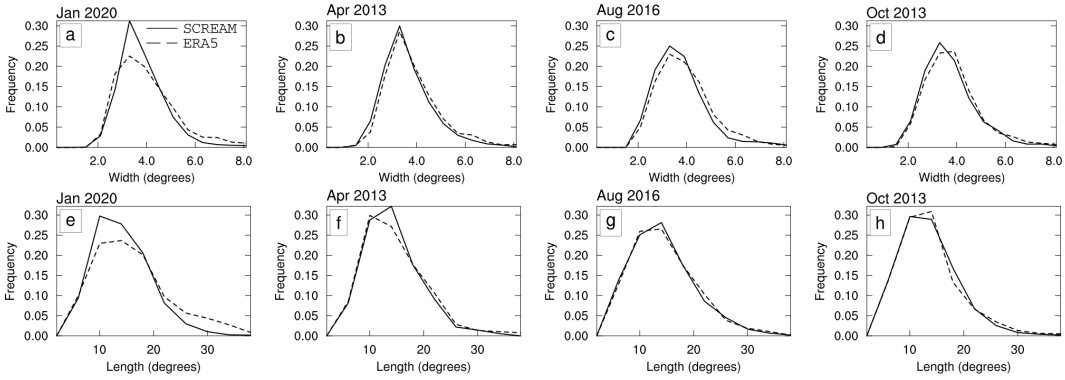


Figure 22. Seasonal histograms of atmospheric river width (a-d) and length (e-h) for SCREAMv1 (solid) and ERA5 (dashed).

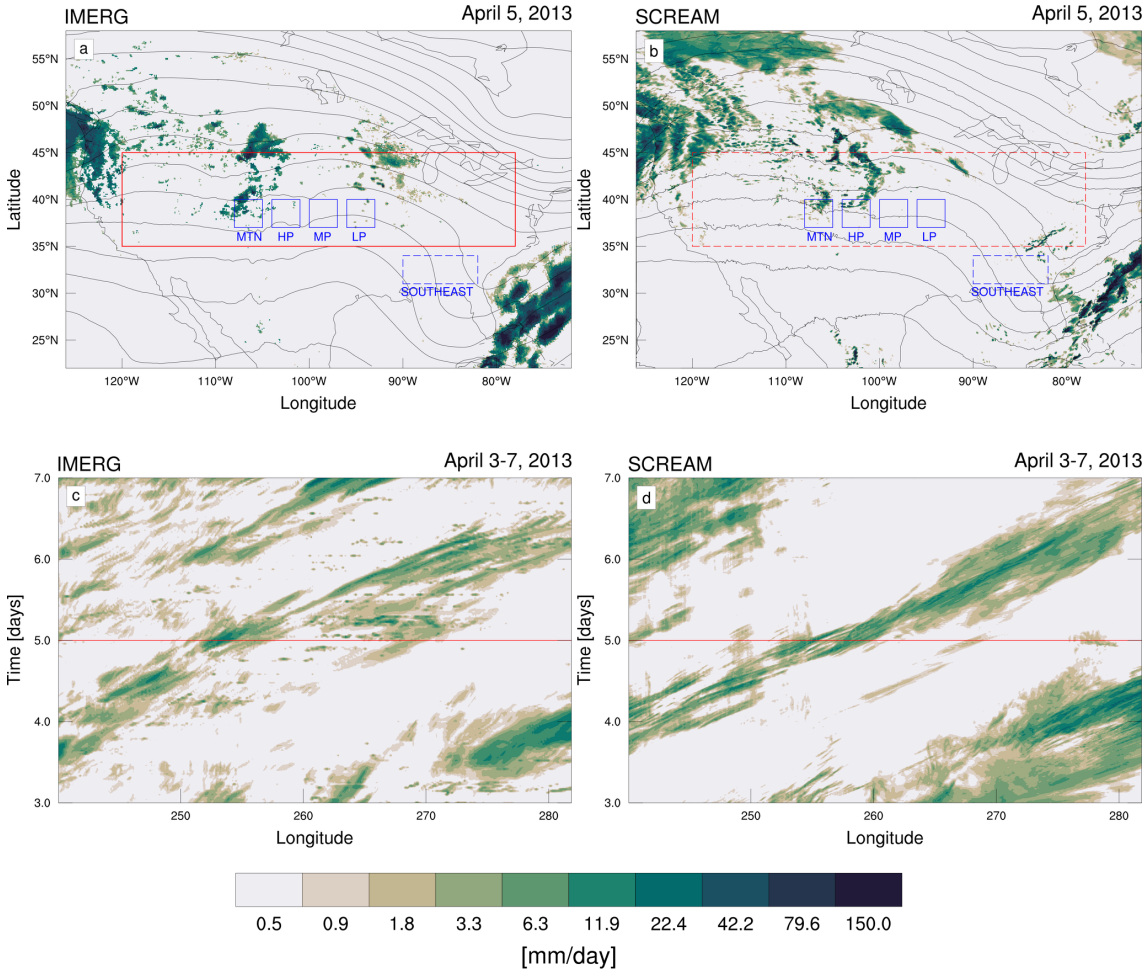


Figure 23. Upper panels show maps of precipitation and 500mb geopotential height from SCREAMv1 (a) and IMERG and ERA5 (b). Lower panels show Hovmoller diagrams of precipitation from SCREAMv1 (c) and IMERG (d) meridionally averaged over the red box in panels a-b.

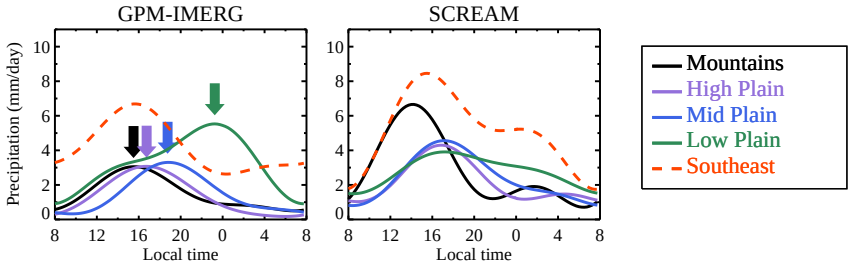


Figure 24. Composite of diurnal precipitation for the selected regions shown above. (a) Left-panel for GPM-IMERG data, b() right-panel for SCREAMv1 results.

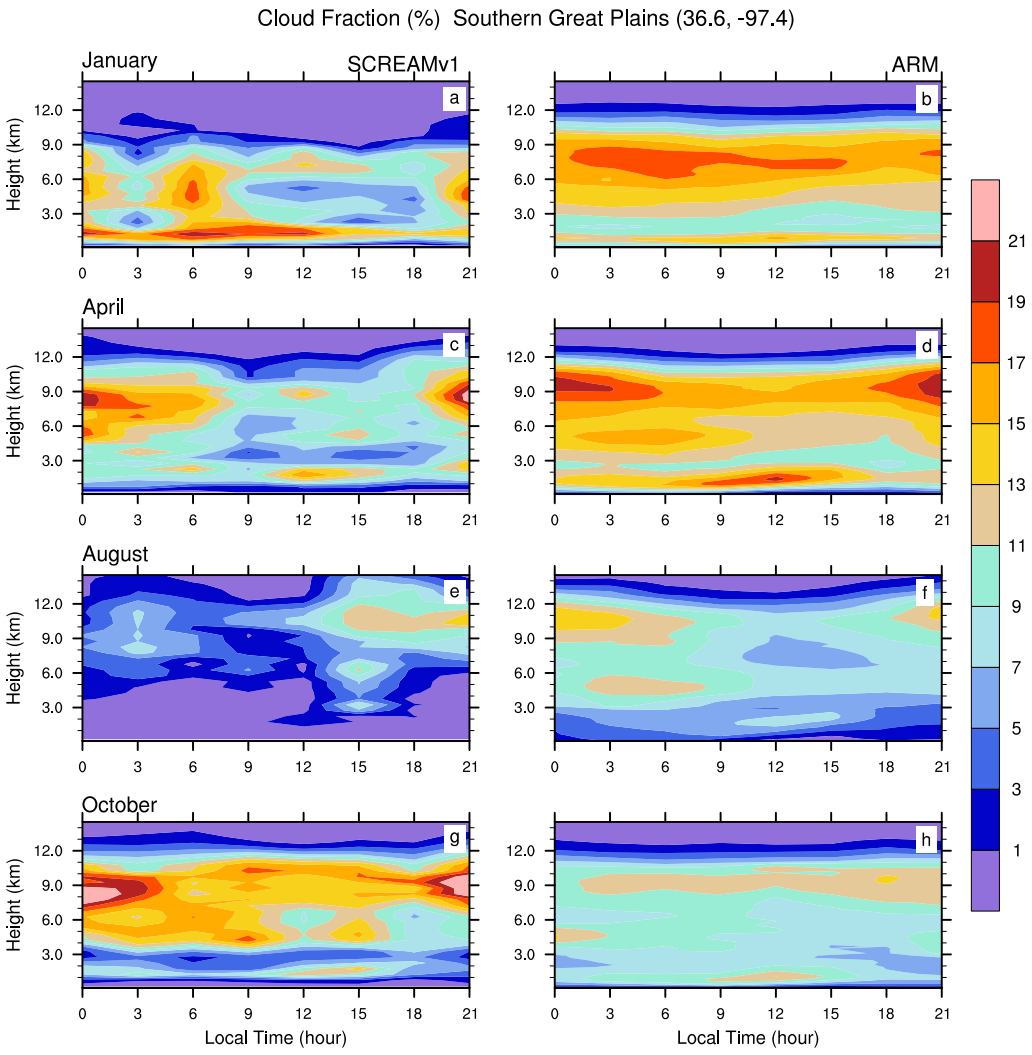


Figure 25. Composite diurnal cycle of cloud fraction (in percent) changing with height from SCREAMv1 (left column) and ARM observations (right column) at the Southern Great Plains (SGP) site located at (36.6N, 97.4W) in January (top row), April (2nd row), August (3rd row) and October (bottom row).

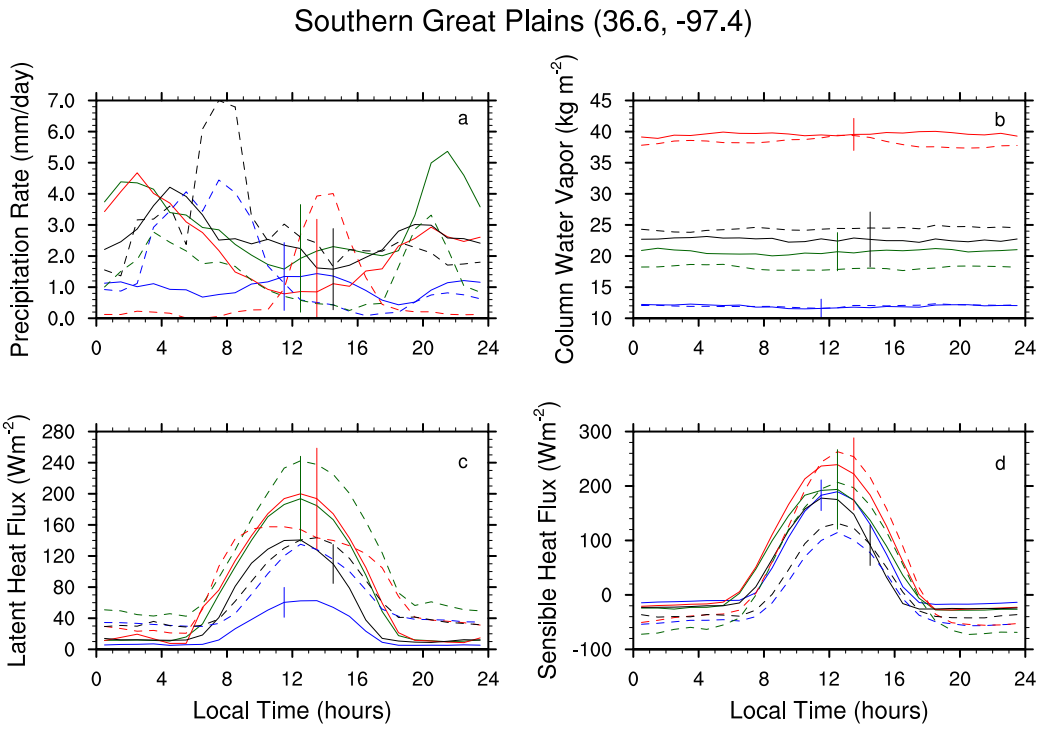


Figure 26. Composite diurnal cycle of selected near-surface variables by SCREAMv1 (dashed lines) and multi-year ARM observations (solid lines) at the Southern Great Plains (SGP) site located at (36.6N, 97.4W) in January (blue), April (dark green), August (red) and October (black): a) precipitation rate; b) column precipitable water vapor; c) latent heat flux; d) sensible heat flux. Vertical lines denote two standard deviations around the the mean diurnal cycle at certain time around local solar noon to provide a sense of the observed variability of the seasonal average diurnal cycles based on multi-year ARM observations.

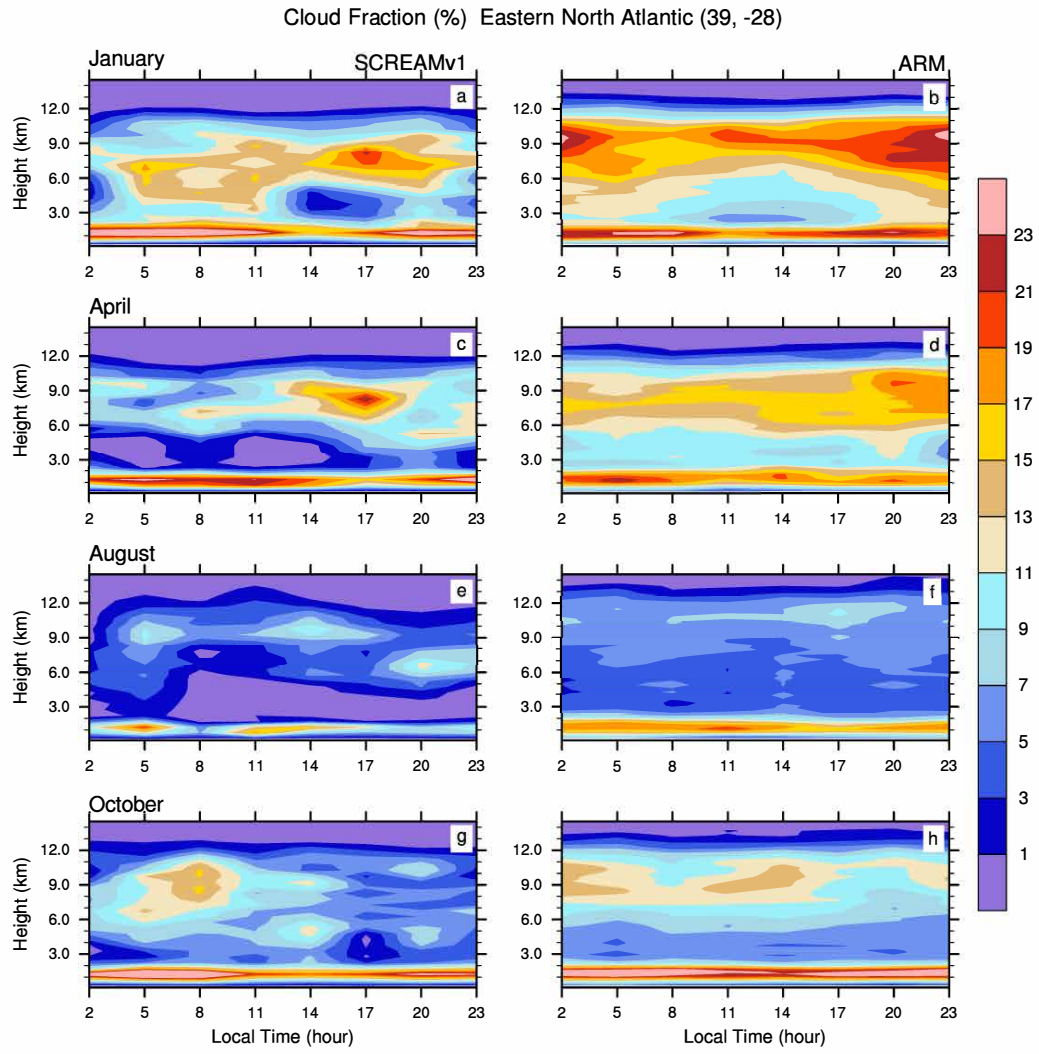


Figure 27. Same as in Figure 25, but for the ARM Eastern North Atlantic (ENA) site located at (39N, 28W).

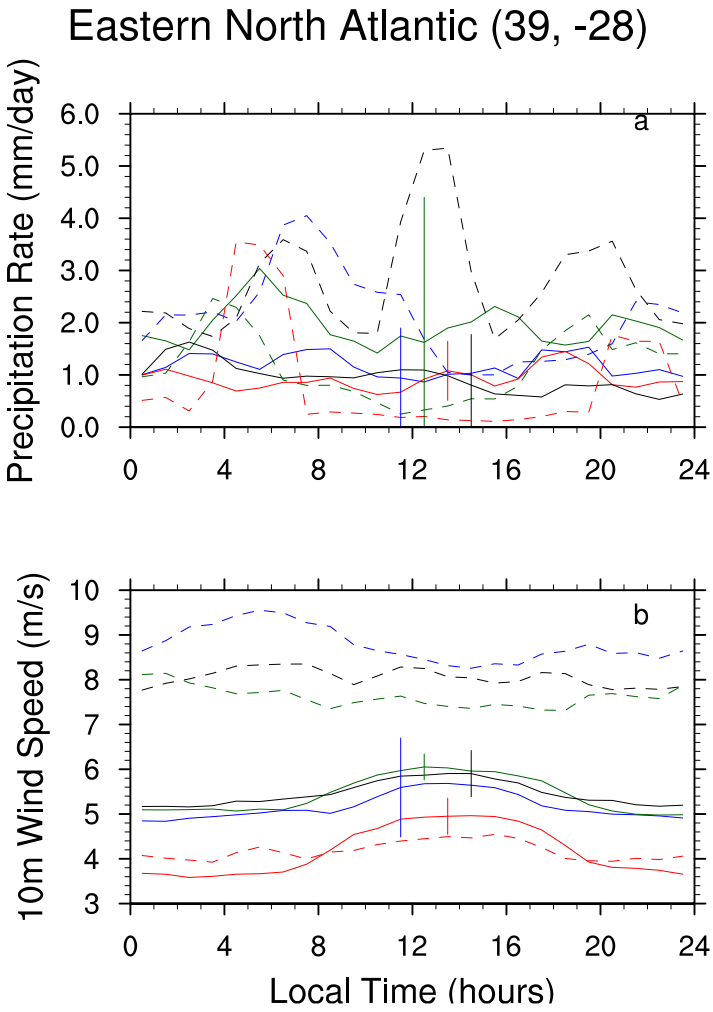


Figure 28. Composite diurnal cycle of surface variables by SCREAMv1 (dashed lines) and ARM observation (solid lines) at the ARM Eastern North Atlantic (ENA) site located at (39N, 28W) in January (blue), April (dark green), August (red) and October (black): a) precipitation rate; b) 10-meter wind speed. The length of the vertical lines stands for two standard deviations (one on each side of the mean diurnal cycle values at certain local time around noon), which denote the interannual variability of the seasonal average diurnal cycles based on multi-year ARM observations.

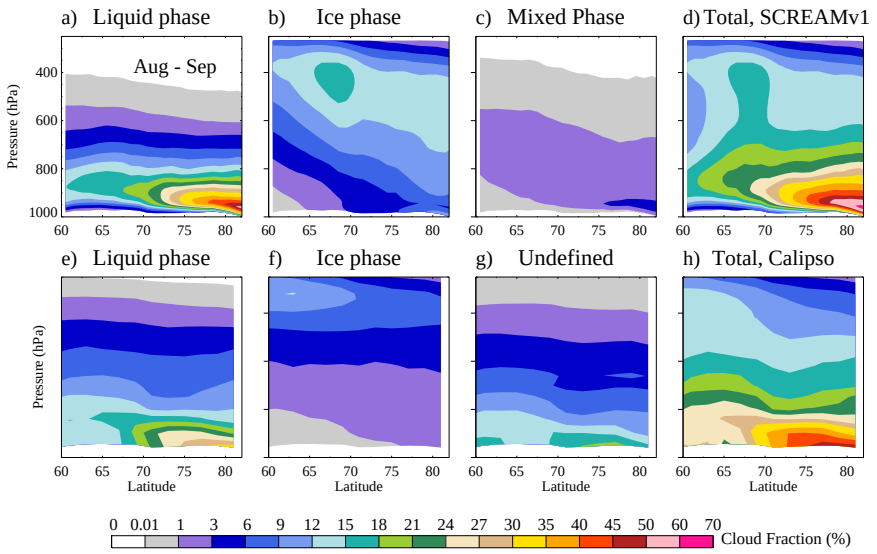


Figure 29. Zonal mean phase-partitioned and total cloud fraction over the Northern polar region for Aug simulation from SCREAMv1 (top) and CALIPSO-GOCCP (bottom)

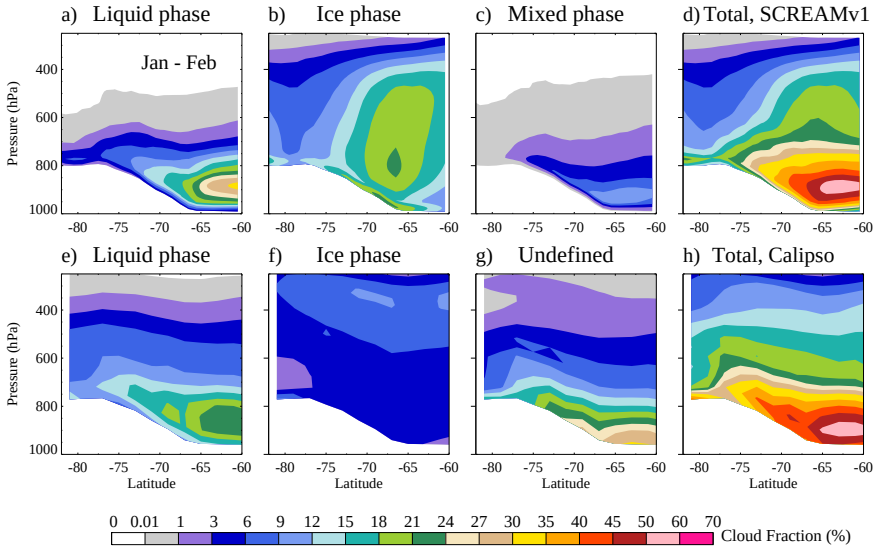


Figure 30. Same as Figure 29, but for Jan-Feb Southern polar region.

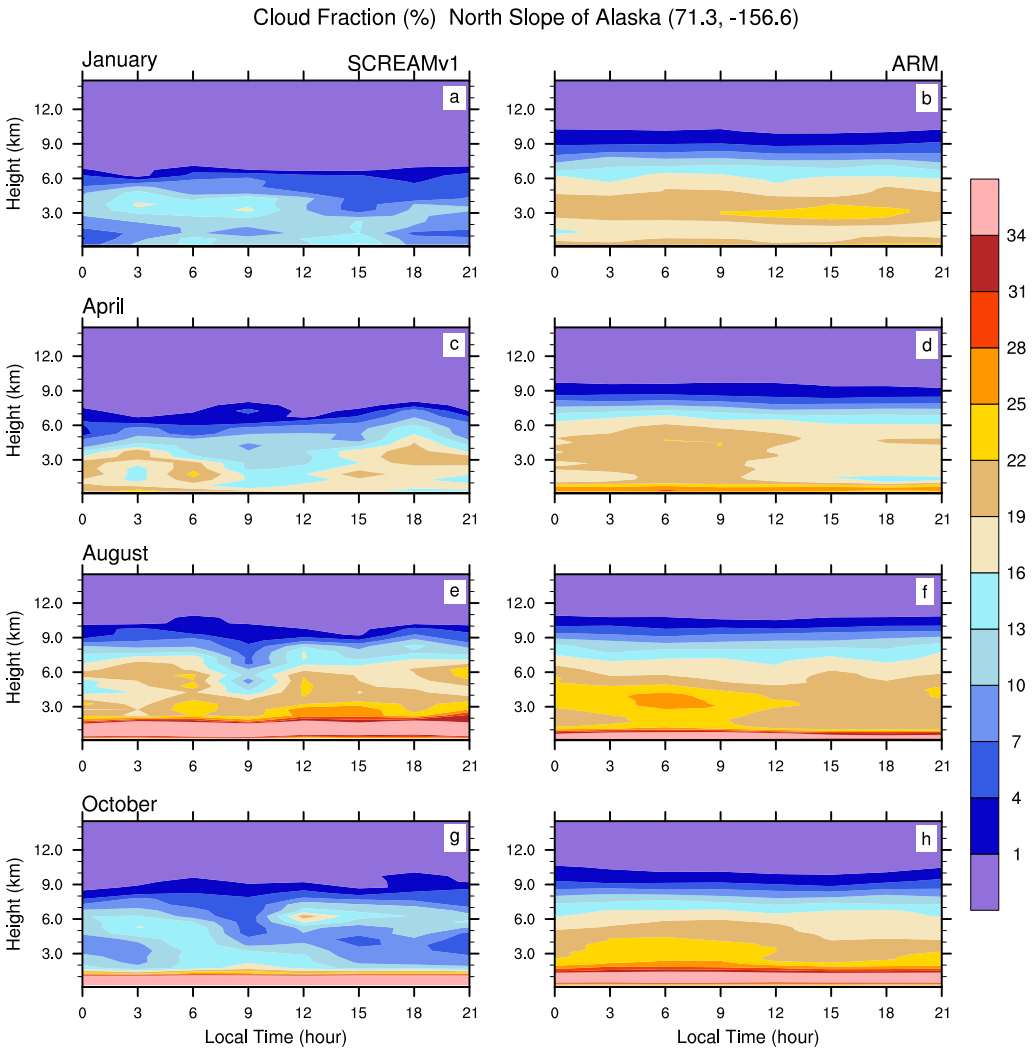


Figure 31. Same as in Figure 25, but for an ARM Arctic site, North Slope of Alaska (NSA) at (71.3N, 156.6W).

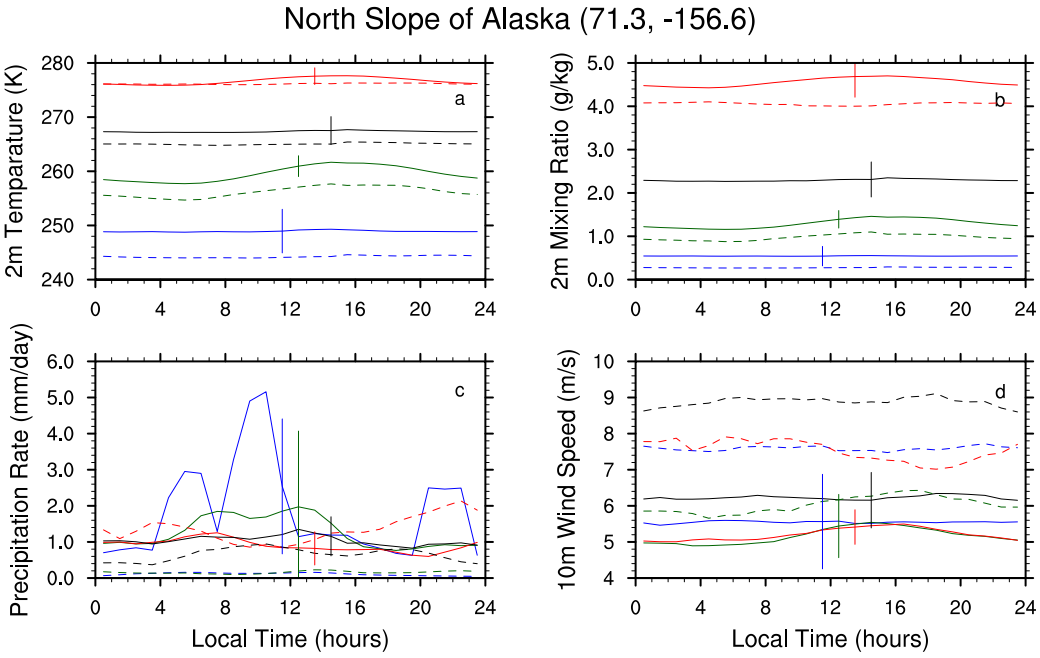


Figure 32. Composite diurnal cycle of surface variables by SCREAMv1 (dashed lines) and ARM observation (solid lines) at the ARM Arctic site, North Slope of Alaska (NSA) at (71.3N, 156.6W) in January (blue), April (dark green), August (red) and October (black): a) 2-meter air temperature; b) 2-meter water vapor mixing ratio; c) precipitation rate; d) 10-meter wind speed. The length of the vertical lines stands for two standard deviations (one on each side of the mean diurnal cycle values at certain local time around noon), which denote the interannual variability of the seasonal average diurnal cycles based on multi-year ARM observations.

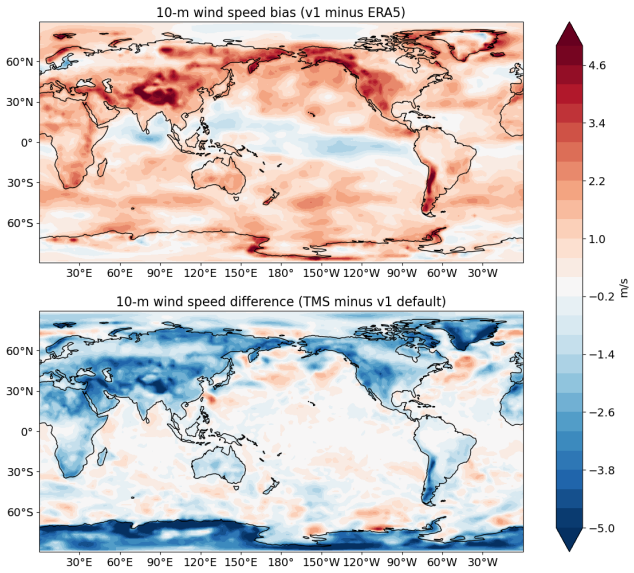


Figure 33. The top panel shows the difference in 10-m wind speed between SCREAMv1 and ERA5 estimates averaged across all four seasons. As with other differences, the last 30 days of the 40 day simulations are used for the comparison. The bottom panel shows the impact of the turbulent mountain stress parameterization (TMS) in the model physics. Since the simulations with TMS only ran for five days, average differences from Day 4 and 5 are shown.

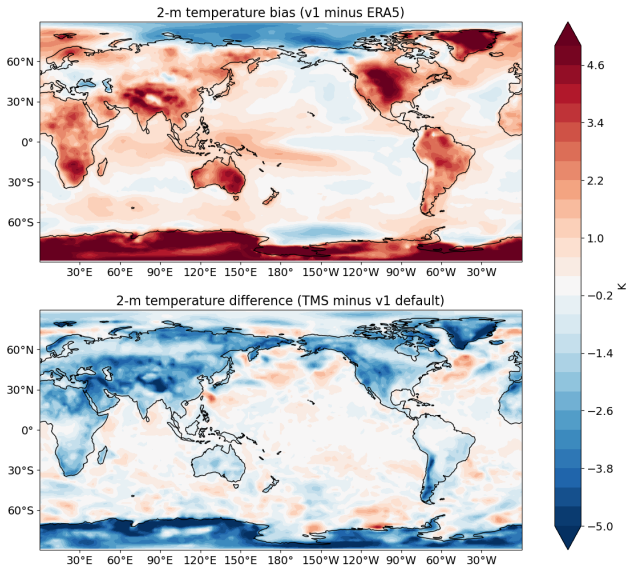


Figure 34. The top panel shows the difference in 2-m air temperature between SCREAMv1 and ERA5 estimates averaged across all four seasons. As with other differences, the last 30 days of the 40 day simulations are used for the comparison. The bottom panel shows the difference in 2-m air temperature between the simulation with the fix to SHOC and the default SCREAMv1 model, also averaged across all four seasons. Since the simulations with the fix only ran for five days, average differences from Day 4 and 5 are shown.

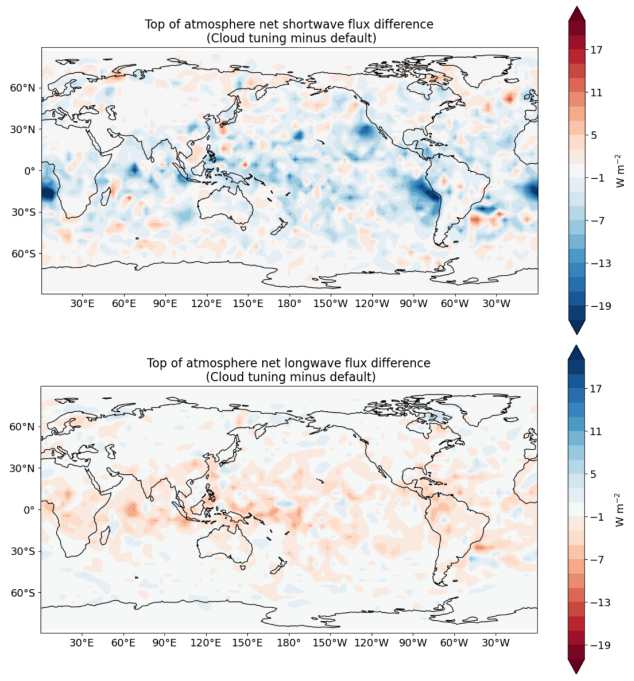


Figure 35. Impact on long- and short–wave cloud forcing across all four seasons as a result of tuning SHOC treatment of stable boundary layer and ice cloud threshold. Figure shows the difference between new cloud tunings and those used for four-seasons analysis.

Figure1.

State from Previous Step:

T_mid, horiz_winds, o3_volume_mix_ratio, phis, ps,
qv, bm, nc, ni, nr, qc, qi, qm, qr, tke, tracers,
T_prev_micro_step, qv_prev_micro_step,
precip_ice_surf_mass, precip_liq_surf_mass

SurfaceCouplingImporter

Computed Fields:
T_2m, qv_2m, sfc_alb_dif_nir, sfc_alb_dif_vis, sfc_alb_dir_nir,
snow_depth_land, surf_evap, surf_lw_flux_up, sfc_alb_dir_vis,
surf_mom_flux, surf_radiative_T, surf_sens_flux, wind_speed_10m

homme

Computed Fields:
omega, p_dry_int, p_dry_mid, p_int, p_mid,
pseudo_density, pseudo_density_dry
Updated Fields:
T_mid, horiz_winds, o3_volume_mix_ratio,
phis, ps, qv, tracers (group)

shoc

Computed Fields:
inv_qc_relvar, pbl_height
Required Fields:
surf_evap, surf_mom_flux, surf_sens_flux,
phis, omega, p_int, p_mid, pseudo_density
Updated Fields:
T_mid, horiz_winds, qv, qc, tke, cldfrac_liq,
eddy_diff_mom, sgs_buoy_flux, tracers (group)

cld_fraction

Computed Fields:
cldfrac_ice, cldfrac_ice_for_analysis,
cldfrac_tot, cldfrac_tot_for_analysis
Required Fields:
qi, cldfrac_liq

spa

Computed Fields:
aero_g_sw, aero_ssa_sw, aero_tau_lw,
aero_tau_sw, nccn
Required Fields:
p_mid

p3

Computed Fields:
precip_ice_surf_mass, precip_liq_surf_mass,
eff_radius_qc, eff_radius_qi, micro_liq_ice_exchange,
micro_vap_ice_exchange, micro_vap_liq_exchange
Required Fields:
p_dry_mid, pseudo_density_dry,
inv_qc_relvar, cldfrac_tot, nccn,
precip_ice_surf_mass, precip_liq_surf_mass
Updated Fields:
T_mid, qv, bm, nc, ni, nr, qc, qi, qm, qr,
T_prev_micro_step, qv_prev_micro_step

rrtmgp

Computed Fields:
LW_clrsky_flux_dn, LW_clrsky_flux_up, LW_flux_dn, LW_flux_up, SW_clrsky_flux_dn,
SW_clrsky_flux_dn_dir, SW_clrsky_flux_up, SW_flux_dn, SW_flux_dn_dir, SW_flux_up, ch4_volume_mix_ratio
cldhgh, cldlow, cldmed, cldtot, co2_volume_mix_ratio, co_volume_mix_ratio, h2o_volume_mix_ratio,
n2_volume_mix_ratio, n2o_volume_mix_ratio, o2_volume_mix_ratio, rad_heating_pdel, sfc_flux_dif_nir,
sfc_flux_dif_vis, sfc_flux_dir_nir, sfc_flux_dir_vis, sfc_flux_lw_dn, sfc_flux_sw_net
Required Fields:
sfc_alb_dif_nir, sfc_alb_dif_vis, sfc_alb_dir_nir, sfc_alb_dir_vis, surf_lw_flux_up qv, p_int, p_mid, qc, qi,
pseudo_density, cldfrac_tot, aero_g_sw, aero_ssa_sw, aero_tau_lw, aero_tau_sw, eff_radius_qc, eff_radius_qi
Updated Fields:
T_mid, o3_volume_mix_ratio

State Saved for Next Step

T_mid, horiz_winds, o3_volume_mix_ratio, phis, ps,
qv, bm, nc, ni, nr, qc, qi, qm, qr, tke, tracers,
T_prev_micro_step, qv_prev_micro_step,
precip_ice_surf_mass, precip_liq_surf_mass

SurfaceCouplingExporter

Required Fields:
T_mid, horiz_winds, phis, qv, p_int, p_mid,
pseudo_density, precip_ice_surf_mass, precip_liq_surf_mass,
sfc_flux_dif_nir, sfc_flux_dif_vis, sfc_flux_dir_nir,
sfc_flux_dir_vis, sfc_flux_lw_dn, sfc_flux_sw_net

Figure2.

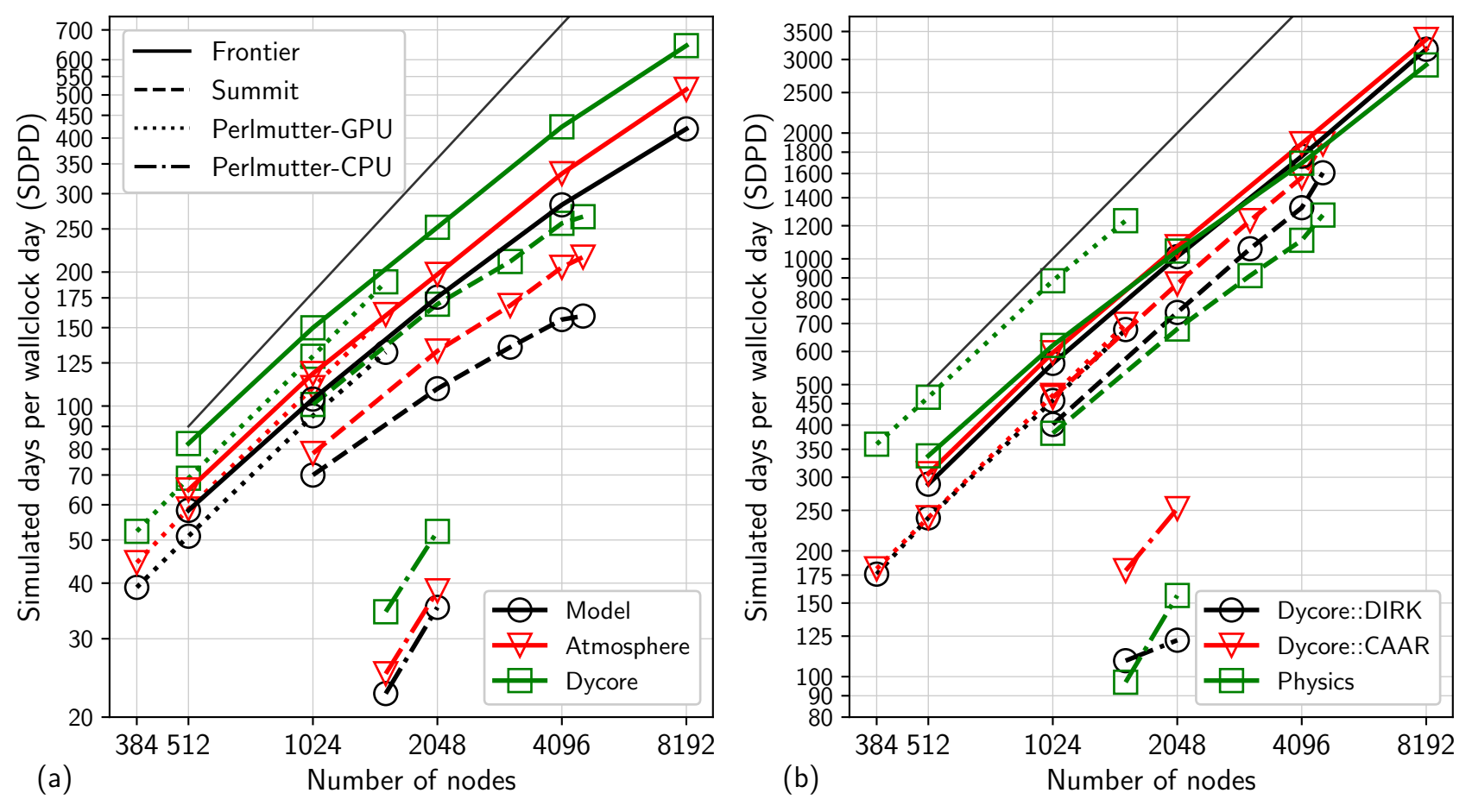

Frontier
Summit
Perlmutter-GPU
Perlmutter-CPU
Model
Atmosphere
Dycore
Simulated days per wallclock day (SDPD)
Number of nodes
(a)
384 512
1024
2048
4096
8192
Dycore::DIRK
Dycore::CAAR
Physics
(b)

Figure3.

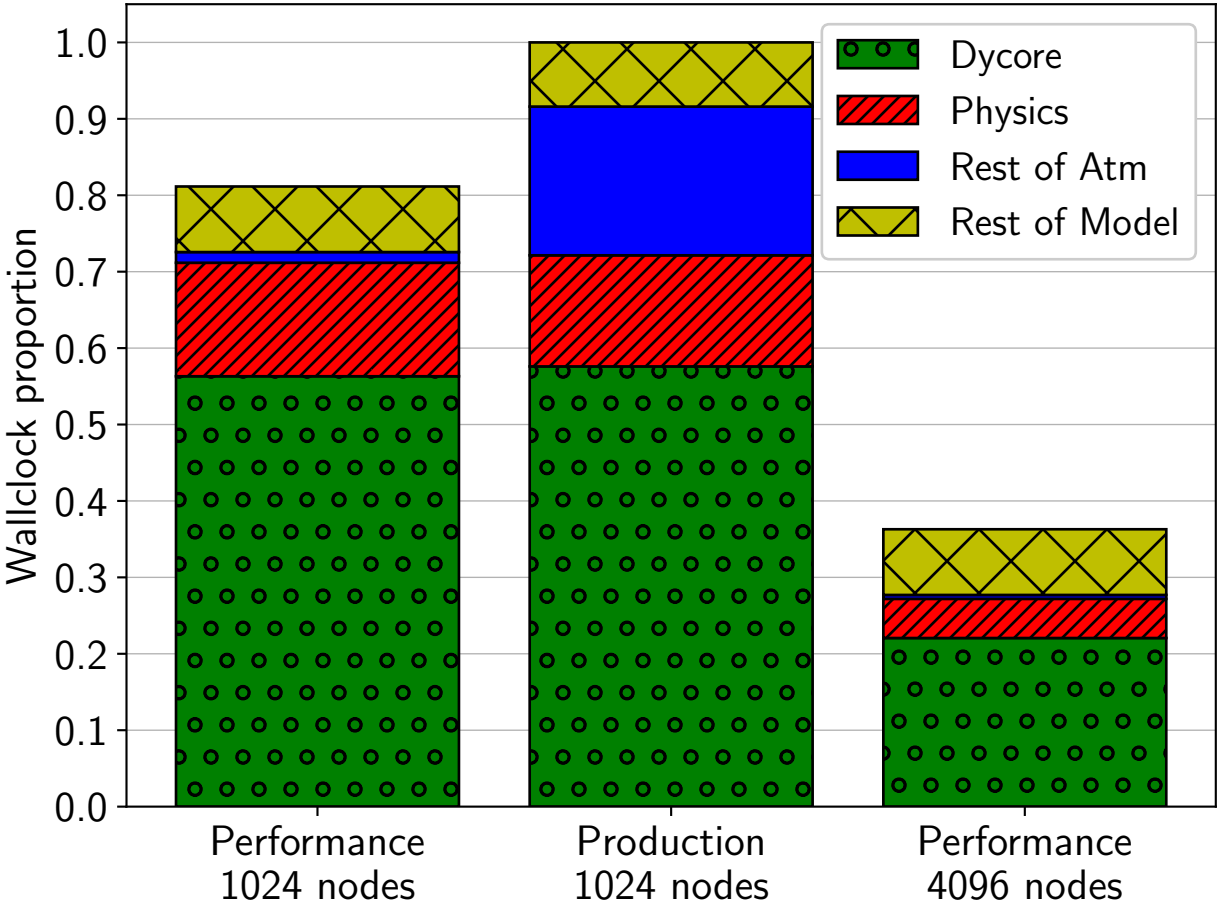

Wallclock proportion
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Dycore
Physics
Rest of Atm
Rest of Model
Performance
1024 nodes
Production
1024 nodes
Performance
4096 nodes

Figure4.

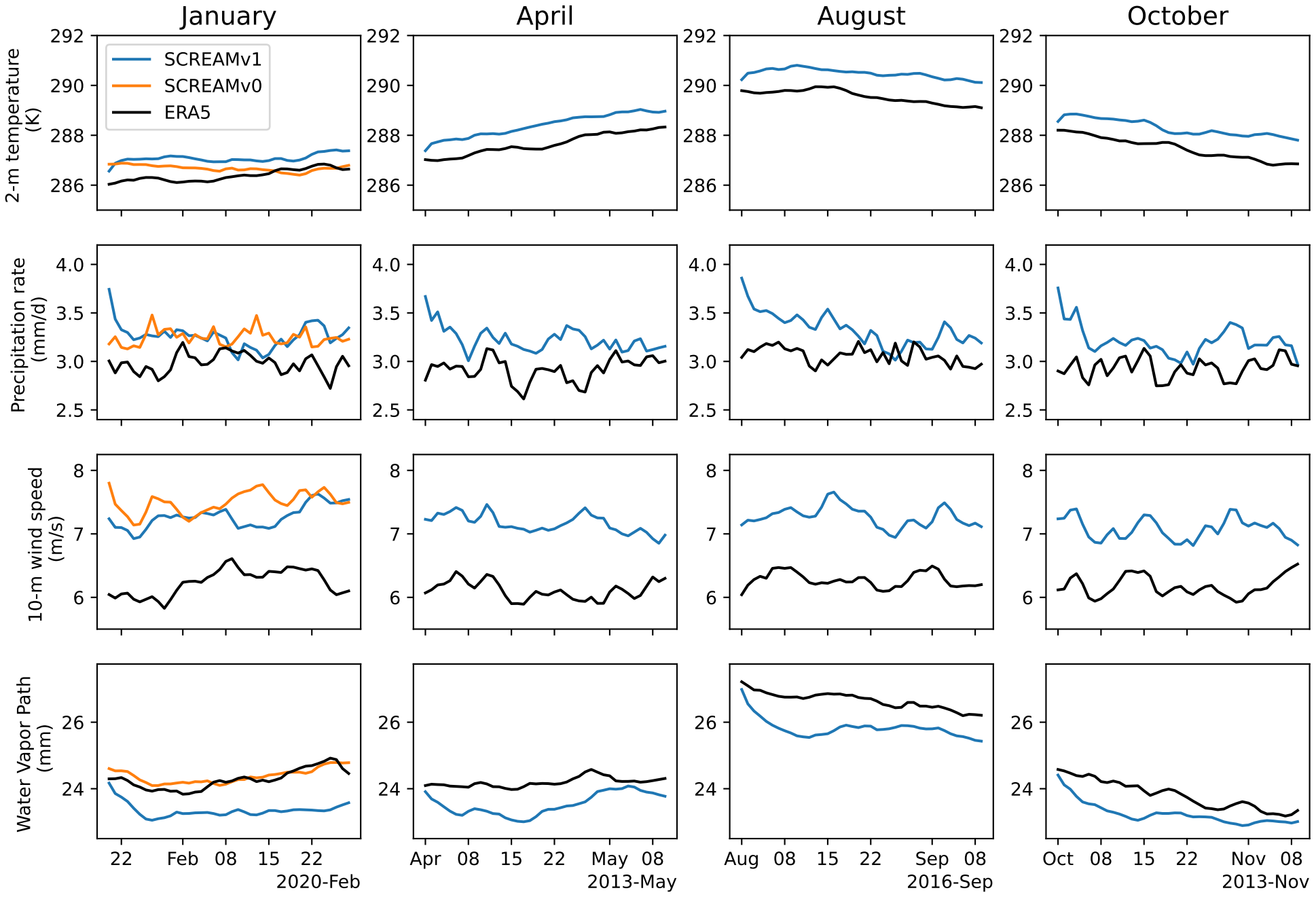

January
April
August
October
SCREAMv1
SCREAMv0
ERA5
2-m temperature (K)
Precipitation rate (mm/d)
10-m wind speed (m/s)
Water Vapor Path (mm)
292
290
288
286
4.0
3.5
3.0
2.5
8
7
6
26
24
22
Feb
08
15
22
2020-Feb
Apr
08
15
22
May
08
2013-May
Aug
08
15
22
Sep
08
2016-Sep
Oct
08
15
22
Nov
08
2013-Nov

Figure5.

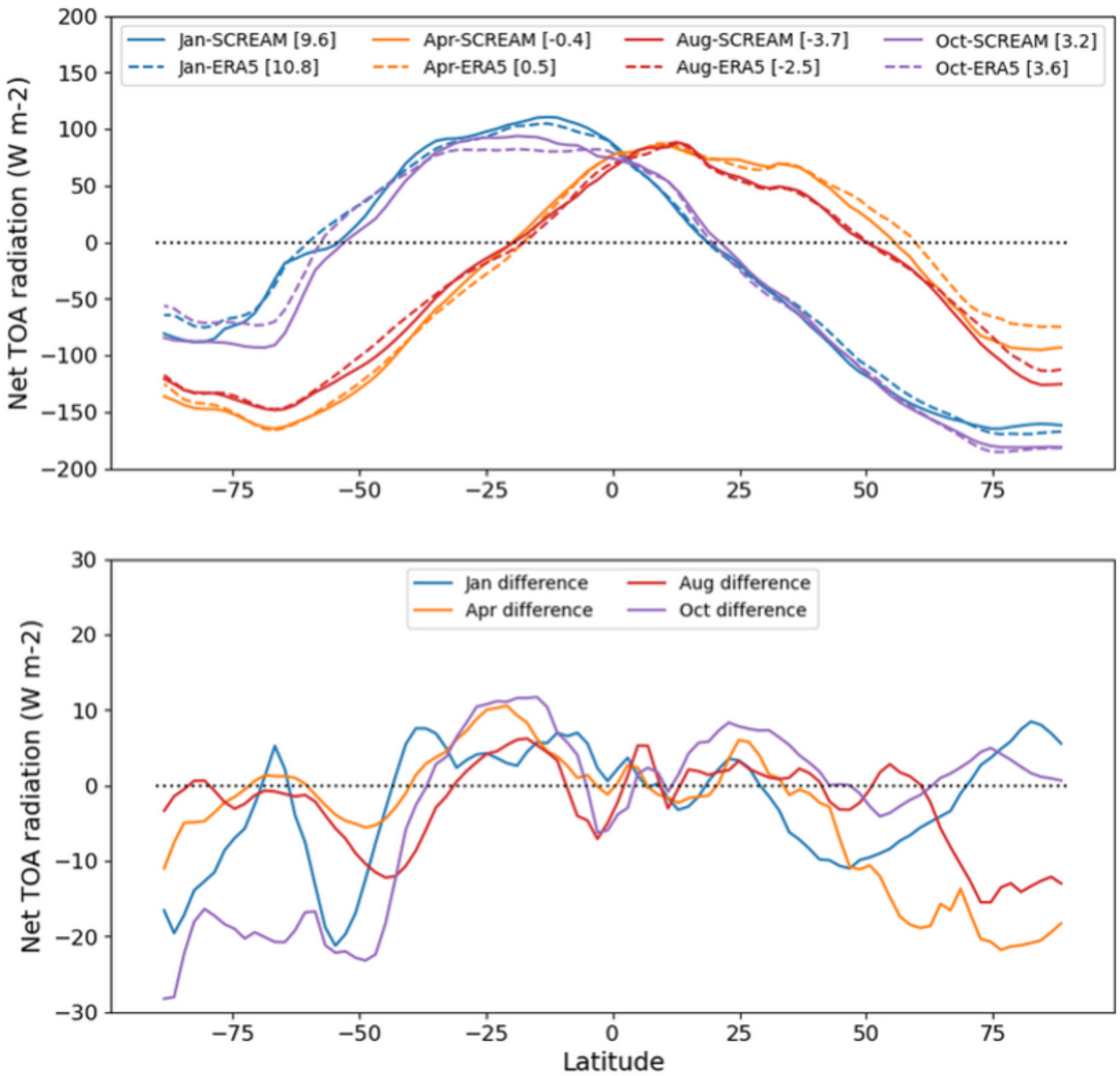

Jan-SCREAM [9.6]
Jan-ERA5 [10.8]
Apr-SCREAM [-0.4]
Apr-ERA5 [0.5]
Aug-SCREAM [-3.7]
Aug-ERA5 [-2.5]
Oct-SCREAM [3.2]
Oct-ERA5 [3.6]
Net TOA radiation (W m-2)
200
150
100
50
0
−50
−100
−150
−200
−75
−50
−25
0
25
50
75
Jan difference
Apr difference
Aug difference
Oct difference
Net TOA radiation (W m-2)
30
20
10
0
−10
−20
−30
−75
−50
−25
0
25
50
75
Latitude

Figure6.

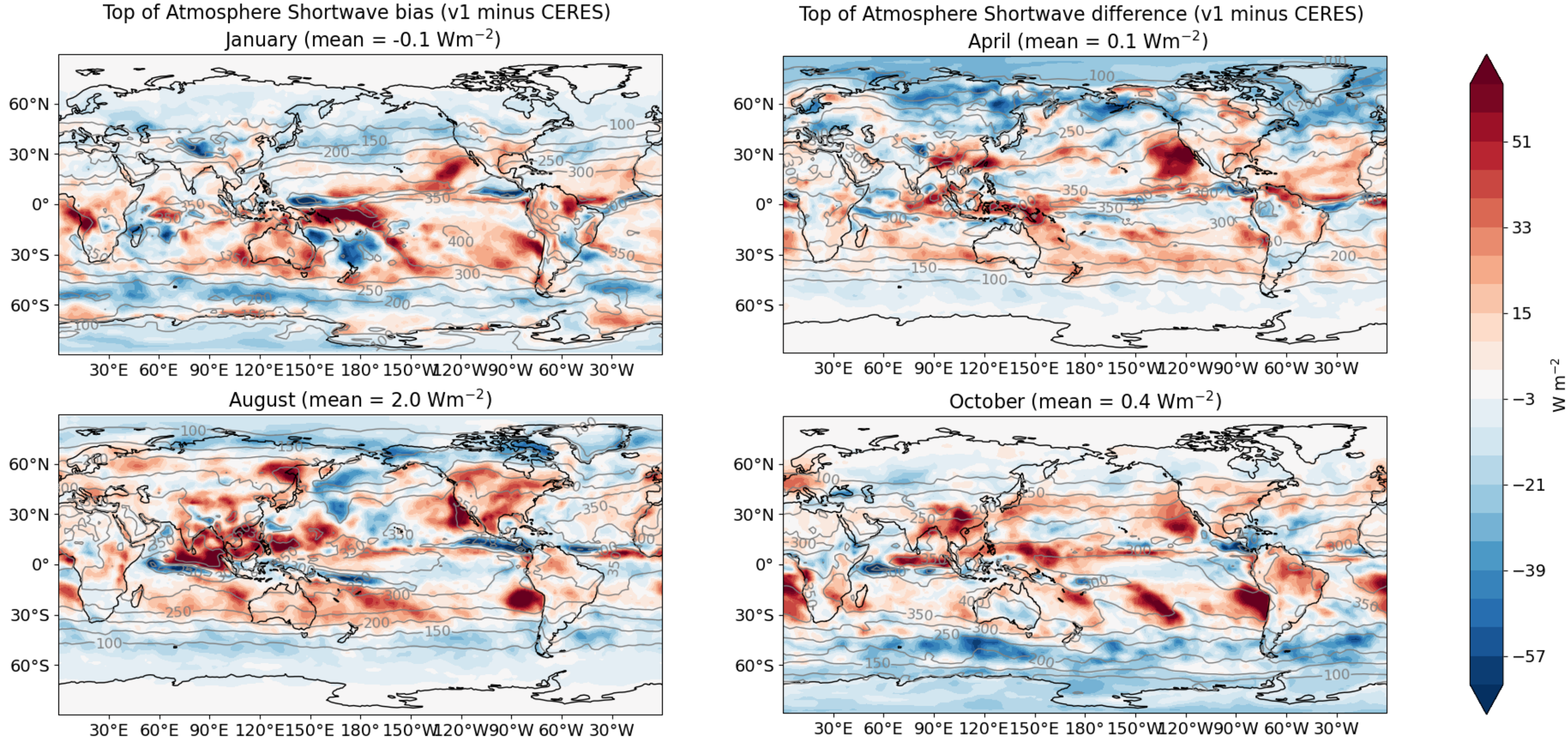

Top of Atmosphere Shortwave bias (v1 minus CERES)
January (mean = -0.1 Wm^{-2})
Top of Atmosphere Shortwave difference (v1 minus CERES)
April (mean = 0.1 Wm^{-2})
August (mean = 2.0 Wm^{-2})
October (mean = 0.4 Wm^{-2})
60°N
30°N
0°
30°S
60°S
30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W
51
33
15
−3
−21
−39
−57
W m^{-2}

Figure7.

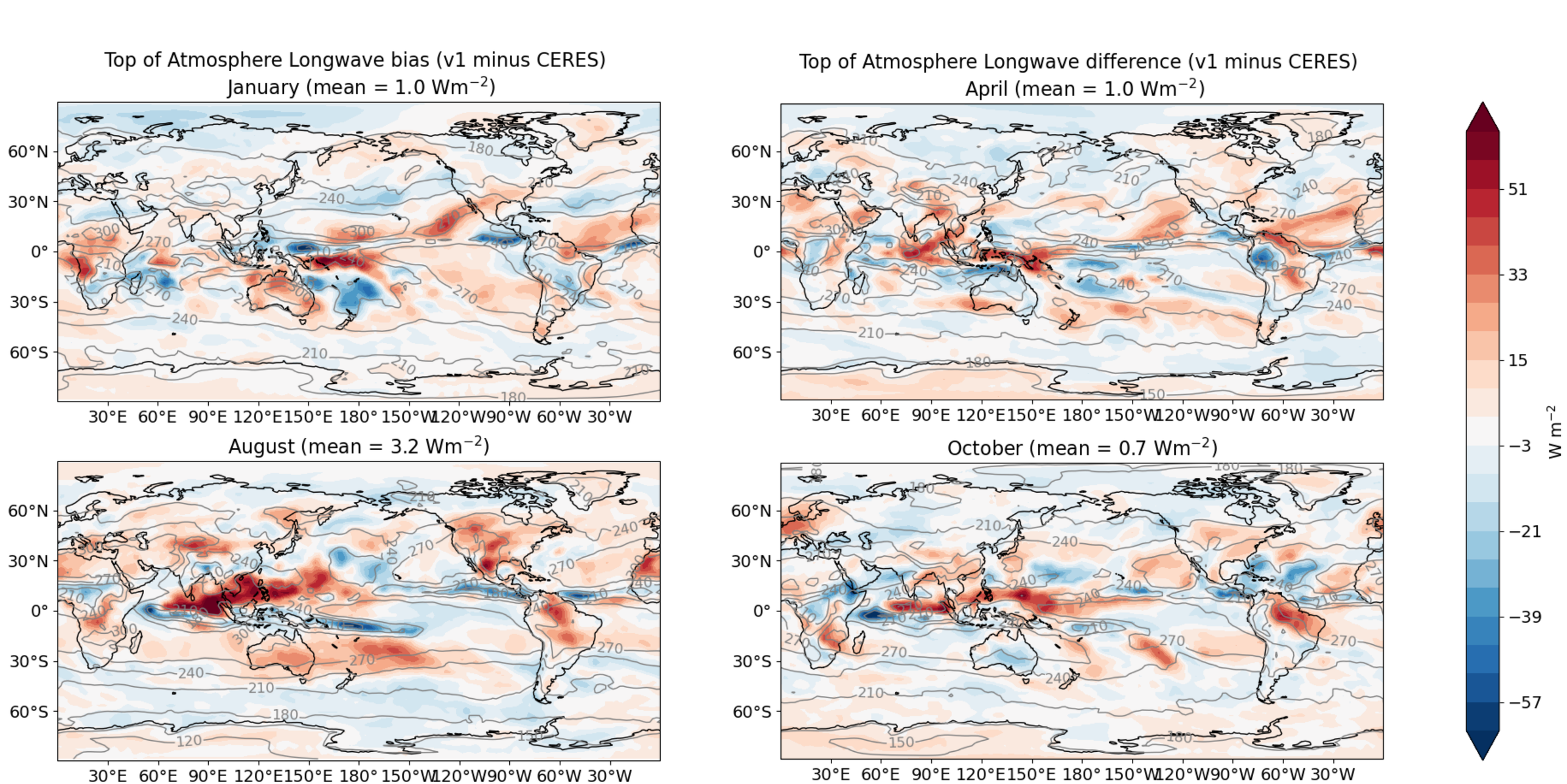

Top of Atmosphere Longwave bias (v1 minus CERES)
January (mean = 1.0 Wm^{-2})
Top of Atmosphere Longwave difference (v1 minus CERES)
April (mean = 1.0 Wm^{-2})
August (mean = 3.2 Wm^{-2})
October (mean = 0.7 Wm^{-2})
60°N
30°N
0°
30°S
60°S
30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W
51
33
15
−3
−21
−39
−57
W m^{-2}

Figure8.

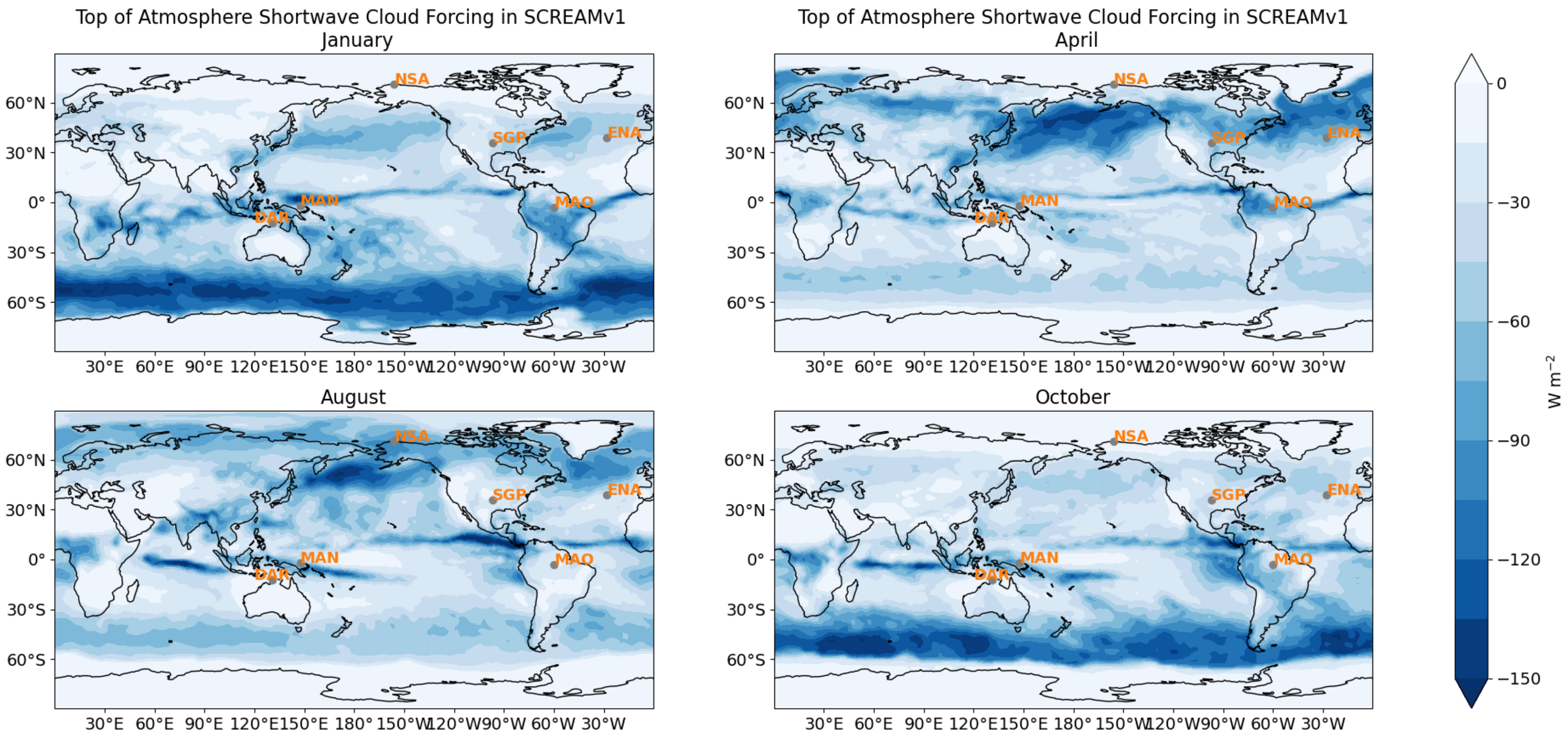

Top of Atmosphere Shortwave Cloud Forcing in SCREAMv1
January
April
August
October
NSA
SGP
ENA
MAN
DAR
MAO
60°N
30°N
0°
30°S
60°S
30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W
0
−30
−60
−90
−120
−150
W m^{-2}

Figure9.

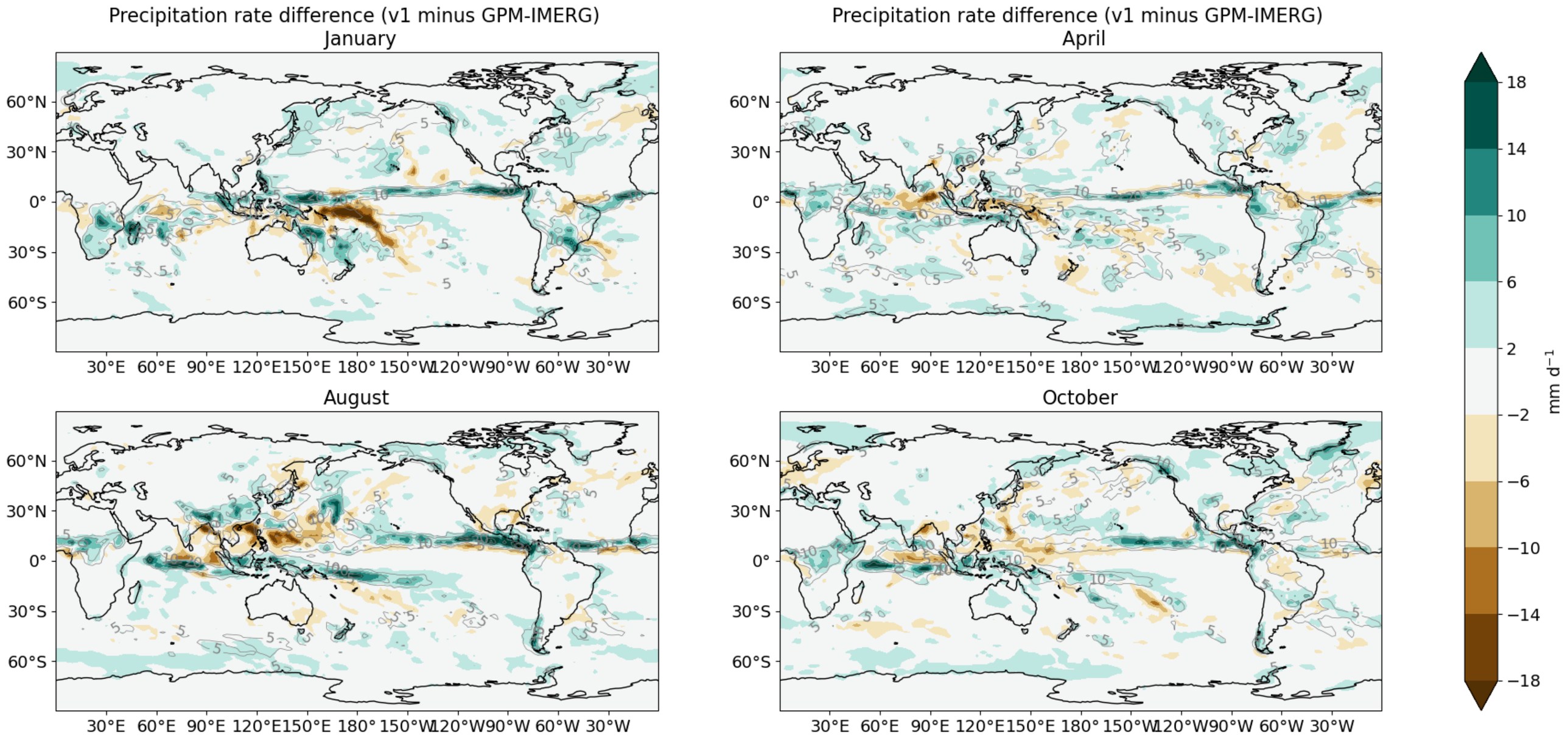

Precipitation rate difference (v1 minus GPM-IMERG)
January
April
August
October
60°N
30°N
0°
30°S
60°S
30°E 60°E 90°E 120°E 150°E 180° 150°W 120°W 90°W 60°W 30°W
18
14
10
6
2
−2
−6
−10
−14
−18
mm d⁻¹

Figure10.

SCREAMv1 minus ERA5 temperature difference

January

April

August

October

Pressure (hPa)

Latitude

K

Figure11.

SCREAMv1 minus ERA5 humidity difference

January

April

August

October

Pressure (hPa)

200
400
600
800
1000

−80 −60 −40 −20 0 20 40 60 80

Latitude

1.9
1.3
0.7
0.1
−0.5
−1.1
−1.7

g kg^{-1}

Figure12.

SCREAMv1 minus ERA5 zonal wind difference

January

April

August

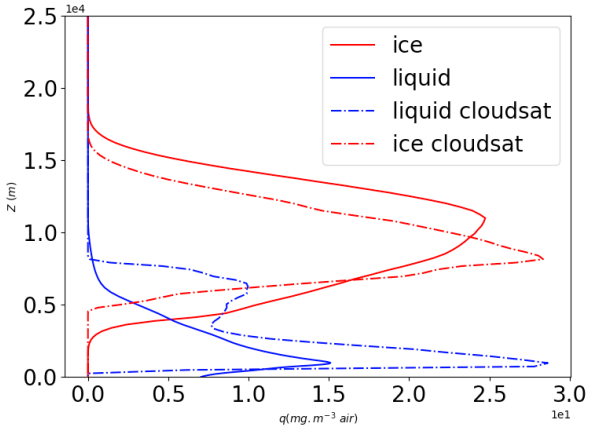
October

Pressure (hPa)

Latitude

m/s

Figure13.


1e4
2.5
2.0
1.5
1.0
0.5
0.0
Z (m)
ice
liquid
liquid cloudsat
ice cloudsat
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1e1
q(mg.m⁻³ air)

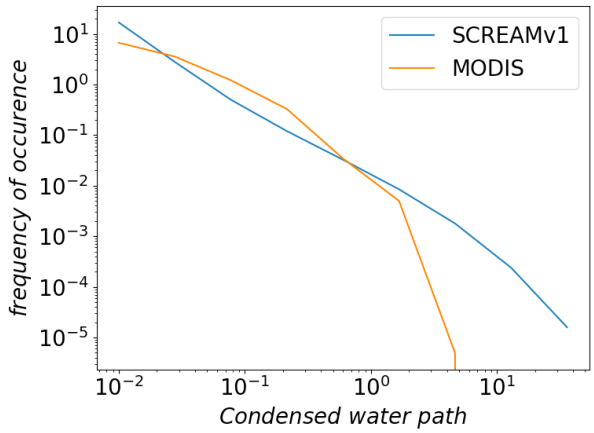

SCREAMv1
MODIS
frequency of occurence
Condensed water path

Figure14.

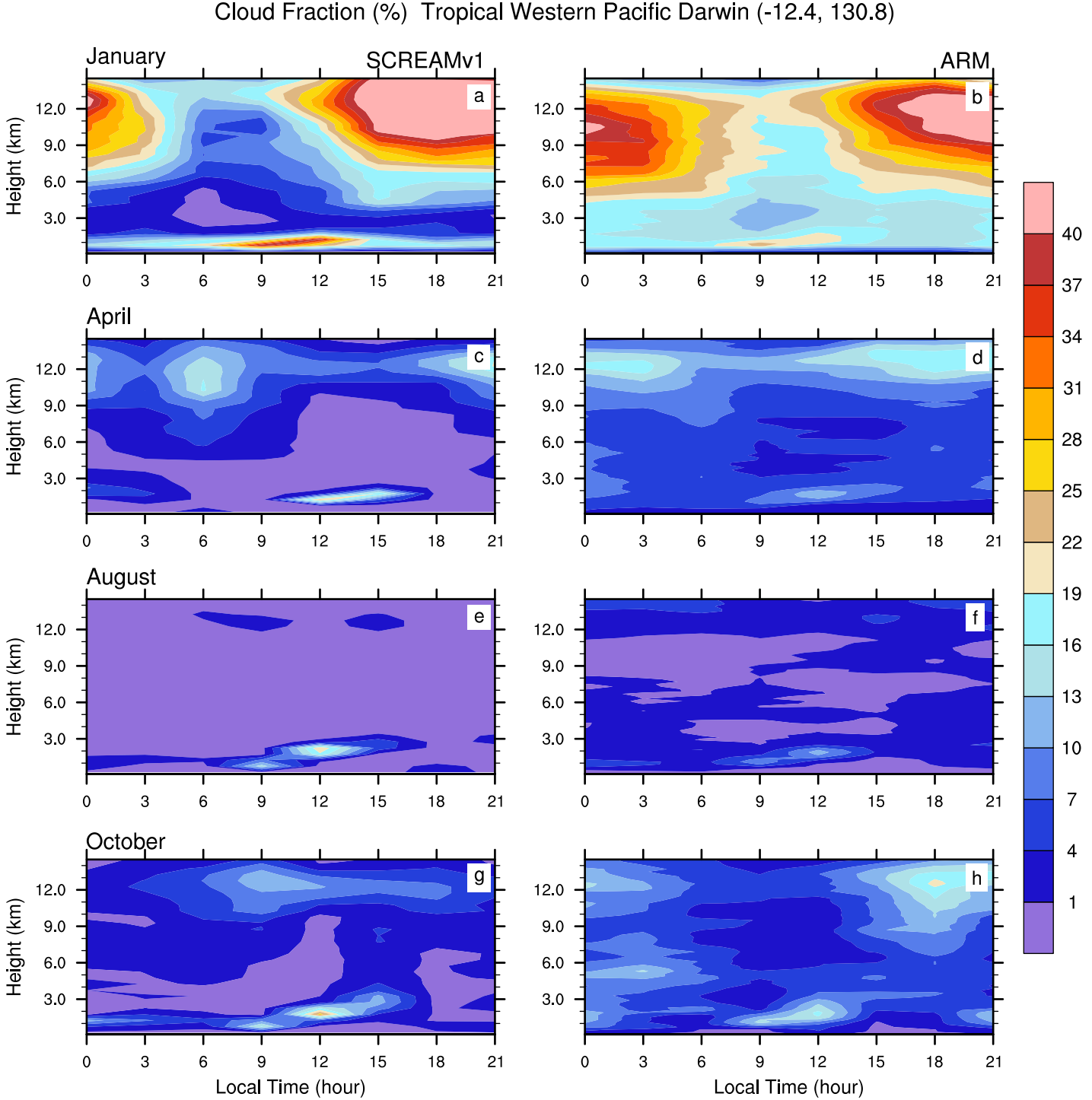

Cloud Fraction (%) Tropical Western Pacific Darwin (-12.4, 130.8)
January
SCREAMv1
ARM
April
August
October
a
b
c
d
e
f
g
h
Height (km)
12.0
9.0
6.0
3.0
0
3
6
9
12
15
18
21
Local Time (hour)
40
37
34
31
28
25
22
19
16
13
10
7
4
1

Figure15.

Tropical Western Pacific Darwin (-12.4, 130.8)

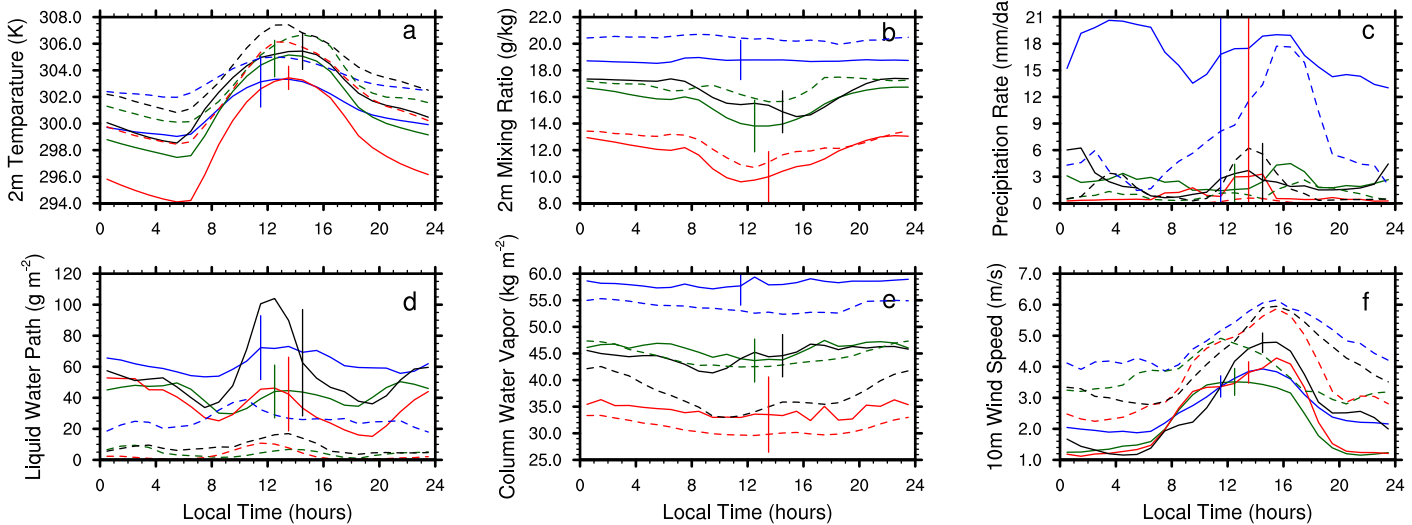


Figure16.

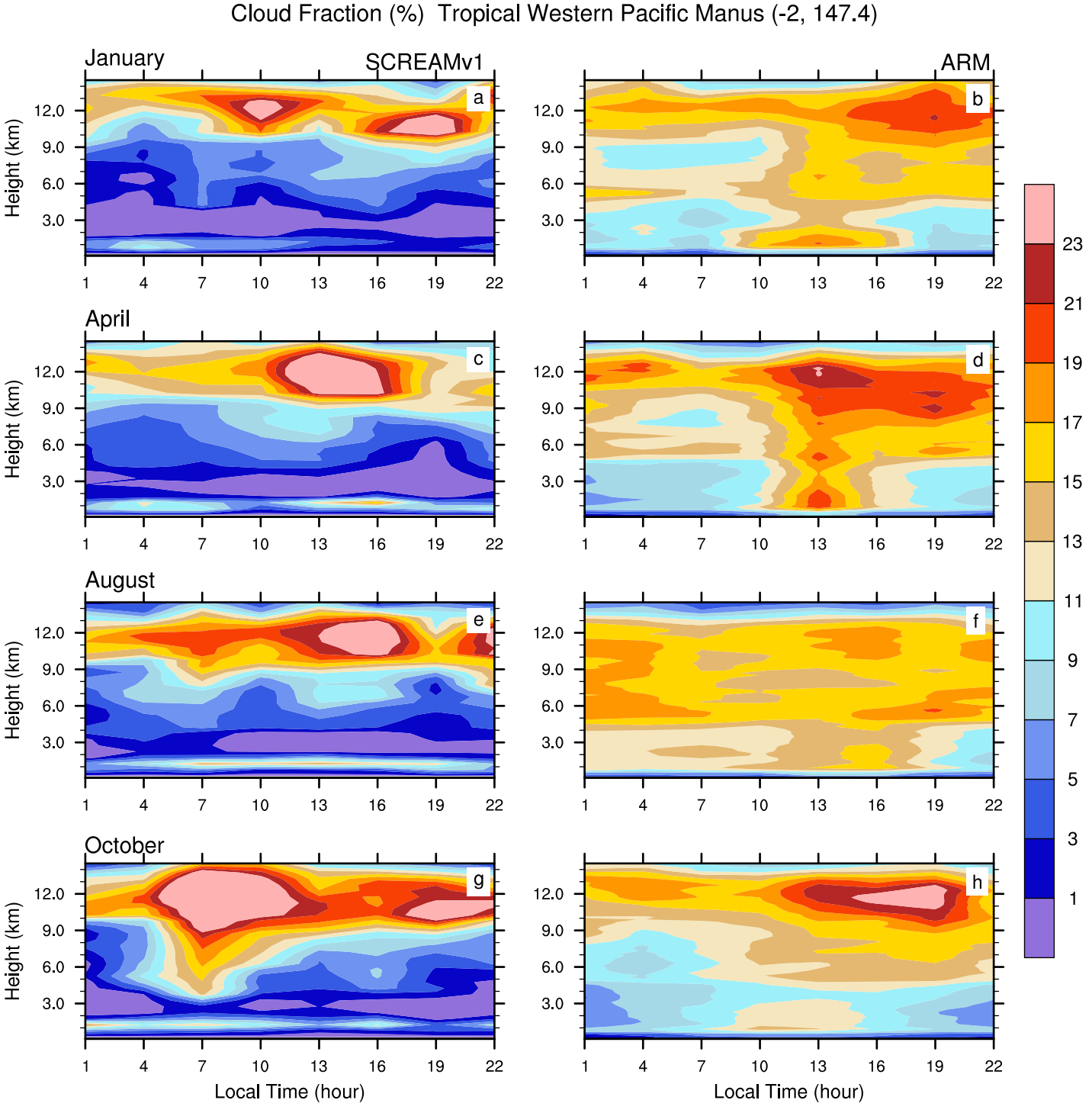

Cloud Fraction (%) Tropical Western Pacific Manus (-2, 147.4)
January
SCREAMv1
ARM
April
August
October
a
b
c
d
e
f
g
h
Height (km)
12.0
9.0
6.0
3.0
1
4
7
10
13
16
19
22
Local Time (hour)
23
21
19
17
15
13
11
9
7
5
3
1

Figure17.

Tropical Western Pacific Manus (-2, 147.4)

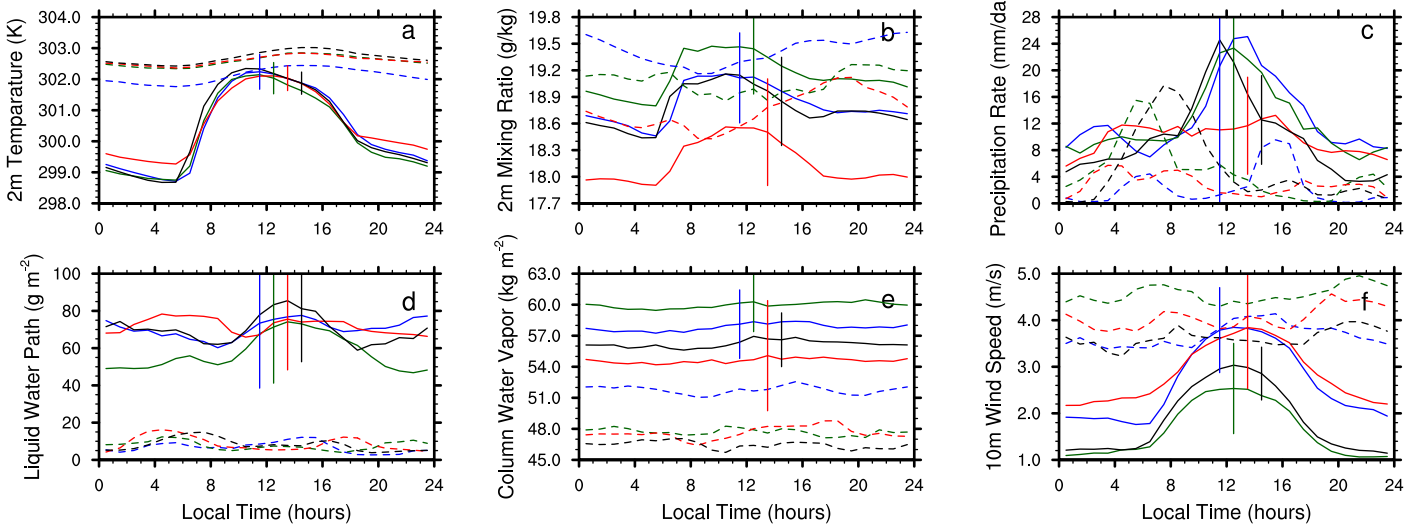


Figure18.

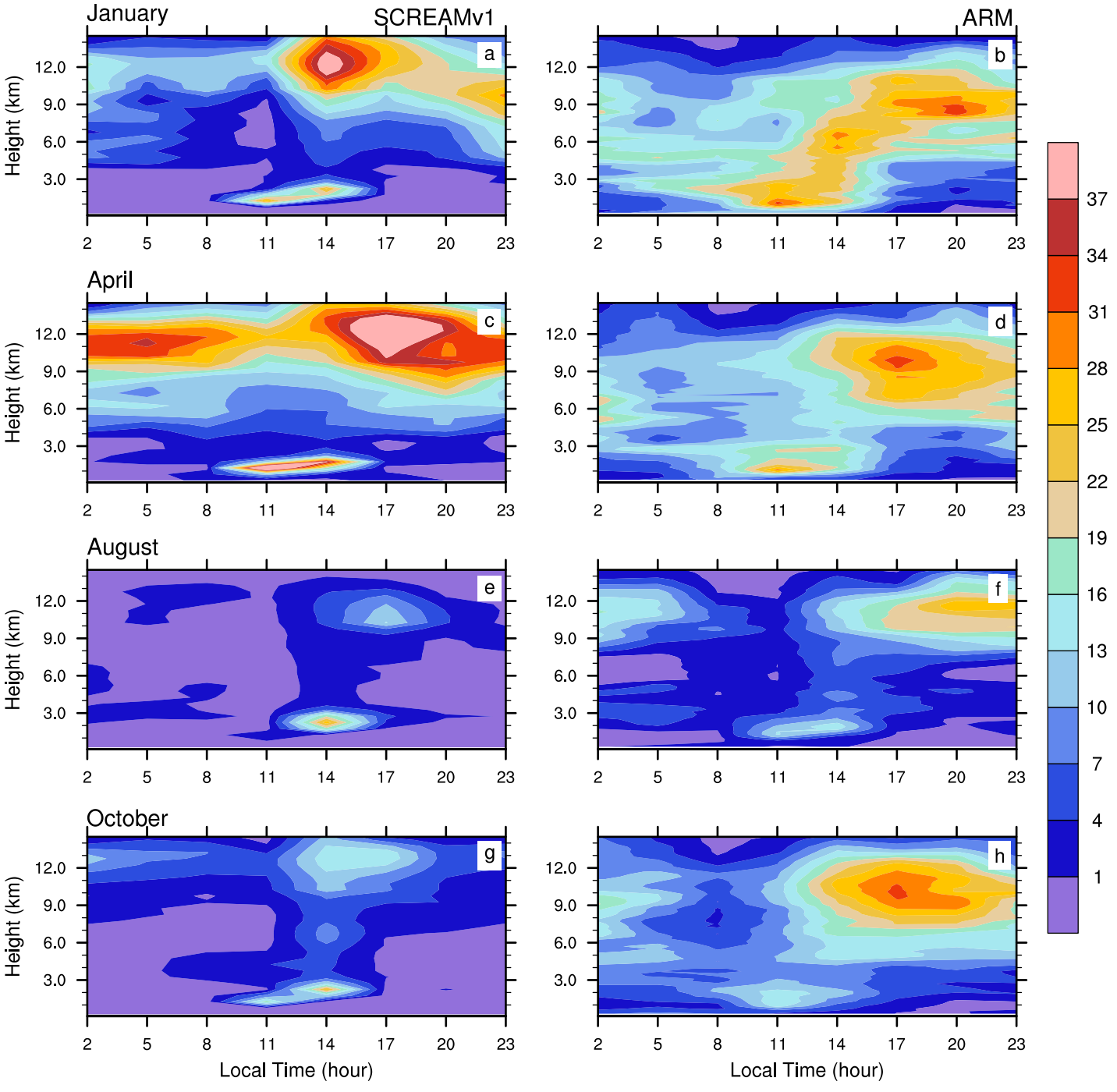

Cloud Fraction (%) GoAmazon (-3.2, -60.5)
January
SCREAMv1
ARM
April
August
October
a
b
c
d
e
f
g
h
Height (km)
3.0
6.0
9.0
12.0
2
5
8
11
14
17
20
23
Local Time (hour)
1
4
7
10
13
16
19
22
25
28
31
34
37

Figure19.

GoAmazon (-3.2, -60.5)

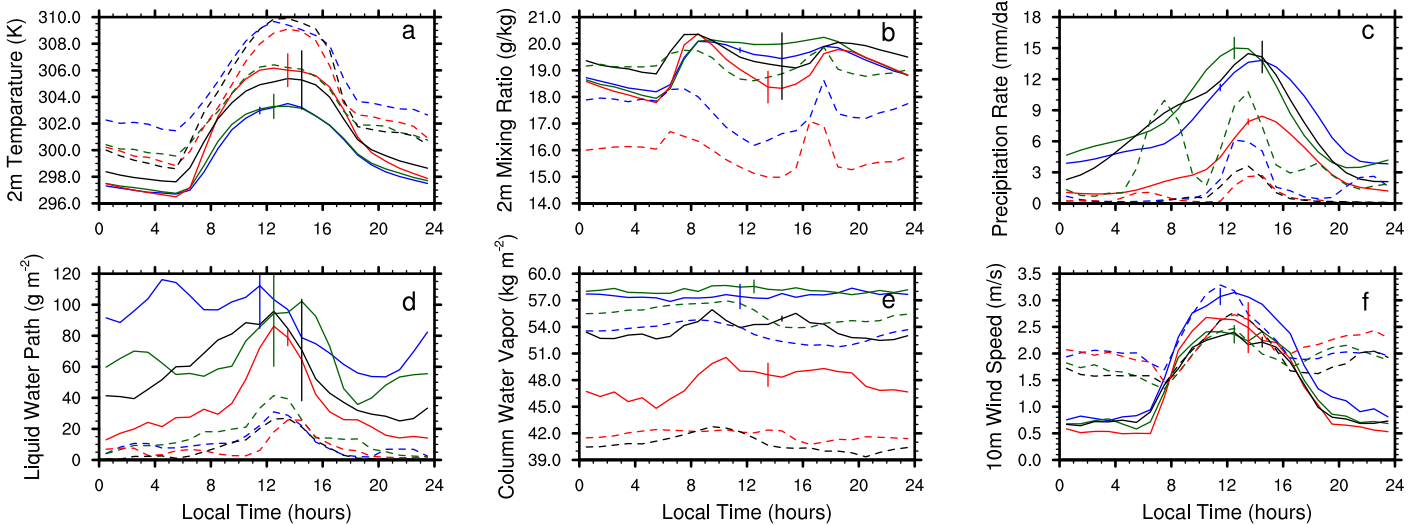


Figure20.

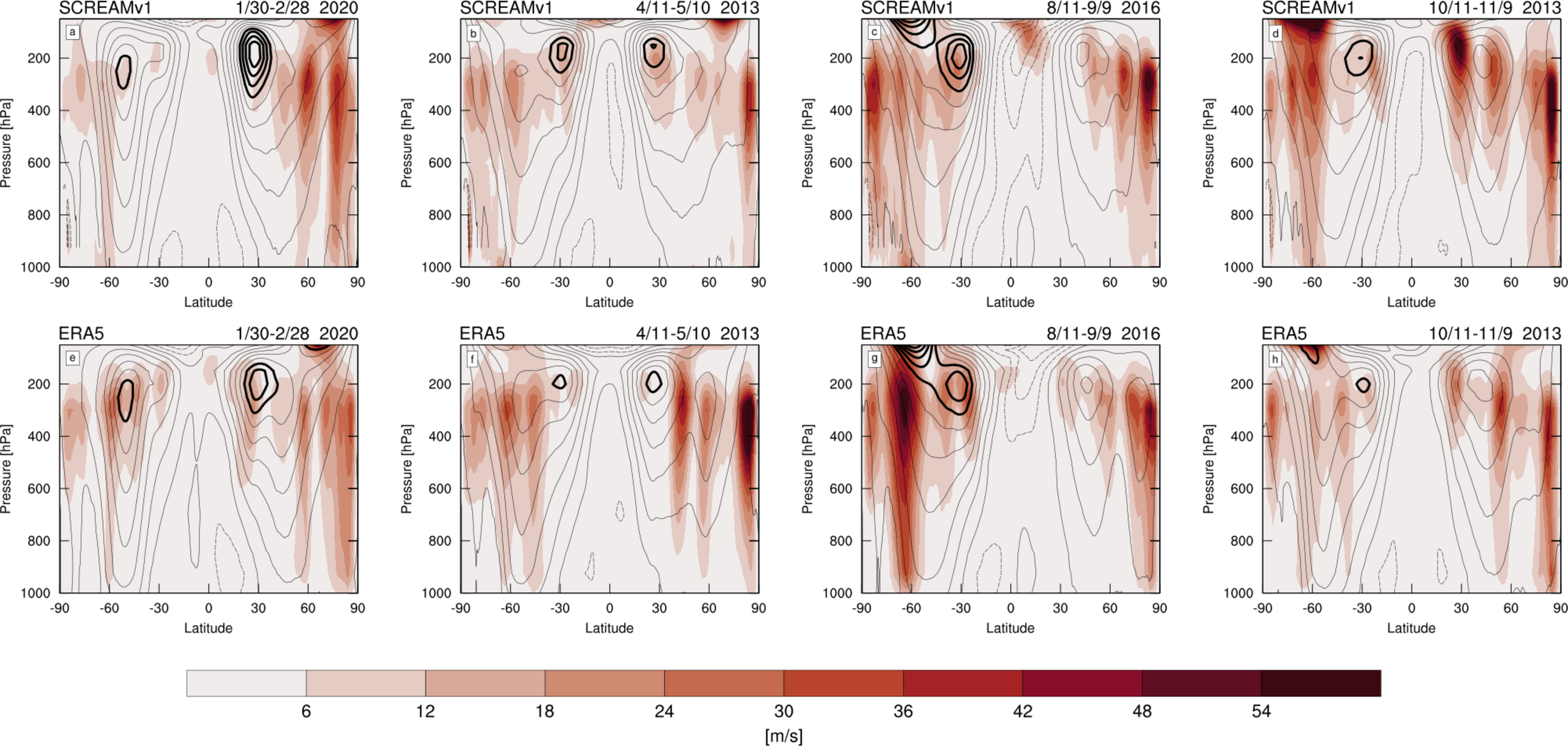
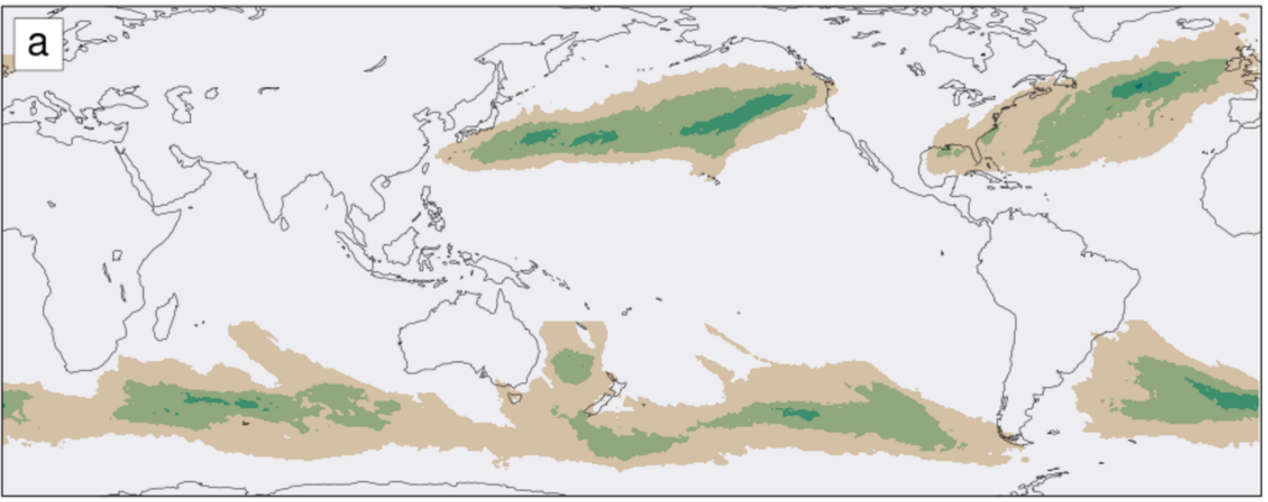
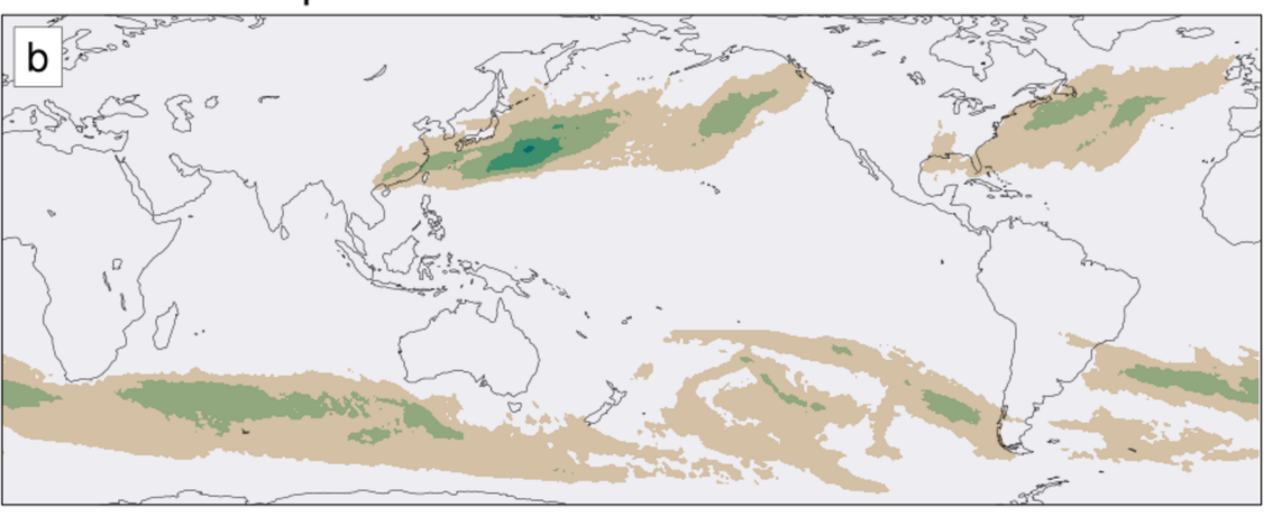

SCREAMv1 1/30-2/28 2020
SCREAMv1 4/11-5/10 2013
SCREAMv1 8/11-9/9 2016
SCREAMv1 10/11-11/9 2013
ERA5 1/30-2/28 2020
ERA5 4/11-5/10 2013
ERA5 8/11-9/9 2016
ERA5 10/11-11/9 2013
a
b
c
d
e
f
g
h
Pressure [hPa]
200
400
600
800
1000
Latitude
-90
-60
-30
0
30
60
90
6
12
18
24
30
36
42
48
54
[m/s]

Figure21.

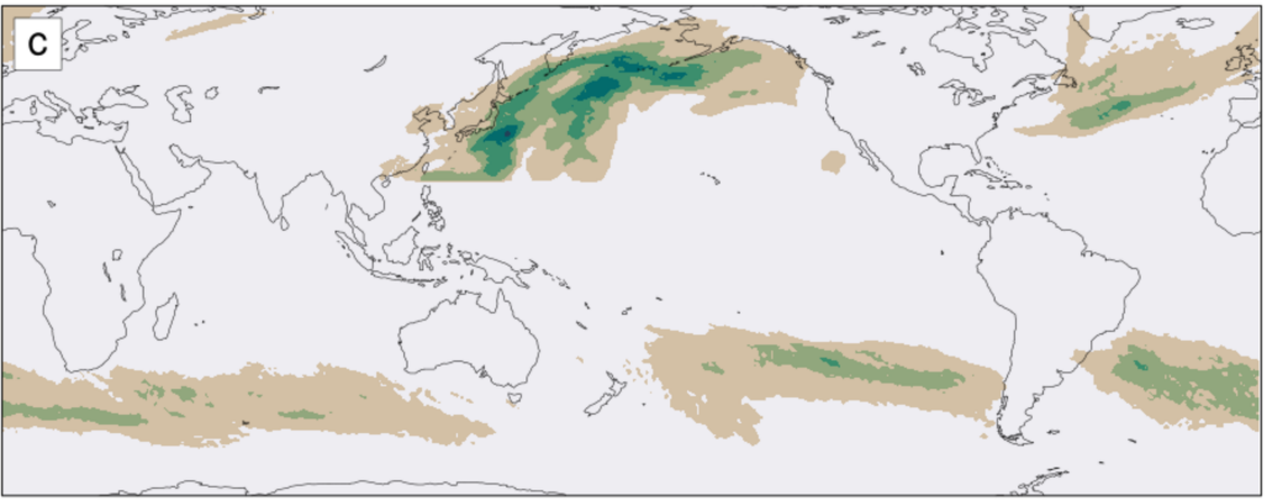
SCREAMv1 Jan 2020



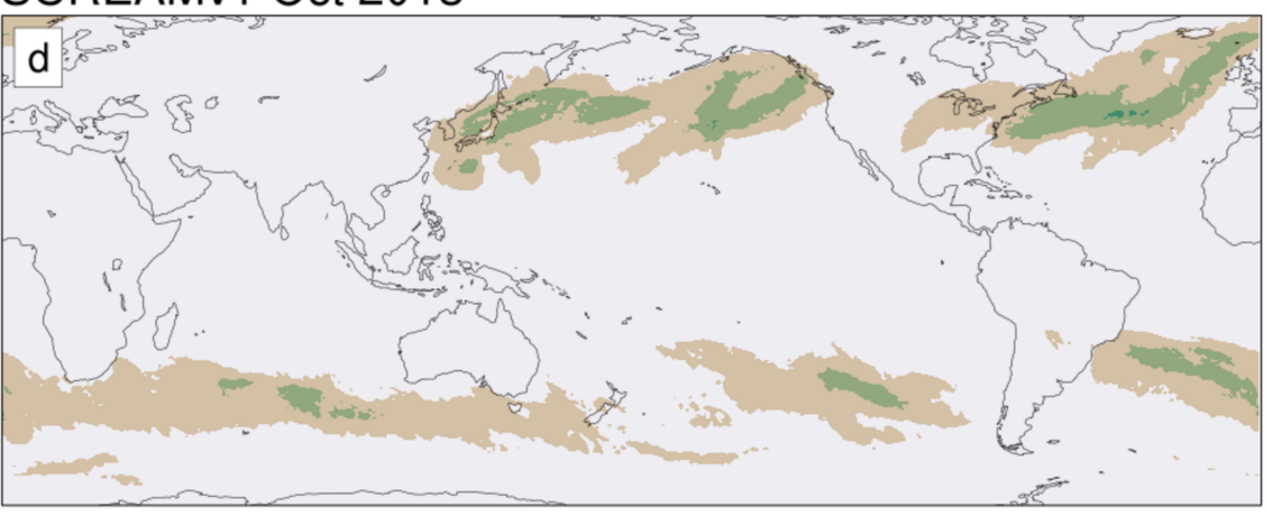
SCREAMv1 Apr 2013



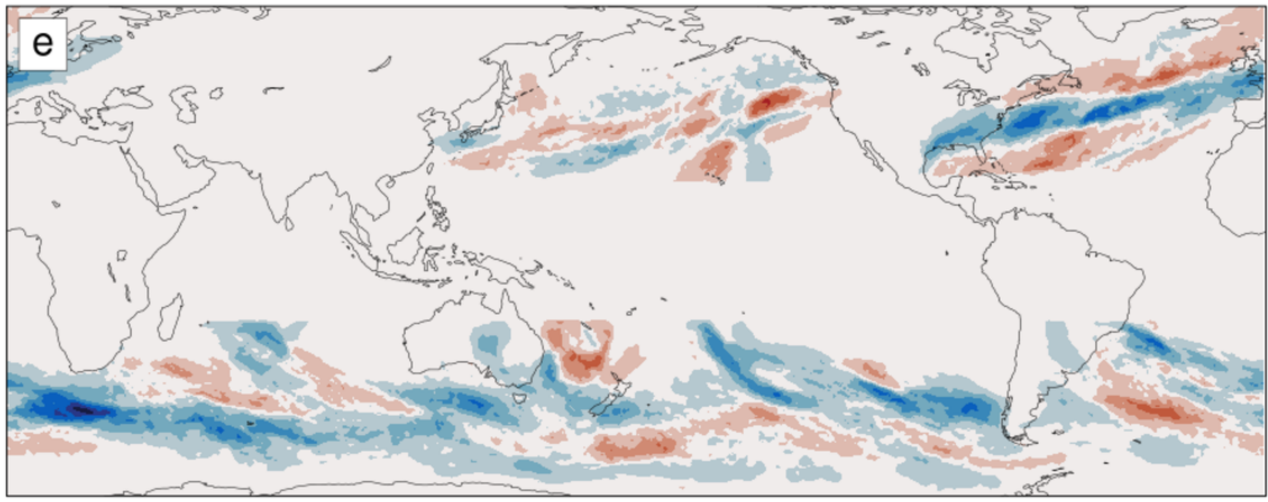
SCREAMv1 Aug 2016



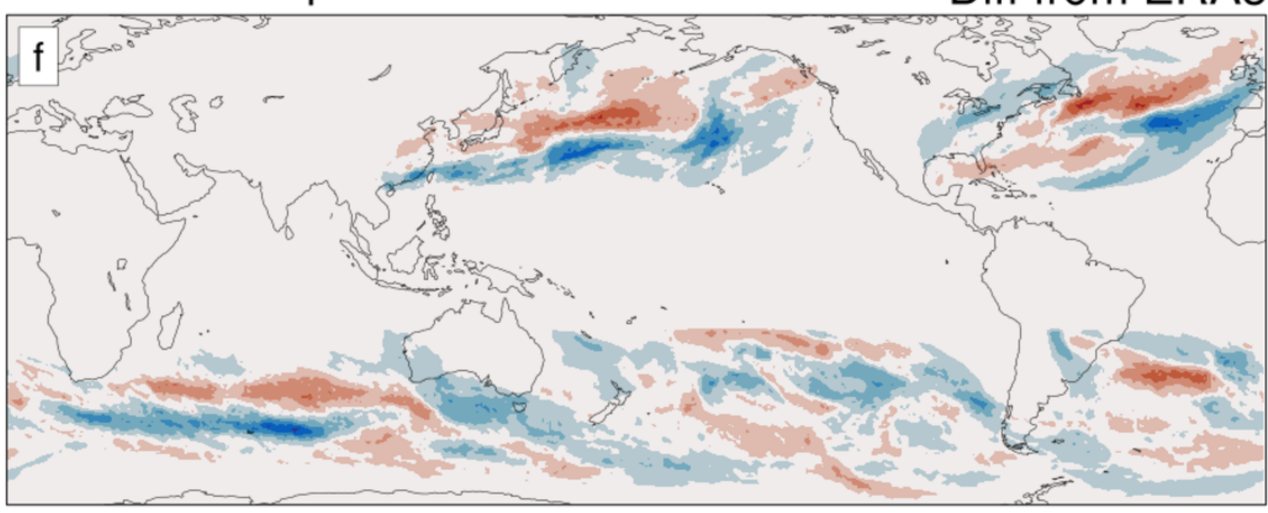
SCREAMv1 Oct 2013



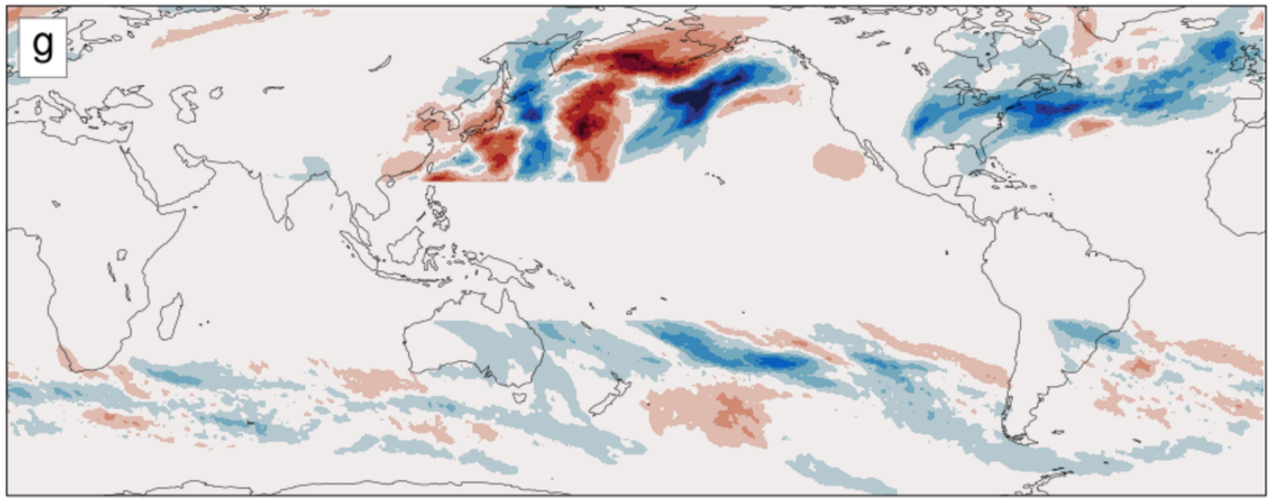
SCREAMv1 Jan 2020 Diff from ERA5



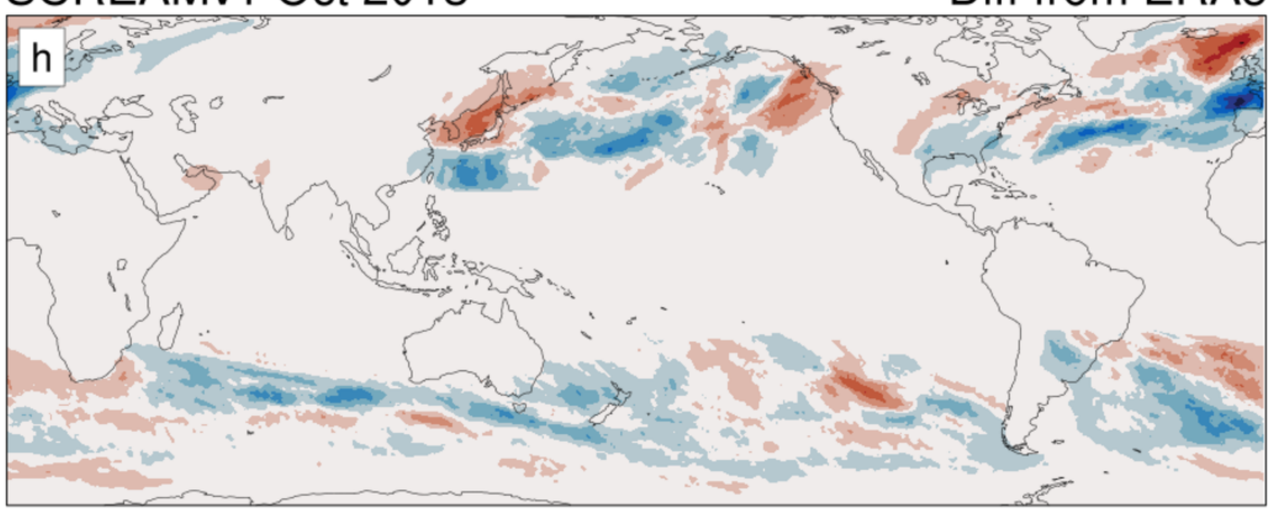
SCREAMv1 Apr 2013 Diff from ERA5



SCREAMv1 Aug 2016 Diff from ERA5



SCREAMv1 Oct 2013 Diff from ERA5



.04 .12 .2 .28 .36 .44

Time Fraction of Atmospheric River

-.22 -.14 -.06 .02 .1 .18

Time Fraction of Atmospheric River

Figure22.

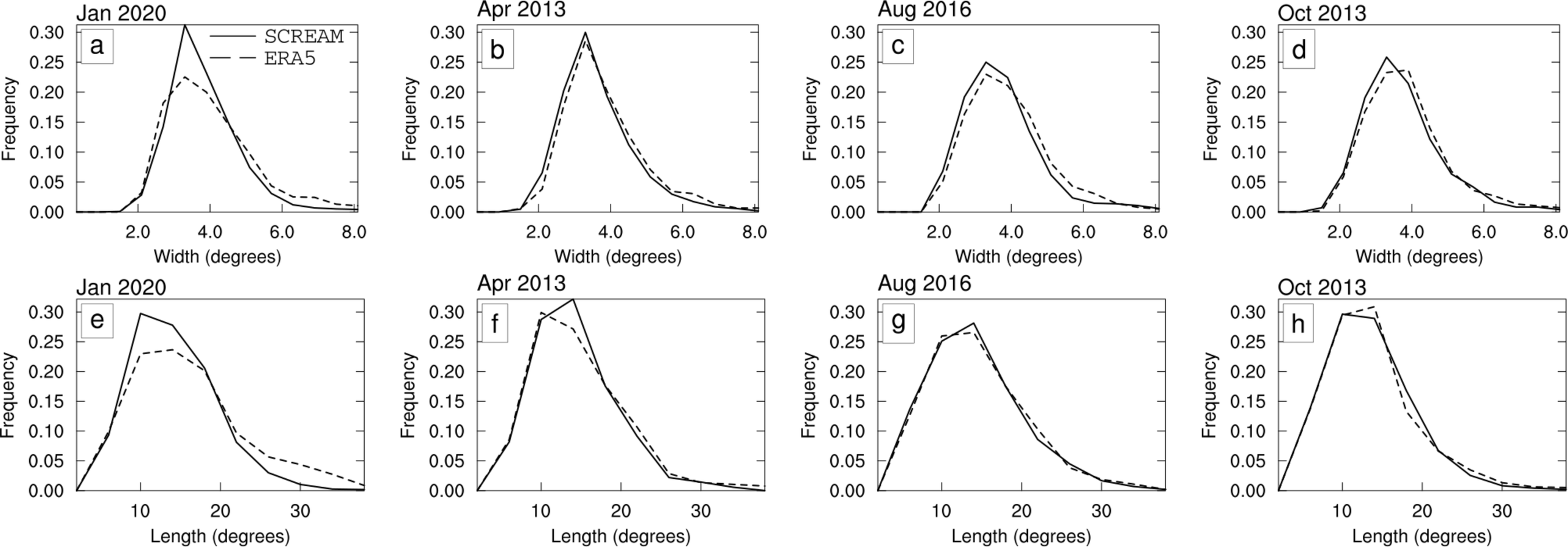

Jan 2020
Apr 2013
Aug 2016
Oct 2013
a
b
c
d
e
f
g
h
SCREAM
ERA5
Frequency
Width (degrees)
Length (degrees)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
2.0
4.0
6.0
8.0
10
20
30

Figure23.

IMERG

April 5, 2013

a

SCREAM

April 5, 2013

b

55°N
50°N
45°N
40°N
35°N
30°N
25°N
Latitude

MTN HP MP LP

SOUTHEAST

120°W 110°W 100°W 90°W 80°W
Longitude

IMERG

April 3-7, 2013

c

SCREAM

April 3-7, 2013

d

7.0
6.0
5.0
4.0
3.0
Time [days]

250 260 270 280
Longitude

0.5 0.9 1.8 3.3 6.3 11.9 22.4 42.2 79.6 150.0

[mm/day]

Figure24.

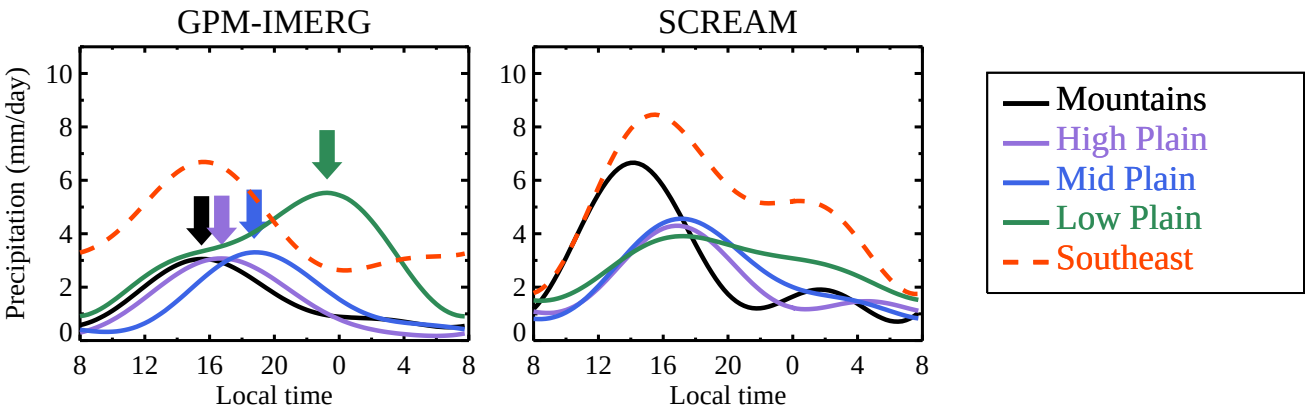

GPM-IMERG
SCREAM
Precipitation (mm/day)
10
8
6
4
2
0
8
12
16
20
0
4
8
Local time
Mountains
High Plain
Mid Plain
Low Plain
Southeast

Figure25.

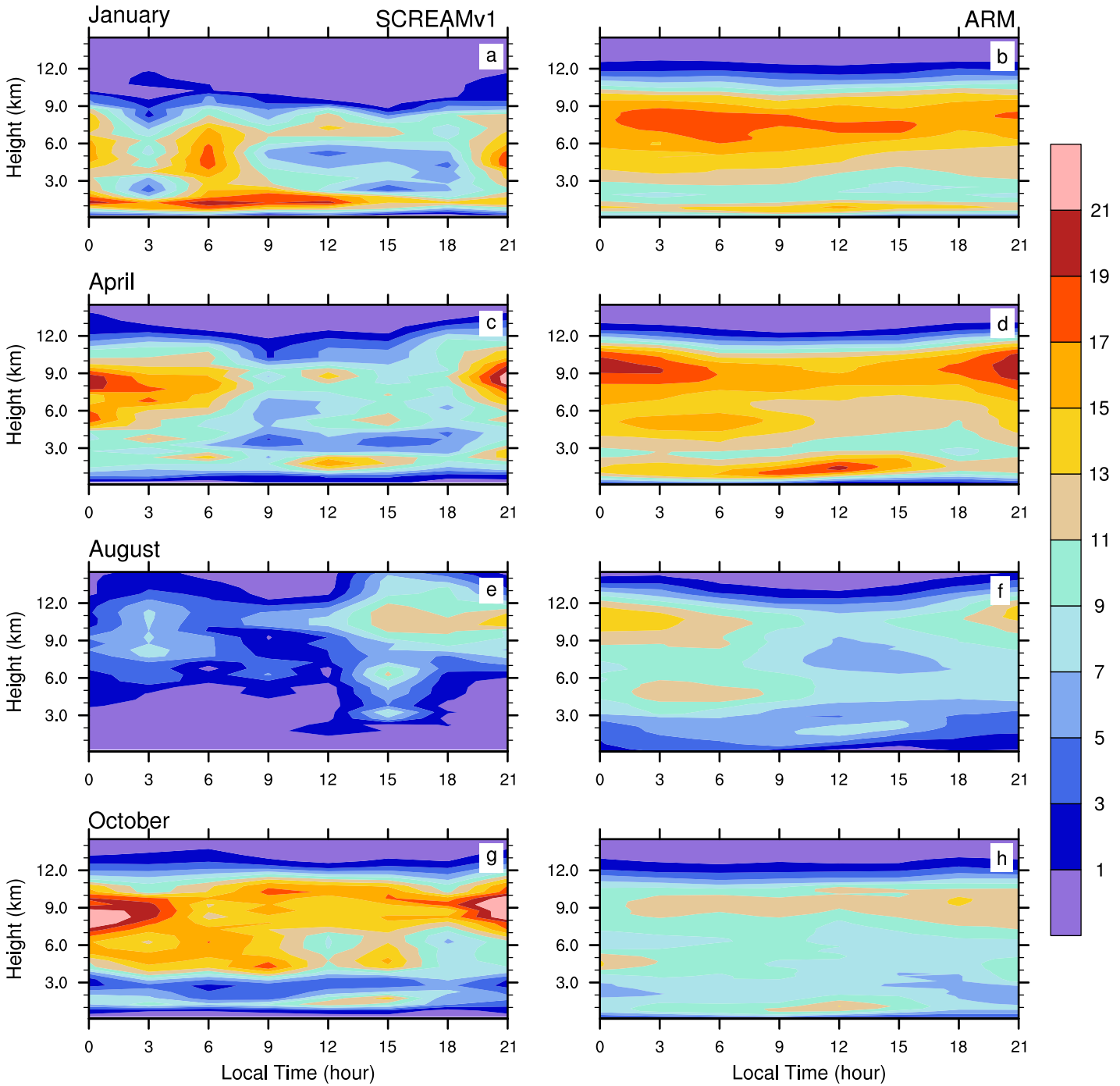

Cloud Fraction (%) Southern Great Plains (36.6, -97.4)
January
SCREAMv1
ARM
April
August
October
a
b
c
d
e
f
g
h
Height (km)
12.0
9.0
6.0
3.0
0
3
6
9
12
15
18
21
Local Time (hour)
21
19
17
15
13
11
9
7
5
3
1

Figure26.

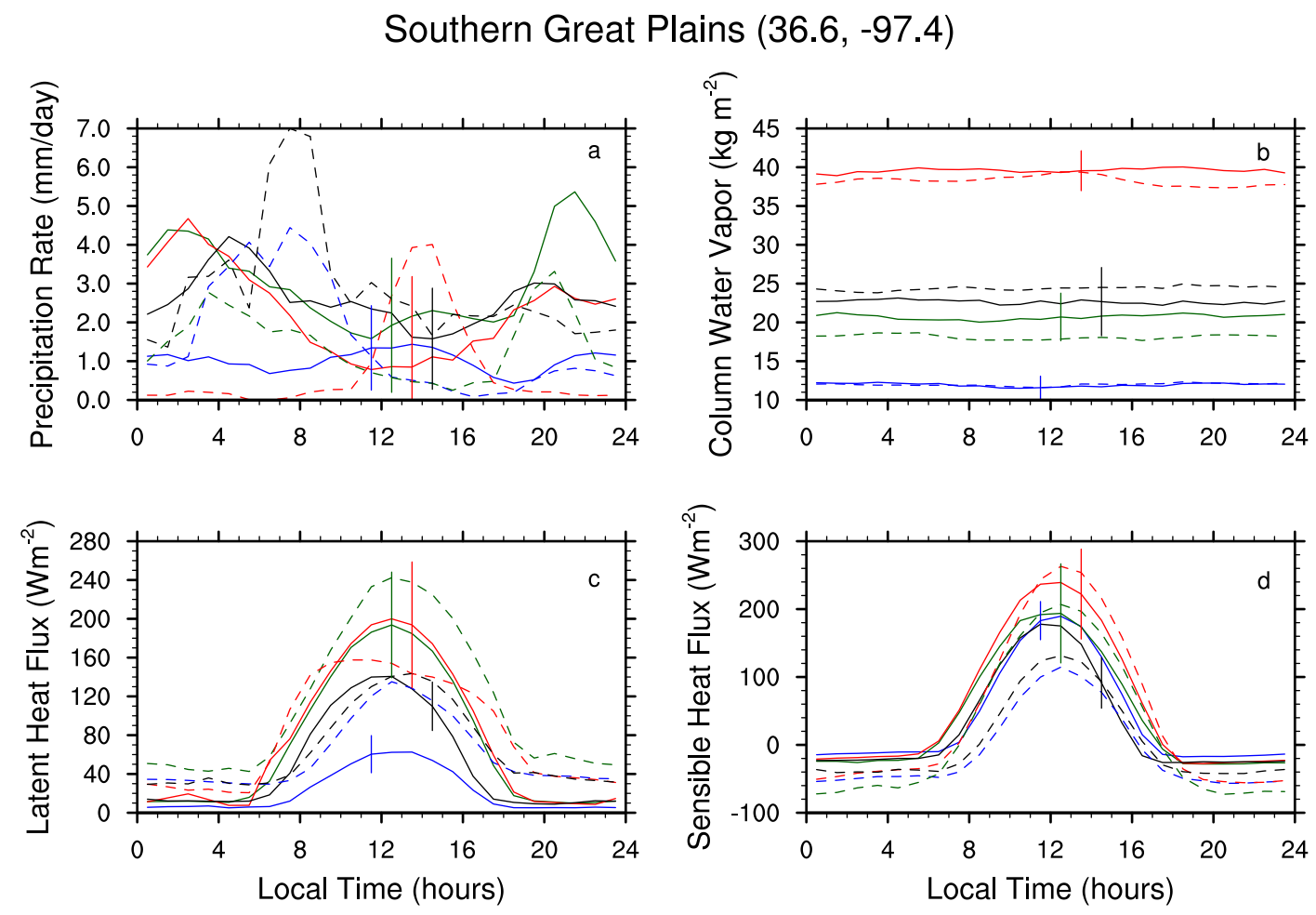

Southern Great Plains (36.6, -97.4)
Precipitation Rate (mm/day)
Column Water Vapor (kg m-2)
Latent Heat Flux (Wm-2)
Sensible Heat Flux (Wm-2)
Local Time (hours)
Local Time (hours)
a
b
c
d

Figure27.

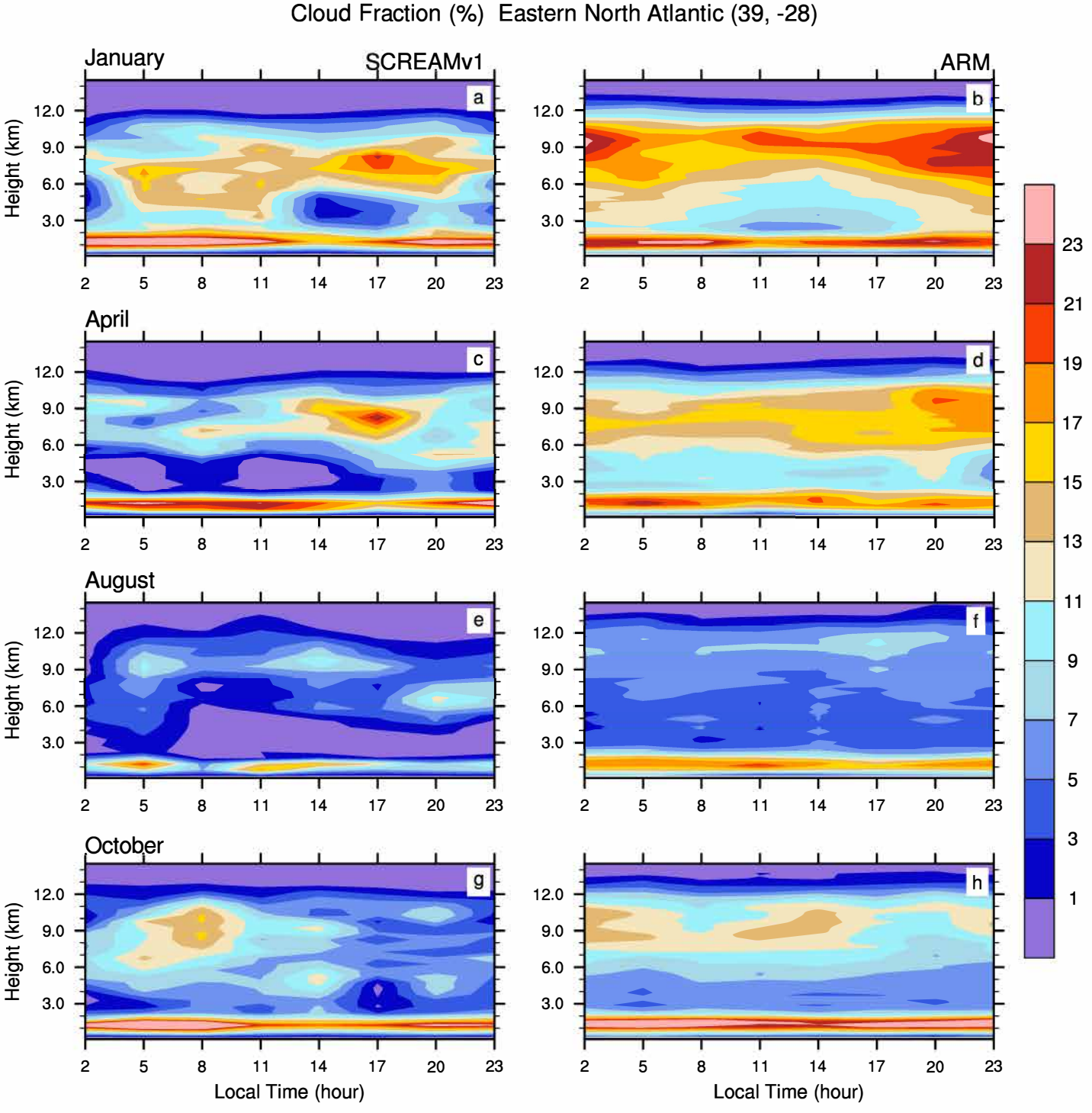

Cloud Fraction (%) Eastern North Atlantic (39, -28)
January
SCREAMv1
ARM
April
August
October
a
b
c
d
e
f
g
h
Height (km)
12.0
9.0
6.0
3.0
2
5
8
11
14
17
20
23
Local Time (hour)
23
21
19
17
15
13
11
9
7
5
3
1

Figure28.

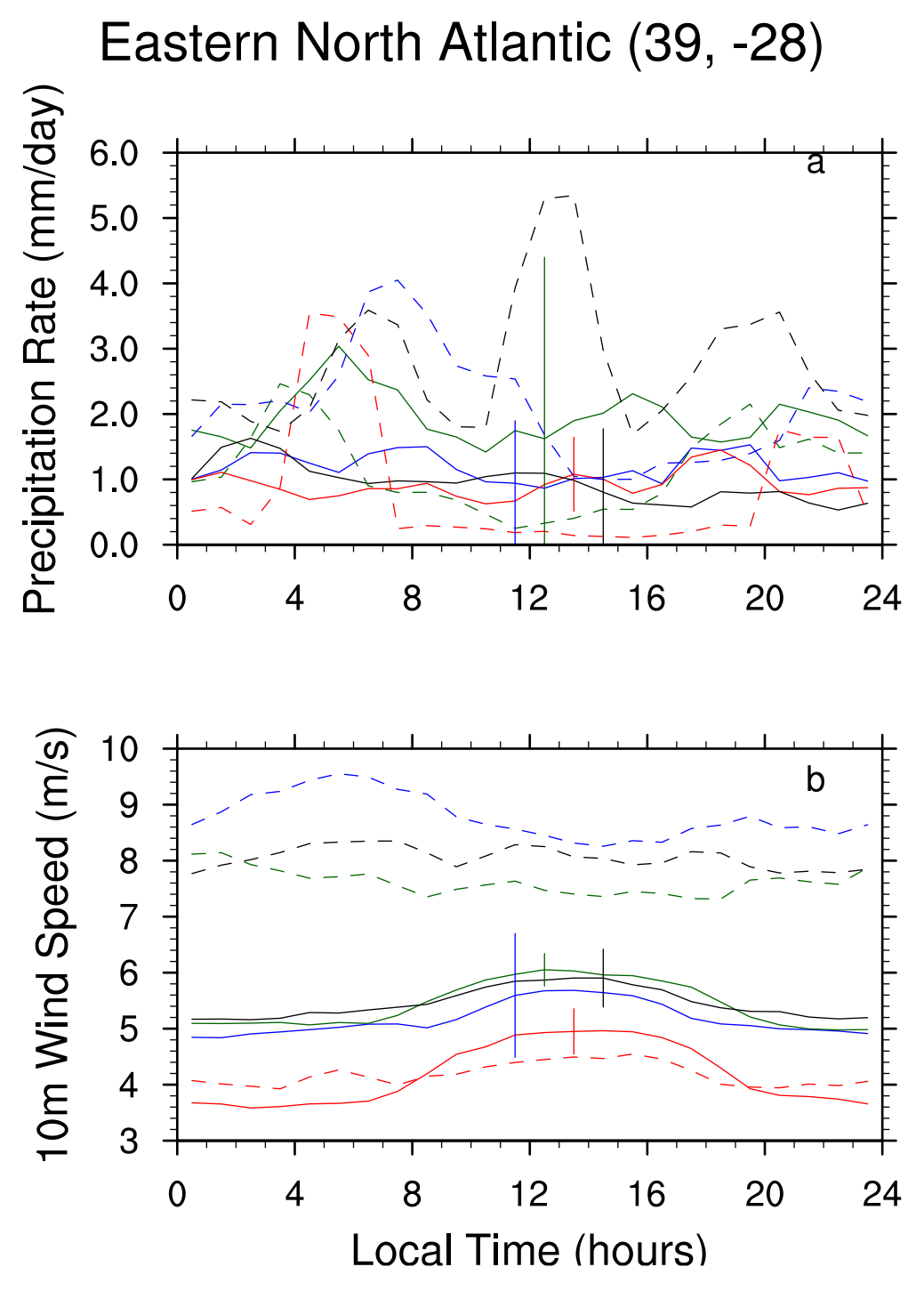

Eastern North Atlantic (39, -28)
Precipitation Rate (mm/day)
6.0
5.0
4.0
3.0
2.0
1.0
0.0
a
0
4
8
12
16
20
24
10m Wind Speed (m/s)
10
9
8
7
6
5
4
3
b
0
4
8
12
16
20
24
Local Time (hours)

Figure29.

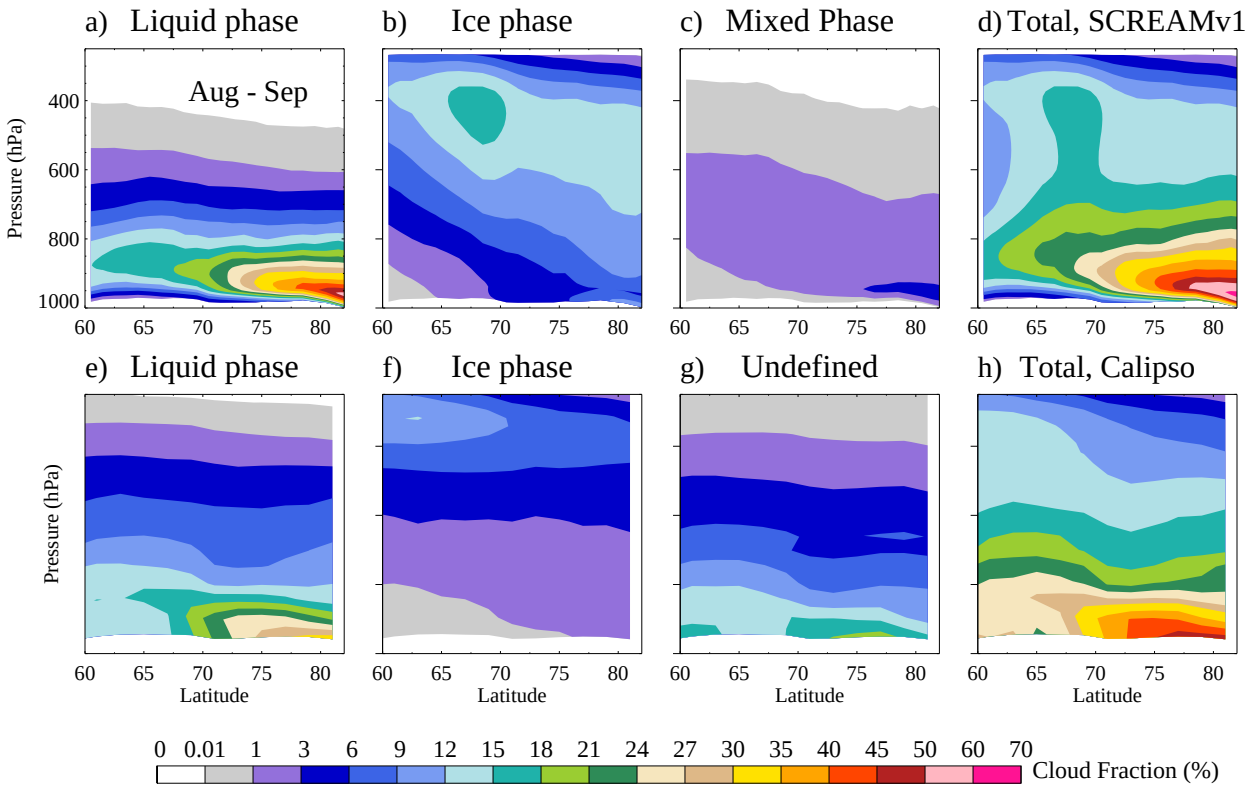

a) Liquid phase
b) Ice phase
c) Mixed Phase
d) Total, SCREAMv1
Aug - Sep
e) Liquid phase
f) Ice phase
g) Undefined
h) Total, Calipso
Pressure (hPa)
400
600
800
1000
60 65 70 75 80
Latitude
0 0.01 1 3 6 9 12 15 18 21 24 27 30 35 40 45 50 60 70
Cloud Fraction (%)

Figure30.

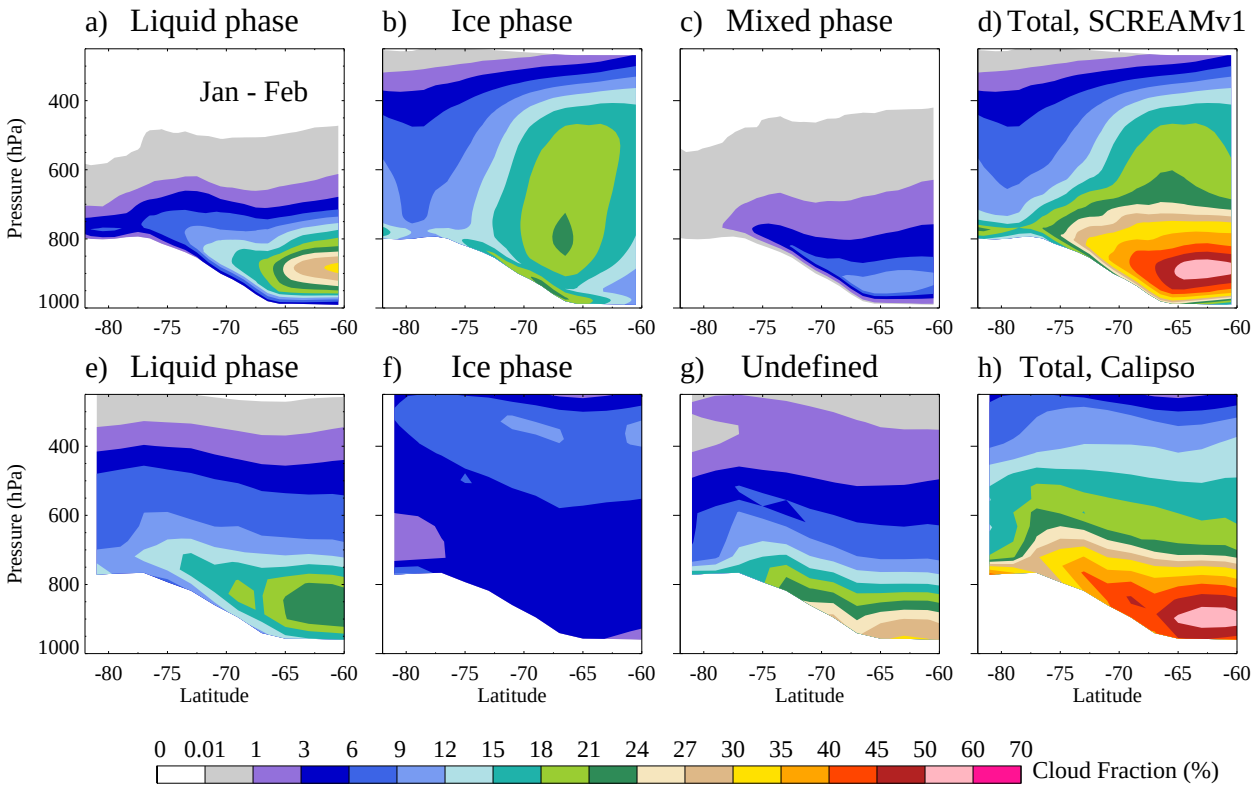

a) Liquid phase
b) Ice phase
c) Mixed phase
d) Total, SCREAMv1
Jan - Feb
e) Liquid phase
f) Ice phase
g) Undefined
h) Total, Calipso
Pressure (hPa)
400
600
800
1000
-80 -75 -70 -65 -60
Latitude
0 0.01 1 3 6 9 12 15 18 21 24 27 30 35 40 45 50 60 70
Cloud Fraction (%)

Figure31.

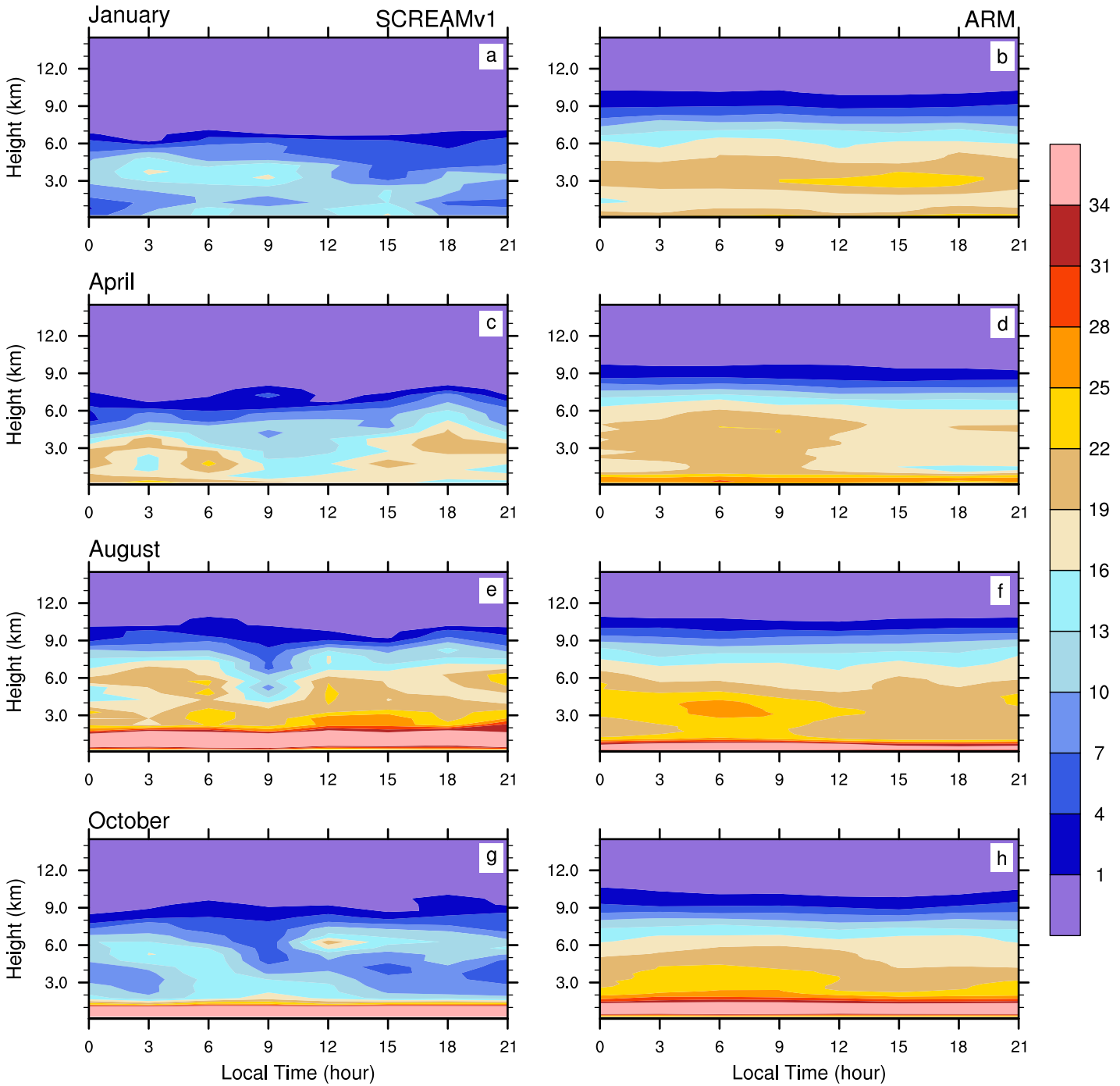

Cloud Fraction (%) North Slope of Alaska (71.3, -156.6)
January
SCREAMv1
ARM
April
August
October
a
b
c
d
e
f
g
h
Height (km)
12.0
9.0
6.0
3.0
0
3
6
9
12
15
18
21
Local Time (hour)
34
31
28
25
22
19
16
13
10
7
4
1

Figure32.

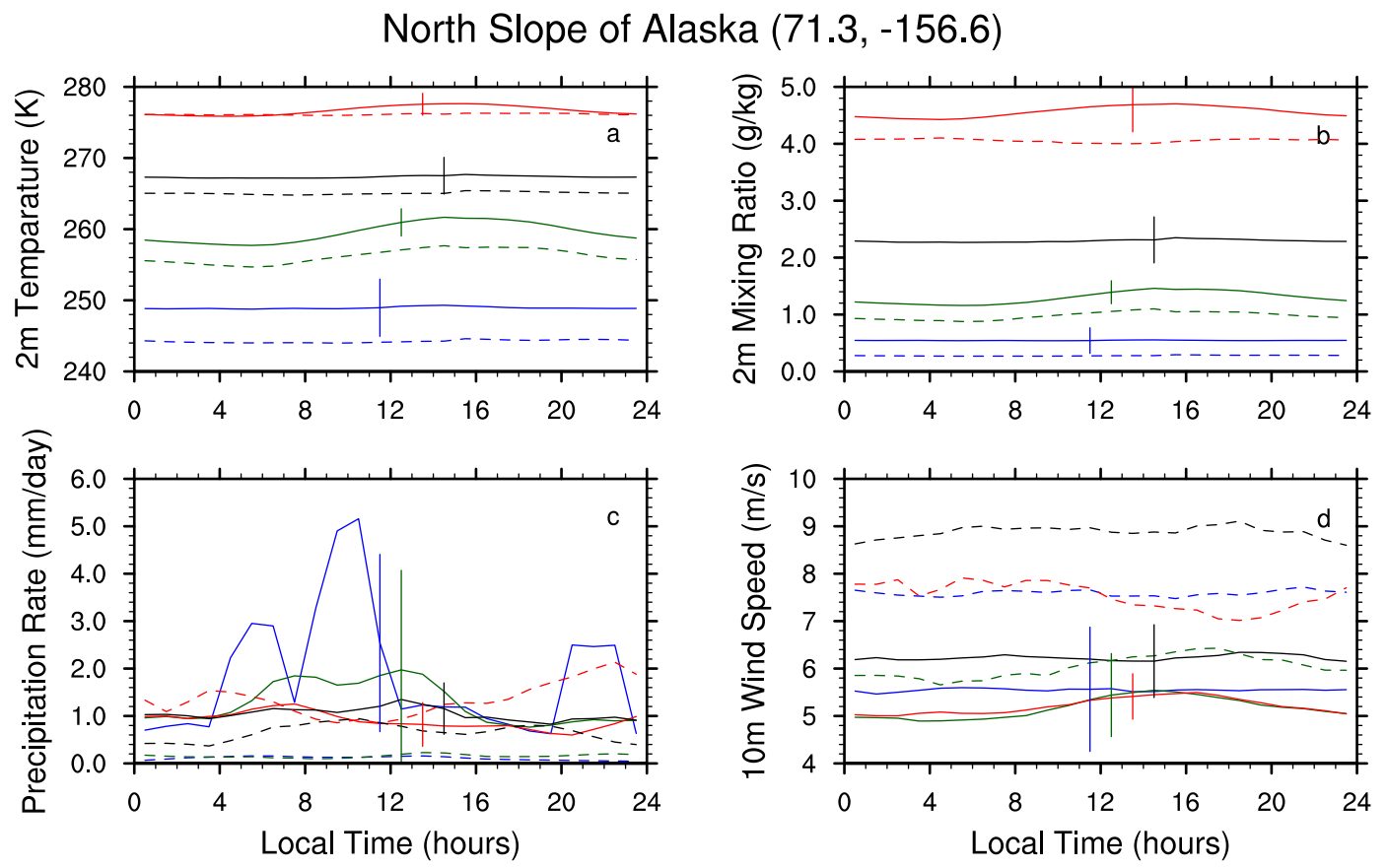

North Slope of Alaska (71.3, -156.6)
2m Temparature (K)
280
270
260
250
240
a
0
4
8
12
16
20
24
2m Mixing Ratio (g/kg)
5.0
4.0
3.0
2.0
1.0
0.0
b
Precipitation Rate (mm/day)
6.0
5.0
4.0
3.0
2.0
1.0
0.0
c
Local Time (hours)
10m Wind Speed (m/s)
10
9
8
7
6
5
4
d
Local Time (hours)

Figure33.

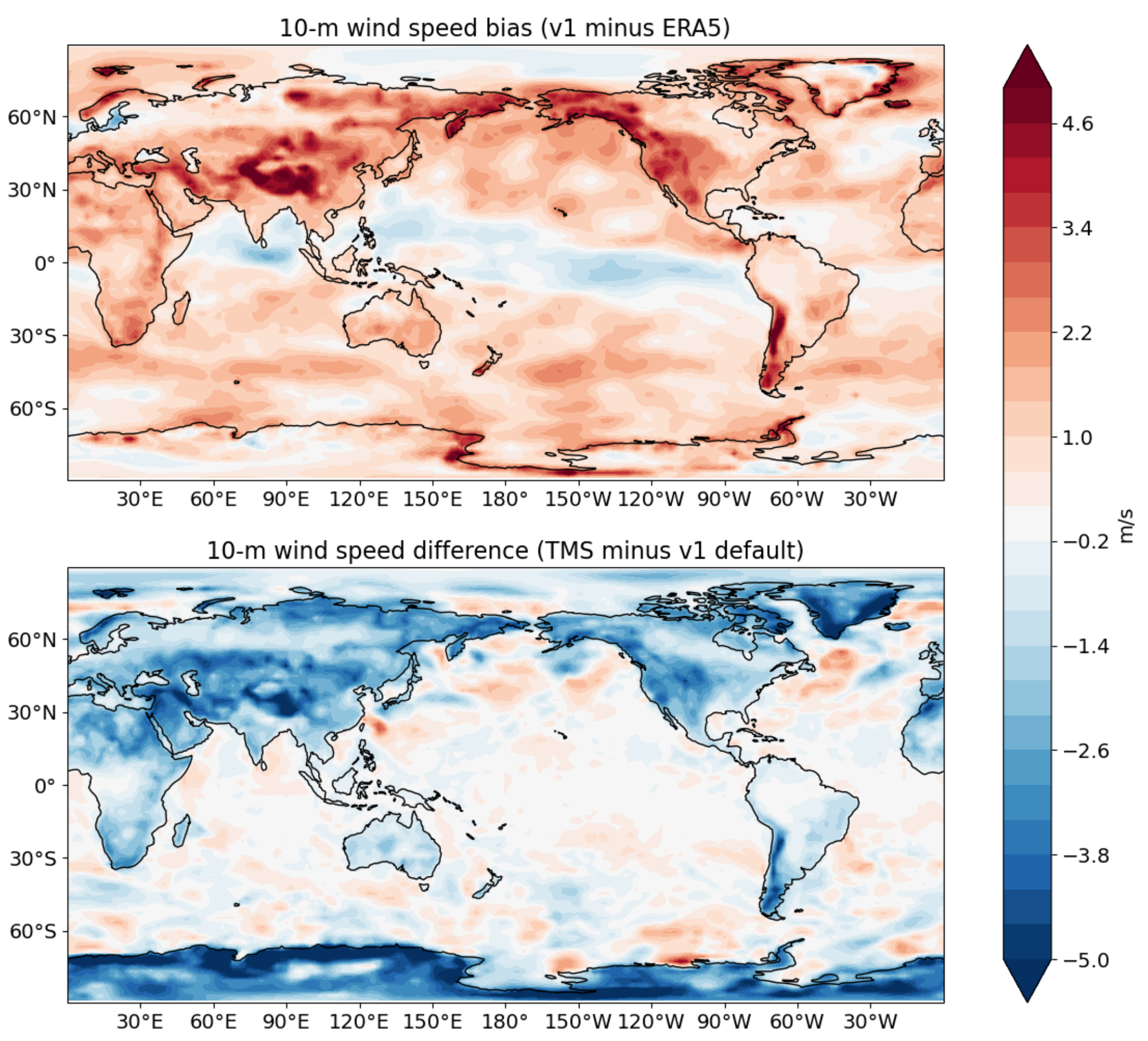

10-m wind speed bias (v1 minus ERA5)
60°N
30°N
0°
30°S
60°S
30°E
60°E
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
10-m wind speed difference (TMS minus v1 default)
60°N
30°N
0°
30°S
60°S
30°E
60°E
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
4.6
3.4
2.2
1.0
−0.2
−1.4
−2.6
−3.8
−5.0
m/s

Figure34.

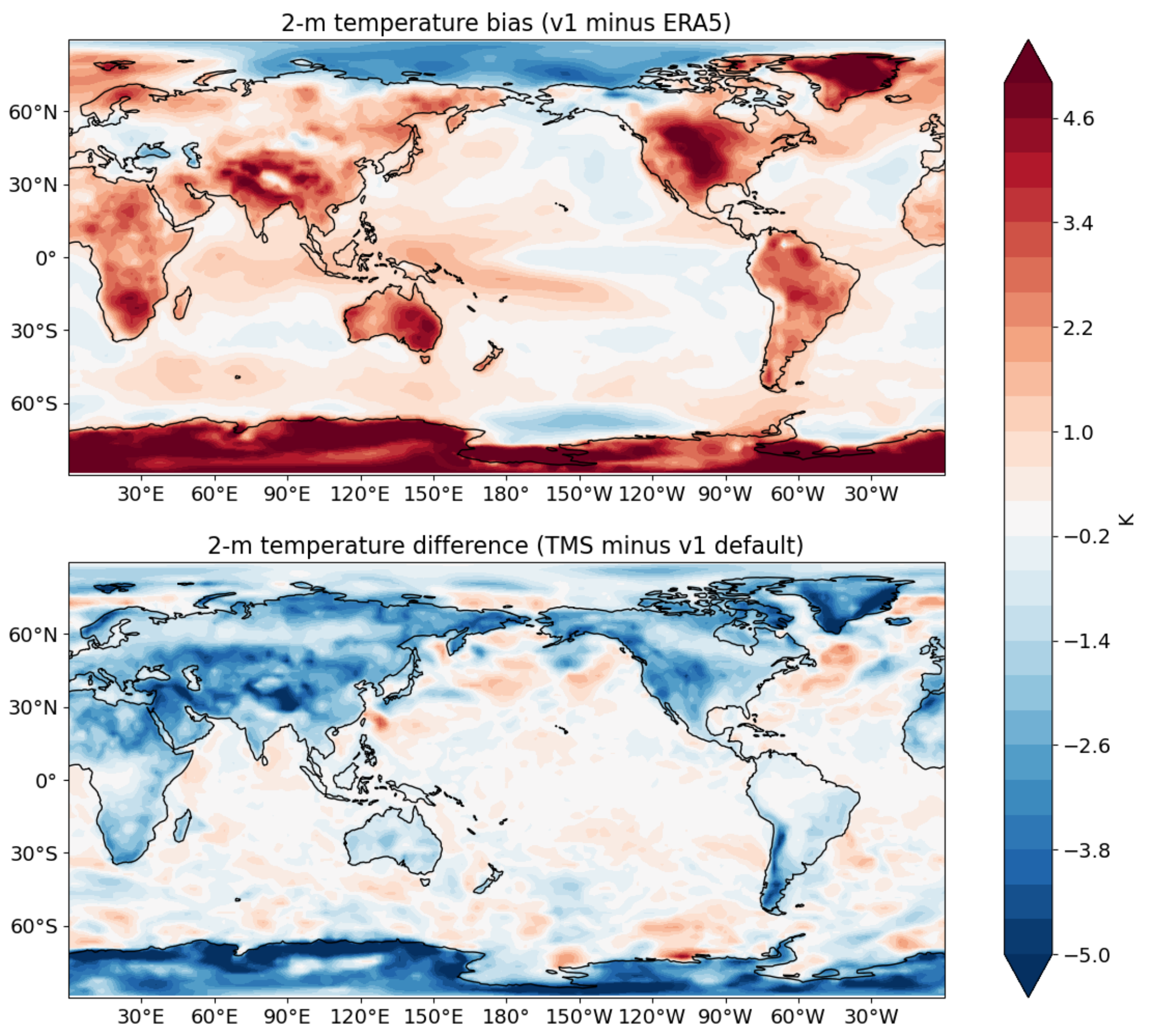
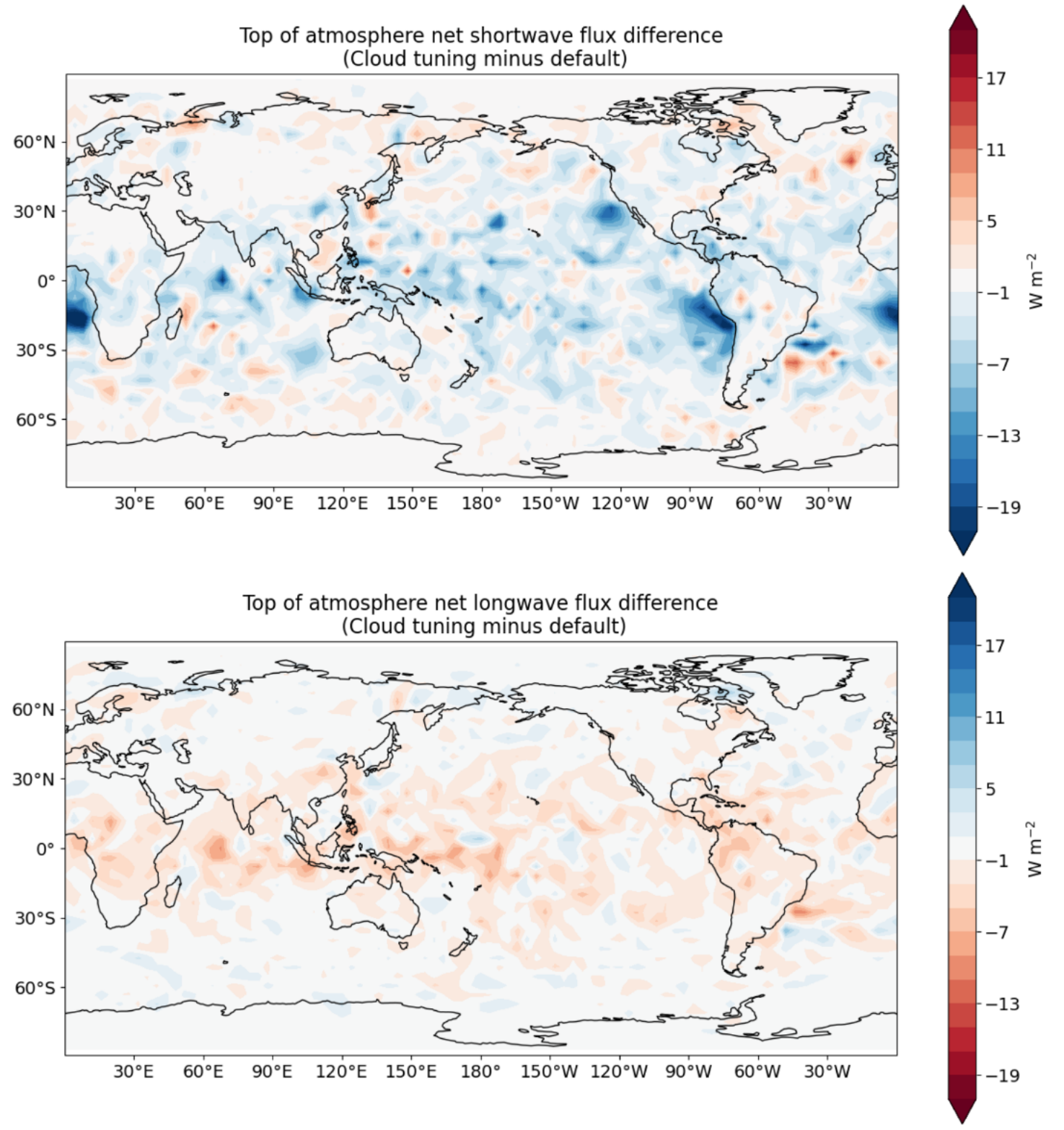

2-m temperature bias (v1 minus ERA5)
60°N
30°N
0°
30°S
60°S
30°E
60°E
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
2-m temperature difference (TMS minus v1 default)
60°N
30°N
0°
30°S
60°S
30°E
60°E
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
4.6
3.4
2.2
1.0
−0.2
−1.4
−2.6
−3.8
−5.0
K

Figure35.


Top of atmosphere net shortwave flux difference
(Cloud tuning minus default)
60°N
30°N
0°
30°S
60°S
30°E
60°E
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
17
11
5
−1
−7
−13
−19
W m⁻²
Top of atmosphere net longwave flux difference
(Cloud tuning minus default)
60°N
30°N
0°
30°S
60°S
30°E
60°E
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
17
11
5
−1
−7
−13
−19
W m⁻²