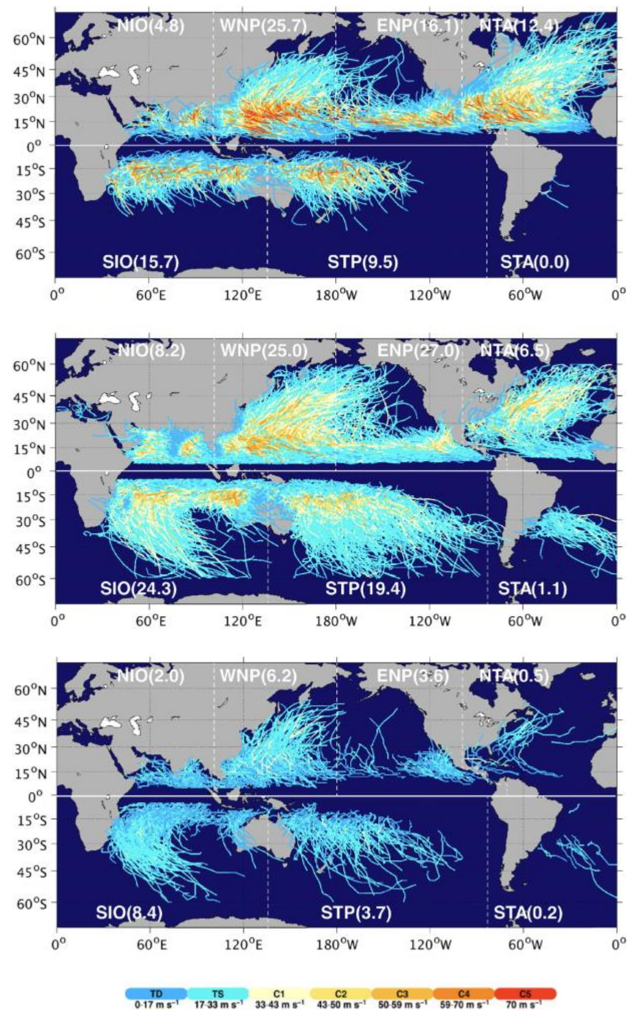
CESM

Community Earth System Model



Proposal for NWSC Resources

PI: Gokhan Danabasoglu, CESM Chief Scientist

Period of Performance: 01 November 2022 – 31 October 2024

Total Request: 826 M Cheyenne (Equivalent) Core-Hours

Table of Contents

Cover image: Observed (top) and simulated Tropical Cyclone (TC) tracks from high-resolution (middle) and low-resolution CESM simulations during the 1950-2018 period (from Chang et al. 2020). Colors indicate storm intensity categories. Vertical dashed lines separate the seven TC basins: North Indian Ocean (NIO), Western North Pacific (WNP), Eastern North Pacific (ENP), North Tropical Atlantic (NTA), South Indian Ocean (SIO), South Tropical Pacific (STP), and South Tropical Atlantic (STA). The numbers in parenthesis indicate the annual-mean TC values. The global annual-means are 82.4, 111.5, and 24.6 for the observations, high-resolution, and low-resolution simulations, respectively.

Acknowledgments

The CESM project is supported primarily by the National Science Foundation (NSF). This material is based upon work supported by the National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the NSF under Cooperative Agreement 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), are provided by the Computational and Information Systems Laboratory (CISL) at NCAR. We thank all the scientists, software engineers, and administrators who contribute to the CESM project.

Introduction

The Community Earth System Model (CESM) is a collaborative, community modeling effort between researchers at the National Center for Atmospheric Research (NCAR), universities, and other national and international research institutions. CESM is used for multiple purposes, including investigations of past and current climate, projections of future climate change, and subseasonal-to-decadal Earth system predictions. To keep CESM at the leading edge of climate and Earth system modeling and related science requires continuous developments, improvements, testing, and subsequent applications of the model to various problems. These efforts have been almost exclusively facilitated through access to the NCAR Computational and Information Systems Laboratory (CISL) computational resources at the NCAR – Wyoming Supercomputer Center (NWSC).

The CESM and its predecessor, the Community Climate System Model (CCSM), have been at the forefront of both national and international efforts to understand and predict the behavior of Earth's climate. Output from numerous simulations using CCSM and CESM are routinely used in thousands of peer-reviewed studies to better understand the processes and mechanisms responsible for climate variability and change. An overwhelming majority of these studies make use of CCSM's and CESM's contributions to the Coupled Model Intercomparison Project phase 3 (CMIP3), phase 5 (CMIP5), and, most recently, phase 6 (CMIP6). CESM2 continues to be among the most realistic climate models in the world based on some metrics that compare the model outputs against present-day observationally-based datasets with all CESM2 configurations submitted to CMIP6 being in the top ten (Fasullo 2020). As a further testament to this point, the three primary manuscripts, Gent et al. (2011), Hurrell et al. (2013), and Danabasoglu et al. (2020) that introduced and described CCSM4, CESM1, and CESM2 have been cited > 2650, > 1850, and > 550 times, respectively, since their publications. Significant NWSC-supported efforts such as the CESM1 Large Ensemble (CESM1-LENS; Kay et al. 2015), its sibling CESM1 Decadal Prediction Large Ensemble (CESM1-DPLE; Yeager et al. 2018), the CESM1 Last Millennium Ensemble (CESM1-LME; Otto-Bliesner et al. 2016), the CESM Stratospheric Aerosol Geoengineering Large Ensemble (GLENS; Tilmes et al. 2018), and Seasonal to Multi-Year Large Ensemble (SMYLE; Yeager et al. 2022) have been key in advancing our understanding of the climate system and its variability and predictability, supplementing CESM's contributions to the CMIPs with community-driven science efforts. CESM source code and simulation output are made freely available to the broad scientific community. Additionally, CCSM and CESM simulations provide the quantitative modeling basis for both national and international assessments of climate, including those of the Intergovernmental Panel on Climate Change (IPCC) and the U.S. Global Change Research Program (USGCRP). Furthermore, CESM provides the National Science Foundation (NSF), its primary sponsor, and the national and international research communities with a well-supported Earth system modeling framework.

CESM's participation in CMIP6, including various Model Intercomparison Projects (MIPs), is its most visible and the broadest community project. The datasets from CESM2 CMIP6 simulations are available on the Earth System Grid Federation (ESGF; https://esgfnode.llnl.gov/search/cmip6). 50 manuscripts describing and analyzing these experiments in detail are collected in the AGU CESM2 Virtual Special Issue (https://agupubs.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)1942-2466.CESM2; also see Danabasoglu and Lamarque 2021). These simulations reach a vast national and international group of researchers who rely on NCAR, and CESM in particular, to perform these simulations. Among the modeling groups which contribute to CMIP6, CESM is very unique in its community involvement and the level of transparency with which it approaches model development. Moreover, many national and international researchers make use of CESM in their research proposals, e.g., submitted to the NSF, counting on the CESM's contributions to the CMIPs. Additionally, NCAR and CESM assist the community with the analysis of CMIP simulations by providing a CMIP Analysis Platform (https://www2.cisl.ucar.edu/resources/cmip-analysis-platform) as well as by making many diagnostics tools available for use of the community such as the Climate Variability and Diagnostics Package (CVDP; available at https://www.cesm.ucar.edu/working_groups/CVC/cvdp/; also see Phillips et al. 2014). High level of scrutiny and analysis of the CESM simulations, in turn, feedback to CESM, promoting further model development as well as enhancing collaborations.

Regarding the CESM source code, we have not had official CESM releases since September 2020 in favor of an open-source community model available via GitHub. This has been a game changer in terms of the community being able to access and contribute to the latest development code bases, providing transparency in several key areas, including issues and discussions related to code contributions being fully public. In addition, personnel costs associated with controlled access have been eliminated.

This proposal is accompanied by a CESM NWSC Use and Accomplishments Report that covers the 01 November 2021 – 31 October 2022 period – as reported on 09 September 2022, corresponding to the second year of our current 2-year allocation cycle. As is customary, the Report presents use and impressive accomplishment summaries for each Working Group (WG) as well as the six Community Projects. Consistent with our Proposal for NWSC Resources submitted in September 2020, these activities fall under two primary overarching priorities: efforts towards creation of the next generation CESM version, CESM3; and performing and providing simulations with CESM2 to enable and support community-driven science to answer fundamental Earth system related questions.

The priorities and goals outlined in the present proposal originate directly from the community of scientists who participate in the CESM project through the CESM WGs and the CESM Scientific Steering Committee (SSC, whose membership primarily consists of scientists from universities and government laboratories). Specifically, to prepare this proposal, each WG reached out to their constituents – beginning in May 2022, with widely distributed emails to discuss model development goals towards the next generation model version, CESM3, and production simulations (mostly with the CESM2 versions) required to address high priority scientific questions, especially those that benefit from analysis and interpretation by the broader community. During drafting of the WG plans, there were substantial discussions and planning both within and across the WGs. Many of the WGs had related discussions during their WG meetings at the 2022 Annual CESM Workshop. The resulting drafts were further reviewed, revised, refined, and prioritized by the CESM Chief Scientist in frequent communication with the WG cochairs and the SSC – when needed – with the goal of producing a coherent and coordinated plan for the use of the NWSC resources over the upcoming period of performance. These reviewed plans and resource requests of the individual WGs and community projects, which are included in the Supplementary Material, then served as the draft source material for further review by the SSC and WG cochairs and liaisons. The responsibility of the SSC in this process was to review the general plans and priorities in its July 2022 meetings and to provide input and guidance on the main development and production activities and the required computing and storage resources. We note that the CESM SSC has been continuously providing such oversight and guidance on CESM objectives and priorities, including development activities towards CESM3. As such, proposed efforts in this proposal, e.g., focus on model development towards CESM3 and evaluations of coupled simulations with the Community Atmosphere Model (CAM) Spectral Element (SE) dynamical core, directly reflect the inputs from the CESM SSC. A similar process was followed during the last several CESM NWSC proposals and, we believe, it has resulted in a coherent overview of the testing, development, and application needs of the broad CESM project.

The guidance from CISL indicated that the first year of our 2-year request would be a hybrid allocation on both Cheyenne and Derecho, taking into account some delays in Derecho's availability. As such, we were informed that we would have 210 M core-hours on Cheyenne during 01 November 2022 – 30 September 2023 and 181 M core-hours on Derecho during 01 June 2023 – 31 October 2023 for a total of 391 M core-hours during the first year of our allocation. We note that there is a 4-month overlap of Cheyenne and Derecho resources during this time. We were also informed that the second year of our allocation would be 435 M core-hours on Derecho for the 01 November 2023 – 31 October 2024 period. Therefore, this request is for 826 M core-hours combined (391 M + 435 M core-hours; see Model Performance section for details of timings on Derecho). In addition, CESM has been allocated about 37 000 hours per month on the GPU portion of Derecho for a total of 0.634 M GPU hours for the duration of this request. We provide some preliminary efforts for partial use of this GPU resource in the GPU Use section below. We certainly understand the need to transition to Derecho during this allocation cycle. Although we do not anticipate major disruptions in our simulations during this transition period, the CESM Software Engineering Working Group (SEWG) will dedicate resources (in collaboration with CISL) to ensure a smooth transition.

Strategy for Addressing Further Delays in Derecho's Arrival

Each WG has prioritized their proposed simulations as high (A), medium (B), and low (C) priority in their Table of Experiments provided in the Supplementary Material. Medium and Low priority simulations[1] request 160.6 M and 23.6 M core-hours, respectively, for a

[1] Despite being listed as medium or low priority simulations, some of these are at the frontiers of science and they will remain so even if we have to delay them by some period of time.

total of 184.2 M core-hours over two years. Based on the allocation plan outlined above, a delay of one month in the availability of Derecho during the Cheyenne – Derecho overlap period (June-September 2023) will reduce our allocation by 36 M core-hours (per month). A delay of one month beginning in October 2023 will reduce our allocation by 17 M core-hours (per month), assuming that we will still have access to Cheyenne at our current monthly rate of ~19 M core-hours. Thus, our initial strategy to accommodate any delays in the arrival of Derecho will be not performing medium and low priority simulations. This path could be followed until about January 2024 by which time accumulated reduction in the core-hours will be roughly 184 M core-hours – equal to the total of medium and low priority simulations. If Derecho is delayed later than January 2024, we will need to reprioritize all our remaining simulations at that time. Simulations *Towards CESM3* will have the highest priority in our considerations. Depending on the general use of Cheyenne, we may also consider running some of our simulations in the Economy queue.

Overarching Priorities and Goals

The primary niche for the CESM Project is to continue to develop and maintain a world-leading, state-of-the-art Earth system model that is fully documented, supported, and freely available, in collaboration with and in support of the broader climate community. This broad niche builds upon CESM's culture of community involvement as well as a tradition of successes as articulated in the Introduction, making CESM the community nexus for Earth system research, both nationally and internationally. Furthermore, given the world-class reputation of CESM, there is trust in the community that each model version is carefully developed as well as exhaustively and carefully assessed to produce the best possible Earth system simulation capability. Within this context, CESM serves a very wide audience in support of both curiosity-driven research and larger collaborative projects across various groups. It has a broad range of applications that include research from paleoclimate through space weather with configurations that vary in complexity as a flexible community modeling framework. Through its unique capabilities, CESM can bridge weather, climate, chemistry, and geospace communities.

To keep CESM at the forefront of Earth system modeling and to enable its use in newly emerging scientific challenges and applications require continual evolution of the coupled system and its components. Such efforts include numerical and parameterization developments; addition of new capabilities and of previously missing physics, chemistry, etc.; improved analysis and diagnostics packages; and new configurations with simpler physics and coarser resolutions at one end, and with higher horizontal and vertical resolutions as well as with regional grid refinement at the other end, just to name a few. Thus, our broad overarching priorities during the 2-year period of performance of this allocation request remain similar to the ones from the previous request:

- Complete the creation of the next generation CESM version, CESM3; and

- Perform and provide simulations with CESM2 as well as with preliminary versions of CESM3 to enable and support community-driven science to answer fundamental Earth system related questions.

The path towards CESM3 had already started prior to the previous allocation cycle and it had continued during the past 2 years. Substantial progress has been made toward developing and incorporating the new component models, i.e., the Modular Ocean Model version 6 (MOM6), the Marine Biogeochemistry Library (MARBL), the Community Atmosphere Model version 7 (CAM7), the Community Ice Sheet Model version 3 (CISM3), the Community Terrestrial Systems Model (CTSM), a river runoff model, and the Community Sea-Ice Model version 6 (CICE6). Significant work has also been completed with increased vertical resolution – both 58 (L58) and 93 (L93) vertical levels – and vertical extent in the atmospheric model to deliver a unified troposphere-middle atmosphere model. Nevertheless, much work remains to be completed in developing, implementing, and tuning parameterizations and new features in each component model as summarized below with further details given in the Supplementary Material.

In their December 2021 meetings, CESM SSC directed us to start preliminary fully-coupled CESM3 simulations to get an initial evaluation of the coupled model with its new components as they are being developed. As detailed in the *Towards CESM3* section below as well as in the Accomplishments Report, this work has already started and will be a major activity during this allocation period as well. So, while resources are being requested by each component model WG in their development allocation for component-model-only simulations, resources for fully-coupled simulations bringing together developmental versions of the component models are being requested under the *Towards CESM3* umbrella.

In parallel with the development efforts towards CESM3, WGs plan to perform simulations to advance Earth system science using various CESM2 configurations. The scope of these experiments is rather broad, reflecting the diversity of the CESM community. To provide a few examples, they include: various simulations with regionally-refined grid configurations to perform, for example, air quality studies over specific regions; studies on detection and attribution of interannual to decadal fluctuations in atmospheric CO_2; simulations of the deployment of carbon dioxide removal (CDR) technologies, considering their effectiveness at sequestering CO_2; targeted experiments to probe the role of particular phenomena in the climate system and the impact of enhanced horizontal resolution; Earth system prediction simulations to advance understanding of the role of land state on subseasonal prediction skill, the role of transbasin interactions in seasonal-to-interannual prediction skill, the drivers of recently observed climate variability, and the role of volcanic forcing in altering near-term climate forecasts; simulations looking into the interactions between land ice and the rest of the climate system at many spatial and temporal scales; investigations of Earth system processes of climate-ice sheet interactions and the climate responses to land-cover and volcanic forcings; simulations to improve understanding of the coupled influence of changing atmospheric circulation, factors affecting Arctic amplification, and influences on changing Arctic precipitation; simulations to understand the factors driving recent climate variations in the Antarctic; experiments targeting atmospheric river conditions, sources of moisture in the polar regions, the physics of polar clouds, and sea ice parameter sensitivity; and various geoengineering experiments.

This request does not include computational resources to perform global high-resolution (coupled) CESM2 or CESM3 simulations with the exception of many regionally-refined grid (in the atmosphere) CESM2 configurations. Simulations that use the International Laboratory for Earth System Prediction (iHESP) high-resolution CESM1.3 model version (0.25° in the atmosphere and land, and 0.1° in the ocean and sea-ice components) will continue to be run at the Texas Advanced Computer Center (TACC). These simulations include high-resolution Earth system predictions. High-resolution CESM1.3 simulations with isotopes will also be performed on an Advanced Science and Discovery (ASD) allocation on Derecho. Taking into consideration the input from the WGs, in its 08 July 2020 Meeting, the CESM SSC unanimously recommended that CESM should not develop a new CESM2 high-resolution version, but instead should focus on CESM3 developments with its limited resources and consider creating a high-resolution CESM3 version in due time. Along these lines, we request resources here to support development and some applications of both eddy-permitting and eddy-resolving configurations of MOM6. In addition, on the atmospheric side, there is a new community (university)-led effort to incorporate the Department of Energy (DOE) Energy Exascale Earth System Model (E3SM)'s version of the CAM SE non-hydrostatic (SE-NH) version which has been developed for both Fortran and C++ software – both available on GitHub, into CESM. This effort is expected to request separate NWSC resources if funded. It is expected that CAM-SE-NH will be able to take full advantage of the GPU portion of Derecho, providing new opportunities for CESM and the community.

We finally note that CESM is involved with several machine learning-related projects that use CESM and led by university and / or private sector collaborators, e.g., the Center for Learning the Earth with Artificial Intelligence and Physics (LEAP) and the Multiscale Machine Learning In Coupled Earth System Modeling (M2LInES), just to name a couple of them. As such, our request for machine learning-related efforts is small in this proposal as our larger efforts are covered under our collaborative projects.

Working Group Research Objectives and Requests

Request for 292.7 M core-hours in Year 1 and 321.6 M core-hours in Year 2

In this section, we very briefly summarize the overall research goals and objectives specific to each WG. In addition, for each WG, we provide the requested computing allocation, split between development (D) and production (P) as well as between Years 1 and 2 – identified as D-Y1, D-Y2, P-Y1, and P-Y2, respectively. All the core-hour requests listed in Table 1 are in millions (M) of Cheyenne (equivalent) core-hours. Further details of each WG's request are available in the Supplementary Material.

As indicated in the Model Performance section below, SEWG provided performance estimates for a wide range of CESM configurations, including for CESM3, to the WGs for their use in their requests. SEWG also provided performance estimates for several configurations running on Derecho (again see the Model Performance section). However, the choice for which estimate to use was left to the WGs to balance between simulation throughput and cost. Computational costs are also significantly impacted by the I/O requirements of a particular simulation. Therefore, while a majority of the requests use the same costs for a given model configuration, different estimates can also be found for the same model configuration across WG requests. In addition, the estimates for new model versions have relied on simple scaling arguments, considering changes, for example, in the number of vertical levels, tracers, or horizontal resolution. For some configurations, particularly for regionally-refined variable resolution simulations, how estimates are determined are provided in the relevant Supplementary Material sections.

	Year 1		Year 2		Totals		
	Dev	Prod	Dev	Prod	Year 1	Year 2	Y1 + Y2
Working Group							
AMWG	35.7	2.9	43.2	2.3	38.6	45.5	84.1
BGCWG	11.7	6.4	8.3	11.9	18.1	20.2	38.3
ChCWG	12.6	14.4	10.4	19.6	27.0	30.0	57.0
CVCWG		27.5		29.1	27.5	29.1	56.6
ESPWG	5.0	17.4	17.2	7.4	22.4	24.6	47.0
LIWG	7.0	13.0	8.9	13.1	20.0	22.0	42.0
LMWG	12.0	15.0	12.5	17.5	27.0	30.0	57.0
OMWG	17.4	9.5	19.6	10.8	26.9	30.4	57.3
PaleoCWG	9.8	18.6	17.6	11.0	28.4	28.6	57.0
PCWG	2.9	18.8	7.8	12.9	21.7	20.7	42.4
SEWG	8.0		10.0		8.0	10.0	18.0
WAWG	14.9	12.2	8.2	22.3	27.1	30.5	57.6
Total WGs	**137.0**	**155.7**	**163.7**	**157.9**	**292.7**	**321.6**	**614.3**
Community Projects							
C1. Towards CESM3	39.0	39.0	44.6	44.6	78.0	89.2	167.2

C2. Dual Polar		15.4		19.3	15.4	19.3	34.7
C3. Emerging Science		5.0		5.0	5.0	5.0	10.0
Total comm	**39.0**	**59.4**	**44.6**	**68.9**	**98.4**	**113.5**	**211.9**
Total WG+comm	**176.0**	**215.1**	**208.3**	**226.8**	**391.1**	**435.1**	**826.2**
Target					391.0	435.0	826.0

Table 1. Complete list of core-hour allocation requests for WG development and production projects and community projects. The entries are in millions (M) of Cheyenne (equivalent) core-hours. The requests for Year 1 and Year 2 as well as the total for both years are provided. Blank entries indicate 0.

Atmosphere Model Working Group (AMWG)

D-Y1: **35.7 M;** D-Y2: **43.2 M;** P-Y1: **2.9 M;** P-Y2: **2.3 M; Total: 84.1 M**

AMWG, a partnership of NCAR and university scientists, utilizes computational resources primarily for the development of the CESM CAM and associated capabilities. This encompasses the advancement of both the representation of the unresolved physical processes in parameterization schemes and the dynamical core processes, including tracer transport. It also covers sensitivity experiments aimed at understanding the many interactions among the represented physical and dynamical processes across climate regimes and multiple timescales.

With this request, AMWG development activity will focus on these broad areas. Specifically, we will: 1) Deliver a unified atmospheric model (CAM7) that incorporates a well-resolved stratosphere as well as improved resolution of boundary layer turbulent processes towards CESM3; 2) Examine substantially new physics and their impact on simulated climate; and 3) Examine atmospheric physics and dynamics in regionally-refined model configurations for domains of special interest in global climate, including a 6-km South America refined configuration, a High-Mountain Asia (HMA)/Monsoon domain encompassing the Tibetan Plateau and Indian subcontinent, and a refined western US configuration. Production activity will be confined to extending ongoing HMA work using existing CAM6 physics in a 32-level, low-top configuration.

Biogeochemistry Working Group (BGCWG)

D-Y1: **11.7 M;** D-Y2: **8.3 M;** P-Y1: **6.4 M;** P-Y2: **11.9 M; Total: 38.3M**

There are several new and significant initiatives related to biogeochemistry planned over the next 2 years. In particular, model development and data assimilation activities will aim to enable CESM to better support questions related to the state of the global carbon cycle, including detection and attribution of interannual to decadal fluctuations in atmospheric CO_2. Related to this will be efforts to explicitly simulate the deployment of CDR technologies, considering their effectiveness at sequestering CO_2 in the context of a dynamic climate system. Furthermore, there is an ambitious effort to implement a prognostic model of fish in MOM6, the Fisheries Size and Functional Type (FEISTY) model, leveraging the representation of lower trophic level dynamics provided by MARBL. FEISTY will enable addressing important actionable research questions by providing an end-to-end framework for explicit predictions of variations in fisheries-relevant biomass distributions from climate forcing and other stressors.

Chemistry Climate Working Group (ChCWG)

D-Y1: **12.6 M;** D-Y2: **10.4 M;** P-Y1: **14.4 M;** P-Y2: **19.6 M; Total: 57.0 M**

The goal of ChCWG is to continue the development of the representation of chemistry and aerosols in CESM and to further our understanding of the interactions between gas-phase chemistry, aerosols, and climate, for past, present, and future climates. The scientific motivation for the proposed development experiments is to a) improve the existing model configurations, including updates in comprehensive chemistry in both troposphere and stratosphere, and the development of more physical aerosols models, and b) to explore different vertical and horizontal resolutions, to develop chemistry-climate configurations that are best suited for the specific scientific studies. The identified configurations will be used for production experiments, to explore present-day and future air quality over multiple scales, to produce community model simulations that can also serve as spin-up experiments for regionally refined experiments, and to understand the role of climate change on composition in both troposphere and lowermost stratosphere.

The development and production simulations requested will improve the representation and chemical prediction of tropospheric and stratospheric composition and air quality, and allow us to participate in multi-model intercomparison activities. We are working on the development of the workhorse model with comprehensive troposphere and stratosphere chemistry. We also aim to develop an advanced chemistry model configuration for air quality studies that will go beyond the standard workhorse model, to further improve the physical representation in the model and allow targeted science experiments. In addition, various regionally refined model grids have been developed to focus on air-quality studies over specific regions. These configurations serve as an initial setup of the Multi-Scale Infrastructure for Chemistry Modeling (MUSICA) experiment suite, which will be supported and further developed with the scientific community.

Climate Variability and Change Working Group (CVCWG)

D-Y1: **0.0 M**; D-Y2: **0.0 M**, P-Y1: **27.5 M;** P-Y2: **29.1 M; Total: 56.6 M**

The goals of CVCWG are (1) to understand and quantify contributions of natural and anthropogenically-forced patterns of climate variability and change in the 20th and 21st centuries and beyond by means of simulations with CESM and its component models and (2) to assess and understand the representation of climate variability and change within CESM. With these model simulations, researchers will be able to investigate mechanisms of climate variability and change, as well as detect and attribute past climate change, and project and predict future climate changes. The CVCWG simulations are motivated by broad community interest and are widely used by national and international research communities. The highest priority for the CVCWG simulations is given to runs that directly benefit the CESM community. CVCWG is a central element of the DOE - NCAR Cooperative Agreement and provides an interface with national (e.g., U.S. National Assessment) and international (e.g., IPCC) climate-change assessment activities. Based on input from the broader community as well as research goals within NCAR, the following priorities have been determined.

The next two years will be a period of model development toward CESM3. As such, WG will continue to focus its efforts on producing datasets and understanding of the behavior of CESM2 or intermediate model versions between CESM2 and CESM3, with a plan to move to production runs with CESM3 in the next proposal cycle (2024-2026). The proposed experiments aim at augmenting the existing suite of CESM2 simulations alongside targeted experiments to probe the role of particular phenomena in the climate system and the impact of enhanced horizontal resolution. The proposed experiments can broadly be categorized into three research areas: (1) exploration of the role of forcings in the climate system; (2) exploration of model behavior at high resolution (horizontal and/or vertical); and (3) exploration of the role of particular phenomena or regions in shaping climate variability and change.

Earth System Prediction Working Group (ESPWG)

D-Y1: **5.0 M**; D-Y2: **17.2 M**, P-Y1: **17.4 M;** P-Y2: **7.4 M; Total: 47.0 M**

ESPWG focuses on understanding the processes responsible for predictability on scales from subseasonal to decadal and filling a very much needed niche of providing the community a framework for performing and analyzing initialized predictions of the Earth system as well as serving as a community nexus for Earth system prediction research.

Experiments proposed by ESPWG focus on facilitating community efforts by expanding large-ensemble production datasets using CESM2 while also exploring the many uncertainties associated with prediction system design, such as, land, ocean, atmosphere initialization, drift, bias, etc. Key science questions that the group will address are (a) how various ocean basins contribute to seasonal to decadal prediction skill, (b) how changes in model physics and system design affect prediction skill, (c) whether online bias correction is a viable method for improving predictions given model systematic errors, and (d) whether initialized predictions can offer insight into the origins of recently observed climate variability. The group’s development efforts will focus on testing the preliminary configuration of CESM3 in the context of subseasonal, seasonal to interannual, and decadal prediction. This will inform how changes in model physics and significantly altered model components affect prediction skill, including prediction skill of key modes of variability. Another focus of development activities will be to leverage the recently completed CESM2-SMYLE dataset as a control for testing several seasonal to decadal system design sensitivities, including initialization and bias adjustment methodologies. The proposed production simulations will build off of existing CESM2 prediction systems to advance understanding of: the role of land state on subseasonal prediction skill, the role of transbasin interactions in seasonal-to-interannual prediction skill, the drivers of recently observed climate variability, the role of volcanic forcing in altering near-term climate forecasts, and the differences between CESM1 and CESM2 decadal prediction performance. We will continue running the prediction systems based on CESM2 in real-time to enable studies of recent extreme events and contribute to various multi-model projects.

Land Ice Working Group (LIWG)

D-Y1: **7.0 M;** D-Y2: **8.9 M;** P-Y1: **13.0 M;** P-Y2: **13.1 M; Total: 42.0 M**

The main objectives of LIWG during this allocation period are to develop and release a new version of the Community Ice Sheet Model, CISM v3.0, with improved physics and dynamics; to complete and validate new CESM capabilities for ice sheet–ocean coupling and multiple ice sheets; to simulate the Greenland Ice Sheet (GrIS), Antarctic Ice Sheet, and paleo ice sheets in past, present, and future climate experiments; and to carry out high-resolution simulations of mountain glaciers and the overlying atmosphere. A goal is to simulate land ice–climate interactions on time scales of decades to millennia, and spatial scales from ~1 km to thousands of km.

CISM development priorities include: testing and validation of new basal friction, subglacial hydrology, and calving schemes; new models and parameterizations to compute sub-ice-shelf melt rates and couple ice sheets to solid-Earth and sea-level models; various numerical and algorithmic improvements; and development of a comprehensive land-ice diagnostic package. The development allocation will also support validation of new capabilities for multiple ice sheets and ice sheet–ocean coupling; standalone and coupled simulations of ice-sheet behavior during the Last Glacial Maximum (LGM) and Holocene; and high-resolution mountain glacier simulations. The production allocation will support simulations of the interactions between land ice and the rest of the climate system at many spatial and temporal scales. The request includes two long paleoclimate simulations with interactive ice sheets: one for the Northern Hemisphere ice sheets during the last glacial inception (120–110 ka) and one for GrIS during the Holocene (9 ka to 1850). We will also explore the effects of variable-resolution grids in studies of Antarctic precipitation and the climate of HMA. We will contribute to several model intercomparison projects. As CESM3 and CISM3 become available, we will couple them and compare the results to earlier runs using CESM2 and CISM2. Some of these experiments are joint projects with AMWG, LMWG, and PaleoCWG.

Land Model Working Group (LMWG)

D-Y1: **12.0 M;** D-Y2: **12.5 M;** P-Y1: **15.0 M;** P-Y2: **17.5 M; Total: 57.0M**

The goals of LMWG are to advance the state-of-the-art in modeling Earth's land surface, its ecosystems, watersheds, and socioeconomic drivers of global environmental change, and to provide a comprehensive understanding of the interactions among physical, chemical, biological, and socioeconomic processes by which people and ecosystems affect, adapt to, and mitigate global environmental change. Land biogeophysical and biogeochemical processes are intimately linked and therefore it is not possible to separate land biogeophysics development from land biogeochemistry development. As in our previous requests, land biogeochemistry model development has been included in the LMWG request.

LMWG continues to pursue an ambitious program of model development. In particular, there are several large development projects that are underway, including a Parameter Perturbation Ensemble (PPE), application of a representative hillslope hydrology model, and building towards a fully supported configuration of the CTSM with the Functionally Assembled Terrestrial Ecosystem Simulator (FATES). These projects will continue into this allocation request along with other development projects on development of a new river model that includes representation of water management, a microbial explicit soil biogeochemistry parameterization, a multi-layer canopy scheme, and continued agriculture model development. Resources for simulations investigating the role of land processes, parameter uncertainty, and their role in climate variability and change in support of LMWG research are also requested.

Ocean Model Working Group (OMWG)

D-Y1: **17.4 M;** D-Y2: **19.6 M;** P-Y1: **9.5 M;** P-Y2: **10.8 M; Total: 57.3 M**

The primary goals of OMWG are to advance the capability and fidelity of the CESM ocean component in support of specific science objectives of the broader CESM community and to conduct curiosity driven research using CESM to advance our understanding of ocean processes and the role of the ocean in the Earth system.

OMWG continues our efforts to provide the ocean component model for CESM3 based on MOM6. During the current allocation cycle, substantial progress was made toward developing a "workhorse" configuration with extensive testing in both forced ocean-sea-ice and fully-coupled simulations. With most of the technical development and baseline tuning of CESM-MOM6 completed, the effort of OMWG for the coming allocation cycle will focus first on assuring the fidelity and robustness of CESM-MOM6 leading to the CESM3 release and second on advancing the capabilities and process representation of the ocean component to address the expanding scope of CESM science objectives. Targeted areas include:

A more complete and flexible resolution hierarchy in both global and regional configurations. While we will continue to assess and improve the workhorse configuration, the need for higher resolution versions of the ocean component to address CESM science objectives is clear. We are requesting resources to support significant development and application of an eddy-permitting global configuration with a nominal resolution of 20-25 km as well as limited testing of an eddy-resolving configuration with a nominal resolution of 8-10 km. The eddy-permitting resolution supports a richer spectrum of climate variability arising from intrinsic ocean variability as well as providing climate information on scales required for applicable science. To further address the latter, several regional configurations that allow downscaling to finer scales in coastal environments are proposed.

Improvements in diapycnal and neutral physics. Work will continue towards replacing the legacy K-profile parameterization of surface boundary layer mixing with a new formulation called the Flux-profile parameterization. This scheme requires tighter coupling to the wind-wave model (WaveWatch 3) and affords a more accurate representation of the mixing of both momentum and scalars. This effort is augmented by research on improving the representation of shear-driven mixing in the stratified ocean interior. Both efforts leverage experiments with a large eddy simulation model (LES) of near surface ocean turbulence for guidance. Improvements in shear-driven mixing are especially targeted at tropical ocean dynamics with consequent impacts on ENSO, and ocean heat and CO_2 uptake. A suite of new parameterizations of mesoscale and sub-mesoscale eddy mixing being developed within the NOAA-NSF Climate Process Team will be evaluated for use in CESM. These include both deterministic and stochastic variants as well as scale-awareness thus fitting well with the efforts towards a flexible resolution hierarchy described above.

Integration with data assimilation system and initialized prediction. An increasing portion of CESM science involves initialized prediction. CESM prediction efforts to date use forced ocean simulations as the basis for initializing hindcasts and forecasts. We will support ESPWG by providing ocean simulations conducted under the Ocean Model Intercomparison Project (OMIP) forcing protocols, using the evolving CESM-MOM6 model. The assessment of the performance of the model in forecast mode will provide valuable feedback on choices on process representation and model configuration leading to the CESM3 release. We are requesting resources to begin conducting both ocean and coupled system data assimilation using the Data Assimilation Research Testbed (DART) framework with MOM6, culminating in a decade-long ocean reanalysis in the later part of the allocation period. In addition to its intrinsic value in ocean climate research, this reanalysis will provide an alternative to the forced-ocean simulations for initializing CESM hindcasts. There is also an intersection with the process representation development described above using stochastic parameterizations as a means of accomplishing ensemble generation and inflation.

Interactions between the ocean and ice shelves. Advances in ocean model algorithms available in MOM6 along with coupling of CISM using the CESM/NUOPC infrastructure open the opportunity for a physically-based representation of the interaction between ice shelves and ocean dynamics in CESM. Initial testing of this capability in idealized simulations conducted under the Marine Ice Sheet-Ocean Model Intercomparison Project (MISOMIP) is ongoing. The support requested in this proposal will be used to begin coupling these components in more realistic configurations first in a regional pan-Antarctic domain, then in the global configuration with one or more of our eddy-permitting (~25 km) resolutions.

While we build toward the CESM3 release with MOM6, research that leverages high value experiments and ensembles with POP2 will continue. These include contributions to international model inter-comparison efforts where POP provides our baseline contribution. Notably, much of this work will be conducted with the eddy-resolving version of CESM-POP2.

Paleoclimate Working Group (PaleoCWG)

D-Y1: **9.8 M;** D-Y2: **17.6 M;** P-Y1: **18.6 M;** P-Y2: **11.0 M; Total: 57.0 M**

PaleoCWG is a consortium of scientists engaged in modeling to understand Earth's past climates and provide a long-term perspective on climate system feedbacks and processes that underlie the transient nature of climate change. PaleoCWG conducts simulations for specific past climate states, designed to explore the relationships between climate forcings, such as variations in atmospheric greenhouse gases, the presence of large continental-scale ice sheets, solar variability and volcanic activity, and feedbacks and processes that control Equilibrium Climate Sensitivity (ECS) and climate responses on a range of temporal and spatial scales. Assessing model simulations for these out-of-sample climate states against paleoclimate reconstructions based on geological and geochemical records is an important element of the PaleoCWG activities.

The overall development and production goals of PaleoCWG are to (1) provide out-of-sample assessment of CESM's physical parameterizations and climate sensitivity through the simulation of key past climates and (2) deliver to the community an expanded set of CESM capabilities and simulations suitable for studying a wide range of fundamental questions of climate and paleoclimate science. PaleoCWG simulations use new configurations and new capabilities of CESM, and, when combined with paleoclimate data, provide unique opportunities to characterize and quantify climate forcing and feedbacks during past climate changes, and provide unique constraints on the future climate projections. PaleoCWG will develop (1) CESM3 and/or CESM2 simulations of the past glacial and high-CO_2 climates and (2) a low-cost, low-resolution CESM3 for climate and paleoclimate applications. PaleoCWG will explore new capabilities in paleoclimate simulations including (1) fast spin-up techniques for spinning up ocean tracers and (2) PPEs for quantifying the model parameter uncertainties. PaleoCWG will investigate Earth system processes of (1) climate-ice sheet interactions and (2) weather and water extremes during the Holocene, and (3) the climate responses to land-cover and volcanic forcings.

Polar Climate Working Group (PCWG)

D-Y1: **2.9 M;** D-Y2: **7.8 M;** P-Y1: **18.8 M;** P-Y2: **12.9 M; Total: 42.4 M**

PCWG is a consortium of scientists who are interested in understanding and modeling Arctic and Antarctic climate and its relationship to global climate. To enable polar science within PCWG and the CESM project as a whole, we request computing resources for both polar-specific CESM parameterization development and for polar-specific CESM scientific research. We anticipate frontier results will emerge from the diversity of activities we propose, and that these results will provide new understanding of polar climate processes and interactions between polar regions and the globe.

Our overall development objective is to ensure that CESM has state-of-the-art abilities to simulate polar climate. Towards this goal, improvements to numerous aspects of the sea ice model within CESM will be developed and tested. This includes simulations to analyze, test, and tune new capabilities available within CICE6. These will improve the simulation of the impacts of snow on sea ice, landfast ice conditions, the floe-size distribution, and wave-sea ice interactions. We expect most of these improvements to be incorporated within CESM3. Additional development work will target new sea ice model capabilities. This includes a machine-learned parameterization to incorporate the thermodynamic influence of high-spatial snow heterogeneity and new parameterizations that will better simulate aspects of the sea ice albedo evolution, including improved representations of the influence of snow and ponding.

The overarching PCWG production goal is to enable important and topical polar science research using CESM. In terms of Arctic research, proposed simulations will improve understanding of the coupled influence of changing atmospheric circulation, factors affecting Arctic amplification, and influences on changing Arctic precipitation. In the Antarctic, simulations are proposed to investigate factors driving recent climate variations. Of relevance to both poles, simulations will target atmospheric river conditions, sources of moisture in the polar regions, the physics of polar clouds, and sea ice parameter sensitivity. Combined, the requested simulations will enable a better understanding of coupled dynamics in the polar regions and interactions between polar regions and the global system.

Software Engineering Working Group (SEWG)

D-Y1: **8.0 M;** D-Y2: **10.0M;** P-Y1: **0.0 M;** P-Y2: **0.0 M; Total: 18.0 M**

The role of SEWG is to coordinate the computational development of the CESM model components, oversee the evolving design of the CESM as new model components, new model grids, and new model physics are added to the system, and at the same time engineer the model system to obtain optimal throughput and efficiency. This continues to be particularly challenging as the number of model configurations, model complexity, and model resolutions are rapidly increasing. Numerous tests are carried out for each new CESM revision on all production platforms to ensure required functionality (such as exact restart capability), correct results (such as bit-for-bit reproducibility where it is expected), tracking of memory and performance metrics (to determine if these have changed relative to the previous revision), and other key production requirements (such as optimizing performance of new revisions, especially where new component science has been introduced). This testing also ensures the robustness of model infrastructure development, such as improvements to the model driver, mediator, tools, and scripts. Computing time is requested to carry out these important functions throughout the various CESM versions that will be generated, both in the CESM2 series and on the path towards CESM3.

Whole Atmosphere Working Group (WAWG)

D-Y1: **14.9 M;** D-Y2: **8.2 M;** P-Y1: **12.2 M;** P-Y2: **22.3 M; Total: 57.6 M**

WAWG promotes the development of a unified sun-to-earth modeling framework with the Whole Atmosphere Community Climate Model (WACCM) and WACCM-X, a version of WACCM extended to the thermosphere-ionosphere (~500 km). Production runs are also proposed to advance science as well as to promote collaborations with the community.

The development proposal is focused primarily on the development, tuning, and evaluation of the upcoming versions of WACCM and WACCM-X, which will adopt the CAM7 physics and vertical grid spacing in the troposphere-stratosphere as well as the SE dynamical core. Development resources are also requested to further develop a thermosphere gravity wave parameterization for WACCM-X as well as geomagnetic storm simulations in WACCM-X coupled to the GAMERA (Grid Agnostic MHD for Extended Research Applications) magnetosphere model. Further development of the extension of the Model for Prediction Across Scales (MPAS) dynamical core into the thermosphere will also be pursued. The production proposal includes evaluation of the CAM7 L93 workhorse model versus WACCM7 L135 for both baseline scenario and geoengineering experiments. Additional simulations are focused on the downward impact of stratospheric sudden warmings, stratospheric aerosol injection, and the impact of circulation changes on climate sensitivity, air quality, and chemistry. Production simulations using WACCM-X target the impact of geoengineering on the space environment, thermosphere gravity waves, and climate projection simulations.

Community Projects

Request for 98.4 M core-hours in Year 1 and 113.5 M core-hours in Year 2

We request allocations for three projects referred to as: *Toward CESM3*, *Dual Polar*, and *Emerging Science*. Among these, the largest allocation (167.2 M core-hours) is requested for *Towards CESM3* for required and anticipated simulations to create a new CESM3 version. *Dual Polar* project, requesting 34.7 M core-hours, was the only project proposed for consideration following a call for such *Community Projects*. The project was reviewed and discusses at the CESM Cochairs Meeting on 16 August 2022. The present version incorporates the suggestions received during that discussion. The project labeled as *Emerging Science* is a modest request that can be allocated to a few relevant emerging science topics that need to be addressed in a timely manner during the course of this allocation cycle.

In the following, summaries of these projects are provided. Further details of each request are available in the Supplementary Material.

C1. Towards CESM3 (167.2 M core-hours)

Requested by: The CESM Chief Scientist

We will devote about 20% of our total allocation, corresponding to 167.2 M core-hours over 2 years, to fully-coupled simulations *Towards CESM3*. The high priority effort here is to perform fully-coupled CESM3 simulations to evaluate the coupled solutions as the new component models are being developed, noting that resources for the development and evaluation of the component-only simulations are requested in individual WG allocations. Ideally, we would like to also perform the so-called Diagnostic, Evaluation, and Characterization of Klima (DECK) experiments that include a long pre-industrial control simulation; an abrupt quadrupling of CO_2 concentration simulation; and a 1% per year CO_2 concentration increase simulation, as well as a LGM simulation to help diagnose the ECS of the new CESM3. The progress on these latter simulations will significantly depend on our progress with the basic development and evaluation of the new CESM3.

C2. Fully-Coupled Simulations with a Dual-Polar Variable Resolution Grid (34.7 M core-hours)

Requested by: AMWG, PCWG, and LIWG

This project aims to simulate the evolution of both the Greenland and Antarctic ice sheets through two-way coupling with CISM, taking advantage of recent enhancements in the CESM infrastructure that allow for multiple ice sheet domains. Furthermore, it will use a new dual-polar atmospheric grid, which has high-resolution regional refinement over both poles. Our previous simulations with the ARCTIC grid (1/4° refinement over the Arctic – the northern half of the new dual polar grid), fully coupled to POP2 and CISM, have shown improved representation of the GrIS and it's simulated mass balance, through more accurately resolving the precipitation and melting processes via high-resolution grids. Such regionally-refined simulations also provide an opportunity to resolve key processes that determine the behavior of atmospheric systems, i.e., atmospheric rivers, that bring more heat to the region and impact the frequency and severity of melting events. The proposed simulations will use CAM6 with its SE dynamical core with 58 vertical levels and the MOM6 ocean model. Resources are requested to perform a 200-year PI control, then branching a 1%CO_2 experiment for 250 years, holding constant the concentrations at 4xCO_2, which occurs around year 140. We also propose running a corresponding 1° configuration in parallel to quantify the impacts of horizontal resolution on the retreat of the GrIS.

C3. Emerging Science (10.0 M core-hours)

Requested by: The CESM Chief Scientist

During the course of our allocation cycles, several science topics naturally emerge that can be addressed using various CESM configurations. The topics in this category are usually time critical. They include, for example, climate impacts of reductions in emissions due to COVID-19 and Australian wildfires of 2019/2020; deciphering impacts of details of how biomass burning emissions are constructed in CMIP6; and investigating scaling properties of CAM at ultra high-resolutions (at km scale). Various sets of ensemble simulations may need additional members or sets of ensembles to further understand the impacts of forcings or a particular parameterization. Clearly, such topics and simulations are of broad community interest, with both national and international assessment implications. Therefore, it behooves CESM to respond to such emerging science topics in a timely fashion where a modest amount of computer time can be allocated for these purposes.

GPU Use

The use of GPUs in CESM is limited; at the resolutions we typically target they aren't beneficial for the high integration rates desired, and analysis workloads are more typically dominated by memory size, not computation or speed, with a few novel exceptions. Nevertheless, we anticipate two efforts that can use the GPU portion of Derecho: i) cross-attractor transformations and ii) hyperparameter exploration (see Supplementary Material for further details). However, these types of projects do not run constantly, and thus aren't a regular, reliable use of GPU hours.

Looking further into the future, we can expect *some* development use of GPUs, especially with the 2023 Casper refresh, which should give us GPUs from multiple vendors, and thus enable a development / testing cycle that ensures our model remains portable and not tied solely to Derecho's planned hardware. Another potential project could be the use of GPUs to massively improve real-time compression of model output data on dedicated I/O tasks. Some initial work on the asynchronous I/O capabilities needed for this is already underway, but both of these large development tasks are conditional on available funding and competing priorities, and are not easily measurable at this moment. Additionally, much of the benefit of I/O on GPUs requires the use of emerging "GPUDirect" storage capabilities – but these are incompatible with Glade while Cheyenne remains in use, rendering it impossible for year 1 of the allocation.

Finally, we anticipate some preliminary GPU use associated with our CAM SE-NH efforts discussed in the Overarching Priorities and Goals section as well as with some CAM physics developments towards the end of year 2 allocation. We can also expect some small amount of GPU use from standard, infrequent development tasks and benchmarking. All of these efforts are in collaboration with CISL colleagues.

In summary, expected GPU use is extraordinarily hard to quantify at this time. It seems very unlikely that CESM will come close to the 186 K hours offered in year 1 without a significant uptick in Machine Learning use or additional funding for large development projects. The use of the year 2 allocation of 448 K hours is less clear, and dependent again on funding for developing new capabilities, the availability of multi-vendor GPU systems which enable new developments in a portable manner, and the status of GPU-enabled tools for analysis that offer advantages over the more flexible CPU-based workflows.

Data Management and Archival Needs

As part of this CESM NWSC request, each WG and Community Project generated estimates of their long-term (order 5 years) storage needs on the Campaign Storage (CS) associated with each proposed development and production experiment set (listed in the Supplementary Material in detail). These estimates are summarized in Table 2 in Terabytes (TB). *It is important to stress that these estimates are only for our anticipated output volume to be stored on CS for long-term archival and that the actual data volume produced by the proposed simulations will be much higher*. *However, many of the datasets, particularly from the development simulations, will not have any archival value.* As in computational time requests, storage estimates for a given configuration may differ depending on the particular output requirements of a simulation. Note that the estimates are for lossless compression and take into account improvements in data compression from the use of the netCDF-4 standard. As shown in Table 2, we anticipate archiving about 8.3 PB of data from the simulations performed at NWSC.

	Year 1		Year 2		Totals		
	Dev	Prod	Dev	Prod	Year 1	Year 2	Y1 + Y2
Working Group							
AMWG	86.0	14.0	101.0	11.0	100.0	112.0	212.0
BGCWG		213.9		354.0	213.9	354.0	567.9
ChCWG	80.0	126.0	80.0	155.5	206.0	235.5	441.5
CVCWG		464.7		413.4	464.7	413.4	878.1
ESPWG	51.0	214.9	156.3	54.1	265.9	210.4	476.3
LIWG	21.0	25.0	37.0	89.0	46.0	126.0	172.0
LMWG		106.8		69.0	106.8	69.0	175.8
OMWG	231.2	104.9	210.7	164.1	336.1	374.8	710.9
PaleoCWG	64.5	271.4	133.0	84.0	335.9	217.0	552.9
PCWG	34.0	206.5	74.0	112.5	240.5	186.5	427.0
SEWG							0.0
WAWG	163.3	53.8	144.4	206.8	217.1	351.2	568.3
Total WG	**731.0**	**1801.9**	**936.4**	**1713.4**	**2532.9**	**2649.8**	**5182.7**

Community Projects							
C1. Towards CESM3	500.0	500.0	1000.0	1000.0	1000.0	2000.0	3000.0
C2. Dual Polar		30.0		45.0	30.0	45.0	75.0
C3. Emerging Science		20.0		20.0	20.0	20.0	40.0
Total comm	**500.0**	**550.0**	**1000.0**	**1065.0**	**1050.0**	**2065.0**	**3115.0**
Total WG+comm	**1231.0**	**2351.9**	**1936.4**	**2778.4**	**3582.9**	**4714.8**	**8297.7**
Workspace Buffer					617.1	1285.2	1902.3
Total					**4200.0**	**6000.0**	**10200.0**

Table 2. Complete list of storage estimates for WG development and production projects and community projects for archiving on CS. The entries are in TB. The estimates for Year 1 and Year 2 as well as the total for both years are provided.

a. A new CESM Data Management and Data Distribution Plan

Our data management, storage, and distribution plans follow the newly updated *Data Management and Data Distribution Plan for the CESM Project* (January 2022) available at https://www.cesm.ucar.edu/management/docs/cesm-data-management-plan-2021.pdf. The Plan documents the procedures for the storage and distribution of data produced by the CESM Project on the NCAR computers with the CESM allocation. These procedures reflect the approaches, standards, and conventions that coordinate the production, post-processing, distribution, and storage of simulation data agreed upon by the CESM SSC and WGs that comprise the CESM Project with input from the relevant CISL personnel. The overall goal of the Plan is to provide the best possible access and ease-of-use of CESM Project data to a broad and diverse user community within the constraints of available resources, following FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

In accordance with the NSF data policy and the NCAR mission, the CESM Project is committed to the timely availability of results from CESM model runs for publication and sharing of the scientific data generated by CESM Project research activities. Open access to CESM data products is essential to the Project. Analysis and interpretation by the broader community promotes scientific discovery, and it leads to new insights into model behavior that feedback into model development efforts. At the same time, however, the efforts of CESM developers and the designers of the scientific simulations performed with the model under the auspices of the CESM WGs need to be recognized by providing them a reasonable amount of time for first access to the simulations. The Plan defines the guidelines to meet these objectives. These procedures thus apply to all CESM data created under the auspices of the CESM WGs, using the NCAR computational and storage resources allocated to the CESM Project. The Plan also provides suggested pathways for the storage of data from CESM simulations performed at NWSC, but not as part of the CESM allocation and from CESM simulations run at other supercomputer centers. Regardless of how they are produced, all CESM datasets are encouraged and expected to follow the naming, distribution, etc. guidelines provided in this Plan.

b. Data archiving

Out of many PBs of total data volume expected to be generated from all proposed experiments, we anticipate archiving about 8.3 PB on CS. Our archival strategy will be as follows:

Development: Output data will be primarily stored on the requested Glade partition (see below). As shown in Table 2, many of the WG development simulations will require no long-term storage. However, we anticipate storing 1.67 PB from development simulations, representing ~16% of our total storage request. Our experience with the CESM2 development cycle taught us that multi-year development work requires keeping a modest set of reference solutions. We similarly expect archiving some of our CESM3 development simulations. Also based on experience, we expect that 1.67 PB will account for approximately 20% of the development output, meaning that 80% of the generated output has no archival value. In addition, the data will likely be removed 24 months after creation, unless retention is requested from the relevant WG co-chairs or the CESM Chief Scientist for the CESM3 development simulations. *One-off* development experiments will be removed more quickly at the PIs and / or WG co-chairs' discretion.

Production and Community Projects: The majority of our long-term storage request with 6.63 PB is for our production and community projects which include various CESM3 development and control simulations as well. Initially, output data will be stored on Glade scratch space (model *history* data), and the majority will be converted to *timeseries* format, using lossless compression. These timeseries data will then be archived to CS. After a period of four years, we will evaluate if certain datasets should be kept for a longer period of time. These data will then be gradually cut back to 50% of their initial volume over a period of two additional years, based on usage and anticipated demand. This data level will be maintained for two more years. Afterward, each WG and / or the CESM Chief Scientist will determine what data are to be removed and at what rate, as the archived data is gradually reduced to an acceptable level, as determined by data archiving costs at the time.

Archive Management: A majority of the experiments will make full use of the existing CESM Experiment Database (see https://csegweb.cgd.ucar.edu/expdb2.0/cgi-bin/expList.cgi). This database contains details about the run configuration and establishes provenance. The database application runs an automated monthly email reminder script triggered off dates stored in the database fields; as such, it will be used to remind all affiliated users with the experiment, including scientific leads and software engineers, to prune their data from CS according to the CESM Data Management and Data Distribution Plan.

c. Data distribution

Development: In general, output data will be made available only to the WG members that are directly involved with the experiments. For WG members that do not have access to CSL resources, these data will be made available via the Climate Data Gateway (CDG), formerly known as the Earth System Grid (ESG), or the NCAR Data Sharing Service, as appropriate.

Production and Community Projects: Output data will be made available according to the guidelines established by the CESM Data Management and Data Distribution Plan (see section (a) above). Initially, access is restricted to the WG members directly involved with the experiments. After a period of no more than 12 months following creation, these data will be made available to the community via the CDG.

d. Data analysis and visualization request

The simulations produced under development and production resources will require considerable analysis and visualization. For these needs, we request access to the Casper data analysis and visualization (DAV) cluster. This will require standard interactive access to these clusters for the WG members. Currently this includes about 200 participating scientists but is subject to change with changing WG members and involvement. We request an allocation of 500,000 hours.

e. Glade project file space (total request: 3.6 PB)

To minimize the usage of CS for storing development results, we request 3.6 PB of CESM Glade project space, a 50% increase from our current 2.4 PB allocation. This estimate is based on our previous experience and our anticipated use. This will also enable efficient access to highly utilized CESM simulation outputs and forcing data used in coupled integrations. The requested increase is vital to expedite analysis of simulations, especially those with many ensemble members and during the *Towards CESM3* phase. This space is collectively managed by the CESM WGs.

f. Campaign Storage space (total request: 8.3 PB + 1.9 PB)

We request 10.2 PB of CS space as summarized in Table 2 and also as discussed above in section (b) for our simulations performed at NWSC. 5.2 PB of this request is for WGs' development and production simulations. We request 3.0 PB for simulations *Towards CESM3* based on our needs during the previous CESM2 development / creation process. We also request 1.9 PB as workspace buffer, again based on our previous experience during the CESM2 development experiments. This buffer space is crucial as we evaluate multiple sets of simulations. We note that we are also performing CESM simulations on non-NWSC supercomputers. Because these simulations are of significant interest to the broader CESM community – following, for example, the extremely successful CESM1 and CESM2 Large Ensemble efforts, we continuously evaluate the feasibility of bringing datasets back so that we can serve them to the NSF-supported CESM user community. This requested buffer space could partially accommodate such datasets as well.

g. Lossy data compression

As indicated above, our data storage estimates use lossless compression which results in an approximately 50% reduction in the data volume for the fully-coupled system. To reduce our data footprint even further, we continue to pursue lossy compression techniques in collaboration with Allison Baker (CISL), Dorit Hammerling (Colorado School of Mines, CSM), and Alex Pinard (CSM). Work-to-date by these collaborators has been very promising, but focused only on the atmospheric model. The newer efforts are now considering all model components, investigating feasibility of various levels (aggressiveness) of lossy compression. Relevant WGs will participate in the evaluation and acceptability of lossy compressed datasets. If a level of lossy compression is deemed scientifically acceptable, it will be adopted for use in future CESM data storage, perhaps during the lifetime of this request.

Model Performance

Many of the production simulations during this allocation period will be done with versions of the model that have very similar performance characteristics to the CESM2 versions used for CMIP6 simulations. Substantial work went into optimizing the performance and throughput of those model versions. Primarily by optimizing the layout of each component and also by improving vectorization and reducing communication costs, a very respectable rate of approximately 30 model years per wall-clock day on 4320 cores on Cheyenne for the 1° workhorse configuration was achieved. In addition, SEWG provided performance estimates for a wide range of configurations with various combinations of simulation years per day and number of cores to the WGs for their use in their requests. These estimates are available at https://csegweb.cgd.ucar.edu/timing/cgi-bin/timings.cgi.

Some new development runs, as well as some production runs later in the two-year allocation period, will use new versions of components or entirely new components that have very different performance characteristics from CESM2 as we move towards CESM3. There are still large uncertainties and many unknowns about the exact configurations of these components and about their performance characteristics. Therefore, we are not attempting to provide detailed model performance estimates for these future model versions. Instead, SEWG has provided performance estimates for a limited set of configurations[2] in the above table, based on our ongoing CESM3 simulations. For other configurations, the computational requests in this proposal account for a general increase

[2] Ocean (MOM6) – sea-ice (CICE5/6) coupled, G-compset; atmosphere-only (CAM6-dev) with L58 (F-compset; and fully-coupled with MOM6, CICE5/6, CAM6-dev and L58 (B-compset).

in model cost based on educated estimates of what is expected, considering, e.g., changes in resolution, atmospheric dynamical core, and new parameterizations in all component models. Once these new configurations have been decided upon, SEWG will work to optimize their performance as they have done in the past for all CESM versions. Finally, when Derecho becomes available, SEWG will dedicate resources to optimize CESM, if needed.

Derecho performance estimates

To provide the WGs with reasonable estimates of performance on Derecho, SEWG ran several comparative cases on both Cheyenne and TACC's Lonestar6, the latter of which uses the same processors and a similar high-speed network as Derecho, making it a very good proxy. An analysis was then made looking at the CAM6, CICE6, and MOM6 components individually, and a fully-coupled pre-industrial control run. The results are summarized in the table below as estimates of percentages of core-hour needs on Derecho compared to Cheyenne for nominal 1° configurations.

Derecho Relative Core-Hour Needs	
CAM6	97.8%
CICE6	99.7%
MOM6	72.2%
PI control	93.4%

For example, we calculate a per-core improvement on Derecho of just 2.2% for CAM6. In other words, a CAM6 run on Derecho, on a similar core count to Cheyenne, is likely to require approximately 98% of the core-hours as it would on Cheyenne. While, broadly speaking, Derecho will have in aggregate many more core-hours available for use, the *per-core* performance is not massively different than on Cheyenne. MOM6 is a notable exception, which would run very efficiently on the new system, which in turn brings the overall cost of coupled runs slightly down.

We can expect small gains from tuning and load balancing once we have greater access to Derecho itself or another proxy, but for this proposal, the above numbers are from real-world tests that mirror expected usage. **Based on this analysis and for simplicity, WGs were instructed to use the same core-hours for a given configuration to be run on Derecho as those on Cheyenne.**

Summary

Earth system models are the most powerful tools for meeting the intellectual challenge of understanding the climate and the Earth system: They are the only scientific tool capable of integrating numerous physical, chemical, and biological processes that determine past, present, and future climates. Furthermore, they are critical for testing hypotheses, confirming understanding, and for making predictions of use to society and policy makers for actionable science. Among a multitude of such Earth system models, CESM is rather unique because its development and applications are determined through strong partnerships with scientists from universities, national laboratories, and other research organizations. As such, CESM enables the investigations of new scientific problems through collaborations with a community continuously increasing in size, empowering many new partnerships.

The present request has two, forward-looking overarching priorities. They are: i) continued efforts towards creation of the next generation CESM version, CESM3; and performing and providing simulations with CESM2 and its components as well as with preliminary versions of CESM3 and its newer components to enable and support community-driven science to answer fundamental Earth system related questions. On the path towards CESM3, all component models have plans for major model development efforts that include numerous advancements in all the component models. During the lifetime of this request, we expect that a vast majority of the proposed developments will be completed and near-final CESM3 versions will be available by early 2024.

To achieve these priorities and goals, the proposal is subdivided into two sub-requests, totaling 826 M Cheyenne (equivalent) core-hours. The first is for WGs' development and production projects for 292.7 M and 321.6 M core-hours, respectively, for a total of 614.3 M core-hours. Of the remaining 211.9 M core-hours, 167.2 M core-hours will be reserved for performing simulations *Towards CESM3*. While the early part of our allocation is on Cheyenne, the latter part will be on Derecho with a 4-month overlap period. As such, we are aware of the need to transition to Derecho during our allocation cycle. Although we do not anticipate major disruptions in our simulations during this transition period because Derecho is very similar to Cheyenne in many aspects, SEWG will dedicate resources (in collaboration with CISL) to ensure a smooth transition. The tables of individual WG development and production simulations presented in the Supplementary Material include a prioritization of proposed simulations in the event that availability of Derecho is further delayed – our proposed plan to accommodate such a delay is included in the Introduction. We note that simulations *Towards CESM3* will have the highest priority in our considerations.

We request 10.2 PB on CS for long-term archival of our datasets after lossless compression. About 4.67 PB of this request is to support our model development efforts, including 3.0 PB request for *Towards CESM3*. Our experience with the CESM2 development cycle taught us that multi-year development work requires storage of a modest set of reference solutions. We similarly expect archiving some of our CESM3 development simulations. For our storage, we will follow the newly-updated CESM Data Management and Distribution Plan. To minimize / optimize storage on CS, we request 3.6 PB on Glade as project space. This request is again based on our previous experience and anticipates the analysis needs of simulations, especially with many ensemble members and CESM3 evaluation simulations.

The NWSC computational resources remain indispensable to carry out our ambitious scientific and model development agenda as we move towards CESM3. The goals and priorities outlined in this proposal emanate directly from the community of scientists who participate in the CESM project through the WGs and the SSC. They were developed, refined, and prioritized after a several month process with the goal of producing a coherent and coordinated plan for the use of the NWSC resources over the upcoming period of performance. The proposed experiments will fill important development and science needs and contribute to the overall project priorities.

References

Chang, P., and Coauthors, 2020: An unprecedented set of high-resolution Earth system simulations for understanding multiscale interactions in climate variability and change. *J. Adv. Mod. Earth Sys.*, **12**, e2020MS002298, doi: 10.1029/2020MS002298.

Danabasoglu, G., and Coauthors, 2020: Community Earth System Model version 2 (CESM2). *J. Adv. Model. Earth Sys.*, **12**, e2019MS001916, doi: 10.1029/2019MS001916.

Danabasoglu, G., and J.-F. Lamarque, 2021: Building a better model to view Earth's interacting processes. *EOS*, **102**, 19-21, doi: 10.1029/2021EO155818.

Fasullo, J. T., 2020: Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1). *Geosci. Mod. Dev.*, **13**, 3627-3642, doi: 10.5194/gmd-13-3627-2020.

Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973-4991, doi: 10.1175/2011JCLI4083.1.

Hurrell, J., and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94,** 1339-1360, doi: 10.1175/BAMS-D-12-00121.1.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333-1349, doi: 10.1175/BAMS-D-13-00255.1.

Otto-Bliesner, B. L., and Coauthors, 2016: Climate variability and change since 850 CE: An ensemble approach with the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **97**, 735-754, doi: 10.1175/BAMS-D-14-00233.1.

Phillips, A. S., C. Deser, and J. Fasullo, 2014: A new tool for evaluating modes of variability in climate models. *EOS*, **95**, 453-455, doi: 10.1002/2014EO490002.

Tilmes, S., and Coauthors, 2018: CESM1(WACCM) Stratospheric Aerosol Geoengineering Large Ensemble (GLENS) project. *Bull. Amer. Meteor. Soc.*, **99**, 2361-2371, doi: 10.1175/BAMS-D-17-0267.1.

Yeager, S. G., and Coauthors, 2018: Predicting near-term changes in the Earth System: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **99**, 1867-1886, doi: 10.1175/BAMS-D-17-0098.1.

Yeager, S. G., and Coauthors, 2022: The Seasonal-to-MultiYear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2. *Geosci. Mod. Dev.*, **15**, 6451-6493, doi: 10.5194/gmd-15-6451-2022.