CMIP6 without the interpolation: Grid-native analysis with Pangeo in the cloud

Julius Busecke[1] and Ryan Abernathey[2]

[1]Princeton University
[2]Columbia University

November 22, 2022

Abstract

The Pangeo project recently introduced large parts of the CMIP6 data archive into the cloud. This enables, for the first time, centralized, reproducible science of state-of-the-art climate simulations without the need to own large storage or a supercomputer as a user. The data itself however, still presents significant challenges for analysis, one of which is applying operations across many models. Two of the major hurdles are different naming/metadata between modeling centers and complex grid layouts, particularly for the ocean components of climate models. Common workflows in the past often included interpolating/remapping desired variables and creating new files, creating organizational burden, and increasing storage requirements. We will demonstrate two Pangeo tools which enable seamless calculation of common operations like vector calculus operators (grad, curl, div) and weighted averages/integrals across a wide range of CMIP6 models directly on the data stored in the cloud. cmip6_preprocessing provides numerous tools to unify naming conventions and parse grid information and metrics (like cell area). This information is used by xgcm to enable finite volume analysis on the native model grids. The combination of both tools facilitates fast analysis while ensuring a reproducible and accurate workflow.

CMIP6 without the interpolation: Grid-native analysis with Pangeo in the cloud

Julius Busecke, Ryan Abernathey

The Pangeo project recently introduced large parts of the CMIP6 data archive into the cloud. This enables, for the first time, centralized, reproducible science of state-of-the-art climate simulations without the need to own large storage or a supercomputer as a user. The data itself however, still presents significant challenges for analysis, one of which is applying operations across many models. Two of the major hurdles are different naming/metadata between modeling centers and complex grid layouts, particularly for the ocean components of climate models. Common workflows in the past often included interpolating/remapping desired variables and creating new files, creating organizational burden, and increasing storage requirements. We will demonstrate two Pangeo tools which enable seamless calculation of common operations like vector calculus operators (grad, curl, div) and weighted averages/integrals across a wide range of CMIP6 models directly on the data stored in the cloud. cmip6_preprocessing provides numerous tools to unify naming conventions and parse grid information and metrics (like cell area). This information is used by xgcm to enable finite volume analysis on the native model grids. The combination of both tools facilitates fast analysis while ensuring a reproducible and accurate workflow.