1   *Supplementary material*

3   **Objective Evaluation of Earth System Models: PCMDI Metrics Package**

4   **(PMP) version 3**

5

6   Jiwoo Lee[1], Peter J. Gleckler[1], Min-Seop Ahn[2,3], Ana Ordonez[1], Paul A. Ullrich[1,4], Kenneth R.

7   Sperber[1,a], Karl E. Taylor[1], Yann Y. Planton[5,6], Eric Guilyardi[7,8], Paul Durack[1], Celine Bonfils[1],

8   Mark D. Zelinka[1], Li-Wei Chao[1], Bo Dong[1], Charles Doutriaux[1], Chengzhu Zhang[1], Tom Vo[1],

9   Jason Boutte[1], Michael F. Wehner[9], Angeline G. Pendergrass[10,11], Daehyun Kim[12], Zeyu Xue[13],

10   Andrew T. Wittenberg[14], and John Krasting[14]

11

12   [1] Lawrence Livermore National Laboratory, Livermore, California, USA

13   [2] NASA Goddard Space Flight Center, Greenbelt, MD, USA

14   [3] ESSIC, University of Maryland, College Park, MD, USA

15   [4] University of California, Davis, Davis, California, USA

16   [5] NOAA Pacific Marine Environmental Laboratory, Seattle, Washington, USA

17   [6] Monash University, Clayton, Australia

18   [7] LOCEAN-IPSL, CNRS-IRD-MNHN-Sorbonne Université, Paris, France

19   [8] National Centre for Atmospheric Science-Climate, University of Reading, Reading, UK

20   [9] Lawrence Berkeley National Laboratory, Berkeley, California, USA

21   [10] Department of Earth and Atmospheric Science, Cornell University, Ithaca, New York, USA

22   [11] National Center for Atmospheric Research, Boulder, Colorado, USA

23   [12] School of Earth and Environmental Sciences, Seoul National University, Seoul, South Korea

24   [13] Pacific Northwest National Laboratory, Richland, WA, USA

25   [14] NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

26   [a] Retired

27

28   Submitted to Geoscientific Model Development (GMD)

29   (November 2023)

30

31   *Corresponding to:* Jiwoo Lee (lee1043@llnl.gov)

32   7000 East Ave, Livermore, California 94550, USA

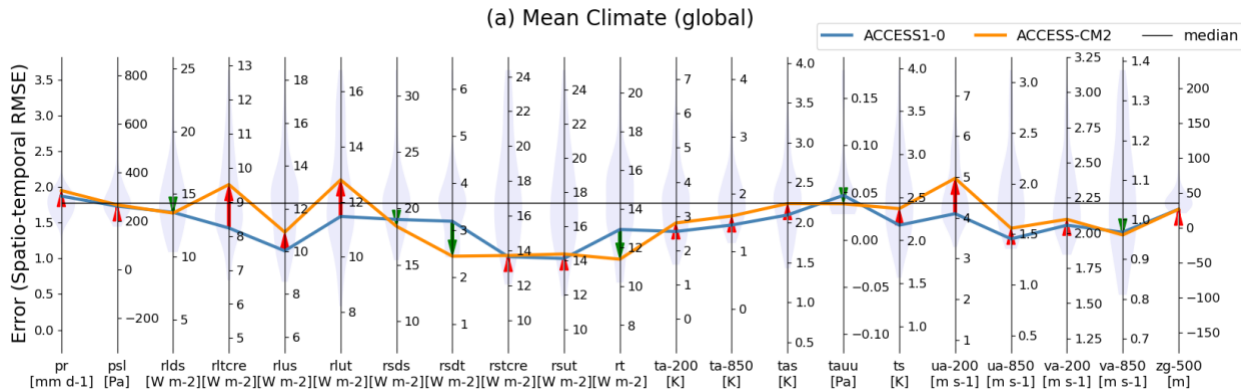
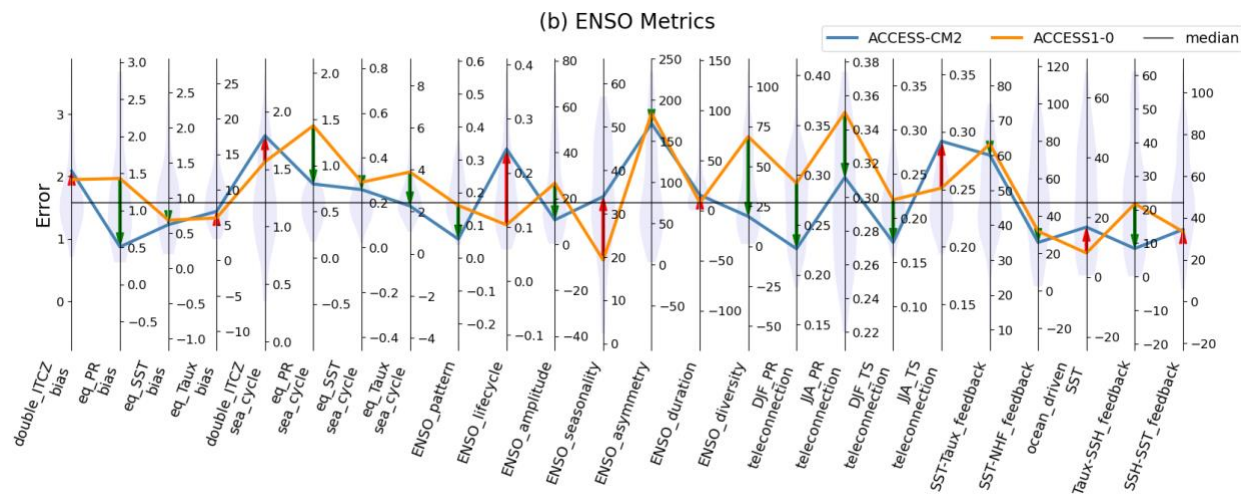
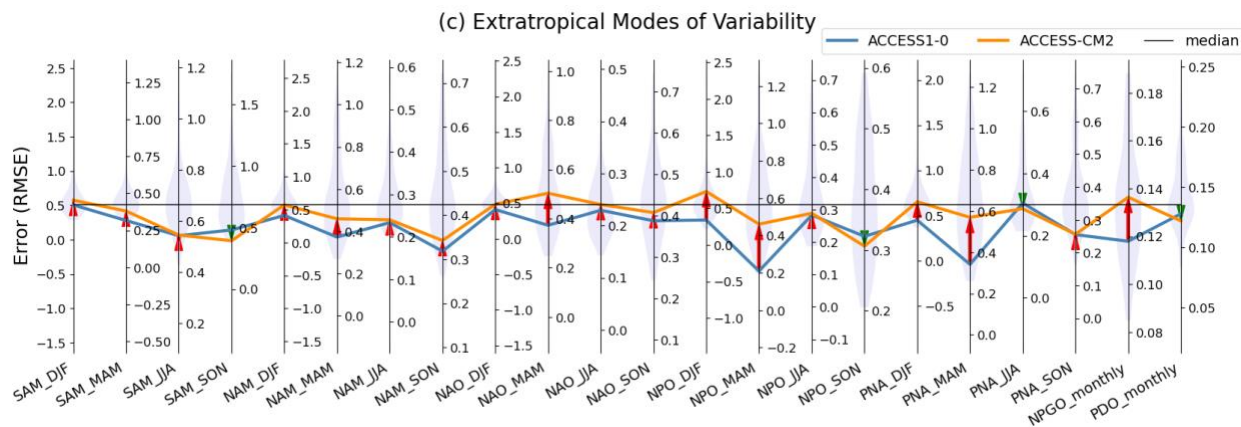
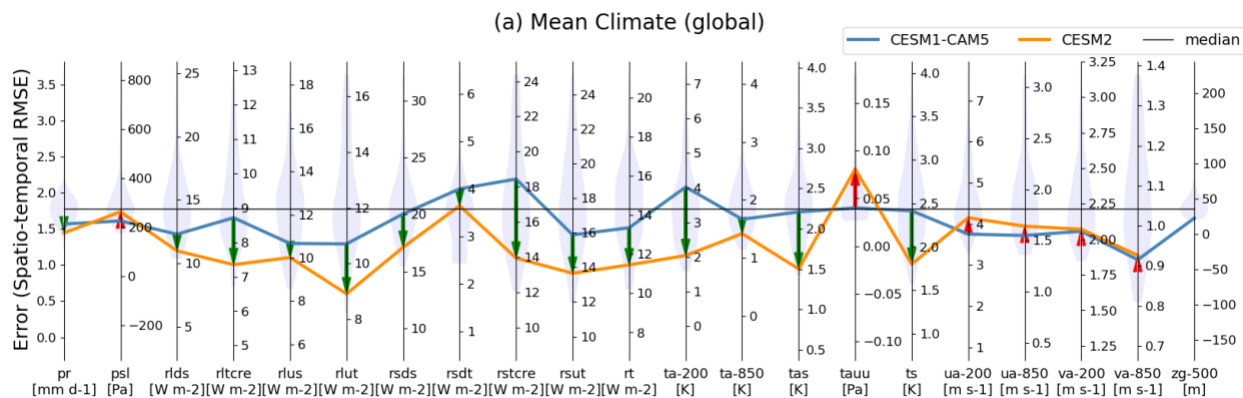
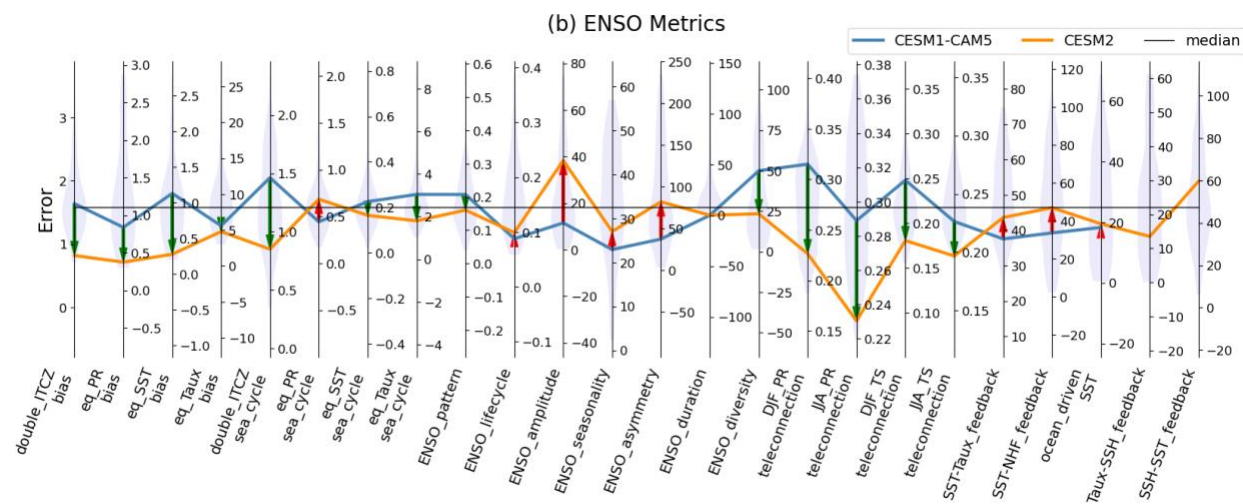
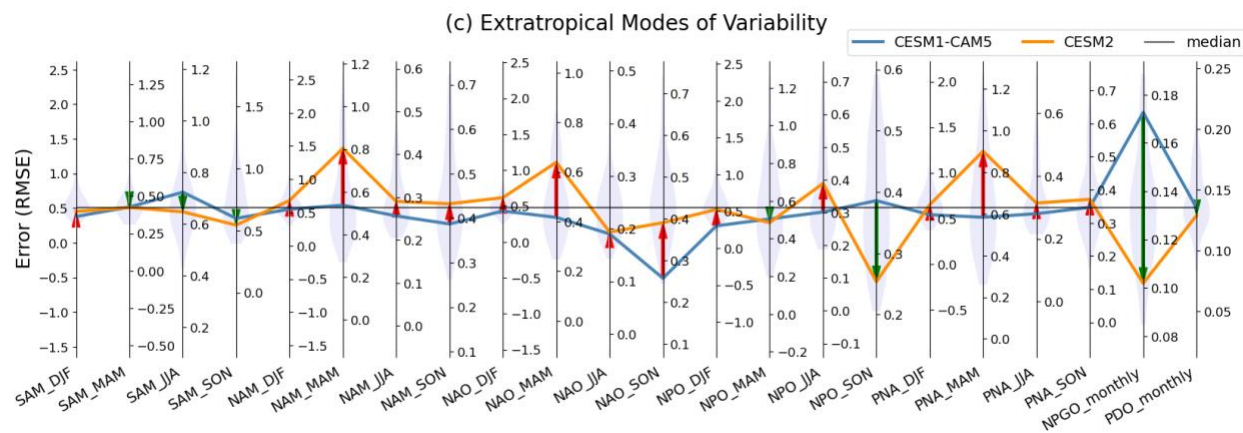
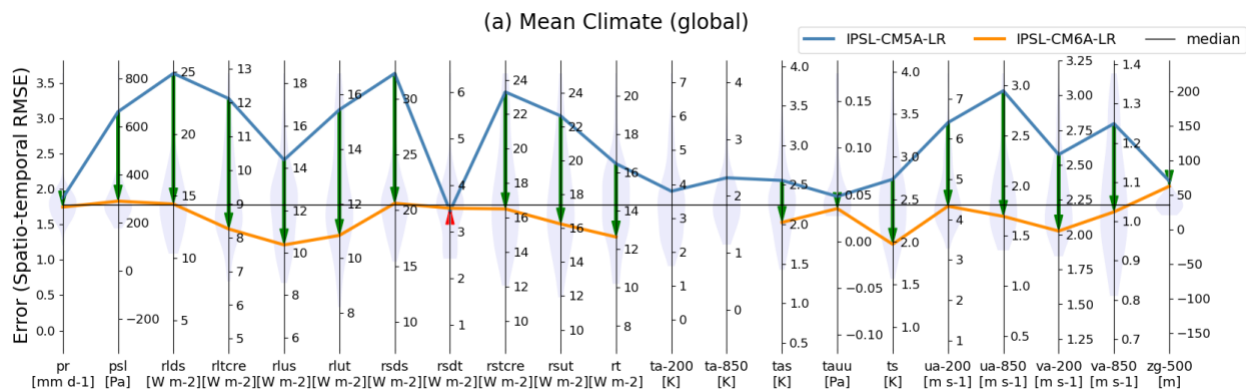
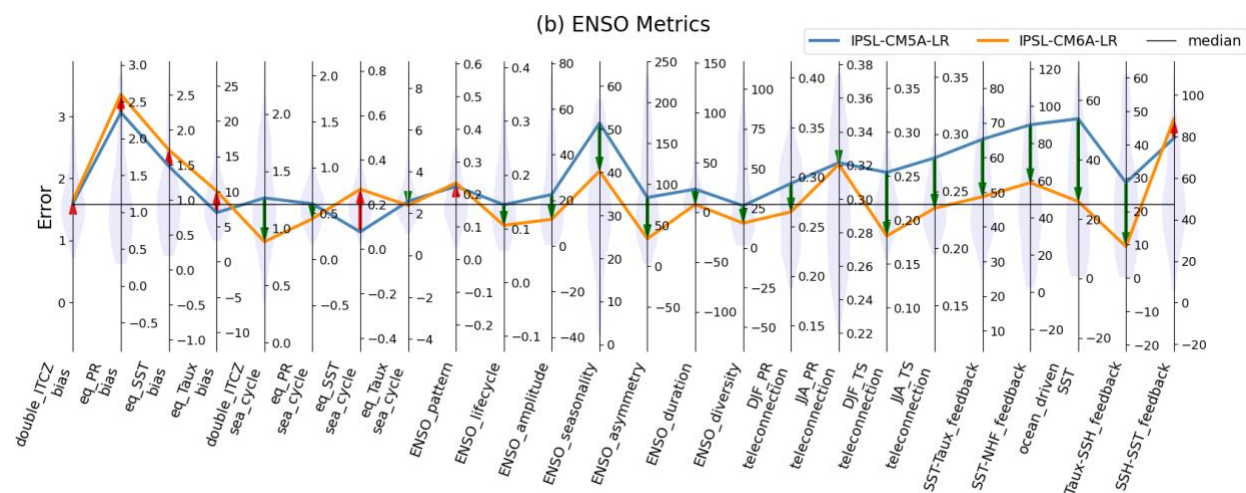
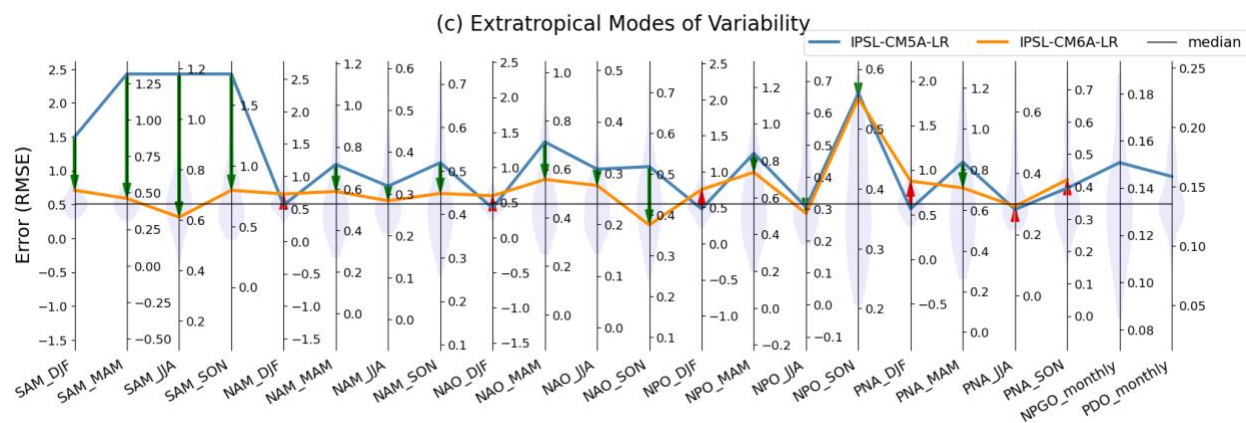



Figure S1. Parallel Coordinate Plot contrasting performance of two different versions of the ACCESS model (i.e., ACCESS-1-0 from CMIP5 and ACCESS-CM2 from CMIP6) in their Historical experiment for errors from (a) mean climate, (b) ENSO, and (c) extratropical modes of variability. Improvement (degradation) in the later version of the model is highlighted as a downward green (upward red) arrow between lines. Middle of each vertical axis is set to the median of the group of benchmarking models (i.e., CMIP5 and CMIP6), with the axis range stretched to maximum distance to either minimum or maximum from the median for visual consistency. The inter-model distributions of model performance are shown as shaded violin plots along each vertical axis.

45

46

47

48 **Figure S2.** Same as Figure S1, but for CESM models (i.e., CESM1-CAM5 from CMIP5 and

49 CESM2 from CMIP6)

50

51
52
53
54   **Figure S3.** Same as Figure S1, but for IPSL models (i.e., IPSL-CM5A-LR from CMIP5 and
55   IPSL-CM6A-LR from CMIP6)