Energy-budget evaluation in initialized decadal hindcasts

Confronting model trajectories with observations

Dominik Stiller (dstiller@uw.edu), **Gregory J. Hakim, Kyle C. Armour**
Department of Atmospheric and Climate Science, University of Washington, Seattle, USA

UNIVERSITY of WASHINGTON

Supplement

Key findings

- Hindcasts beat uninitialized simulations up to 1–2-year leads at outgoing longwave radiation and ocean heat content (OHC) up to 700 m, likely mediated by ENSO (Fig. 2).
- Skill is undetermined for net energy imbalance and absorbed shortwave radiation (Fig. 2), and for OHC below 700 m due to observational limitations.
- Initialization adds local but not detectably global OHC skill vs persistence (Figs. 2, 3).
- Hindcasts miss the internal variability behind the early-2000s warming slowdown (Fig. 4). They reproduce the decay but not growth of the 2015/16 El Niño (Fig. 5).

1. Introduction

- Coupled models are usually evaluated on the forced response and the statistics of internal variability. Trajectories and events cannot be compared directly, because the model is not in phase with the real world.
- Initialization from observations **synchronizes internal variability**, enabling 1) event-level model evaluation and 2) estimation of predictability limits.
- Earth's energy budget couples top-of-atmosphere (TOA) radiation and OHC through ocean heat uptake and the pattern-dependent radiative response, but **models struggle to reproduce both** (Olonscheck & Rugenstein, 2024; Kuhlbrodt et al., 2023).
- **We confront the energy budget in initialized decadal hindcasts with observations**:

Do predictions gain the right amount of energy,
and store it in the right places?

2. Methods and Data

- Hindcasts: 10-year *dcppA-hindcast* (Boer et al., 2016) from CanESM5, CESM1-CAM5, CESM2-CAM6, EC-Earth3, IPSL-CM6A-LR, MIROC6, NorCPM1 (starts: 1960–2015). Initialization methods differ (e.g., anomalies vs full field, reanalysis vs data assimilation).
- Verification: Berkeley Earth for surface temperature, ERSST for sea-surface temperature, CERES EBAF for TOA radiation (since 2001), EN4/IAP/Ishii/Levitus for OHC.

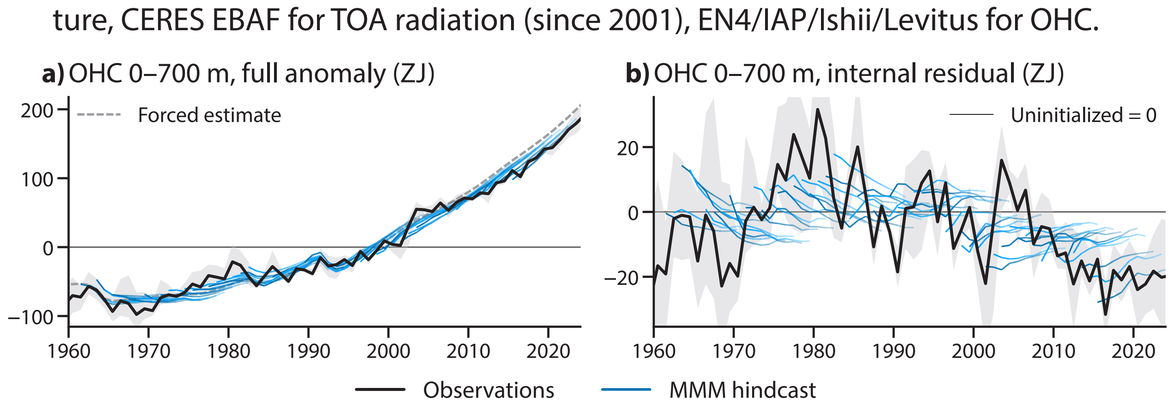

Figure 1 Bias and drift are removed with lead-dependent climatology. The residual isolates internal variability by removing the forced response estimated from uninitialized large ensembles (Smith et al., 2019). Curves are anomalies relative to 1981–2010.

3. Prediction skill

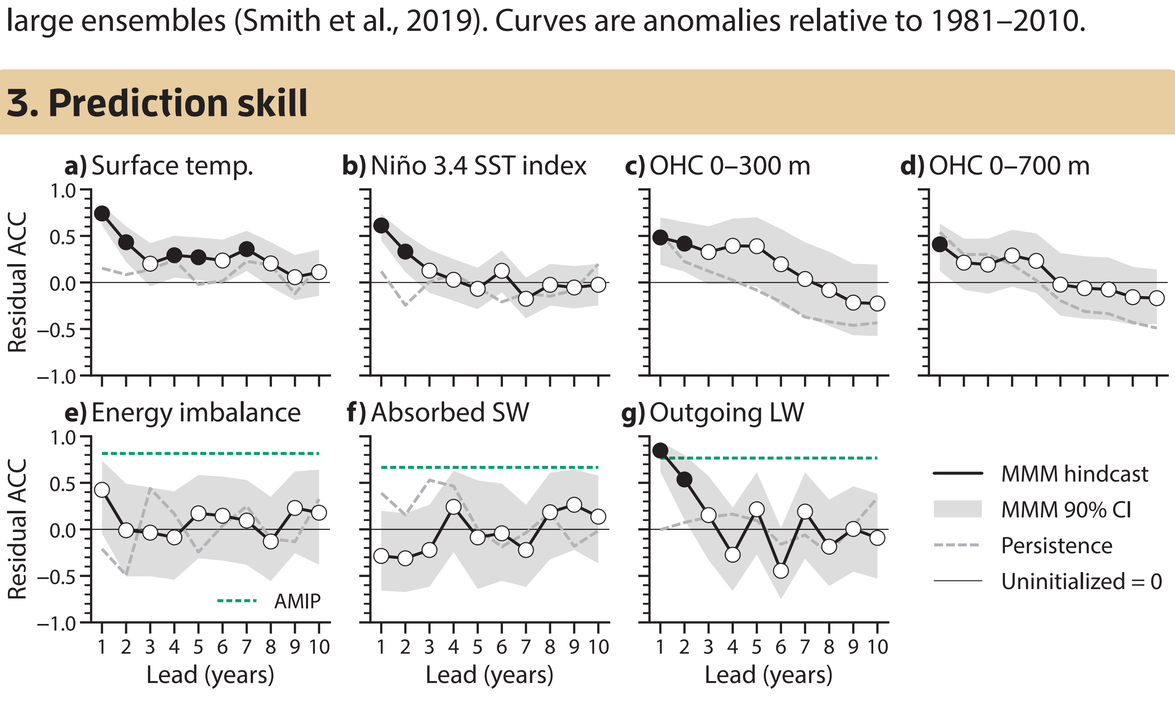

Figure 2 Prediction skill in internal variability of global means. ACC = anomaly correlation coefficient. AMIP = prescribed-SST simulations. MMM = multi-model mean.

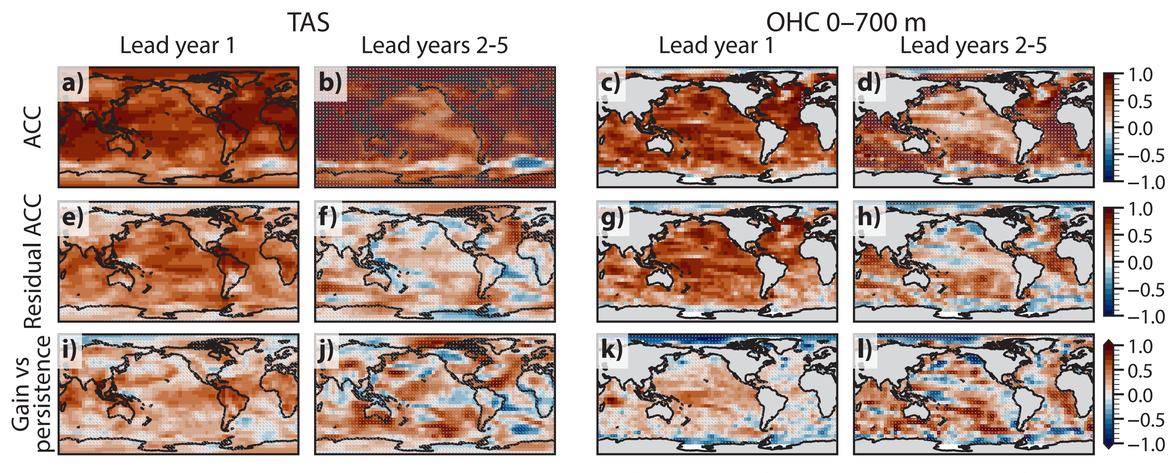

Figure 3 Prediction skill in internal variability of gridded fields. Unmarked = field-significant, · = locally significant only, ╲ = not significant, × = untestable.

4. Case studies: Early-2000s slowdown and 2015/16 El Niño

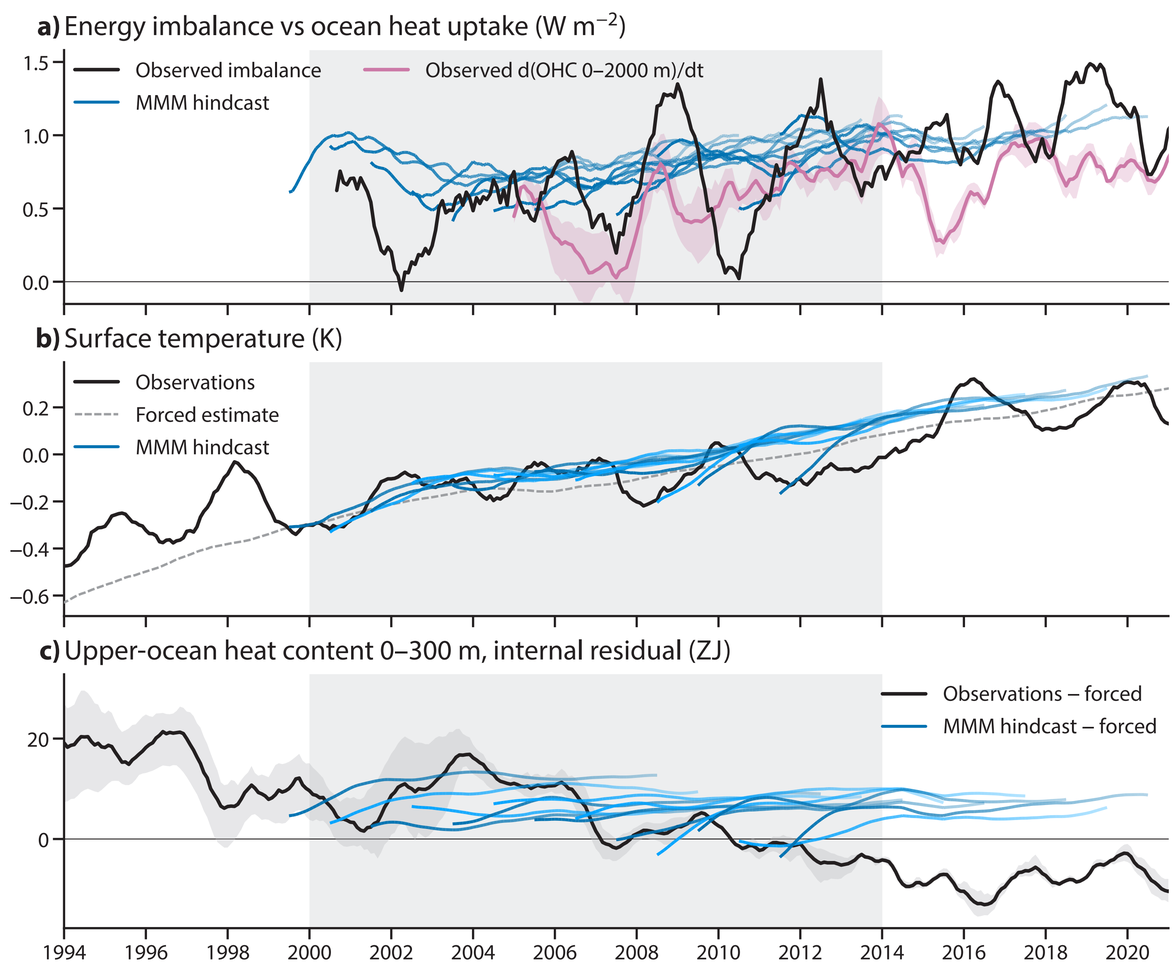

Figure 4 Early-2000s warming slowdown, a decade of stalled surface warming despite sustained ocean heat uptake. Anomalies in b) and c) are relative to 2001–2020. All series have a 12-month running mean, dOHC/dt a 36-month rolling trend.

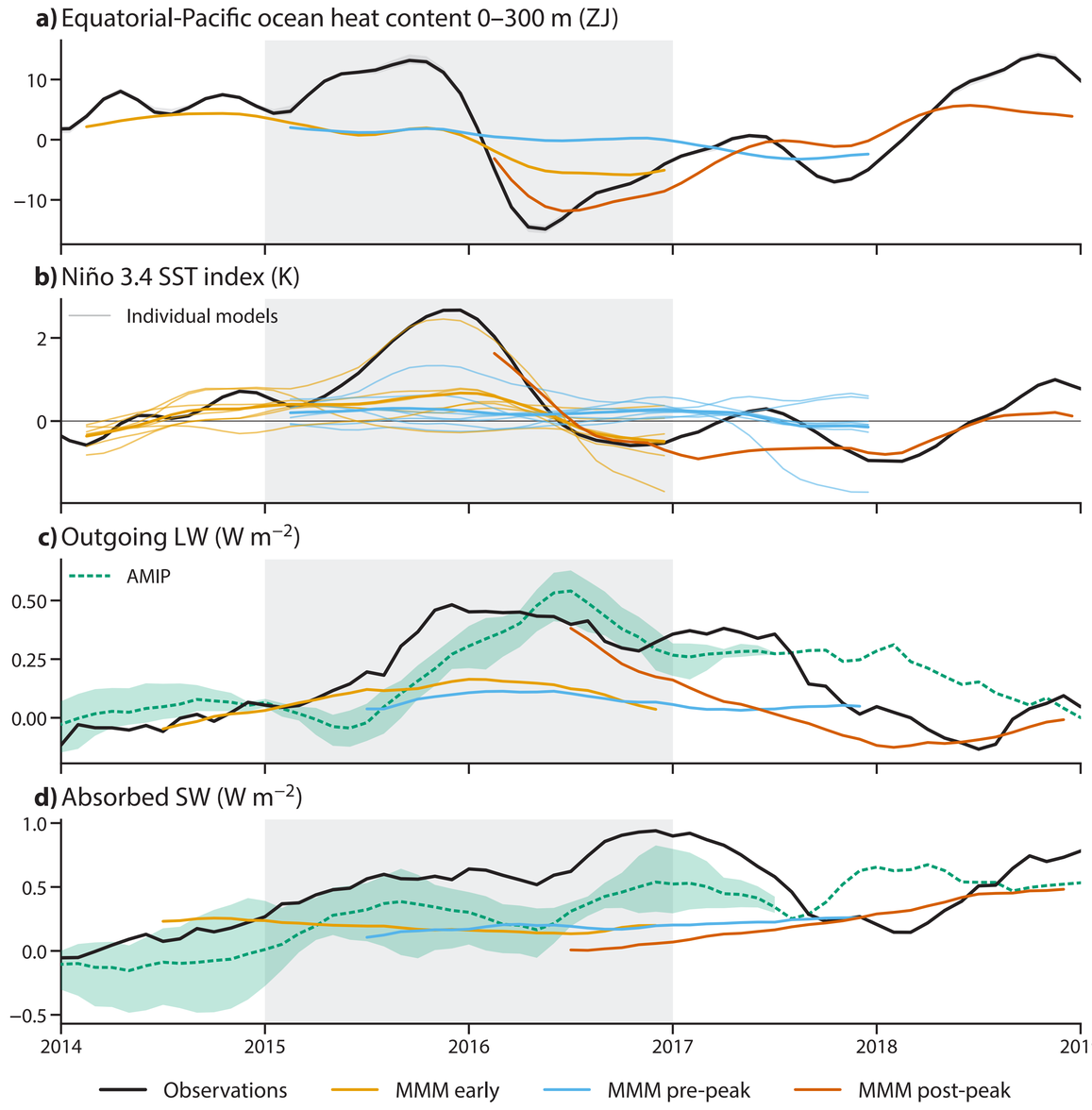

Figure 5 2015/16 El Niño, the strongest one of the CERES era. OHC (integrated over 5°S–5°N, 120°E–80°W, approximating warm water volume) and Niño 3.4 have a 3-month running mean, radiation has a 12-month one; anomalies are relative to 2001–2020.

5. What's next?

- Focus on predictions at seasonal-to-interannual timescales (e.g., CESM2 SMYLE), where models and observations can be more confidently compared.
- Diagnose mechanisms of model error that break the energy-budget chain of OHC distribution → SST pattern → TOA radiation → heat accumulation. We hypothesize that energy-budget consistency also improves surface-temperature skill.
- Predictions are skillful despite large drift and model error. Constraining budget consistency in models and initialization is an open opportunity.
- Perform hindcasts using differentiable and computationally cheap machine-learning climate emulators, then implement strongly coupled data assimilation.