**Capabilities and remaining challenges for SCREAM Evaluation**

*Chris Terai, Ben Hillman, Paul Ulrich, Weiran Liu, Hassan Beydoun, Peter Bogenschutz, Chris Golaz, Noel Keen, Wuyin Lin, Mark Taylor, Peter Caldwell*

We present the current capabilities and remaining challenges in evaluating SCREAM’s 3km resolution global model. When evaluating SCREAM against observations, we face two challenges. First, the large data output (100MB per 2D snapshot) slows our workflow, stresses storage, and limits movement of data. To obtain quick evaluations, we often first regrid most output variables to a coarser resolution before evaluating the output at the global scale. We also support the capability to plot the output on the model’s native grid for a more detailed examination. To address data storage limits, we leverage the model’s ability to write limited-area, high temporal frequency output to evaluate three-dimensional fields and process rates when we identify issues or biases that appear at smaller spatial and shorter temporal scales.

The computational cost of running the simulation has so far limited our test simulations to a couple days at a time. These short simulations keep us from using traditional evaluation strategies and diagnostics that compare simulation output to monthly-mean or climatological observations. A forecast diagnostic package that uses hourly ERA5 fields for comparison is being developed and tested, providing a quick evaluation of the model’s forecast skill both globally and regionally. Even in these short simulations, comparison between select model and observational fields suggests strengths and areas of improvement for the model.

Finally, we will discuss remaining challenges and questions that need to be addressed to more effectively evaluate the 3km model.