

## Description of dataset and statistical models

- Dataset: *multiresolution\_params\_and\_metrics.npy*
- Extra trees statistical models: *energy\_balance\_random\_forest.pkl* and *precipitation\_random\_forest.pkl*
- Code to read in statistical models and make predictions: *predict.py*

**Purpose:** Given Latin hypercube samples of 11 climate model input parameter values across three different resolutions (1 degree, 2 degrees and 4 degrees), determine the parameter value combinations that cause ensemble variability in global, annual average precipitation and energy balance at each resolution.

Use the simulations to train statistical models that can make high-resolution predictions of global, annual average precipitation and energy balance, leveraging fewer computationally expensive low-resolution simulations.

**Source:** The CAM5 ensemble used to generate this dataset was generated under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This work was supported by the DOE Office of Science through the Scientific Discovery Through Advanced Computing (SciDAC) project on Multiscale Methods for Accurate, Efficient, and Scale-Aware Models of the Earth System. This dataset is released under IM number LLNL-MI-746647.

**Relevant Information:** This dataset contains the values of global, annual average precipitation and energy balance computed from several multi-resolution climate ensembles. Ensemble members were constructed using a Latin hypercube method in LLNL's UQ Pipeline software system that sampled 11 important model parameters within their allowable ranges of variation in version 5 of the Community Atmosphere Model component of the Community Earth System Model (CESM). The dataset contains the parameter values and the corresponding values of global, annual average precipitation and energy balance for 906 simulations. There are 22 one-at-a-time (OAT) simulations per resolution, generated by setting each parameter to its high and low values, keeping the others at their default value. There is one independent Latin hypercube sampling (LHS) ensemble per resolution, with the sample size reduced by a factor of two between resolutions. The third experiment used 320 new LHS samples at 4 degrees, and then reused half and a quarter of the same samples at 2 degrees and 1 degree, respectively. This experiment is called LHS "matched" because a set of 80 parameter values is identical across resolutions. This information is summarized in the table below.

Resolution	# OAT runs	# LHS runs (unmatched)	# LHS runs (matched)	# Total runs
LOW (4 degrees)	22	160	320	502
MEDIUM (2 degrees)	22	80	160	262
HIGH (1 degree)	22	40	80	142
Total	66	280	560	906

## Attribute Information:

### multiresolution\_params\_and\_metrics.npy

*Columns 1-11:* values of 11 CAM5 parameters scaled in the interval [0, 1], appearing in the following order: cldfrc\_rhminl, micromg\_dcs, eddydiff\_a2l, uwshcu\_criqc, uwshcu\_rkm, uwshcu\_rpen, zmconv\_alfa, zmconv\_c0\_ocn, zmconv\_dmpdz, zmconv\_ke, zmconv\_tau.

*Column 12:* the resolution of the simulation. Values are 1.0, 0.25 or 0.0625 for 4 degrees, 2 degrees and 1 degree respectively.

*Column 13:* Global annual average energy balance in units of W/m<sup>2</sup>

*Column 14:* Global annual average precipitation in units of m/s

*Rows 1-22:* 22-member low-resolution OAT ensemble

*Rows 23 – 44:* 22-member medium-resolution OAT ensemble

*Rows 45 – 66:* 22-member high-resolution OAT ensemble

*Rows 67 – 226:* 160-member independent low-resolution LHS ensemble

*Rows 227 – 306:* 80-member independent medium-resolution LHS ensemble

*Rows 307 – 346:* 40-member independent high-resolution LHS ensemble

*Rows 347 – 666:* 320-member matched low-resolution LHS ensemble

*Rows 667 – 826:* 160-member matched medium-resolution LHS ensemble

*Rows 827 – 906:* 80-member matched high-resolution LHS ensemble

### energy\_balance\_random\_forest.pkl and precipitation\_random\_forest.pkl

Extra trees regressors trained on the dataset described above. See the supplementary material of paper cited below for more information.

### predict.py

This script reads in the extra trees regressors energy\_balance\_random\_forest.pkl and precipitation\_random\_forest.pkl, and makes a prediction of global annual average energy balance and precipitation respectively, at a user-specified resolution.

## Acknowledgments:

Please cite the following paper:

*Machine Learning Predictions of Multi-Resolution Climate Model Ensembles– Anderson, G. J. and Lucas, D. D. (GRL: in review). IM release number: LLNL-JRNL-744178.*