SCC25 and IndySCC25
Boosting Earth System Model Outputs And Saving PetaBytes in Their Storage Using Exascale Climate Emulators

Aug 14th 2025

Task 1: Mean Removal Trend

As the first step in building the Climate Emulator, each team is provided with three years of global surface temperature data in `NetCDF` format. The objective of this task is to remove the mean trend from the data and compute accuracy metrics. This task focuses only on **accuracy** (no performance scores).

Data and Software Setup

1. Download and install `ExaGeoStatCPP` by cloning the repository from:
   `https://github.com/ecrc/exageostatcpp`
   Use the `stage-zero-files-generation` branch for this task.

2. The code will run on a single CPU node (**no GPU required**).

3. Run the `Example_Stage_Zero` program using the following command:

```
./examples/stage-zero/Example_Stage_Zero \
  --kernel=trend_model \
  --data-path=/home/netcdf_files/ \
  --forcing-data-path=/home/forcing_new.csv \
  --lts=200 \
  --lb=0.001 \
  --ub=0.95 \
  --starting-theta=0.9 \
  --stage-zero \
  --cores=20 \
  --gpus=0 \
  --dts=200 \
  --resultspath=/home/outcpp/ \
  --startyear=2019 \
  --endyear=2021 \
  --numlocs=1440 \
  --max-mle-iterations=30 \
  --tolerance=7
```

`forcing_new.csv` is available in the main directory of the repository.

Expected Output

- The code will generate new measurements after removing the mean trend in the `/home/outcpp/` directory.
- Each output file corresponds to a single hour of data. For the given three `NetCDF` files (years 2019, 2020, and 2021), the total expected number of files is:

$$3 \times 365 \times 24 = 26{,}280$$

- Inside the `/home/outcpp/` directory, a file named `params.csv` will also be generated. This file contains 25 columns and 1440 rows, representing a subset of the grid with 1440 longitudes and 721 latitudes. The data is ordered starting from longitude 0, latitude $0, 1, 2, \ldots, 721$, then longitude 1, latitude $0, 1, 2, \ldots, 721$, and so on. The columns correspond to the spatial parameters in Equation (2) of the paper: `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10793148`

$$\rho,\ \sigma^2,\ \beta_0,\ \beta_1,\ \beta_2,\ a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}\ b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}$$

Reporting Requirements

Each team must submit:

1. The Mean Squared Error (MSE) between the original Y_t (measurements) from the `NetCDF` files (all three years) and the reconstructed Y_t from Equation (1) in the paper.
2. Plots of the norm of Y_t in Equation (1):
   - As a function of space.
   - As a function of time.

Contact Information

For any questions regarding this task, please contact:

1. sameh.abdulah@kaust.edu.sa
2. zubair.khalid@lums.edu.pk
3. qinglei.cao@slu.edu