Parallel Climate Analysis
using structured data

github.com/laitifranz/parallel-climate-analysis

UNIVERSITÀ
DI TRENTO

Laiti Francesco - Lobba Davide

High Performance Computing for Data Science
Academic Year 2022/2023

Dataset

Dataset card

- Dimensions
  - Latitude
  - Longitude
  - Time
- Variables
  - Latitude
  - Longitude
  - Time
  - Precipitation flux
- Format
  - NetCDF

- For better benchmarks
  - Merge multiple datasets along time dimension using cdo library

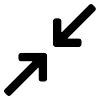
- New temporal window available for precipitation flux

Serial implementation

Serial implementation

READING

- Open .nc file
- Retrieve dimension IDs and variable IDs
- 2-dim matrix for temporal data
- 2-dim matrix for output data



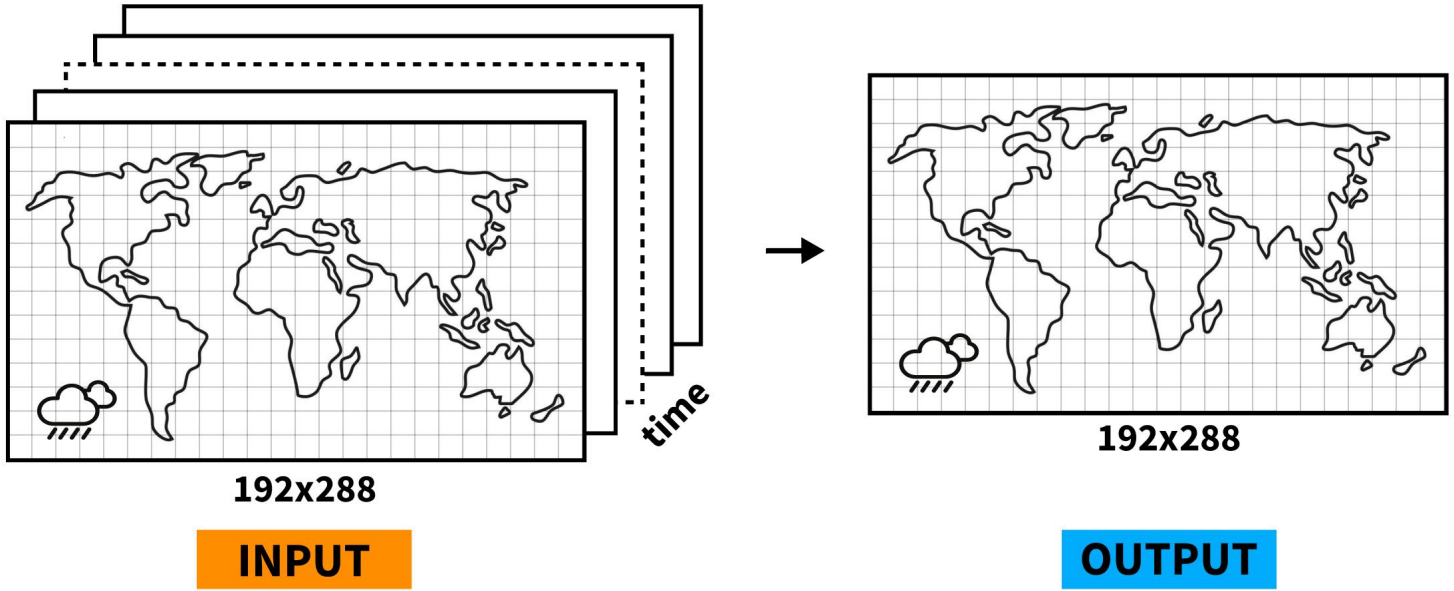
REDUCING

- Looping along time dimension
- Retrieve matrix regarding precipitation flux
- Nested loop sum up to the output matrix the temporal one

WRITING

- Average output matrix
- Write a new netCDF file

Serial implementation



Serial implementation graphically represented

Parallel implementation

MPI and Hybrid Parallelization

MPI

READING

- Open .nc file
- Retrieve dimension IDs and variable IDs
- 2-dim matrix for temporal data
- 2-dim matrix for output data

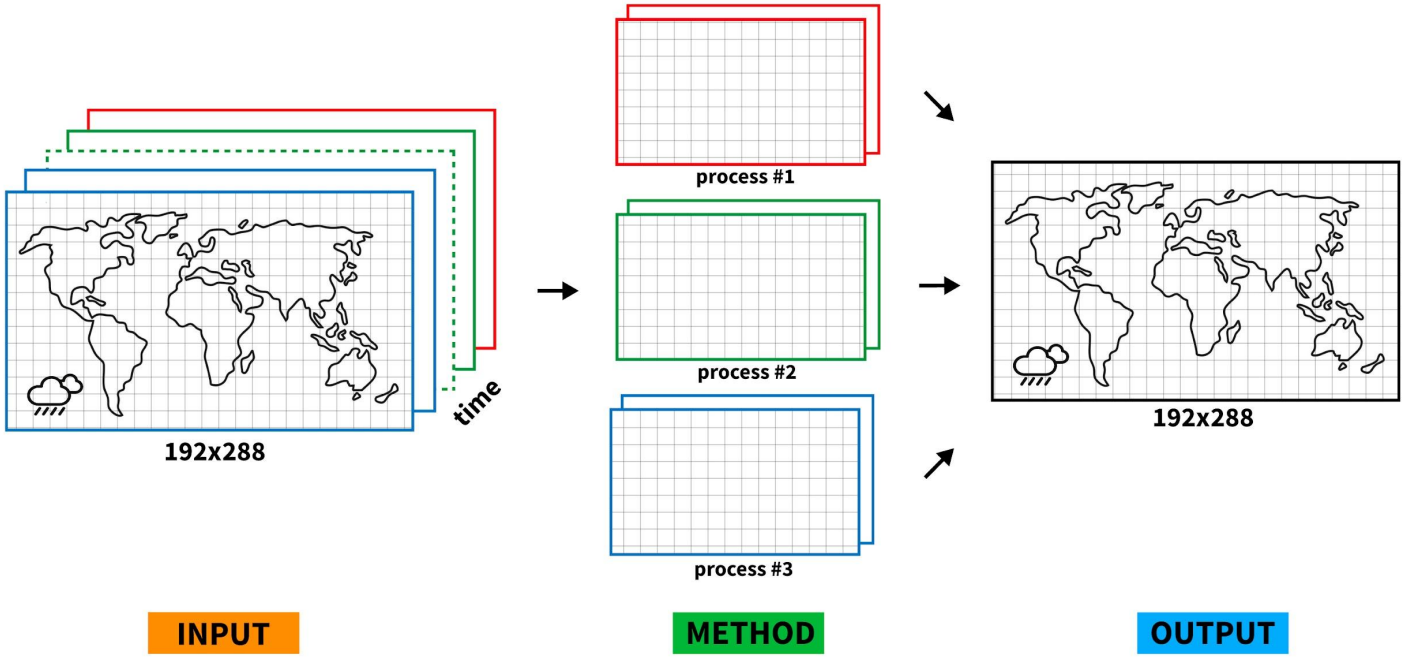
REDUCING

- Split the time dimension across the available processes
- For each process:
  - Loop along time dimension
  - Compute the local precipitation flux matrix
- Reduce and compute the final precipitation flux matrix

WRITING

- Only in the root process:
  - Average output matrix
  - Write a new netCDF file

MPI



MPI implementation graphically represented

Hybrid parallelization

READING

- Open .nc file
- Retrieve dimension IDs and variable IDs
- 2-dim matrix for temporal data
- 2-dim matrix for output data

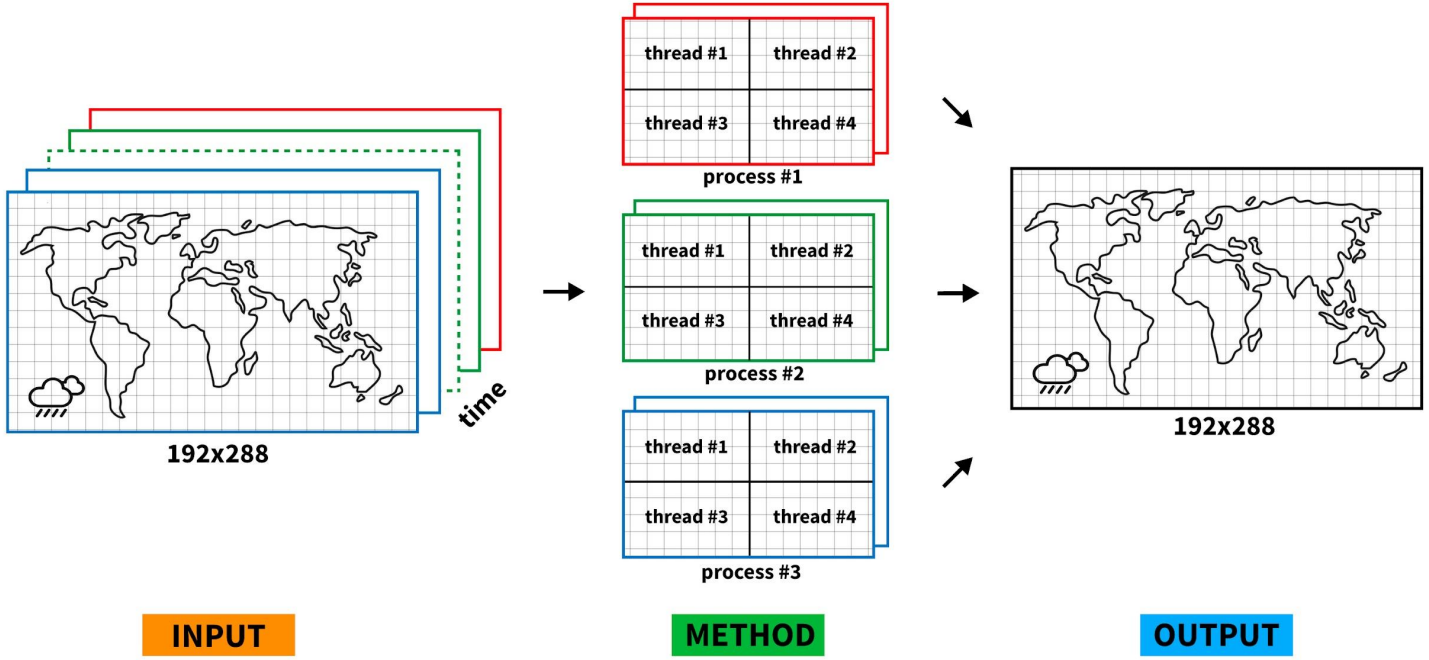
REDUCING

- Split time dimension across the available processes and the grid across the available threads
- For each process:
  - Loop along time dimension
- For each thread:
  - Loop along lat and lon dimensions
  - Compute the local matrix precipitation flux matrix
- Reduce and compute the final precipitation flux matrix

WRITING

- Only in the root process:
  - Average output matrix
  - Write a new netCDF file

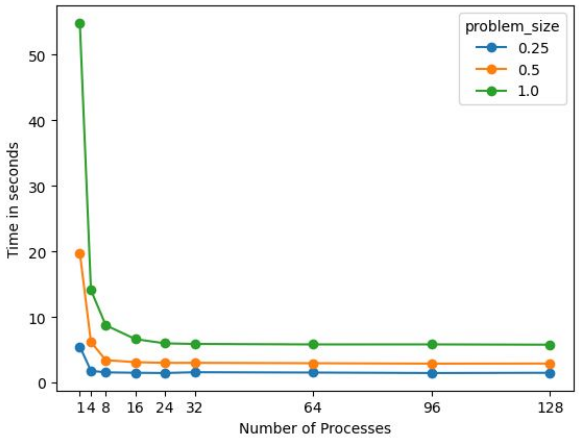
Hybrid parallelization



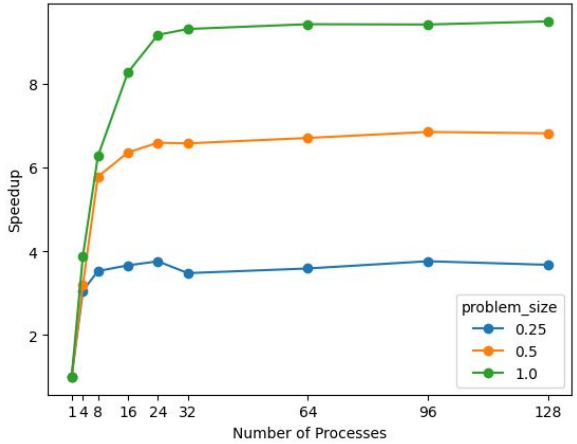
Hybrid implementation graphically represented

Performance & Benchmarking

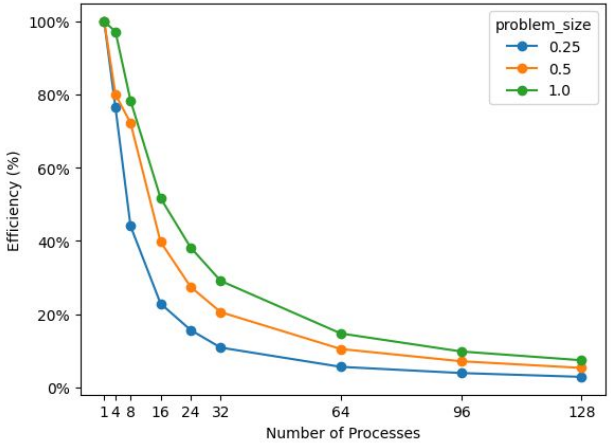
MPI using different problem size



Time vs. Number of processes

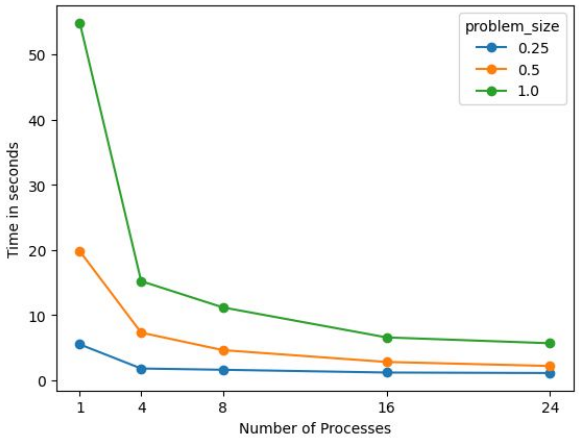


Speedup vs. Number of processes

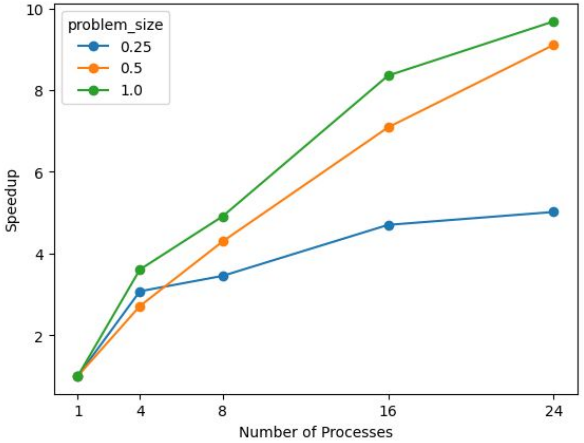


Efficiency vs. Number of processes

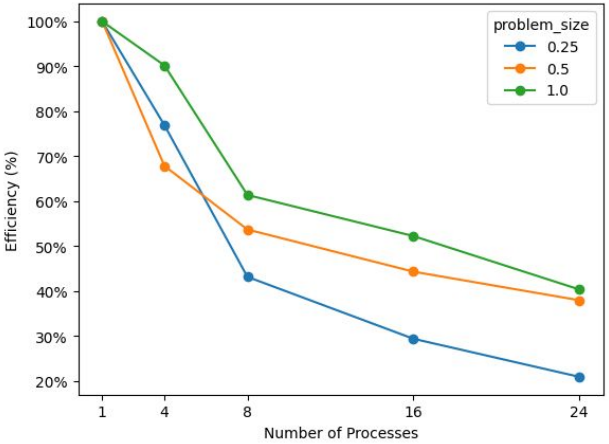
Hybrid parallelization using different problem size



Time vs. Number of processes

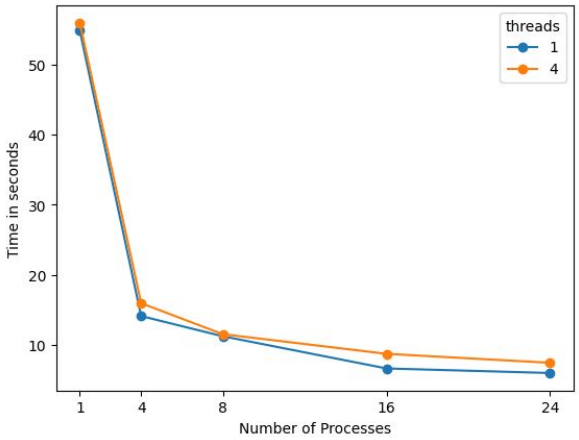


Speedup vs. Number of processes

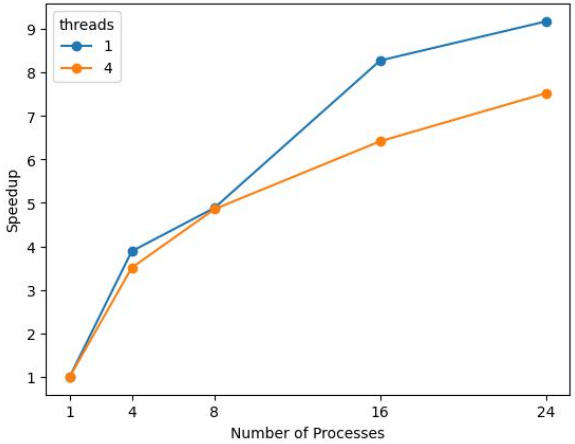


Efficiency vs. Number of processes

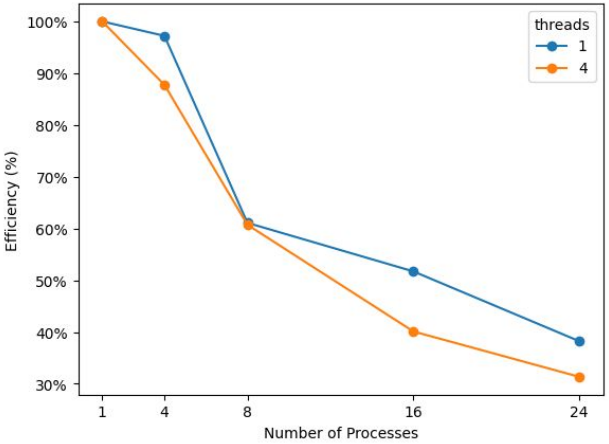
Comparison using problem size = 1.0



Time vs. Number of processes



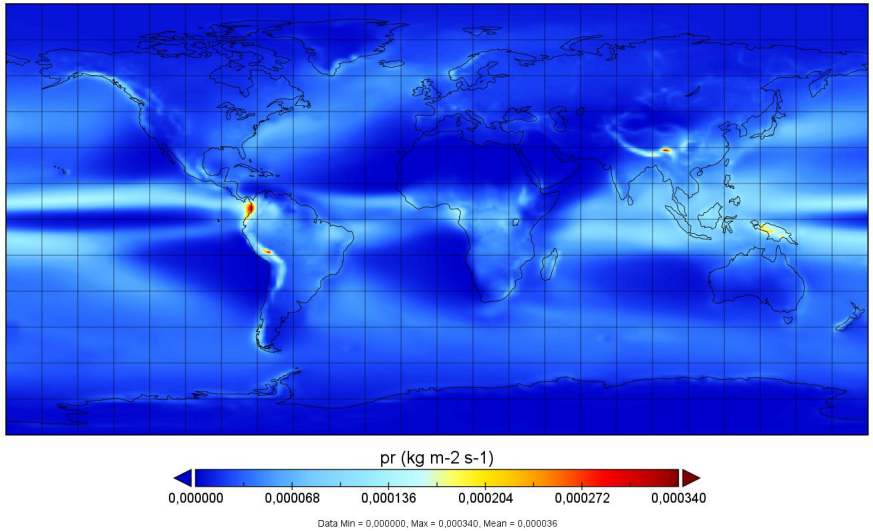
Speedup vs. Number of processes



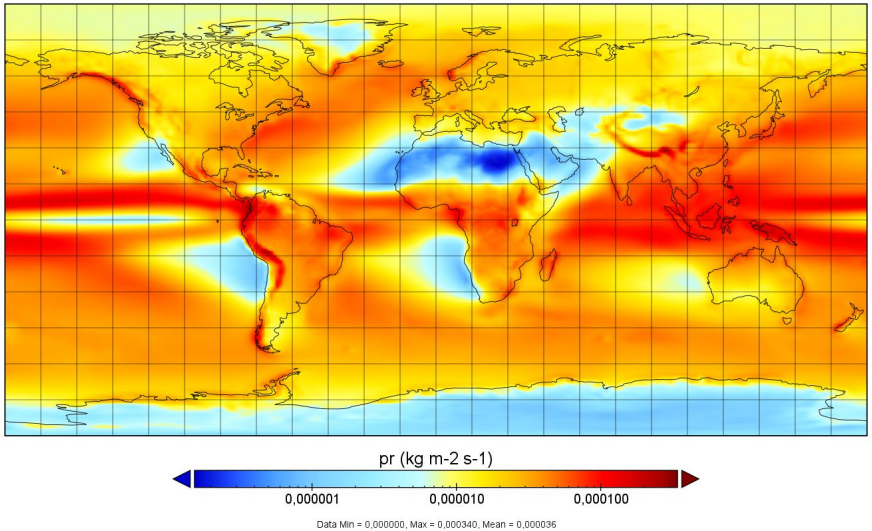
Efficiency vs. Number of processes

Results using Panoply

Standard scale



Logarithmic scale



Mean precipitation flux from 1850-01-01 to 2014-12-31 (60225 days)

Parallel Climate Analysis
using structured data

UNIVERSITÀ
DI TRENTO

Laiti Francesco - Lobba Davide

High Performance Computing for Data Science
Academic Year 2022/2023