Coupling km-Scale Earth System Model to Hierarchical Output for Analysis-Ready Dataset

Nils-Arne Dreier[1] and Siddhant Tibrewal[2]

[1]*German Climate Computing Center (DKRZ), Hamburg, Germany*
[2]*Max-Planck-Institute for Meteorology, Hamburg, Germany*

Introduction

Kilometer-scale Earth System Model (ESM) simulations produce petabyte-scale outputs that are difficult to access, analyse, and share due to their size, heterogeneity, and the overhead of ad-hoc workflows. We introduce **Hiopy** (Hierarchical Output in Python), a lightweight in-situ output component, that is coupled to the ICON model with YAC (Yet Another Coupler) and writes multi-resolution, self-describing datasets directly to the cloud in Zarr format.

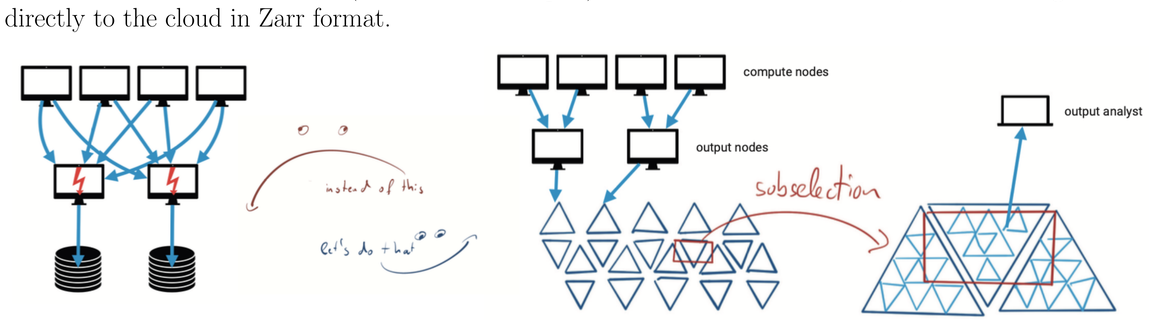

Fig. 1: Conceptual sketch of the Hiopy system. Courtesy of Tobias Kölling.

Hierarchical Healpix

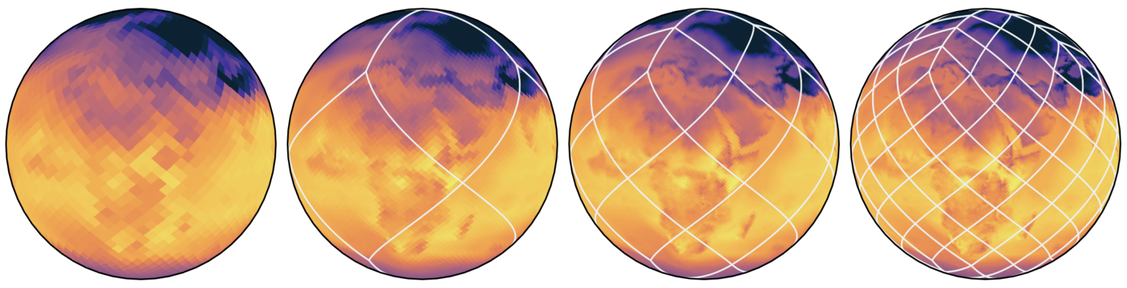

Fig. 2: Hierarchical subdivision of the HEALPix pixelization scheme.

The HEALPix grid provides a hierarchical pixelization of the sphere, where each pixel at a given resolution can be subdivided into four child pixels at the next finer resolution.

Zarr

Zarr is a cloud-optimised format for storing chunked, compressed N-dimensional arrays. It is self-describing, supports parallel reads and writes, and maps naturally to object storage systems such as S3, where datasets can be streamed directly over HTTPS without any data movement. Hiopy writes datasets directly in Zarr format, aligning model domain decomposition with Zarr chunking to minimise communication overhead and enable efficient parallel I/O without post-processing bottlenecks. Because data is divided into small, independently retrievable chunks, analysis tools can load only the subset of data relevant to a given query, dramatically reducing bandwidth and memory requirements even for petabyte-scale datasets.

Easy analysis

The Zarr datasets produced by Hiopy are directly accessible via standard Python libraries such as xarray and zarr, enabling analysis without any preprocessing or data movement. Users can open datasets with a single line of code and leverage lazy loading to work with petabyte-scale data on modest hardware.

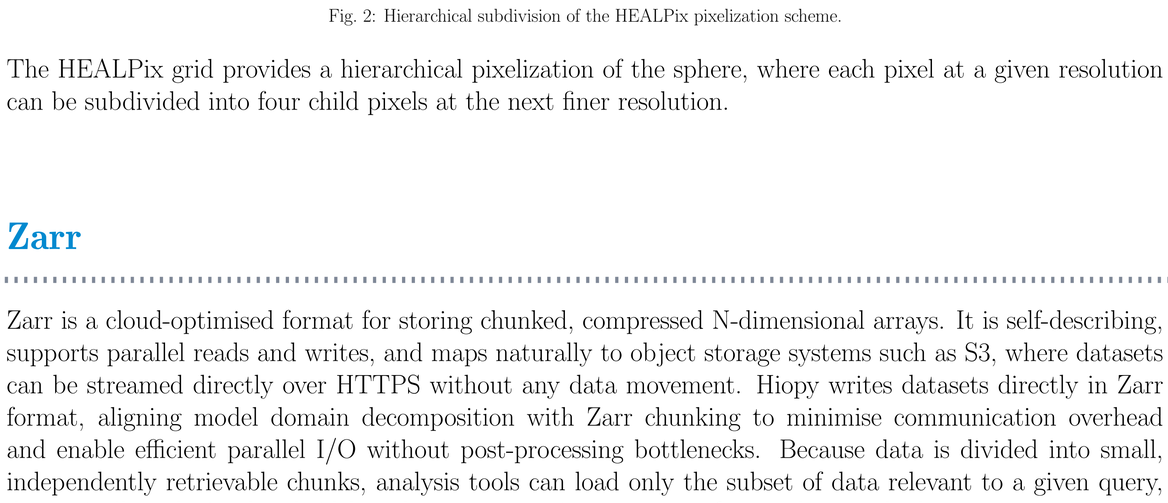

Fig. 3: Example plot showing a subsection of the global domain at high resolution.

Implementation

Hiopy computes hierarchical temporal and spatial aggregates on-the-fly, aligning model domain decomposition with Zarr chunking to minimise communication, balance workload across processes, and preserve simulation throughput. The framework also streamlines metadata handling, eliminates redundant buffers, and streams data straight to its final storage location. Hiopy supports native ICON grids, regular latitude-longitude grids, and HEALPix, and has been validated by generating publicly accessible, analysis-ready datasets from several kilometre-scale ESM projects.

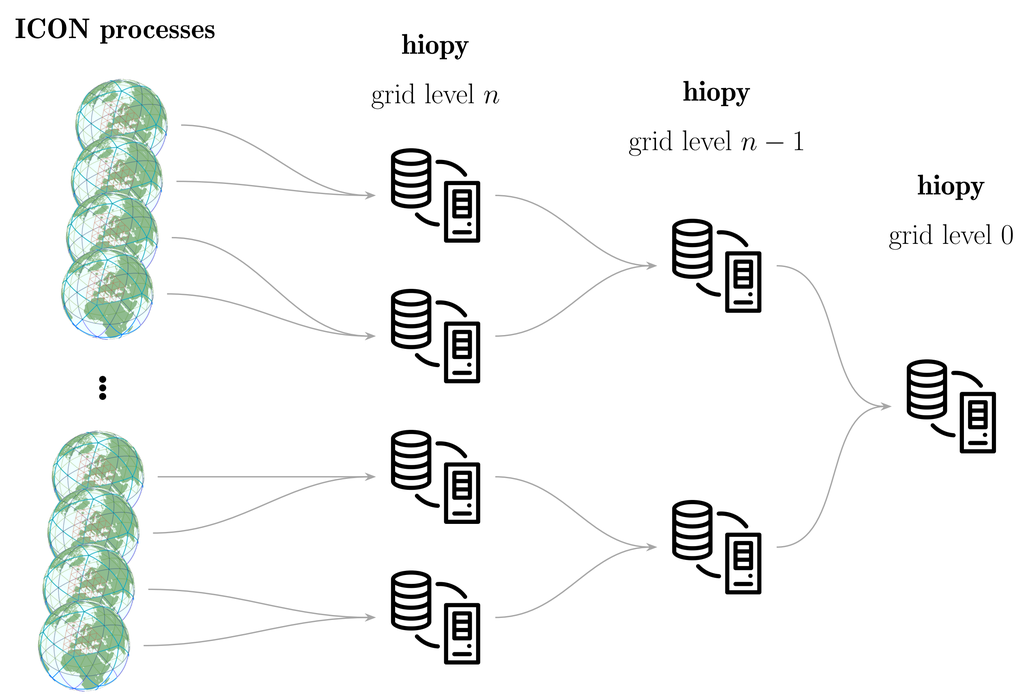

Fig. 4: Schematic overview of Hiopy's communication pattern.

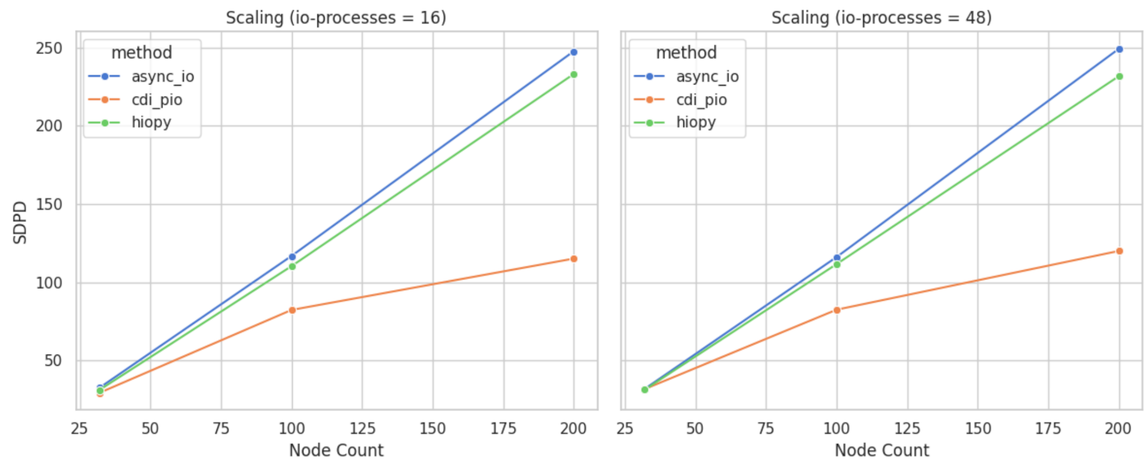

Fig. 5: Performance Results at Scale

Showcases

- **NextGEMS Cycle 4** (`icon_ngc4008`): 5 km global coupled simulation (up to HEALPix level 10), spanning 30 years and producing approximately 5 petabytes of analysis-ready data, computed on Levante.
- **Global Hackathon** (`icon_d3hp003`): A 2.5 km global atmosphere-only simulation (up to HEALPix level 11) conducted over 14 months, yielding 273 terabytes of data. This enabled interactive exploration by a large international user group without any preprocessing.
- **xSPIES** (`icon_xps005`): A 1.25 km global coupled simulation spanning 14 months with 680 terabytes of data—one of the highest-resolution ESM runs to date—with Hiopy-enabled output running on Jupiter.

Conclusion

This work demonstrates a practical, scalable addition to the high-resolution climate-modelling software stack, enabling seamless, cost-effective access from coarse to native resolution without post-processing bottlenecks.

References

[1] T. Kölling and L. Kluft. *Building useful datasets for Earth System Model output*. EGU24-17141. Vienna, Austria, 2024. DOI: 10.5194/egusphere-egu24-17141.

[2] S. Tibrewal and N.-A. Dreier. *Coupling km-Scale Earth System Model to Hierarchical Output for Analysis-Ready Dataset*. EGU General Assembly 2026, Vienna, Austria, 3–8 May 2026. EGU26-9822. 2026. DOI: 10.5194/egusphere-egu26-9822.

DKRZ DEUTSCHES KLIMARECHENZENTRUM

MAX-PLANCK-INSTITUT FÜR METEOROLOGIE

WARM WORLD