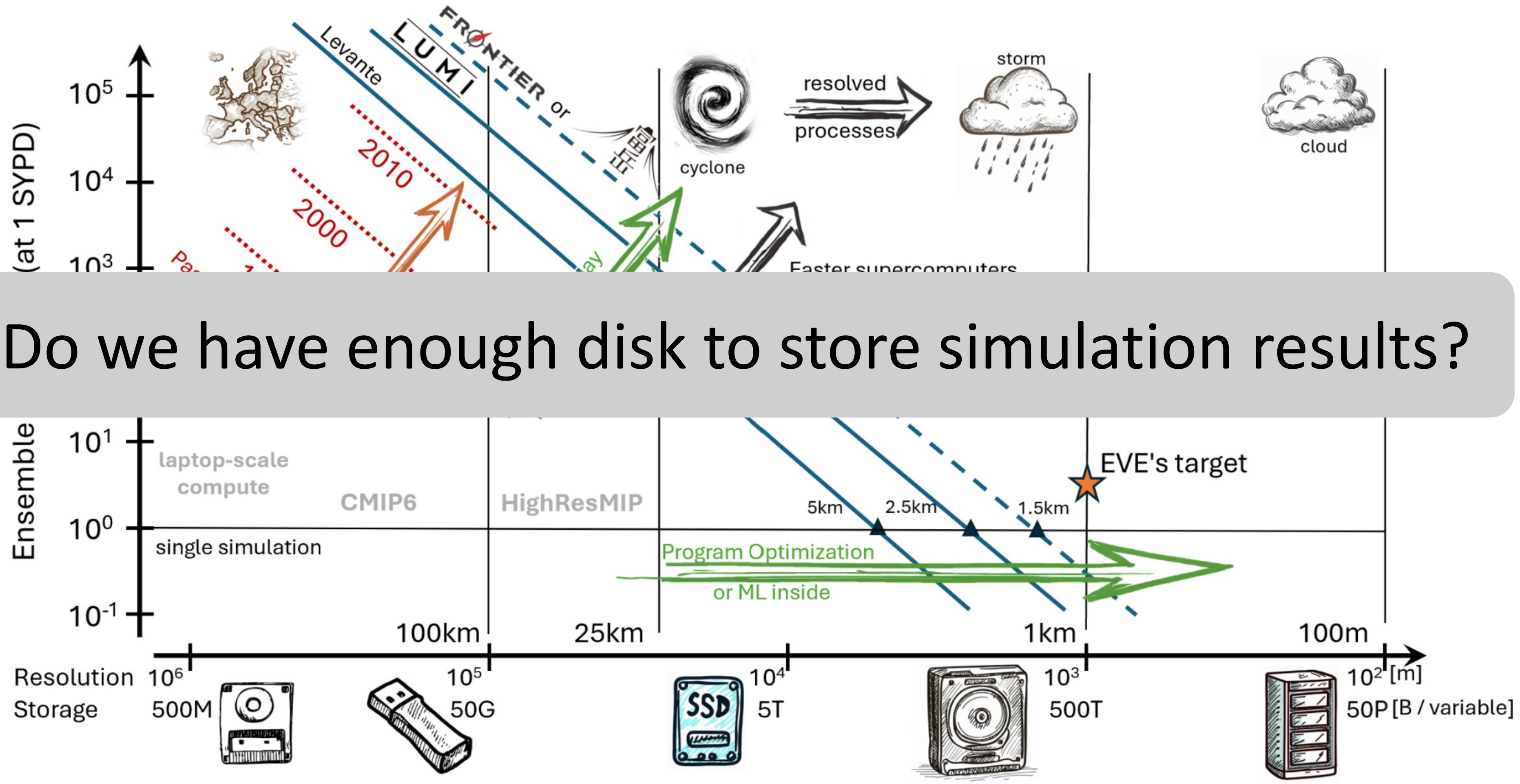
LANGWEN HUANG, LUIGI FUSCO, FLORIAN SCHEIDL, JAN ZIBELL, MICHAEL ARMAND SPRENGER, SEBASTIAN SCHEMM, TORSTEN HOEFLER

EBCC: Error Bounded Compressor for Climate data

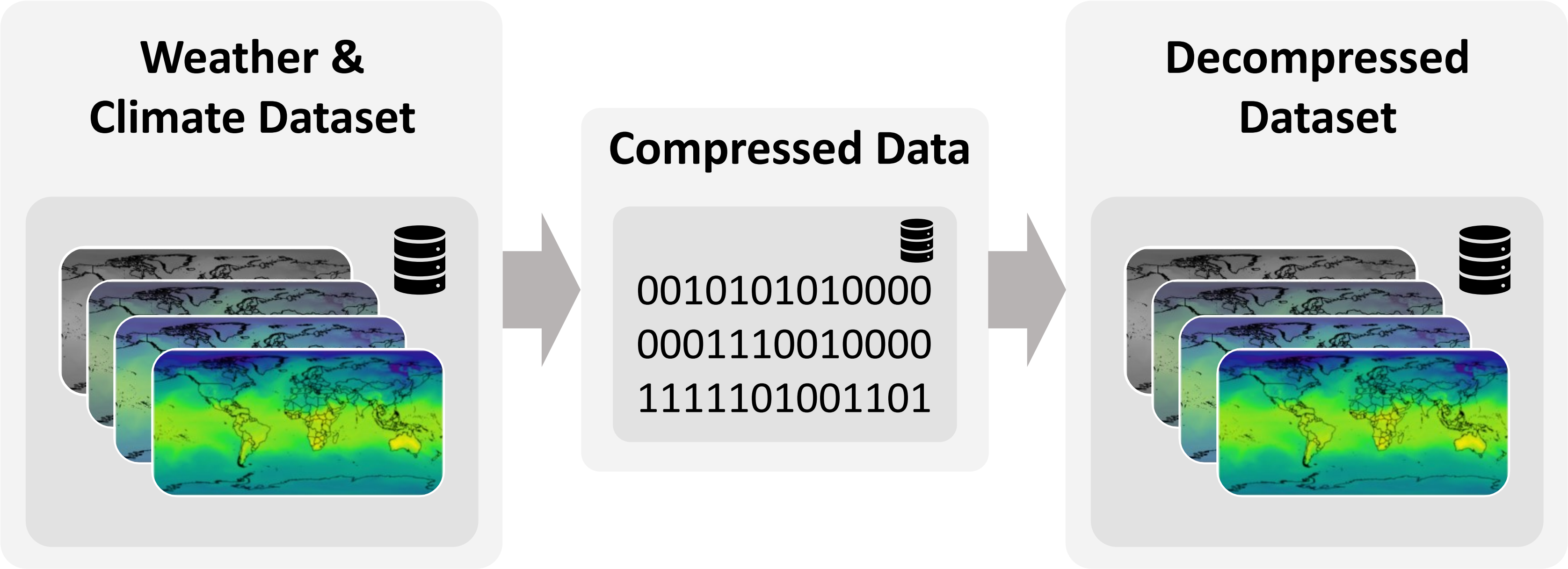
Scaling of climate simulation



Do we have enough disk to store simulation results?

Hoefler, T., Stevens, B., Prein, A.F., Baehr, J., Schulthess, T., Stocker, T.F., Taylor, J., Klocke, D., Manninen, P., Forster, P.M. and Kölling, T., 2023. Earth Virtualization Engines: a technical perspective. *Computing in Science & Engineering*, *25*(3), pp.50-59.

Overview



Compression Methods

Lossless compression

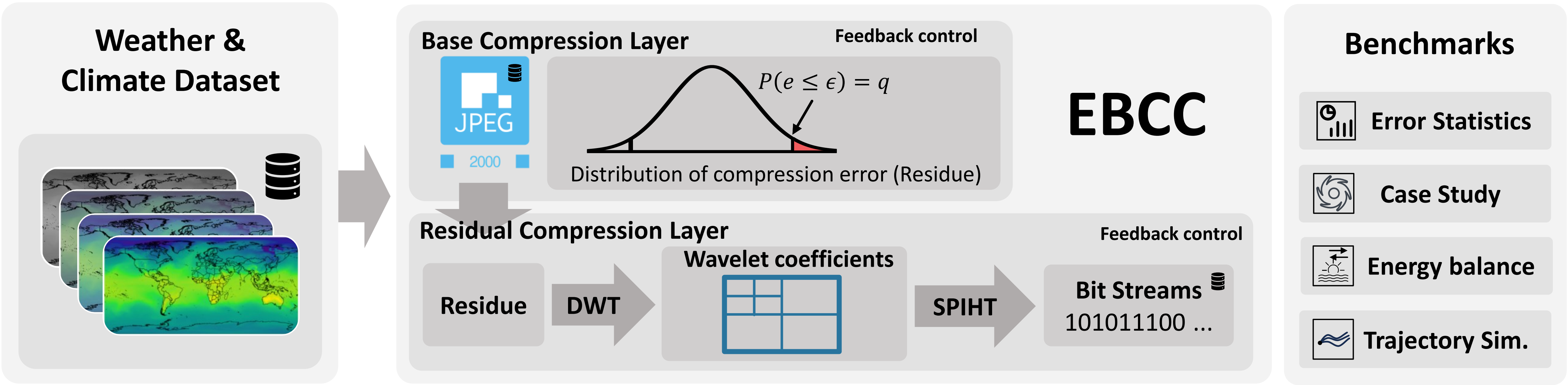
- Exact reconstruction
- Compression ratio (CR): 2 - 5x

Lossy compression

- Decompressed ~ Original
- Can achieve arbitrary CR
- High CR ↔ High error
- Hard to justify an acceptable error level

We will focus on max error bounded lossy compression.

Overview



$q = 1 \Leftrightarrow$ **Error bounded JPEG2000**

Compression Mode:

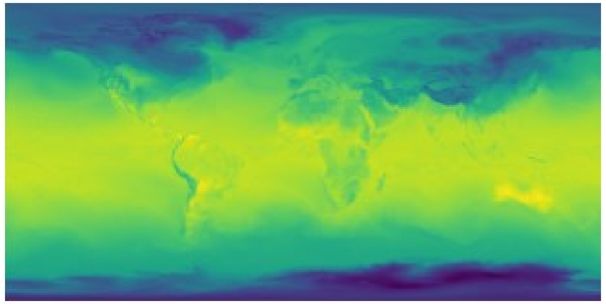
- Max absolute error target
- Range-relative max absolute error target
  max_abs_err/(max – min)

Icon of JPEG2000 taken from https://jpeg.org/jpeg2000/

Method: Base Compression Layer

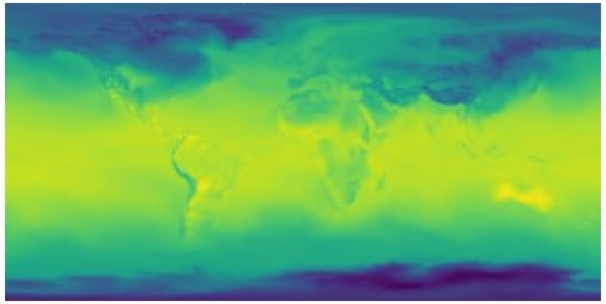
Compression Ratio = 500x

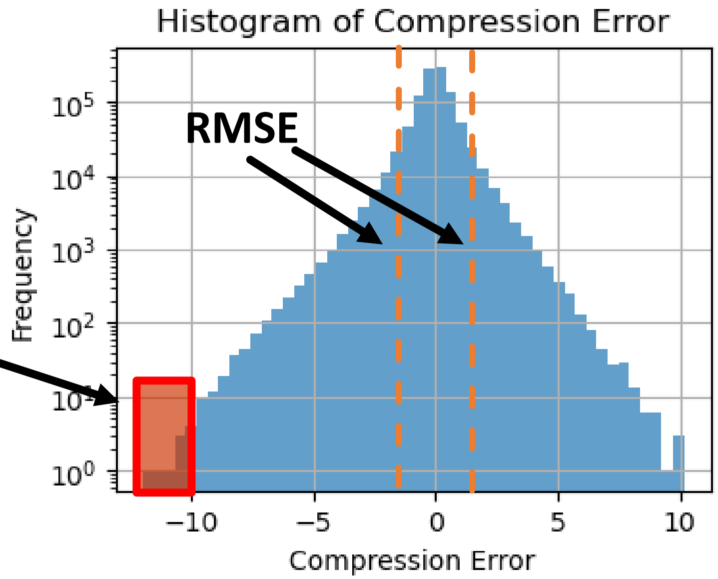
Original Data



JPEG 2000

Compressed Data

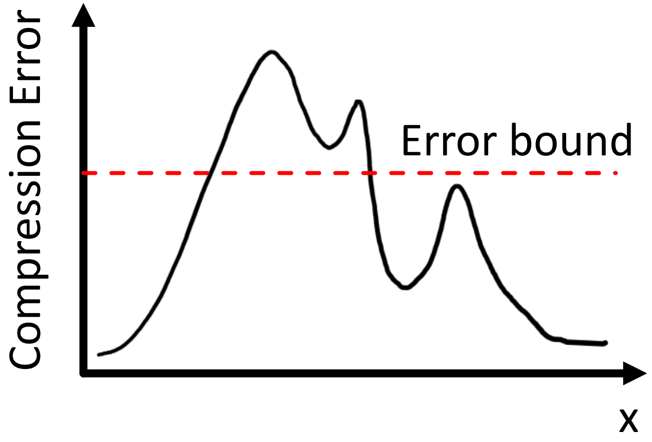




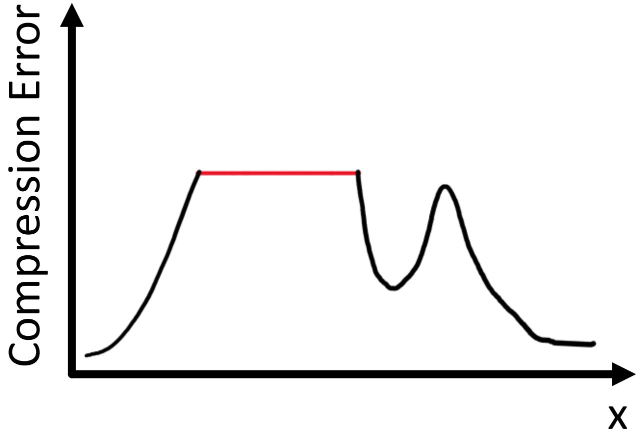
Excessive Errors

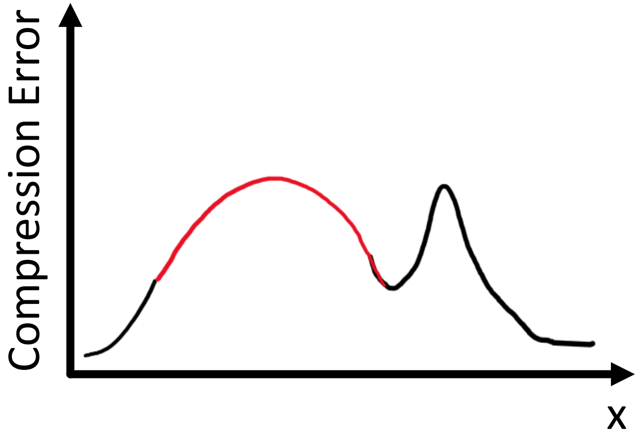
JPEG2000 is good at minimizing RMSE at a given CR, but excessive errors much larger than RMSE may occur!

Method: Residual Compression Layer



1D Slice of Compression Error





Idea: log down errors that exceed the error bound.

Hard Clip

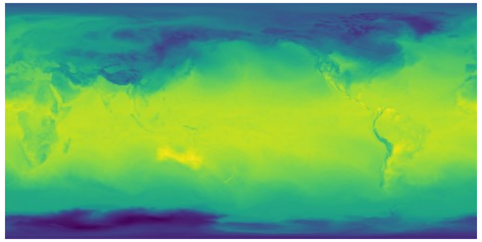
\+ Require less bits
\+ Easy to implement
\- Abrupt change in gradient
\- Zero gradient on clipping plateau

Soft Clip

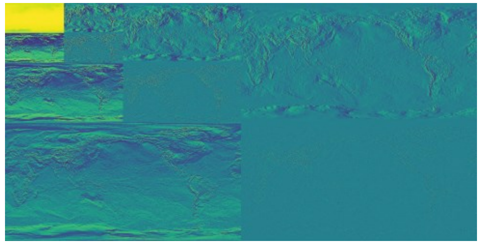
\+ Smooth
\+ Does not introduce unphysical artifacts
\- Require more bits

Implement soft clipping by 'smart thresolding' wavelet coefficients.

Method: Residual Compression Layer



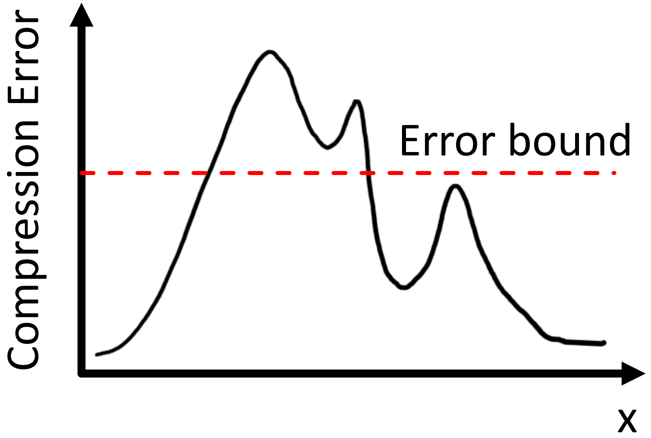
Wavelet Transform



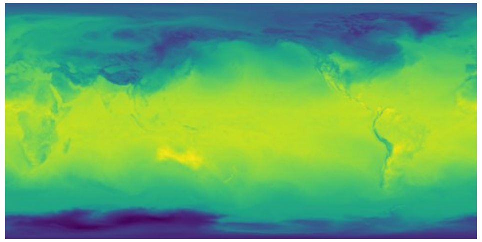
SPIHT

01110000010111000001100001101110O
1110101011100110010000101011001 00
1111001100101011111111001101111 11
0011111001110101111001011010101 01
1100100001010010101111111100

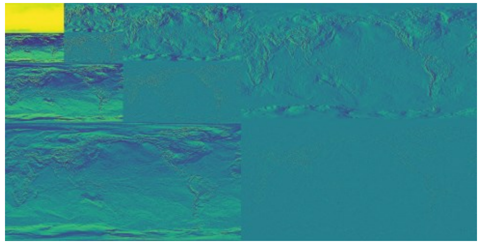
01110000



Method: Residual Compression Layer



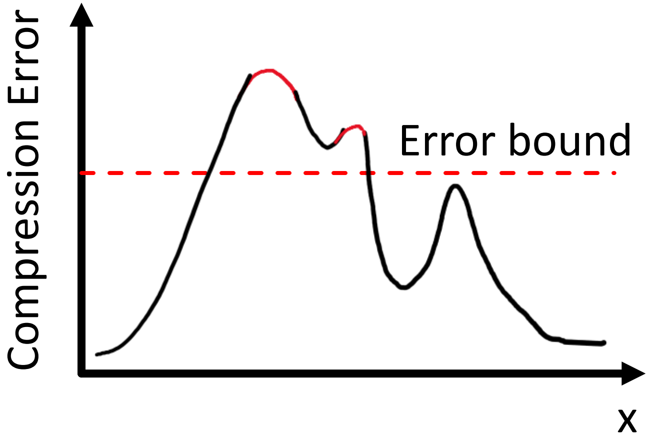
Wavelet Transform



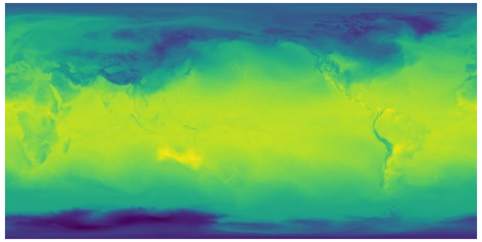
SPIHT

011100000101110000011000011011100
111010101110011001000010101100100
11110011001010111111100110111111
001111100111010111100101101010101
11001000010100101011111111100

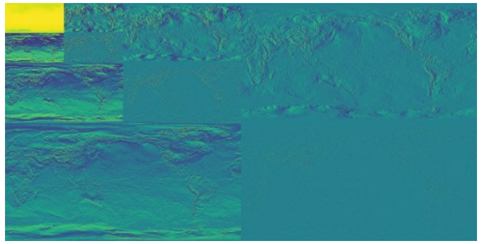
01110000**01011100000110000**
110111001110101011100110



Method: Residual Compression Layer



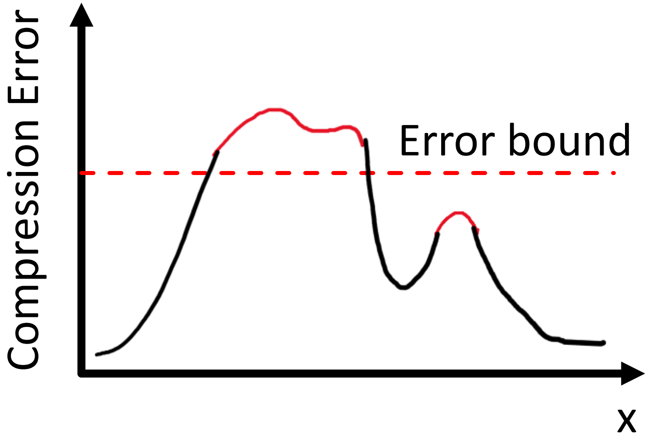
Wavelet Transform



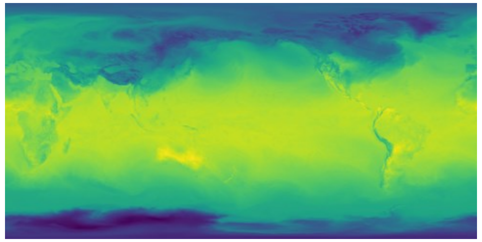
SPIHT

0111000001011100000110000110111001
1101010111001100100001010110010011
1110011001010111111100110111111001
1111001110101111001011010101011100
100001010010101111111100

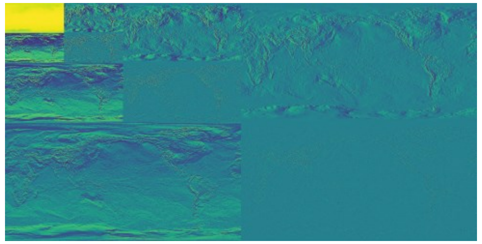
0111000001011100000110000
110111001110101011100110**0**
1000010101100100111100110
0101011111111001101111111



Method: Residual Compression Layer



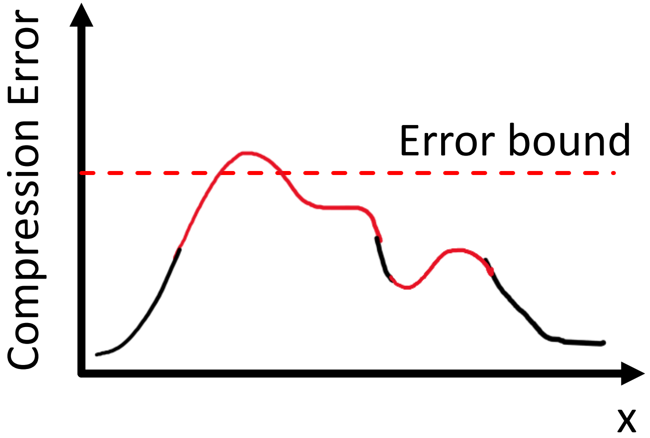
Wavelet Transform



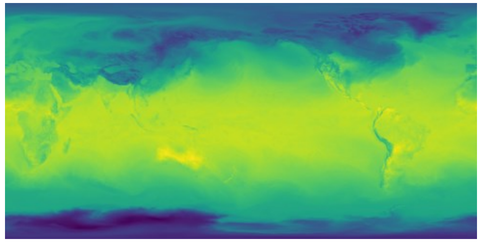
SPIHT

01110000010111000001100001101100
11101010111001100100001010110010
0111100110010101111111110011011111
1001111100111010111100101101010101
11001000010100101011111111100

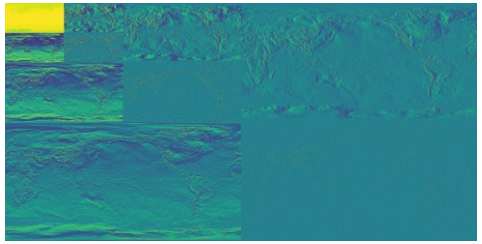
01110000010111000000110000
11011100111010101110011 00
10000101011001001111100110
0101011111111001101111110**0**
0111110011101011110010110
1010101



Method: Residual Compression Layer



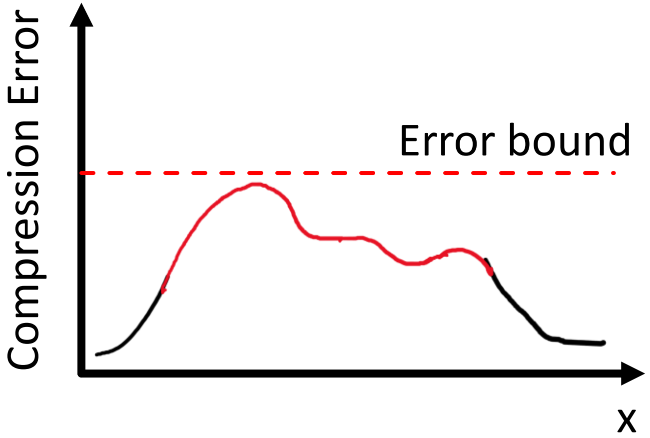
Wavelet Transform



SPIHT

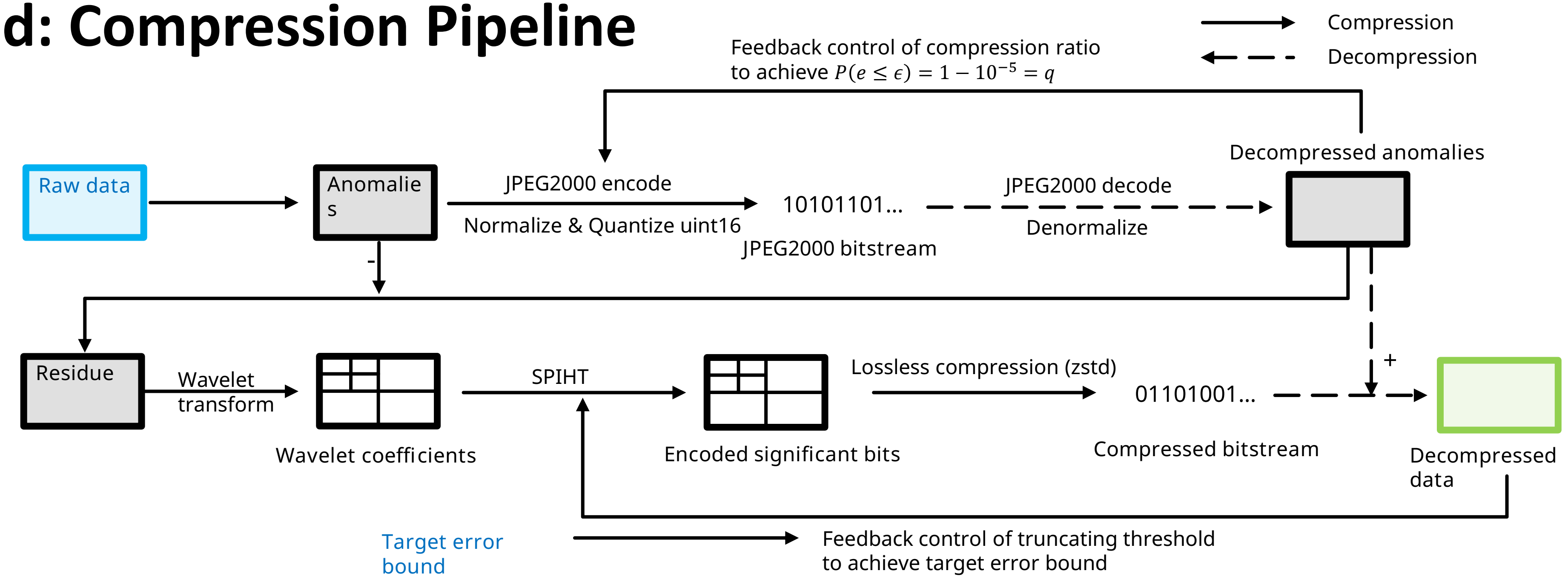
011100000101110000011000011011100
111010101110011001000010101100100
111100110010101111111100110111111
001111100111010111100101101010101
11001000010100101011111111100

0111000001011100000110000
1101110011101010111001100
1000010101100100111100110
0101011111111001101111110
0111110011101011110010110
1010101**110010000101001010**
1111



OK, error bound satisfied, store current sequence.

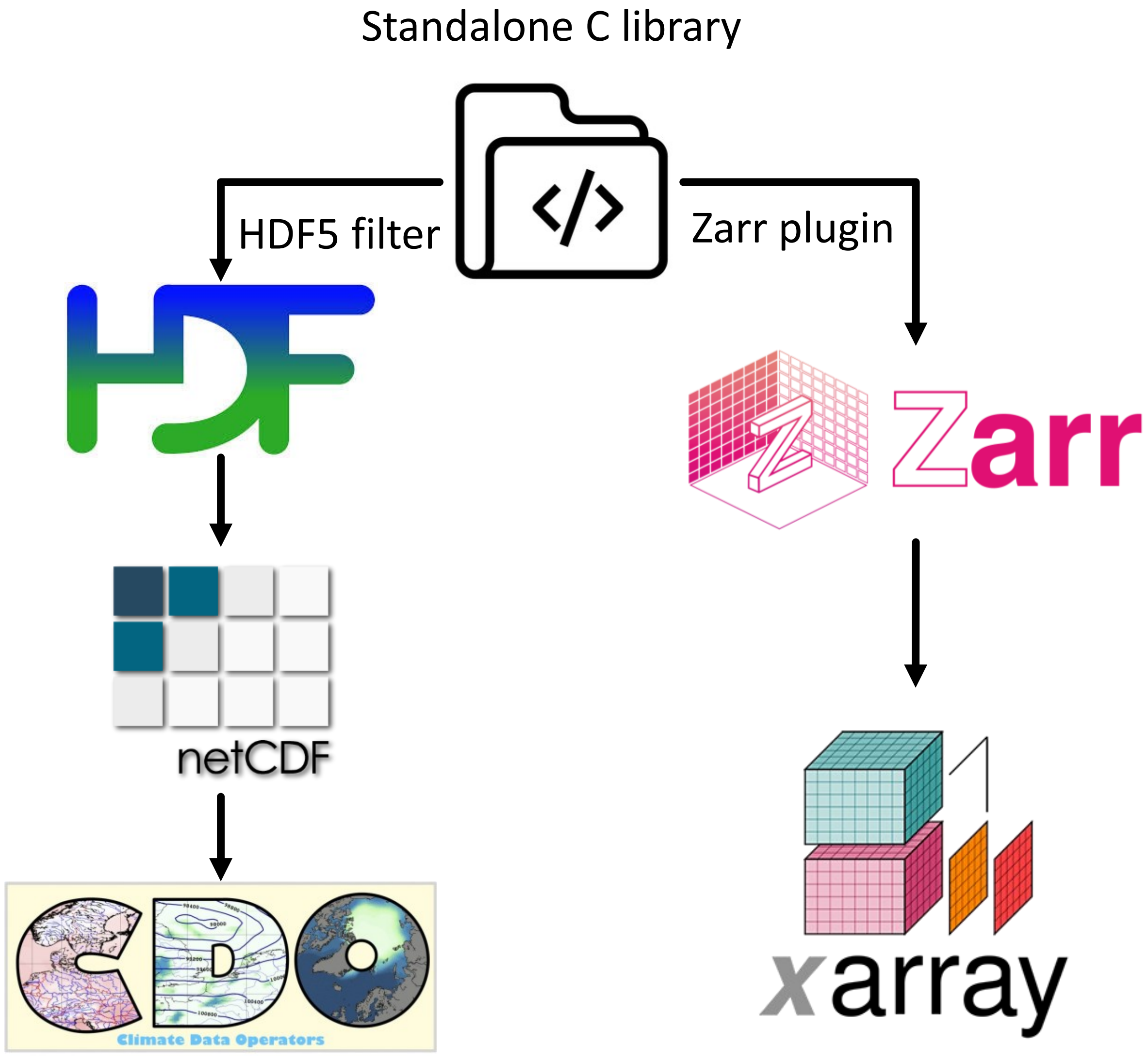
Method: Compression Pipeline



- **For every 2D slices of data e.g. with shape (nlon, nlat)**
  - compress bulk part with JPEG2K
  - feedback loop controlling JPEG2K compression ratio so that error target > $q_{(1\text{-}\mathbf{1e\text{-}5})}$(JP2K compression error)
    only less than 1e-5 fraction of the compressor error is higher than target
  - Encode residue using wavelet transform
  - Encode wavelet coeffs using SPIHT coding
  - feedback loop to find the shortest truncate point of SPIHT sequence so that error < error target

Method: Integration with existing toolchains

Repo: https://github.com/spcl/ebcc



Implementation

- Pure C (OpenJPEG + SPIHT + HDF5 filter)
- Efficient SPIHT implementation

Compression with oneliner

- HDF5/netCDF: filter API
- Zarr: numcodecs compression plugin
- CDO

```
cdo --filter FILTER_ID, `python filter_wrapper.py --base_cr=1000
 --height=721 --width=1440` copy temperature.nc comp.nc
```

- Xarray

```
ds.to_zarr("comp.zarr", encoding={"foo": {"compressor":
 J2KFilter(JP2SPWV_FilterOpts(base_cr=1000,height=721,
 width=1440).hdf_filter_opts)}})
```

Transparent decompression

HDF5_PLUGIN_PATH=<path/to/filter>

- HDF5: h5dump -d /temperature comp.nc
- netCDF: ncdump -v temperature comp.nc
- CDO: cdo info comp.nc
- Zarr: zarr.open('comp.zarr', mode='r')
- Xarray: xarray.open_zarr('comp.zarr')

Icons are taken from https://www.hdfgroup.org/solutions/hdf5 , https://www.unidata.ucar.edu/software/netcdf , https://code.mpimet.mpg.de/projects/cdo , https://zarr.readthedocs.io/en/stable , https://xarray.pydata.org

Benchmarks

Error Statistics

- Data: 12 variables, 37 pressure levels, 2 timesteps
- Error metrics: SSIM, histogram, spectrum
- Compression and decompression throughput

Case Study

- Visual inspection at Tropical Cyclones
- Directly compressed u-wind
- Derived divergence from compressed wind

Error Target: Range relative max error = (max error / (max – min))

Compressors: EBCC, SZ3, SZ, SPERR

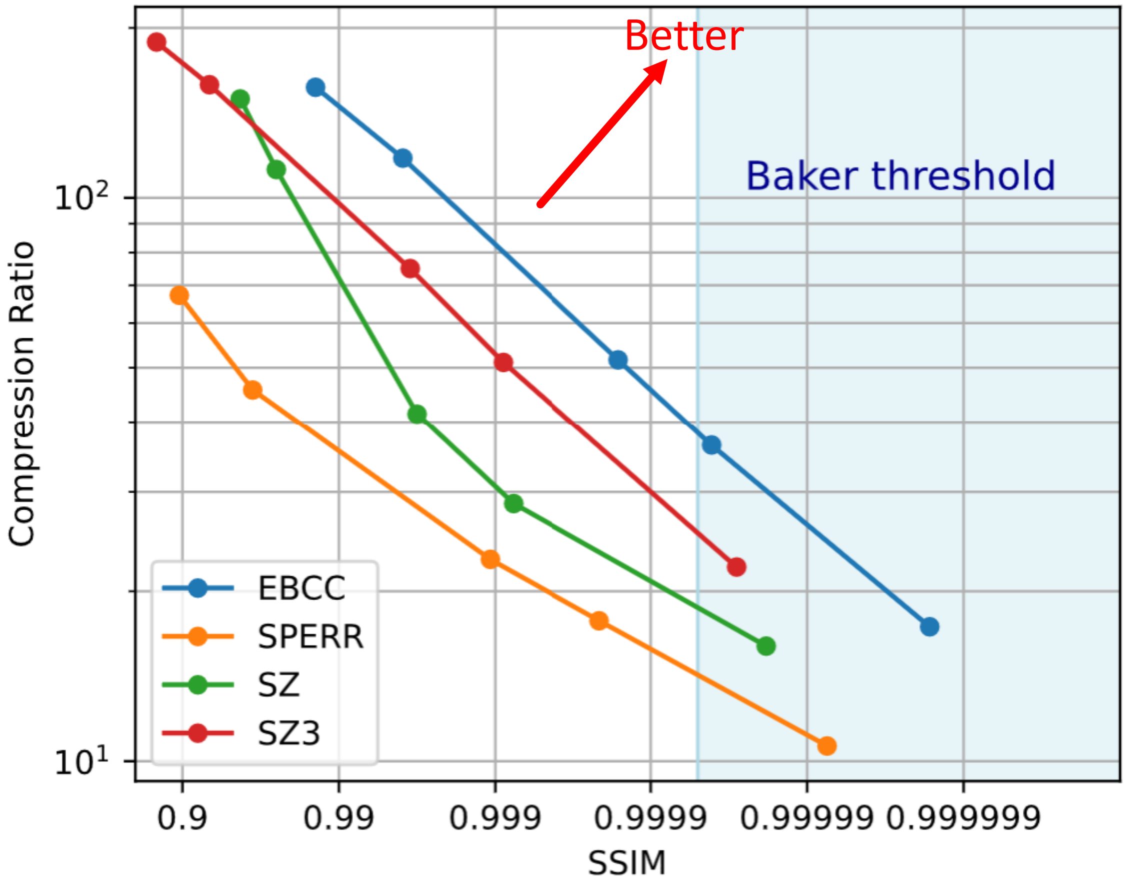
SZ3: Liang, X., Zhao, K., Di, S., Li, S., Underwood, R., Gok, A.M., Tian, J., Deng, J., Calhoun, J.C., Tao, D. and Chen, Z., 2022. Sz3: A modular framework for composing prediction-based error-bounded lossy compressors. *IEEE Transactions on Big Data*, *9*(2), pp.485-498.

SZ: Di, S. and Cappello, F., 2016, May. Fast error-bounded lossy HPC data compression with SZ. In *2016 ieee international parallel and distributed processing symposium (ipdps)* (pp. 730-739). IEEE.

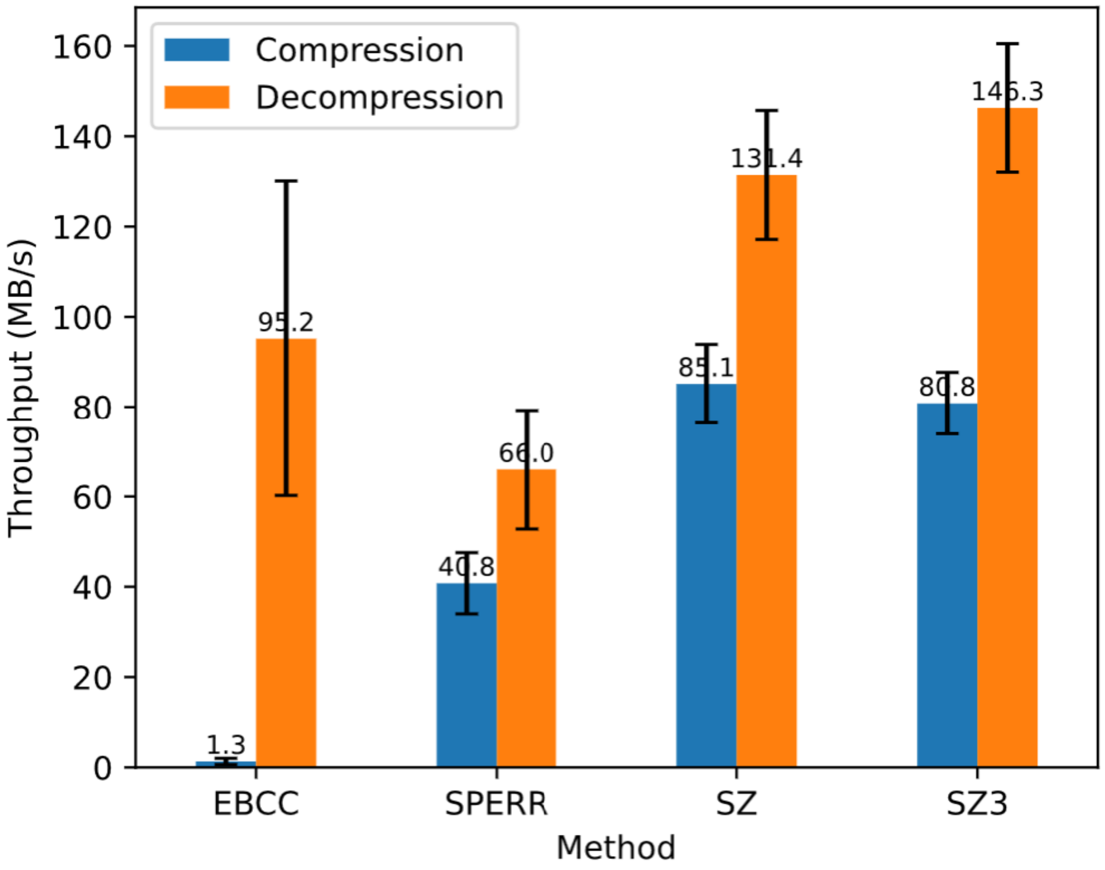
SPERR: Li, S., Lindstrom, P. and Clyne, J., 2023, May. Lossy scientific data compression with sperr. In *2023 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* (pp. 1007-1017). IEEE.

Benchmark: Basic Statistics

Rate-Distortion

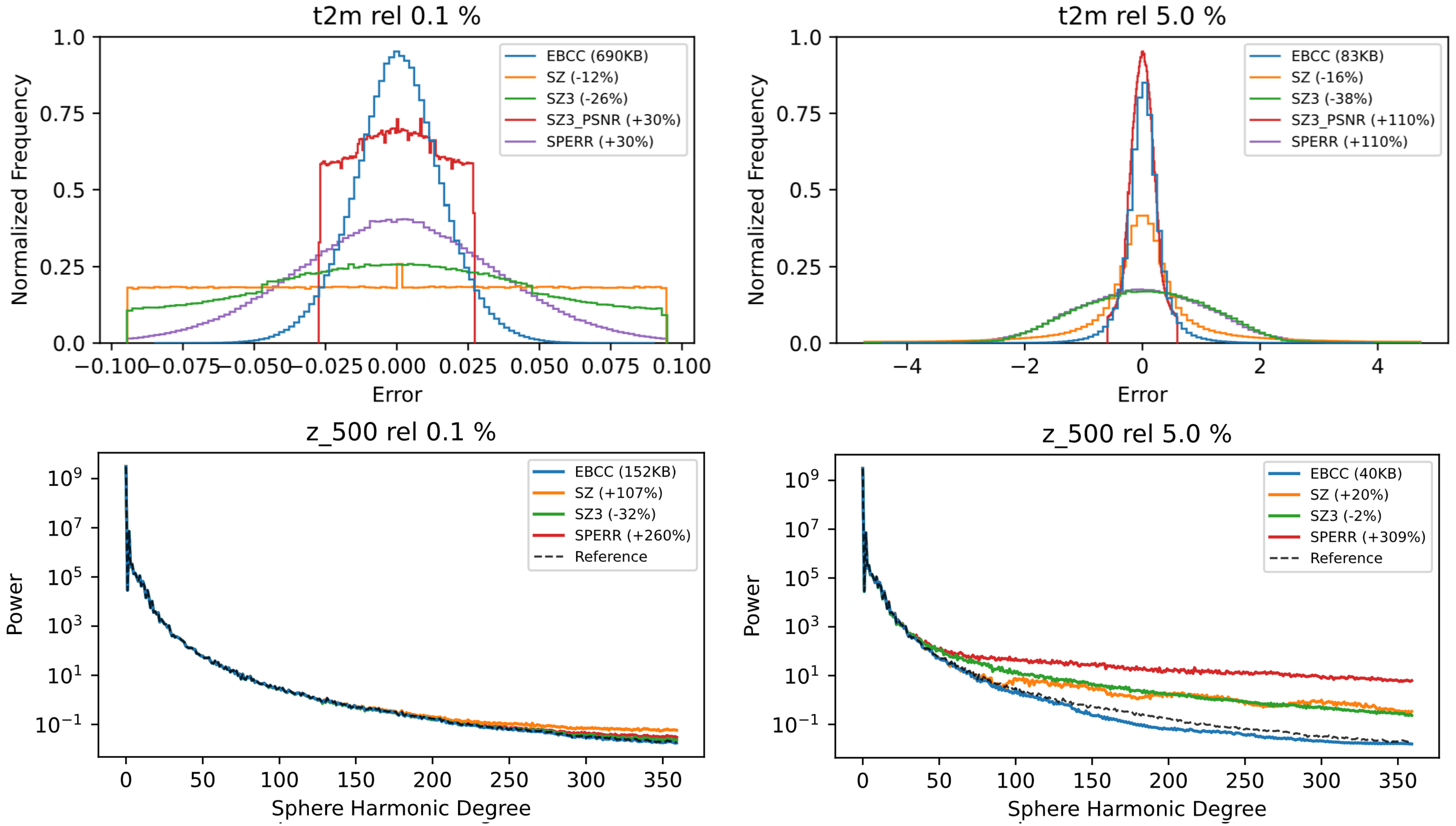


Throughput

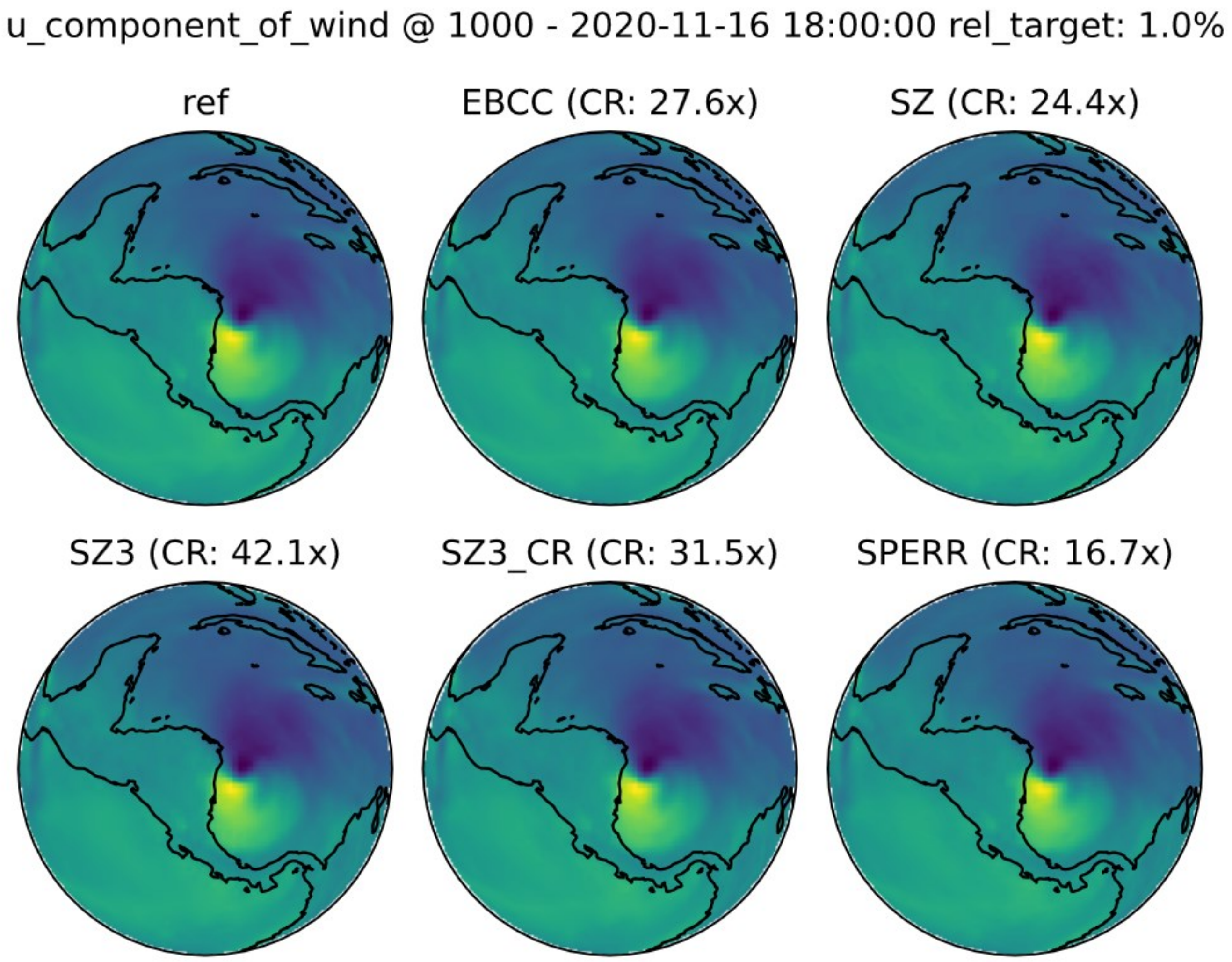


Baker, A.H., Hammerling, D.M. and Turton, T.L., 2019, June. Evaluating image quality measures to assess the impact of lossy data compression applied to climate simulation data. In Computer Graphics Forum (Vol. 38, No. 3, pp. 517-528).

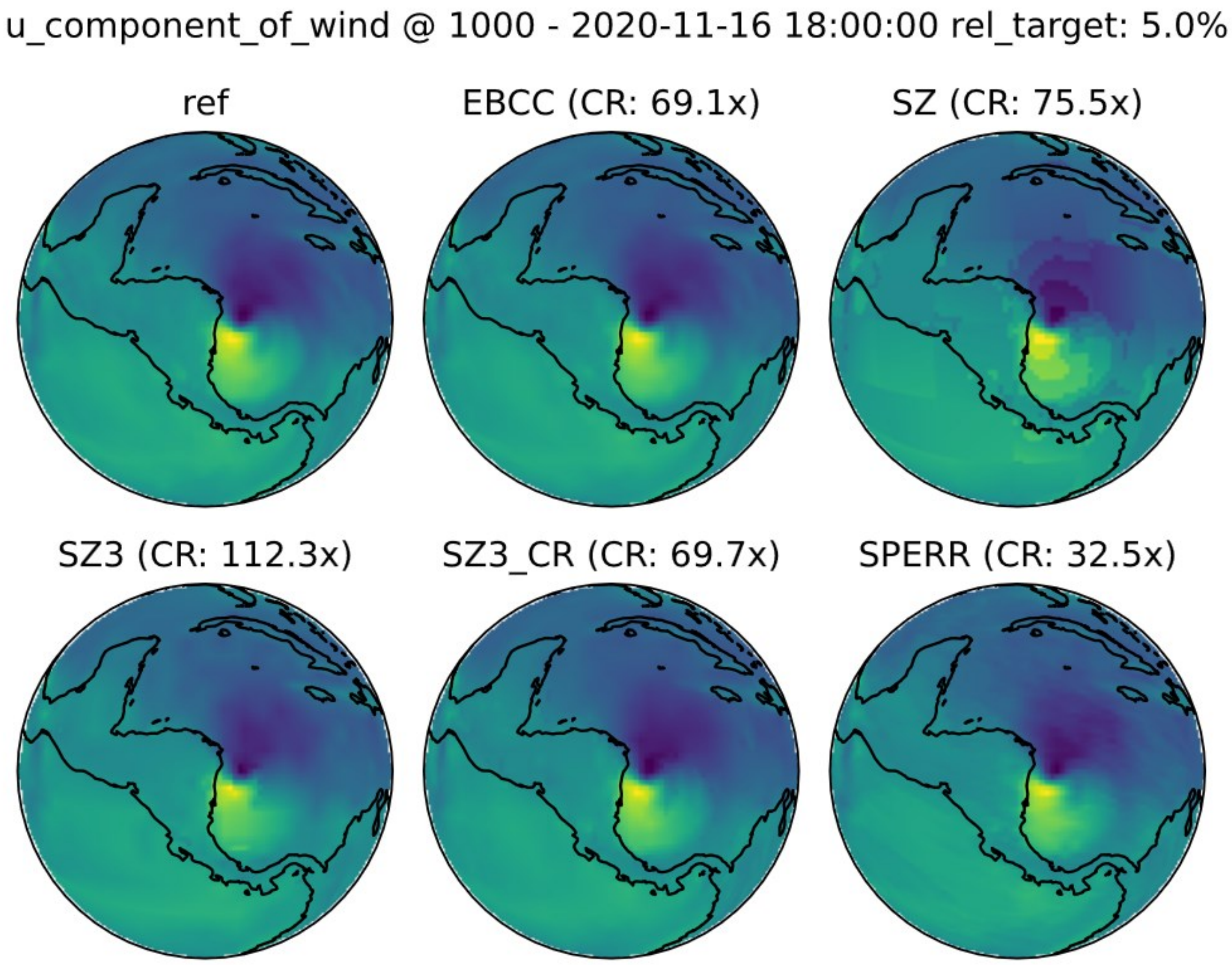
Benchmark: Basic Statistics



Benchmark: Case study

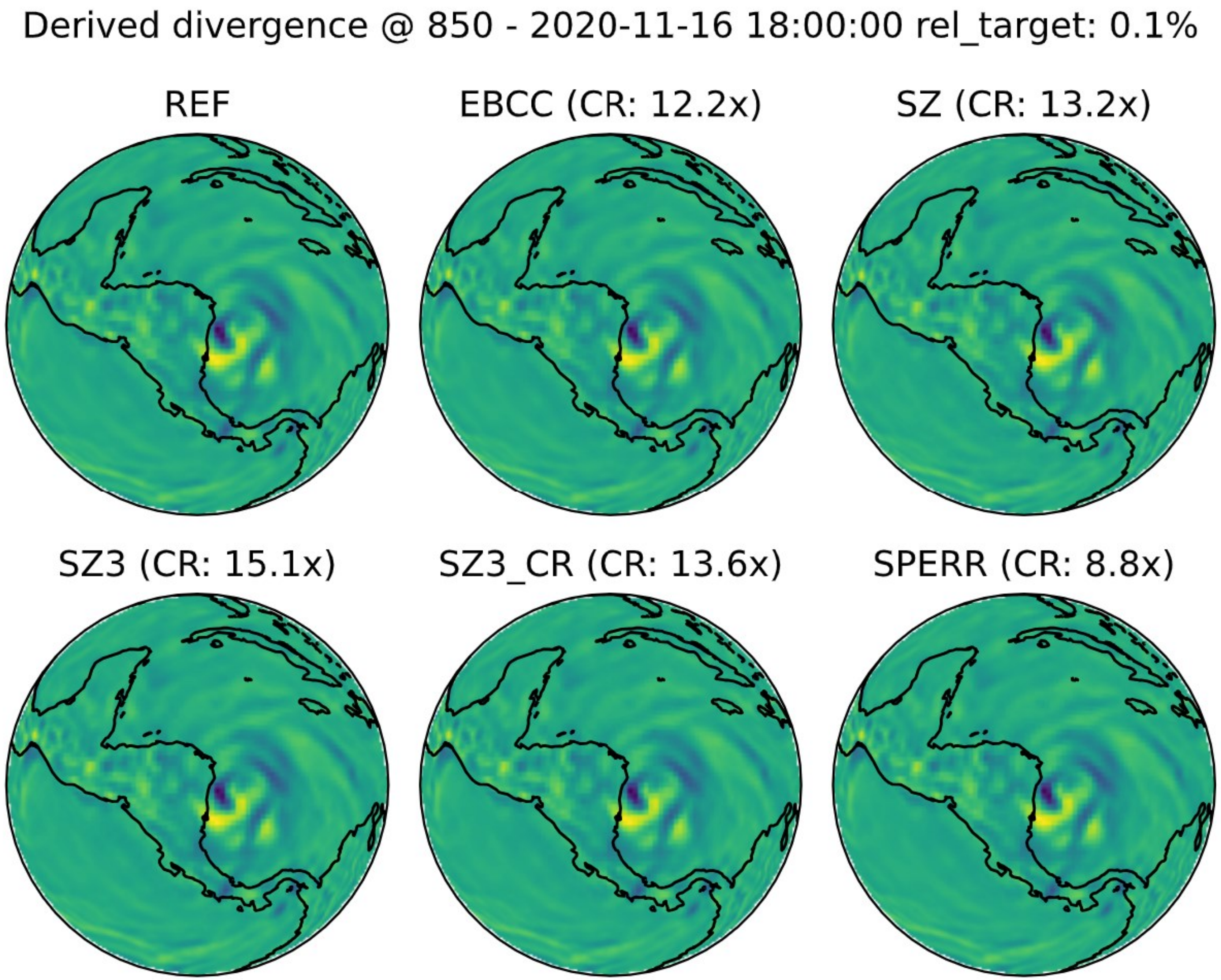


Benchmark: Case study



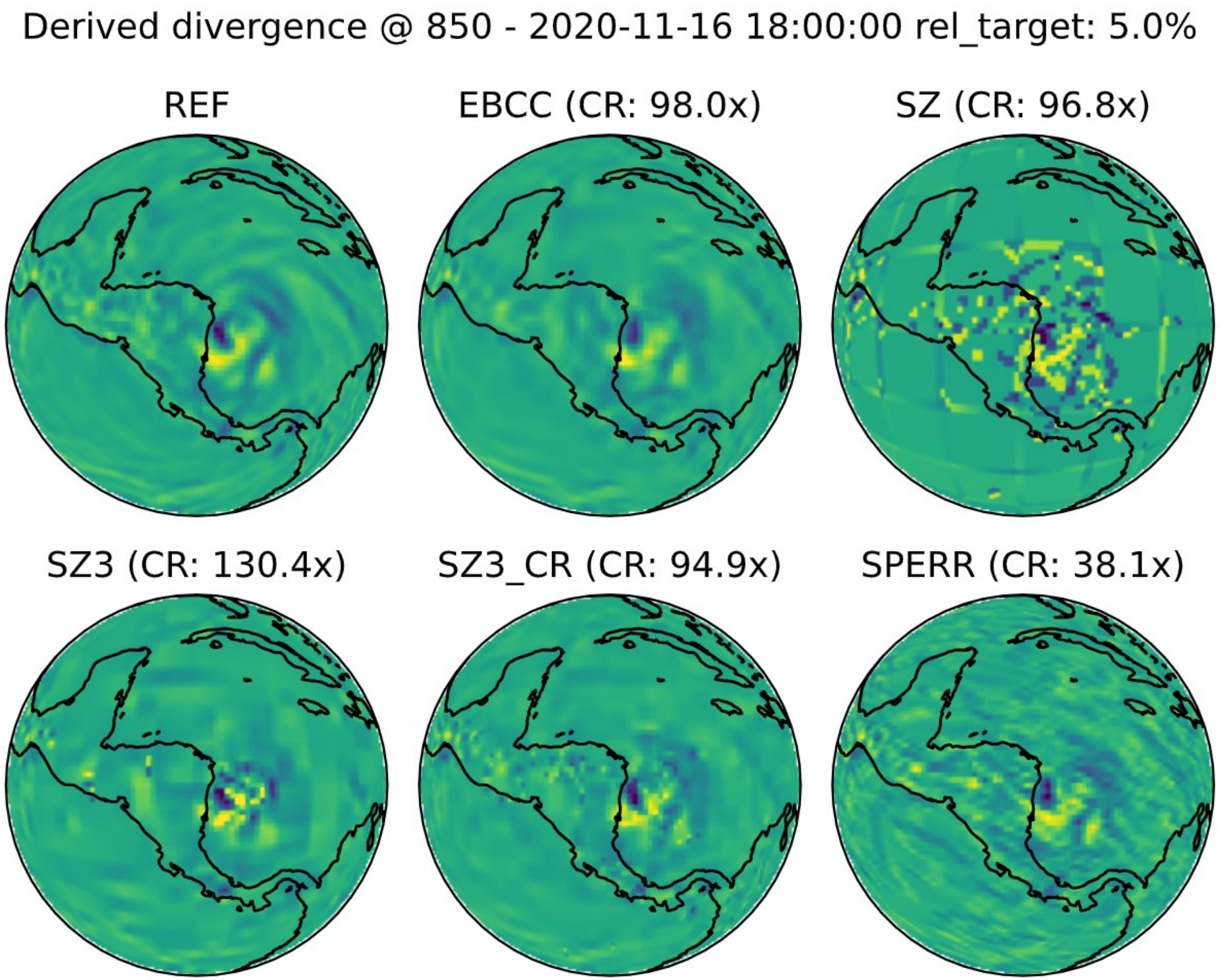
Benchmark: Case study on derived variables

Divergence of horizontal winds (du/dx + dv/dy)



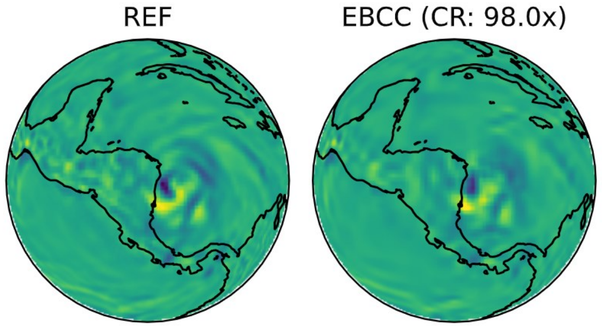
Benchmark: Case study on derived variables

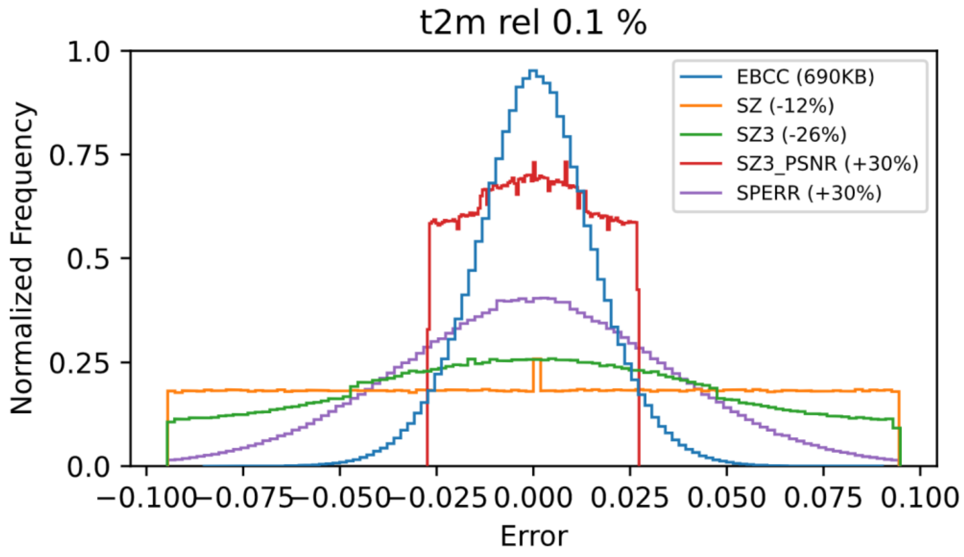
Divergence of horizontal winds (du/dx + dv/dy)

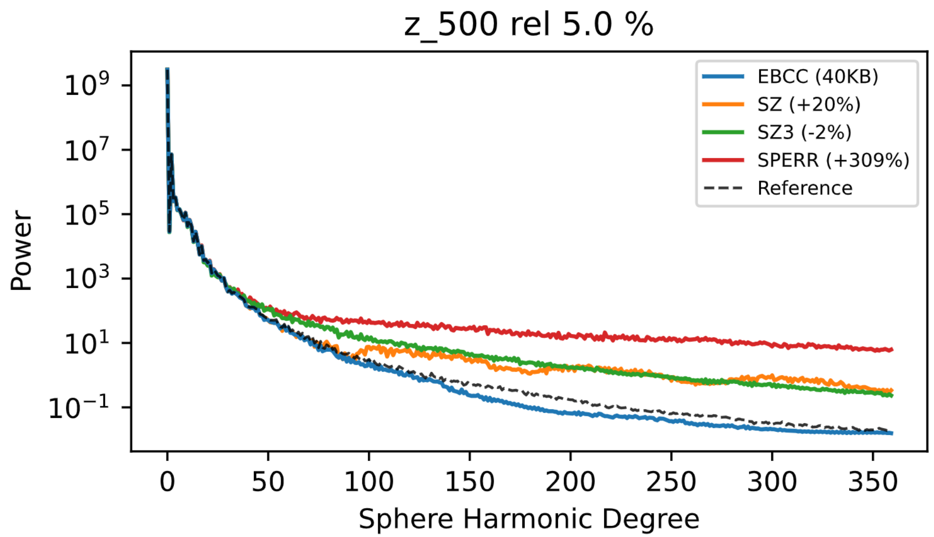


Summary

- **EBCC works well at both high accuracy (10x CR) – high compression ratio (100x) regime**
- **Higher SSIM at every CR**
- **High accuracy regime**
  - Error concentrated more towards 0
  - Errors are normal distributed instead of uniform distributed within error bound
  - Stdev ~ 0.1 error bound
  - Match spectrum to the high freq end
- **High compression ratio regime**
  - Less artifact introduced
  - Smoothen the data -> less power on high freq
  - Smoothen gradients instead of adding artifacts







Benchmarks

Error Statistics

- Data:12 variables, 37 pressure levels, 2 timesteps
- Error metrics: SSIM, histogram, spectrum
- Compression and decompression throughput

Case Study

- Visual inspection at Tropical Cyclones
- Directly compressed u-wind
- Derived divergence from compressed wind

Energy balance

- Data: 10 variables, 37 pressure levels, 3 months
- Closure of atmosphere energy budget
- Moist static energy framework
- Test deviation of zonal mean residue

Trajectory Sim.

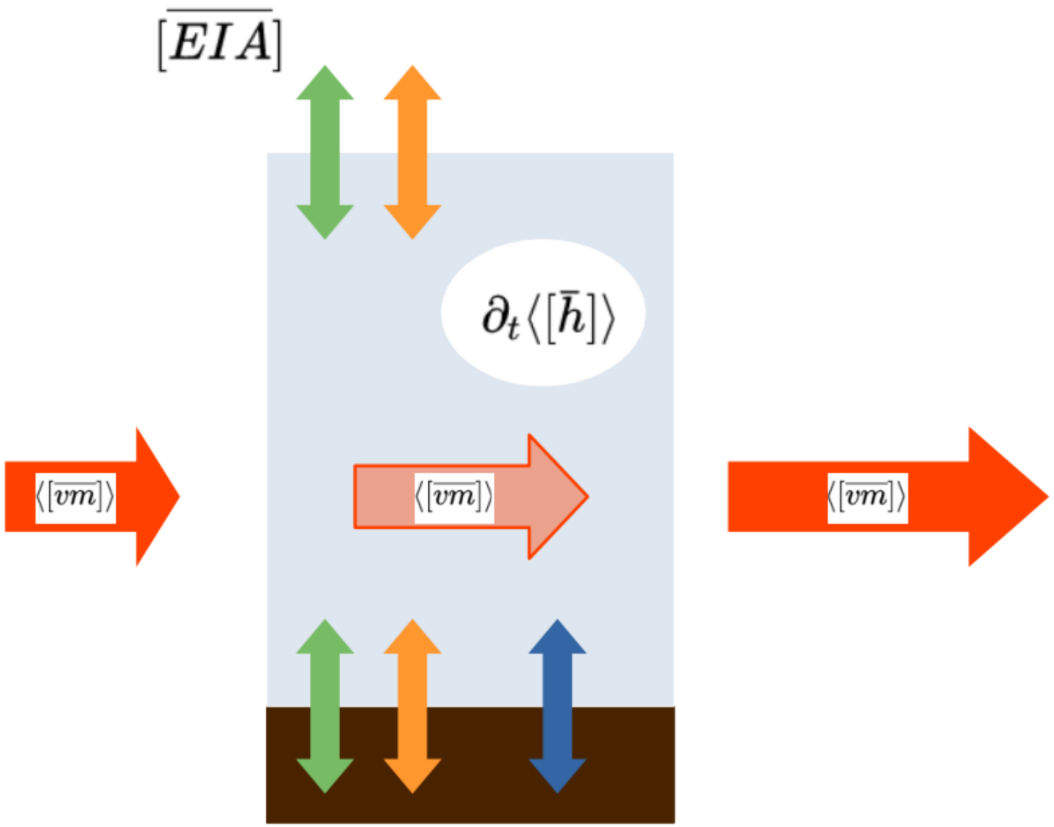
- Data: 3D wind speed, 1 week
- Trajectory simulation with compressed wind
- Test deviation of trajectories and particle distributions

Error Target: Range relative max error = (max error / (max – min))

Compressors: EBCC, SZ3

Benchmark: Closure of Atmospheric Energy Budget

$$\boldsymbol{F_e} = \partial_y\langle[\overline{vm}]\rangle - \left([\overline{EIA}] - \partial_t\left\langle[\bar{h}]\right\rangle\right) \approx 0$$

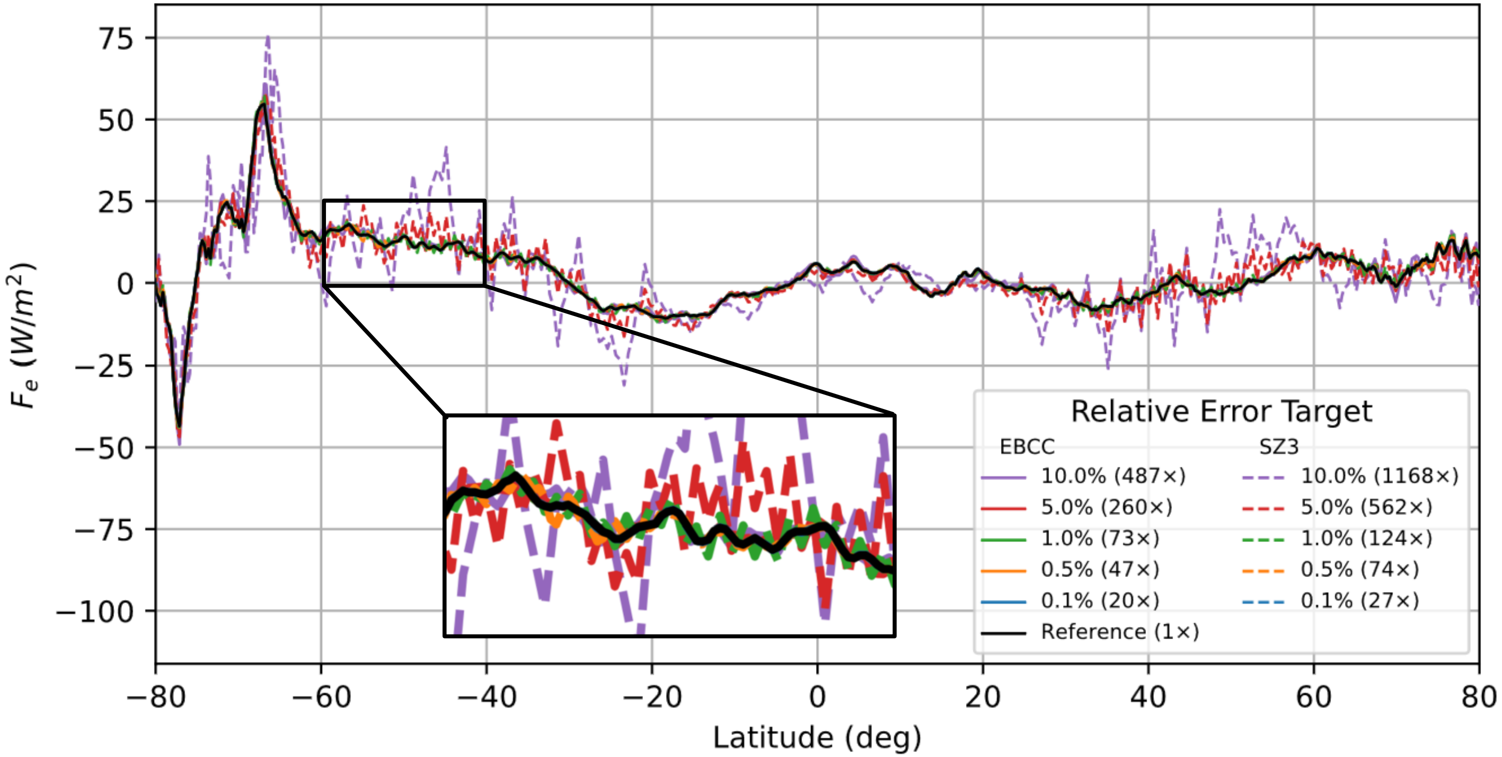


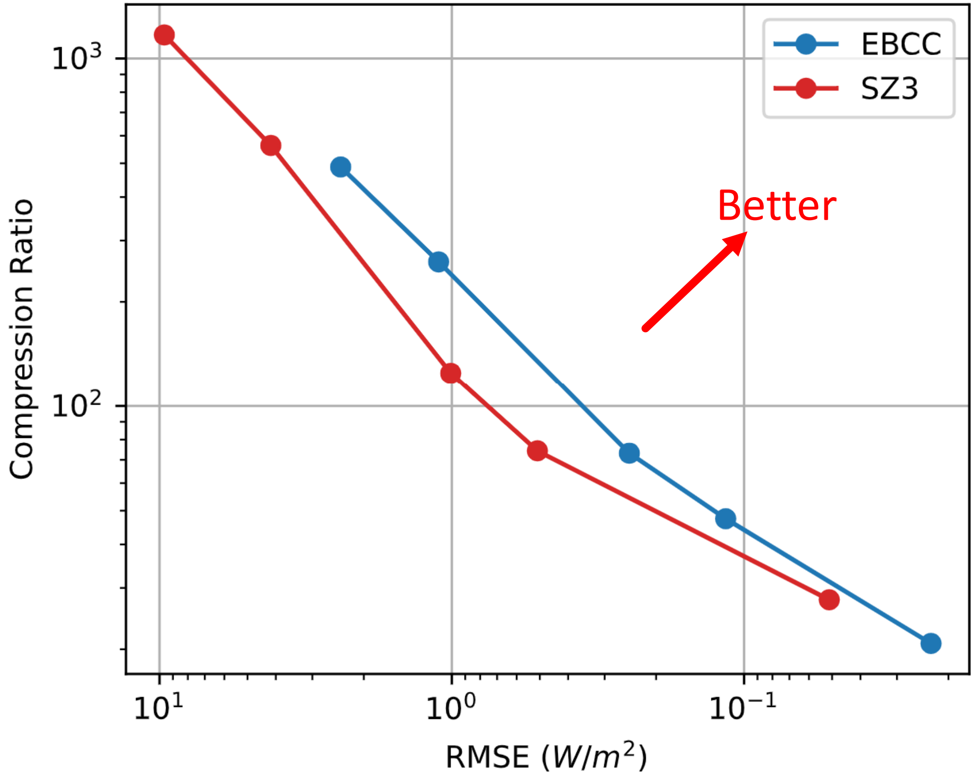
v	meridional wind velocity
$m = c_pT + Lq + \Phi$	MSE
$h = c_pT + Lq$	thermal energy
EIA	Energy input atmosphere from radiative and surface fluxes
$\partial_y(\cdot) \equiv \partial_\phi\{\cos\phi(\cdot)\}/(a\cos\phi)$	meridional divergence
$\langle\cdot\rangle$	mass-weighted vertical integration
$[\cdot]$	zonal mean
$\bar{x}$	temporal mean

- **Compressed variables:** v-wind, temperature, specific humidity, geopotential, net radiative & heat fluxes
- **Time:** 2016/08 – 2016/10, per hour
- **Data size:** 1.2TB

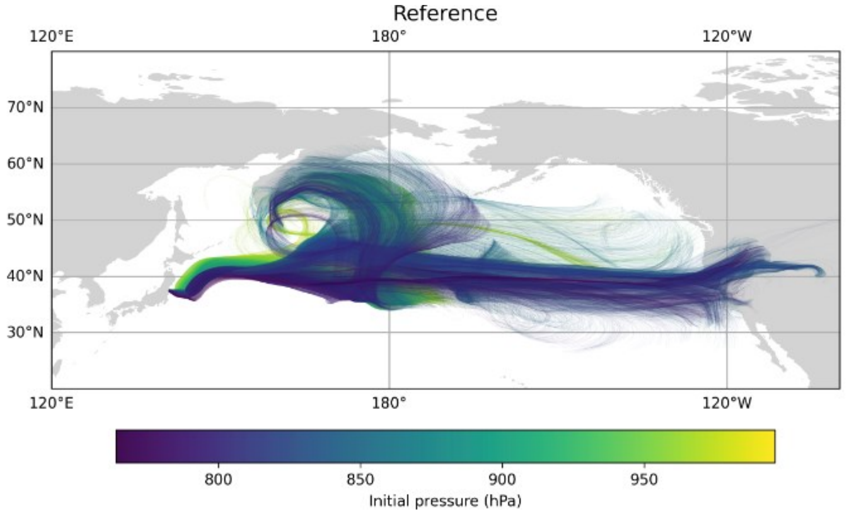
Benchmark: Closure of Atmospheric Energy Budget

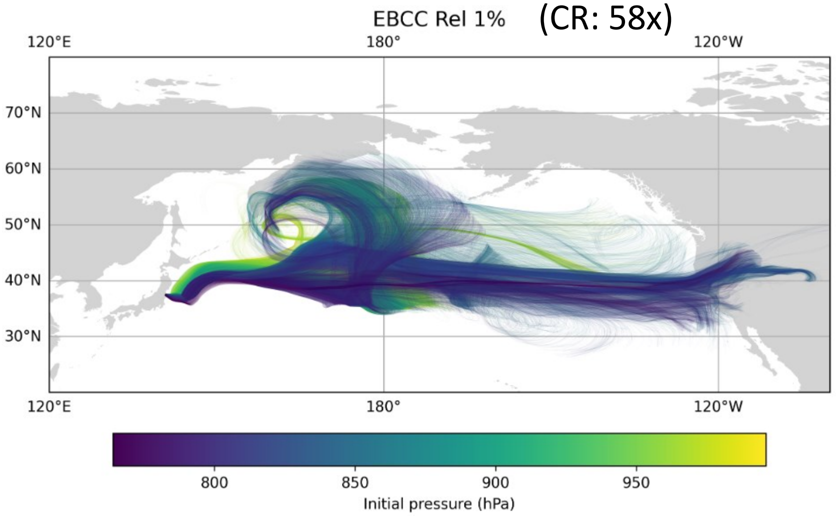
$$F_e = \partial_y \langle [\overline{vm}] \rangle - \left([\overline{EIA}] - \partial_t \left\langle [\bar{h}] \right\rangle \right) \approx 0$$

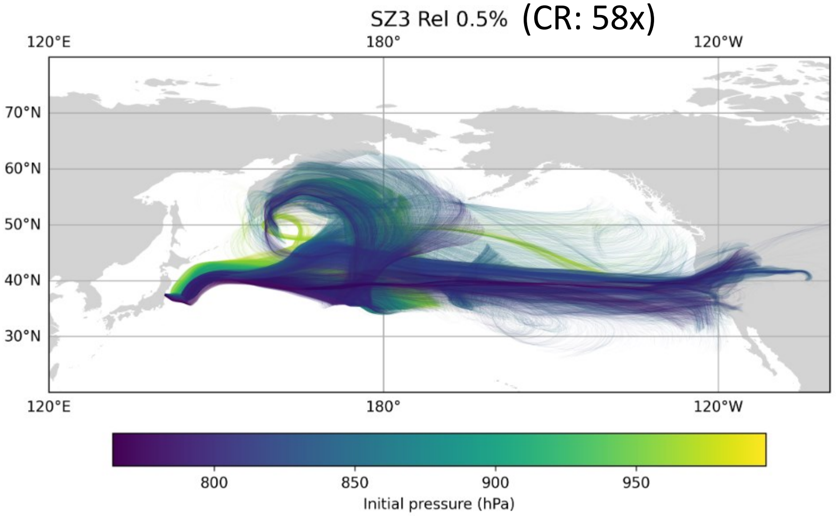


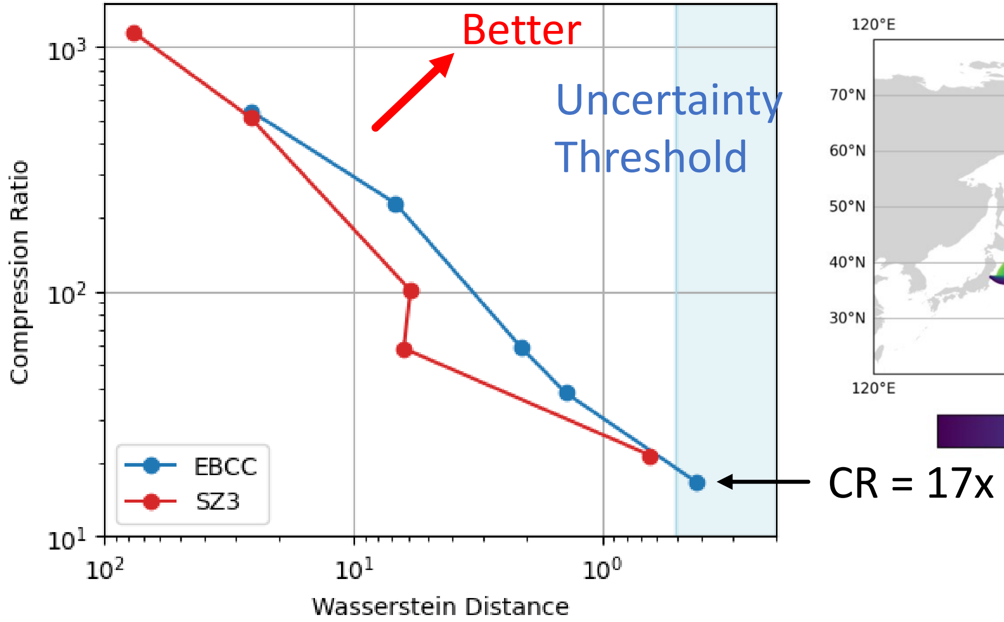


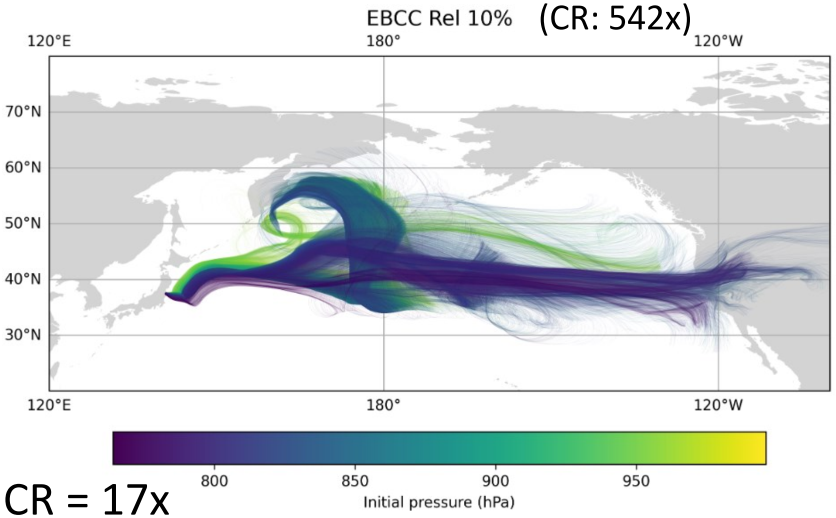
Benchmark: Trajectory simulation on compressed data

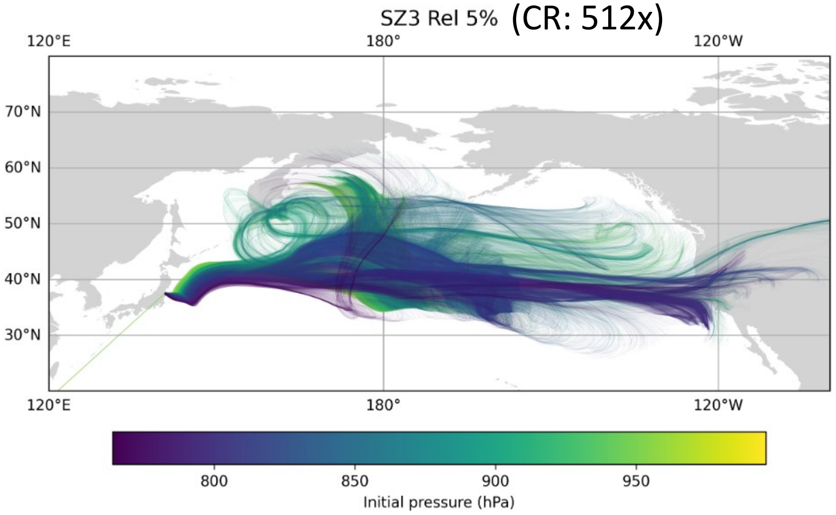










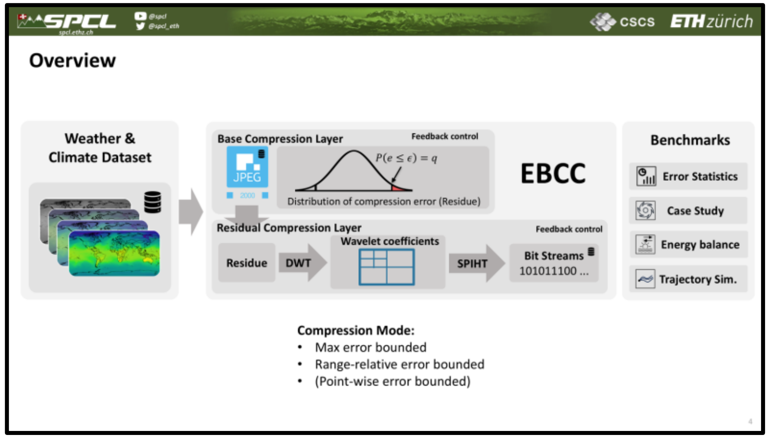


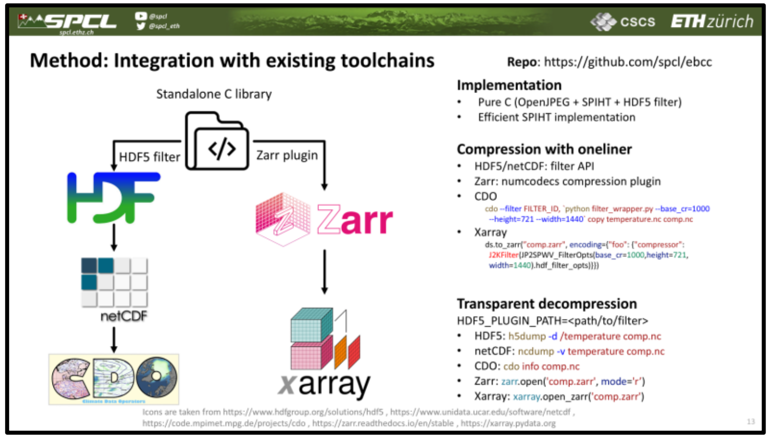
What are the acceptable compression ratios?

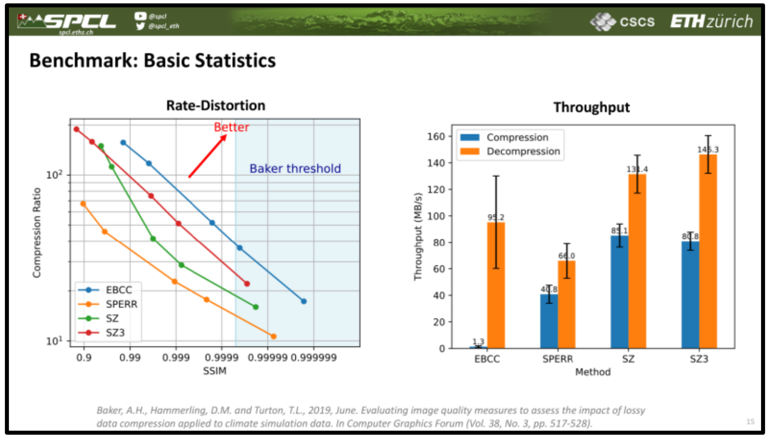
- **Visual inspection (rough), Zonal/global mean**
  - Errors cancel out when calculating mean
  - Range relative max error target: 1% - 5%, Compression ratio: 50x – 300x
- **Visual inspection (accurate): 0.99995 threshold (Baker et al., 2019)**
  - Perceptually identical
  - Range relative max error target: 0.5%, Compression ratio: 40x
- **Calculate derived variables (divergence, vorticity, relative humidity)**
  - Range relative max error target: 0.1% - 1%, Compression ratio: 15x – 50x
- **Perform trajectory simulation**
  - Range relative max error target: 0.1%, Compression ratio: 17x

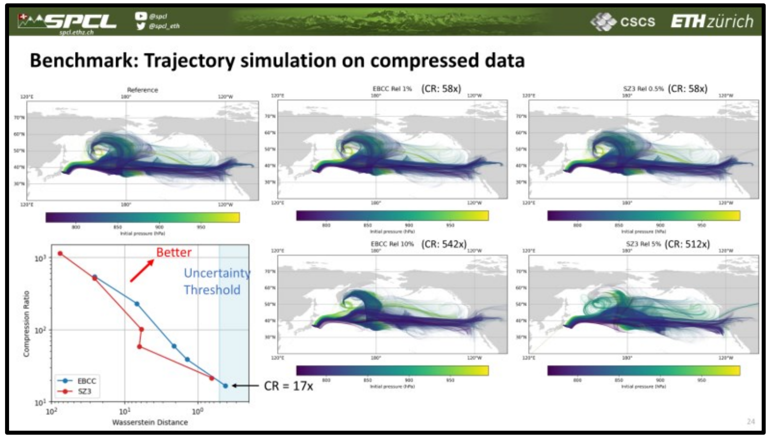
Compression ratios are estimated under 0.25° ERA5

Conclusions









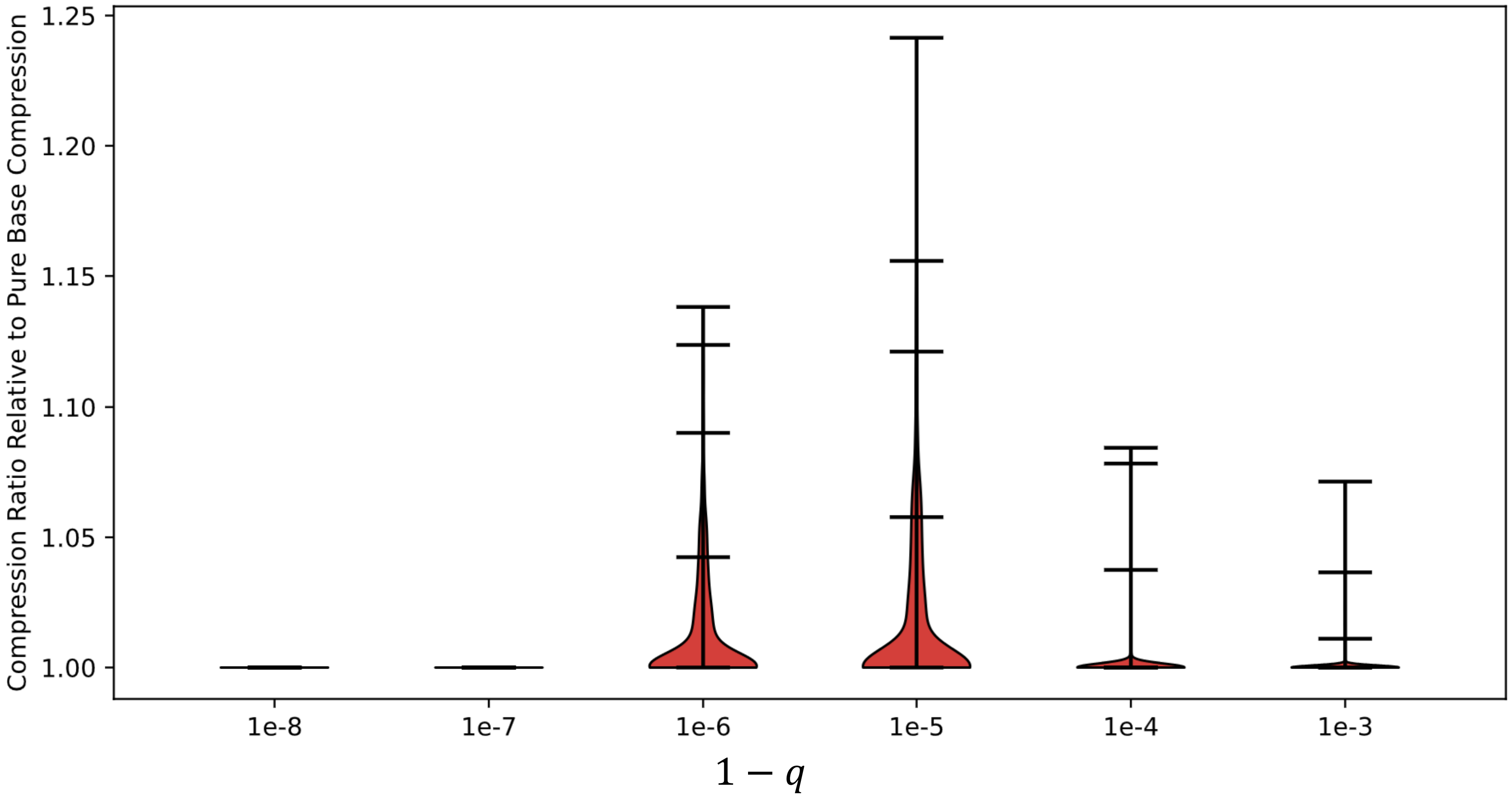
More of SPCL's research:

youtube.com/@spcl	210+ Talks
twitter.com/spcl_eth	1.6K+ Followers
github.com/spcl	5.6K+ Stars

… or spcl.ethz.ch

Github Repo

Ablation study : is EBCC better than error bounded JPEG2000?



EBCC is at least as good as error bounded JPEG2000, at most 25% higher in CR.