HTTPS://GITHUB.COM/FTHPC/CORRELATION_COMPRESSIBILITY

Argonne NATIONAL LABORATORY

STATISTICAL PREDICTION OF LOSSY COMPRESSION RATIOS FOR 3D SCIENTIFIC DATA

DAVID KRASOWSKA
Clemson University

ADVISORS
Julie Bessac
Robert Underwood

COLLABORATORS
Jon Calhoun
Sheng Di
Franck Cappello
Arkaprabha Ganguli

19 July, 2022
Lemont, IL

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

WHY USE COMPRESSION IN HPC?

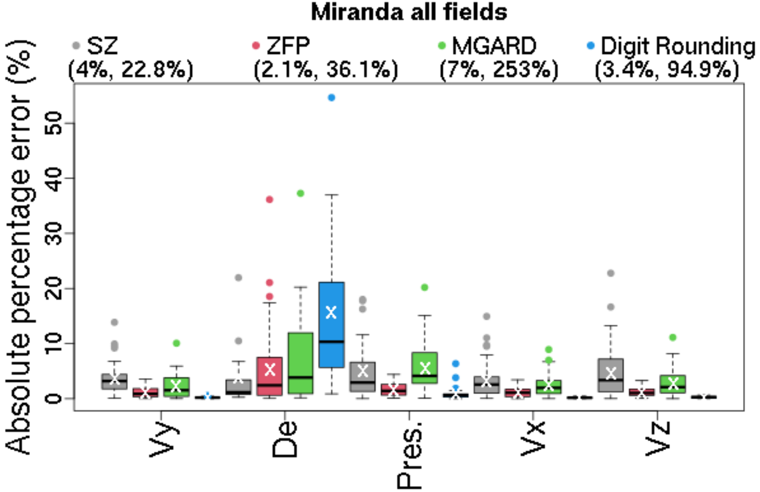
- HPC applications require lots of storage and memory throughput
- Compression allows for larger problem sizes to be ran while accelerating I/O time
- Checkpoint snapshots of an application's state

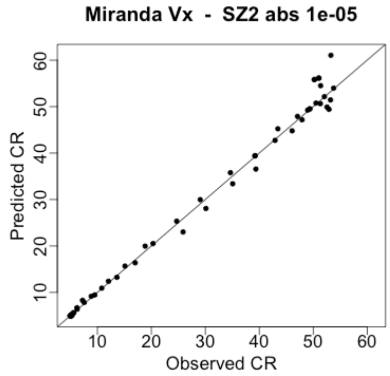
OUR PREVIOUS WORK

CR estimation for 2D datasets

- Statistical predictors with notions of correlation, entropy and lossyness
  - Truncated SVD
  - Quantized entropy

$$\log(\mathrm{CR}) = a + b \times \log(\text{q-ent}) + c \times \log\left(\frac{\text{SVD-trunc}}{\sigma}\right) + d \times \log(\text{q-ent}) \times \log\left(\frac{\text{SVD-trunc}}{\sigma}\right) + \epsilon, \quad (1)$$





EXTENSION TO 3D DATASETS

Completed this summer

- Extend statistical predictors to 3D
- Higher order SVD (HOSVD)

Algorithm 1 High-order SVD on tensor X of order N

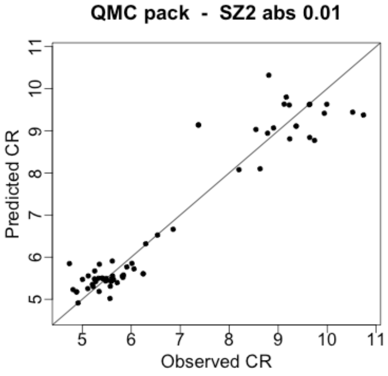
for $i = 1, ..., N$ **do**
  $X_{(i)} \leftarrow$ unfolding(X, mode=i)
  $X_{(i)} \leftarrow U^{(i)} D^{(i)} (V^{(i)})^T$
  $A^{(i)} \leftarrow$ left singular vectors $U^{(i)}$ of $X_{(i)}$
end for

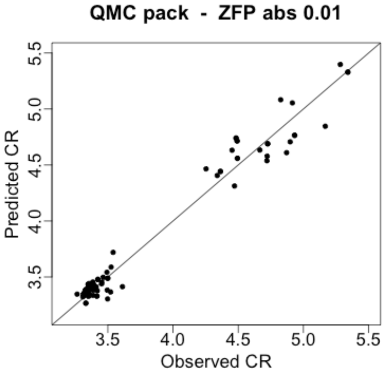
- Used the same regression model used for 2D datasets

RESULTS FOR 3D

- Comparable results to the 2D method counterpart

Compressor	MPE (median percentage error)	10% Quantile	90% Quantile
SZ2	4.5%	3.2%	5.7%
ZFP	1.7%	1.3%	3.5%
MGARD	0.6%	0.4%	1.3%
Bit Grooming	7.4%	5%	9.3%
TTHRESH	24.8	15.7%	27.7%





CONCLUSIONS

- Ability to predict CRs in 3D remains competitive with our method
- Flexible across compressors, error bounds, and datasets

CURRENT WORK

- Currently working on sampling-based approaches to reduce computational costs
  - Generate training samples from blocks of the data
  - Estimate CR using the samples and our predictors

QUESTIONS?

CONTACT: LINKTR.EE/KRASOW
DKRASOW@CLEMSON.EDU