

Langwen Huang, Torsten Hoefler

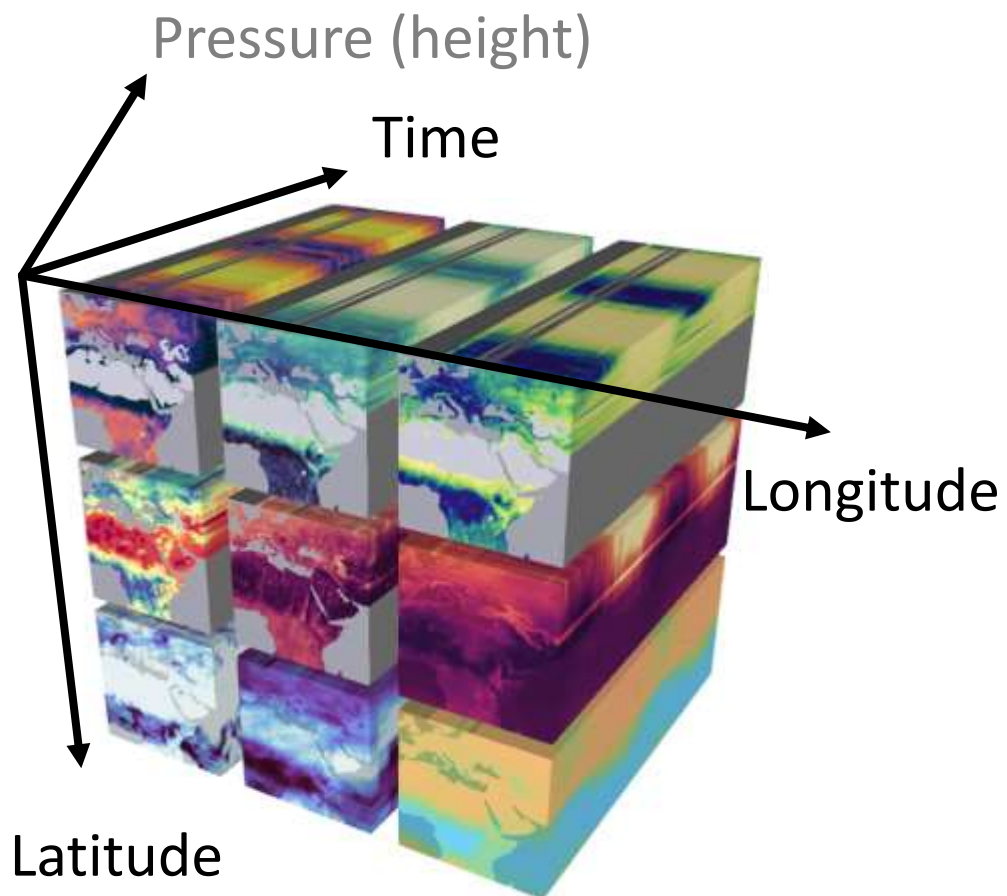
Compressing Multidimensional Weather and Climate Data into Neural Networks



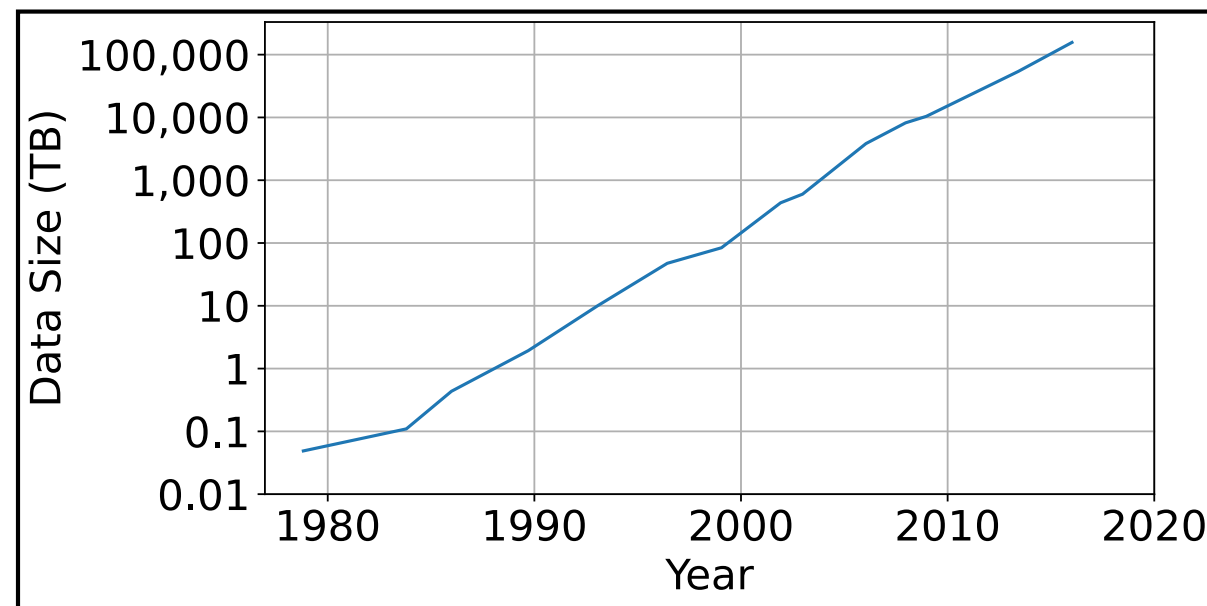
Background: Climate Change



Production and Consumption of Weather & Climate Data



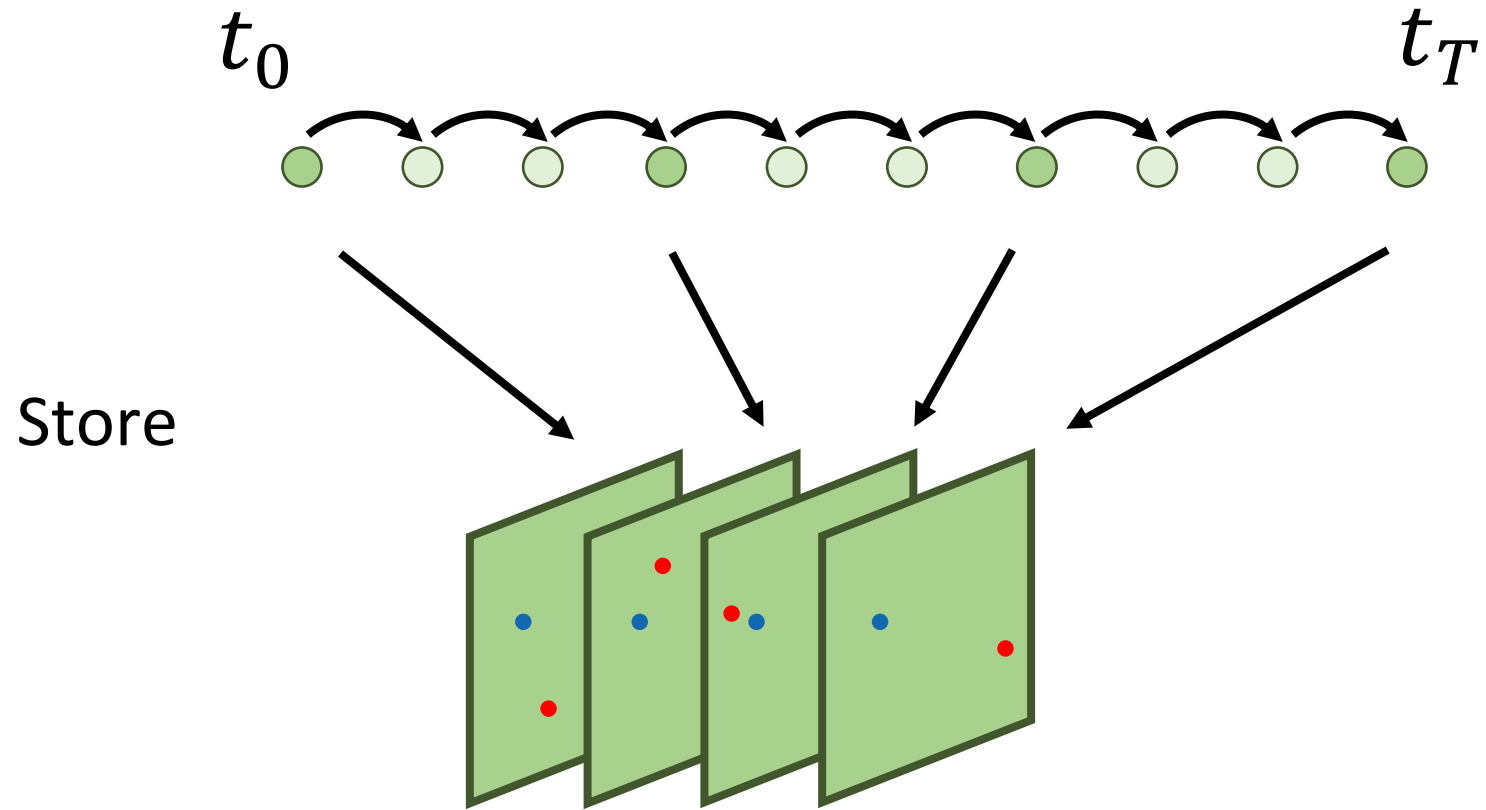
Most weather & climate data are stored as multidimensional arrays.



The data archive is growing **exponentially** in ECMWF! [1]

Production and Consumption of Weather & Climate Data

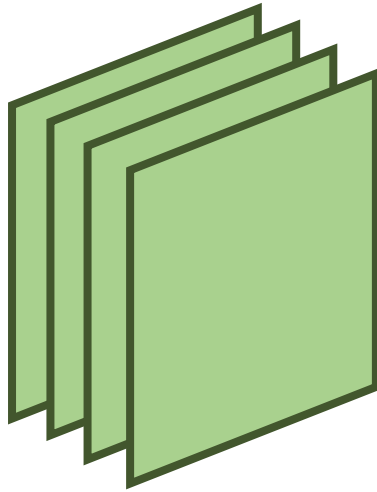
Simulation



Store

Access pattern is often **strided** or even **random**

Compression with a Neural Representation Approach

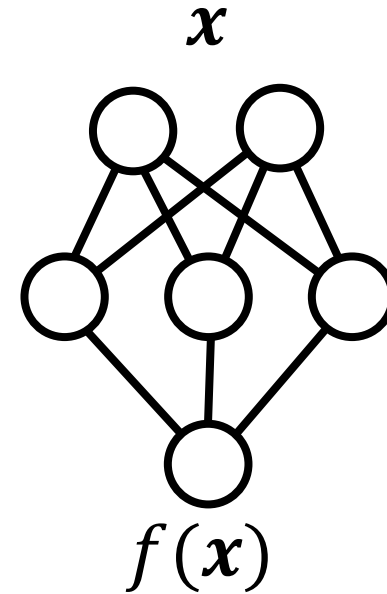


Multidimensional Data

Compress/Train

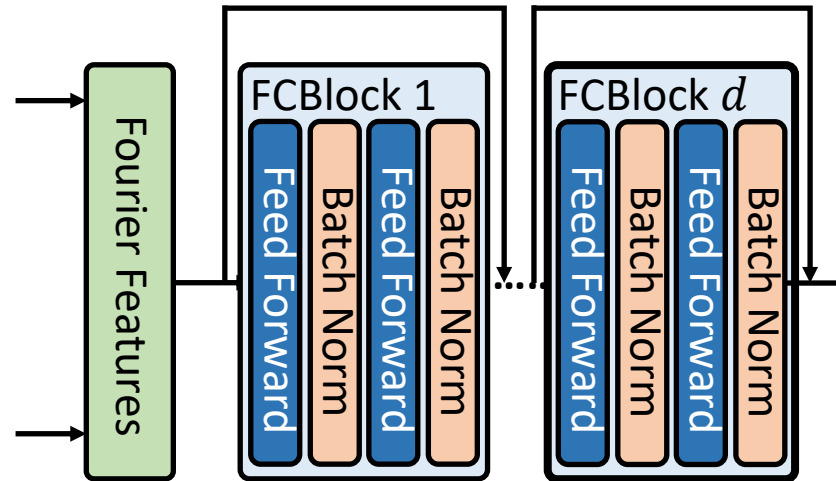


300 x – 3,000 x
 15.6GB → 13.8MB



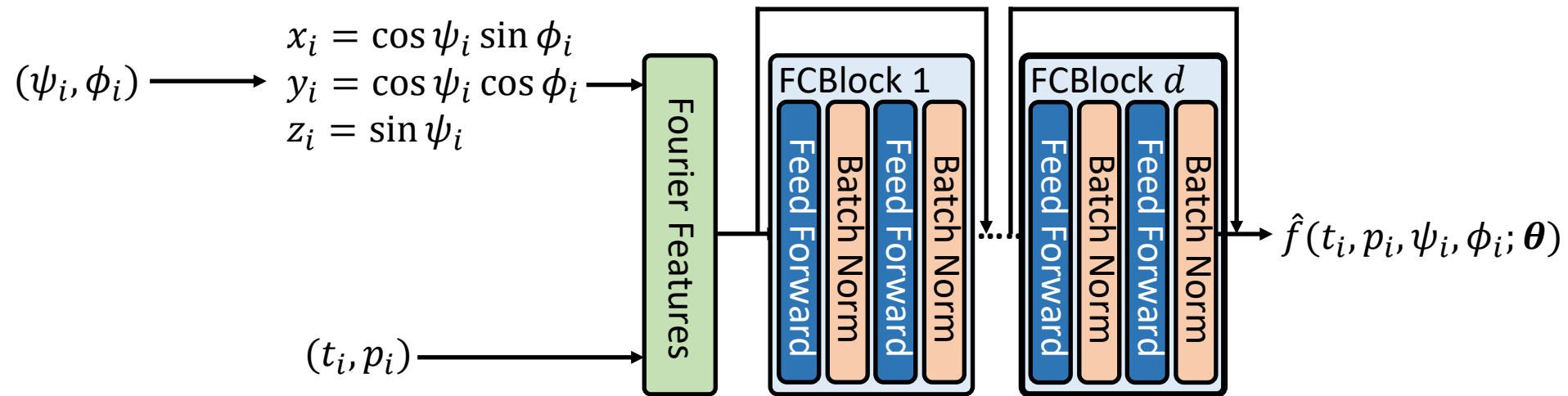
Neural Representation

Neural Network Structure



Neural Network Structure

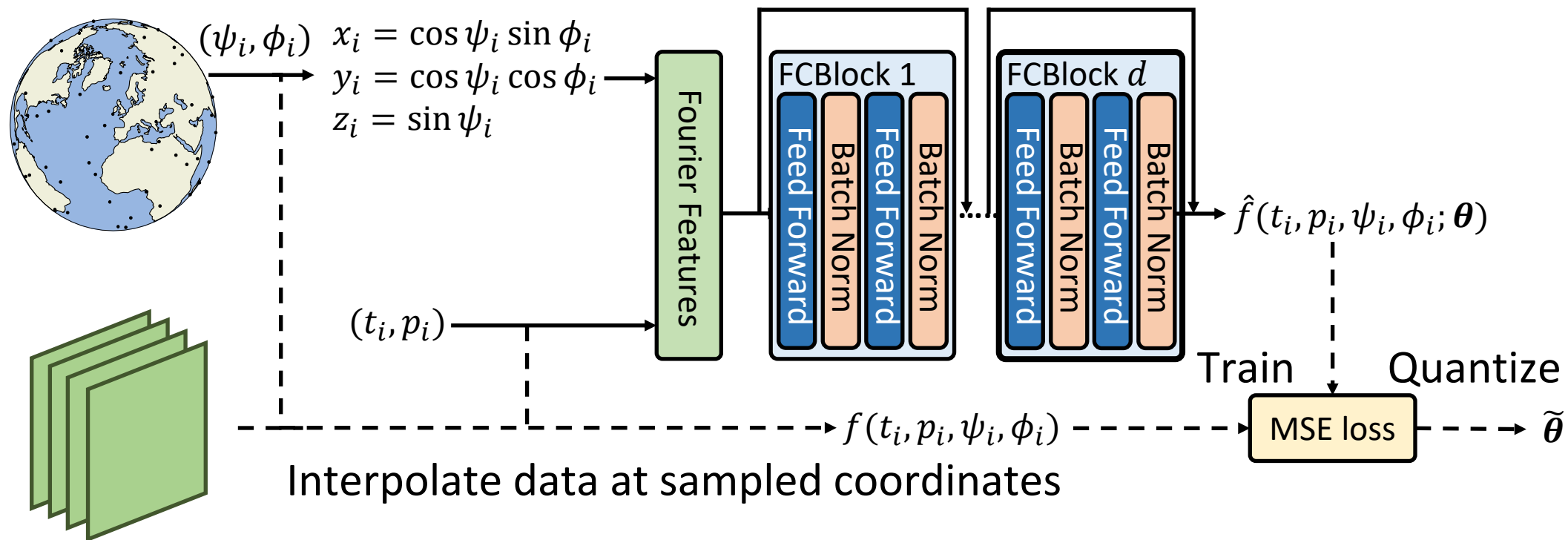
Decompression / Inference















- On-demand decompression
- Fully utilize GPUs

Neural Network Structure

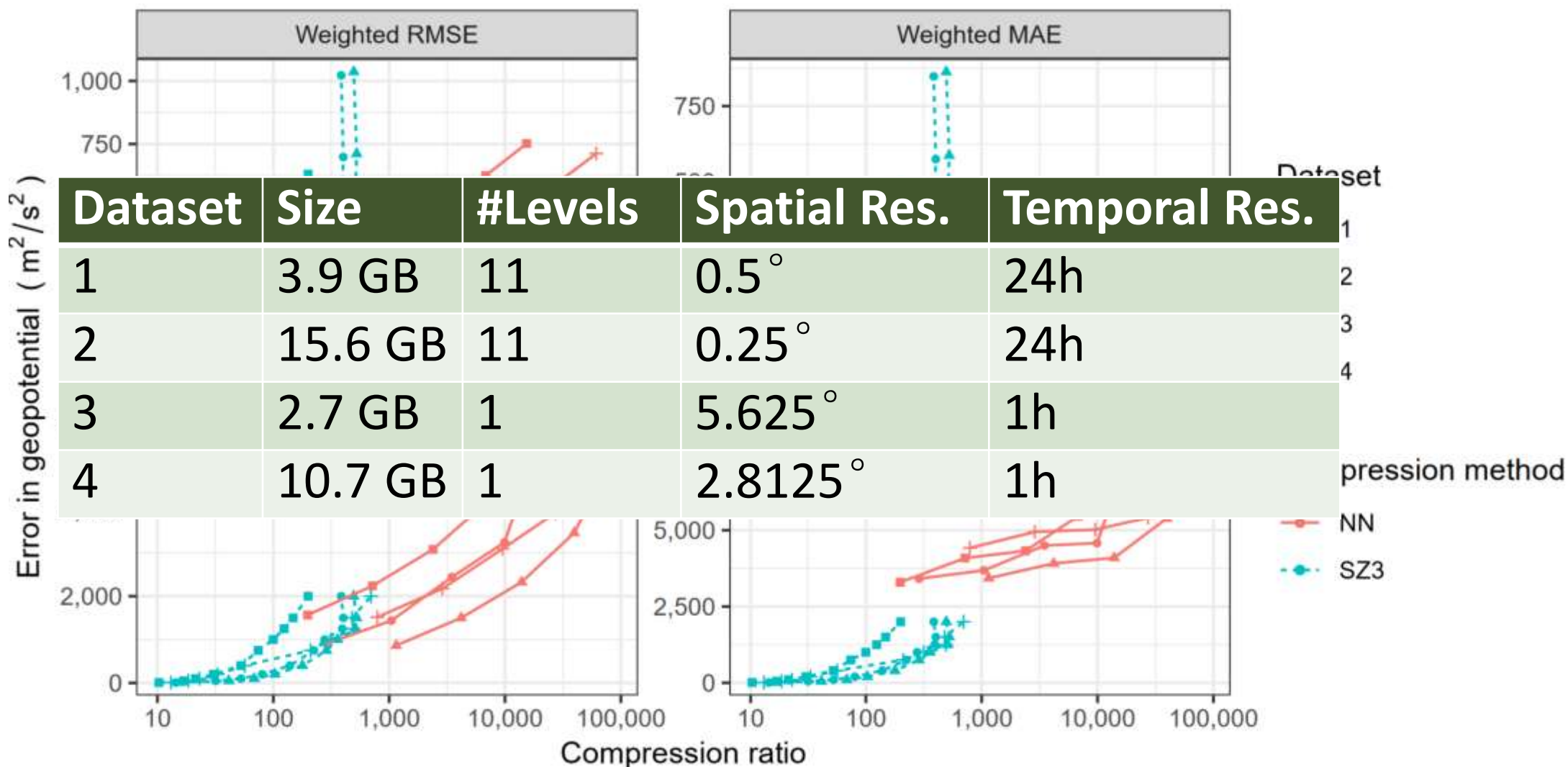
Compression / Training



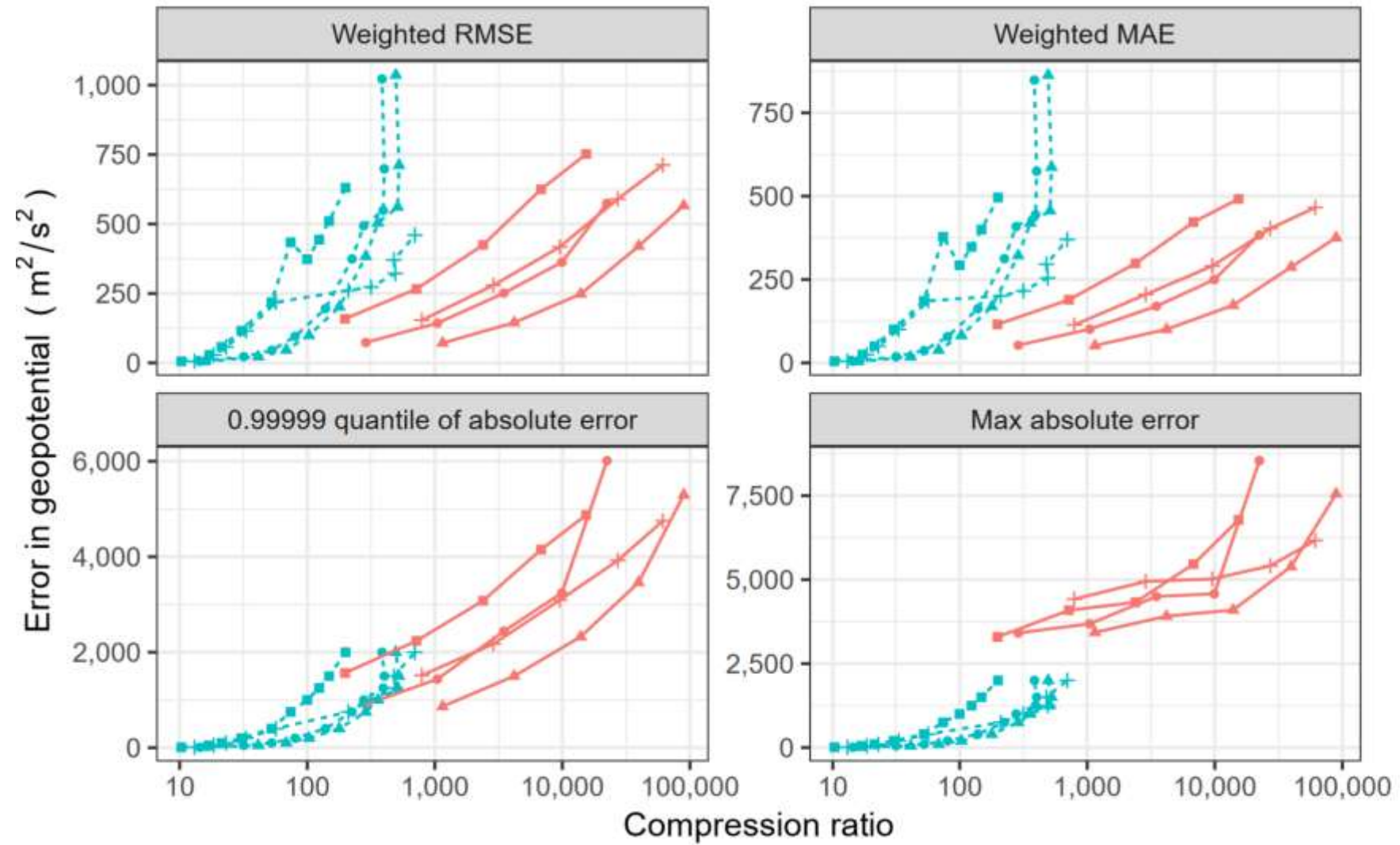
Comparison with Existing Methods

Method	Compression Ratio	Comp. Speed	Decompression Continuous Access	Decompression Random Access
ZFP [2]	< 10 x			
TTHRESH [3]	< 300 x			
SZ3 [4]	< 400 x			
NN (Ours)	300 x – 3,000 x			

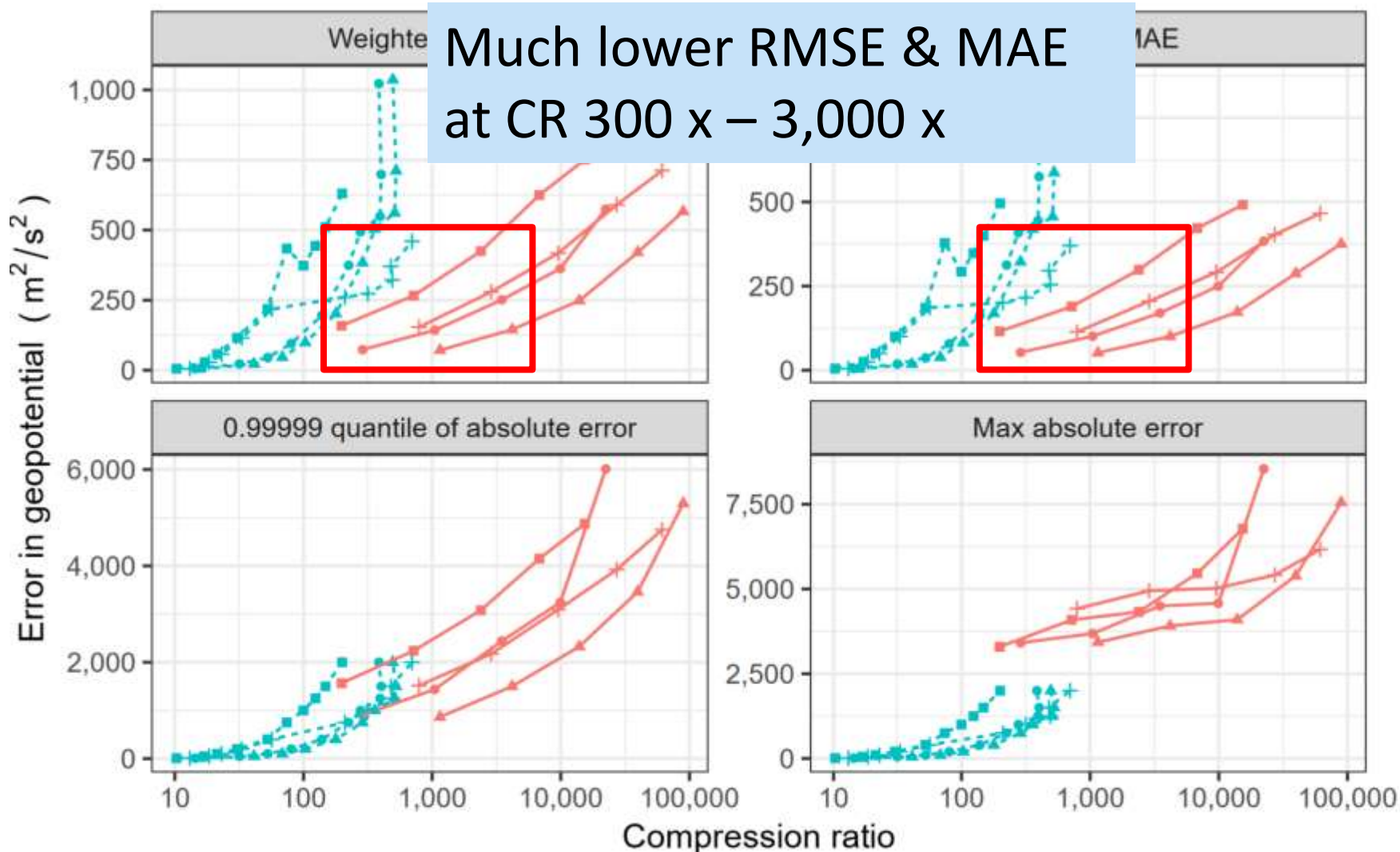
Evaluation: Compression Error



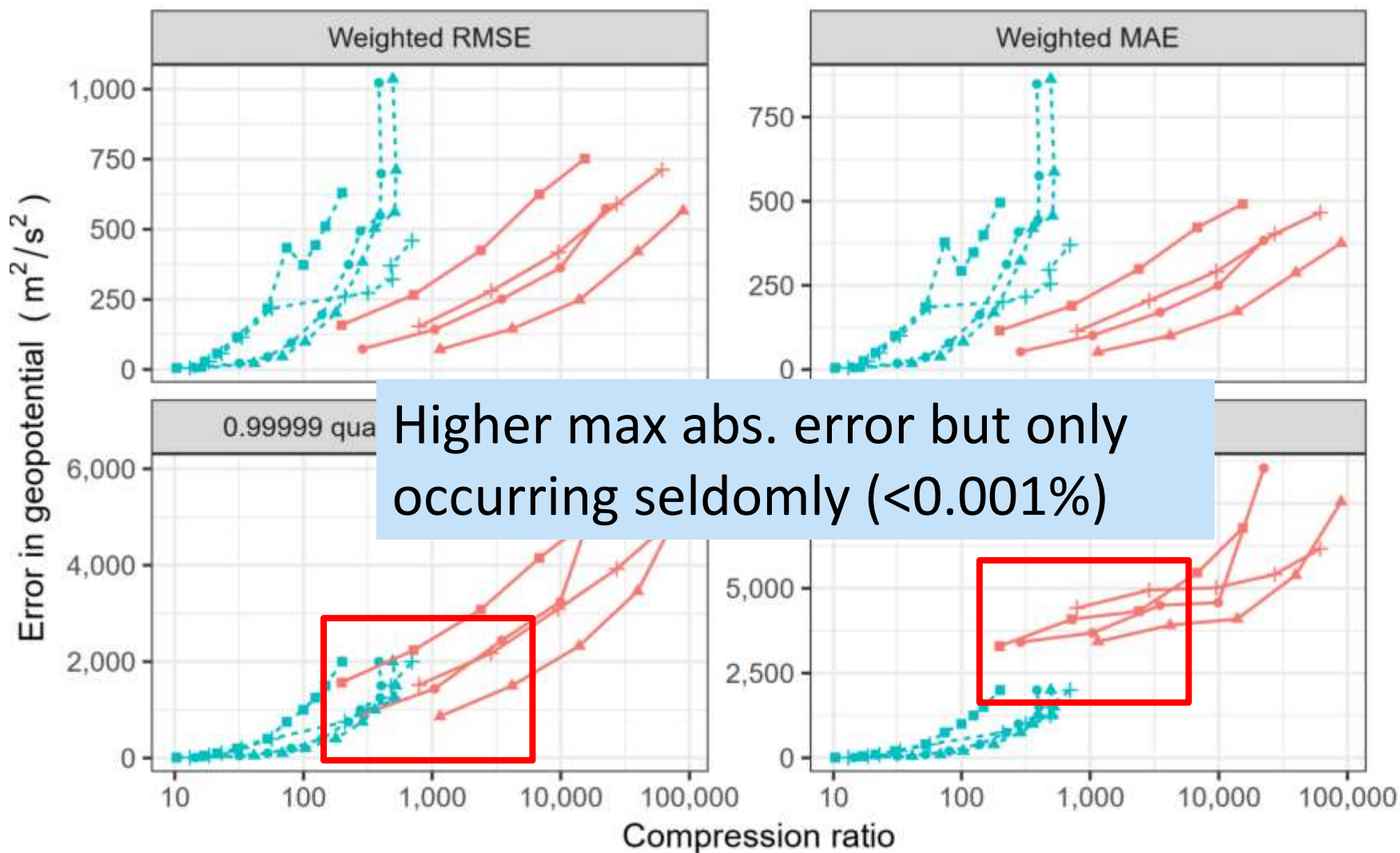
Evaluation: Compression Error



Evaluation: Compression Error

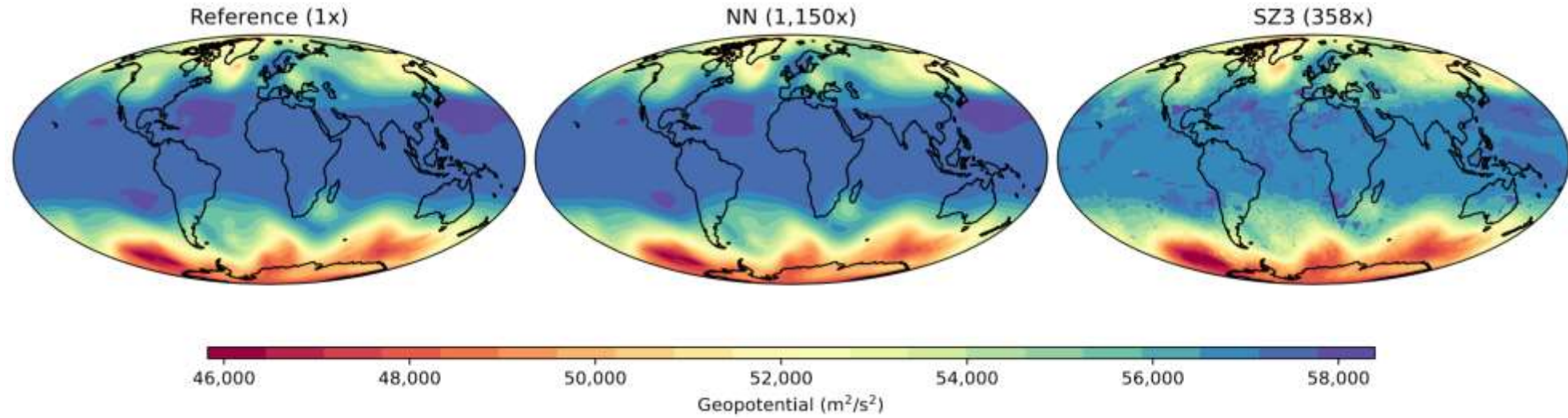


Evaluation: Compression Error



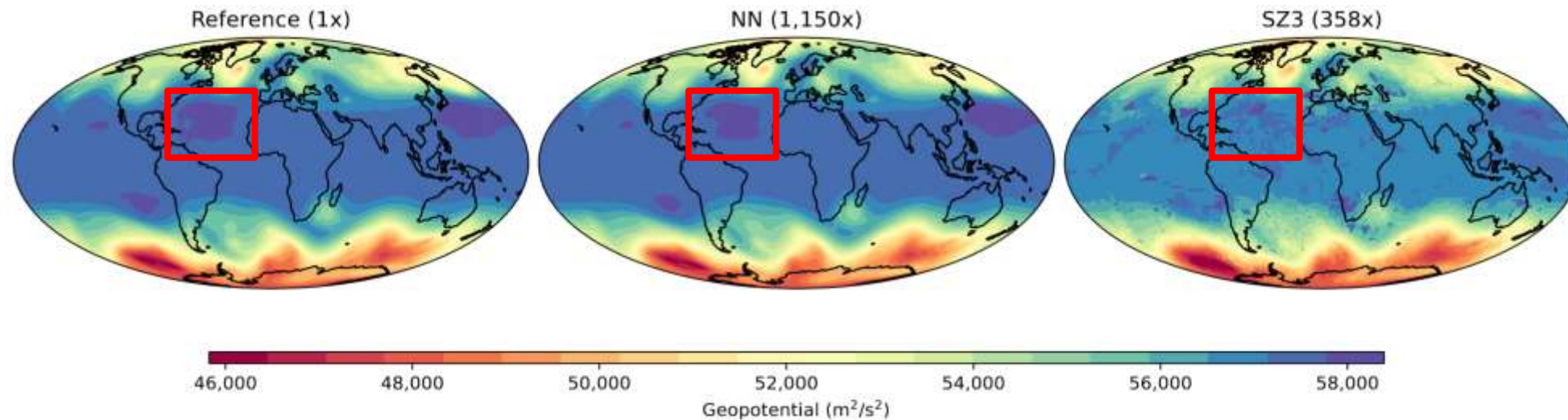
Evaluation: Case Study

Geopotential at 500hPa, 2016 Oct 5th



Evaluation: Case Study

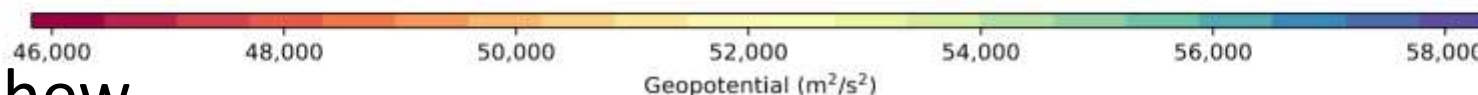
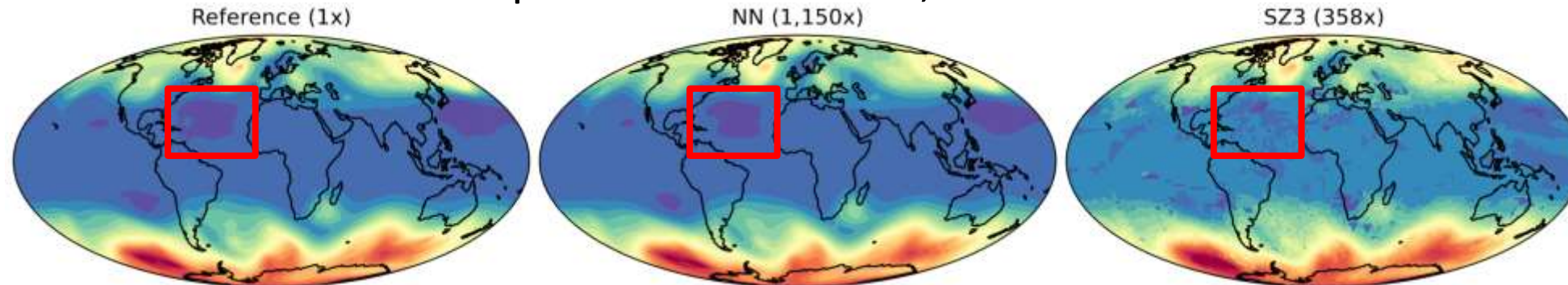
Geopotential at 500hPa, 2016 Oct 5th



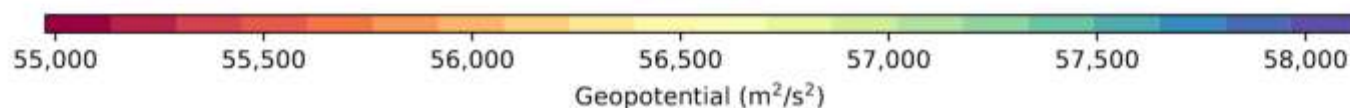
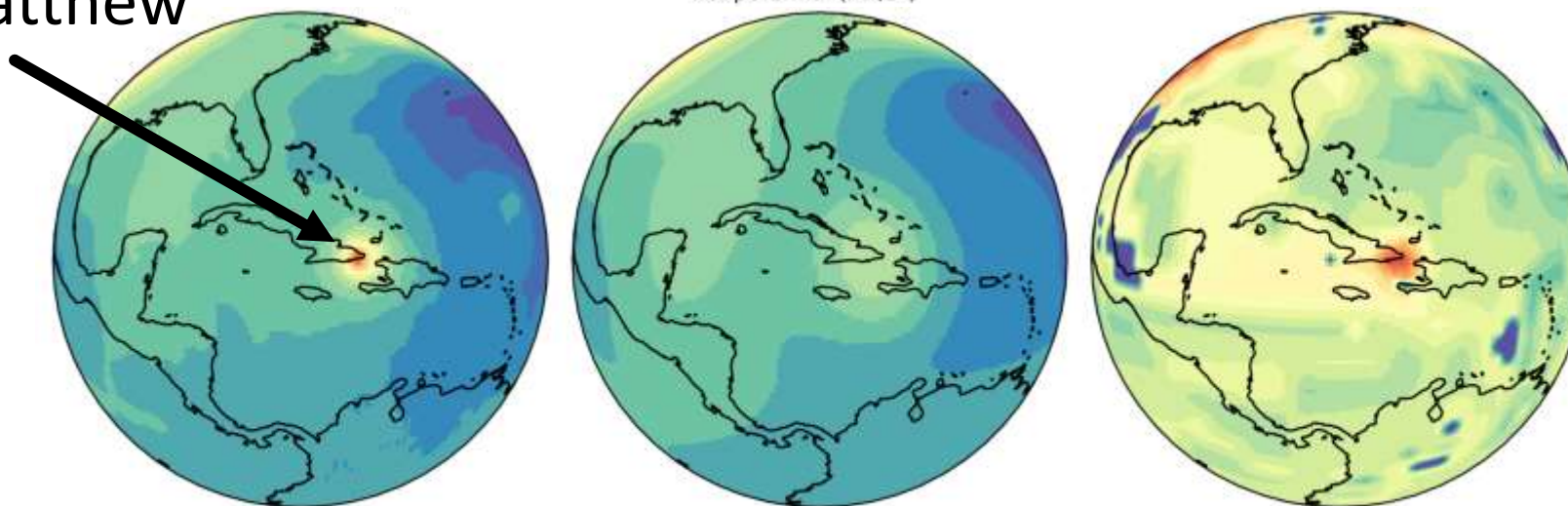
Preserves general shapes of important events and average values without introducing significant artifacts

Evaluation: Case Study

Geopotential at 500hPa, 2016 Oct 5th

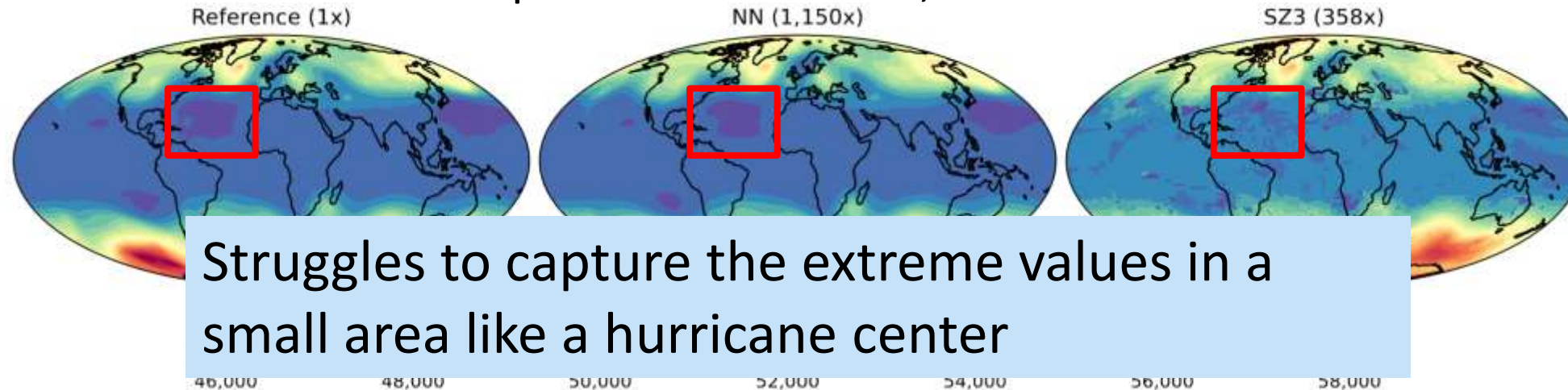


Hurricane Matthew



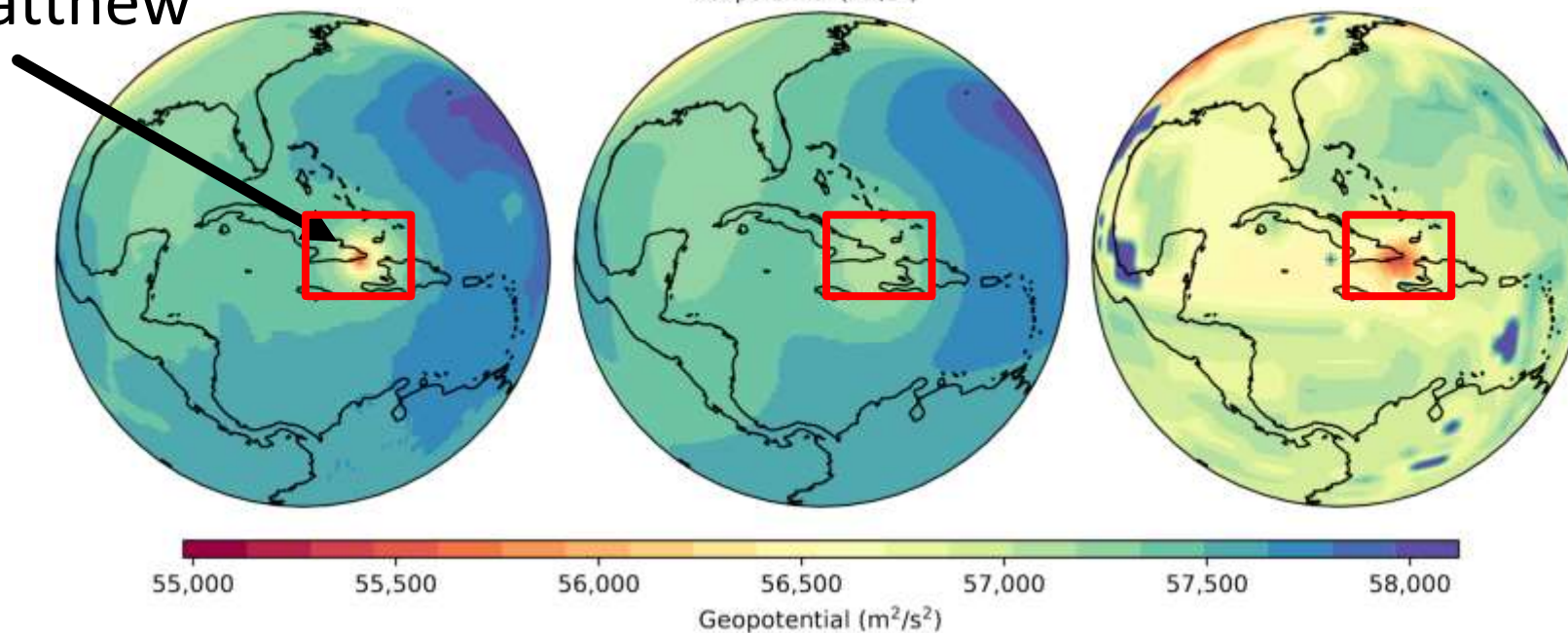
Evaluation: Case Study

Geopotential at 500hPa, 2016 Oct 5th



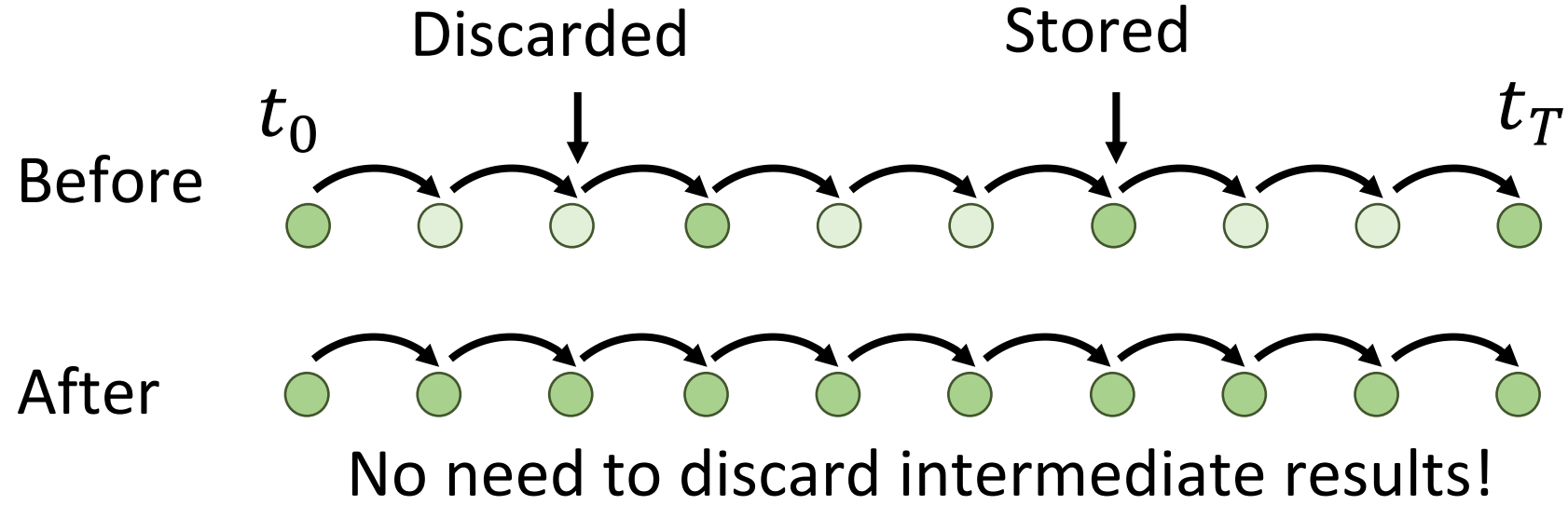
Struggles to capture the extreme values in a small area like a hurricane center

Hurricane Matthew



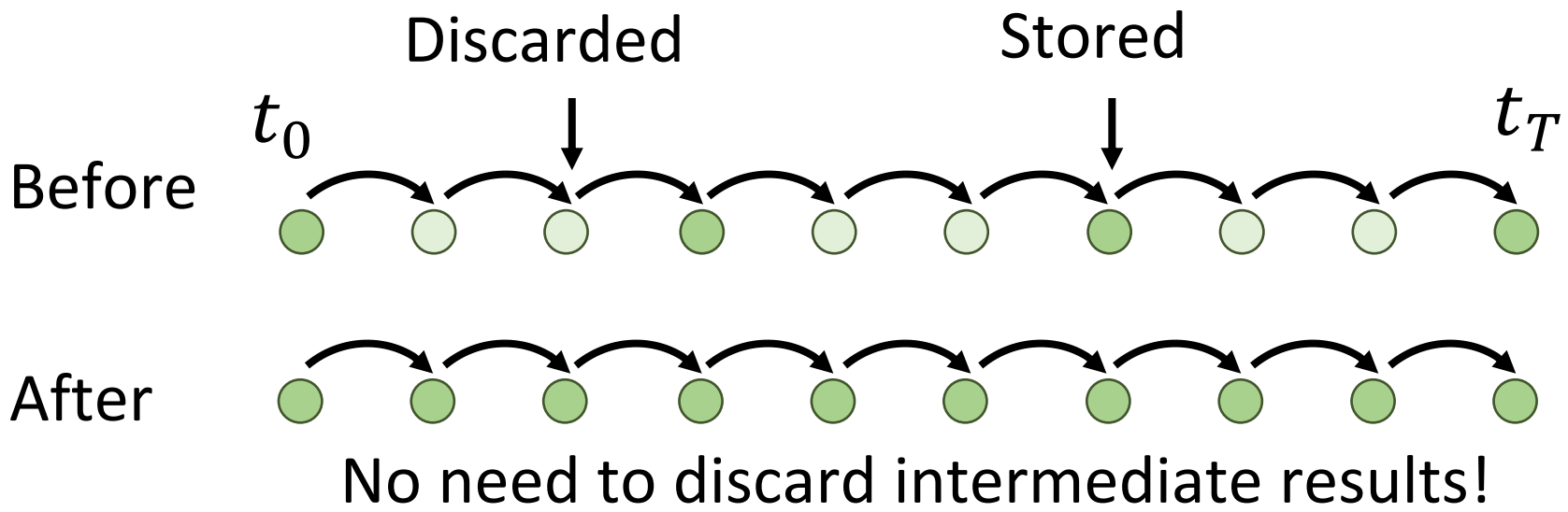
Applications

Store More Data

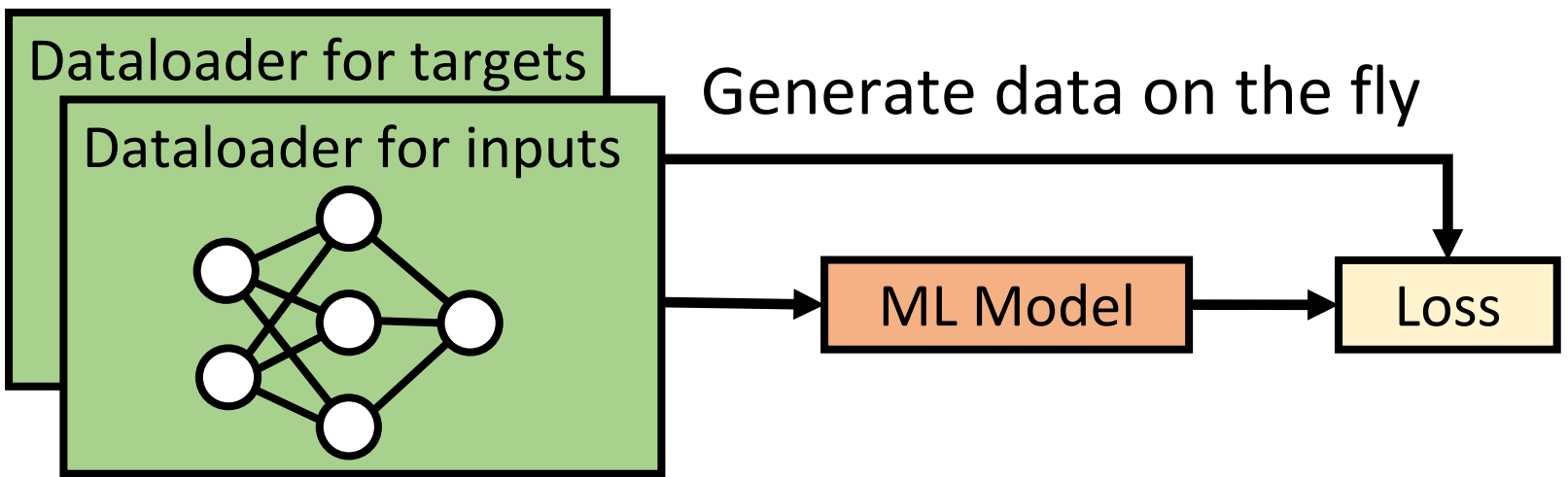


Applications

Store More Data



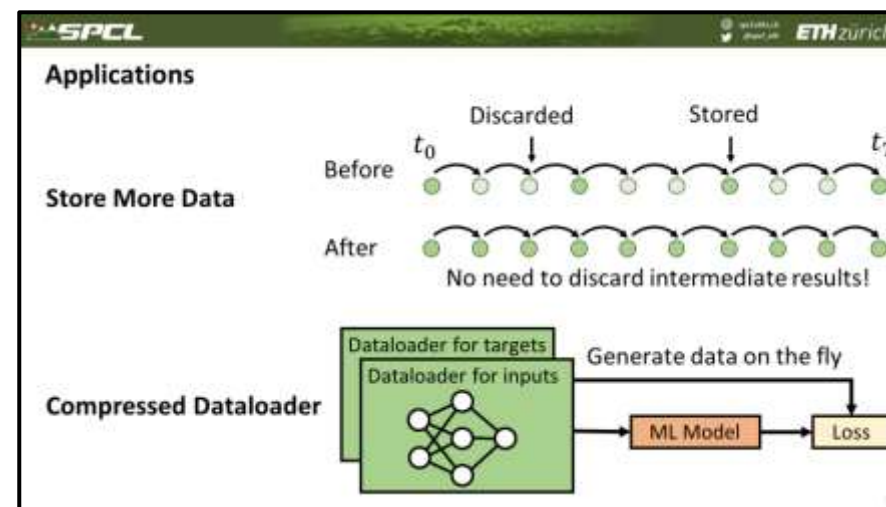
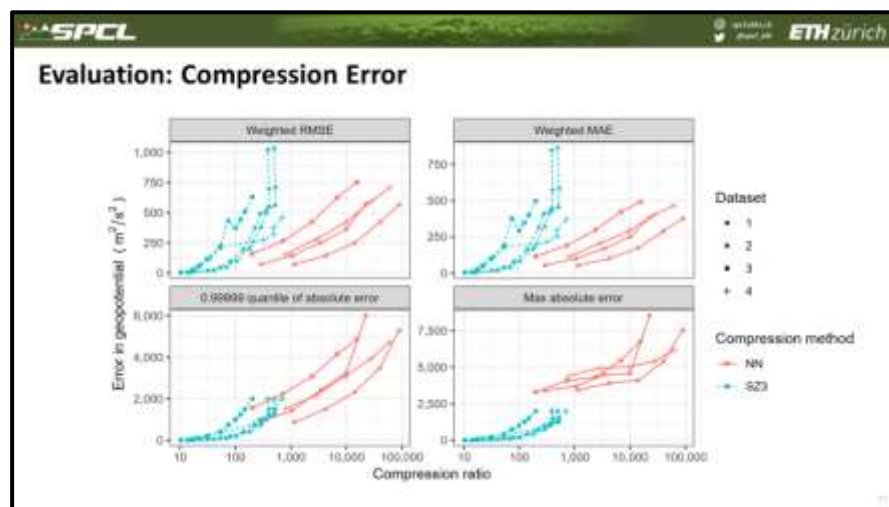
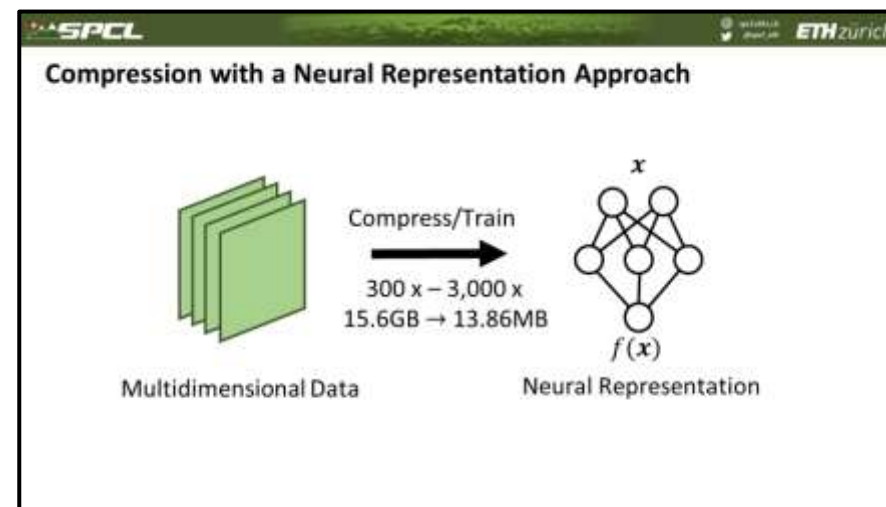
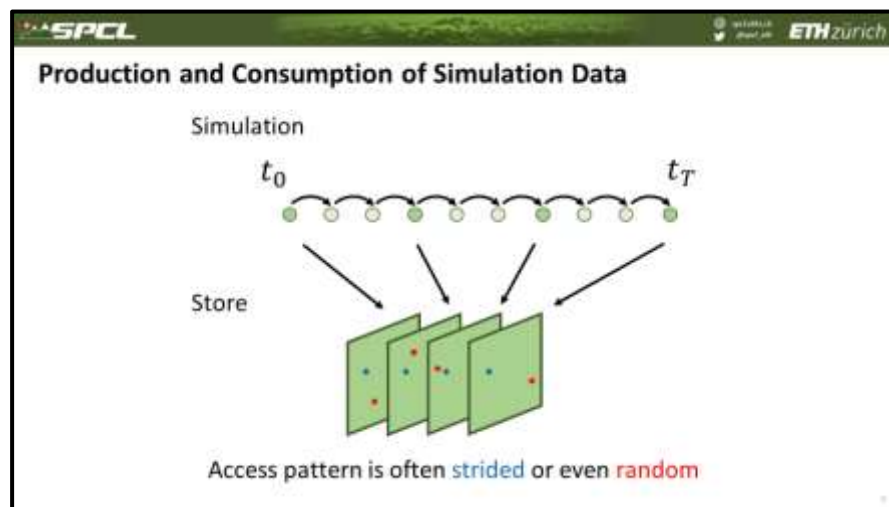
Compressed Dataloader



Applications

Storage	Dataset 3	C.R.	Discarded		Stored	
			Weighted RMSE error (test set)			
			Z at 500 hPa (m ² /s ²)	T at 850 hPa (K)		
	Original	1 x	632.9	2.906		
	NN Compressed	198 x	637.3 (+0.7%)	2.944 (+1.3%)		
	SZ3 Compressed	71 x	650.6 (+2.8%)	2.985 (+2.7%)		
Compression	Dataset 4					
	Original	1 x	688.8	2.834		
	NN Compressed	790 x	697.3 (+1.2%)	2.888 (+1.9%)		
	SZ3 Compressed	106 x	702.9 (+2.0%)	2.887 (+1.9%)		

Summary



Reference

- [1] “ECMWF's Vision for Big Data, AI and Cloud Computing,” 2019.
- [2] Lindstrom, Peter. ‘Fixed-Rate Compressed Floating-Point Arrays’. *IEEE Transactions on Visualization and Computer Graphics* 20, no. 12 (2014): 2674–83.
- [3] Ballester-Ripoll, Rafael, Peter Lindstrom, and Renato Pajarola. ‘TTHRESH: Tensor Compression for Multidimensional Visual Data’. *IEEE Transactions on Visualization and Computer Graphics* 26, no. 9 (2019): 2891–2903.
- [4] Liang, Xin, Kai Zhao, Sheng Di, Sihuan Li, Robert Underwood, Ali M. Gok, Jiannan Tian, et al. ‘SZ3: A Modular Framework for Composing Prediction-Based Error-Bounded Lossy Compressors’. *IEEE Transactions on Big Data*, 2022.

Thank you!




Paper and Poster



Github Repo



More of SPCL's research:

-  youtube.com/@spcl **150+ Talks**
-  twitter.com/spcl_eth **1.2K+ Followers**
-  github.com/spcl **2K+ Stars**

... or spcl.ethz.ch

