Recovering from Silent Data Corruption through Spatial Prediction

Sarah Placke

Dr. Jon Calhoun (Advisor)

CLEMSON UNIVERSITY

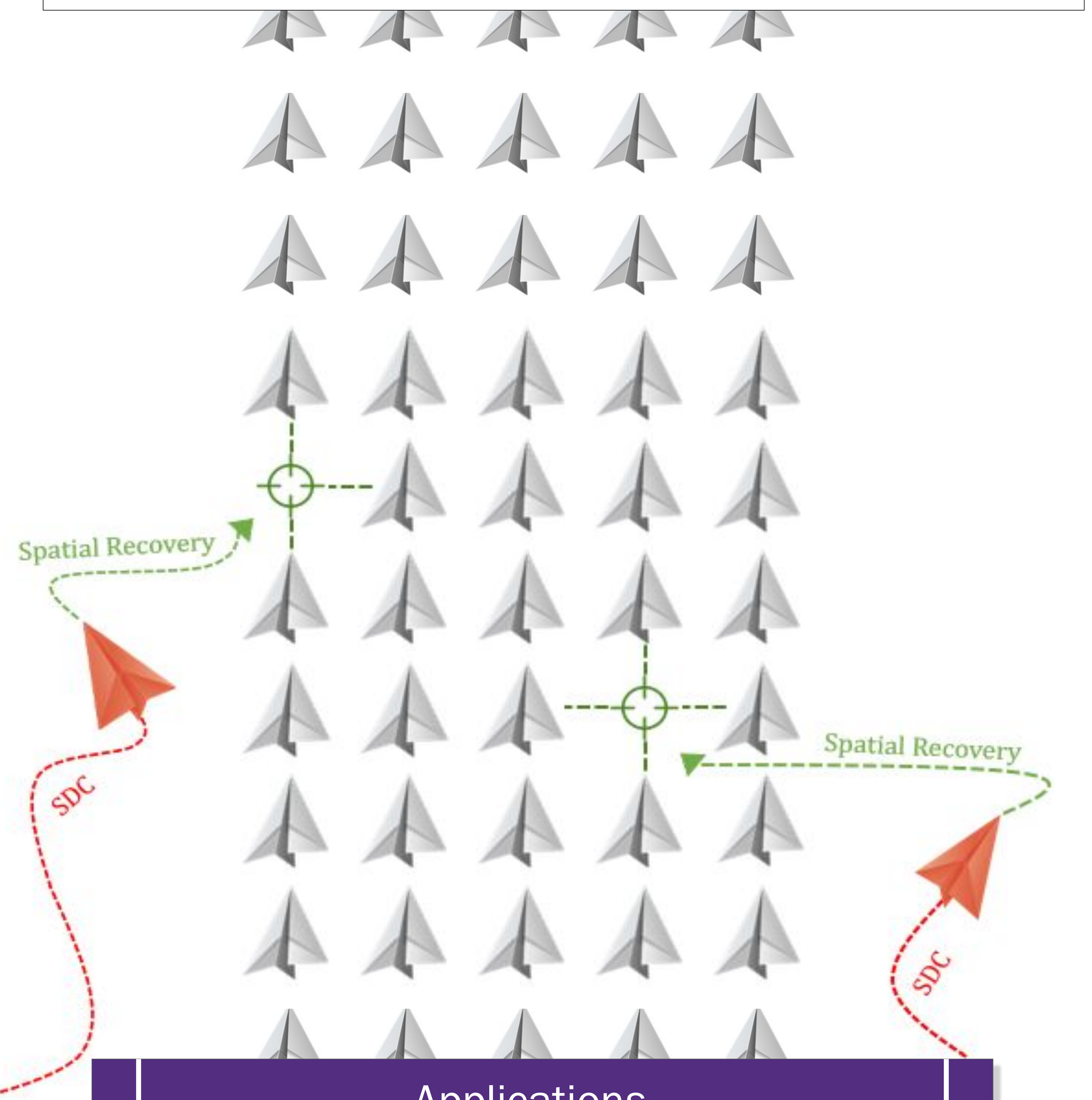
Abstract

- High-performance computing applications are central to advancement in science and engineering, but transient upset events become more likely as system size grows and hardware components are run with near-threshold voltages.
- Silent Data Corruption can silently perturb HPC data. This corruption can lead to invalid large-scale application results.
- **We explore the use of spatial similarity to recover from silent data corruption. We test eight reconstruction methods that utilize local data to implement low-cost spatial recovery.**

Experimental Methods

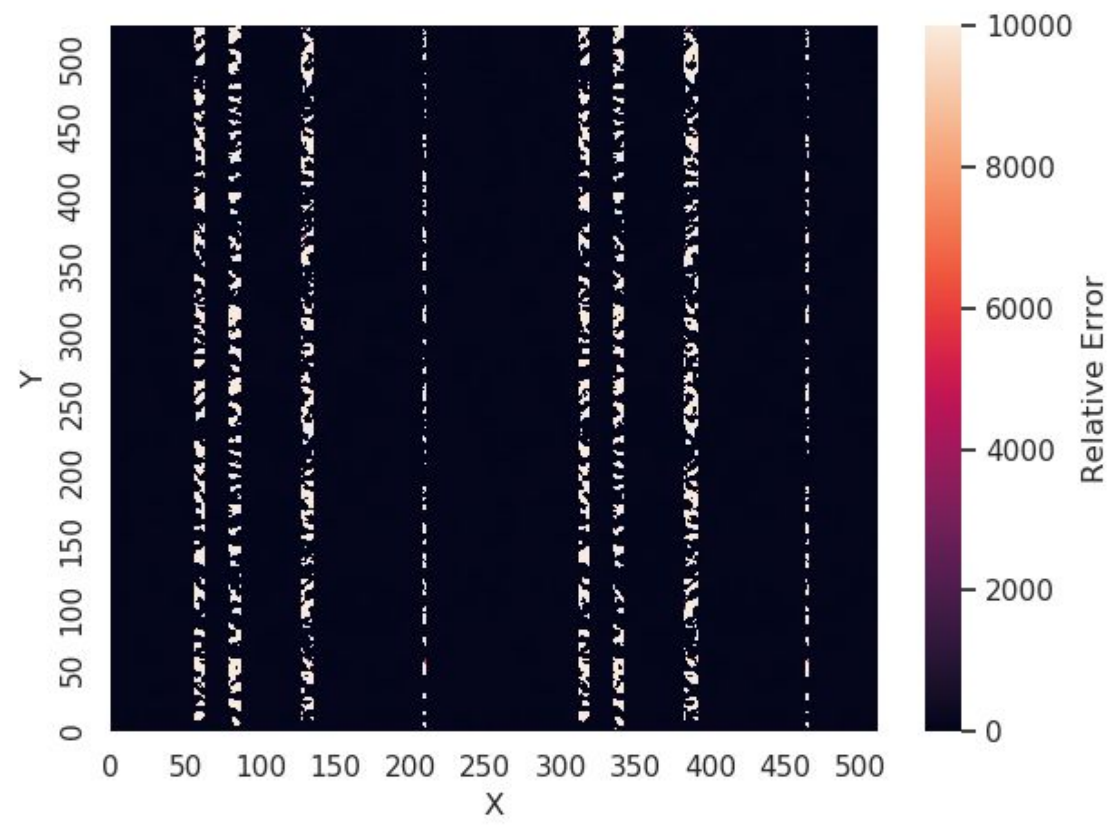
N-Layer Lorenzo Predictors (N = 1,2,3,4)

Average, Quadratic, Power, and Linear Regression



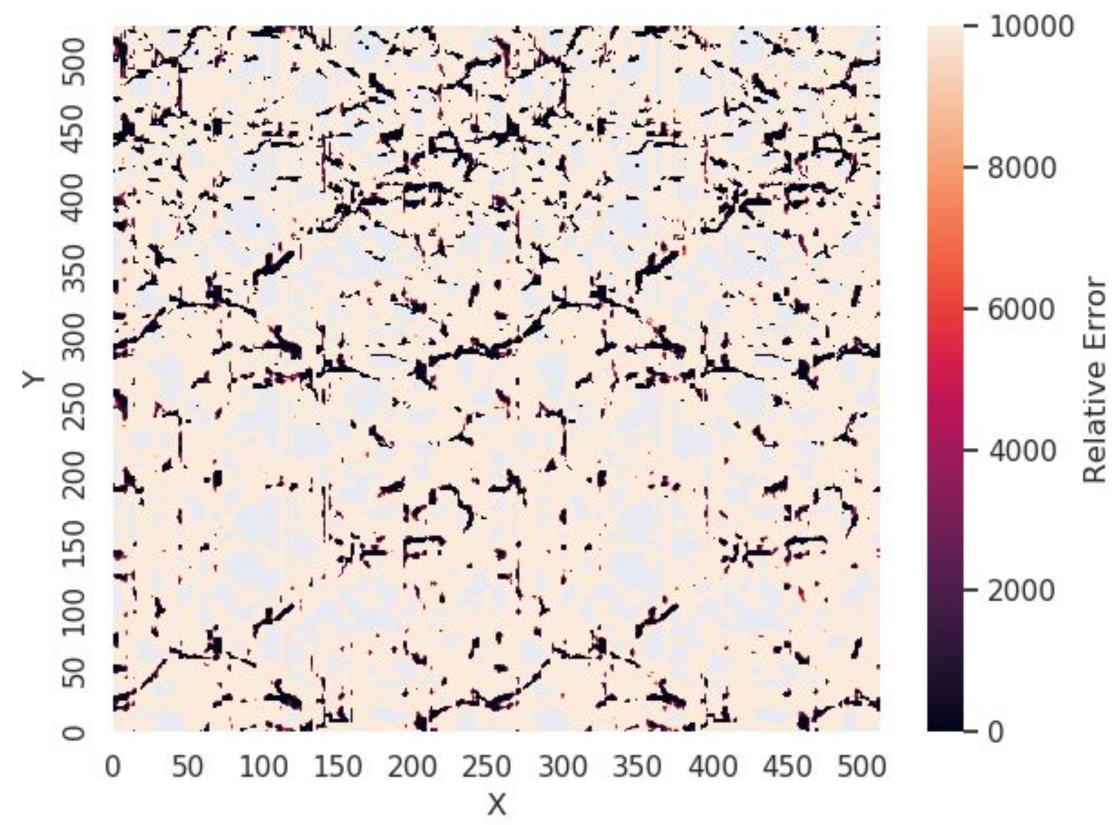
Results

Linear Regression's Relative Error based on Positional Coordinates at slice Z = 200.0



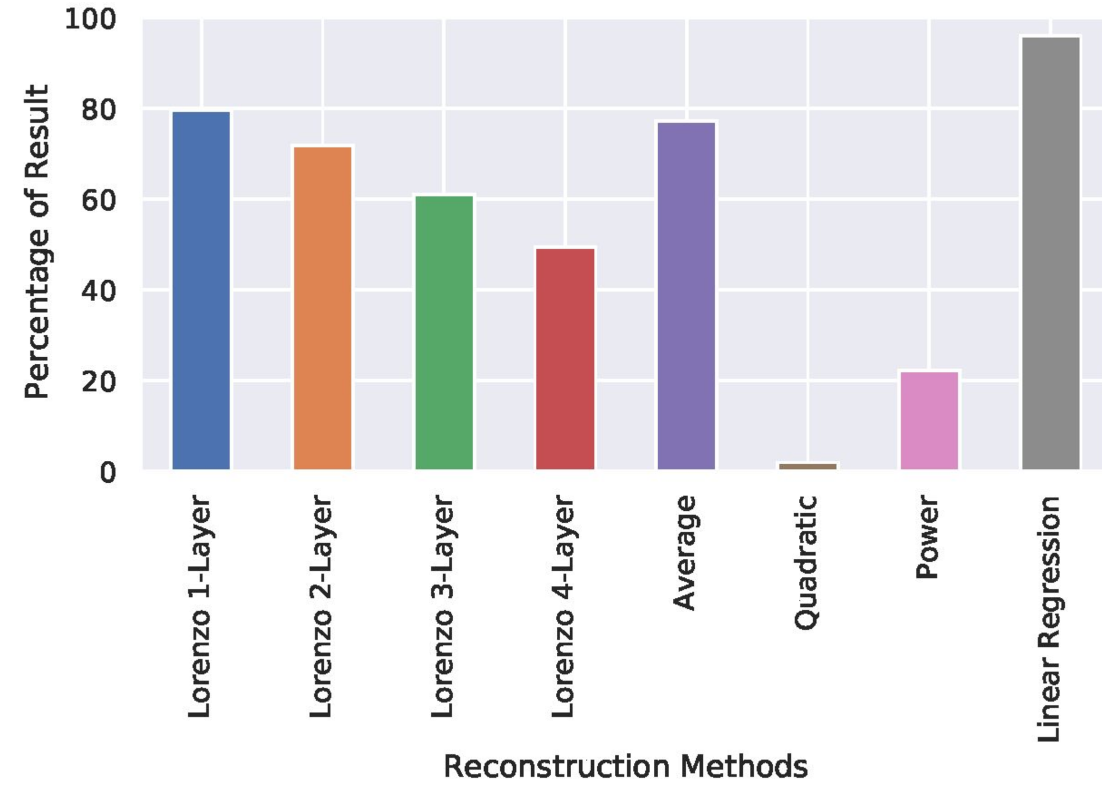
This figure shows the spatial smoothness of baryon density from the NYX data set in relation to the Linear Regression reconstruction model.

Quadratic's Relative Error based on Positional Coordinates at slice Z = 200.0



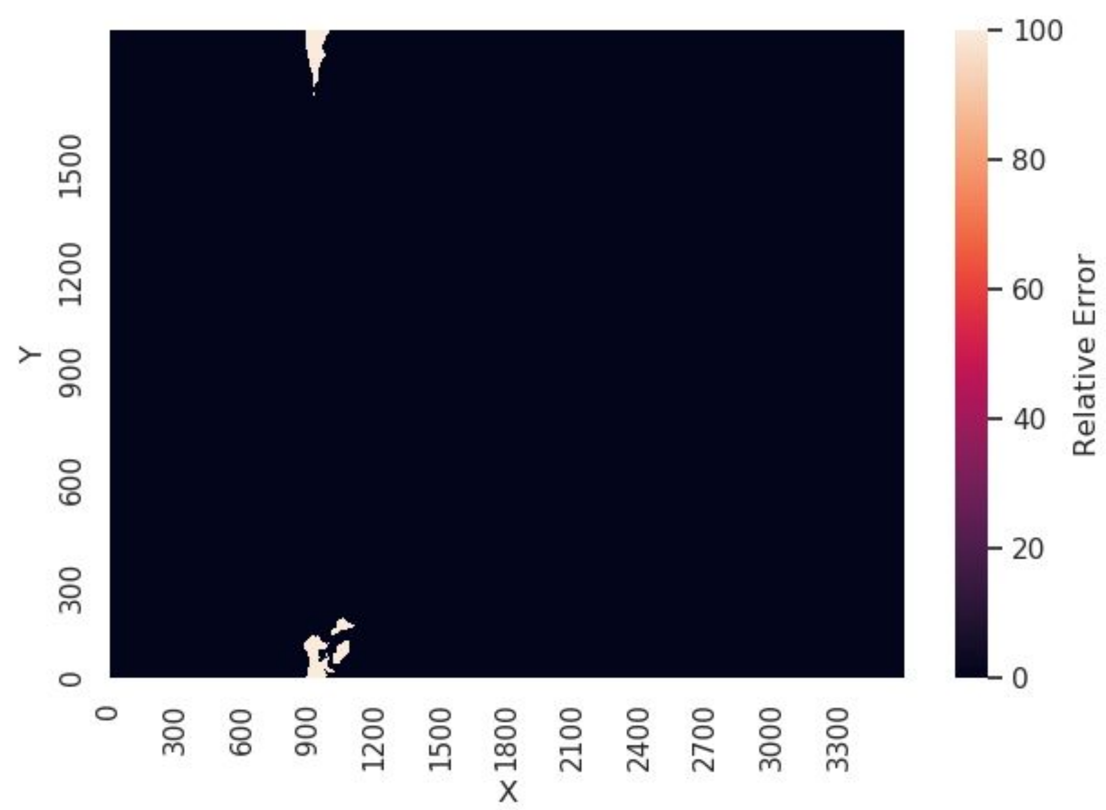
This figure shows the spatial smoothness of baryon density from the NYX data set in relation to the Quadratic reconstruction model.

Percent of Predictions Within 1% of the Correct Value



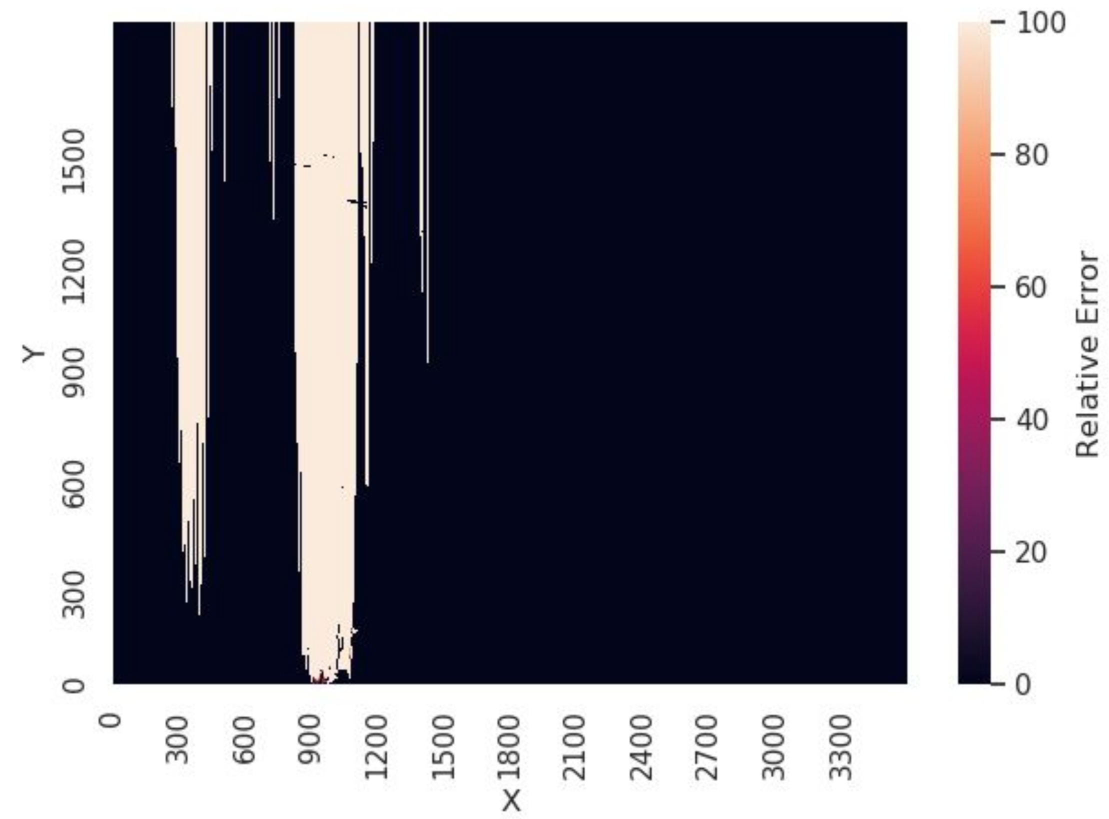
This figure demonstrates that linear regression has the most accuracy for baryon density from the NYX data set.

Linear Regressions Relative Error based on Positional Coordinates



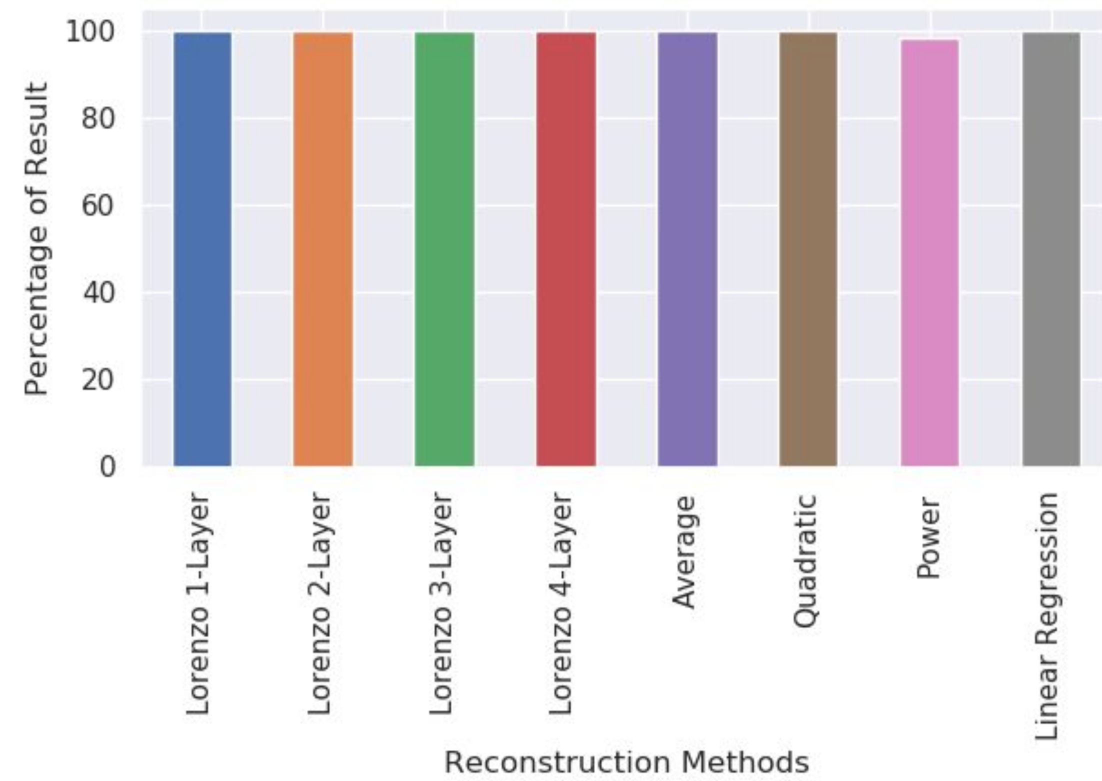
This figure shows the spatial smoothness of AEROD from the CESM data set in relation to the Linear Regression reconstruction model.

Power's Relative Error based on Positional Coordinates



This figure shows the spatial smoothness of AEROD from the CESM data set in relation to the Power reconstruction model.

Percent of Predictions Within 1% of the Correct Value



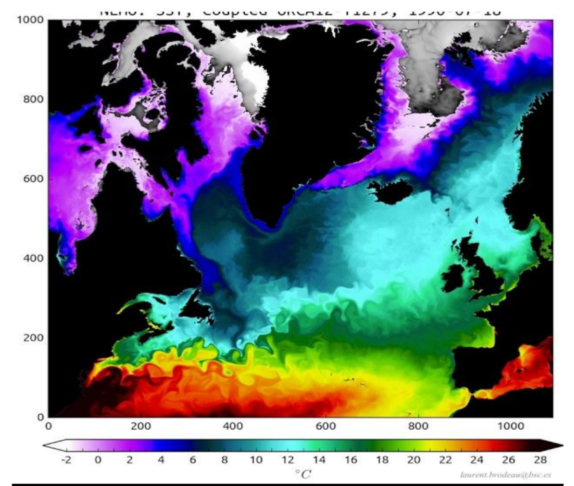
This figure demonstrates that linear regression has the most accuracy for AEROD from the CESM data set.

Applications

CESM

The Computer Earth System Model (CESM) models how Earth's climate may respond to change. It is a numerical simulation that considers atmospheric, ocean, ice, land surface, carbon cycle, along with other components. CESM is used to model potential past, present, and future Earth systems.

Data set dimensions: 1800 x 3600



NYX

NYX is an N-body cosmological simulation that incorporates adaptive mesh hydrodynamics. It considers baryon density, dark matter density, temperature, velocities in the x, y, and z direction, and other characteristics. As a 'cosmic web' simulator, NYX is used to explore the fundamentals of physics and astrophysics.

Data set dimensions: 512 x512x 512

Conclusion

- Results show that the linear regression-fit is the most accurate reconstruction method with over 90% of its predictions within 1% of the correct value.
- Discrepancies between individual reconstruction method accuracy decrease in proportion to the data set's spatial smoothness.
- Spatial recovery is an effective tool in mitigating the negative influences of SDC.

ACKNOWLEDGEMENTS
Clemson University is acknowledged for generous allotment of compute time on the Palmetto cluster. This material is based upon work supported by the National Science Foundation under Grant No. SHF-1910197.