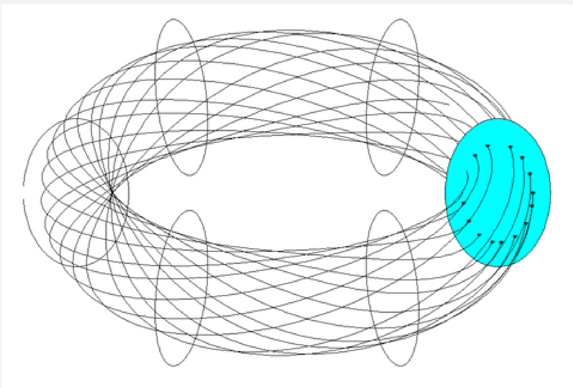
Compressing Unstructured Mesh Data using Spline Fits, Compressed Sensing, and Regression Methods

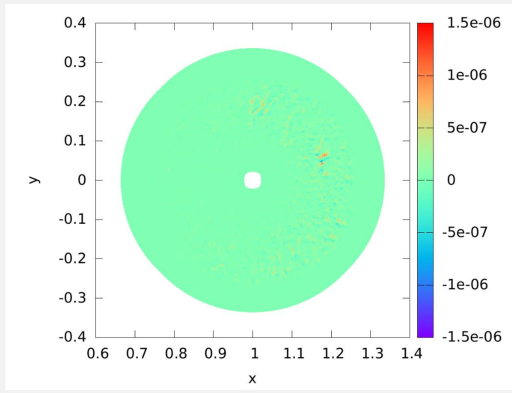
Chandrika Kamath and Ya Ju Fan

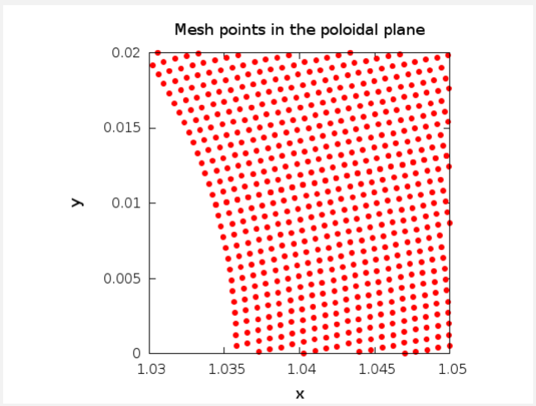
Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore , CA

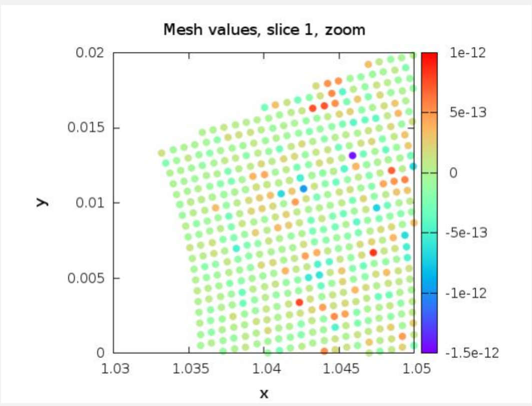
Compressing unstructured mesh data from computer simulations is very challenging. Using a test problem from a fusion simulation, we compare three methods on the accuracy of reconstruction and the reduction in data size. Our results indicate that compressed sensing works well if the data are sparse, regression requires regions with nearly uniform values, and spline fits, with the fewest restrictions on the data, performs the best.

A motivating example from a fusion simulation

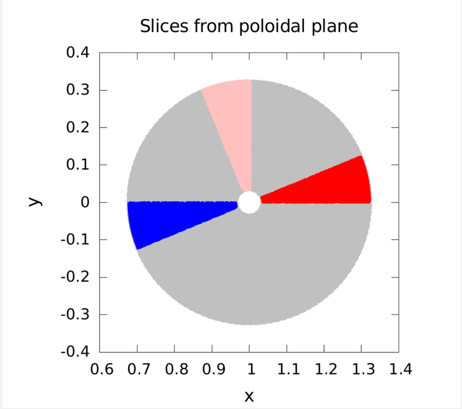






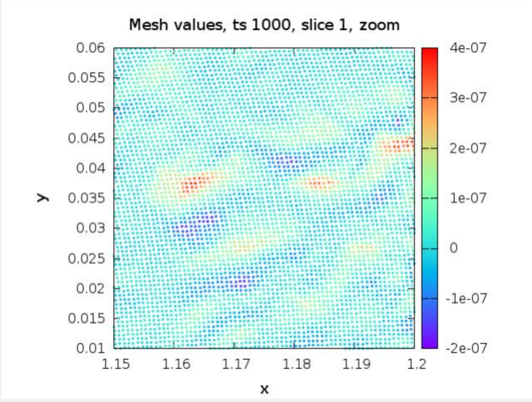


Clockwise from top left: A schematic of a three-dimensional toroid representing the tokamak, with a poloidal plane high-lighted. The variable of interest at time step 1500 on the poloidal plane, with 591,745 grid points. Slices 1, 5, and 9, in red, pink, and blue, respectively, used in our work, each with 37,000 grid points



Challenges to the compression:

- unstructured mesh
- double precision data
- positive and negative values
- large variation in nearby values
- large range of values within and across time steps



Description of the methods

Scale the data at each time step by the absolute of the maximum value and extract the three slices. Apply the methods to each slice and evaluate metrics for different parameters.

Spline fits using ISABELA:

- Split data into fixed size windows of size W.
- Sort the data and fit B-splines with C coefficients.
- Store relative error between estimated and actual values when errors exceed threshold E.

Compressive sensing using GPSR:

- To introduce sparsity, set all scaled values in range [-0.03,0.03] to zero
- Apply the gradient projection for sparse reconstruction (GPSR) method, with a user defined value of *m*, the number of measurements

Machine learning using locally-weighted kernel regression (LWKR):

- Choose a percentage, S, of samples randomly as far apart from each other.
- Use the values at these samples as the training set at each time step.
- Use locally weighted kernel regression to predict values at remaining points.
- Identify a fixed percentage (HERR %) of sample points with highest errors; write out values along with indices.

Performance metrics for comparing methods

Coefficient of determination between actual data, *a*, and reconstructed data, *p*:

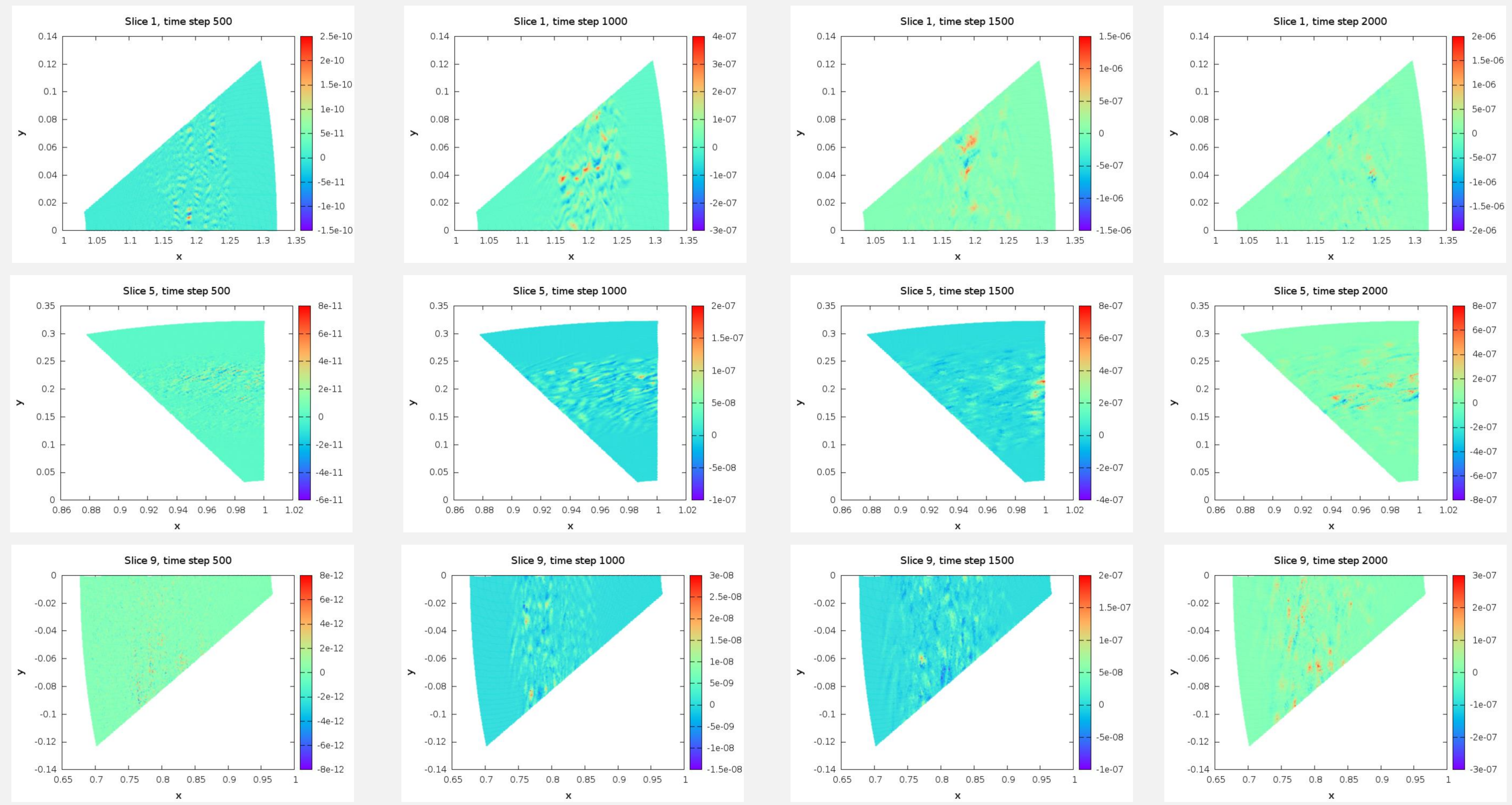
$$R^2 = \frac{(n\sum a_i p_i - \sum a_i \sum p_i)^2}{\left(n\sum a_i^2 - (\sum a_i)^2\right)\left(n\sum p_i^2 - (\sum p_i)^2\right)}$$

Output reduction factor::

$$\text{ORF} = \frac{\textit{size of gzipped original data}}{\textit{size of gzipped compressed data}}$$

Description of the data from the simulation

Time step 500 | Time step 1000 | Time step 1500 | Time step 2000



Top row: slice 1; middle row: slice 5; bottom row: slice 9

Results of compression and comparison of methods

Time Step 1500, Slice:	ISABELA			GPSR			LWKR		
	parameters	R^2	ORF	parameters (m/n)	R^2	ORF	parameters	R^2	ORF
1	W1024C30E1	0.99997	3.4460	0.6	0.91684	3.8925	S10HERR15	0.98601	3.0699
	W512C30E1	0.99997	3.1077	0.7	0.97974	3.6201	S20HERR8	0.98225	3.1776
	W1024C30E10	0.99876	4.8278	0.8	0.99926	3.6346	S25HERR5	0.97868	3.1716
5	W1024C30E1	0.99997	3.2194	0.6	0.99992	3.3394	S10HERR15	0.97657	3.0629
	W512C30E1	0.99997	2.8799	0.7	0.99995	3.5796	S20HERR8	0.96793	3.1709
	W1024C30E10	0.99792	4.4911	0.8	0.99996	3.5182	S25HERR5	0.96193	3.1624
9	W1024C30E1	0.99997	3.2455	0.6	0.99999	3.3443	S10HERR15	0.95877	3.0375
	W512C30E1	0.99997	2.8922	0.7	0.99999	3.1114	S20HERR8	0.94501	3.1514
	W1024C30E10	0.99791	4.5455	0.8	0.99999	3.1217	S25HERR5	0.93345	3.1489

- ISABELA: Reducing window size reduces ORF as more windows result in more coefficients. Increasing error threshold increases ORF and decreases R^2 as more error is tolerated. Results are independent of the time step or slices.
- Compressed sensing (GPSR): The more sparse slice 9 has perfect reconstruction, but not slice 1 with fewer non-zeros. Increasing the number of measurements improves R^2 but usually lowers ORF.
- Regression using LWKR: Accuracy improves if we use a lower sampling rate but keep a higher percentage of the high-error points. Data that has large regions with roughly constant values (slice 1 at mid-time) gives better reconstruction.

LLNL-POST-761299: This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This work is part of the ExaDM project, funded by Dr. Lucy Nowell, program manager, ASCR/DOE.