An *In Situ* Approach for Approximating Complex Computer Simulations and Identifying Important Time Steps

Kary Myers[1], Earl Lawrence[1], Michael Fugate[1], Claire M. Bowen[2], Lawrence Ticknor[1], Jon Woodring[3], Joanne Wendelberger[1], and Jim Ahrens[3]

1. Statistical Sciences, Los Alamos National Laboratory 2. Applied and Computational Mathematics and Statistics, University of Notre Dame 3. Applied Computer Science, Los Alamos National Laboratory

What is the Problem?

As computer simulations continue to increase in size and complexity, scientists are moving applications toward exascale computing (i.e. 10^{18} FLOPS or Floating-point Operations Per Second). However, when analyzing these simulations, their output might exceed both the storage capacity and the bandwidth required for transfer to storage. One solution is embedding some of the analysis during the simulation, often called *in situ*.

Since *in situ* adds computational burden to the simulation, we seek lightweight analysis that satisfies the following:

1. Identify important time steps of the simulation.
2. Significantly reduce the amount of data needed to move and store to preserve these time steps.
3. Facilitate future exploration of the stored reduced data (post processing).

How are Important Time Steps Identified?

At each time step of the simulation, the data is broken into three temporal regions:

- **curr**: Time steps currently classified by a linear fit.
- **buff**: B time steps newly computed by the simulation and stored in the buffer.
- **past**: Time steps no longer under consideration.

We can identify important time steps using piecewise linear fitting. This method verifies whether:

- One line (**curr** ∪ **buff**) fits the data, determining no change in the data e.g. Figure 1.
- Two lines (one line for **curr**, one for **buff**) fit the data better, determining there is a change in the data e.g. Figure 2.

Each time step is tested whether two lines provide a better fit than one line by calculating the F Statistic and comparing the associated p-value to α, the level of significance.

Piecewise Linear Fits Example

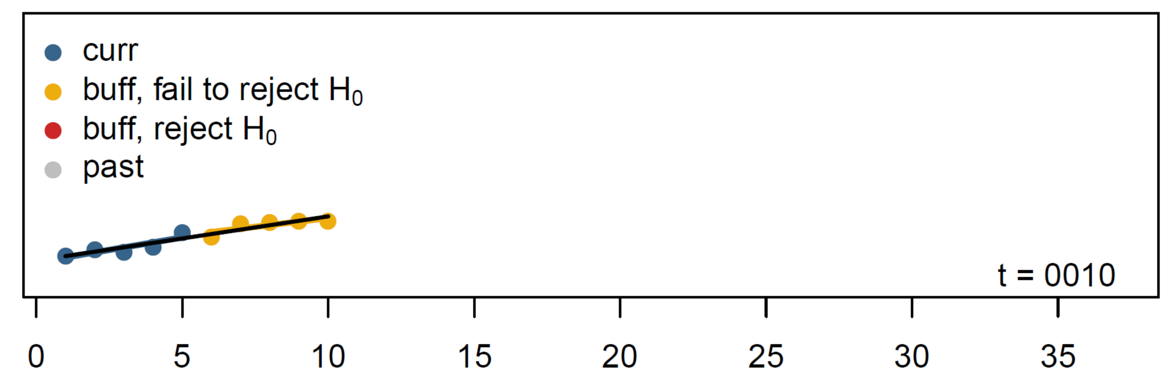

Figure 1: One line (H_0: **curr** ∪ **buff**, no change in the data) is a better fit than two.

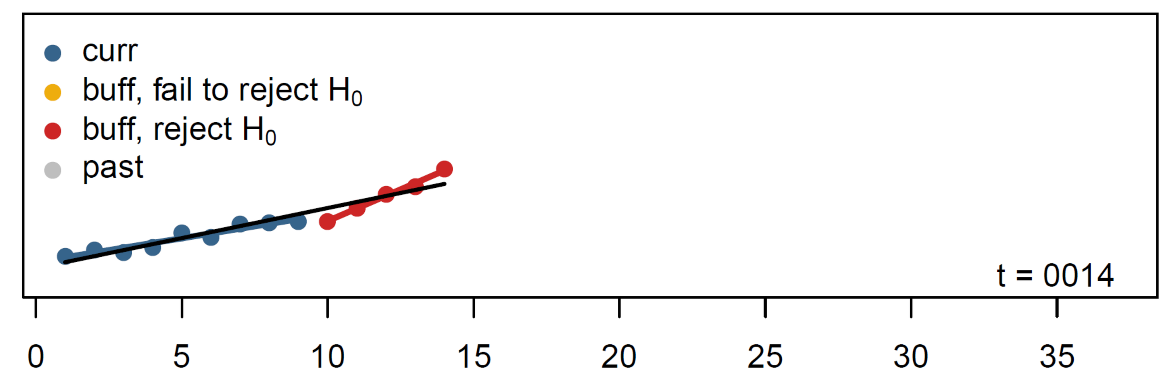

Figure 2: Two lines (one line for **curr**, one for **buff**, change in the data) is a better fit than one.

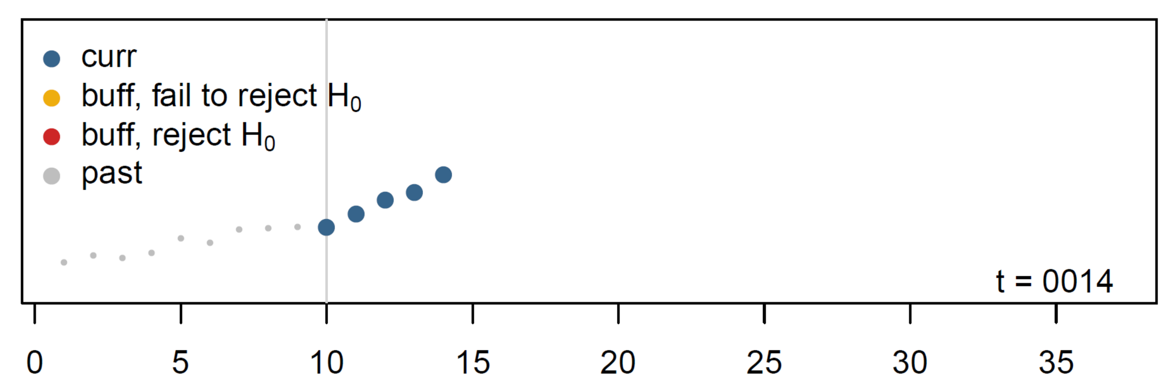

Figure 3: Since two lines is a better fit than one, the data in **curr** is partitioned (**past**) and sufficient statistics are gathered on **past**.

What Information Should be Stored at Each Partition?

The following six values are the sufficient statistics that are the minimum amount of information needed to capture the data in each partition:

$$\theta = \sum t_i, \Theta = \sum t_i^2, \psi = \sum y_i, \Psi = \sum y_i^2, \tau = \sum t_i y_i, \text{ and } T$$

where y_i is the data value and t_i is the time steps. T is the total number of time steps under consideration. These values can be used to calculate the sum of residuals squared (RSS) and the coefficients of the linear regression model $(\hat{\beta}_0, \hat{\beta}_1)$ for each partition.

$$RSS = \Psi - \frac{1}{T}\psi^2 - \frac{\left(\tau - \frac{\theta\psi}{T}\right)^2}{\Theta - \theta^2/T} \qquad \hat{\beta}_0 = \frac{1}{T}(\psi - \hat{\beta}_1\theta) \qquad \hat{\beta}_1 = \frac{\tau - \theta\psi/T}{\Theta - \theta^2/T}$$

How does Piecewise Linear Fits Perform?

To provide a concrete case study, we used a computer simulation run by Korycansky et al. (2009) in support of NASA's 2009 Lunar Crater Observation and Sensing Satellite mission (LCROSS). The plots show the change of pressure over time in a certain region.

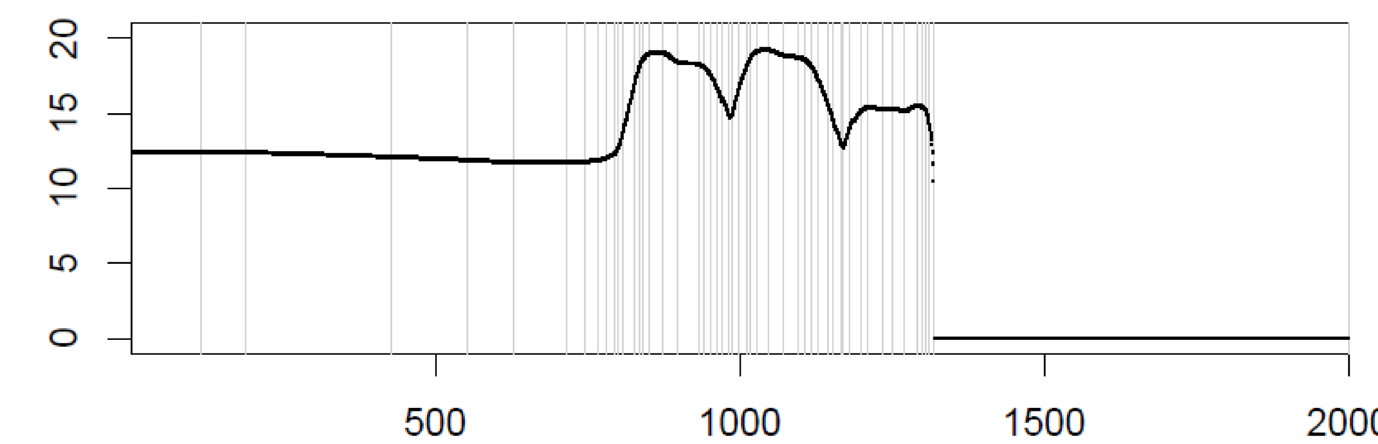

Figure 4: $\text{Buff} = 5, \alpha = 5 \times 10^{-4}$, 51 partitions selected.

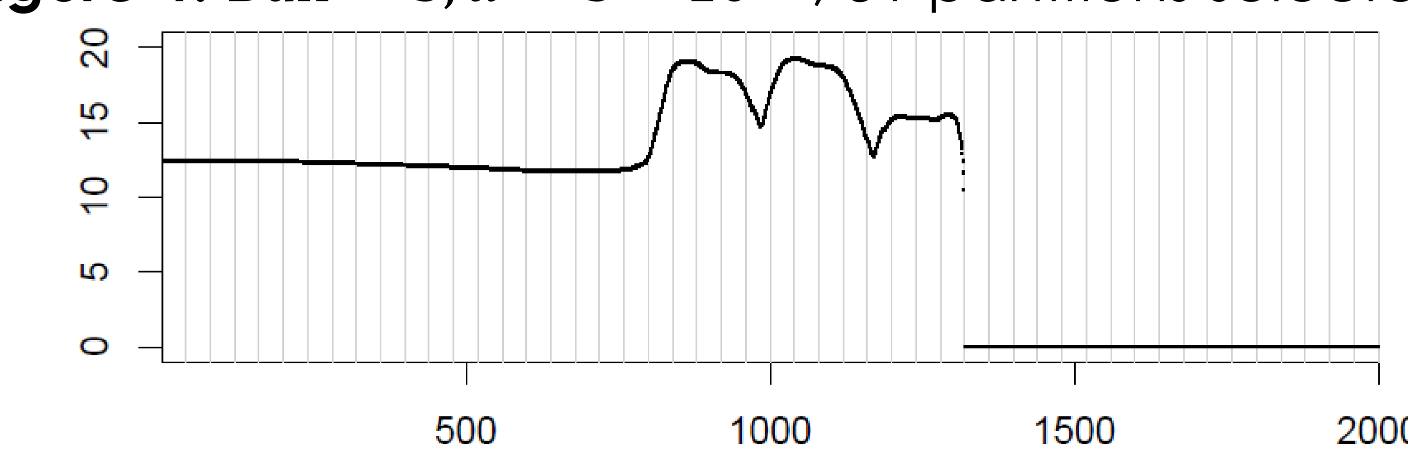

Figure 5: 50 evenly spaced partitions.

The total error of the linear approximations in Figure 4 is 1.16 while Figure 5 has an error of 208.23. Therefore, the piecewise linear fit approach has a smaller error and is more computationally efficient than storing all 2000 time steps of pressure in the LCROSS simulation.

Conclusions and Future Work

The piecewise linear fit method is a simple, computationally efficient, statistical method that can be:

- Embedded in a complex computer simulation to identify important time steps.
- Dramatically reduce data storage requirements.
- Facilitate later reconstruction and analysis of the simulation.

An obvious extension to consider is using a more complicated model than the piecewise linear fits of scalars.

Reference: Korycansky, D., et al. (2009). Predictions for the LCROSS mission. *Meteoritics & Planetary Science*, 44(4): 603-620.

LA-UR-15-25631