Eurographics Symposium on Parallel Graphics and Visualization (2016)
W. Bethel, E. Gobbetti (Editors)

Showcase: Data Mining Tornadogenesis Precursors

G. Foss[1] , A. McGovern[2] , C. Potvin[3] , G. Abram[1] , A. Bowen[1] , N. Hulkoti[1] and A. Kaul[1]

[1]Texas Advanced Computing Center, University of Texas at Austin
[2]School of Computer Science and School of Meteorology, University of Oklahoma
[3]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and National Oceanic and Atmospheric Administration/National Severe Storms Lab

Abstract

We investigate the value of 3-D visualization to data mining techniques for identifying tornadogenesis precursors in supercell thunderstorm simulations. We've found results will assist defining storm objects extracted and input to the data mining. The video shows samples of updrafts, downdrafts, cold pools, and regions of strong vorticity.

This submission results from an in-progress component of Amy McGovern's research in spatiotemporal data mining [MIW*14] to identify tornado precursors in simulated supercell thunderstorms, the origin of most violent and long-lived tornados. The data mining techniques operate on 4-D storm features (objects) extracted from the simulations and considered potentially important to whether or not a tornado forms. We explore the use of 3-D visualization techniques in creating and refining the definitions of these objects. The goal is for the automatically extracted objects to match visually identified objects.

Simulations are generated with the Bryan Cloud Model 1 (CM1) on 4096 cores for 4.5 hours on a Cray XC30 at the National Institute for Computational Sciences. Each approximately 5 TB dataset consists of 301 timesteps, valid every 30 seconds. This submission's data includes 9 scalar and 2 vector variables, structured on a rectilinear 1216x1216 grid with 100-meter horizontal resolution and 98 irregularly spaced vertical layers, increasing in depth from 40 meters at bottom to 500 meters at model top: 16985 meters. Intrigued by initial visualization results, we also calculated helicity and vertical gradients of pressure perturbations, both of which are important to supercells.

The data requires two operations to convert it from simulation output to a ParaViewer readable form: co-locating variables at grid vertices, and performing a transformation from grid to world coordinates. The latter converts the data from compact regular representation to an unstructured one. To avoid a pre-processing step producing substantially larger files, and to accommodate parallel I/O, we developed a custom reader which performs the conversion on the fly.

The project is a visually driven hunt through very complex 3-D data so we used an interactive approach with ParaView's GUI vs. its scripting interface. Freely orienting 2-D slices and clips in the 3-D environment revealed some patterns, while other possibly tornado-related factors appeared by comparing 3-D contours, volume rendering, and streamlines. A tolerable interactive approach requires optimum performance thus several variations of compute and memory were tried. For example, in serial, updating a timestep in one typical scene took over 2.5 minutes and soon inadequate memory caused ParaView to crash. Using 10 nodes, 1 process each, updated a timestep in under 40 seconds. The table shows rendering times of a representative scene/timestep, on TACC's Stampede visualization nodes (each 16 cores, 32 GB, and 1 NVIDIA Tesla K20m with a Mellanox FDR Infiniband interconnect).

nodes:	1	4	8	16	32
1 process:	267.35	105.18	45.68	26.83	17.11
4 processes:	219.88	103.81	42.89	20.22	13.42

Table 1: *serial (upper left) and parallel (in seconds)*

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. The visualization was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program.

References

[MIW*14] McGovern A., II D. G., Williams J., Brown R., Basara J.: Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning. *Machine Learning 95*, 1 (2014), 27–50. 1

DOI: 10.2312/pgv.20161186