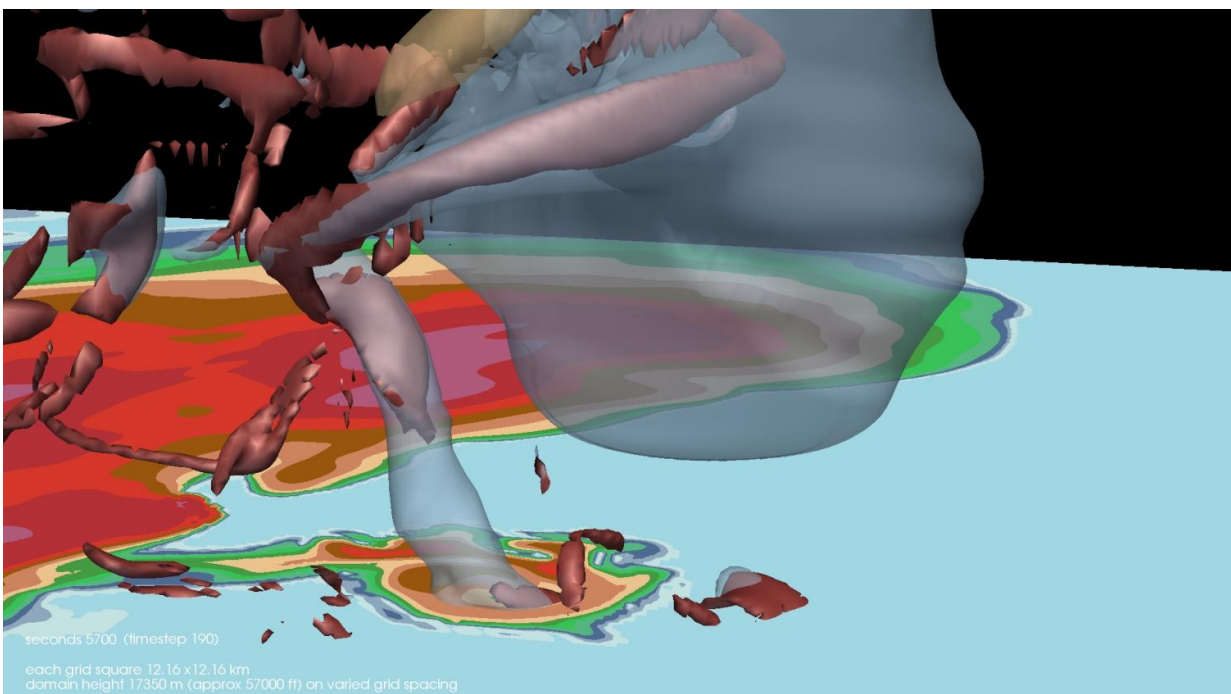# How XSEDE's ECSS program is helping researchers understand tornados with scientific visualization and data mining

October 6 2016

Negative gradient threshold (blue) depicts primary (big part) and low-level (lower tube part) updrafts, Strong vorticity (red) is wrapping into the primary updraft. The scientists hypothesize the evolution of this gradient differs between tornadic and tornado failure storms. Credit: Greg Foss, Texas Advanced Computing Center, The University of Texas at Austin

A dark, greenish sky...a loud roar, similar to a freight train...low-lying

clouds - if you see approaching storms or any of the danger signs, take shelter immediately. A tornado might be in your path!

Amy McGovern, a computer scientist at the University of Oklahoma, has been studying tornadoes, nature's most violent storms for eight years. She uses computational thinking to help understand and solve these scientific problems. Computational thinking is a way of solving problems, designing systems, and understanding human behavior that draws on concepts fundamental to computer science. In science and engineering, computational thinking is an essential part of the way people think about and understand the world.

"Since 2008, we've been trying to understand the formation of tornadoes, what causes tornadoes, and why some storms generate tornadoes and other storms don't," McGovern said. "Weather is a real world application where we can really make a difference to people." She wants to find solutions that are useful.

Specifically, she is trying to identify precursors of tornadoes in supercell simulations by generating high resolution simulations of these thunderstorms. Supercell storms, sometimes referred to as rotating thunderstorms, are a special kind of single cell thunderstorm that can persist for many hours. They are responsible for nearly all of the significant tornadoes produced in the U.S. and for most of the hailstorms larger than golf ball size. McGovern would like to generate as many as 100 different supercell simulations during this project.

In addition to high resolution simulations, McGovern is also using a combination of data mining and visualization techniques as she explores the factors that separate tornado formation from tornado failure.

Studying tornadoes and violent weather comes with a high learning curve, as it requires the application of science and technology to predict

the state of the atmosphere for a given location. When McGovern first started the research with a National Science Foundation (NSF) Career Grant, she had to attend several classes so that she would understand more about meteorology, the interdisciplinary scientific study of the atmosphere. She worked closely with meteorology students who taught her about the atmosphere, and she, in turn, taught them about computer science. They went back and forth until they understood each other.

The early research generated by the NSF Career Grant resulted in developing data mining software and developing initial techniques on lower resolution simulations.

"Now, we're trying to make high resolution simulations of super cell storms, or tornadoes," McGovern said. "What we get with the simulations are the fundamental variables of whatever our resolution is—we've been doing 100 meter x 100 meter cubes—there's no way you can get that kind of data without doing simulations. We're getting the fundamental variables like pressure, temperature and wind, and we're doing that for a lot of storms, some of which will generate tornadoes and some that won't. The idea is to do data mining and visualization to figure out what the difference is between the two."

Corey Potvin, a research scientist with the OU Cooperative Institute for Mesoscale Meteorological Studies and the NOAA National Severe Storms Laboratory, said: "I knew nothing about data mining until I started working with Amy on this project. I've enjoyed learning about the data mining techniques from Amy, and in turn teaching her about current understanding of tornadogenesis. It's a very fun and rewarding process. Both topics are so deep that you really need experts in both fields to tackle this problem successfully."

The process to do this research requires five steps: 1) Running the simulations; 2) Post-processing the simulation to merge the data; 3)

Visualizing the data (to ensure the simulation was valid); 4) Extracting the meta-data; and 5) Data mining (discovering patterns or new knowledge in very large data sets).

McGovern's research is related to the National Oceanic and Atmospheric Administration's (NOAA) Warn-on-Forecast program, tasked to increase tornado, severe thunderstorm, and flash flood warning lead times to reduce loss of life, injury, and damage to the economy. NOAA believes the current yearly-averaged tornado lead times are reaching a plateau, and a new approach is needed. "My ideal goal would be to find something that no one has thought of...discovering new science," McGovern said.

According to the National Weather Service, on average, nearly three out of every four tornado warning issues are false alarms. How do we reduce the false alarm rate for tornado prediction and increase warning lead time? This is a question that McGovern asks on a continual basis.

Right now the lead time is about 15 minutes on average for every tornado, but McGovern and team want to be able to better predict it. They're trying to do this by issuing warnings based on probabilities from the weather forecast rather than issuing warnings based on weather that is already about to form. "Once the weather is already starting to form, you won't get a two hour lead time," McGovern said.

How Is the Extreme Science and Engineering Discovery Environment (XSEDE) Helping?

When asked about XSEDE, McGovern replied: "XSEDE is fabulous. We've been using XSEDE resources for years. I started out with the resources at my university and then quickly outgrew what they had. They pointed me to XSEDE. I started out at NICS using Darter, and when that

went away, I started using Stampede at TACC. These resources are fundamental...you can't do this kind of data mining on your PC."

Stampede is one of the largest, most capable high-performance computing (HPC) systems in the world. McGovern says she is one of the few people that's using HPC and data mining for severe weather.

In addition to using XSEDE's Stampede for high resolution simulations, McGovern is taking advantage of XSEDE's Extended Collaborative Service and Support (ECSS) program. McGovern has been working with Greg Foss at the Texas Advanced Computing Center (TACC) for visualization expertise. ECSS experts like Foss are available for collaborations lasting months to a year to help researchers fundamentally advance their use of XSEDE resources. Foss is an expert in scientific visualization, a field devoted to graphically illustrating scientific data to enable scientists and non-scientists to understand, illustrate, and glean insight from their data.

"Greg comes at the problem from a completely different perspective, and provides new ways of looking at the data that you wouldn't have thought of in the beginning," McGovern noted. "Once you get into a domain, it's easy to think, 'This is the only way to look at it,' but then someone else comes along and asks, 'Why are you doing it like that?'"

Serving as a bridge between science research and the lay person, Foss says he enjoys working through XSEDE and highlighting the value and validity of the program. "I believe in our mission and I believe in the visualization field. It's quite a sense of accomplishment to help our users and even be a part of the science." For this project, Foss says that he's learned more about all of the aspects of weather than any of his six past weather projects.

"Ultimately, we're trying to discover if a 3-D visualization approach will

be a useful data mining tool for violent weather testbeds," Foss said.

The Process

Foss, McGovern and Potvin are working together to find storm features (objects) in the tornado simulations. Weather simulations must capture as much of the state of the atmosphere at each time step as possible, and this results in a tremendous amount of data. For example, in one of their first simulated data sets, McGovern had to sort through 352 million data points per time step.

"Since you can't save all of this data or mine it, you try to find the high level features because you don't want to do that for 100 simulations by hand, which is the traditional method of studying storms in meteorology. There's no way you can take traditional analysis techniques to that set and find an answer. Data mining is designed to help us sift through that data set and find statistical patterns that are causing tornadoes. The simulations are run in Fortran, the post-processing is in Python, and the data mining is in Java or Python," she said.

For Foss, this project is unique. "Instead of designing an interesting way to present the storm event, we're applying visualization techniques, ideas, and training to data mining. We're using the process to explore ways to identify individual storm features, and characteristics that wouldn't be found with other data mining methods," Foss said.

McGovern built the first dataset using Darter (decommissioned as of June 2016) at the National Institute for Computation Sciences. The beginning of the process was writing out different variables and transferring this huge dataset (approx. 5.7 TB from one simulation) to TACC. Then, Foss recruited another TACC visualization specialist, Greg Abram, to program a custom data reader for ParaView, and the visualization work could commence. "ParaView is the software I've been

using the most at TACC," he said. "Once I get the data and can read it correctly, my goal is to build scenes from different variables, looking for critical values and viewpoints that the researchers confirm are accurate and useful, and hopefully end up with something visually interesting," Foss said.

There are variables expected in a storm simulation, such as velocity (direction, distance). "This is the first project where I used velocity's vertical component to model strong updrafts, key indicators of violent weather. The visualization process isn't just building models- it's allowing a viewer to see the data, so it's important a scene communicates accurately, and doesn't mislead or confuse," Foss said.

"Visualizing these datasets is important because they are extremely complex, making it difficult to pull out the most physically important information, in this case, processes contributing to tornadogenesis," Potvin said. "The graphics will be used to help develop the storm feature (objects) definitions for the data mining, to ensure automatically extracted objects match visually and subjectively, and to develop definitions for new objects."

Potvin continued, "Using the visualizations to guide our object definitions is critical, he said. "We need the data mining to "know" about the storm features that most matter to whether a tornado forms so that it can tell us about the relationships between those features and tornadogenesis. My primary role in the project is to use my familiarity with current conceptual models of these storms to help isolate features known to be important to tornadoes, and to guide identification of features that aren't typically examined in simulations but that may actually play an important role in tornadogenesis."

So far, Foss's visualization work has identified six weather features, data mining 'objects' that can potentially be used to learn about tornados and

violent weather: hook echoes, bounded weak echo regions, updrafts, cold pools, helicity with regions of strong vorticity, and vertical pressure perturbation gradients. The results came from Foss exploring variables by experimenting with different values and various models, and looking for consistent patterns and interesting structures over the life of the simulated storm.

The goal is to compare these simulated storms with real storms. In real weather, you can't actually see these objects. The simulated data sets turn the tornado or storm into what would be actual objects. In real weather, you can't see an updraft, for example.

"Greg's visualizations are of much higher quality than what meteorologists typically use, since most of us lack the skills and computational resources," Potvin said. "The four-dimensionality and high resolution provide a much fuller perspective on how storms and tornadoes evolve. I've been studying these storms for over a decade, and these visualizations have changed the way I think about them."

In summary, the XSEDE ECSS team investigated computationally intensive datasets using 3D visualization techniques and an interactive user interface with datamining for identifying 'tornadogenesis' precursors in supercell thunderstorm simulations. They found that the results will assist in defining storm features extracted and input to the data mining: ensuring automatically extracted objects match visually identified ones.

"We strongly believe that using data science methods will enable us to discover new knowledge about the formation of tornados," McGovern concluded.

Provided by University of Texas at Austin

Citation: How XSEDE's ECSS program is helping researchers understand tornados with scientific visualization and data mining (2016, October 6) retrieved 3 May 2024 from https://phys.org/news/2016-10-xsede-ecss-tornados-scientific-visualization.html