

## Science Stories

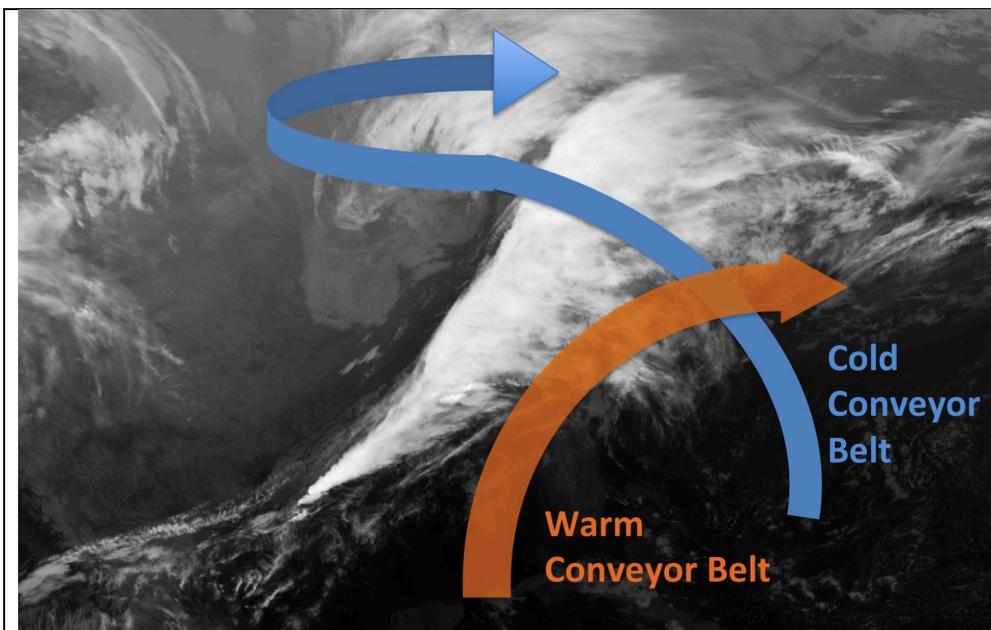
Discover More about XSEDE-enabled science, programs, trainings, and more.

### Key Points

• See how XSEDE supports science

« [Back \(https://www.xsede.org/news/science-stories?p\\_p\\_id=101\\_INSTANCE\\_9JovW1UTN10Q&p\\_p\\_lifecycle=0&p\\_p\\_state=normal&p\\_p\\_mode=view&p\\_p\\_col\\_id=column-2&p\\_p\\_col\\_count=1\)](https://www.xsede.org/news/science-stories?p_p_id=101_INSTANCE_9JovW1UTN10Q&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1)

## Dangerous Curves Ahead



Satellite image of a comma-shaped cloud showing how the warm conveyor belt (red) and cold conveyor belt (blue) winds combine to create the distinctive shape.

### Related Links

- [User News \(/user-news\)](#)
- [Newsletters \(/newsletters\)](#)
- [Podcasts \(/podcasts\)](#)
- [For Media \(/for-media\)](#)

### Contact Information

- [XSEDE External Relations \(mailto:er@xsede.org\)](mailto:er@xsede.org)
- [XSEDE Help Desk \(mailto:help@xsede.org\)](mailto:help@xsede.org)

### Machine Learning Enables Scientists to Spot "Comma-Shaped Clouds,"

#### Extreme Weather

By: Ken Chiacchia, Pittsburgh Supercomputing Center

Meteorologists can get time-critical help in spotting dangerous cloud formations using artificial intelligence (AI), according to scientists at Penn State and AccuWeather Inc. The team used a "machine learning" program running on XSEDE supercomputers to recognize a typical cloud formation known as comma-shaped clouds in satellite images. Their results detected up to 99 percent of the comma-shaped clouds and 64 percent of ensuing storms in 2011 and 2012 satellite images over the U.S. Their hope is to develop an accurate early warning system so that storm warnings can be issued more quickly than possible today.

#### Why It's Important:

Between 1980 and 2018, according to the National Oceanic and Atmospheric Administration, 238 particularly severe U.S. weather and climate disasters—including severe storms, floods and other weather problems—caused more than \$1.5 trillion in damages. Severe storms alone kill hundreds of people a year. It's safe to say that quicker storm warnings would save lives and money.

When meteorologists started using satellite photos to help them predict the weather in the 1970s, they got a surprise. Before a storm developed, they saw, the clouds in that area took on the shape of a comma. These comma-shaped clouds varied in size from tiny to enormous, with some spanning the continental U.S. from north to south. Before long, though, the meteorologists discovered they had too much of a good thing. The data from satellite photos expanded enormously, quickly overwhelming any ability for humans to monitor them in real time. The big comma-shaped formations were obvious—but smaller ones could still threaten life and property, and were easy to miss.

*"As computer scientists, we were really fascinated by this new field of data science, which we could leverage to make the computers more efficient in solving real-world problems. The issue was to define a problem that had enough data to work with. But not just a lot of data. The data needed to be manually labeled, in order for the computer to learn ... Meteorology is one of the rare fields where you can find a huge amount of data that are so complex people don't know how to leverage it—and the data are indexed [labeled]."*—James Wang, Penn State

Enter James Wang of Pennsylvania State University. Wang is a computer scientist who uses AI to solve "Big Data" problems. Around 2008, he was looking for a good scientific application to inspire and test machine-learning technology, in which the computer starts from data that has been labeled by humans, and experiments many, many times on new data until it can reproduce the humans' performance. Detecting severe thunderstorms was a perfect problem. The data are enormous, with tens of thousands of weather stations worldwide and satellites generating gigabytes of high-resolution images, radar data and other data every day. But they have also been carefully reviewed and labeled with "ground truth" by expert meteorologists. Plenty of accurate examples of developing storms were available for the computer to start with.

#### **How XSEDE Helped:**

Following up on earlier work by Wang's team, and working with Stephen Wistar and his colleagues at AccuWeather Inc., Wang and his graduate student, Rachel Zheng, started with the satellite images and storm records from the years 2008, 2011, and 2012. It was a perfect machine learning dataset. With the help of meteorologists, the team had labeled all the comma-shaped clouds. The storms were a matter of record. But it was also a huge dataset, with more than 50,000 images. To take this problem on, the team realized that they needed a supercomputer with a particular mix of power and compatibility with Big Data. They turned to the XSEDE-allocated Bridges system at PSC. Bridges was designed to be particularly friendly to data science. The scientists also employed the XSEDE-allocated system Comet at the San Diego Supercomputer Center at UC San Diego to build up a baseline model of their "machine learning" artificial intelligence (AI) program.

Zheng used the first 250 days of 2008 as the training data, with the comma-shaped clouds and storms labeled. That allowed the AI program to learn how to recognize comma-shaped clouds, as well as to spot signs that a comma-shaped cloud would develop into a storm. These formations are easy for humans to identify, but they vary just enough from cloud to cloud that learning to spot them was a major task for the machine. Because even an AI running on a supercomputer can't process all the weather data from that span of time, Zheng made the AI focus on two factors. One was the shape of a given cloud system. The other was how it changed over time. The two together proved reliable indications of a comma-shaped cloud.

*"The Bridges system works very well with machine learning and image processing. As a result, it took less than 40 seconds for the trained system to process the image data and get a result ... Our experience was very good because you have this CPU and large-memory 'multiple choice' that I can choose for a specific job. Even though I process the two steps on different nodes, the data can transfer very quickly."*—Rachel Zheng, Penn State

Zheng used Bridges' CPU nodes, which are particularly good for the machine learning techniques used in the work. She also used the system's large-memory nodes, which allowed her to process big image files quickly. After her program had learned to spot comma-shaped clouds, she used the last 116 days of data from 2008 as a cross-validation set, to make sure the learning had worked properly. Finally, she let the program loose on the 2011 and 2012 data, without any labels, to see how well it predicted the known storms of those years.

*"I still do one operational forecast a week at AccuWeather. We're using all kinds of data, all kinds of screens, taking phone calls from clients ... [The machine learning] technique could be useful for calling attention to, say, a comma-shaped disturbance along a cold front in Georgia that none of us saw because we were looking at the snow storm in Michigan. It won't replace meteorologists, but it's a tool that can say, 'You'd better look over here, because something's going on ... that four hours from now will be a severe thunderstorm.'"—Stephen Wistar, AccuWeather Inc.*

The results were encouraging. The program detected over 99 percent of the comma-shaped clouds correctly. This is an important number. If the machine is going to do a first pass on the data before the humans see it, it can't miss many of them. Equally encouraging, by comparing comma-shaped clouds that did and didn't lead to storms in 2008, the AI predicted where storms were going to form in the 2011-2012 data more than 64 percent of the time, using satellite images alone. Best of all, the system could produce its predictions in about 40 seconds. That's fast enough to really help meteorologists spot storm risks, and way faster than humans could review the images. The scientists reported their results in the journal *IEEE Transactions on Geoscience and Remote Sensing* in January 2019. Currently, they're using Bridges' graphics processing units (GPUs) to investigate whether a type of AI called deep learning, which involves the machine investigating many layers of factors, can improve on their results, particularly when multiple types of weather data are considered together.

Tweet Like 0

5/21/19

successes (<https://www.xsede.org/news/science-stories?successes>)  
 psc (<https://www.xsede.org/news/science-stories?psc>)  
 bridges (<https://www.xsede.org/news/science-stories?bridges>)  
 meteorology (<https://www.xsede.org/news/science-stories?meteorology>)  
 comet (<https://www.xsede.org/news/science-stories?comet>)  
 sdsc (<https://www.xsede.org/news/science-stories?sdsc>)  
 ai (<https://www.xsede.org/news/science-stories?ai>)  
 earth science (<https://www.xsede.org/news/science-stories?earth+science>)



© 2018 XSEDE All Rights Reserved  
 The Extreme Science and Engineering Discovery Environment (XSEDE)  
 is supported by the National Science Foundation

Contact Consulting Ticket Site Map  
 (mailto:help@xsede.org) (mailto:help@xsede.org) (web/site/site-map)