

Simulation and Analysis of Severe Thunderstorms on Blue Waters

Kelton Halbert

University of Wisconsin - Madison
Cooperative Institute for Meteorological Satellite Studies

On behalf of PI **Dr. Leigh Orf**
Blue Waters Symposium
Sunriver, OR
June 4, 2019



We use Blue Waters to simulate violent,
long-lived tornadoes



The Problem...

- Supercell thunderstorms produce the strongest (EF-4 and EF-5) tornadoes, which are responsible for the largest losses to life and property.
- We still do not fully understand why a small subset of supercells produce violent, long-track tornadoes. Less than 1% of tornadoes are classified as violent.
- Even worse, the general processes that lead to any tornado development are not well understood.
- Collecting observations from supercells/tornadoes is challenging and often temporally or spatially coarse.
- Goal: Use high-resolution simulations to understand the internal workings of supercells that produce violent tornadoes well enough that the forecasting of these events may eventually be improved.

The Problem...

- Supercell thunderstorms produce the strongest (EF-4 and EF-5) tornadoes, which are responsible for the largest losses to life and property.
- We still do not fully understand why a small subset of supercells produce violent, long-track tornadoes. Less than 1% of tornadoes are classified as violent.
- Even worse, the general processes that lead to any tornado development are not well understood.
- Collecting observations from supercells/tornadoes is challenging and often temporally or spatially coarse.
- Goal: Use high-resolution simulations to understand the internal workings of supercells that produce violent tornadoes well enough that the forecasting of these events may eventually be improved.

The Problem...

- Supercell thunderstorms produce the strongest (EF-4 and EF-5) tornadoes, which are responsible for the largest losses to life and property.
- We still do not fully understand why a small subset of supercells produce violent, long-track tornadoes. Less than 1% of tornadoes are classified as violent.
- Even worse, the general processes that lead to any tornado development are not well understood.
- Collecting observations from supercells/tornadoes is challenging and often temporally or spatially coarse.
- Goal: Use high-resolution simulations to understand the internal workings of supercells that produce violent tornadoes well enough that the forecasting of these events may eventually be improved.

The Problem...

- Supercell thunderstorms produce the strongest (EF-4 and EF-5) tornadoes, which are responsible for the largest losses to life and property.
- We still do not fully understand why a small subset of supercells produce violent, long-track tornadoes. Less than 1% of tornadoes are classified as violent.
- Even worse, the general processes that lead to any tornado development are not well understood.
- Collecting observations from supercells/tornadoes is challenging and often temporally or spatially coarse.
- Goal: Use high-resolution simulations to understand the internal workings of supercells that produce violent tornadoes well enough that the forecasting of these events may eventually be improved.

The Problem...

- Supercell thunderstorms produce the strongest (EF-4 and EF-5) tornadoes, which are responsible for the largest losses to life and property.
- We still do not fully understand why a small subset of supercells produce violent, long-track tornadoes. Less than 1% of tornadoes are classified as violent.
- Even worse, the general processes that lead to any tornado development are not well understood.
- Collecting observations from supercells/tornadoes is challenging and often temporally or spatially coarse.
- Goal: Use high-resolution simulations to understand the internal workings of supercells that produce violent tornadoes well enough that the forecasting of these events may eventually be improved.

Why It Matters...

- Currently, the National Weather Service false alarm rate for tornado warnings is about 70%.
- While incredibly rare, there are still cases where tornadoes occur without warning.
- Majority of tornado related fatalities are from violent (EF-4 and EF-5) tornadoes.
- Despite advances in operational and computational meteorology, the processes that create violent tornadoes are obscured due to lack of data resolution... until Blue Waters.

Why It Matters

Why It Matters...

- Currently, the National Weather Service false alarm rate for tornado warnings is about 70%.
- While incredibly rare, there are still cases where tornadoes occur without warning.
- Majority of tornado related fatalities are from violent (EF-4 and EF-5) tornadoes.
- Despite advances in operational and computational meteorology, the processes that create violent tornadoes are obscured due to lack of data resolution... until Blue Waters.

Why It Matters

Why It Matters...

- Currently, the National Weather Service false alarm rate for tornado warnings is about 70%.
- While incredibly rare, there are still cases where tornadoes occur without warning.
- Majority of tornado related fatalities are from violent (EF-4 and EF-5) tornadoes.
- Despite advances in operational and computational meteorology, the processes that create violent tornadoes are obscured due to lack of data resolution... until Blue Waters.

Why It Matters...

- Currently, the National Weather Service false alarm rate for tornado warnings is about 70%.
- While incredibly rare, there are still cases where tornadoes occur without warning.
- Majority of tornado related fatalities are from violent (EF-4 and EF-5) tornadoes.
- Despite advances in operational and computational meteorology, the processes that create violent tornadoes are obscured due to lack of data resolution... until Blue Waters.

The Challenges...

- Data Storage: Not everything needs to be stored with 32 or 64 bits of floating point precision. The ZFP plugin for HDF5 allows for us to specify a minimum desired accuracy for a given array.
- Model I/O: Modifications to the model I/O driver were necessary for performance. Instead of communicating arrays when writing, each node dumps its tile to disk as an HDF5 file. Each HDF5 file is self describing of the whole dataset, meaning it can be patched back together. This set of tools to write and reconstruct the dataset is called LOFS.
- In order to save space, only the bare minimum fields are saved to disk. Important derived fields are calculated post-hoc with LOFS and GPU/CUDA code.
- Analysis: In order to do quantitative analysis with passive Lagrangian tracers, GPU/CUDA code was developed to integrate millions of trajectories on the data that's saved every model time step.

The Challenges...

- Data Storage: Not everything needs to be stored with 32 or 64 bits of floating

Data Storage: Not everything needs to be stored with 32 or 64 bits of floating point precision. The ZFP plugin for HDF5 allows for us to specify a minimum desired accuracy for a given array.

- Model I/O: Modifications to the model I/O driver were necessary for performance. Instead of communicating arrays when writing, each node dumps its tile to disk as an HDF5 file. Each HDF5 file is self describing of the whole dataset, meaning it can be patched back together. This set of tools to write and reconstruct the dataset is called LOFS.
- In order to save space, only the bare minimum fields are saved to disk. Important derived fields are calculated post-hoc with LOFS and GPU/CUDA code.
- Analysis: In order to do quantitative analysis with passive Lagrangian tracers, GPU/CUDA code was developed to integrate millions of trajectories on the data that's saved every model time step.

The Challenges...

- Data Storage: Not everything needs to be stored with 32 or 64 bits of floating

point precision. The ZFP plugin for HDF5 allows for us to specify a minimum desired accuracy for a given array.

- Model I/O: Modifications to the model I/O driver were necessary for performance. Instead of communicating arrays when writing, each node dumps its tile to disk as an HDF5 file. Each HDF5 file is self describing of the whole dataset, meaning it can be patched back together. This set of tools to write and reconstruct the dataset is called LOFS.
- In order to save space, only the bare minimum fields are saved to disk. Important derived fields are calculated post-hoc with LOFS and GPU/CUDA code.
- Analysis: In order to do quantitative analysis with passive Lagrangian tracers, GPU/CUDA code was developed to integrate millions of trajectories on the data that's saved every model time step.

The Challenges...

- Data Storage: Not everything needs to be stored with 32 or 64 bits of floating

point precision. The ZFP plugin for HDF5 allows for us to specify a minimum desired accuracy for a given array.

- Model I/O: Modifications to the model I/O driver were necessary for performance. Instead of communicating arrays when writing, each node dumps its tile to disk as an HDF5 file. Each HDF5 file is self describing of the whole dataset, meaning it can be patched back together. This set of tools to write and reconstruct the dataset is called LOFS.
- In order to save space, only the bare minimum fields are saved to disk. Important derived fields are calculated post-hoc with LOFS and GPU/CUDA code.
- Analysis: In order to do quantitative analysis with passive Lagrangian tracers, GPU/CUDA code was developed to integrate millions of trajectories on the data that's saved every model time step.

Previous Work on Blue Waters...

Initial work focused on modifying the model (CM1) to improve I/O performance

- Initial work focused on modifying the model (CMI) to improve I/O performance (LOFS) and set up an efficient data visualization pipeline
- At the end of the first allocation in 2013, a successful simulation of an EF-5 tornado was produced at 30 meter resolution.
- Our 30 meter resolution dataset is 67 TB, containing 32,400 times saved every model time step throughout the life of the EF-5 tornado.
- Additional simulations of the same environment have been done at 20 and 15 meters
- Simulation of new supercell environments have yielded another EF-5 tornado (April 27, 2011 tornado outbreak; e.g. Tuscaloosa, AL)

Previous Work on Blue Waters...

- Initial work focused on modifying the model (CMI) to improve I/O performance

(LOFS) and set up an efficient data visualization pipeline

- At the end of the first allocation in 2013, a successful simulation of an EF-5 tornado was produced at 30 meter resolution.
- Our 30 meter resolution dataset is 67 TB, containing 32,400 times saved every model time step throughout the life of the EF-5 tornado.
- Additional simulations of the same environment have been done at 20 and 15 meters
- Simulation of new supercell environments have yielded another EF-5 tornado (April 27, 2011 tornado outbreak; e.g. Tuscaloosa, AL)

Previous Work on Blue Waters...

- Initial work focused on modifying the model (CMI) to improve I/O performance

(LOFS) and set up an efficient data visualization pipeline

- At the end of the first allocation in 2013, a successful simulation of an EF-5 tornado was produced at 30 meter resolution.
- Our 30 meter resolution dataset is 67 TB, containing 32,400 times saved every model time step throughout the life of the EF-5 tornado.
- Additional simulations of the same environment have been done at 20 and 15 meters
- Simulation of new supercell environments have yielded another EF-5 tornado (April 27, 2011 tornado outbreak; e.g. Tuscaloosa, AL)

Previous Work on Blue Waters...

- Initial work focused on modifying the model (CMI) to improve I/O performance

(LOFS) and set up an efficient data visualization pipeline

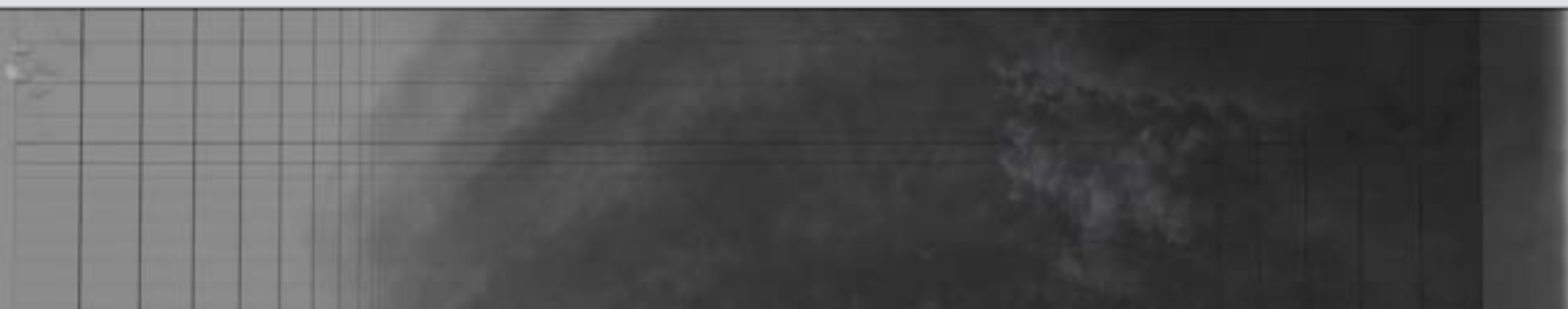
- At the end of the first allocation in 2013, a successful simulation of an EF-5 tornado was produced at 30 meter resolution.
- Our 30 meter resolution dataset is 67 TB, containing 32,400 times saved every model time step throughout the life of the EF-5 tornado.
- Additional simulations of the same environment have been done at 20 and 15 meters
- Simulation of new supercell environments have yielded another EF-5 tornado (April 27, 2011 tornado outbreak; e.g. Tuscaloosa, AL)

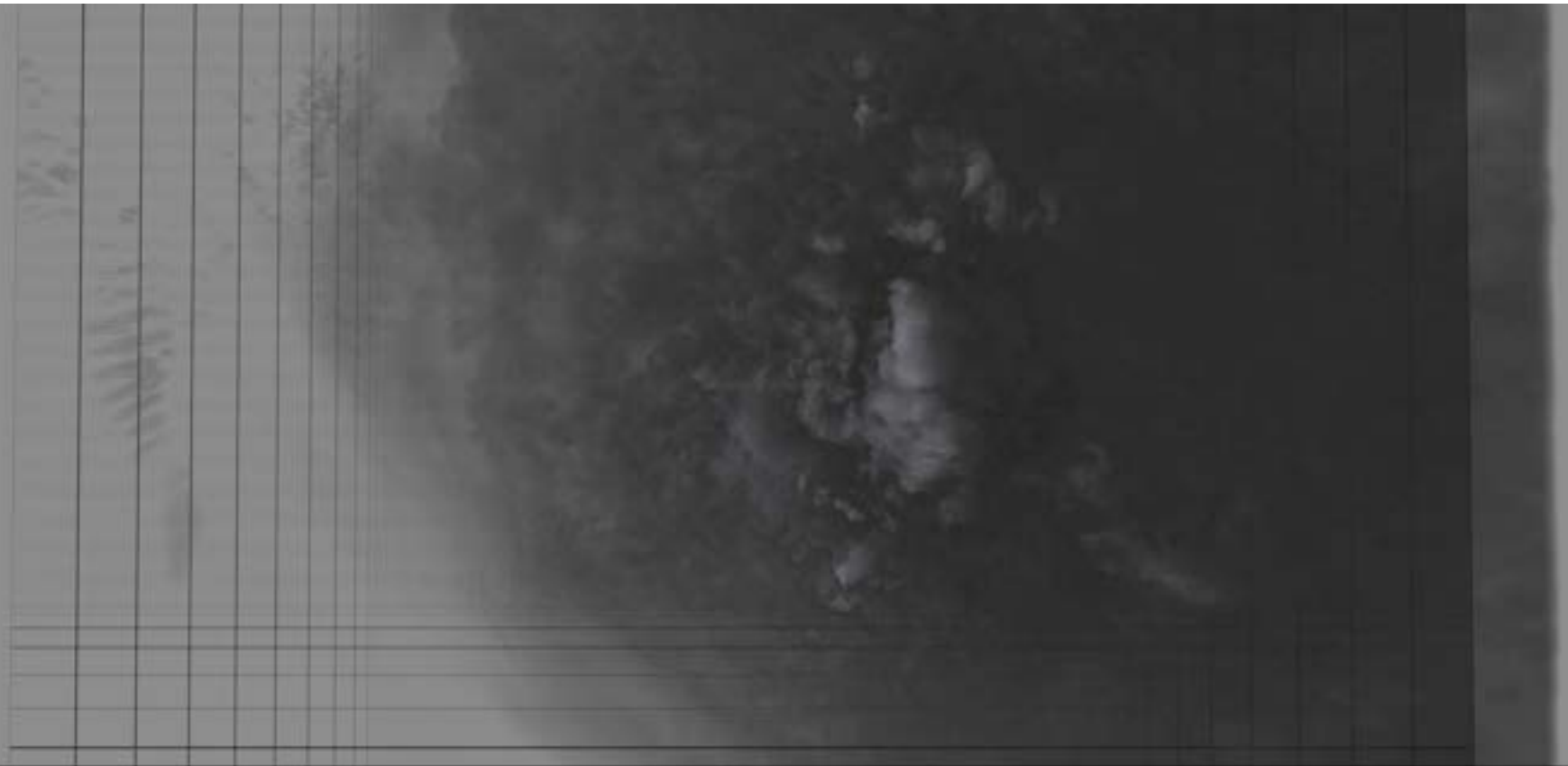
Previous Work on Blue Waters...

- Initial work focused on modifying the model (CMI) to improve I/O performance (LOFS) and set up an efficient data visualization pipeline

(EOR) and set up an efficient data visualization pipeline

- At the end of the first allocation in 2013, a successful simulation of an EF-5 tornado was produced at 30 meter resolution.
- Our 30 meter resolution dataset is 67 TB, containing 32,400 times saved every model time step throughout the life of the EF-5 tornado.
- Additional simulations of the same environment have been done at 20 and 15 meters
- Simulation of new supercell environments have yielded another EF-5 tornado (April 27, 2011 tornado outbreak; e.g. Tuscaloosa, AL)





Current Work on Blue Waters...

- Kick everything up to 11...
- New "hero" simulation at 10 meter resolution

- 112km x 112km x 20km domain
- 250,000,000,000 grid points
- 1/20s model time step
- A single model restart checkpoint weighs ~35 TB. Turning on gzip compression reduced to ~10TB.
- Started running in increments in April, only in the last few weeks has the simulation reached the point where it produces the tornado. Jobs use over half of Blue Waters. Simulation still in progress - currently at T=6,000s.
- More than just a "one and done" simulation, data will also be saved for post processing via GPU code and the custom HDF5 file system LOFS.

Current Work on Blue Waters...

- Kick everything up to 11...
- New "hero" simulation at 10 meter resolution

- 112km x 112km x 20km domain
- 250,000,000,000 grid points
- 1/20s model time step

- A single model restart checkpoint weighs ~35 TB. Turning on gzip compression reduced to ~10TB.

- Started running in increments in April, only in the last few weeks has the simulation reached the point where it produces the tornado. Jobs use over half of Blue Waters. Simulation still in progress - currently at T=6,000s.

- More than just a "one and done" simulation, data will also be saved for post processing via GPU code and the custom HDF5 file system LOFS.

Current Work on Blue Waters...

- Kick everything up to 11...

- New "hero" simulation at 10 meter resolution
 - 112km x 112km x 20km domain

- 250,000,000,000 grid points
- 1/20s model time step
- A single model restart checkpoint weighs ~35 TB. Turning on gzip compression reduced to ~10TB.
- Started running in increments in April, only in the last few weeks has the simulation reached the point where it produces the tornado. Jobs use over half of Blue Waters. Simulation still in progress - currently at $T=6,000s$.
- More than just a "one and done" simulation, data will also be saved for post processing via GPU code and the custom HDF5 file system LOFS.

Current Work on Blue Waters...

- Kick everything up to 11...
- New "hero" simulation at 10 meter resolution
 - 112km x 112km x 20km domain

- 250,000,000,000 grid points
- 1/20s model time step
- A single model restart checkpoint weighs ~35 TB. Turning on gzip compression reduced to ~10TB.
- Started running in increments in April, only in the last few weeks has the simulation reached the point where it produces the tornado. Jobs use over half of Blue Waters. Simulation still in progress - currently at T=6,000s.
- More than just a “one and done” simulation, data will also be saved for post processing via GPU code and the custom HDF5 file system LOFS.

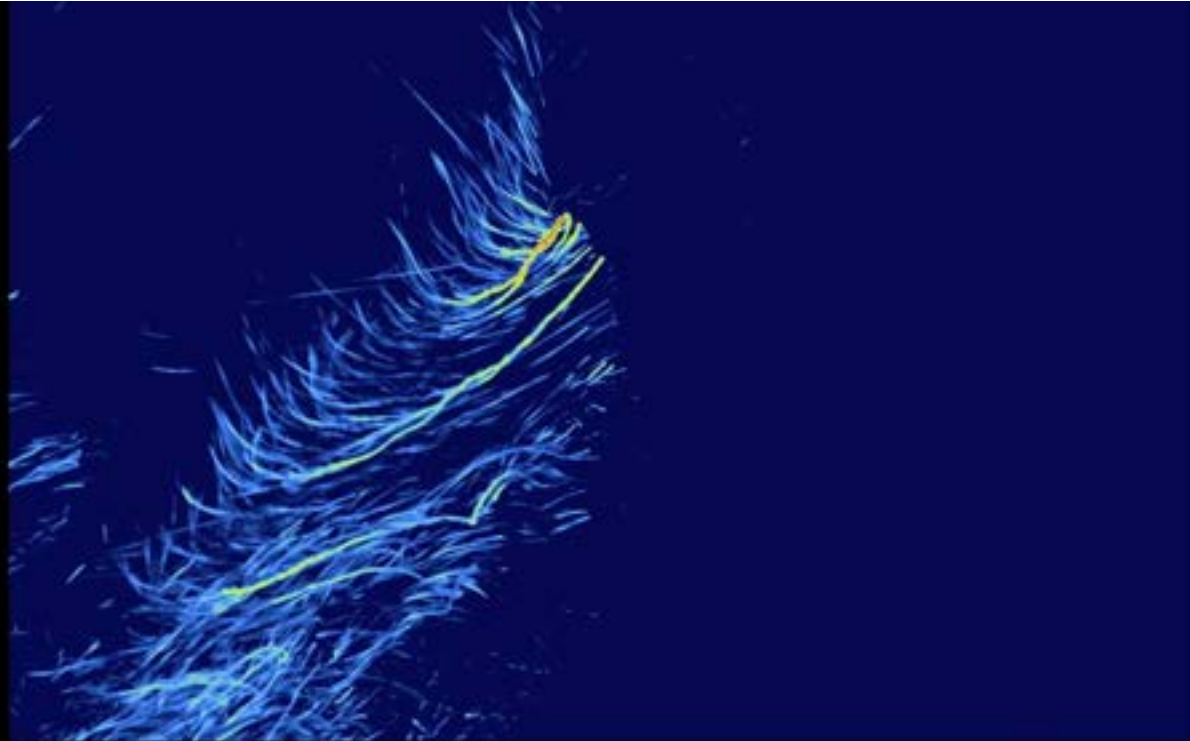
Current Work on Blue Waters...

- Kick everything up to 11...
- New “hero” simulation at 10 meter resolution
 - 112km x 112km x 20km domain

- 250,000,000,000 grid points
- 1/20s model time step
- A single model restart checkpoint weighs ~ 35 TB. Turning on gzip compression reduced to ~ 10 TB.
- Started running in increments in April, only in the last few weeks has the simulation reached the point where it produces the tornado. Jobs use over half of Blue Waters. Simulation still in progress - currently at $T=6,000$ s.
- More than just a “one and done” simulation, data will also be saved for post processing via GPU code and the custom HDF5 file system LOFS.

**Vertical vorticity maxima
at lowest level**





Some Parting Thoughts...

- Blue Waters has enabled breakthrough science in supercell/tornado modeling. As

of now, the 10 meter simulation is the highest resolution study of a supercell and tornado ever conducted.

- Blue Waters staff was instrumental in getting things like ZFP compiled, handling node failures, reservations, and more.
- Blue Waters will continue to be instrumental in creating and saving this 10 meter simulation dataset, and possibly more.

Acknowledgements...

This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort at the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications (NCSA). Leigh Orf is partly supported by NSF

supercomputing Applications (NCSA). Leigh Oris is partly supported by NSF grants 1550405, 1514973, 1663954, 1832327, and the Space Science and Engineering Center at the University of Wisconsin - Madison.

Collaborators on this and related projects: Dr. Catherine Finley (UND); Dr. Bruce Lee; Dr. Robert Wilhelmson (UIUC/NCSA); Dr. Lawrence Frank (UCSD); Dr. Morgan O'Neill (Stanford); Hank Schyma ("Pecos Hank"), storm videographer from Houston; Kelton Halbert and Austin Dixon, UW grad students, Michelle Elmore, SLU grad student.

Questions?

Simulation and Analysis of Severe Thunderstorms on Blue Waters

Kelton Halbert

University of Wisconsin - Madison

Cooperative Institute for Meteorological Satellite Studies

On behalf of PI **Dr. Leigh Orf**
Blue Waters Symposium
Sunriver, OR
June 4, 2019



We use Blue Waters to simulate violent,
long-lived tornadoes

