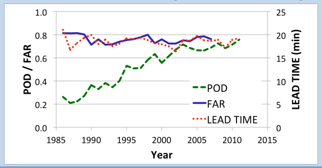
Using Machine Learning Techniques to Investigate Tornadogenesis

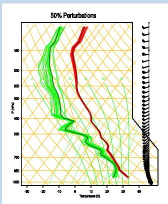
Amy McGovern[1], Corey K. Potvin[2], Rodger A. Brown[3]
1: Schools of Computer Science and Meteorology, University of Oklahoma, amcgovern@ou.edu
2. CIMMS/NSSL 3: NOAA/National Severe Storms Laboratory (NSSL)

Motivation

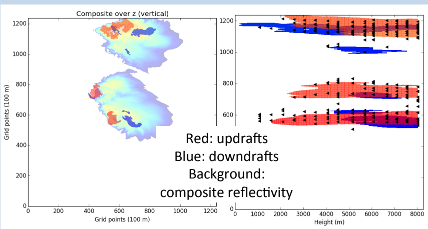
- Tornado warning lead times and POD have stalled in recent years
- Computational capabilities have increased, providing finer resolution simulations, but data cannot be analyzed using traditional techniques
- Machine learning & data mining can find patterns in large data sets
- **Goal: identify precursors of tornadoes in high-resolution simulations of supercells and tornadoes using machine learning/data mining**



Method



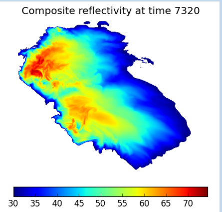
- Generate high-resolution simulations of supercell thunderstorms by varying initial environmental conditions from known tornadic environments (Cintineo and Stensrud 2013, Dahl 2014)
- Identify and extract high-level objects



- Apply machine learning/data mining to the high-level data
- Analyze and verify results

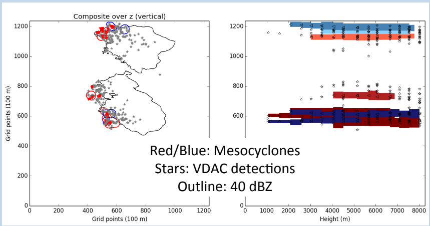
High-resolution simulations

- CM1 (Bryan and Fritsch 2002)
- 100 m horizontal resolution in center (125 km)
- Nested grid to 400 m at edges
- Stretched vertical grid (40 m to 500 m)
- 1536 x 1536 x 99 ≈ 234 million model grid points
- Data saved every grid point every 30 seconds of simulated time



Identifying High-level objects

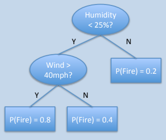
- Most objects extracted using thresholds
- Rotating objects use VDAC (Potvin, 2013)
- Mesocylone requires: 1 detection in 2-4 km and at least 4 total in 0-8 km, all greater than 750 m in diameter
- Tornado object requires: 1 VDAC detection at 100 m and at least one more detection between 500 m and 3km, all with radius less than 500 m



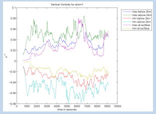
- Tornadic storm: low-level mesocyclone object (lowest height <= 2 km) co-located with a tornado object, both of which last at least 60 seconds
- Tornado-failure storm: low-level mesocylone object without a tornado object
- Training data taken from 30 minutes prior to start of tornado or low-level mesocyclone

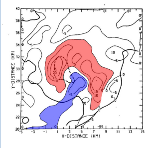
Name	Definition
Updraft	$w \geq 15 ms^{-1}$
Downdraft	$w \leq -6 ms^{-1}$
Radar reflectivity	$dBZ \geq 40$
Intense reflectivity	$dBZ \geq 60$
Cold pool	$\Theta' \leq -2°K$
Positive/negative vertical vorticity	$\zeta \geq 0.05 s^{-1}$
Convergence	$\nabla \geq 0.05 s^{-1}$
Positive/negative tilting	$T \geq 0.03 s^{-2}$
Positive/negative stretching	$S \geq 0.03 s^{-1}$
Positive/negative mesocyclone	VDAC (10 ms^{-1})
Positive/negative tornado	VDAC (10 ms^{-1})

Spatiotemporal Relational Random Forests



- SRRF is a Random Forest style algorithm composed of Spatiotemporal Relational Probability Trees (SRPT) (McGovern et al 2014 & to appear)
- SRPT: Decision tree type algorithm, probability estimation tree
- Each tree can make spatial, temporal, and spatiotemporal splits
  - Scalar Attribute: Is there an updraft with maximum speeds >= 25 m/s?
  - Temporal Exists: Is there a storm object that lasted at least 20 minutes?
  - Array Attribute: Did the updraft ever get larger than 20km^3?
  - Array Attribute Temporal Duration: Did the updraft maintain a volume of at least 20 km^3 for 5 minutes?
  - Array Partial Derivative: Did the strength of the updraft double within 5 minutes?
  - Scalar Field Attribute: Is the temperature in the near storm environment always greater than 80?
  - Shapes:
    - Can split on the shape of 2 and 3 dimensional objects
    - Can split on how 2 and 3D shapes change over time
    - Can split on the spatial relationships between 2 and 3D shapes
  - Time:
    - Can split on the temporal relationship between pairs of objects
    - Did a downdraft with a strength of at least -10 ms^{-1} appear before the updraft reached a strength of 20 ms^{-1}?



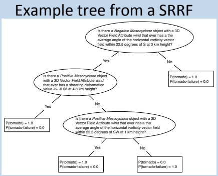


Adapted from Brandes (1984)

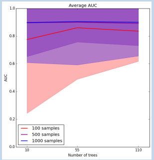
Preliminary Results

- Data: 34 tornadic storms, 21 tornado-failures
- Results averaged over 30 runs (different training/test sets)
- Measured performance using Area Under Receiver Operator Curve (AUC)

Example tree from a SRRF



Overall performance



Most frequently chosen question types and objects

Question type	Object	Percentage
Vector Field Vorticity	Positive Mesocyclone	24.2%
Vector Field Vorticity	Negative Mesocyclone	12.7%
Vector Field Deformation	Positive Mesocyclone	10.4%
Vector Field Divergence	Positive Mesocyclone	6.5%
Vector Field Deformation	Negative Mesocyclone	5.1%
Array Temporal Duration	Updraft	5.1%
Array Shapelet	Updraft	4.8%
Vector Field Vorticity	Updraft	3.4%

Current work

- Refining object definitions (tornadoes, gust fronts, etc)
- Generating additional simulations of both strong and weak storms
- Refining our post-processing and data mining pipeline to handle the volume of big data

References

- Bryan, G. H., and J. M. Fritsch, 2002: A benchmark simulation for moist nonhydrostatic numerical models. Mon. Wea. Rev., 130, 2917–2928.
- Cintineo, R. M., and D. J. Stensrud, 2013: On the predictability of supercell thunderstorm evolution. J. Atmos. Sci., 70, 1993–2011.
- Dahl, B., 2014: Sensitivity of Vortex Production to Small Environmental Perturbations in High-Resolution. Supercell Simulations. School of Meteorology, University of Oklahoma.
- McGovern, Amy and Gagne II, David J. and Williams, John K. and Brown, Rodger A. and Basara, Jeffrey B. , 2014. Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning. Machine Learning. Vol 95, Issue 1, pages 27-50.
- McGovern, Amy and Potvin, Corey and Brown, Rodger A., to appear. Using Large-scale Machine Learning to Improve our Understanding of the Formation of Tornadoes, invited chapter in Large-Scale Machine Learning in the Earth Sciences.
- Potvin, C. K., 2013: A variational method for detecting and characterizing convective vortices in Cartesian wind fields. Mon. Wea. Rev., 141, 3102–3115.

Acknowledgements

This work was funded by the National Science Foundation under Grant No. NSF/IIS/0746816 and related REU supplements and NSF/AGS/1261776 and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce.