Data Mining of High-Resolution Storm-Scale Datasets

PIs: Travis Smith, Chris Karstens, Jimmy Correia, Kiel Ortega
U. Of Oklahoma / CIMMS (+ NOAA/NSSL&SPC)
NOAA Collaborators

• Lans Rothfusz, Alan Gerard (National Severe Storms Laboratory / Hazardous Weather Testbed)
• NWS Storm Prediction Center staff
• many NWSFO staff
• NCEI
Current warning paradigm:

- Make extrapolative prediction based on radar and storm spotter observations
- “Warn-on-Detection”

Forecasting A Continuum of Environmental Threats (FACETs):

- Continuously updating flow of information
- Storm-scale ensembles (“Warn on Forecast”)
- Probabilistic Hazard Information
Goals

Measure model improvements at the scale of individual thunderstorms (short term predictability).

Understand the strengths and limitation in the models’ simulation of storms and storm evolution over a diverse spectrum of convective modes.

Demonstrate in Hazardous Weather Testbed.
Project Plan

1. Acquire and prepare data sets
2. Develop and refine software
3. Objectively classify convective storms
   • Observational data
   • Model output
4. Compare observed and model storm-typing and severity
5. Real-time testing and evaluation with forecasters
Acquire/Prepare Data Sets

Multi-Year Reanalysis of Remotely Sensed Storms (MYRORSS; pronounced “mirrors”)

Storm Prediction Center database of convective modes

NSSL WRF storm-object data set
Multi-Radar Multi-Sensor System

• Radars in network supplement each other:
  • Overlapping coverage
  • Fills in gaps in radar coverage
  • Increased sampling frequency
  • Seamless, consistent

+ other neighboring radars
Examples: Multi-sensor data fields

- Show physical relationships between data fields from multiple sensors
- Storm tracks and trends can be generated at any spatial scale, for any data fields
5-minute, 0.01x0.01 degree resolution, 35 vertical levels
Multi-Year Reanalysis of Remotely Sensed Storms (MYRORSS)

- 15+ years of storm statistics
- Data Mining
- MRMS & MYRORSS are foundational to the effort to improve NWS warning services

Days per Year: Any Hail

Days per Year: Severe Hail
Storm classification

Disorganized
- Discrete
  - In Cluster
  - In Line

QLCS
- Bow Echo
  - Supercell
    - Right Mover
      - Discrete
        - In Cluster
        - In Line
    - Supercell
      - Left Mover
        - Discrete
        - In Cluster
        - In Line

Based on:
We supplement Smith et al. 2012, including weak severe, non-severe, and synthetic verification measures.
Develop / Refine Software

Storm object identification / tracking

Machine learning

Visualization / blending of MRMS + model
Multi-scale storm "cluster" identification
Multi-scale storm "cluster" identification

200 km²
Multi-scale storm “cluster” identification

2000 km²
Multi-scale storm “cluster” identification
### Storm classification inputs from MYRORSS / MRMS

<table>
<thead>
<tr>
<th>Storm Attribute</th>
<th>Max 30 Minute MESH</th>
</tr>
</thead>
<tbody>
<tr>
<td>-20 C Merged Reflectivity</td>
<td>Most Unstable CAPE</td>
</tr>
<tr>
<td>0 C Merged Reflectivity</td>
<td>Most Unstable LCL Height</td>
</tr>
<tr>
<td>Aspect Ratio</td>
<td>Probability of Severe Hail (POSH)</td>
</tr>
<tr>
<td>0-2 km Merged Azimuthal Shear</td>
<td>Quality Controlled Merged Reflectivity Composite</td>
</tr>
<tr>
<td>3-6 km Merged Azimuthal Shear</td>
<td>Severe Hail Index (SHI)</td>
</tr>
<tr>
<td>0-6 km Shear Magnitude</td>
<td>Storm Size</td>
</tr>
<tr>
<td>0-1 km Storm Relative Helicity</td>
<td>Surface CAPE</td>
</tr>
<tr>
<td>0-3 km Storm Relative Helicity</td>
<td>Surface Dewpoint</td>
</tr>
<tr>
<td>Longevity</td>
<td>Surface Temperature</td>
</tr>
<tr>
<td>Maximum Expected Size of Hail (MESH)</td>
<td>Vertically Integrated Liquid (VIL)</td>
</tr>
</tbody>
</table>

**Note:** The table lists various storm attributes and their corresponding classification inputs from the MYRORSS/ MRMS system.
Storm classification: Example Decision Tree
Machine Learning

Tested on storm lifetime forecast

- gradient boosted regression trees
- random forests
- logistical regression
- AdaBoosting

Trend of the expected storm lifetime.
1) Reflectivity (or other image that can be clustered)

2) Storm Clustering and Tracking

3) Then extract storm properties:
   - Other MRMS data for each cluster (radar, satellite, lightning)
   - Background environment (from NWP model analysis)

12 analyses / hour
X 24 hours / day
X 365 days / year
-----------------------
105,120 analyses / year

Millions of storms / year
HWT Real-Time Visualization

Boxes are frequency of WoF UH > 50 m²/s²

WoF updraft helicity predictions

MRMS supercell object (contour)
1. Acquire/prepare data sets
   • Multi-Year Reanalysis of Remotely Sensed Storms (MYRORSS; pronounced “mirrors”)
   • Storm Prediction Center database of convective modes (+ supplemental data)
   • NSSL WRF storm-object data set

2. Develop / refine software
   • Storm object identification / tracking
   • Machine learning
   • Visualization / blending of MRMS + model
Objectively classify storms

- Machine learning algorithms applied to storm object data
- Distributions of storm type and lifetime based on environment

Compare observed and model storm-typing and severity

Real-time testing and evaluation with forecasters (Spring 2017)