Ensemble modeling of storm interaction with XSEDE

Brian Jewett*
Jay Alameda
Robert Wilhelmson
Ann Syrowski

University of Illinois
Atmospheric Sciences & NCSA

*bjewett@illinois.edu

presentation online: rt.atmos.uiuc.edu/xsede12
Overview

• Introduction
• Our scientific investigation
• Lessons learned: science
• Lessons learned: computing
• Lessons learned: data
• Supercomputing science wish list
In context...

35 YEARS OF MICROPROCESSOR TREND DATA

Chuck Moore - AMD Fellow. Data from M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, C. Batten
A brief history of cloud modeling

- **1970s**: 3-D clouds
  - *first three-dimensional simulations of clouds; key physics & numerics*
  - Steiner 1973, Schlesinger ’75
  - Klemp & Wilhelmson 1978

\[
\Delta x = \Delta y = 1 \text{ km}
\]

Run: 80 min

Grid: 24 x 24 x 20

Rainwater

X-Z slice

Klemp and Wilhelmson, 1978

1 GHz

Klemp and Wilhelmson, 1978

XSEDE’12

University of Illinois Atmospheric Sciences
A brief history of cloud modeling

- **1980s**: 3-D squall lines; sensitivity studies
  - *realism, larger phenomena*
    - Rotunno et al. 1988

- **1970s**: 3-D clouds

\[ \Delta x = \Delta y = 2 \text{ km} \]

Run: 8 hours

Grid: 90 x 90 x 25

\( z = 0 \text{ km} \)
\( z = 4 \text{ km} \)

Rotunno, Klemp and Weisman, 1988
A brief history of cloud modeling

- **1990s**: larger problems
  - *in-depth investigation of multi-scale phenomena*

- **1980s**: 3-D squall lines

- **1970s**: 3-D clouds

\[
\Delta x = \Delta y = 2 \text{ km}
\]

Run: 6 hours

Grid: 300 x 300 x 24

Weisman and Davis, 1998
A brief history of cloud modeling

- **1970s**: 3-D clouds
- **1980s**: 3-D squall lines
- **1990s**: larger problems
- **2000s**: forecasting storms
  - spring forecasts; resolution

\[ \Delta x = \Delta y = 2, 4 \text{ km} \]

Run: 30 hours

Grid: 1500 x 1320 x 51

Kain et al., 2008

University of Illinois Atmospheric Sciences
A brief history of cloud modeling

- **Real-time** forecasts of convection are now being done at 1-2 km grid spacing.
- **Simulation** studies are at 10s-100s of meters.
- **Very large simulations** planned for systems coming online, e.g. Blue Waters at NCSA.

- **2010s**: higher resolution
- **2000s**: forecasting storms
- **1990s**: larger problems
- **1980s**: 3-D squall lines
- **1970s**: 3-D clouds
Resolution

- Convective modeling:
  - on larger domains
  - at higher resolution
  - not always better

- More detail ...
  - see right.
  - are the solutions converged?

Droegemeier et al. 1994
Resolution

- Convective modeling:
  - on larger domains
  - at higher resolution
    - not always better
- More detail ...
  - see right.
  - converged?
- Bryan et al.
  - Turbulence considerations: we should use $\Delta \sim 100m$
Storm interaction study

M.S., Ann Syrowski, 2012
Where we stand today

We have a reasonable understanding of the large-scale conditions associated with major tornado outbreaks.
Where we stand today

SPC forecasts for April 14, 2012 – from a week prior to the day of the event. link
Statement of the problem

• But – a particular tornadic storm ...
  • can be much more difficult to forecast
  • this contributes to tornado watches over a wide area.

• Our thunderstorm knowledge ...
  • is based largely on how an isolated storm would evolve without other storms, fronts, etc. nearby
  • Not a good assumption!
  • That’s where our work comes in.
Science questions

• How does the presence of a nearby cell influence the severity of a storm?
  • Does it matter?
  • If so, why?
  • How?

• We call the development of one storm in the vicinity of another *storm interaction*.
  • This is a catch-all for a number of possible scenarios.

Shabbott and Markowski (2006)
Storm interaction

• 1996 tornado outbreak

Jacksonville, IL tornadic storm, (B. Lee)

Midwest severe reports – 4/19/96

- Over 30 tornadoes struck IL
- Some were strong (F2-F3)
- Most tornadoes were short-lived
Storm interaction

- 1996 tornado outbreak
- Analysis of radar data

Splitting.
Storm interaction

- 1996 tornado outbreak
- Analysis of radar data

... becomes two.

Splitting.

Radar reflectivity - western IL (red = heaviest rain/hail)
Storm interaction

- 1996 tornado outbreak
- Analysis of radar data

Merging.

Radar reflectivity - western IL (red = heaviest rain/hail)
Storm interaction

- 1996 tornado outbreak
- Analysis of radar data

... become one.

Merging.
Storm interaction

- 1996 tornado outbreak
- Analysis of radar data

Newly strengthened storms ...

Intensification.

Radar reflectivity - western IL (red = heaviest rain/hail)
Storm interaction

- 1996 tornado outbreak
- Analysis of radar data

... later produce F2-F3 tornadoes.

... to tornadic strength.
Experimental design

- Simulate *many* interaction scenarios
  - but only of *two* storm cells at a time.

![Diagram showing interaction scenarios](image)
Experimental design

- **Weather Research and Forecasting Model (WRF) v3.3**
- **Control:** Single isolated storm
- **51 two-storm simulations**
- $\Delta x = \Delta y = 0.54 \text{ km} \quad \text{(we are not at tornado-scale)}$
- 90 vertical levels
- Run for 5 hours
Data, runs, and post-processing

• **540-m simulations**
  - Domain 256 x 280 x 90; run 5 hr
  - Raw output saved every minute
  - Total ensemble run data: 15 TB
  - Total ensemble run SUs: 19,600
  - Total ensemble post SUs: 8,200

• **60-m simulation**
  - Data saved every 12 sec; run 5 hr
  - Total run data: 17 TB

• **Post-processing**
  - after the fact
  - need to integrate into a comprehensive workflow.
Storm interaction: strong case

- 5-hr run
- “Radar” view, top-down
- Initial storms split, interact
Metric: maximum updraft

Among the 51 simulations, storms of comparable updraft strength were produced (good!)
Metric: peak rotation @ ground

Distribution of surface vertical vorticity, per second. All runs exhibited strong *mesocyclones.*
Metric: peak rotation, sorted

Finding #1: almost any case with a nearby storm is stronger.
Cases of interest at 30 min

WRF_20: Max rotation 0.11 s\(^{-1}\)

WRF_04: Max rotation 0.11 s\(^{-1}\)

WRF_03: Max rotation 0.06 s\(^{-1}\)

CONTROL: 0.06 s\(^{-1}\)
Finding #2: small differences early => large differences later.
Finding #3: rotation centers merge; this is a key mechanism for intensification.

Vertical vorticity ($s^{-1}$)

- 0.1
- 0.027
- -0.046

Vertical vorticity ($s^{-1}$)
Animation of 3-D rotation, surface wind.

Viewpoint is initially straight down, then looking to $-Y$ (east to west).

**STRONG CASE**

Vertical vorticity
- $0.016 \, \text{s}^{-1}$

3-D vorticity
- $0.035 \, \text{s}^{-1}$

Vertical velocity
Finding #4: highly unsteady nature of the “forward flank” downdraft ahead of the storm.

Numerous pulses can be seen, driven by rain-cooled air.
Lessons learned: Science

Cell interaction can greatly alter the behavior, and potential severity, of thunderstorms.

1. Small early differences grow significantly

2. Rotation centers develop and merge with the primary mesocyclone in episodes of storm intensification. This may be important for warning lead time.

3. The downdrafts are particularly unsteady, and their location and timing appears key to rotation intensity.
Lessons learned: Computing

• Resource stability
  • We hope for computational resources that persist for 2-3 years.
    • This research moved among 5 HPCs in a 3-year period.
    • Our results are highly sensitive to small changes
      • moving to a new architecture requires rerunning the ensemble.

• Workflows
  • Running post-processing after the fact is incredibly inefficient, and
    with shifting to multiple centers, required much data movement.
    • We would be very willing to collaborate or help
      drive use cases for these types of problems.
    • Our workflow orchestration is not very application specific.

• Planning
  • We need to plan for analysis / data as thoroughly as for simulations.
Lessons learned: Data (1)

• Overall –
  • We (I) underestimate data needs, despite good intentions!

• Data needs
  • for code compilation / maintenance (home directory, ≥ 10 GB)
  • short term storage during run
  • medium term storage (1-2 yr) for analysis
  • long term storage (data subset) while paper is in review
  • extended term storage for limited subset
  • Data capacity/throughput/stability ideally would be considered in future HPC planning.

• Archivability
  • We desperately need stable, predictable long-term storage.
  • Could this be a high priority for future HPC planning?
Lessons learned: Data (2)

• Architecture
  • SSD (Gordon, Ember) / memory flash storage (Blacklight) is very useful for enhancing run data movement (likely: post)

• Post-processing
  • We are among those who produce select large data files but also very large numbers of small files.
    • I understand the latter is problematic for high-performance file systems.
    • Our post-processing is partly serial and partly (trivially) parallel.
  • We have experienced significant slowdown (we were the cause!) on parallel file systems during our POST step.
  • We would welcome (1) help in improving our POST design, or (2) a greater emphasis on handling these types of transfers in future HPCs.
Lessons learned: Data (3)

• Temporal resolution
  • Our data needs are going UP!
  • Every minute @ 0.5-km grid spacing is almost certainly inadequate
    • especially for statistics, trajectory calculations
  • Tornado-scale problems will mandate MUCH more data saved.
    • This could greatly increase execution wallclock, post needs.

• Data storage policies
  • More uniformity among centers would be welcome
  • Not all policies are clearly stated
    • “project” storage is very helpful!
  • Home directory space < 10 GB is not practical for our use.
Supercomputing wish list

• We are concerned there is a difference of need / outlook between application sciences and cyberinfrastructure planning at NSF.

• Our needs are foremost
  • stable platforms, > 2 years
  • stable data storage and access
  • HPC stability and performance.

• Some NSF-funded projects in my department may leave XSEDE due to stability / persistence concerns.

• We are concerned about future HPC availability

• Resource allocations
  • Could use some more information
  • Trial allocations/transfers to benchmark before production?