J9.3 AN AUTOMATED SYSTEM FOR PROCESSING THE MULTI-YEAR REANALYSIS OF REMOTELY SENSED STORMS (MYRORSS)

John Cintineo*
University of Oklahoma, Norman, Oklahoma

Travis Smith, Valliappa Lakshmanan
NOAA National Severe Storms Laboratory; Cooperative Institute of Mesoscale Meteorological Studies, Norman, Oklahoma

Steve Ansari
NOAA National Climatic Data Center, Asheville, North Carolina

ABSTRACT

The Multi-Year Reanalysis Of Remotely-Sensed Storms (MYRORSS) is a cooperative endeavor between the National Oceanic and Atmospheric Administration's (NOAA) National Severe Storms Laboratory (NSSL) and the National Climatic Data Center (NCDC) to reconstruct and evaluate numerical model output and radar products derived from 15 years of WSR-88D data over the coterminous U.S. (CONUS). The end result of this research will be a rich dataset with a diverse range of applications, including severe weather diagnosis and climatological information.

An automated system has been developed by the NSSL for processing level-II radar data provided by NCDC. The system operates in a multiple-machine framework, with an already existing network of computers. This procedure maximizes processing power on each machine and utilizes idle time from other users' computers. The Warning Decision Support System – Integrated Information (WDSS-II) suite of programs are used to process and quality-control the data.

The system operates in three main phases. The first phase is single-radar processing from each individual radar in the CONUS, which includes quality-control of the level-II data and creation of velocity-derived products. The second phase creates blended 3D reflectivity fields as well as 2D fields derived from radial velocity data. The third phase runs post-processing algorithms on merged meteorological fields and ingested near storm environment data provided by the Rapid Update Cycle model (RUC).

Applications for this work are discussed and preliminary results are shown for one year of CONUS NEXRAD and RUC data.

1. INTRODUCTION

The NEXRAD era has now eclipsed 15 years, and is beginning to near the end of its intended 20-year lifespan. There are 134 WSR-88D radars throughout the CONUS, which have been operating nearly constantly during this time. NCDC has archived the level-II products for each of these radars. These raw weather radar data provide an opportunity for a wealth of data mining analyses. However, there are two main problems with the data:

1. The data are fragmented and inconsistent, meaning there are many missing or partial volume scans.
2. Blending the radar data to make it more useful is very resource intensive.

To tackle these issues and process the extremely large weather radar dataset, an automated system with sufficient hardware resources is required.

With such a system in place, a plethora of investigations may begin to use

*Corresponding author address:
John Cintineo, 120 David L. Bored Blvd., Norman, OK 73072
e-mail: john.cintineo@ou.edu

the radar data and data from other observation platforms. Some examples include: a climatology of severe weather events, such has hail, tornadoes, high-winds, and flash floods; convective initiation studies, which may incorporate satellite or lightning data in unison with the radar data; and hydrological studies based on radar-derived rainfall estimates.

2. COMPUTING FRAMEWORK

The main computational tool used to carry out the processing of the radar data is the Warning Decision Support System – Integrated Information (WDSS-II) [1]. WDSS-II provides several useful algorithms for processing radar data. These include a reflectivity quality control algorithm, the ability to de-alias radial velocity from a Rapid Update Cycle (RUC) model sounding, and a "merging" or "blending" algorithm to provide better estimates of radar moments where there is over-lapping coverage of WSR-88Ds. The blending also mitigates the "cone of silence" over a radar site, and provides coverage over one radar that may be off-line, if an adjacent radar has over-lapping coverage. The blended radar data will have a spatial resolution of 0.01^o x 0.01^o (approximately 1km x 1km in mid-latitudes) and a temporal resolution of initially 5 minutes. WDSS-II also has the ability to run "multi-radar/multi-sensor" algorithms, which incorporate model data to proxy the environment, or potentially lightning and satellite data. These derived-products will provide researchers with a historical, observational dataset that has very high spatio-temporal resolution.

With WDSS-II serving as the backbone software for processing level-II data, the automated system for MYRORSS provides the parallel processing.

3. HARDWARE CONFIGURATION AND PROCESSING AUTOMATION

The current configuration of hardware consists of 20 desktop machines (8 to 16GB RAM and 4 to 8 processing cores), 6 server machines (16 or 32GB RAM and 4 processing cores), 54TB raid storage disk, and 60TB of additional storage (thirty 2TB external hard drives).

Three of the servers are used for the most computationally intensive portion of processing, the merging of reflectivity over the CONUS in 24-hour increments. 2 other servers have the task of running post-processing algorithms on the processed radar and RUC data, which includes algorithms such as the multi-radar Hail Detection Algorithm (HDA). The remaining server acts as the "master" server, in charge of job distribution to the single-radar processing machines. It also has the 54TB raid mounted to it.

The 20 desktop machines are used for single-radar processing. Each "job" sent to a computer contains instructions to process a specific radar for a specific amount of time (1 hour for super-res data and an 8-hour block for legacy data). This stage of processing includes quality-control of reflectivity using a neural network in WDSS-II, de-aliasing of radial velocity data, and computing several derived products, such as azimuthal shear and a pre-merged reflectivity. All processed products are sent back to the main storage raid, except the pre-merged reflectivity, which is sent to one of the 3 servers for merging. Each job is set to a low priority on the computer, since these machines have other users, including other NSSL scientists and seasonal projects, such as the Hazardous Weather Testbed (HWT) Spring Experiment and the Severe Hazards Analysis and Verification Experiment (SHAVE).

The merging servers blend radar data from 134 CONUS radars over a contiguous 24 hours, using a distance-weighted scheme. This step requires much computer memory and time—6 to 12 hours of real-time to merge 24 hours of archived data (depending on the resolution and coverage of the data over the CONUS). Using 3 servers to accomplish this provides much needed parallel-processing capability for a very time-consuming task.

Once complete, the 24-hour 3D merged QC reflectivity grids are stored on the raid, and are a most useful product for multi-radar/multi-sensor algorithms. At this point, the two other servers will now pull the merged reflectivity grids and near-storm environment data (NSE, derived from the RUC model) to compute a multitude of derived products. These include maximum expected size of hail (MESH), probability of severe hail (POSH), vertically integrated liquid (VIL), echo top heights, isothermal reflectivities, layer-average reflectivities, low-level and mid-level azimuthal shears, and more. Upon creation, These derived products are also sent and stored on the main raid, but will eventually be housed at NCDC's Severe Weather Data Inventory (SWDI) for public distribution [2]. Figure 1 schematically illustrates the flow of processing for MYRORSS.

Neglected to be mentioned until now, the very first step in the entire process is to download raw level-II data from NCDC. It takes approximately 1 week to download an entire month of radar data for the CONUS. Processing that month's worth of data takes approximately 7 to 12 days for super-res, and 3-5 days for legacy data.

4. APPLICATIONS

The first application of these data investigated locally at NSSL is a MESH climatology, which is meant to be a more objective hail climatology than NCDC's Storm Data [3]. Only one full year of data has been processed, and the results will be discussed. The data are very preliminary, and are meant to show readers the possibility of the fields produced by MYRORSS, and not actually provide very meaningful climatological information regarding the occurrence of hail in the United States. A much more thorough analysis and multi-year database would be required for scientific conclusions to be drawn.

From a year's worth of MESH grids (2009), daily MESH tracks were created, by taking the maximum MESH at every grid point over a 24-hour time span. Figure 2 shows a subset of a daily MESH grid over the Midwest U.S. The contours represent maximum radar-derived hail sizes in millimeters. Once the daily MESH grids are computed for an entire year, a daily "count" field may be synthesized. Figure 3 displays this grid for 2009. The first thing to notice is the very large coverage and high resolution map produced by using the multi-radar/multi-sensor products. The blue portions represent grid cells where MESH exceeded 21 mm on 1 or 2 days in 2009; the green portions represent cells where MESH exceeded 21 mm for 3, 4, or 5 days. There are only a few small regions where MESH exceeded 21 mm for 6 or more days. The threshold of 21 mm was chosen since it was found to be the best calibrated MESH value to delineate hail and no-hail values (analysis not shown). The HDA was designed such that 75% of hail fall would be below the MESH value, therefore, we do not get a one-to-one correspondence with actual hail fall.

Despite the quality-control and radar blending, the data are still not pristine. There are errors that were not caught by the reflectivity QC algorithm such as electronic interference at certain radials and anomalous propagation. Regardless, the main signal is there: that most of the hail in the U.S. for a given year falls in the Great Plains region. The fact that this count field is reasonable lends credence to the prospect of creating a complete radar-derived hail climatology for the CONUS during the NEXRAD era.

5. CONCLUSION

An automated system for processing MYRORSS has been established and is currently running on raw level-II data provided by NCDC. Its ability to process the data in parallel makes MYRORSS feasible. The extremely high-resolution, 3D, merged QC reflectivity dataset over the CONUS provides boundless data-mining opportunities. Analysis has already begun on one year of processed data, which includes a radar-derived hail climatology and a radar-derived circulation climatology. Other studies that may be investigated in the near future include a high wind climatology, convective initiation detection, and gridded multi-sensor flash flood climatology and/or forecast guidance.

6. REFERENCES

1. V. Lakshmanan, T. Smith, G. J. Stumpf, and K. Hondl, “The warning decision support system - integrated information," *Weather and Forecasting*, vol. 22, no. 3, pp. 596-612, 2007.

2. NOAA’s Severe Weather Data Inventory (SWDI):
http://www.ncdc.noaa.gov/swdi/#Intro

3. NCDC Storm Data:
http://www.ncdc.noaa.gov/oa/climate/sd/

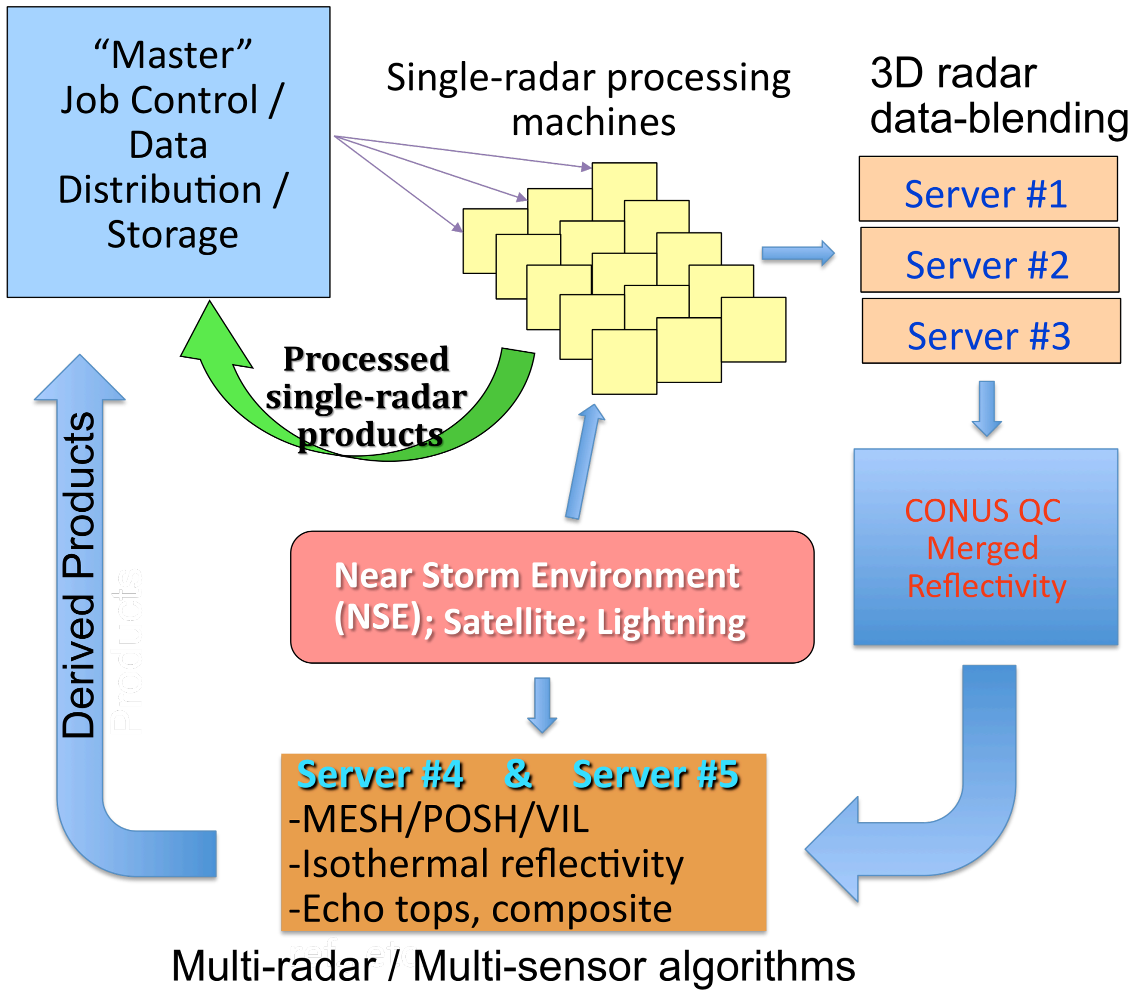

Figure 1: Schematic of the flow of data and processing. Jobs all originate with the "Master" server. Processed data fields are stored on a 54TB raid disk, which is mounted to the Master server.

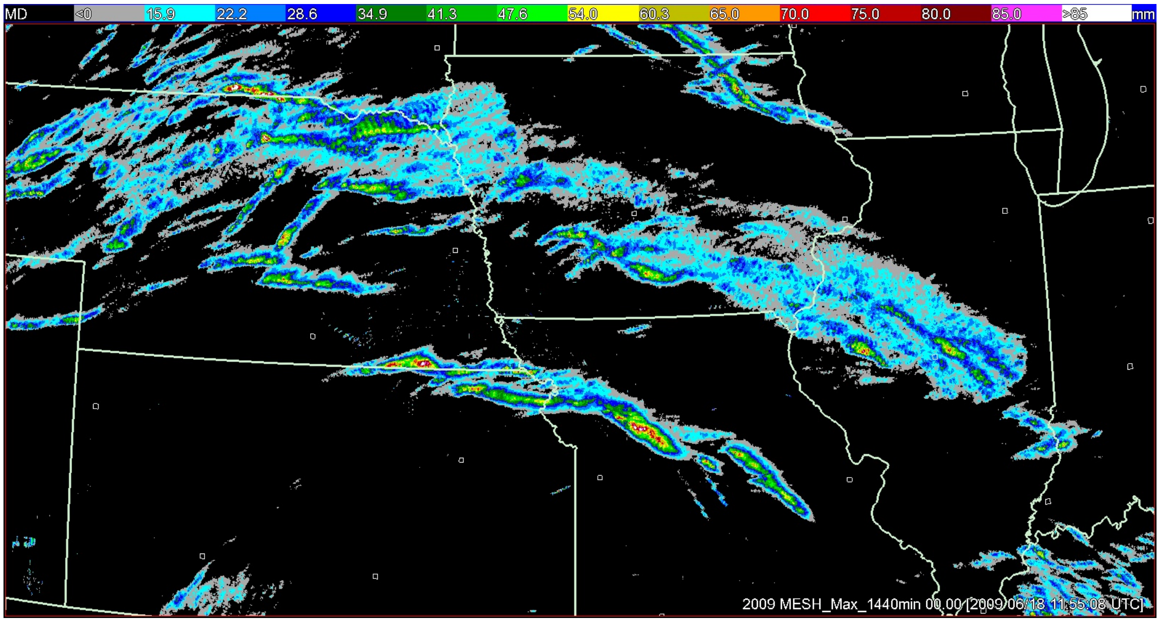

Figure 2: Maximum MESH swaths over a 24-hour period (18 June 2009) for a portion of the Midwest U.S. This field is aggregated to create “count” maps for a year.

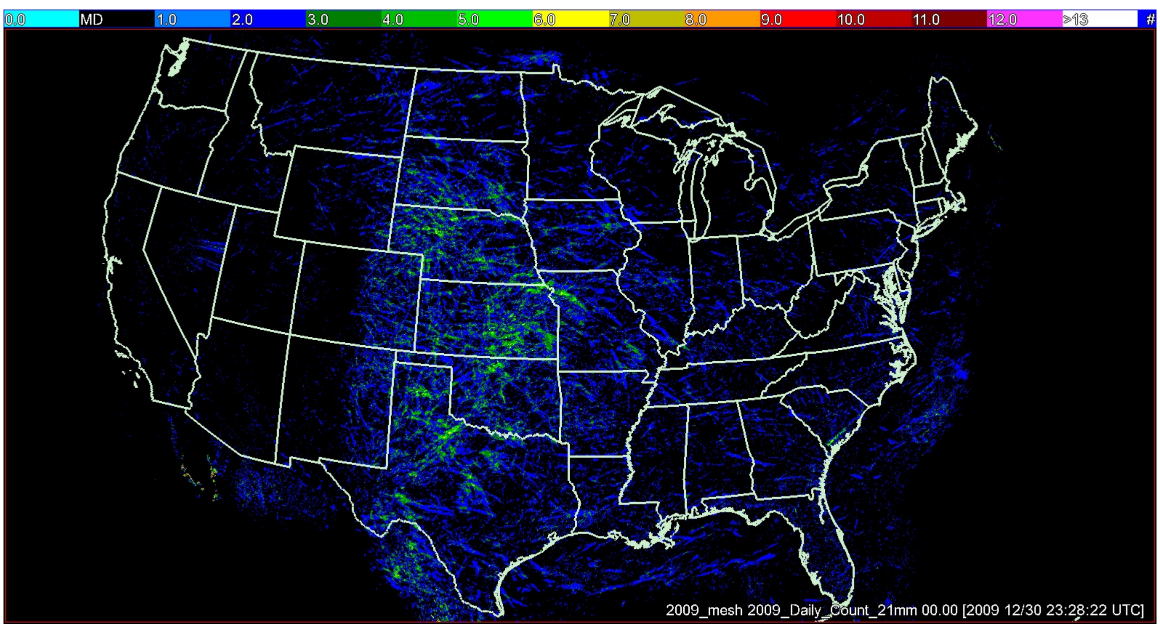

Figure 3: “Count” map for 2009. Colors represent the number of days where MESH ≥ 21mm. Blues are 1 to 2 days and greens are 3-5 days.