Towards Better Utilization of NEXRAD Data in Hydrology: an Overview of Hydro-NEXRAD

Witold F. Krajewski[1], Anton Kruger[1], Ramon Lawrence[6], James A. Smith[2], A. Allen Bradley[1], Matthias Steiner[4], Mary Lynn Baeck[2], Mohan K. Ramamurthy[3], Jeffrey Weber[3], Stephen A. DelGreco[5], Bong-Chul Seo[1], Piotr Domaszczynski[1], Charles Gunyon[1] and Radoslaw Goska[1]

[1]IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, IA 52242
[2] Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544
[3]Unidata Program Center, UCAR, Boulder, CO 80307
[4]National Center for Atmospheric Research, UCAR, Boulder, CO 80307
[5]National Climatic Data Center, Ashville, NC 28801
[6]Computer Science, University of British Columbia Okanagan, B.C., Canada, V1V 1V7

Abstract

With a very modest investment in computer hardware and the open source local data manger (LDM) software from UCAR's Unidata Program Center, an individual researcher can receive a variety of NEXRAD Level III gridded rainfall products, and the unprocessed Level II data in real-time from most NEXRAD radars. Additionally, the National Climatic Data Center has vast archives of these products and Level II data. Still, significant obstacles remain in order to unlock the full potential of the data. One set of obstacles is related to effective management of multi-terrabyte data sets: storing, compressing, and backing up. A second set of obstacles, for hydrologists and hydrometeorologists in particular, is that the NEXRAD Level III products are not well suited for application in hydrology. There is a strong need for the generation of high-quality products directly from the Level II data with well-documented steps that include quality control, removal of false echoes, rainfall estimation algorithms with variety of corrections, coordinate conversion and georeferencing, conversion to a convenient data format(s), and integration with GIS. For hydrologists it is imperative that these procedures are basin-centered as opposed to radar-centered. Thirdly, the amount of data present in a multi-year, multi-radar dataset is such that simple cataloging and indexing of the data is not sufficient. Rather, sophisticated metadata extraction and management techniques are required. The authors describe and discuss the Hydro-NEXRAD software system that addresses the above three challenges. With support from the National Science Foundation through its ITR program, the authors are developing a basin-centered framework for addressing all these issues in a comprehensive manner, tailored specifically for use of NEXRAD data in hydrology and hydrometeorology. Through a flexible web interface users can search a large metadata database base, managed by a

relational database, for subsets of interest. Well-chosen and documented defaults are provided for the flow from unprocessed NEXRAD data to basin-centered rainfall estimates at a desired space-time resolution. In addition to the web interface, there are web services that provide access to scripts and compiled programs.

Introduction

The main objective of the Hydro-NEXRAD project is to increase the use of NEXRAD data in hydrologic research. The project is a joint effort of researchers from the University of Iowa, Princeton University, Unidata Center Program of the Universities Corporation for Atmospheric Research (UCAR) and the National Climatic Data Center of the National Oceanic and Atmospheric Administration (NOAA). While there are readily available NEXRAD radar-rainfall products developed by the National Weather Service (NWS), their resolution is fixed to the HRAP grid (Reed and Maidment 1999) and hourly time scale and thus limits the applications for which these products are useful. At the same time, developing custom products directly from the Weather Surveillance Radar (WSR-88D) collected raw reflectivity and Doppler velocity (Level II) data requires expertise that is not widespread in the hydrologic and engineering community and not easy to quickly acquire. Hydro-NEXRAD is a demonstration of a framework and the information technology tools that overcome the above constraints. We have been developing this Internet-based and browser-compatible software for access, search, selection, and specification of customized radar-rainfall products based on WSR-88D radar reflectivity Level II data.

In this paper we highlight and briefly discuss the main aspects of the project. To provide such capability we are developing a software system consisting of the following main elements:

1. Efficient storage and fast read time Level II data format (Kruger and Krajewski 1997);
2. A relational database that enables flexible data storage organization (Kruger et al. 2006);
3. Hydrologic basin centric metadata;
4. Level II data quality control;
5. A set of modular radar-rainfall estimation algorithms;
6. A set of utilities for final product generation and dissemination;
7. Graphical User Interface that allows users (research hydrologists) to specify the products they needs;
8. Documentation of the entire system.

In the following sections we elaborate on the functionality and other aspects of each element and provide current status of its development. Two companion papers focus on the metadata calculations (Kruger et al. 2007, this volume) and the radar-rainfall algorithms (Krajewski et al. 2007, this volume).

Components of the Hydro-NEXRAD System

1. *Efficient data format*

We decided to use the Run Length Encoding (RLE) format we had developed for radar data several years ago (Kruger and Krajewski 1997). Despite fast progress in computer storage and processing speed, this lossless format continues to offer advantages as compared to other popular data compression utilities. Portability, fast read times, and storage efficiency equaling or exceeding those of *gzip* are the main reasons for our decision. Also, in the process of converting Level II data from its native to the RLE format we perform a number of quality control checks identifying corrupt files and headers and marking them with a system of flags. This ensures robustness of our overall system. The RLE format is applied on the level of volume scan Level II data. Additional advantage of the RLE format is that we don't need to read the entire file when processing information and generating rainfall products for a basin, often a small subsection of the entire radar umbrella. This results in processing speed gains. In contrast, use of *gzip* format requires decoding entire volume scan data every time a piece of information is needed that is contained within.

2. *Relational database*

Researchers usually manage radar data using a hierarchical file and directory system, organizing radar volume files by radar and date. NEXRAD radars produce large amounts of data, namely a ~3.5 MB volume file every 5–7 minutes, and there are 140+ such radars. The file system-approach suffices for small datasets, but quickly becomes unwieldy as the data set grows. Organizations such as the National Climatic Data Center also follow this approach, but employ sophisticated hardware such as robotic tape loaders, large RAIDs and so on. To help navigate a file-based archive, one can maintain a catalog of "interesting" portions of the data, such as severe rain events. This enables researchers to ignore uninteresting parts of the data.

A key idea in Hydro-NEXRAD is that of metadata—data about (NEXRAD) data—managed in a relational database. It greatly eases the management of the data and allows researchers to search for, and find interesting subsets in a very flexible manner. Briefly, for each of the raw (Level II) NEXRAD volume files that a radar produces, we compute descriptive statistics. For example, the areal coverage statistic is the area of the radar image that has reflectivity (Z) values above a certain threshold. Such statistics, and many other pieces of information about Level II files, including their location on the file system, are managed in a relational database. Files are located across several file systems, at many different physical locations. As Level II files are moved between locations, the database table entries are updated accordingly.

Web services play an important role in Hydro-NEXRAD. The relational database and the data are behind web servers. All access to the data and the relational database are through these servers. A client (human, compiled program, or script) accesses the data through a two-step process: (a) query the database, which returns a URL to the data, and (b) request the web server to serve up the data. The original concept and implementation follows on earlier work documented in Kruger et al. (2006). Currently, our system provides ad-hoc web services using the PHP and JSP GCI engines of the web servers. Future work includes moving to standard interfaces (SOAP, OWL, etc.) We elaborate on the computation of the metadata in the companion paper Kruger et al. (2007, this volume.)

3. Basin-centric metadata

Hydrologic studies are often organized around basins, i.e., land units defined by a point on a stream channel network that all draining water from the units passes through. In the United States surface waters are organized into a hierarchy system of basins defined by the Unites States Geologic Survey (USGS). Basins of different sizes are assigned a Hydrologic Unite Code (HUC) that provides unique identification and defines membership within larger units. Large basins, e.g., the Upper Mississippi River Basin (USGS HUC 0708) is covered by many WSR-88D NEXRAD radars and contain smaller basins. The numbering is hierarchical: watersheds with 4-digit HUC are comprised of smaller, 6-digit HUCs, and each 6-digit HUC watershed is comprised of 8-digit HUC watersheds and so on. Lower level units are as small as 10 km^2. Clearly, many such units are within each of the WSR-88D radar coverage.

Since there are many more small basins than large basins, it is likely that most hydrologic studies require rainfall information from a small portion of a given radar umbrella. To facilitate fast search for required data we have developed a system of storing information relevant to each basin. To accomplish this we used the first four levels of the USGS system. As a result we have developed database for each of 2199 4th level (8-digit HUC) basins. We have constructed an indexing system that links each basin with each radar. Thus, a user can quickly determine how many and which radars are "looking" over each basin. Each basin is associated with a latitude/longitude box for which a user is likely to request precipitation products. Each basin is also assigned a polar box for each relevant radar. The box defines the azimuth and range of data required for processing. We elaborate on the specific rainfall quantities we calculate in Kruger et al. 2007, this volume.)

4. Data quality control

Since the main product of the Hydro-NEXRAD system is precipitation, radar echo due to other phenomena should be identified and excluded from further processing. This should be accomplished at the volume scan level, before Level II data are used for precipitation estimation.

To address this problem we have implemented the algorithm of Steiner and Smith (2002). The algorithm is applied at each volume scan as it considers the vertical extend and structure of the radar echo in its classification decisions. The first time the use of the algorithm is requested, (e.g., at data ingest and processing time) a mask is produced that classifies each pixel in the base scan as precipitation or non-precipitation. Subsequent users of the volume scan data have an option of applying or not applying the mask.

We have also implemented an option that uses a CAPPI (Constant Altitude Plan Position Indicator). To avoid sharp boundaries between various antenna elevation scans, we use a smoothing kernel. Such calculation of the CAPPI also helps mitigate the effect of ground clutter and anomalous propagation echo.

5. *Rainfall estimation algorithms*

Radar-based rainfall estimation algorithms differ in complexity and performance (e.g., Fulton et al. 1997; Ciach et at. 1998; Anagnostou and Krajewski 1999). Unfortunately, regarding their performance there is no system in place that would enable objective evaluation. As a consequence, there is no consensus on what is the best algorithm and optimality criterion. To address this situation our strategy is to provide users with flexibility in selecting different options and parameter values for the algorithms. While it is impossible to be fully comprehensive in providing such options, our system will offer a lot of flexibility, way more than what is possible with the NWS "official" Precipitation Processing System (PPS) (Fulton et al. 1997).

Among different algorithm options users will be able to specify a "quick look" algorithm, a default algorithm, quasi-PPS algorithm, corrections for advection and range effects, different *Z-R* parameters, hail cap and no-rain thresholds, and hybrid scan construction parameters, among others.

We considered providing an option to reproduce the PPS results as very fundamental but this turned out to be a "mission impossible." While we have received the source code for the PPS from the Office of Hydrology of the NWS, it is buried within a much larger piece of software called *CODE* developed by the NEXRAD agencies and their private contractors. There is no stand-alone PPS available for use outside of the NWS. Also, the PPS is constantly changing and it is hard to keep track of all the minor fixes and modifications. As a result, the closest we have come to reproducing the PPS results is to about 5%. This is the reason why we refer to this option as pseudo-PPS.

To facilitate the mix-and-match approach we have developed basic modules for the following elements: hybrid scan construction using the concept of Constant Altitude Plan Position Indicator (CAPPI) and kernel smoothing to avoid ring appearance common in long term accumulations of the PPS products, rainfall rate calculations, rainfall accumulation, advection correction that improves rainfall accumulation by ensuring that pixels are not skipped over under certain combinations of storm velocity and product resolution; and range correction. We provide more details about the algorithms in the companion paper by Krajewski et al. (2007, this volume.)

We continue performing extensive tests of the algorithms and the codes running them on multi month periods of Level II data to make sure that they don't crash under varied data conditions.

6. *Final product utilities*

To provide users with further flexibility, our system will output precipitation products in several coordinate systems and resolutions. While the operational NEXRAD precipitation products are provided on the so-called Hydrologic Rainfall Analysis Project (HRAP) grid, which is about 4 km by 4 km, many distributed hydrologic models of basin processes require rainfall input on much finer resolution, e.g., 1 km by 1 km. Some models require input on latitude/longitude grid while others work using a local Cartesian system.

Our software accommodates many of these scenarios. Users can specify HRAP, Super-HRAP (1/16 of HRAP), latitude/longitude, and NASA's Land Data Assimilation System (LDAS) grids. Since these grid systems are fixed, we have developed lookup tables for each basin for each grid that allow fast projection of the precipitation products generated in polar coordinates for a relevant radar onto the grid. We included three options to make the projection (i.e. interpolation): nearest neighbor, simple averaging, and weighted averaging that account for the radar beam pattern. Use of different options implies different computational effort even using the lookup tables.

Once the products are generated and ready to be transferred to the user they are formatted as netCDF files, and can be used in many different applications, including Geographic Information Systems. We have developed a file naming and header structure convention so that one can tell at a glance what product and with what options is contained in the file. The naming convention also facilitates file level manipulation by various utility scripts.

7. Graphical User Interface

To facilitate users' interaction with our database and the algorithms we have designed a web-browser based GUI. Using it users can locate their basin or domain of interest, radars that cover it, visualize the grid on which the final products will be provided, find cases (data periods) of interest using metadata based searches, and specify algorithm options. A messaging system guides the non-expert (in radar meteorology) users through issues relevant to the user's request. For example, basin located very close to the radar site might be subject to frequent ground clutter contamination while basin located at the fringes of converge might suffer from sever under detection of precipitation.

We have developed a browser-based GUI (see Figure 1 for a screen shot). It communicates with a map server to provide user-specified map detail at four levels of zooming (first level is the entire United States). The mapserver is maintained at Unidata. The main elements shown on the maps are the hydrologic basin boundaries according to the USGS classification system. Users can select large basins that contain their basin of interest (hydrologist usually are familiar with this information). Zooming allows finding the basin of interest and selecting it for further processing. Next, users specify real-time data stream or archived data search. At the current time our focus is on the archived data but it will be relatively easy in the future to activate the real-time option. After selecting the basin or region of interest, the user then searches for the cases (periods) of interest. This is done using metadata. Once the periods are selected, the user specifies algorithm options (discussed by Krajewski et al. 2007, this volume). Again, description of these options is provided in plain language, avoiding radar jargon, so that hydrologist can think in terms of rainfall variable and not so much the radar context. Specification of the grid resolution and formatting option completes the dialog. What follows (behind the scenes) is automatic building of a script that executes the user's request. At this stage we provide the user with an estimate of the time that it takes to complete the request.

Currently, it takes about one day to process data and produce hourly products for one year worth of Level II data. This represents 10-20 times improvement vs. using CODE (see above) from the Office of Hydrology of the NWS.

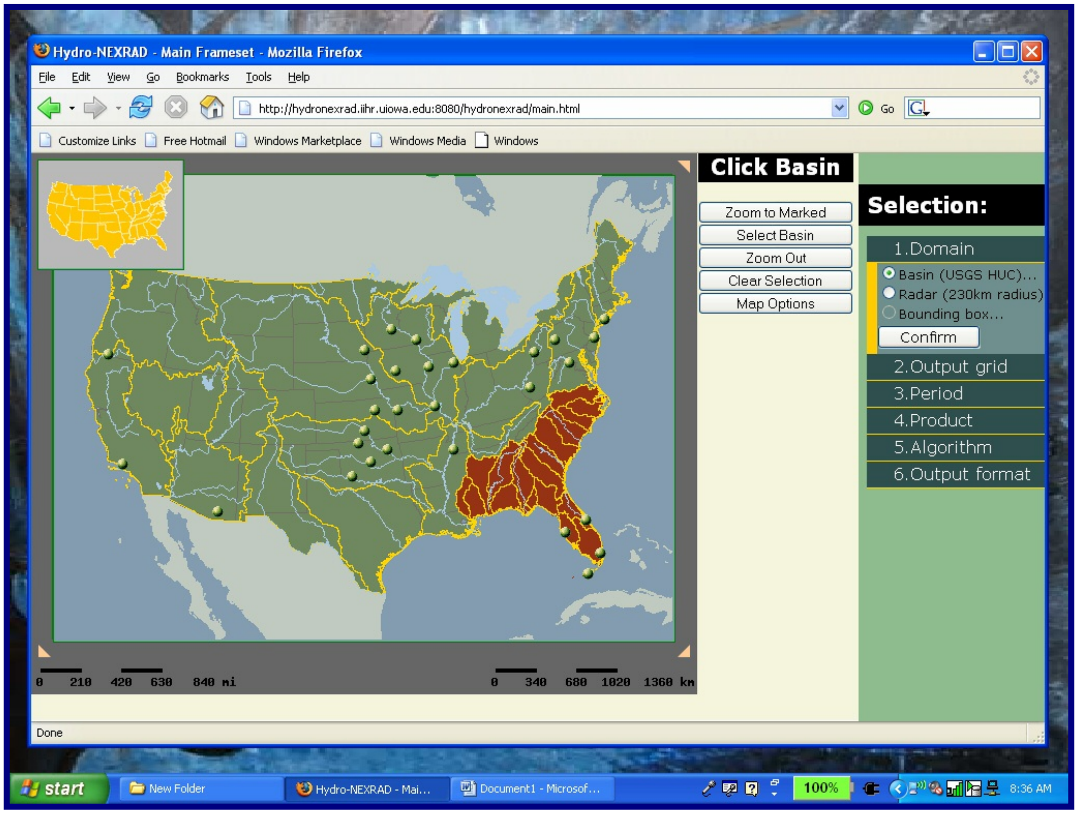

Figure 1. A screen shot of the Hydro-NEXRAD GUI. The menu on the RHS shows the main phases of product selection. The menu expands showing available options when the particular phase becomes relevant (active).

Summary

Hydro-NEXRAD is over-the-Internet-accessible software that provides custom radar-rainfall maps. The software can also be used to create real-time data stream and rainfall products generation over a specified region at a specified spatial grid. Currently, we collect data from some 30 WSR-88D radars (see Figure 1) around the country and have close to 100 radar-years of Level II data in our 150 TB database.

At the present time the product are radar-only estimates of rainfall but could be easily merged with rain gauge estimates to reduce systematic and random errors involved in the measurement and estimation process. Discussing the problem of optimal merging is beyond the scope of this paper, but any minimum error variance procedure obviously requires quantified knowledge of products uncertainty. An uncertainty model limited to the official radar-only products generated by the NWS PPS is described by Ciach et al. (2007).

Acknowledgments: This work was supported with by the National Science Foundation Award ATM 0427422.

References

Anagnostou, E.N. and W.F. Krajewski, Real-time radar rainfall estimation. Part 1: algorithm formulation, *Journal of Atmospheric and Oceanic Technology*, 16(2), 189-197, 1999.

Ciach, G.J., W.F. Krajewski, E.N. Anagnostou, J.R. McCollum, M.L. Baeck, J.A. Smith, and A. Kruger, Radar rainfall estimation for ground validation studies of the Tropical Rainfall Measuring Mission, *Journal of Applied Meteorology*, 36(6), 735-747, 1997. Ciach, G.J., W.F. Krajewski, G. Villarini, Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data, submitted to *Journal of Hydrometeorology*, 2007.

Fulton, R.A., J.P. Breidenbach, D.J. Seo, and D.A. Miller, WSR-88D rainfall algorithm. *Weather and Forecasting*, 13, 377-395, 1998.

Krajewski, W.F., B.C. Seo, A. Kruger, P. Domaszczynski, G. Villarini and C. Gunyon, Hydro-NEXRAD radar-rainfall estimation algorithm development, testing and evaluation, In preprints of ASCE's World Environmental and Water Resources Congress 2007, Tampa, Florida, May 15-19, 2007.

Kruger, A., W.F. Krajewski, P. Domaszczynski, C. Gunyon, J.A. Smith and A.A. Bradley, Hydro-NEXRAD metadata computation and use, In preprints of ASCE's World Environmental and Water Resources Congress 2007, Tampa, Florida, May 15-19, 2007.

Kruger, A. and W.F. Krajewski, Efficient storage of weather radar data, *Software Practice and Experience*, 27(6), 623-635, 1997.

Kruger, A., R. Lawrence, and E.C. Dragut, Building a Terabyte NEXRAD radar database for hydrometeorology research, *Computers & Geosciences*, 32(2), 247–258, doi:10.1016/j.cageo.2005.06.001, 2006.

Grecu, M., and W.F. Krajewski, An efficient methodology for detection of anomalous propagation echoes in radar reflectivity data using neural networks, *Journal of Oceanic and Atmospheric Technology*, 17(2), 121-129, 2000.

Reed, S.M., and D.R. Maidment, Coordinate transformations for using NEXRAD data in GIS-based hydrologic modeling, *Journal of Hydrologic Engineering* 4, 174-182, 1999.

Steiner, M., and J.A. Smith, Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data, *Journal of Atmospheric and Oceanic Technology*, 19(5), 673-686, 2002.

Vignal, B. and W.F. Krajewski, Large sample evaluation of two methods to correct range-dependent error for WSR-88D rainfall estimates, *Journal of Hydrometeorology*, 2(5), 490-504, 2001.