Hydro-NEXRAD: metadata computation and use

Anton Kruger, Witold F. Krajewski, Piotr Domaszczynski and
James A. Smith

ABSTRACT

To enable the efficient use of the Hydro-NEXRAD system in hydrologic research, the authors are developing and computing a set of metadata. Vast archives of radar-centric NEXRAD data, the highly variable spatial and temporal characteristics of precipitation events, and the existence of anomalous data complicate the process of selecting subsets of the data for specific investigation. Well-defined, pre-computed, radar-centric, basin-centric and point (for a selected set of rain gauge locations) metadata offer the possibility of efficiently selecting data subsets that meet user-specified criteria. Metadata have to be computed quickly so that they do not slow down the data ingest process, but they do not have to be highly accurate. Separating rainfall from clear air conditions and heavy rainfall from light precipitation is often sufficient. In this paper, the authors discuss the technical challenges of computing metadata for the vast archives of NEXRAD data and basic data quality control incorporated into that process. Users of Hydro-NEXRAD will be able to query the database using different criteria based on the metadata or visually inspect the time series of the metadata.

Key words | Hydro-NEXRAD, metadata

Anton Kruger (corresponding author)
Witold F. Krajewski
Piotr Domaszczynski
IIHR - Hydroscience and Engineering,
The University of Iowa,
Iowa City, IA 52242-1585,
USA
Tel.: +1 319 335 5237
Fax: +1 319 335 5238
E-mail: *anton-kruger@uiowa.edu*

James A. Smith
Civil and Environmental Engineering,
Princeton University,
Princeton, NJ 08544,
USA

INTRODUCTION

With the increasing variety and quality of available atmospheric measurements, researchers must have good access to data generated by scientific institutions.

For example, since its commencement almost 20 years ago, the NEXRAD system has been providing precipitation data with high spatial and temporal resolution for the entire United States. NEXRAD uses Doppler capabilities to detect precipitation intensity and enable forecasters to examine wind speed and direction associated with severe thunderstorms. Vast archives of high resolution precipitation data have the potential to increase our understanding of many hydrologic processes. However, the existing data format and the high terabyte volume of data require a high level of expertise to navigate and limit the broader use of NEXRAD data in hydrologic studies (Krajewski *et al.* 2009).

A challenging problem is to develop the ability to extract meaningful subsets of data that are applicable to a given research problem. We call this the data-finding problem. It is related to, but different from, data mining which is generally understood as the process of looking for hidden relationships in a large dataset (Grossman *et al.* 2001). A subset of interest may represent less than 10% of the total size of the dataset, but the task of finding a relevant subset is becoming an increasingly major endeavor.

The difficulty of managing the metadata increases with the size of the NEXRAD dataset. There is simply too much data for a user to manually locate an interesting subset of the archive. Catalogs have some utility, but they are static and tied to preconceived notions of what is important in a dataset. In Hydro-NEXRAD, we compute a small set of fairly generic metadata or descriptive statistics about the underlying NEXRAD data and then manage the metadata in a relational database. Metadata organized in a database automatically obtains functionality native to all database

doi: 10.2166/hydro.2010.057

stored data, and by that can be queried in flexible ways by users to create "catalogs" of data relevant to their needs.

To reflect the basin-centered perspective of hydrologic applications, a logical metadata tool would be time series of basin-averaged rainfall rate. We could integrate the basin framework in a variety of ways, e.g. river network files, products derived from seamless Digital Elevation Models, user-specified masks and standard point locations on river networks including United States Geological Survey stations, etc. The domain scientists will determine other metadata such as anomalous propagation and ground clutter masks that flag suspect pixels.

Solving the selection problem is not trivial. The authors of the Hydro-NEXRAD project decided to use a metadata approach as the most suitable for hydrologic applications. Since the Hydro-NEXRAD is a prototype system, its metadata component should also be seen as a prototype approach. We use data from 40 NEXRAD radars, and the set of metadata is limited to metadata which, to our understanding, best serves users with hydrologic focus. A set of point, radar-centric and basin-centric metadata has been defined and is being computed and stored in the Hydro-NEXRAD relational database. The purpose of this paper is to present a methodology and discuss challenges of the metadata computation. All Hydro-NEXRAD metadata can be subdivided into three groups: radar metadata, basin metadata and point metadata. The metadata computation methodology and a detailed description of each metadata group follows.

METADATA COMPUTATION METHODOLOGY

For many scientific initiatives, radar-measured precipitation is the main source of rainfall information. Searching vast archives of high spatial and temporal resolution NEXRAD data can be tedious and time-consuming. In many hydrologic studies, it is the basin rather than the radar spatial domain that is of hydrologic interest. Relating radar data to basin metadata is part of the metadata computation methodology. A well-defined set of metadata can clearly identify periods of time when radars recorded high reflectivity values over significant coverage areas. Such information can significantly improve the process of data subset selection.

The authors have specified requirements for metadata to be a valuable descriptor of available data from a hydrologic point of view. Metadata are computed separately for radars, USGS hydrologic units (basin metadata) and points. Computations of metadata should be fast enough not to inhibit the data ingest process.

Radar metadata

Depending on the Volume Coverage Pattern (VCP), all NEXRAD radars generate Level II files with time resolutions ranging from 4 to 12 min. Within the Hydro-NEXRAD system, each Level II file is converted into an RLE format file (Kruger & Krajewski 1997). The RLE format allows for significant file size reduction and quality control of the original data. Based on each RLE radar file, a Constant Altitude Plan Position Indicator (CAPPI) scan is calculated. A CAPPI scan is a fixed structure of 360 rays, each consisting of 230, 1 km, grid cells. Using a 1.5 km CAPPI scan, as opposed to a base scan (0.5° elevation), for metadata calculation is necessary to eliminate ground clutter echoes, especially in radar proximity. From a CAPPI scan, a set of radar-based metadata is computed. We discuss members of this set below.

Percentage coverage with reflectivity above a threshold

Each cell from a CAPPI scan is checked for its reflectivity value. If the value exceeds any of the four threshold levels (20 dBZ, 30 dBZ, 40 dBZ and 50 dBZ), the corresponding metadata value is increased accordingly. When all cells have been checked, final metadata values are normalized by the total radar coverage area.

An example formula for percentage coverage, C, of reflectivity above threshold Z_{thresh} is presented as

$$C(Z_{\text{thresh}}) = \frac{100\%}{A_R} \sum_{i=1}^{360} \sum_{j=1}^{230} \left(A_{ij} \cdot T_{ij}(Z_{\text{thresh}}) \right) \quad (1)$$

where A_R is the nominal radar coverage area based on 230 km range, A_{ij} is cell area (m^2) at azimuth i and range j, and T_{ij} is an indicator function defined as

$$T_{ij}(Z_{\text{thresh}}) = \begin{cases} 1 \text{ if } Z_{ij} \geq Z_{\text{thresh}} \\ 0 \text{ if } Z_{ij} < Z_{\text{thresh}} \end{cases} \quad (2)$$

where Z_{ij} is the radar reflectivity value at azimuth i and range j.

Selected metadata thresholds can be qualitatively related to different precipitation regimes: 20 dBZ approximates the rain/no rain threshold; 30 dBZ corresponds to light rain; a threshold of 40 dBZ indicates heavy rain and values >50 dBZ indicate heavy rain, and possibly hail. Each metadata can be viewed in the Hydro-NEXRAD system in the form of a time series, as presented in Figure 1.

This relation is valid for NEXRAD WSR-88D radars where conversion from reflectivity values (dBZ) to rainfall rate (mm/h) is based on the Z–R power law with coefficients $a = 300$, $b = 1.4$:

$$Z = a \cdot R^b \tag{3}$$

where Z has units of $\mathrm{mm^6\,m^{-3}}$ and R is in $\mathrm{mm\,h^{-1}}$.

Maximum observed reflectivity value

From a CAPPI scan, a cell with the highest reflectivity (dBZ) value is selected:

$$Z_{\mathrm{MAX}} = \max\left\{\left(Z_{ij}\right)\right\} \tag{4}$$

where Z_{ij} is the reflectivity value at ith range and jth azimuth.

Note that reflectivity values greater than 55 dBZ can often be attributed to the “bright band” phenomenon, AP or clutter (buildings, birds, terrain, trees, etc.).

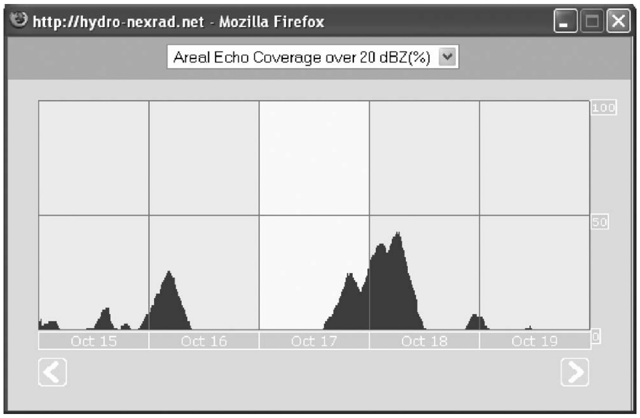


Figure 1 | Example of percentage coverage with reflectivity above 20 dBZ metadata time series.

Detected anomalous propagation echo

Using an algorithm developed by Steiner (Steiner & Smith 2002), each radar file in the Hydro-NEXRAD system undergoes an anomalous propagation echo (AP) detection process. Radar cells are checked for possible AP echoes using vertical and horizontal echo structures. Detected Anomalous Propagation metadata stores the total number of radar cells classified as AP and can be used as an indicator of weather conditions supporting AP occurrence:

$$N_{AP} = \sum_{i=1}^{360}\sum_{j=1}^{230}\left(AP_{ij}\right) \tag{5}$$

where AP_{ij} is an indicator function defined as in (2) but operated on the results of the classification (“AP” or “no AP”) determined by the detection algorithm.

Echo advection vector

Using CAPPI scan information from two consecutive radar files, separated by a maximum 20 min time interval, the Hydro-NEXRAD metadata system estimates advection vectors and shows the magnitude and direction of horizontal echo movement. To expedite the calculations, only one vector per radar field is computed. The method is based on Fabry *et al.* (1994).

Volume Coverage Pattern (VCP)

During the data acquisition process, the WSR-88D radar antenna is controlled by an automatic scanning program. VCP is selected to optimize data gathering for a given meteorological situation. Since VCP information is not embedded in the RLE format, information on VCP is extracted from each radar file based on the number of radar elevation scans, the PRF sequence used and the scan time. VCP information is sufficient to understand the radar volume structure for a given scan.

Volume scan duration

Depending on the Volume Coverage Pattern (VCP), it can take approximately 4 to 12 min to complete one volume scan. This does not hold true for custom scanning strategies, which are not uncommon in extreme weather events. To give more insight into radar scanning times, the authors decided to extract this information and make it accessible to Hydro-NEXRAD users.

Missing radar data

Missing radar data is designed to provide information on completeness of radar information over time. Missing radar data can be caused by technical problems, routine radar calibration and maintenance, corrupt data due to hardware or software problems, and other problems. When such vast amounts of data are analyzed, it is highly beneficial to provide users with information identifying when data is not available.

Radar metadata daily aggregates

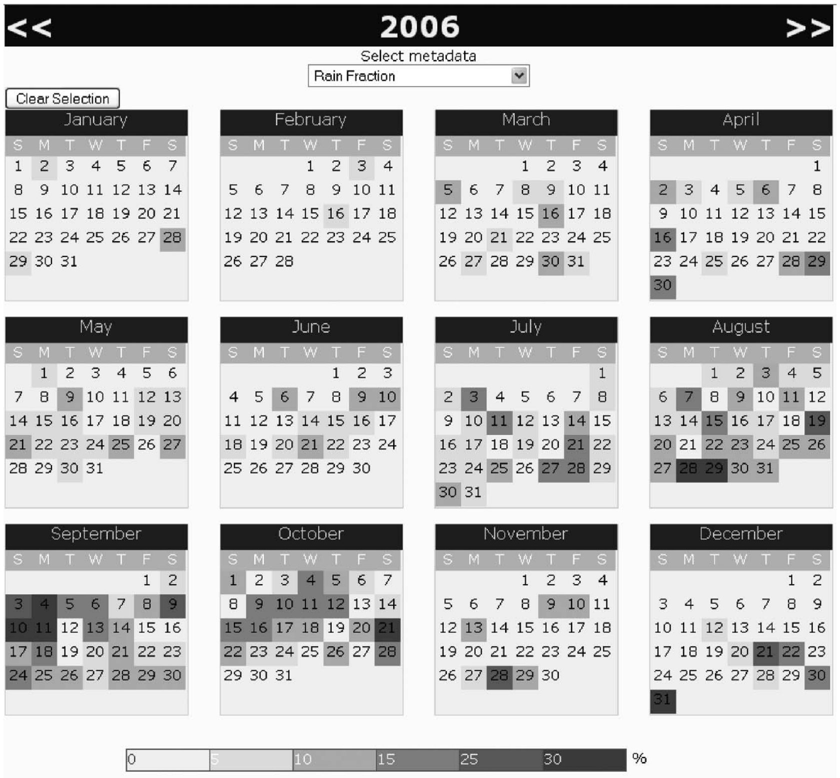
To enable metadata analysis on different timescales, daily metadata aggregates are computed. This allows for the display of metadata in a convenient calendar view (Figure 2). In the following subsection, we define a set of radar metadata daily aggregates.

Daily percentage coverage with reflectivity above a threshold

This group of daily aggregated metadata is based on "percentage coverage with reflectivity above a threshold" (previously computed) metadata. Metadata values from a given radar and for a given day are combined using the following expression:

$$DC(Z_{\text{thresh}}) = \frac{1}{N}\sum_{i=1}^{n}(C_i(Z_{\text{thresh}})) \tag{6}$$

where n is the number of $C(Z_{\text{thresh}})$ metadata occurrence during a given day, N is a number of $C(Z_{\text{thresh}})$ metadata with non-zero values and C_i is the ith occurrence value. This procedure is repeated for all of the thresholds, i.e. 20 dBZ, 30 dBZ, 40 dBZ and 50 dBZ.



2006

Select metadata: Rain Fraction

Figure 2 | Calendar view of Hydro-NEXRAD metadata. Darker color indicates higher metadata value. This visual inspection tool can help identify days with significant precipitation events.

Daily maximum reflectivity value

From the time series of maximum observed reflectivity value (Z_{MAX}) metadata, the highest daily value is selected:

$$DZ_{\mathrm{MAX}} = \max\{(Z_{\mathrm{MAX}_i})\}. \tag{7}$$

Daily rain fraction

Expressed as a percentage, the portion of a day when echoes classified as precipitation were recorded by a radar. To compute this value, two previously computed metadata are used: percentage coverage above 20 dBZ (as rain/no-rain indicator) and the volume scan duration:

$$\mathrm{DRF} = 100\% \frac{1}{1440}\sum_{i=1}^{n}(\mathrm{Volume\,Duration}_i) \tag{8}$$

where Volume Duration$_i$ is a metadata value representing volume scan duration time and n is a number of "percentage coverage with reflectivity above 20 dBZ" metadata occurrences within one day.

Daily heavy rain fraction

Using a 40 dBZ threshold as an indicator of heavy rainfall, the same procedure used to compute the daily rain fraction is applied to estimate the portion of daily radar records representing heavy rainfall:

$$\mathrm{DHRF} = 100\% \frac{1}{1440}\sum_{i=1}^{n}(\mathrm{Volume\,Duration}_i) \tag{9}$$

Completeness of daily radar data

These metadata represent the aggregation of missing radar data periods to a daily scale:

$$\mathrm{Missing\,Data} = \frac{100\%}{1440}\sum_{i=1}^{n}(\mathrm{Missing\,Radar\,Data}_i) \tag{10}$$

where Missing Radar Data$_i$ represents the duration of the ith missing data period.

Basin metadata

Hydrologic unit codes (HUC) developed by the USGS are a way of identifying all of the drainage basins in the United States in a nested arrangement from largest (regions) to smallest (cataloging units). For each of these hydrologic units, the Hydro-NEXRAD system computes a set of metadata with a fixed 10-min temporal resolution. Since every basin has its own irregular shape and size, and multiple radars can cover the same area (Figure 3), a multistep procedure has been adopted to allow for efficient and accurate computation of metadata.

First, a reflectivity mosaic is generated. A Hydro-NEXRAD mosaic is a 1′ longitude by 1′ latitude grid covering the continental United States. Each grid cell has an index and is designed to hold reflectivity values. CAPPI products from radar files are used to fill the mosaic with reflectivity values. WSR-88D NEXRAD radars are not synchronized with each other in time, so we have chosen 10-min time windows for calculating basin metadata. Thus, every 10 min there is a new value that characterizes rainfall (or no rainfall) for each cell in the mosaic. All HUC basins are identified in the Hydro-NEXRAD system as a set of mosaic cells (Figure 4). Pre-computed look-up tables, storing mosaic-to-basin relationships, make the process of metadata computation fast. In the following we define basin metadata.

Mean areal precipitation

From all the cells constituting a given basin, cells with reflectivity values greater than 20 dBZ (rain/no rain threshold) are selected and converted to rain rate using the $Z-R$ power law with coefficient $a = 300$ and $b = 1.4$. We use only one set of parameters for the $Z-R$ power law.

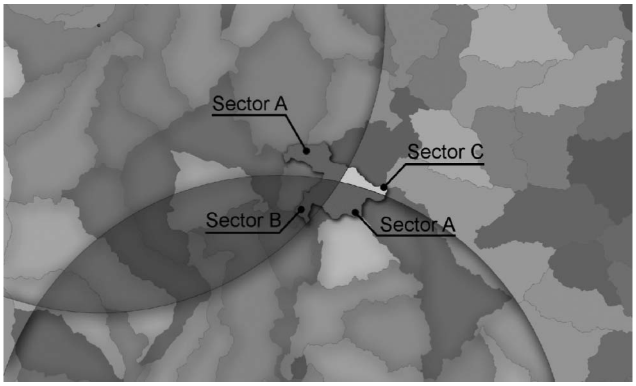


Figure 3 | NEXRAD radars overlapping over a basin. Metadata gives information on what part of a basin is covered with radar information.

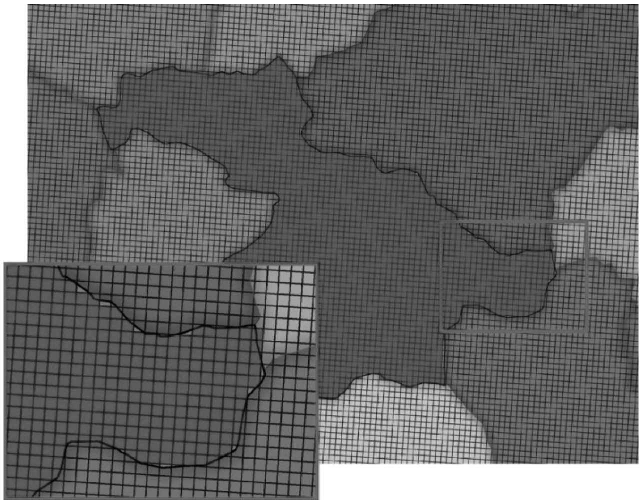


Figure 4 | Identification of a basin on Hydro-NEXRAD mosaic. Basin masks, stored as look-up tables, speed up the process of metadata computation.

This is in agreement with our philosophy of fast metadata computation and eliminates the problem of stratiform versus convective precipitation identification at this stage. The estimated value is then normalized by the total number of cells constituting a basin N_B:

$$R_{\mathrm{MAP}} = \frac{1}{N_B}\sum_{i=1}^{n}(RR_i \cdot T_i(Z_{\mathrm{thresh}})) \tag{11}$$

where RR_i represents the rainfall rate value of the ith basin cell and T_i is an indicator function defined by (2).

Conditional mean areal precipitation

This metadata estimates mean precipitation over the precipitating area. The symbolic definition is similar to the mean areal precipitation definition, but its normalization is done by taking the number of cells where precipitation was observed (N_R):

$$R_{\mathrm{CMAP}} = \frac{1}{N_R}\sum_{i=1}^{n}(RR_i \cdot T_i(Z_{\mathrm{thresh}})). \tag{12}$$

Convective conditional mean areal precipitation

To indicate strong convective precipitation, only cells with reflectivity values greater than 40 dBZ are considered and normalized by the total precipitating area:

$$R_{\mathrm{CMAP}} = \frac{1}{N_R}\sum_{i=1}^{n}(RR_i \cdot T_i(Z_{\mathrm{thresh}})). \tag{13}$$

Fractional basin coverage

Metadata closely related to mean areal precipitation represent the percentage coverage of a basin with echoes of reflectivity greater than 20 dBZ:

$$\mathrm{FBC} = \frac{100\%}{N_B}\sum_{i=1}^{n}(T_i(Z_{\mathrm{thresh}})) \tag{14}$$

where T_i is an index function defined by (2).

Convective fractional basin coverage

This estimate shows how much of a basin area is covered with strong convective precipitation. To estimate convective fractional basin coverage, the sum of all the cells with reflectivity values greater than 40 dBZ is normalized with the total number of cells constituting a basin and expressed in a percentage:

$$\mathrm{CFBC} = \frac{100\%}{N_B}\sum_{i=1}^{n}(T_i(Z_{\mathrm{thres}})) \tag{15}$$

where T_i is an indicator function defined by (2).

Maximum observed rainfall rate

For a given basin, the highest observed rainfall rate is stored with 10-min temporal resolution:

$$R_{\mathrm{MAX}} = \max\{(RR)\}. \tag{16}$$

Location of maximum observed rainfall rate

This metadata stores the location of the maximum observed rainfall rate in geographical coordinates (decimal degrees of latitude and longitude). The estimated location corresponds to the center of the cell with the highest rainfall rate value.

Basin coverage with radar information

The Hydro-NEXRAD system stores data from approximately 40 NEXRAD radars. These radars provide precipitation information for some 800 basins (USGS cataloging units), some of which are only partially covered by radar data. Additionally, radars undergo routine maintenance, during which time no radar data is available. Initially, mosaic cells are set to a "no value" flag. After the process of filling mosaic cells with radar data is over, each basin is checked for the number of "empty cells." This number is then normalized by the total number of cells constituting a basin to calculate what part of a basin area does not have radar information:

$$BC = \frac{1}{N_B}\sum_{i=1}^{n}(T_i) \tag{17}$$

where T_i is an indicator function defined by (2) and operated on the existence of valid values.

Basin metadata daily aggregates

Similarly to radar metadata, these basin metadata are also aggregated to a daily scale to allow for efficient metadata display and analysis. Below, we present the mathematical definitions of basin metadata daily aggregates.

Total precipitated water per day

The expression for this metadata has been designed to calculate how much water (in m^3) is precipitated over a basin in a given day:

$$\mathrm{TPW} = \frac{0.001}{6}\sum_{i=0}^{n}(R_{\mathrm{MAP}i}\cdot A_B) \tag{18}$$

where $R_{\mathrm{MAP}i}$ is the ith observed mean areal precipitation metadata and A_B is the basin's area expressed in m^2. A_B information is calculated based on USGS data and is stored in the Hydro-NEXRAD database.

Daily conditional average rainfall rate

Based on conditional mean areal precipitation (R_{CMAP}) metadata, we can estimate an average daily rainfall rate by summing all observed R_{CMAP} metadata and normalizing the sum by the total number of instances greater than zero:

$$DR_{\mathrm{CMAP}} = \frac{1}{N}\sum_{i=0}^{n}(R_{\mathrm{CMAP}i}\cdot T_i(R_{\mathrm{thresh}})) \tag{19}$$

where R_{CMAP} is the observed conditional mean areal precipitation rate, N is equal to the number of all instances of $R_{\mathrm{CMAP}} > 0$ in a day and T_i is an indicator function defined as

$$T_i(\mathrm{R_{thresh}}) = \begin{cases} 1 \text{ if } R_i \geq R_{\mathrm{thresh}} \\ 0 \text{ if } R_i < R_{\mathrm{thresh}} \end{cases} \tag{20}$$

where R_i is the rainfall rate value at basin bin i and $R_{\mathrm{thresh}} = 0$.

Daily maximum basin coverage

To indicate rainfall events when substation basin coverage occurs, we select the highest value of fractional basin coverage metadata (FBC) observed during a day:

$$\mathrm{DFBC_{MAX}} = \max\left\{\sum_{i=1}^{n}(\mathrm{FBC}_i)\right\}. \tag{21}$$

Daily maximum convective basin coverage

This daily metadata is similar to the daily maximum basin coverage, but it operates on CFBC metadata and indicates days where strong convective precipitation was observed over a substantial basin area:

$$\mathrm{DCFBC_{MAX}} = \max\left\{\sum_{i=1}^{n}(\mathrm{CFBC}_i)\right\}. \tag{22}$$

Daily maximum observed rainfall rate

For a given basin, the highest observed rainfall rate observed during a given day is stored:

$$DR_{\mathrm{MAX}} = \max\left\{\sum_{i=1}^{n}(R_{\mathrm{MAX}i})\right\}. \tag{23}$$

Location of maximum observed rainfall

This metadata stores the location of the maximum observed rainfall rate in geographical coordinates (decimal degrees of latitude and longitude). The estimated location corresponds to the center of the cell with the highest rainfall rate value observed over a basin during a given day.

Daily average basin coverage with radar information

Over a period of one day, a basin's coverage with radar information may change. This can be caused by radar malfunction, incomplete datasets, partial basin coverage, etc. To better indicate a basin's coverage with radar information during a day, Hydro-NEXRAD computes an average value of basin coverage with radar information:

$$\mathrm{DBC}_{\mathrm{AVG}} = \frac{1}{N}\sum_{i=0}^{n}(BC_i). \tag{24}$$

Point metadata

In the Hydro-NEXRAD system, the locations of approximately 1000 rain gauges are stored. The selected gauges are members of the ASOS network (ASOS Program Office 1992), Oklahoma Mesonet and Oklahoma Micronet that have Hydro-NEXRAD radar coverage. For each of these locations, point metadata is computed with 10-min temporal resolution. The geographical location of a rain gauge is converted to cell indices of the pre-computed mosaic file. Point metadata for a single location has the form of a 3 × 3 matrix to provide information about rainfall directly over a point and its neighboring cells (Figure 5). Such formulation of point metadata allows users to obtain more information about the spatial extent of rainfall experienced over a rain gauge and creates the possibility of adjusting radar data with rain gauge data.

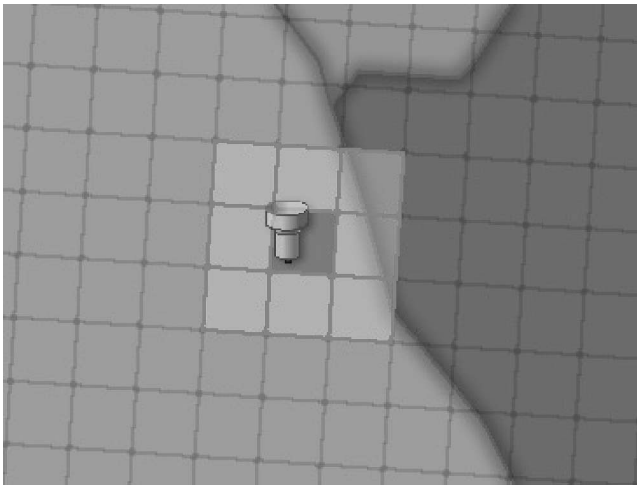


Figure 5 | Point metadata conceptual image with a rain gauge location indicated.

USING METADATA

Metadata is used to easily identify the relevant part of the dataset. As the dataset grows, finding relevant data becomes an important issue. There is simply too much data for a user to search for relevant portions manually. Furthermore, "relevant" has different meanings to different users. One person may be interested in periods of heavy precipitation, another may be developing an algorithm for anomalous propagation (Krajewski & Vignal 2001), while another may be developing algorithms related to bright band (Borga *et al.* 1997). All these users need to identify subsets of the dataset.

Our solution is to generate descriptive statistics (metadata) for each data file as well as for basins and points. These statistics can then be used to precisely identify the subset of the data with the required properties. Users and user-defined software scripts can query the database and receive a list of files or time stamps that matches their query. We worked with domain experts (hydrologists and hydrometeorologists) to help create the metadata. As an example of using metadata, consider the following English-language query:

> Find all KDVN radar volume files with more that 20% of radar coverage showing reflectivity values above 30 dBZ but maximum observed reflectivity less than 55 dBZ.

Here we used areal coverage with precipitation above 30 dBZ and maximum reflectivity to create a more sophisticated composite search. It is not hard to imagine many other powerful searches based on a few basic descriptors.

The metadata described so far are volume-centric and location-specific. Useful as this is, one can increase the value dramatically by including a temporal or spatial context for data from multiple radars. Thus, with a temporal context the query above becomes (see the part in bold):

> Find when, **within the last 48 h**, mean areal precipitation over the Upper Cedar basin (07080201) was above 20 mm/h with rainfall covering more than 50% of basin area.

It is important to note that these algorithms will operate on the metadata to identify data of interest and not on the data itself, which greatly improves retrieval time. Execution time of a metadata query depends on its complexity, but for most cases should not exceed one hour, whereas raw data read time for a comparable query can result in days to weeks of computation time.

Time series metadata

Time series of metadata provide an opportunity for researchers to explore data on a fine temporal scale. However, the different spatial scales of NEXRAD WSR-88D radars, hydrologic basins and single points suggest that each should be observed individually. As can be seen in Figure 6, a different spatial extent leads to different characters of observed rainfall and thus to different metadata time series. Analyzing metadata at a scale comparable to the area of interest increases accuracy and efficiency of the data search process.

Metadata daily aggregates

Converting metadata time series to single-value daily representations has to be done carefully so as not to disguise the hydrological significance of events. In the Hydro-NEXRAD system, three aggregation techniques are used. The first technique is conditional averaging of metadata values. Conditional averaging allows for short-term events of high intensity to be accordingly represented as single values on a daily scale projection. The second technique selects maximum values from metadata time series. This method is very useful in identifying extreme events in both the spatial extent of precipitation over a domain and extreme rainfall intensities. It is not uncommon for NEXRAD radars to record extremely high reflectivity values. These high values can be attributed to the "bright band" phenomenon but also to clutter. The third technique represents a summation of all observed metadata values. Summation is used to calculate total precipitated water per day.

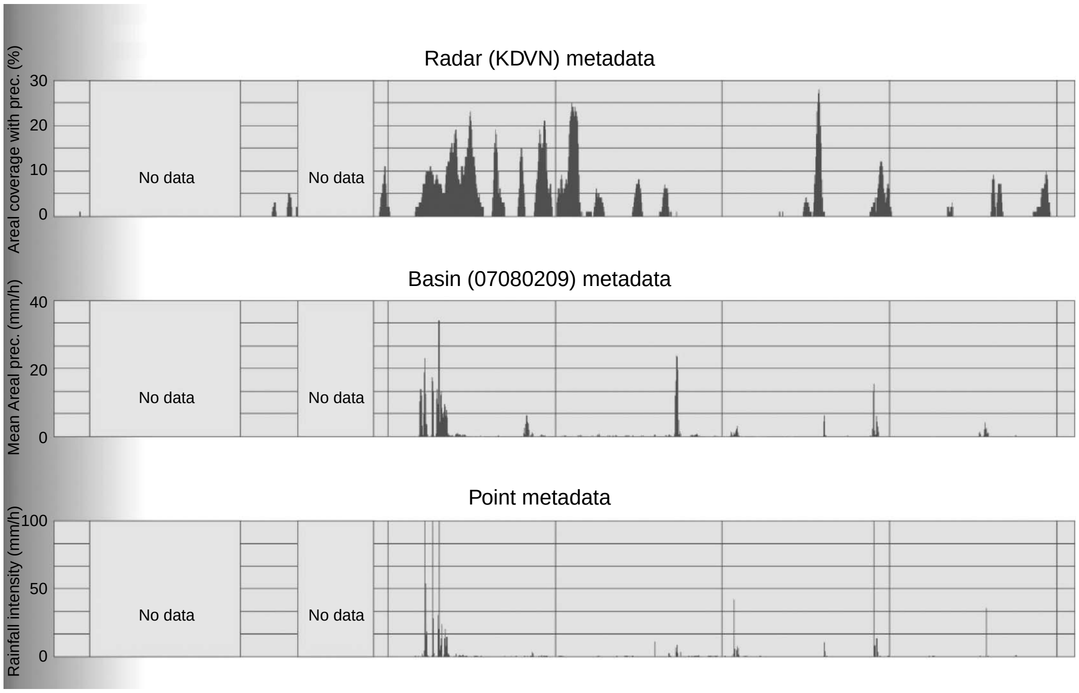


Figure 6 | Comparison of metadata over different spatial scales.

CONCLUSIONS

In the development process of the Hydro-NEXRAD system, the authors assumed that metadata does not have to be highly accurate, but it does have to be computed quickly. Stored in the Hydro-NEXRAD database, pre-computed metadata enhances the possibility of making an efficient selection of a user-specified subset of NEXRAD-based precipitation data. Selected methods included necessary quality control and allowed for a fast and accurate computation. The rich set of metadata, available to Hydro-NEXRAD users in the form of time series plots and a colored calendar view, greatly improves the search and selection of data subsets and provides data quality overview and quality control. The modular architecture of the metadata computation algorithm makes it easy to expand the existing set of metadata in the future.

REFERENCES

ASOS Program Office 1992 *Temperature/Dewpoint Sensor. Automated Surface Observing System Site Technical Manual S100*. AAI Systems Management, Inc., Las Vegas, Nevada.

Borga, M., Anagnostou, E. N. & Krajewski, W. F. 1997 A simulation approach for validation of a brightband correction method. *J. Appl. Meteorol.* **36** (11), 1507–1518.

Fabry, F., Bellon, A., Duncan, M. R. & Austin, G. L. 1994 High resolution rainfall measurements by radar for very small basins: the sampling problem reexamined. *J. Hydrol.* **161**, 415–428.

Grossman, R. L., Kamath C., Kegelmeyer, P., Kumar, V. & Namburu, R. (eds) 2001 *Data Mining for Scientific and Engineering Applications*. Kluwer, Amsterdam.

Krajewski, W. F., Kruger, A., Smith, J. A., Lawrence, R., Gunyon, C., Goska, R., Seo, B., Domaszczynski, P., Beack, M. L., Ramamurthy, M. K., Weber, J., Bradley, A. A., DelGreco, S. A. & Steiner, M. 2009 Towards better utilization of NEXRAD Data in hydrology: an overview of Hydro-NEXRAD. In *Proceedings of the 2007 World Environmental and Water Resources Congress, May 15–19, 2007, Tampa, Florida; Sponsored Environmental and Water Resources Institute (EWRI) of ASCE* (ed. K. C. Kabbes). ASCE, Reston, VA, doi: 10.1061/40927(243)288.

Krajewski, W. F. & Vignal, B. 2001 Evaluation of anomalous propagation echo detection in WSR-88D data: a large sample case study. *J. Atmos. Oceanic Technol.* **18** (5), 807–814.

Kruger, A. & Krajewski, W. F. 1997 Efficient storage of weather radar data. *Softw. Pract. Exp.* **27** (6), 623–635.

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

First received 29 July 2009; accepted in revised form 11 December 2009. Available online 29 April 2010