EOSC 510 - Term Project: Assessing Similarity of Western Canadian River Basins Using Runoff and Spatial Distribution of Precipitation

Dan Kovacek (35402767)

2020-04-06

Contents

1 Abstract

Hydrological analysis is often undertaken on the basis that historical records from one basin can be used as a proxy for the characterization of long-term flow characteristics in an ungauged watershed of reasonably similar physiography and location. Across British Columbia and Alberta, mountainous river basins exhibit a large amount of localized variability, confounding efforts to develop relationships that apply accurately at the regional level. In this study, historical weather radar imagery is gathered for 144 basins in British Columbia and Alberta. Spatial distribution of precipitation is estimated using a sample of precipitation events identified in summer and fall between 2007 and 2018. Self organizing maps (SOM) are used to identify similarity in spatial distribution of precipitation. SOMs are also applied to measured runoff at the same catchments to identify dominant runoff patterns. The precipitation and runoff based SOMs are compared. The results show…

2 Introduction

The characterization of water resources is a critical step in support of natural resource project proposals in Canada. Hydrological analysis is critical to the planning, design, permitting, and operation of mines and hydropower facilities in particular. The standard of resource engineering practice in assessing precipitation and runoff is often limited to quantification at seasonal or annual levels due to the limitations in the spatiotemporal resolution of available data. Single point-in-space measurements are routinely applied to represent average precipitation across large areas, yet spatial distribution of precipitation is known to be highly variable. As a result, there is considerable uncertainty associated with applying precipitation measurements to an ungauged basin of interest for analytical purposes, and derivative metrics such as runoff ratio and evapotranspiration are accordingly highly uncertain. Similarly, regional relationships are applied in practice to adjust long-term estimates from one location to other ungauged locations some distance away.

Weather radar approximates the density of precipitation by measuring the reflection of transmitted microwave pulses off of water droplets in the air, enabling the estimation of spatial distribution of precipitation with reasonable resolution across large areas. (Thorndahl et al. 2017) details the advantages of precipitation radar over traditional measurement methods, and as well discusses the difficulties in collecting, processing, and interpreting radar data and the limitations and uncertainty inherent in its use. The goal of this research is to investigate the capacity of weather radar data for estimating runoff at ungauged locations. Historical weather radar data from five stations in British Columbia (BC) and Alberta (AB) are combined with concurrent streamflow data from 144 hydrometric stations within sensing range of the radar stations in order to compare precipitation with resultant runoff.

3 Data

3.1 Historical Weather Radar

Environment Canada (EC) provides free access to historical weather radar data as far back as 2007 for some radar stations, though programmatic access is not provided at the time of writing. Historical weather radar was as a result obtained by a web-crawling script targeting time periods associated with summer and fall runoff events. Radar images from five radar stations in BC and AB were collected based on their coverage of mountainous basins. The measurement range of radar is described by a circle with a radius of 250km centred at the radar station. An example radar image is shown in Figure 1

The resolution of radar imagery corresponds to 1 pixel for every $1km^2$, and the centre pixel represents the radar location. It is these two pieces of information that allow a reasonable projection of a coordinate system onto the radar image which is unreferenced in the form it is acquired. The processing of radar data is discussed further in Section X. Note the information layers shown in Figure 1, such as place names and concentric circles, are embedded in the image, causing issues for some catchments.

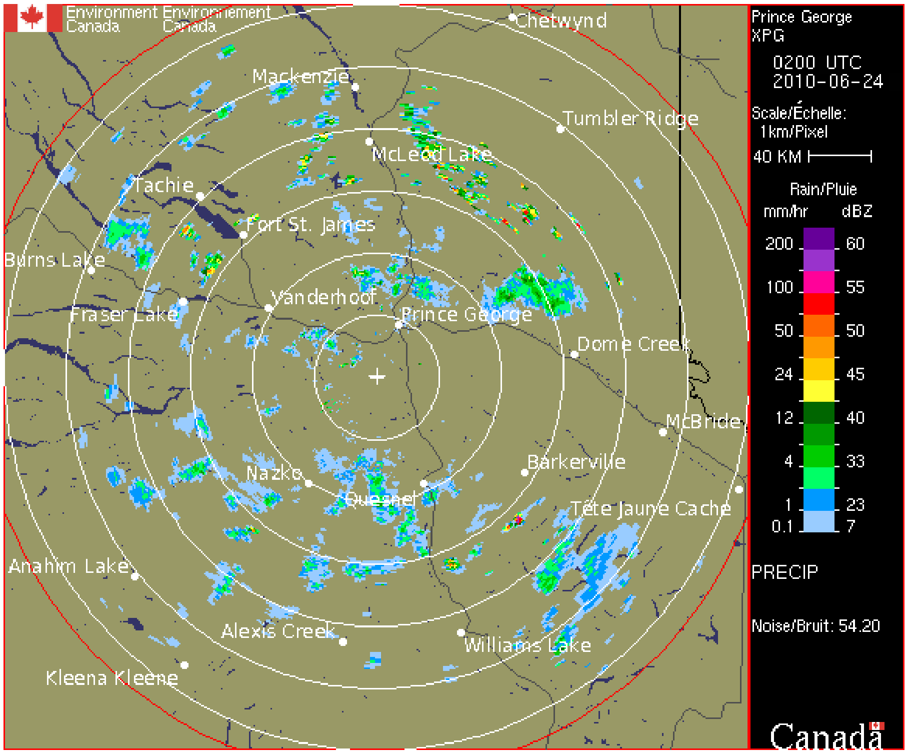

Figure 1: Example Radar Image (Prince George, BC. Source: Environment Canada)

3.2 Daily Average Streamflow (Runoff)

The Water Survey of Canada (WSC) provides open, programmatic access to historical daily average streamflow records for over 8000 active and inactive hydrometric stations across Canada. WSC provides open access to a database file (HYDAT) containing all historical WSC streamflow data, and the latest available database file is used in this study. An example streamflow time series is shown in Figure 2. The Bokeh data visualization library (Bokeh Development Team 2020) for the Python programming language is used for plotting Figure 2 and subsequent figures.

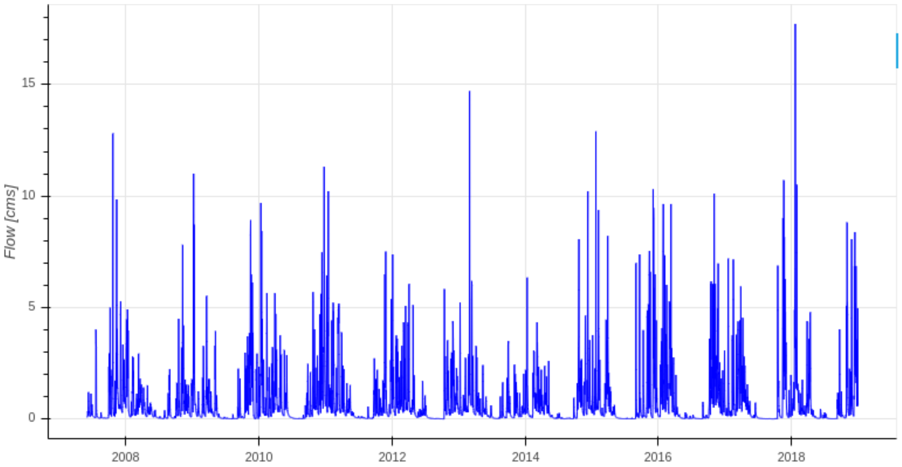

Figure 2: Example Daily Average Flow Timeseries (WSC 08HB048: Carnation Creek at the Mouth)

The HYDAT database is filtered to include stations in BC and Alberta that fall within the sensing range of a radar station, and to include stations with historical record concurrent with the weather radar stations. Since the radius of radar measurement is limited to 250 km, and because information layers embedded in the radar images obstruct some areas of the images, the WSC stations were also filtered to include stations capturing a drainage area of less than $1000 km^2$. A total of 141 stations were found to fit the filtering criteria. The smallest WSC catchments are in the order of $10 km^2$ which correspond to a mere ~10 pixels of the radar image. As a result, the accuracy of the smallest catchments is expected to be poor. The WSC stations satisfying the above criteria are plotted on Figure 3.

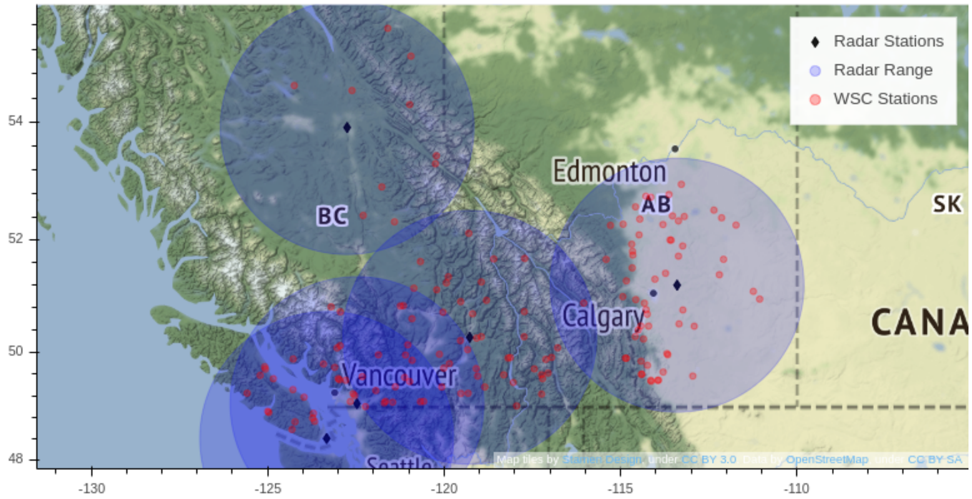

Figure 3: Radar Stations and Range (Approximate), and WSC Stations within Radar Coverage

3.3 Catchment Boundaries

Geographic polygons corresponding to most of the WSC hydrometric stations are available from the Government of Canada's Open Data Platform. Given the low resolution of the radar images, the shape files are believed to be suitable for the intended purpose of extracting from the radar images the pixels corresponding to each catchment of interest. An example catchment boundary with its corresponding station location is shown in Figure 4

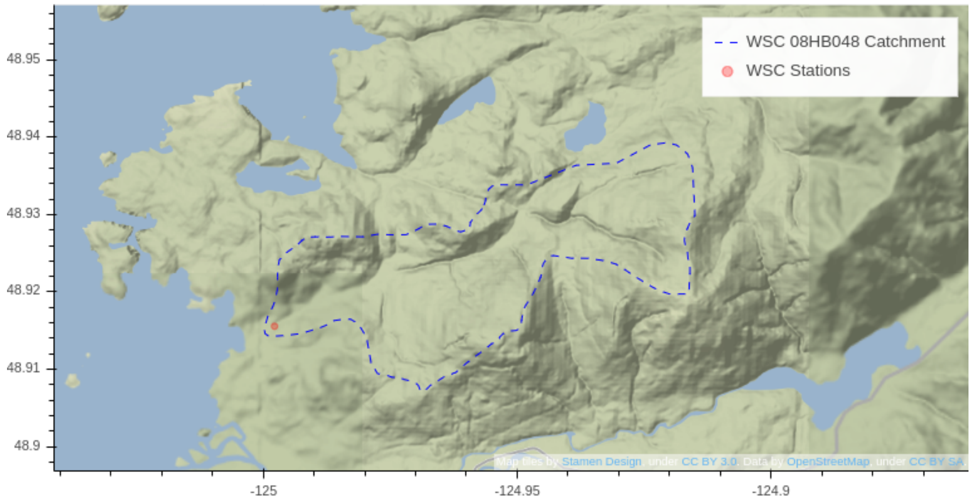
WSC 08HB048 Catchment
WSC Stations
48.95
48.94
48.93
48.92
48.91
48.9
-125
-124.95
-124.9
Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA

Figure 4: Catchment boundary for WSC 08HB048: Carnation Creek at the Mouth

4 Methodology

The primary research question involves the comparison of weather radar data and streamflow data to ultimately find clustering patterns of watersheds based on spatial distribution of rainfall, physiographic characteristics, and geographic location. The data acquisition process necessitates an additional analytical step which will first be addressed to set the context for the subsequent analysis. The remainder of data pre-processing is then described, including extraction of precipitation data from radar images corresponding to WSC station catchments, and the reconstruction of runoff hydrographs. Finally, the methods used to evaluate the predictive power of radar data are presented.

4.1 Data Acquisition and Preprocessing

4.1.1 Radar Image Acquisition

A limiting step in the data acquisition process is radar imagery retrieval. The number of server requests required to capture 12 years of radar images at 10-minute intervals at five different stations is in the order of 3×10^6, which is an excessive imposition on a free service, and it invites a ban from use. Focusing the study to summer and fall months simplifies the interpretation of radar data by avoiding precipitation as snow, but it does not alone reduce the number of server requests to a viable level. An anomaly detection (AD) algorithm is used to identify isolated runoff events between June and October (inclusive) in order to reduce the total number of radar image requests to a reasonable number. However, the execution time of a sufficiently complex AD algorithm could negate the (time) cost savings from a reduced number of image requests, and a sufficiently sensitive AD algorithm could label every oscillation in the input signal, no matter how minute, as a ‘runoff event’, resulting in more image requests and longer AD algorithm execution time. A tradeoff then exists between running the anomaly detection algorithm to reduce the total radar image server calls to a viable number, the time required for the AD algorithm execution, and the number of runoff events identified. The sample size for meaningful analysis adds a minimum constraint on the number of events identified. Details of the AD algorithm are discussed in the subsequent section before returning to the question of radar image preprocessing.

4.1.2 Anomaly Detection for Identifying Runoff Events

The identification of individual runoff events in streamflow signals is a difficult problem, and generalizing the problem across catchments of highly variable characteristics adds to the complexity. For the specific use case of this study, it is important to identify a sufficient number of samples (true positives) in order to support meaningful analysis in the subsequent steps addressing the primary research question, which requires representative estimates of spatial distribution of precipitation. False negatives (missing events) are considered less important than false positives (identifying a runoff event where there isn’t one) in reducing the quality of the dataset, as false positives tend to result in biased outcomes corresponding to either 0 or infinite runoff ratio.

An initial approach to identifying runoff events used a simple anomaly detection (AD) algorithm based on moving windows and a percentile-based outlier threshold on the differenced flow series. This moving-window approach proved ineffective at generalizing across all watersheds, and the results were difficult to validate and interpret. A modified approach using PCA was successful at generalizing across watersheds and at capturing data that were easier to validate. Figure 5 shows an example of the AD algorithm identifying three events in the period of interest (June to October, inclusive). A sample of the total events identified is plotted in Figure 6 in a small-multiples format to facilitate the validation the runoff events identified by the AD algorithm for each station.

The principal components (PCs) comprising a minimum 90% of the variance in lagged data are then used to create a time series corresponding to the Mahalanobis distance (MD) – the Euclidean distance from each row of data points (each detrended observation and its n lags) to the corresponding PCs. A threshold Mahalanobis distance then represents some magnitude of deviation from the PC within a timeframe correspondent with the number of lags (components). The runoff events identified by the AD algorithm correspond to the timestamps where the MD crosses the threshold (in either direction).

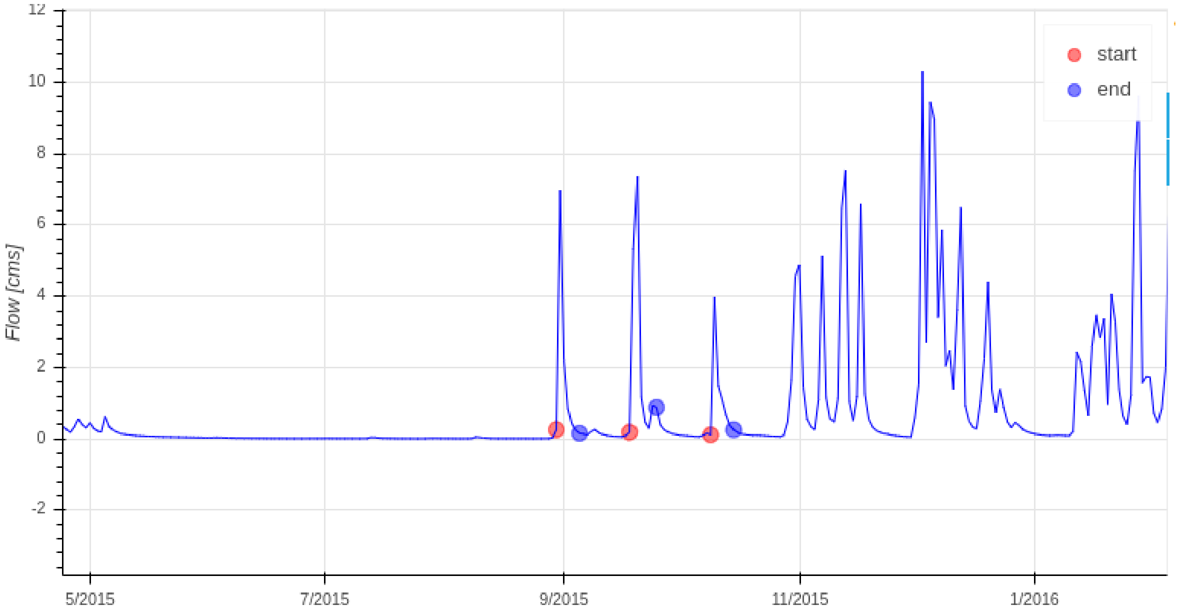

Figure 5: Example AD Results (Timeseries) for WSC 08HB048: Carnation Creek at the Mouth

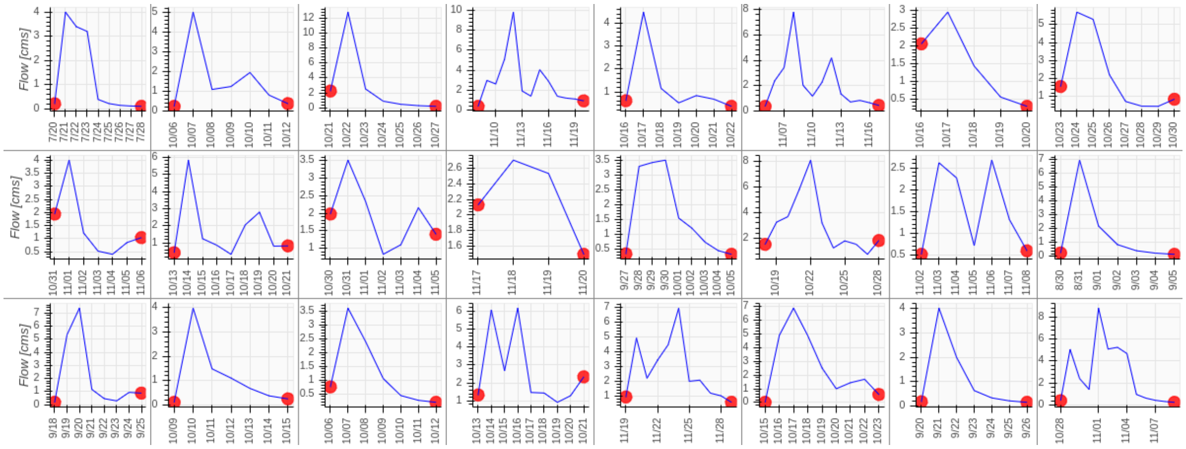

Figure 6: Example AD Results (Event Hydrographs) for WSC 08HB048: Carnation Creek at the Mouth

The PCA-based AD method takes a daily average runoff time series, and builds a matrix of some number of lag periods proportional to the size of the catchment. Using up to 15 lag periods, or 15 days, is expected to be suitable for the time of concentration and hydrograph response of basins up to 1000 km^2. Principal Component Analysis (PCA) is then applied to reduce the number of lag series to those components describing a minimum of 90% of the variance in the data. Figure 7 shows that most of the time just 2 components are required to meet the 90% variance target, however the expected AD performance is a maximum for between 4 and 5 components, and the confidence interval highlights the large amount of variance in the data for between 2 and 6 PCA components. Much of this variance is expected to result from the runoff signals themselves, as for example few runoff events will be identified from June to October (inclusive) in the semi-arid climate of the BC interior.

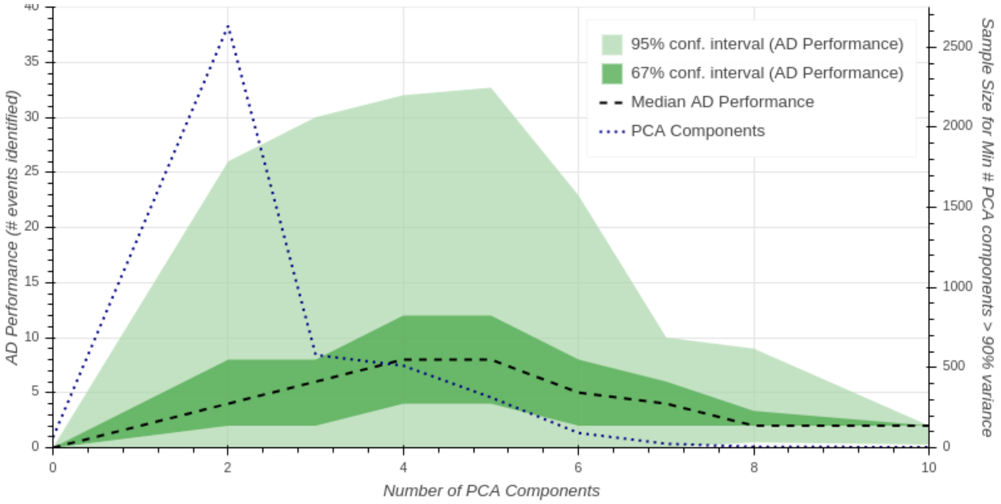

Figure 7: PCA Components and AD Performance

4.1.3 Sensitivity to Training Period

The AD algorithm is supervised in the sense that a training period is provided as an input. Initial testing of the AD algorithm demonstrated a high level of sensitivity to the training period selected. Combining the random selection of a single year (2007-2018) for input training with a random selection of 1-12 months (inclusive) yields a total search space of roughly 50 thousand alternatives. The execution time of the AD algorithm is such that evaluating the entire search space is intractable for practical purposes. Better efficiency in code may be possible, however the main function of the AD algorithm uses the well-optimized Tensorflow Python library (Abadi et al. 2015).

Monte Carlo (MC) simulation is used to illustrate the variability in AD performance as a function of training period selection. Figure 8 shows the variability in number of events identified by the AD algorithm across all WSC stations based on random selection. Note that the results of each station are each comprised of 50 trials of randomly selected training parameters, and the MC simulation represents 1000 random selections from the aggregate results. Since the training inputs are not continuous variables, a gradient-based search method cannot be applied directly, however the long tail of the distribution highlights the opportunity for an improved search method, which remains to be addressed in future work.

4.1.4 Radar Image Preprocessing

With a set of runoff events identified by the AD algorithm for each WSC station, start and end times of the runoff events are are used to form a series of queries to retrieve concurrent radar images from the nearest radar station. Once the radar images are retrieved, a matrix is constructed with the same shape as the image (in pixels), and populated with values corresponding to the azimuthal equidistant projection of the radar image. The known coordinates of the radar station correspond to the centre

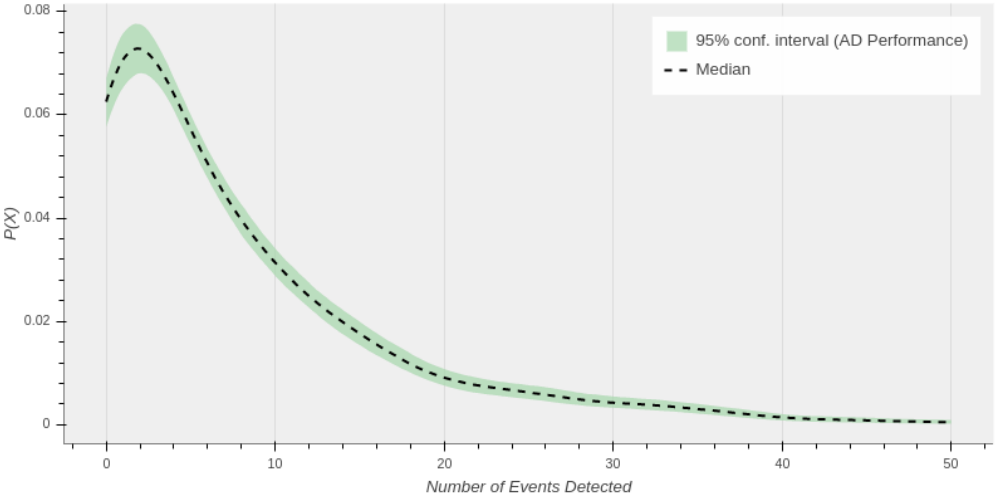

Figure 8: MC simulation: 1000 random selections of training period (KDE probability density function fit)

pixel, which is used to reproject the entire matrix to a projection consistent with that of the catchment basin geometry. Each basin geometry is retrieved and used to create a boolean 'mask' such that the radar image pixels representing each basin can be captured in a batch process. The projection error associated with conversion from the radar projection to the basin geometry projection is well within the image resolution of $1km^2$ per pixel, so the error is neglected. An example of a radar image mask representing a basin is shown in Figure 9.

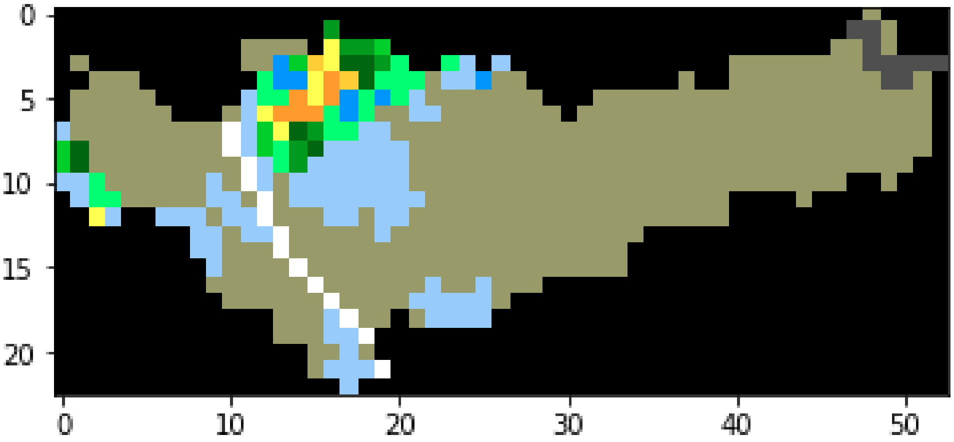

Figure 9: Sample Basin Captured from Radar Image (WSC 05BL014)

As shown in Figure 1, rainfall intensity is mapped to an array of 14 colours unique to representing precipitation. The final step in the radar image processing is to map the colour values in the masked radar images to their corresponding precipitation intensity. The result is a series of matrices representing the volume of water that fell on each pixel or cell comprising the basin at each time step. Reducing the time dimension by summing and normalizing yields the spatial distribution of precipitation across all events captured. The results are discussed in the subsequent section.

5 Results and Discussion

-having all the data would eliminate the costly AD step, negate the issues with false negatives, and give a more complete picture of total precip.

5.1 Spatial Distribution of Precipitation

Summing the precipitation time series and applying a colour map to the normalized output volume yields a representation of the spatial distribution of precipitation for the sample events captured. Figure 9 shows a grid plot of a subsample of 100 WSC basins scaled to similar size to emphasize the differences in colour patterns, where yellow represents less precipitation and blue represents more precipitation.

Vestiges of the information layer are apparent in some of the basins in Figure 10 where parts of place names and the concentric rings from the image block the radar information. In other basins, non-uniformities are evident (08MG001, 08HA003) suggesting orographic effects. Spotted patterns (08NL071) suggest precipitation falling in convective cells in one or few events in the summer season, though this could also be attributable to noise or other interference in the radar data.

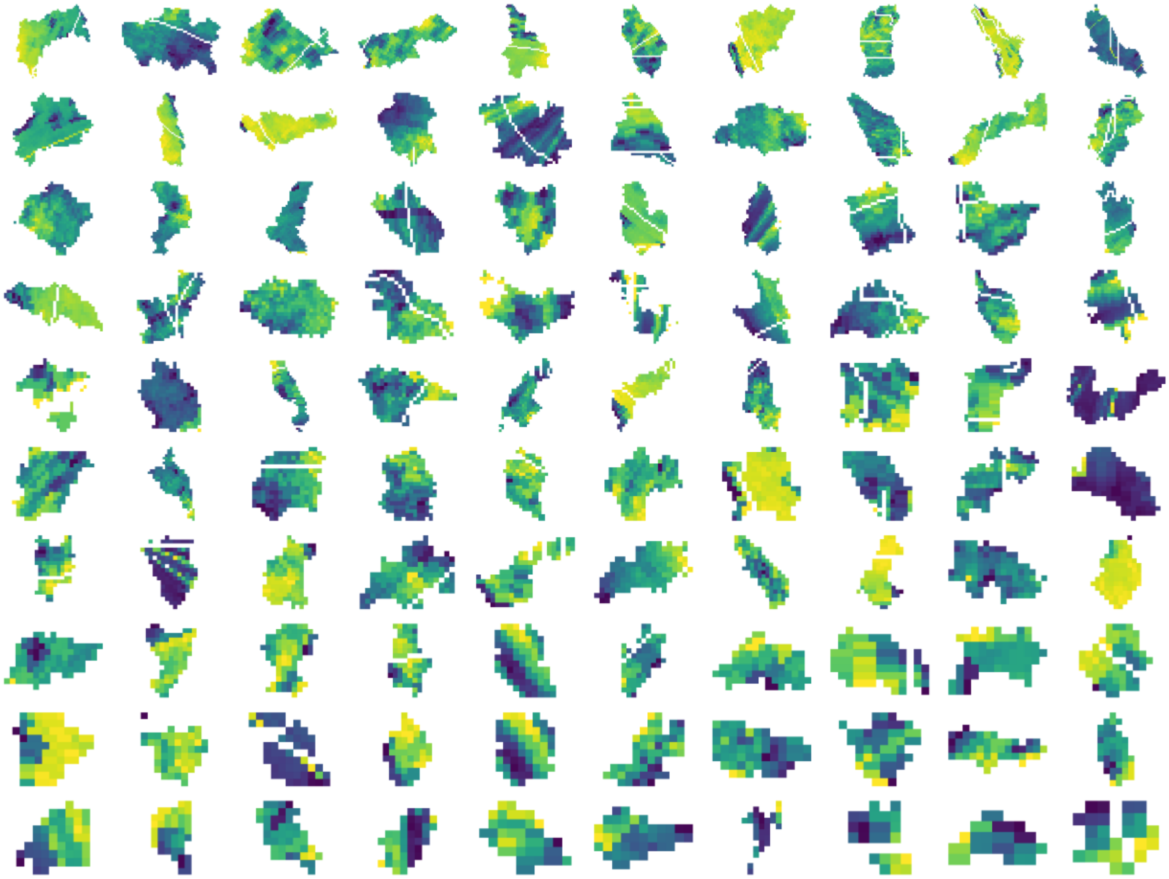

Figure 10: Qualitative Representation of Spatial Distributions of Precipitation in Basins of Western Canada

5.2 Hydrograph Reconstruction

5.3 Clustering Spatial Precipitation Distribution

Often basins of interest to a researcher are ungauged. In the case of estimating water resources for ungauged basins, information from gauged basins is projected to the location of interest based on similarities in physiography. (Obedkoff, Sustainable Resource Management, and Branch 2003) divides the hydrology of British Columbia into many subregions on the basis of there existing some level of homogeneity in runoff characteristics at a local level. The hydrologic zones in BC are typically aligned with the coast as the source of air moisture is predominantly the Pacific Ocean. Measured runoff statistics such as long-term averages and extremes then form the basis for the region for the purpose of estimating runoff at ungauged locations.

6 Conclusions

6.1 Future Work

References

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, et al. 2015. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems." http://tensorflow.org/.

Bokeh Development Team. 2020. *Bokeh: Python Library for Interactive Visualization*. https://bokeh.pydata.org/en/latest/.

Obedkoff, W., British Columbia. Ministry of Sustainable Resource Management, and British Columbia. Aquatic Information Branch. 2003. *Streamflow in the Lower Mainland and Vancouver Island*. Government of British Columbia. https://books.google.ca/books?id=II7iAAAACAAJ.

Thorndahl, Søren, Thomas Einfalt, Patrick Willems, Jesper Ellerbæk Nielsen, Marie-claire ten Veldhuis, Karsten Arnbjerg-Nielsen, Michael R. Rasmussen, and Peter Molnar. 2017. "Weather Radar Rainfall Data in Urban Hydrology." *Hydrology and Earth System Sciences* 21 (3): 1359–80.