Streamflow Modeling Using Weather Data

Tyler Wilson and Gage Paskiewich

Background; Why Model Streamflow?

- Many communities are heavily impacted by streamflows
    - Drinking water
    - Flood risk
- Changing climate is impacting streamflow patterns
    - Particularly in glacial watersheds
- Streamflow forecasts can be useful in community planning and mitigating consequences of severe weather

Research Question

- How well can we predict daily BC streamflow data based on remote sensing climatic data?
- Can this be used when data from stream gauge stations is unavailable?
- How do our predictions compare across different streamflow clusters?

Experiment:

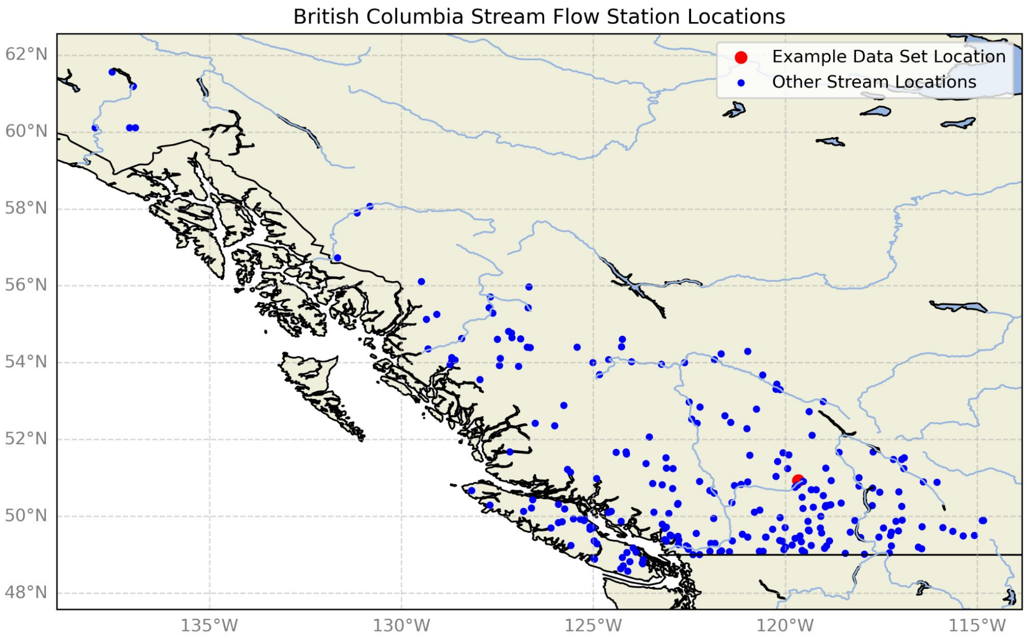
- Using historical weather data, predict the next day’s streamflow values.

Data Sources

- Daily streamflow values (HYDAT; Water Survey of Canada HYDAT data)
- Remote sensing data (ECMWF; European Centre for Medium-Range Weather Forecasts) via ERA5 reanalysis
  Includes:
  - Precipitation
  - Snowfall
  - Temperature

Streamflow Data

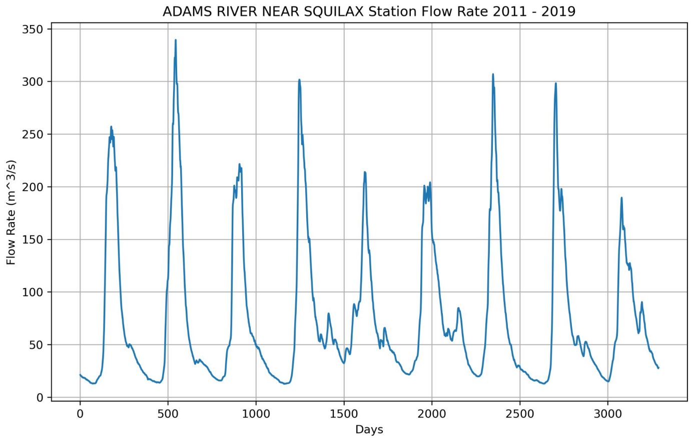
- British Columbia wide database for stream flow stations
    - 241 streams used in investigation
- Flow rates recorded every 5 minutes
- Trimmed data to focus on 2011-2019



Streamflow Data

- Downsampled to get daily values
- Streams with a gap of more than 30 days were excluded

Pre-Processed Example:



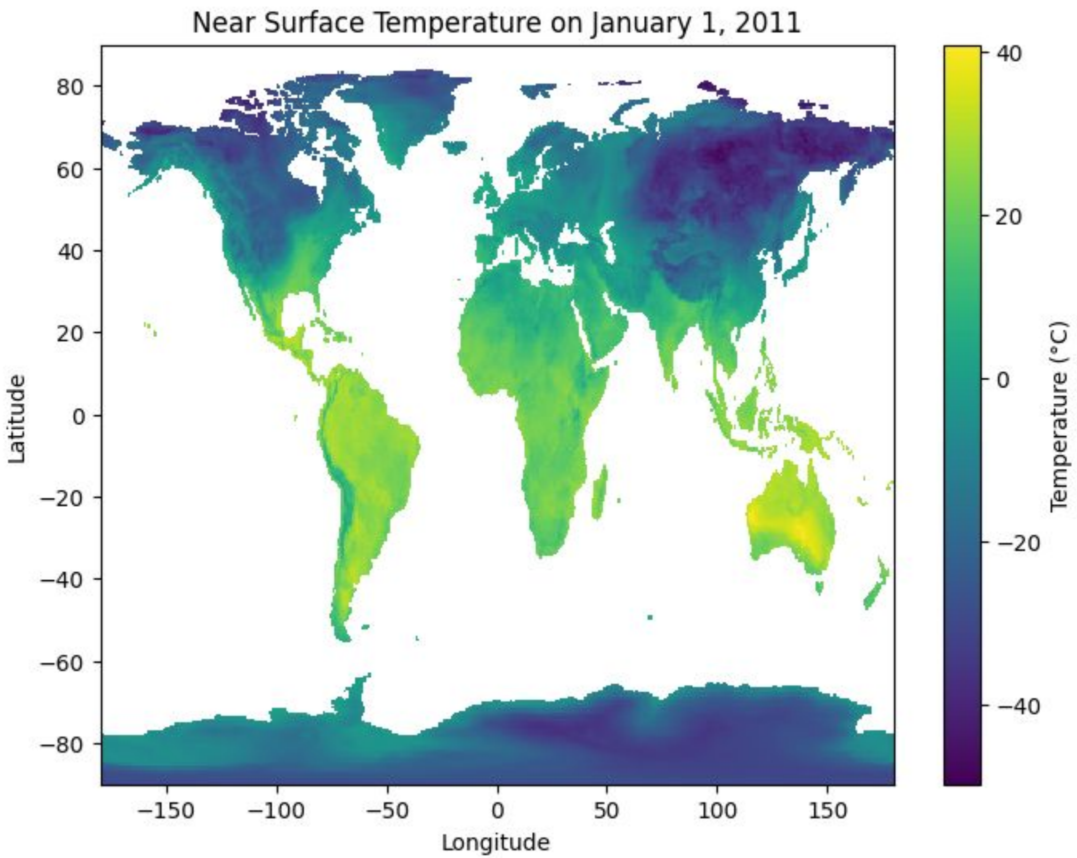
Raw Data Example:

# Discharge.Working@08DB014.20110101.csv				
#				
# Time-series identifier: Discharge.Working@08DB014				
# Location: ADAMS RIVER NEAR SQUILAX				
# UTC offset: (UTC-08:00)				
# Value units: m^3/s				
# Value parameter: Discharge				
# Interpolation type: Instantaneous Values				
# Time-series type: Derived				
#				
# Export options: Corrected signal from 2011-01-01T00:00:00Z to End of Record				
#				
# CSV data starts at line 15.				
#				
ISO 8601 UTC	Timestamp (UTC-08:00)	Value	Approval L	Grade
2011-01-01T08:00:00Z	2011-01-01 0:00	21.293	Approved	-1
2011-01-01T08:00:15Z	2011-01-01 0:00	21.293	Approved	-1
2011-01-01T08:05:15Z	2011-01-01 0:05	21.589	Approved	-1
2011-01-01T08:10:15Z	2011-01-01 0:10	21.86	Approved	-1
2011-01-01T08:15:15Z	2011-01-01 0:15	21.404	Approved	-1
2011-01-01T08:20:15Z	2011-01-01 0:20	20.96	Approved	-1
2011-01-01T08:25:15Z	2011-01-01 0:25	21.145	Approved	-1
2011-01-01T08:30:15Z	2011-01-01 0:30	21.293	Approved	-1
2011-01-01T08:35:15Z	2011-01-01 0:35	21.145	Approved	-1
2011-01-01T08:40:15Z	2011-01-01 0:40	21.33	Approved	-1

Source: Water Survey of Canada HYDAT data

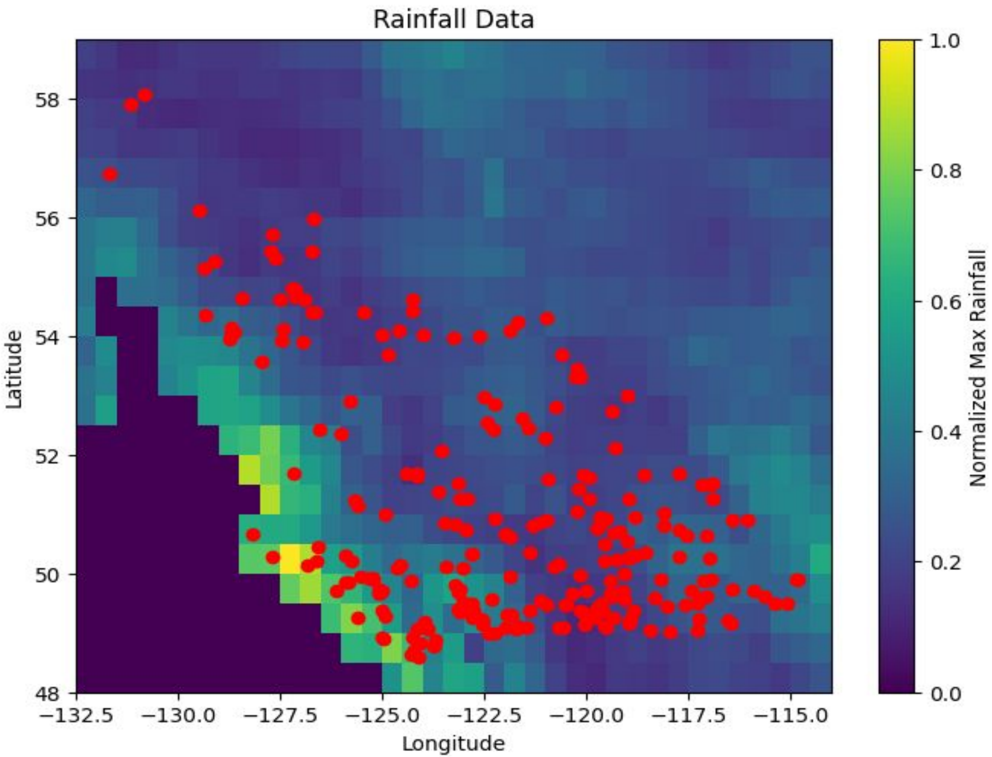
Climate Data

- Global raster with 0.75 degree resolution
- Data recorded hourly from 2011 to 2019
- Contains many different parameters
  - Rainfall
  - Snowfall
  - Temperature
  - Humidity
  - Etc.



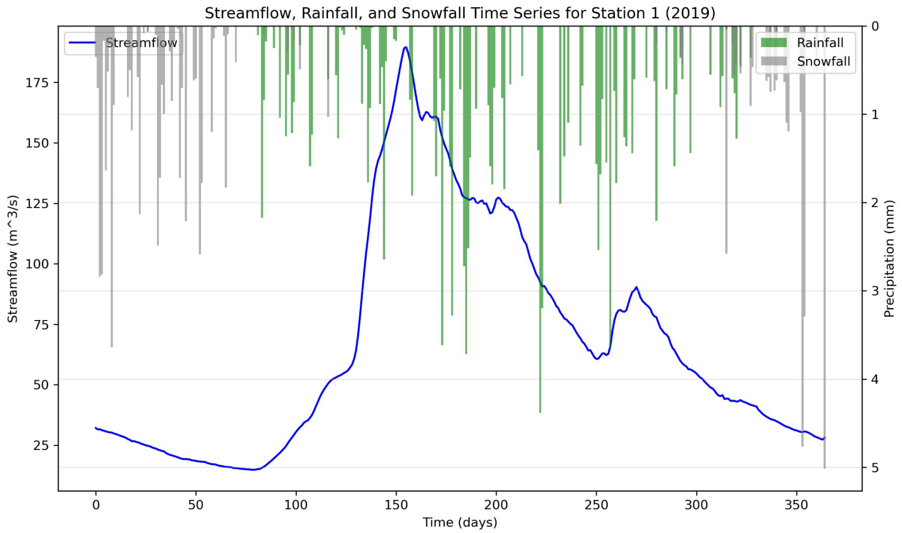
Climate Data

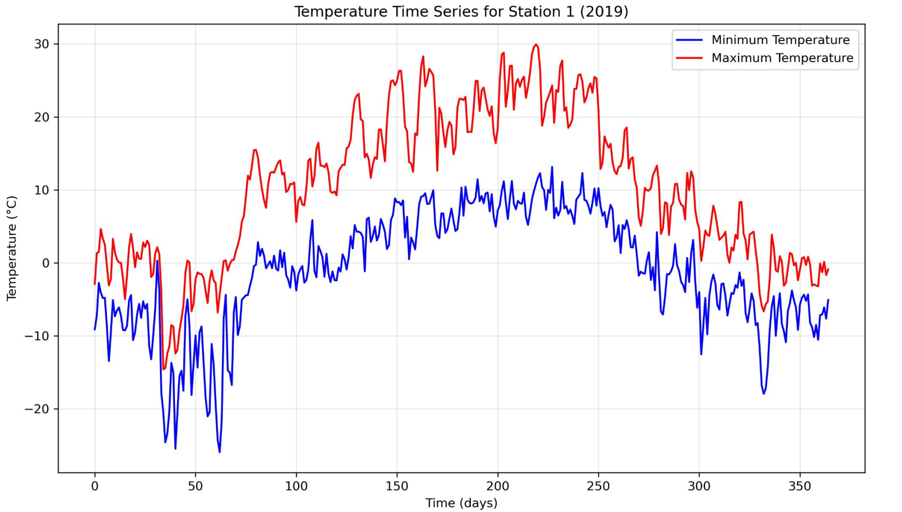
- Cropped to only lat/lon points surrounding gauge stations
- Downsampled to get daily values
- Interpolated NAN values over the ocean
- Selected features
    - Rainfall flux
    - Snowfall flux
    - Temperature



Combining Data

- These climatic datasets combined are clear to influence streamflow
- See example to the right:
  - Low flow in the winter with little precipitation the stream
  - Spring melt and some rain on snow event creates peak flow
  - Fall has a smaller peak from late season rainfall



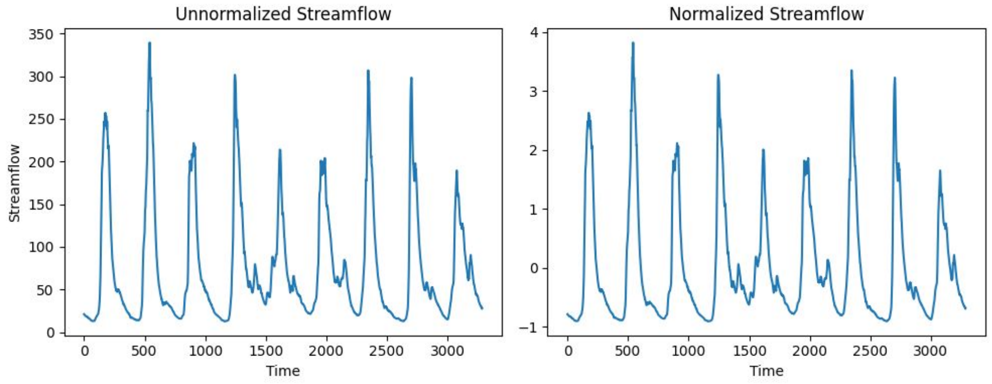


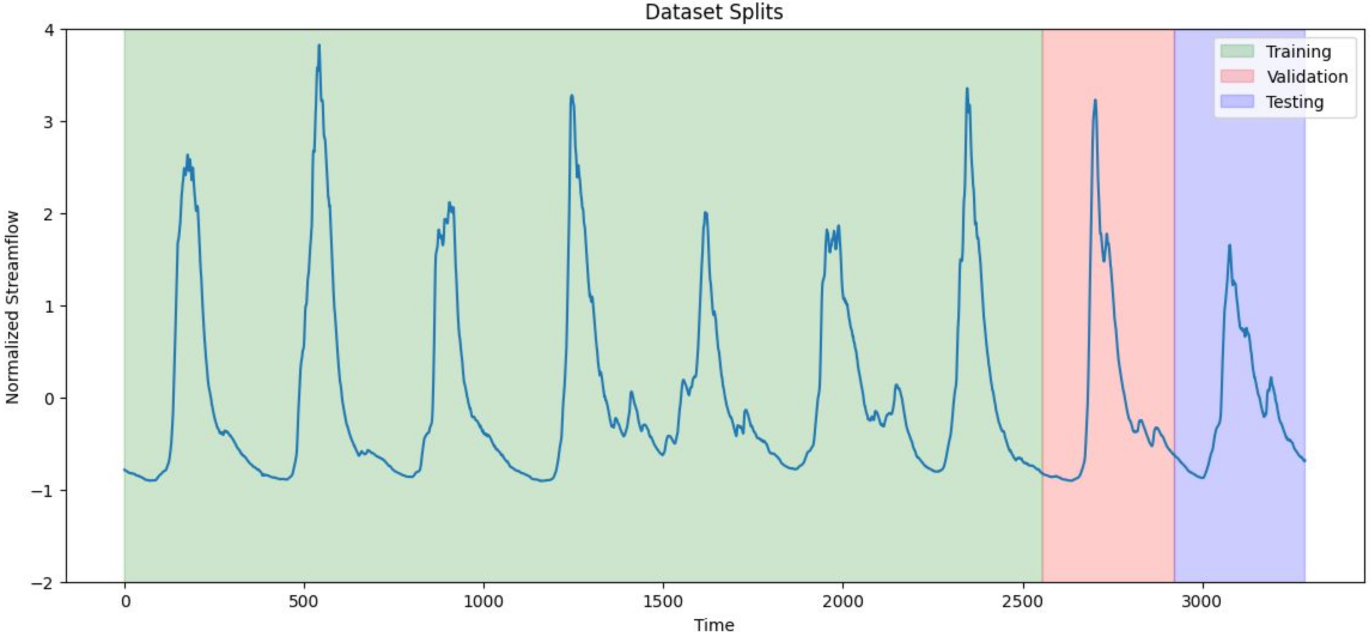
Experiment

1. Train two different models
   a. Neural network
   b. Deep learning network
2. Evaluate and compare models
3. Examine any spatial trends in model performance

Dataset Preparation

1. Normalization
2. Define data splits





Model Evaluation

- Utilize Nash–Sutcliffe model efficiency coefficient (NSE) for model quality
  - Value of 1 is well fit, 0 is as predictive as the mean of the time-series

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T}\left(Q_o^t - Q_m^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t - \overline{Q}_o\right)^2}$$

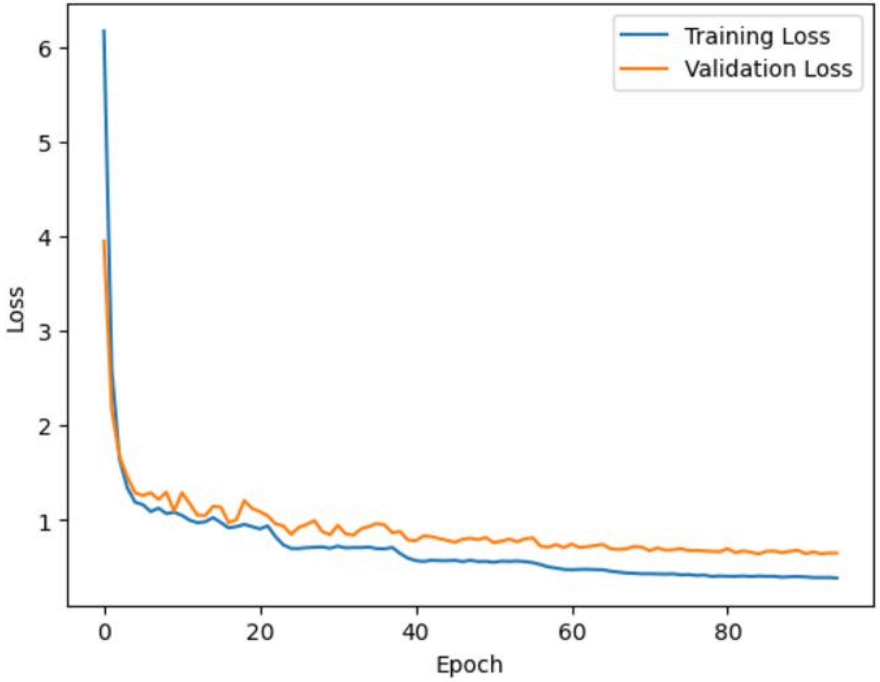
$\overline{Q}_o$ ~ Mean flow rate

Q_m^t ~ Predicted flow rate at time t

Q_o^t ~ Observed flow rate at time t

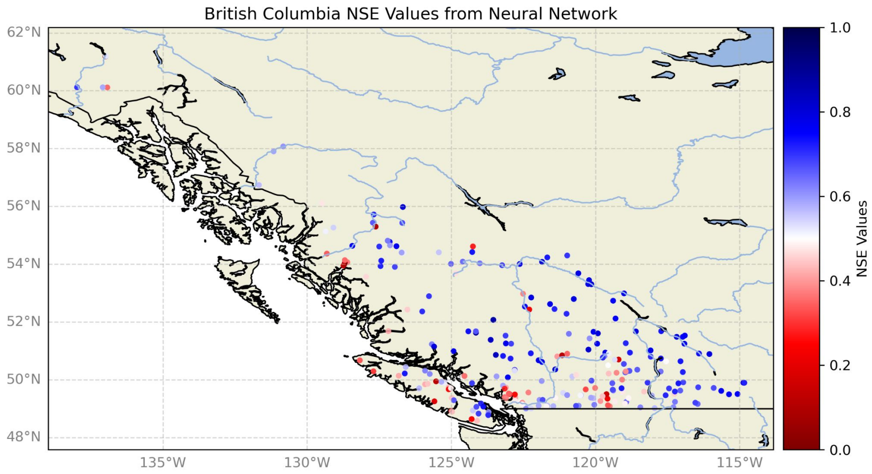
Methodology - Neural Network

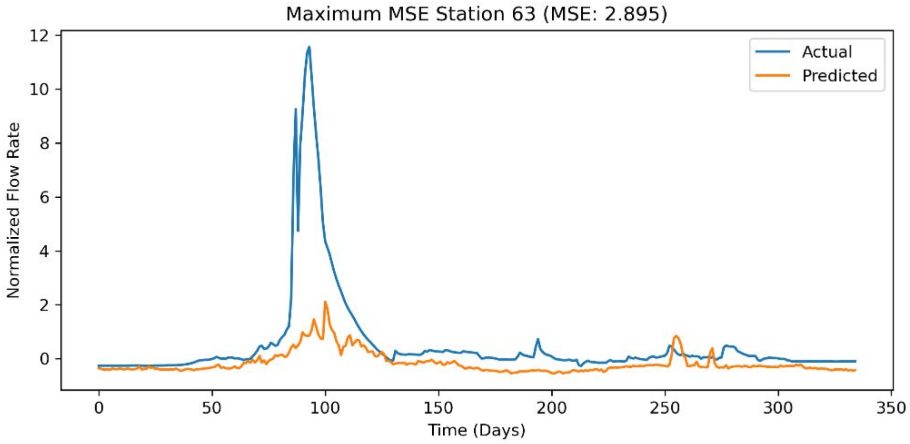
1. Time window
2. Neural network definition
3. Overfitting reduction
4. Training
5. Evaluation

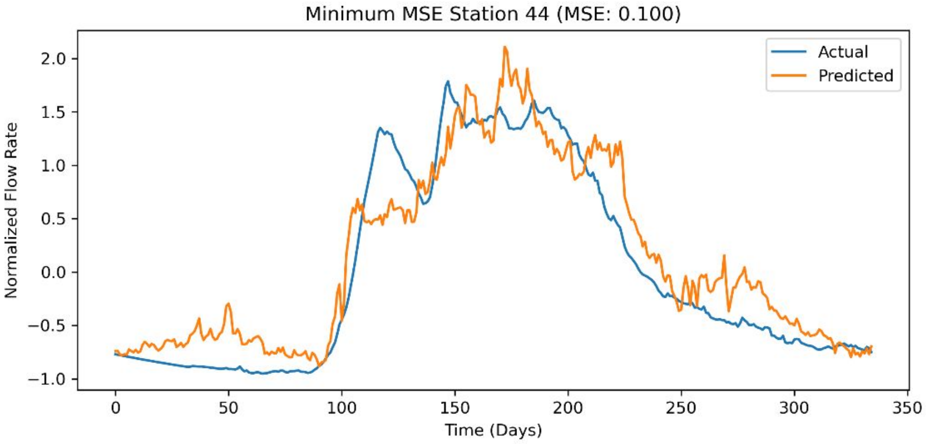


Results - Neural Network

- More accurate on low sensitivity streams
- Mean NSE: 0.5506
- NSE values lower in lower mainland







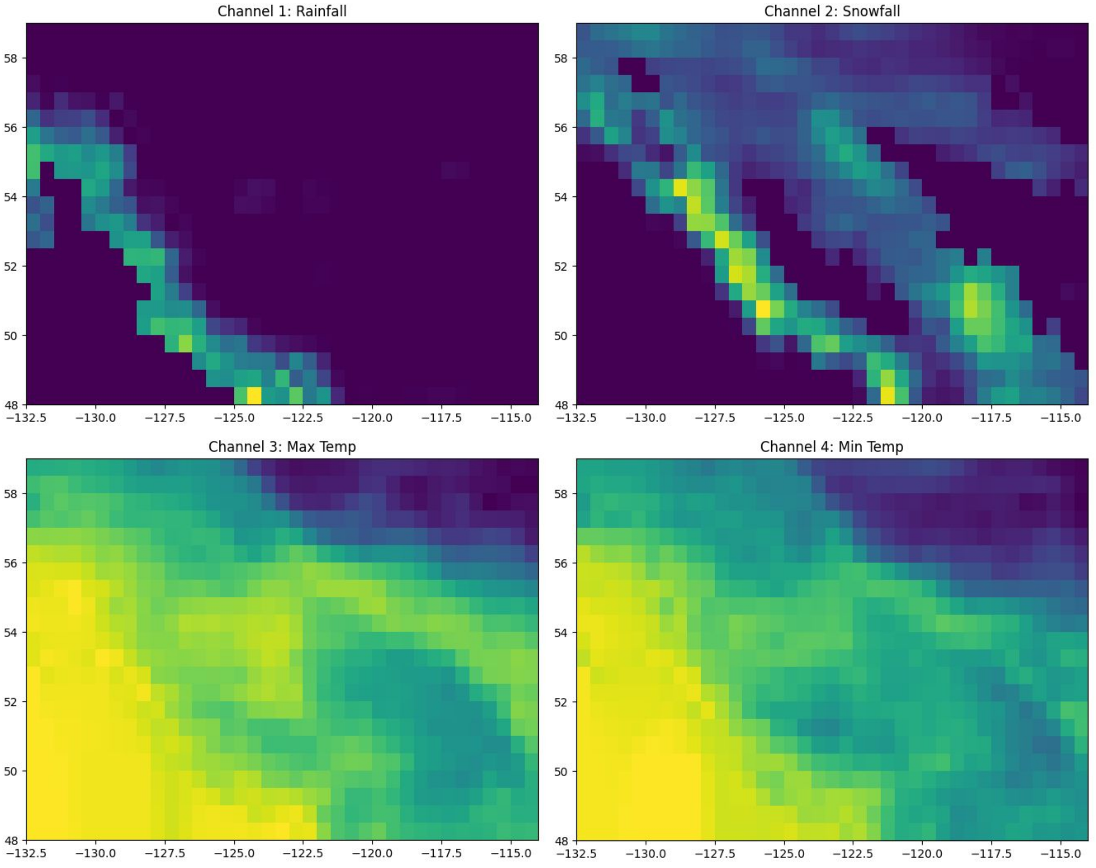
Methodology - Deep Learning Model

1. Input Data
2. Model Type
3. Architecture
4. Model tuning
5. Evaluation

Deep Learning - Model Input

- Raster with 4 channels
  - 22x37 size
- Repeated in time Nt times
  - Averaged by day, week, or month
- Final input shape:

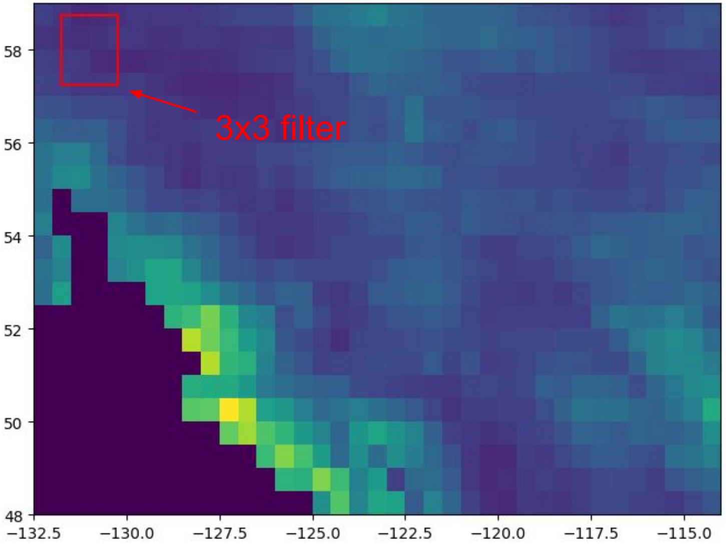
$$N_t \times 22 \times 37 \times 4$$



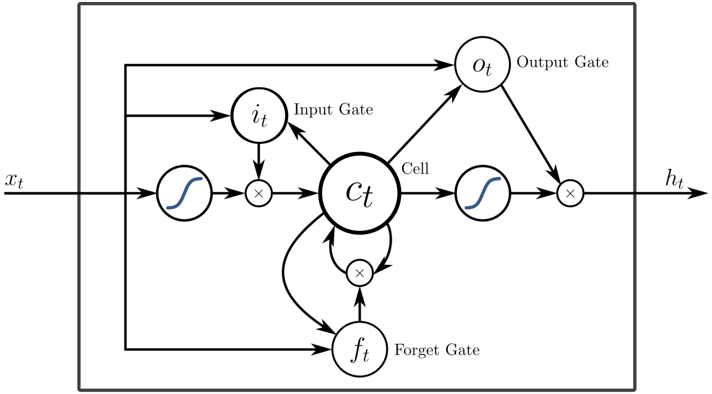
Deep Learning - Model Type

- Convolutional Neural Network (CNN)
  - Captures spatial trends in the raster
- Long Short Term Memory (LSTM)
  - Handles temporal trends in the data
- Dense Neural Network
  - Used for the final prediction step

Convolution input

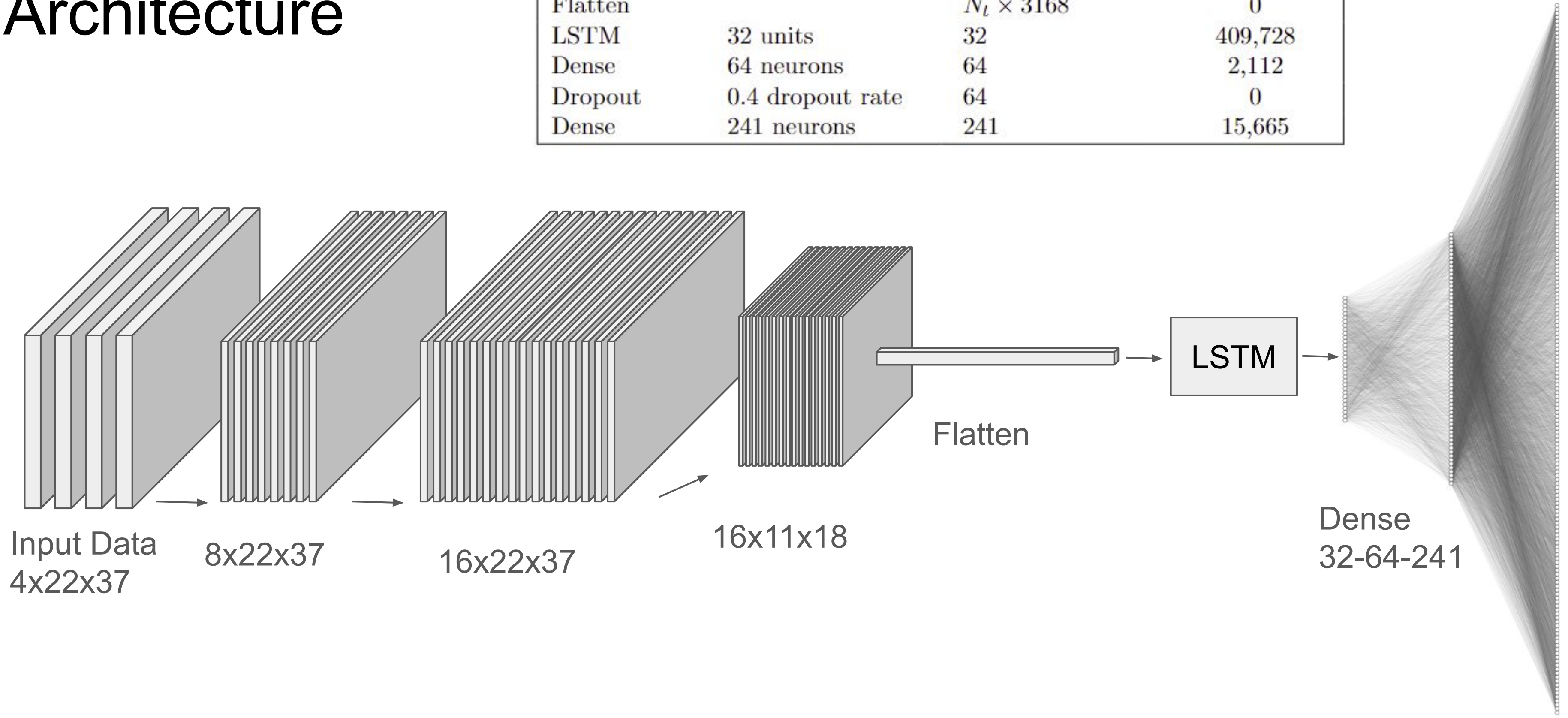


LSTM cell



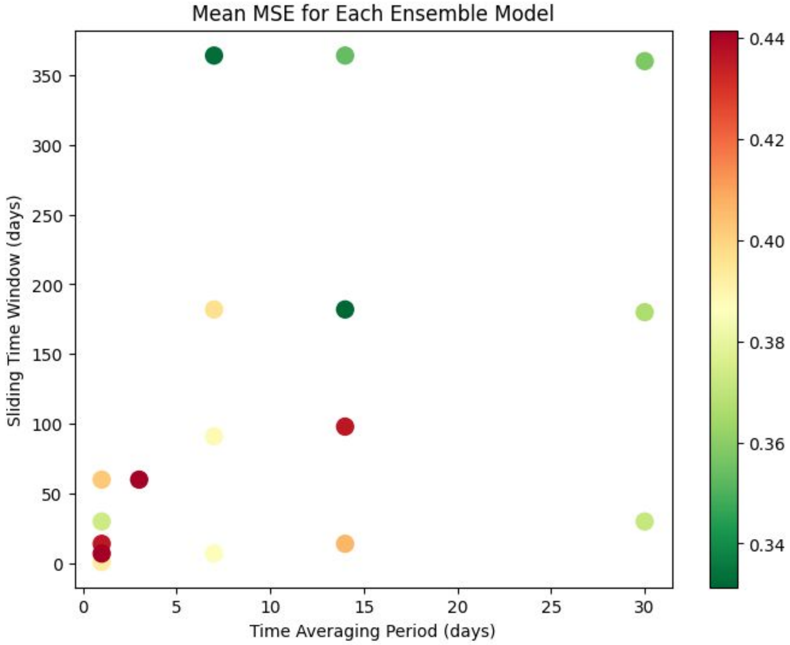
CNN-LSTM Architecture

Layer Type	Description	Output Shape	# Parameters
Input	Weather data	$N_\iota \times 22 \times 37 \times 4$	0
Convolutional	8 filters, 3×3 size	$N_\iota \times 22 \times 37 \times 8$	296
Convolutional	16 filters, 3×3 size	$N_\iota \times 22 \times 37 \times 16$	1,168
Max Pool	2×2 pool size	$N_\iota \times 11 \times 18 \times 16$	0
Flatten		$N_\iota \times 3168$	0
LSTM	32 units	32	409,728
Dense	64 neurons	64	2,112
Dropout	0.4 dropout rate	64	0
Dense	241 neurons	241	15,665



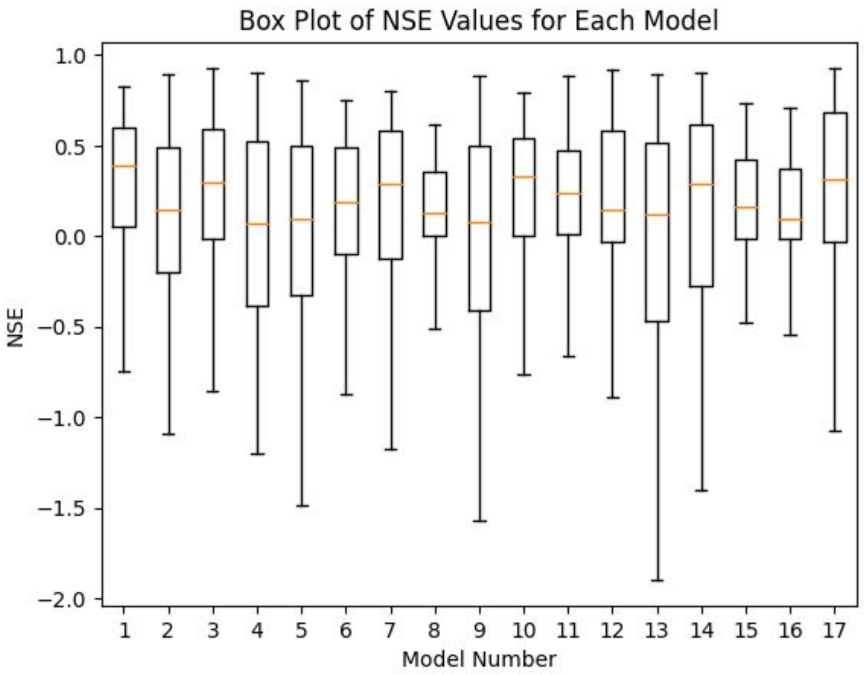
Deep Learning - Model Tuning

- Dataset loaded/trained in small batches
  - Limits RAM usage
  - Improves model generalization
- Early stopping based on validation data
- Self-adjusting learning rate
- Created an ensemble of different length sliding windows and time averaging periods
  - Done to capture different temporal patterns at different time scales



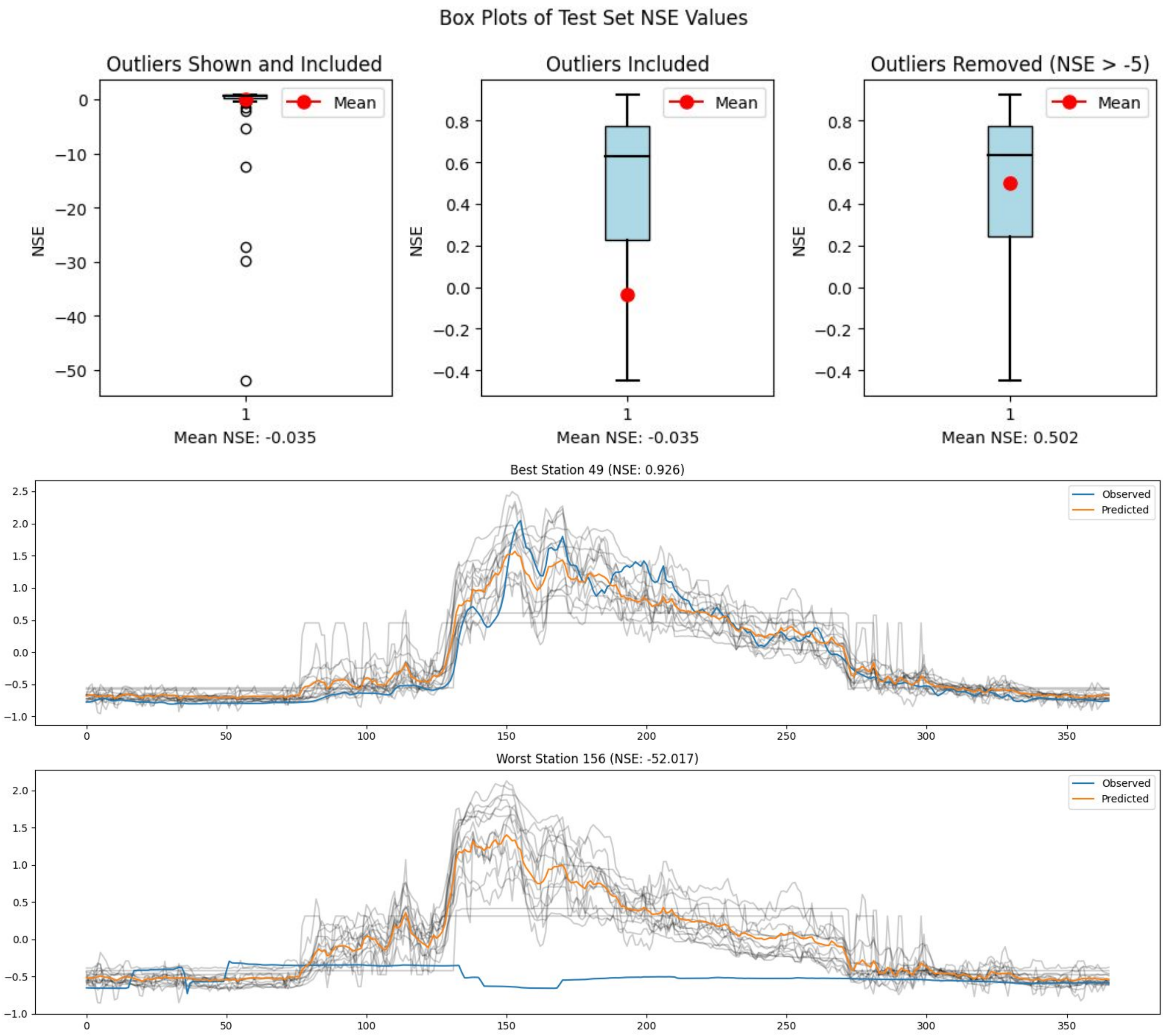
Deep Learning - Ensemble Creation

- Computed best NSE on validation set
- Lowest two Q1 scores dropped



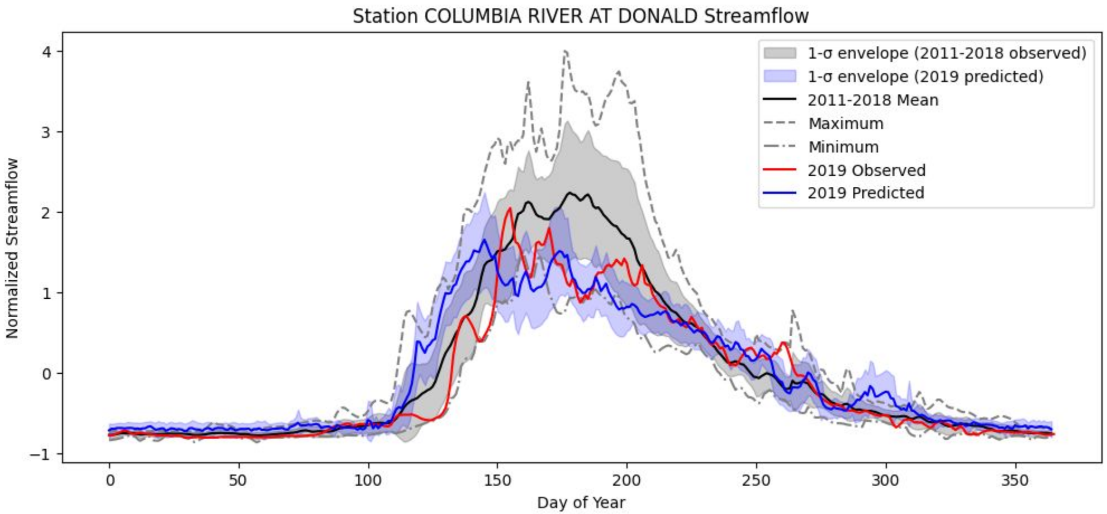
Test Set Results

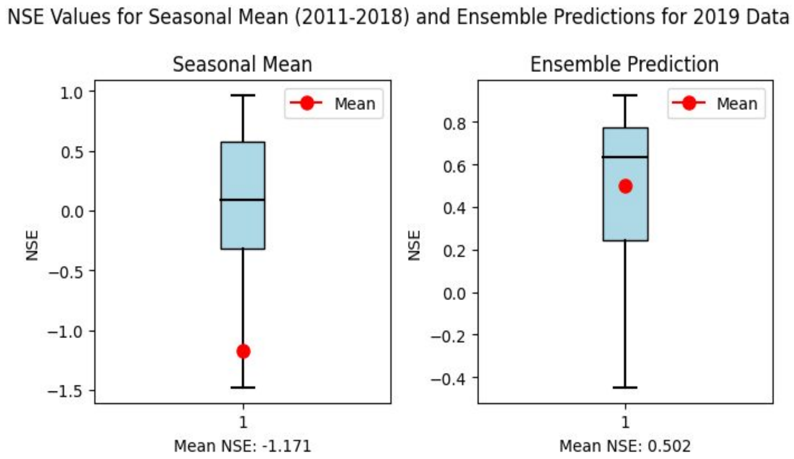
- Somewhat polarizing results
- Mean NSE dominated by a few outliers
  - Could be caused by poor quality gauge data

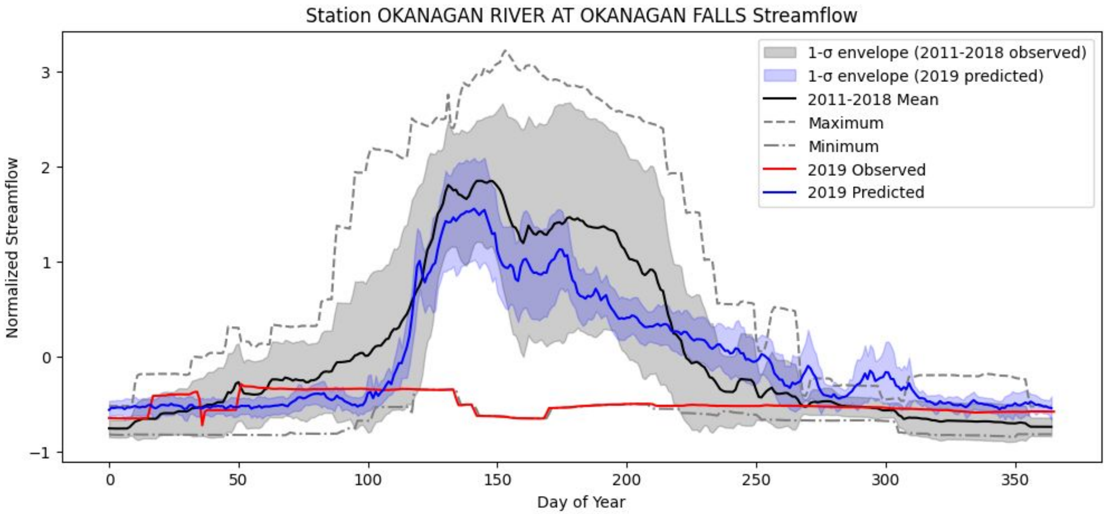


Test Set Results

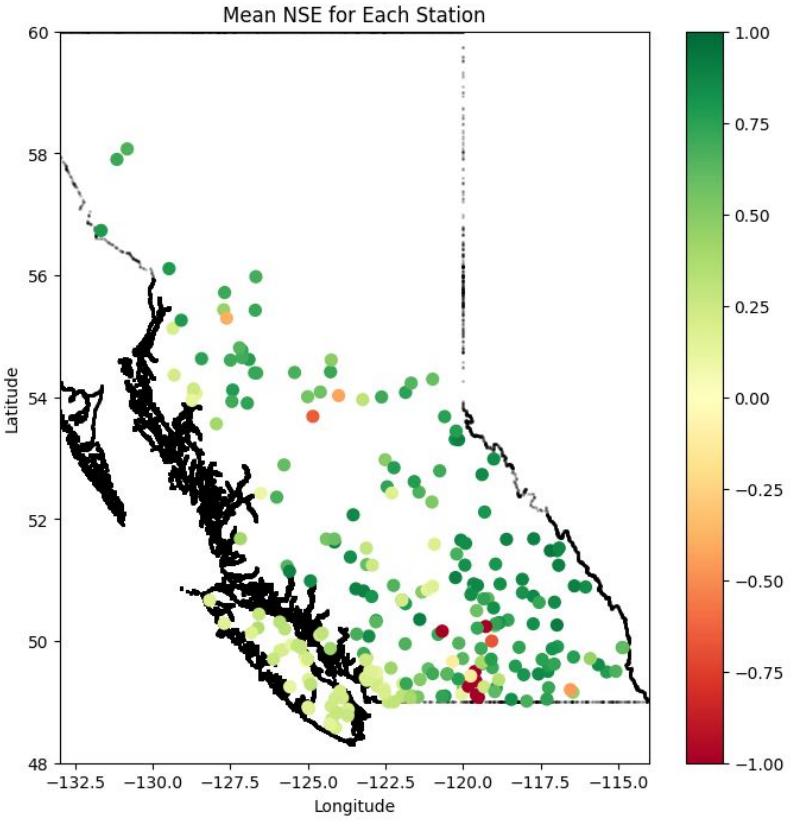
- Outlier can be attributed to abnormal yearly pattern
- Model outperforms baseline of seasonal mean

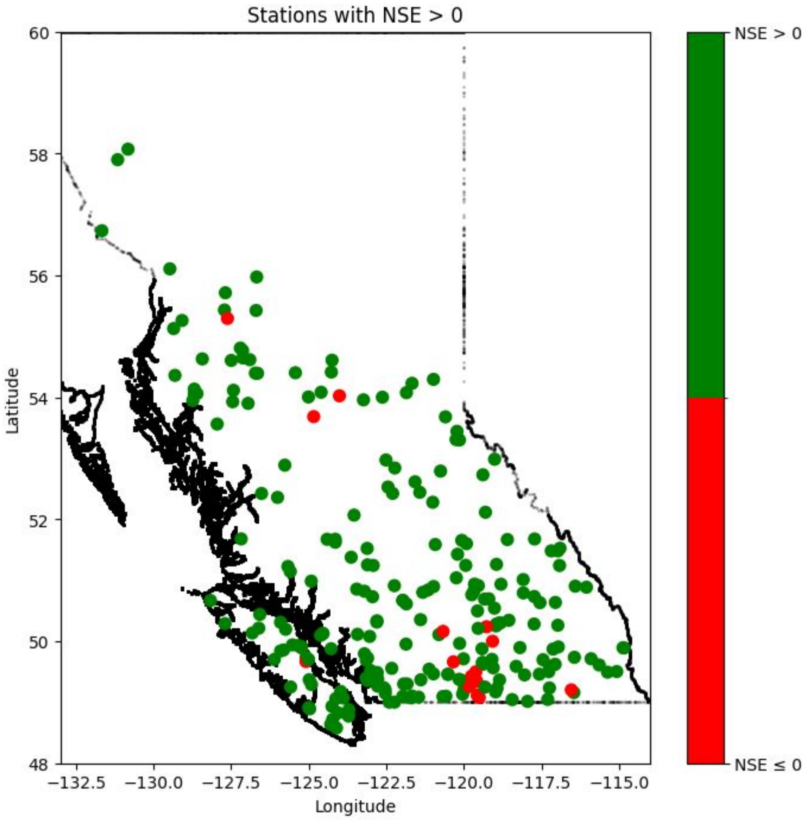






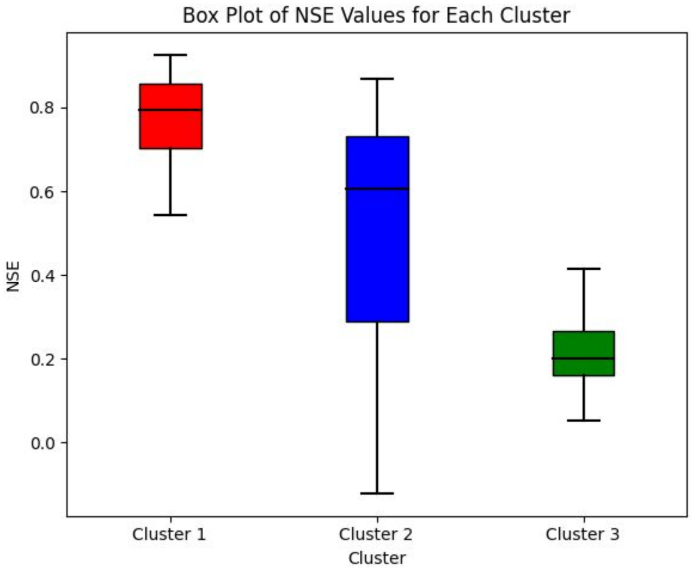
Test Set Results - Spatial Trends

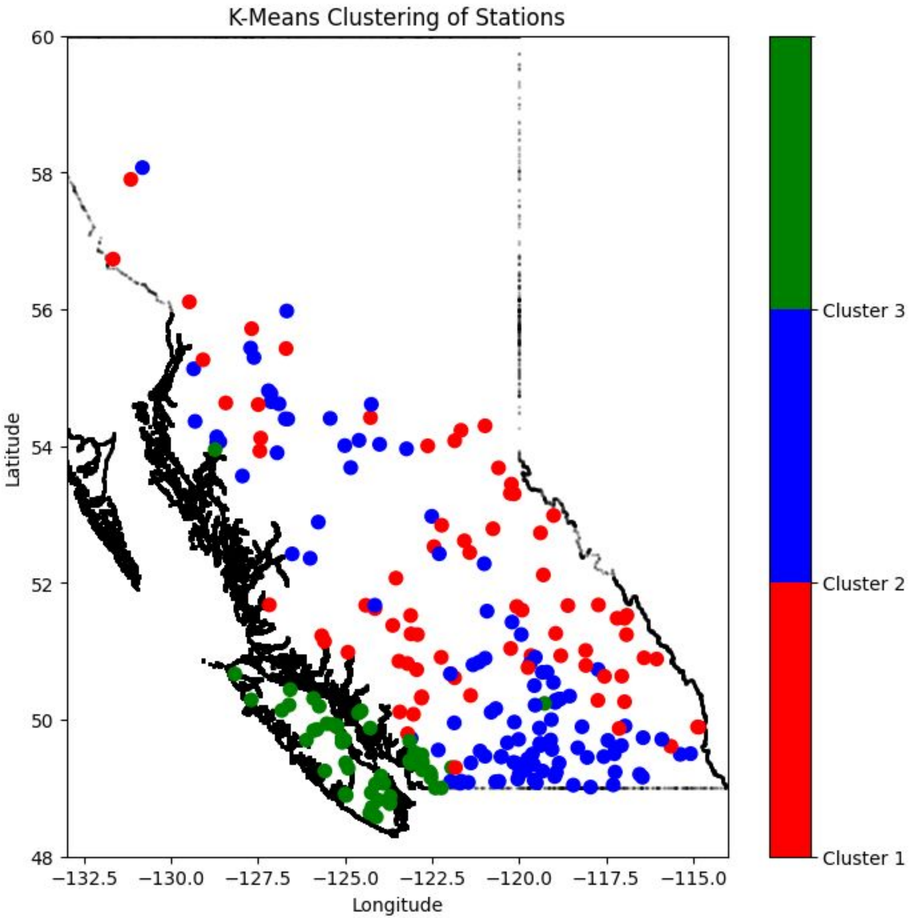




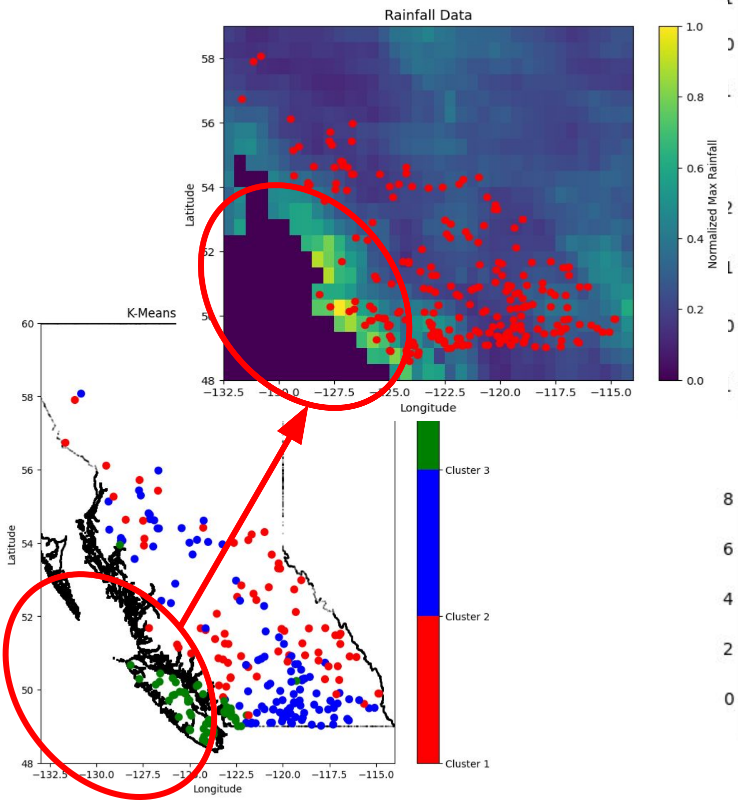
Clustering of Stations

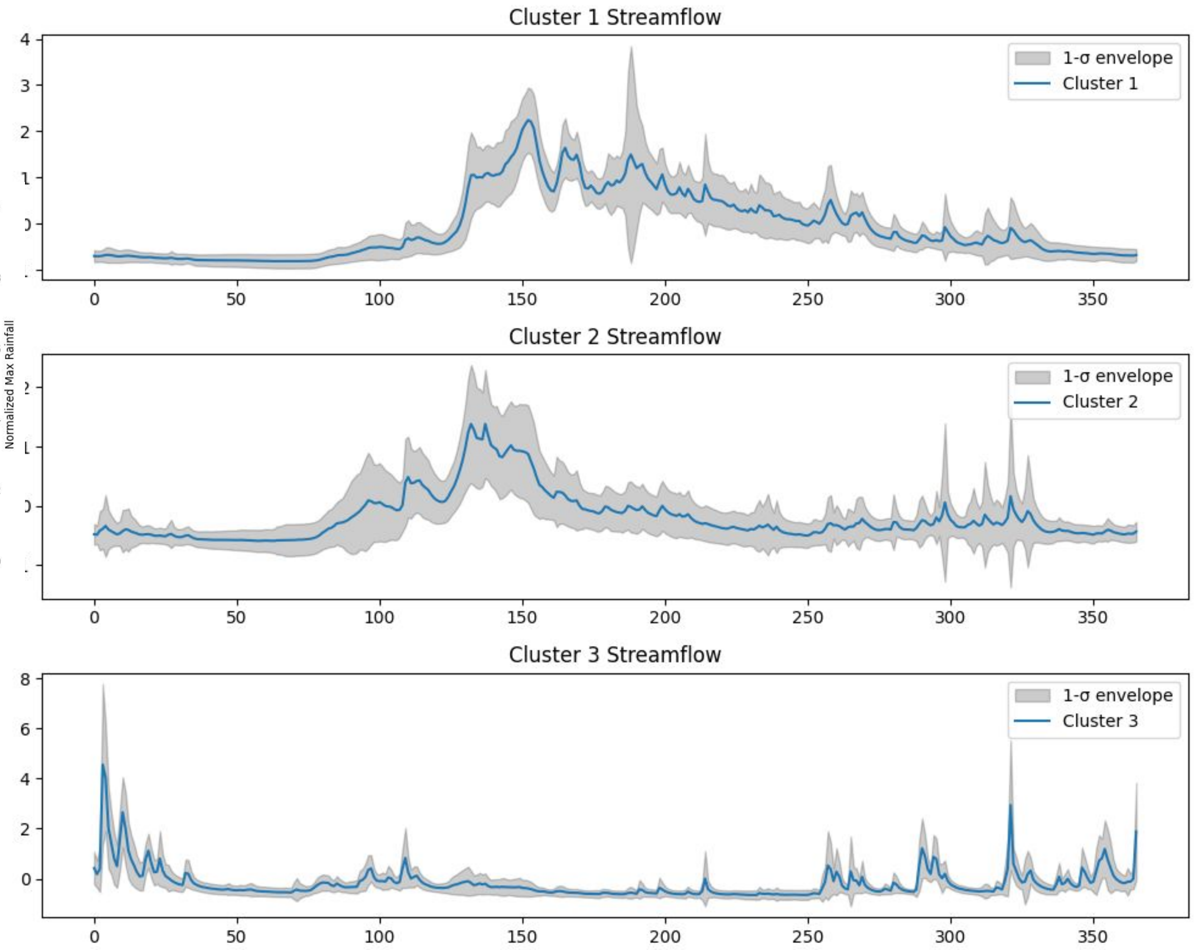
- Clustered using K-means and streamwater test data





Cluster Results





Limitations - Modelling

- Model complexity was limited by available compute resources
  - Model was optimized as much for efficiency as it was performance
- Many alternative architectures that can be explored
  - Hard to find the most ideal architecture for the problem at hand
- CNN had trouble with interpolated raster data
- Model struggled with detecting large changes in streamflow
  - This would not be ideal for preliminary detection of potential flood risk

Limitations - General

- Some streams are not properly mapped to locations
  - This resulted in the data not being used in the study
- Large data gaps resulted in streams not being included
- Relying on the network to learn spatial relationships
  - Locations of the gauge stations were not used

Conclusion

- Both models performed similarly
- The models could predict daily BC streamflow well based on remote sensing climatic data
- These models are able to be used when stream gauge stations are unavailable, for non-peak flow seasons
- Clustering identified three main clusters
  - Two of which featured abnormally high peak flow in the spring and early summer respectively, spatially corresponding to snow melt areas in the interior, producing higher NSE values
  - And one of which having sporadic peak flows in the winter, spatially corresponding to areas with rainy winters on Vancouver Island and the Lower Mainland, producing lower NSE values

Future Work

- Room for improvement with deep learning architecture
- Relate findings to glaciation of watersheds
- Apply climatic data corresponding to the entire watershed above stream gauge

References

Papers:

Anderson, S. and Radic, V. (2022): https://hess.copernicus.org/articles/26/795/2022/

Anderson, S. and Radic, V. (2022): https://www.frontiersin.org/journals/water/articles/10.3389/frwa.2022.934709/full#B5

Nash, J. E.; Sutcliffe, J. V. (1970): https://doi.org/10.1016%2F0022-1694%2870%2990255-6

Data:

ECMWF (2024): https://cds.climate.copernicus.eu/datasets/derived-near-surface-meteorological-variables?tab=overview

HYDAT (2024): https://collaboration.cmc.ec.gc.ca/cmc/hydrometrics/www/

Natural Resources Canada (2024):
https://natural-resources.canada.ca/science-and-data/science-and-research/geomatics/topographic-tools-and-data/watershed-boundaries/20973

Stats Canada (2021):
https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21