A Web Platform for Dynamical Streamflow Prediction using Machine Learning and Deep Learning Methods

CPSC 8810: Mining Massive Data

Hydroinformatics Group:

Sadegh Sadeghi Tabas

Nushrat Humaira

Pawan Madanan

Siddish P Rao

Meghan Patil

Github:

https://github.com/sadeghitabas/CPSC8810-Mining-Massive-Data

Outline

- Introduction
  - Problem Specification
  - Motivation and Novelty
- Methodology
  - Case Studies and Data
  - Data Driven Models
    - MLP
    - LSTM
    - Hybrid Model
- Web Platform
- Experimental Results
- Future Forecasting
- Conclusion

Introduction

Importance of Watershed Hydrological Modeling

- Flood Peaks
- Management of Water Resources
- Design of Hydropower Plant
- Planning of Irrigation Schemes
- Extension of Streamflow Records and Imputing Missing Values
- Prediction of Low Flows

Modeling Approaches

- Theory Driven Models (Conceptual Models and Physically-Based Methods)
- Data Driven Models
  - Classic Methods (AR, MA, ARMA, ARIMA and so on)
  - Machine Learning Methods (SVM, MLP, Random Forest and so on)
  - Deep Learning Models (Vanilla RNN, LSTM, GRU)

Motivation and Novelty

Proposing a Web designed Platform which is able to:

- Show Watershed Boundaries in different Scales
- Train and Test different ML and DL approaches for Streamflow Simulation
- Forecast Streamflow for Near Future
- Ability to Simulate Runoff at Global Scale (North America, South America and Africa)

Methodology

Case Studies

- North America
- South America
- Africa

Data Sources:

- GRDC Database
- NCDC Database
- CAMELS Database

➔ (Datasets paired based on watershed boundary and geographical proximity)

Methodology

Why Supervised Learning?

Data Driven Models:

- Rainfall-runoff processes are quite complex making physical models unreliable
- Due to their complex nature, data driven models best suited for this task

➔ Three different ML and DL models used:

- ◆ Multi-Layer Perceptron(MLP)
- ◆ Long Short-Term Memory(LSTM)
- ◆ Hybrid Convolution Neural Network-LSTM (CNN-LSTM)

MLP:

- A class of feed forward ANN with nonlinear activation functions
- Model was trained for 10,00 iterations with 20 hidden layers
- We have selected the Rectified Linear Unit (ReLU) transform function

LSTM:

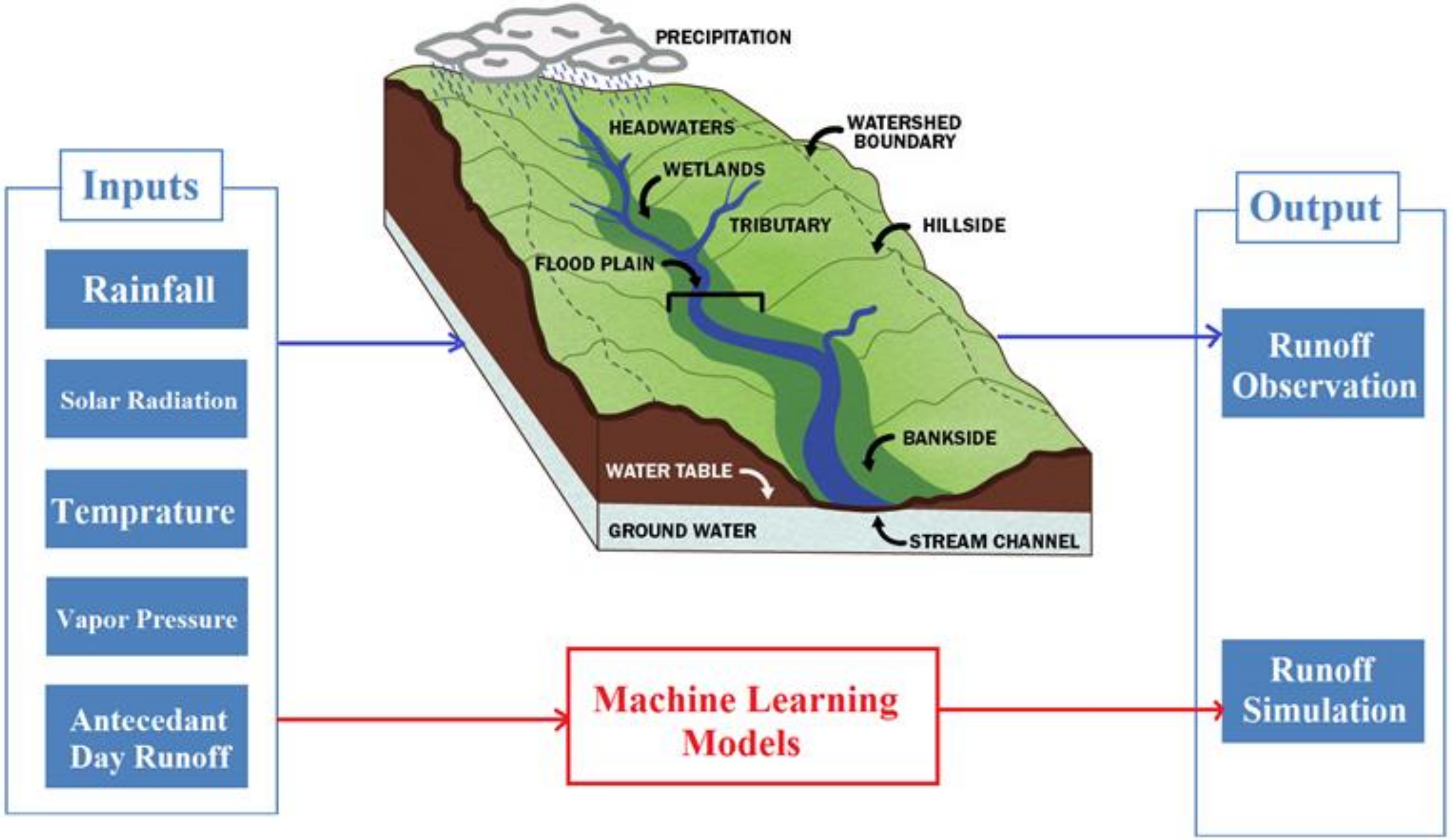
- A memory based method as it is a type of recurrent neural networks
- 30 hidden layers
- Transform function: Rectified Linear Unit (ReLU)

Hybrid Model (LSTM-CNN):

- A hybrid model with CNN (Conv1D) in the lower combined with LSTM in the following layer with a fully connected dense layer for output
- The kernel size was assumed as 2
- Transform function: Rectified Linear Unit (ReLU)
- Mean Squared Error as loss function

Model Structure

$$Q_t^s = f\,(P_t, T_t^{\max}, T_t^{\min}, Srad_t, VP_t, Q_{t-1}^o)$$



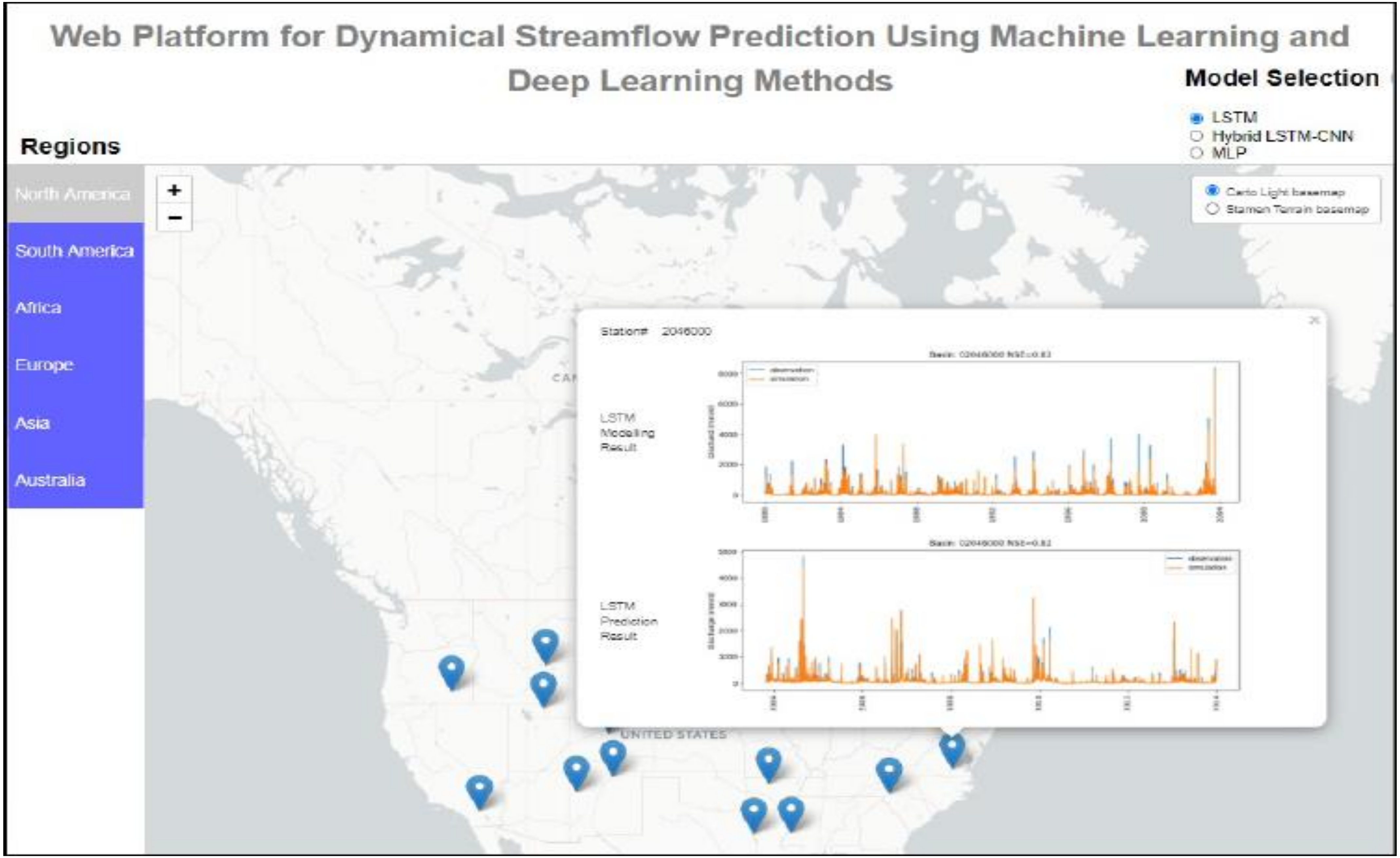
Web Platform

Interface:

- Created with HTML and JavaScript
- JavaScript to design popups and legend design
- Leaflet library in JavaScript provides interactive display elements for geographic information

Modules:

- Model selection option
- Region selection option
- Simulated Runoff Time-series Visualization
- Forecasting Runoff Time-Series Visualization



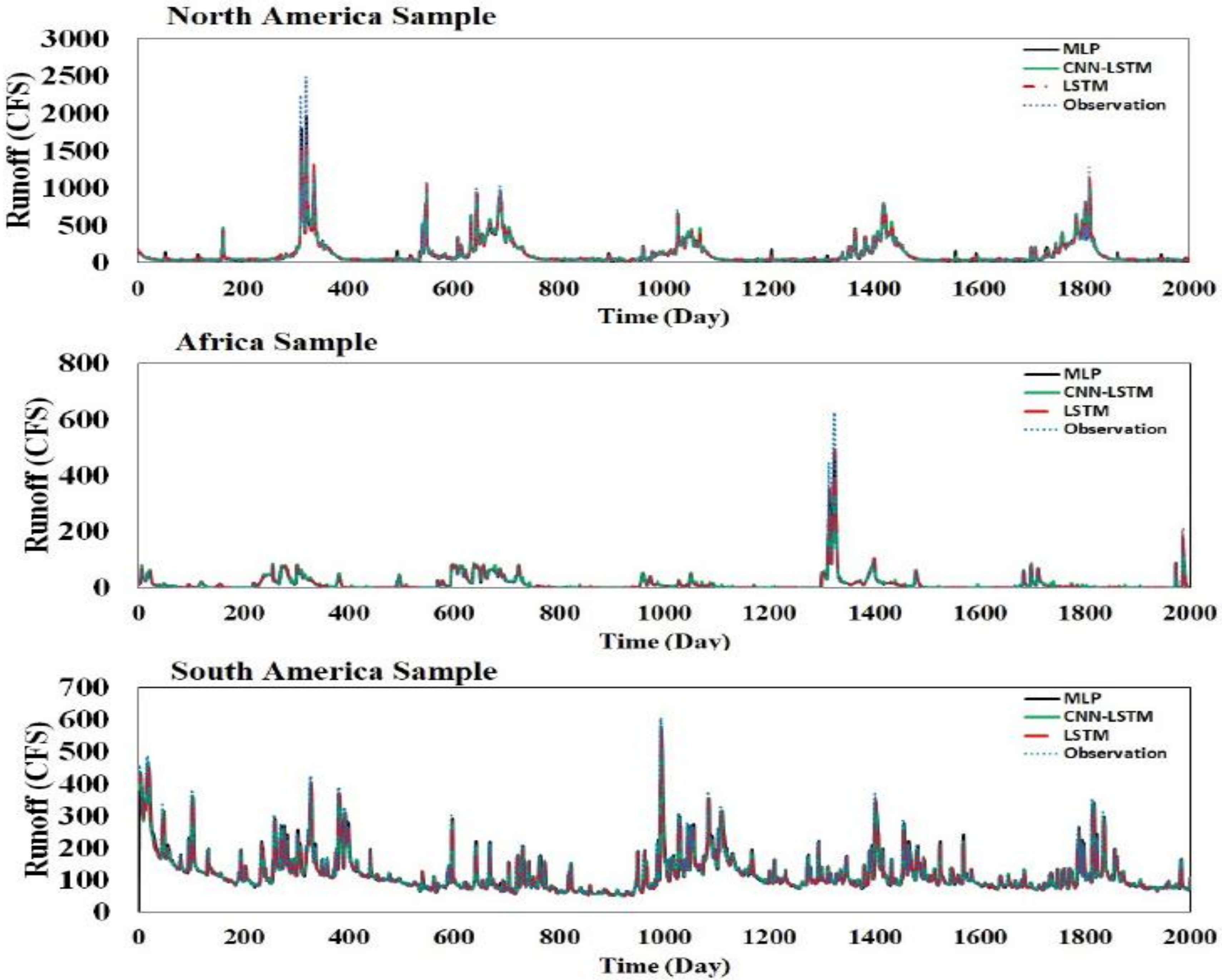
North America Case Study as an example

Experimental Results

Metrics used for evaluation:

- Nash-Sutcliffe Efficiency - domain $(-\infty, 1]$
- Kling-Gupta Efficiency - domain $(-\infty, 1]$
- Transformed Root Mean Squared Error - domain $[0, \infty)$

Region	Model	NSE	KGE	TRMSE	NSE (Mean)
North America	**MLP**	0.82	0.88	0.98	0.79
	LSTM	0.82	0.87	1.11	0.8
	CNN-LSTM	0.84	0.9	0.95	0.81
South America	**MLP**	0.85	0.9	0.67	0.6
	LSTM	0.86	0.9	0.66	0.6
	CNN-LSTM	0.86	0.9	0.66	0.62
Africa	**MLP**	0.75	0.79	0.79	0.75
	LSTM	0.74	0.82	0.81	0.77
	CNN-LSTM	0.7	0.69	1.74	0.77


North America Sample
Runoff (CFS)
Time (Day)
MLP
CNN-LSTM
LSTM
Observation
Africa Sample
Runoff (CFS)
Time (Day)
MLP
CNN-LSTM
LSTM
Observation
South America Sample
Runoff (CFS)
Time (Day)
MLP
CNN-LSTM
LSTM
Observation

Future Forecasting

Streamflow forecasts for the next 3 months for basin 0204600

Conclusion

- Data Driven methods showed satisfactory results.
- LSTM and CNN-LSTM had better performance compare to the MLP as they are memory-based methods
- ML methods cannot replace physical modelling, but strongly complement and enrich it.
- Base flow Separation for the future works is suggested.

Thanks for your attention!

Questions?

Github:

https://github.com/sadeghitabas/CPSC8810-Mining-Massive-Data