**Comparative Analysis of Loss Functions**

The selection of an appropriate loss function is crucial in the training of a machine learning model as it directly influences how well the model learns to perform its task. The loss function quantifies the difference between the predicted outputs of the model and the actual target values, providing a measure that the training algorithm seeks to minimize. Different loss functions can drive the model's learning in various ways, depending on the nature of the problem and the characteristics of the data. For instance, the Mean Squared Error (MSE) is commonly used in regression tasks because it emphasizes larger errors due to its squaring of the error terms, making it sensitive to outliers. In contrast, the Nash-Sutcliffe Efficiency (NSE), a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance, is particularly useful in hydrological and other environmental modeling where the assessment of predictive accuracy in relation to observed variability is crucial. The MSE and NSE loss function can be written as follows:

(1)

(2)

Where , , and are observation, simulation and mean flow, respectively.

To compare performance of NSE and MSE on building LSTM model which can be used to predict streamflow in multiple basins and ungauged basins, LSTM model have been trained and evaluated on different basins using MSE and NSE loss functions. However, To compare the results of calibrated model which has used either NSE or MSE, we have used the Kling–Gupta efficiency (KGE). The Kling–Gupta efficiency (KGE) serves as a widely recognized measure of fit within hydrological science, utilized for evaluating how well simulations align with observational data. KGE offers enhancements over commonly employed indicators, including the coefficient of determination and the NSE coefficient. KGE can be written as follows:

Where r, µ and are correlation coefficient, mean, and standard deviation, respectively.

We have evaluated the loss function using KGE metric on 01022500, 01013500, 01030500, 01031500, and 01047000 basins which are defined by USGS.

**analysis/results with preliminary plots**

The following table shows the best results of loss functions on evaluation sets on different basins after 50 epoch. We also computed the KGE metric to compare the capability of different loss functions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Basin | MSE | KGE | NSE | KGE |
| 01022500 | 1.42 | 0.79 | 0.76 | 0.79 |
| 01013500 | 0.66 | 0.92 | 0.89 | 0.94 |
| 01030500 | 1.28 | 0.84 | 0.82 | 0.86 |
| 01031500 | 2.38 | 0.82 | 0.81 | 0.80 |
| 01047000 | 3.49 | 0.75 | 0.75 | 0.75 |

LSTM model has been trained on the period starting from 1980-10-01 to 1995-09-30 and evaluated on the period starting from 1995-10-01 to 2000-10-01. To compare the performance of LSTM mode using different loss function, the LSTM streamflow prediction on evaluation set have been shown using NSE and MSE loss functions on multiple basins in the following figure.

