Are Machine Learning methods robust enough for hydrological modeling under changing conditions?

Carolina Natel de Moura[1,2], Jan Seibert[2], Miriam Rita Moro Mine[1], Ricardo Carvalho de Almeida[1]

carolina.nateldemoura@geo.uzh.ch

[1]Departament of Hydraulic and Sanitation
Federal University of Parana
Parana, Brazil

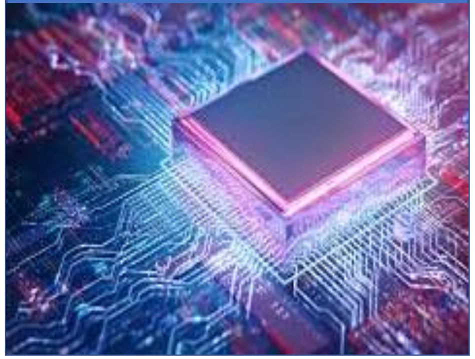
[2]Department of Geography
University of Zürich
Zürich, Switzerland

Problem formulation

Big Data explosion

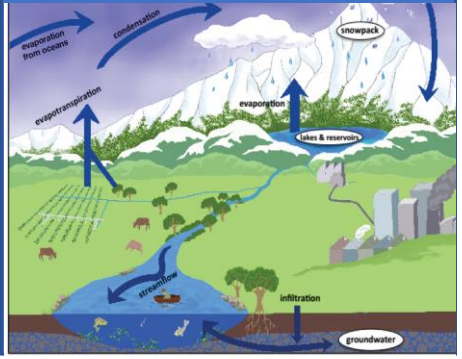
- Remote sensing
- Radars
- New large - scale datasets
- Citizen Science

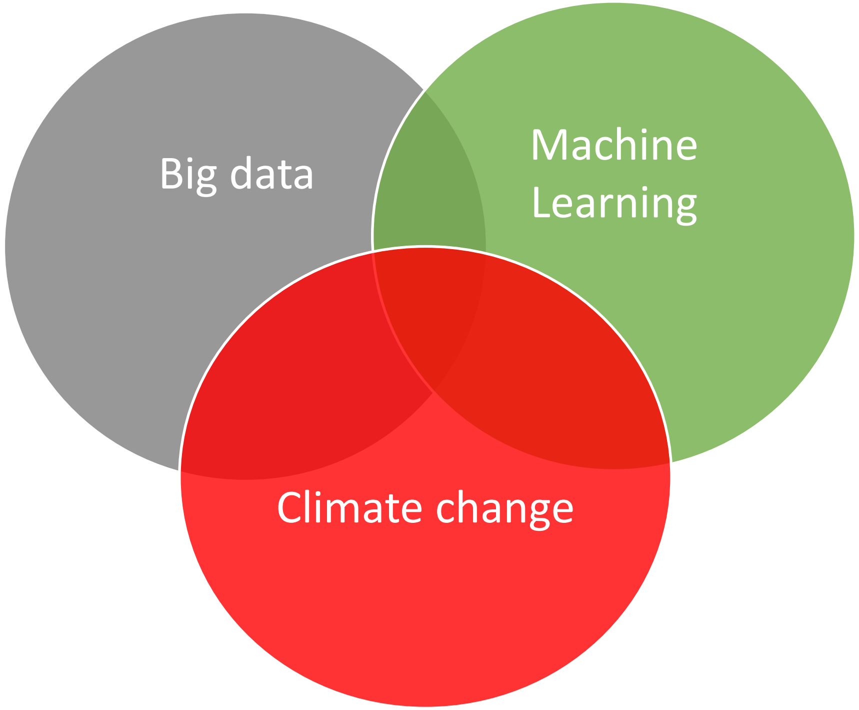
Advancement in Machine Learning algorithms and computational power



Climate change impacts on water resources

- Likely to change the hydrological processes patterns





Big data

Machine Learning

Climate change

Are Machine Learning methods robust enough for hydrological modeling under changing conditions?

Are Machine Learning methods robust enough for hydrological modeling under changing conditions?

The short answer is YES

The long answer is:

Compared to a traditional hydrological model, a data-driven model can be considered good enough for simulations in non-stationary conditions

- Besides that, as longer the dataset length used in calibration, better the model performance is (slide 9 and 11)
- The learning transfer from a model calibrated in a dry period to a wetter period is more difficult than the contrary (slide 13)
- There was an underestimation of the peak flows by the LSTM model in the Spring, period when there is contribution of snow melting in the total discharge (slide 14)

How did we test our hypothesis?

- Benchmarking a data-driven model over a traditional Hydrological model
- Comparing the model performance under non-stationary condition against the stationary condition
- Testing the impact of the dataset length used in calibration on the model performance

Study area & Data

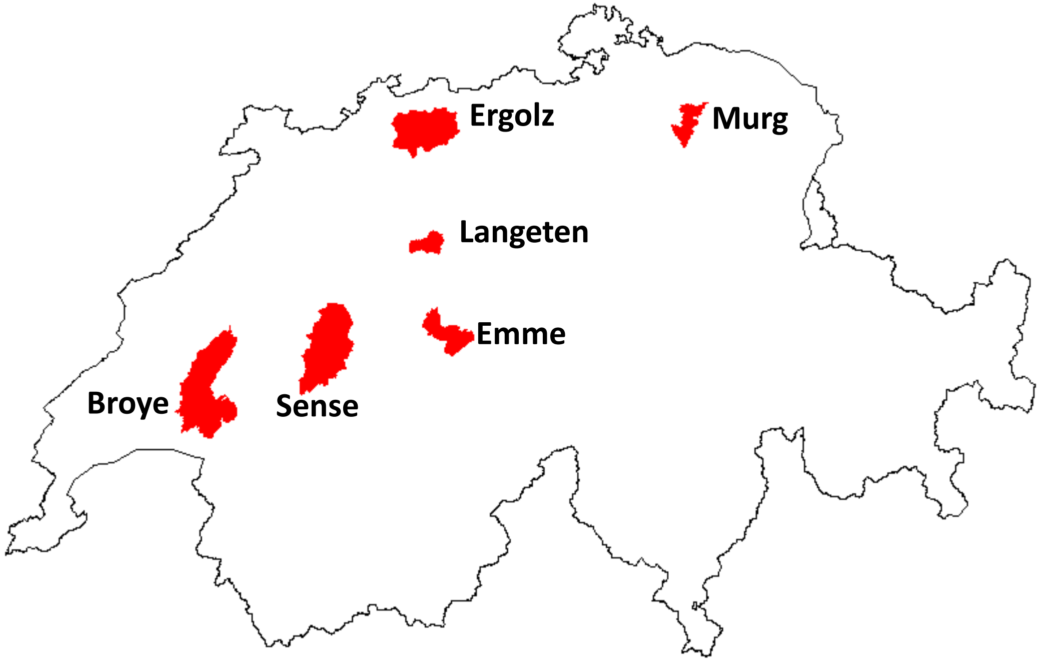


Figure 1. Catchment locations in Switzerland

Six snow-influenced catchments
Area: ~ 60 to 400 km²
Mean altitude: ~ 500 to 1200 m.a.s.l

Future changes expected for Switzerland

- Increase in mean and maximum floods for most of Switzerland
  Near future: 5 – 24%
  Far future (25 – 49%)
- Different signal for Southern alpine catchments: mean annual floods decrease in the far future

Source: Köplin et al., 2014

Data required

- Daily air temperature (°C)
- Precipitation (mm.d^{-1})
- Discharge (mm.d^{-1})
- Monthly long-term potential evapotranspiration rates (mm).

Machine Learning model

- Long Short - Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997)

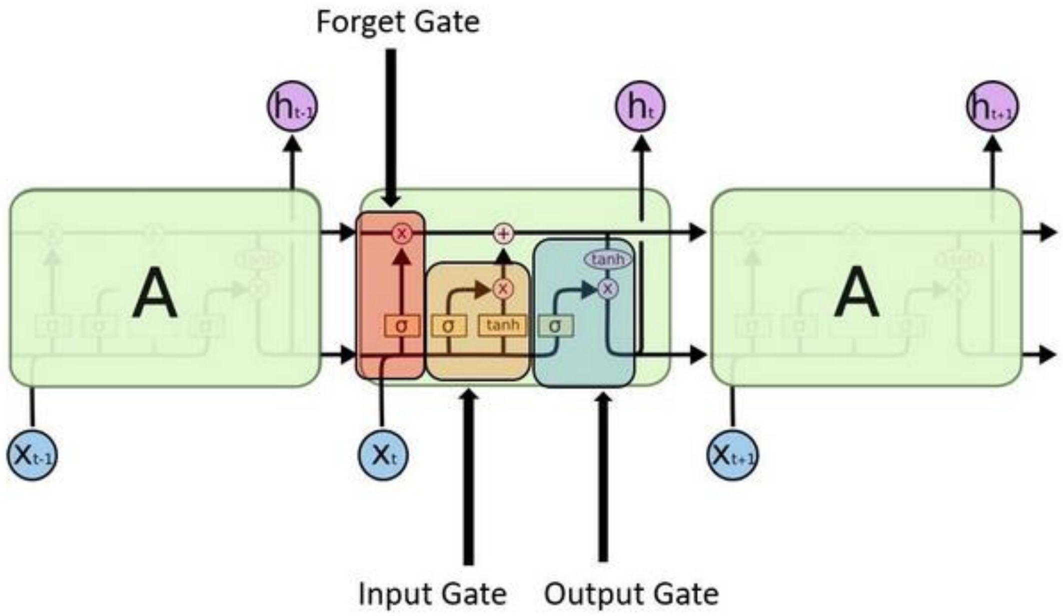


Figure 2. LSTM model representation.
Source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

LSTM model structure*

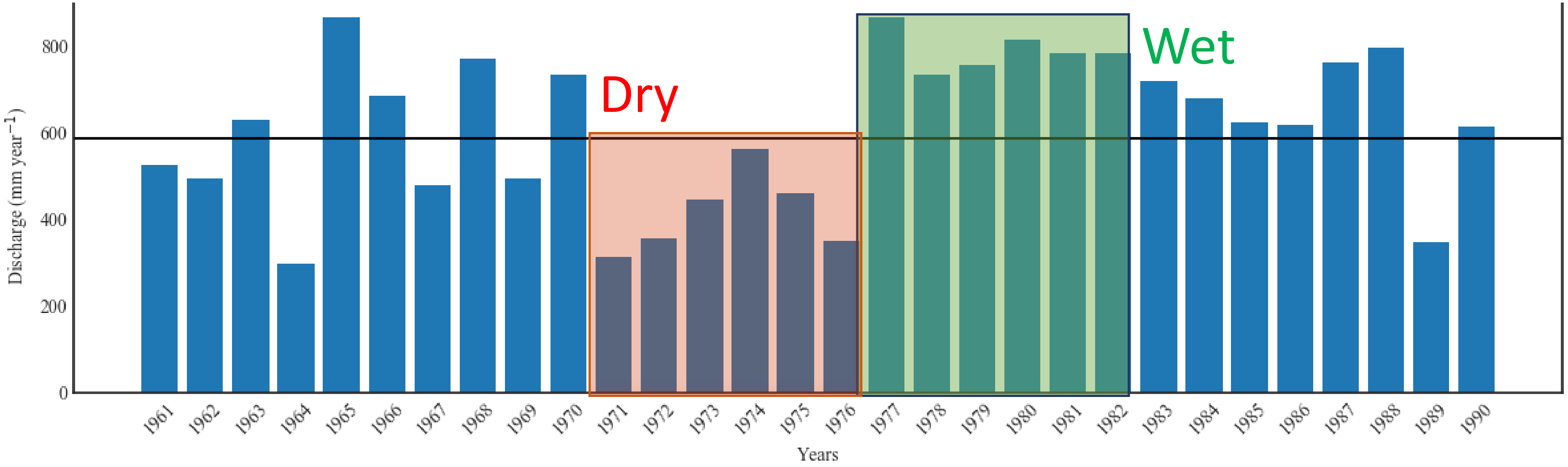
- Number of neurons: 50
- Number of steps: 365
- Number of epochs: 50
- Number of repetitions: 20
- Batch size division: 12
- Dropout rate: 0%

- The network structure was determined through a process of trial and error

Experimental design: the cases under analysis

For the **stationary case,** we selected continuous periods of data containing both dry and wet years, which we split in calibration and validation period (50/50).

For the **non-stationarity case**, we used the Differential Split Sample Test - DSST (Klemes, 1986) to split the data set in dry and wet periods. The model was calibrated in a **dry** period and validated in a **wet** one, and vice-versa.



Streamflow mean difference of 50% between the different climate periods. Compatible with the streamflow changes expected for Switzerland

Figure 3. Example of the dataset split in dry and wet periods. The horizontal line represents the long-term average anual discharge.

Experimental design: the models

- **LSTM**: data driven-model

Benchmarking the performance of the LSTM model over a conceptual model:

- **Lower benchmark**: ensemble mean of simulations using 1,000 random parameter sets for the HBV model
- **Upper benchmark**: automatic calibration of the HBV model through a Genetic Algorithm within feasible parameter ranges.

➢ Does the dataset length used in calibration affects the model performance?

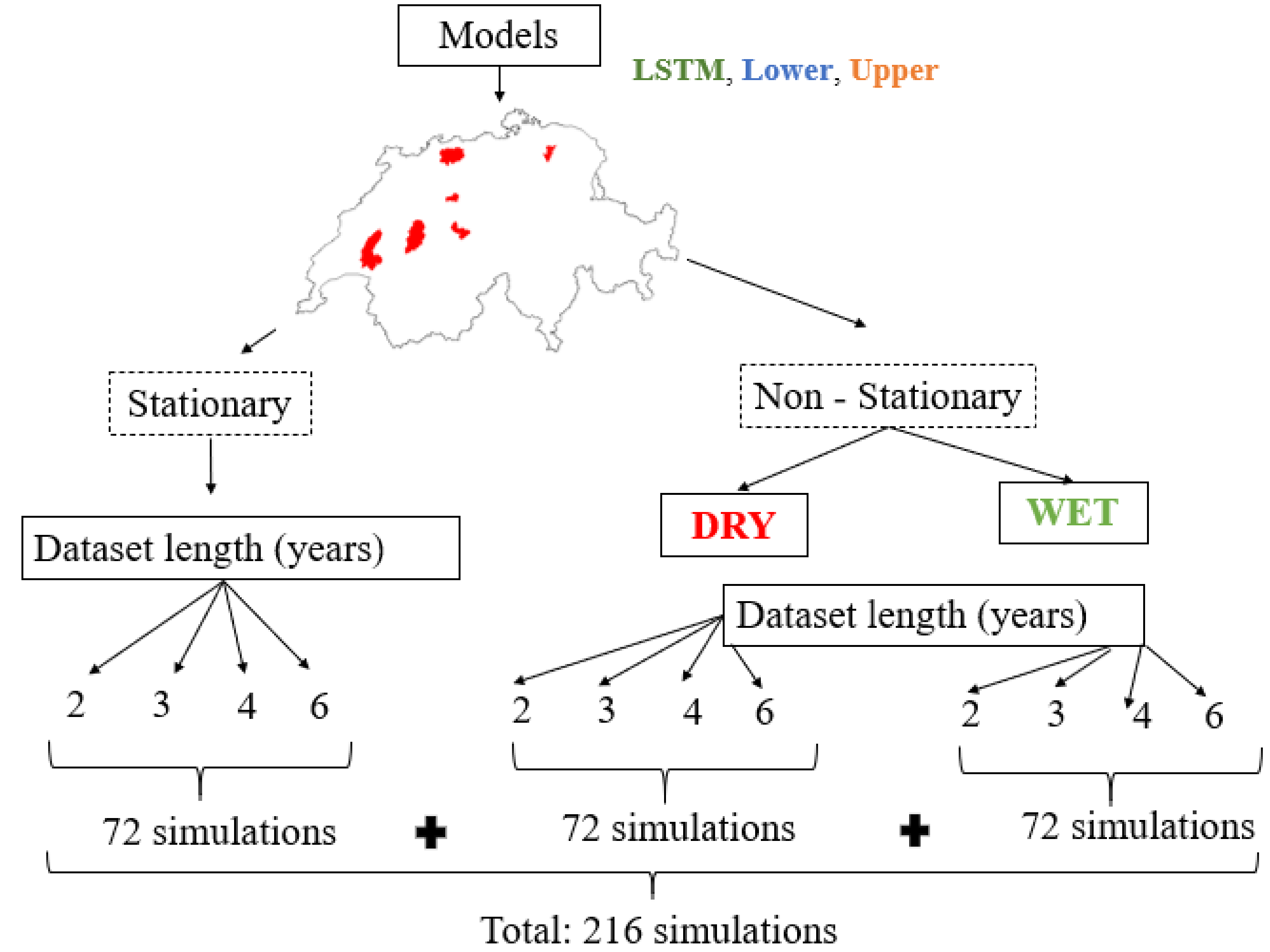


Figure 4. Methodology flowchart.

Model performance by case and dataset length
Nash-Sutcliffe Efficiency

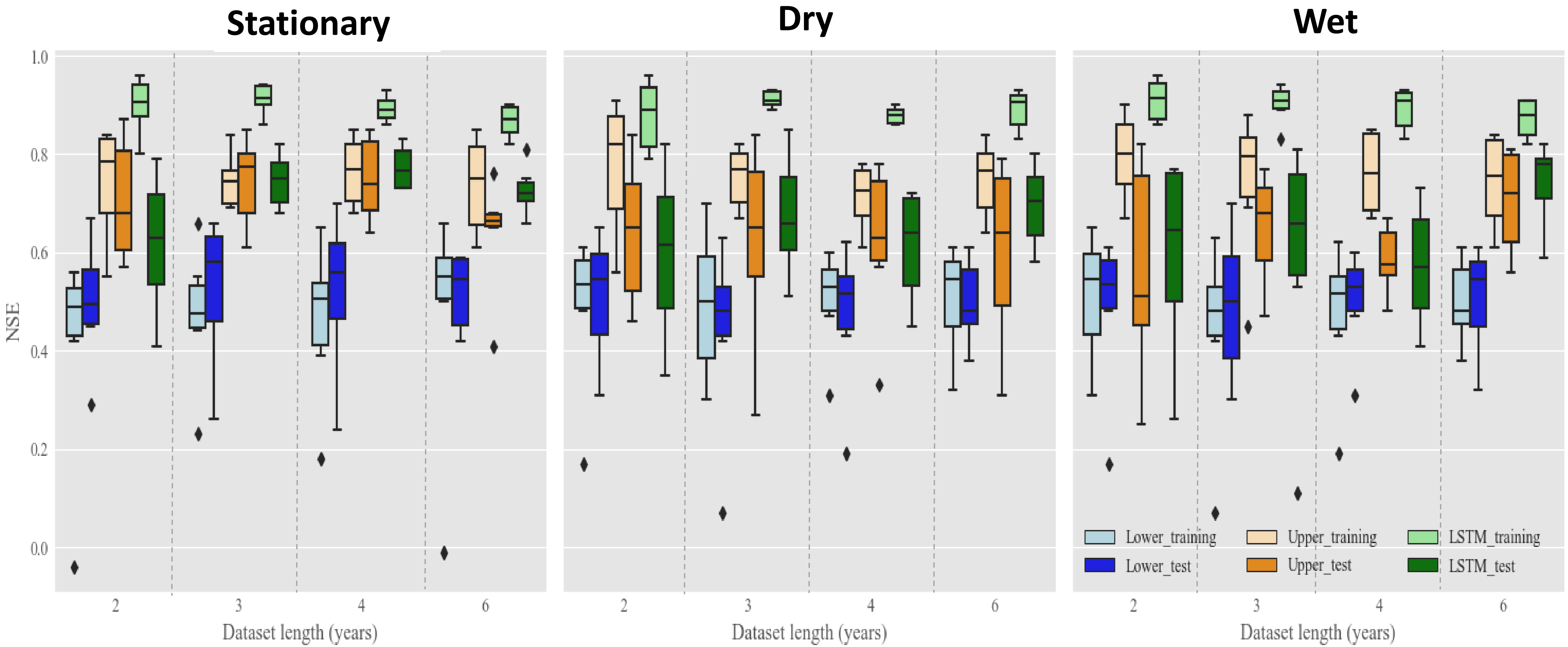
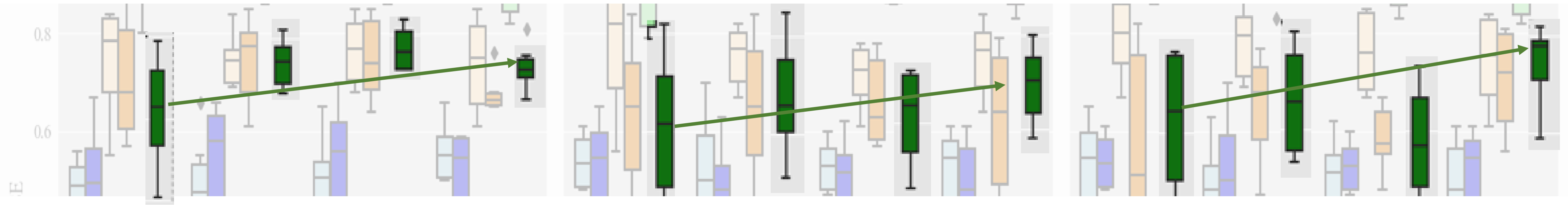


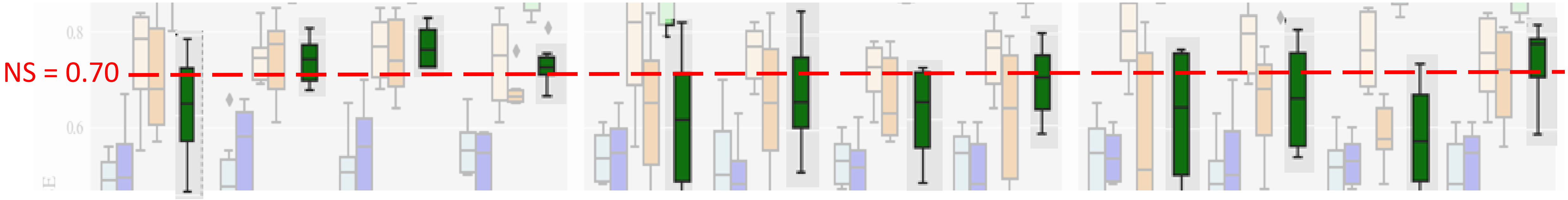
Figure 5. Boxplot of the model performance (NSE) in stationary and non-stationary conditions using LSTM for streamflow prediction.

Model performance

Stationary **Dry** **Wet**



Positive correlation between LSTM performance and dataset length for model generalization



Simulations for non-stationary conditions performed slightly poorer than those for stationary conditions
Less accentuated using 6 years in calibration.

Percentage of the time in which the LSTM model has higher performance than Lower and Upper benchmarks

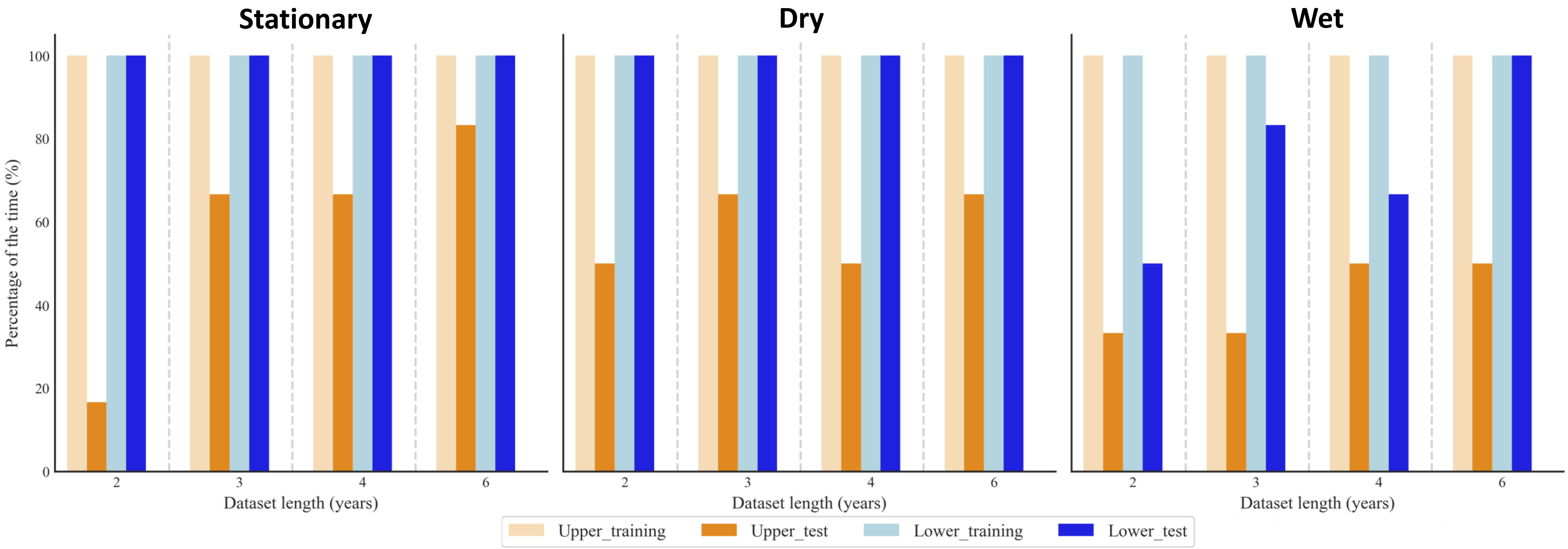
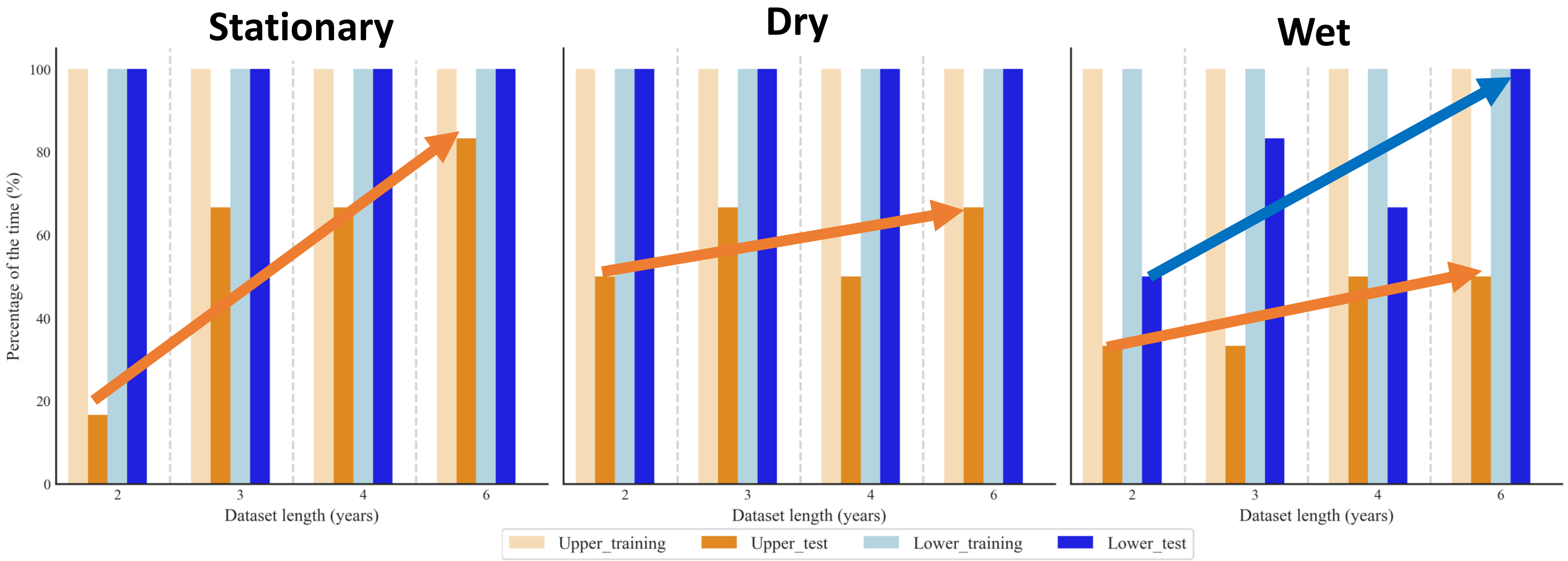


Figure 6. Percentage of the time in which the LSTM model has higher performance than the Lower and Upper benchmarks, based on the NSE, for stationary and non-stationary conditions.

Percentage of the time in which the LSTM model has higher performance than Lower and Upper benchmarks

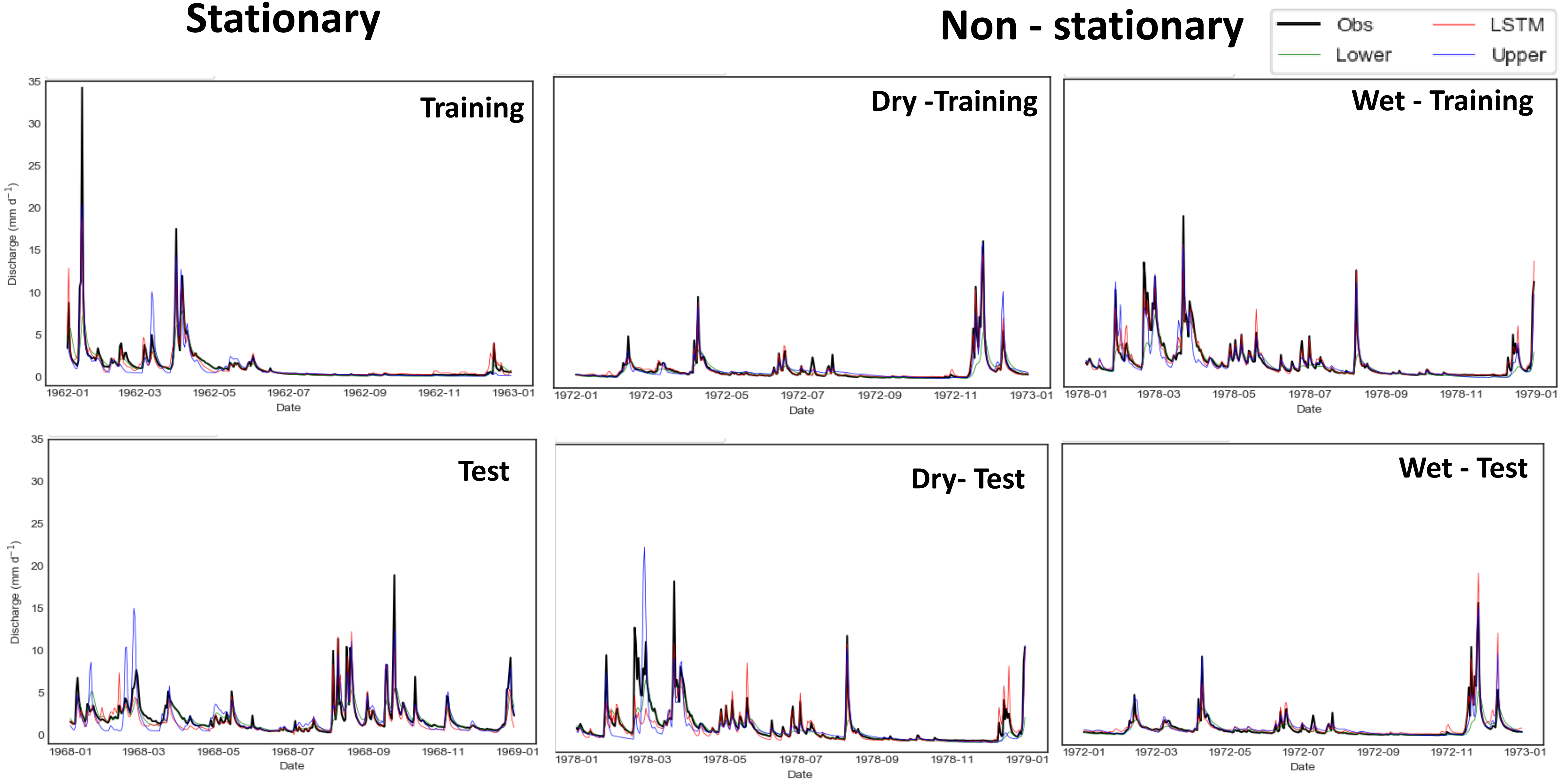


- LSTM performance is higher than the Lower and Upper benchmark in 100% of the time for the training period (light orange and light blue columns)
- The relative performance increases with the increase of the dataset length used in training (more evident for the stationary case)
- The lesser contribution of the dataset length in model performance for non-stationary case may be explained by the limitation of the data provided for the learning process (only dry or wet periods used in training). Less additional information about the hydrological processes are being provided

Hydrograms – Broye catchment

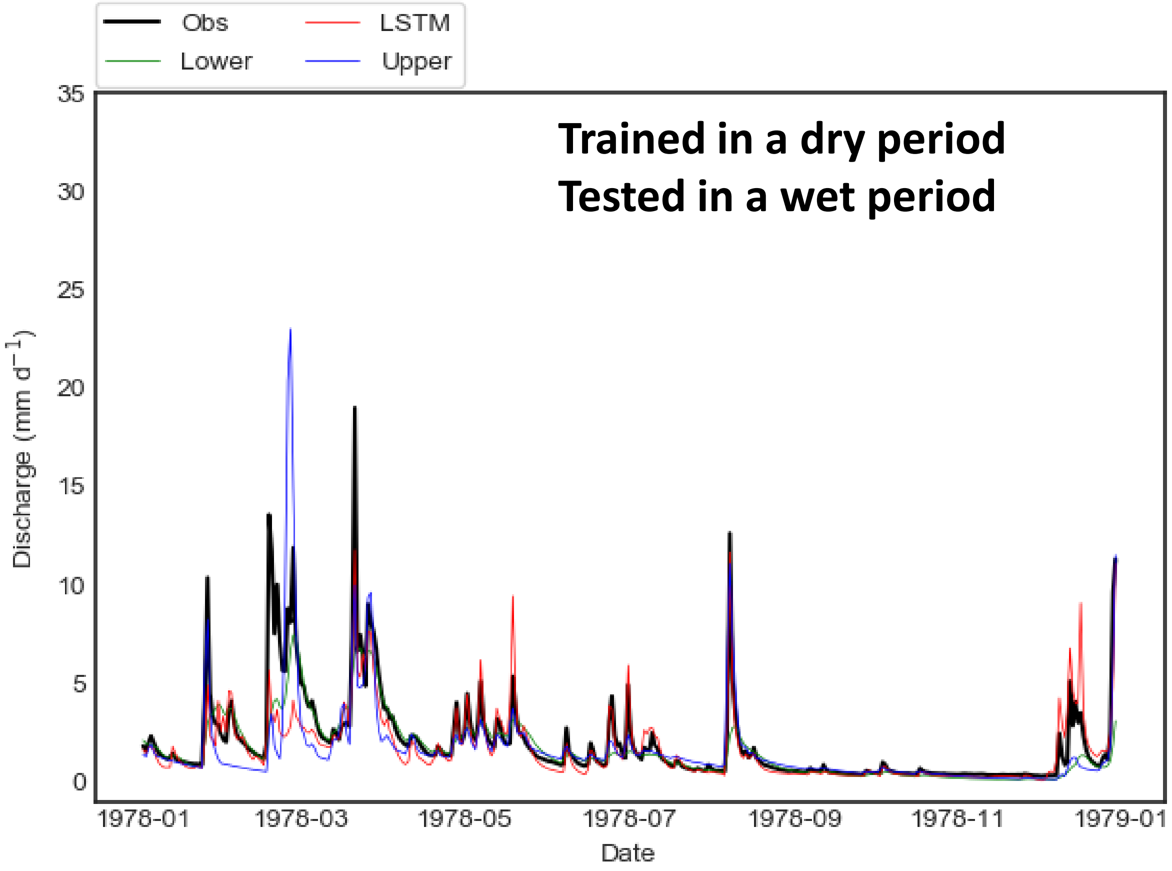
Stationary

Non - stationary

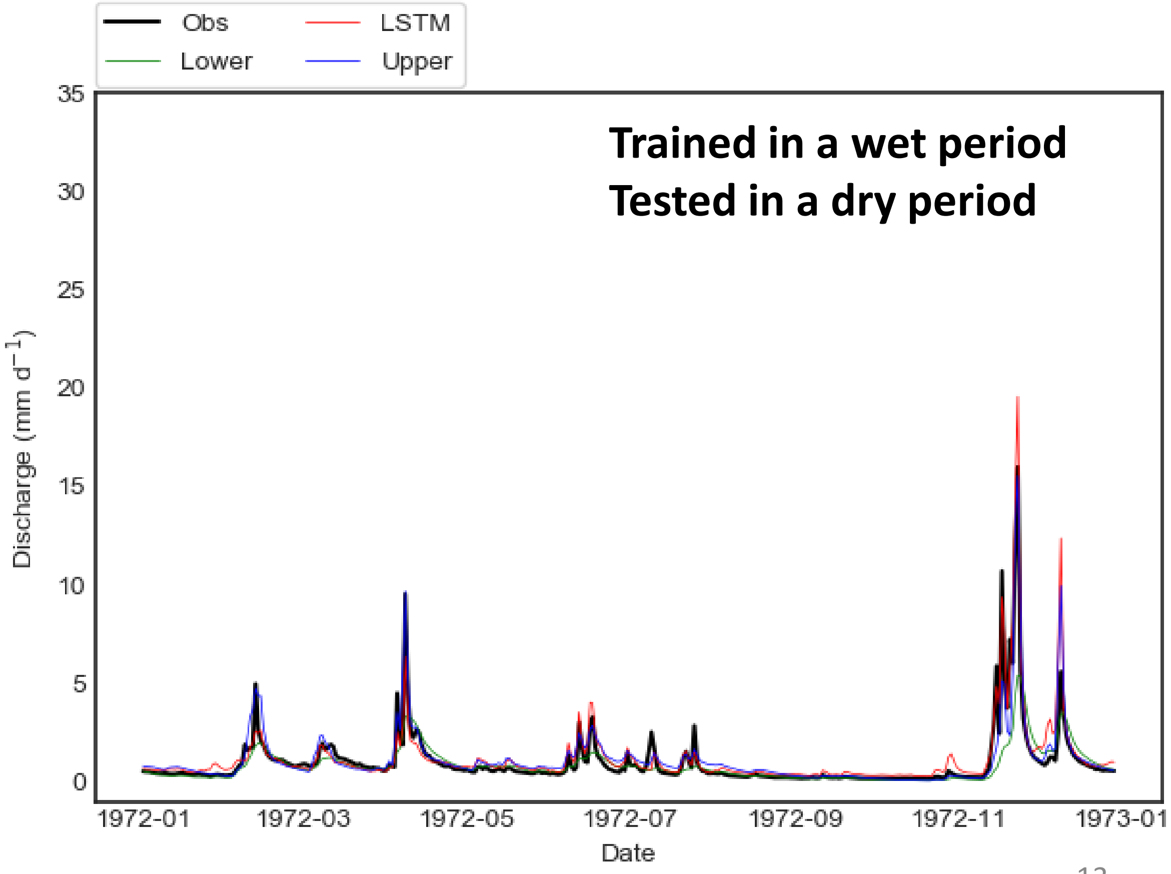


Learning transfer – Non-stationary

The learning transfer from a model calibrated in a dry period to a wetter period is more difficult than the contrary.

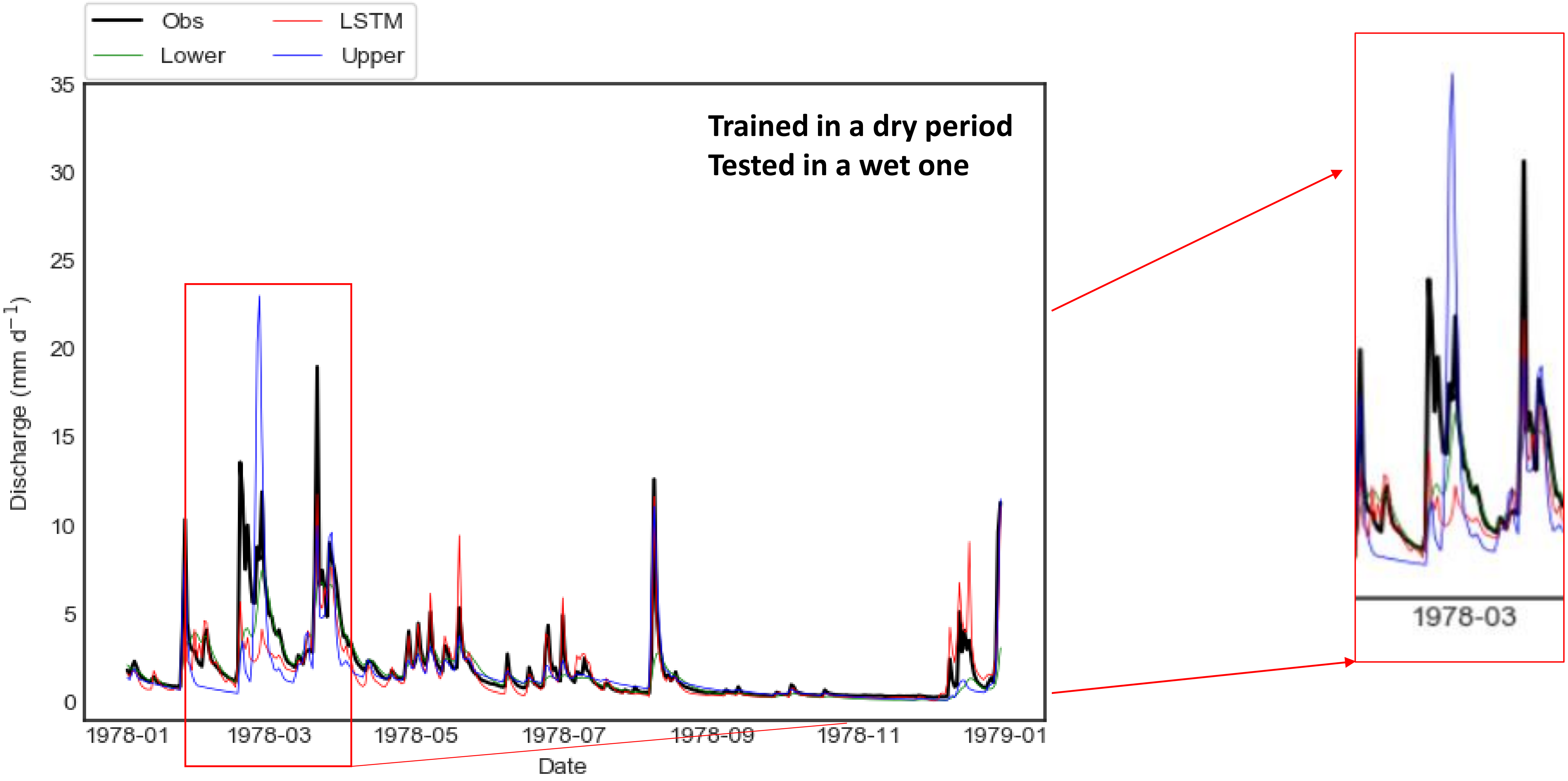


Trained in a dry period
Tested in a wet period



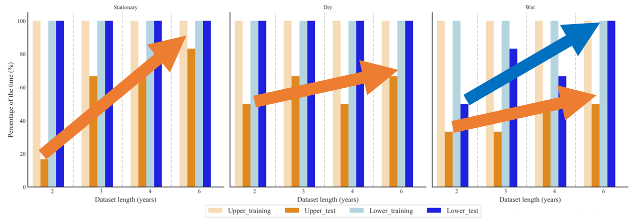
Trained in a wet period
Tested in a dry period

Poor representation of the snow-melting period

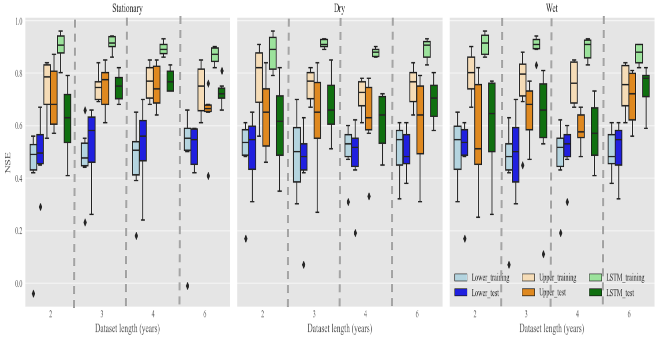


Trained in a dry period
Tested in a wet one

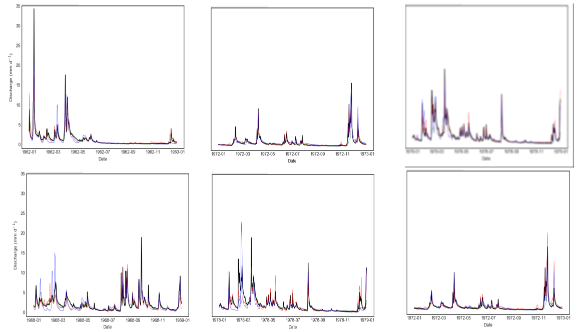
Final conclusions



- Training set: The LSTM performs better than the Lower and Upper benchmarks for both stationary and non-stationary conditions, independently of the dataset length used in calibration
- However, for the Test set, the LSTM seems to need a larger dataset length to beat the Upper benchmark performance, for both the stationary and non-stationary cases



- Simulations for non-stationary conditions performed slightly poorer than those for stationary conditions (less observed using longer dataset length in calibration)
- Positive correlation between LSTM performance and dataset length for model generalization



- The generalization of the model calibrated from a dry period to a wetter period is more difficult than the contrary
- There was an underestimation of the peak flows by the LSTM model and difficult to represent the streamflow in the Spring, period when there is contribution of snow melting.

Take home messages

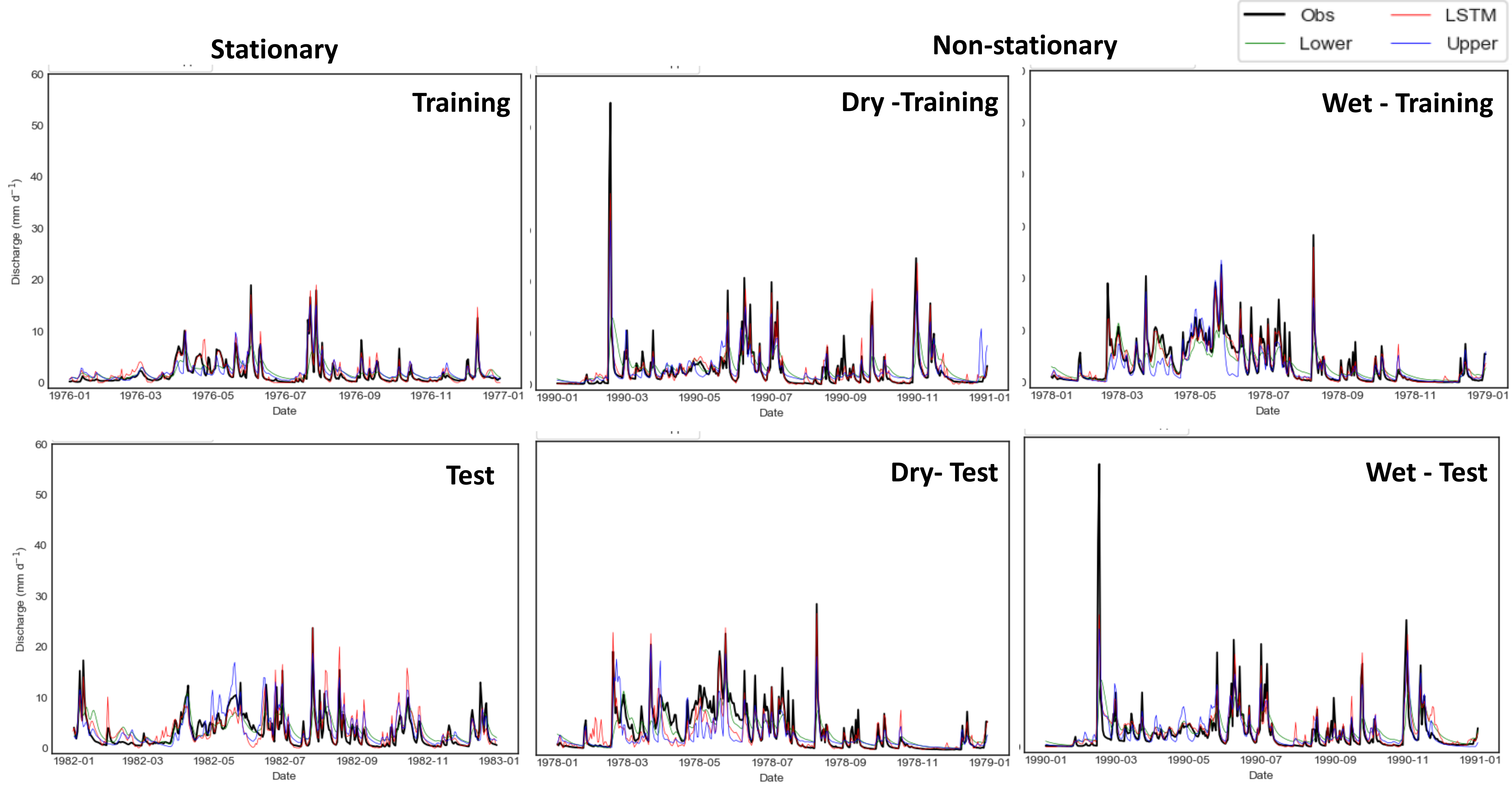
- The LSTM model can be applied for climate change assessments, when the series are likely to be non-stationary

- The dataset length is an important factor on the model performance

- Other input data can be added to the model structure in order to improve the representation of the snow melting period, or the use of a hybrid model, accounting for the snow routine

- The interpretability of the neural network cells is a plus in the application of LSTM model in Hydrology, to be explored further

- **Emme**



Stationary

Non-stationary

Training

Dry -Training

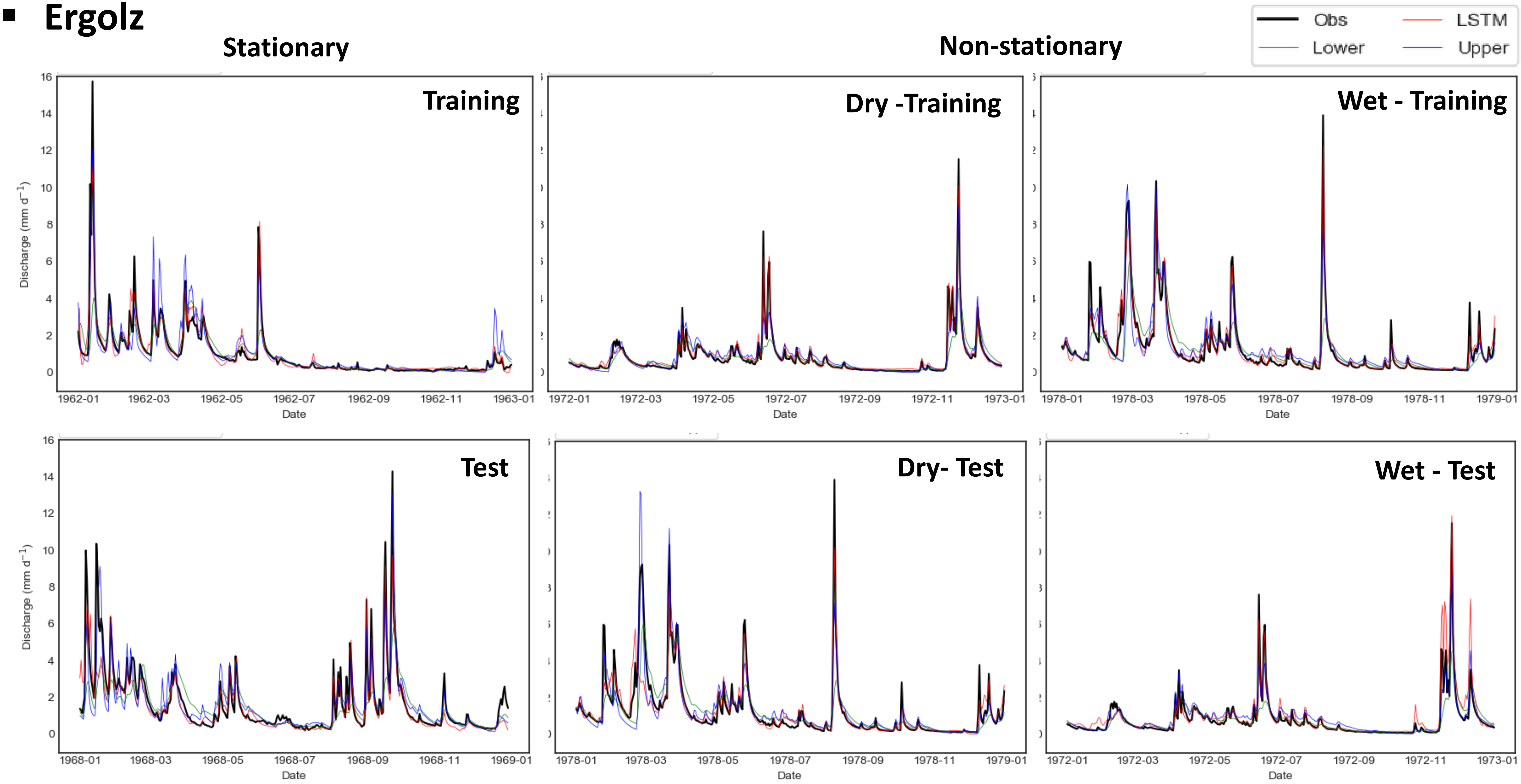
Wet - Training

Test

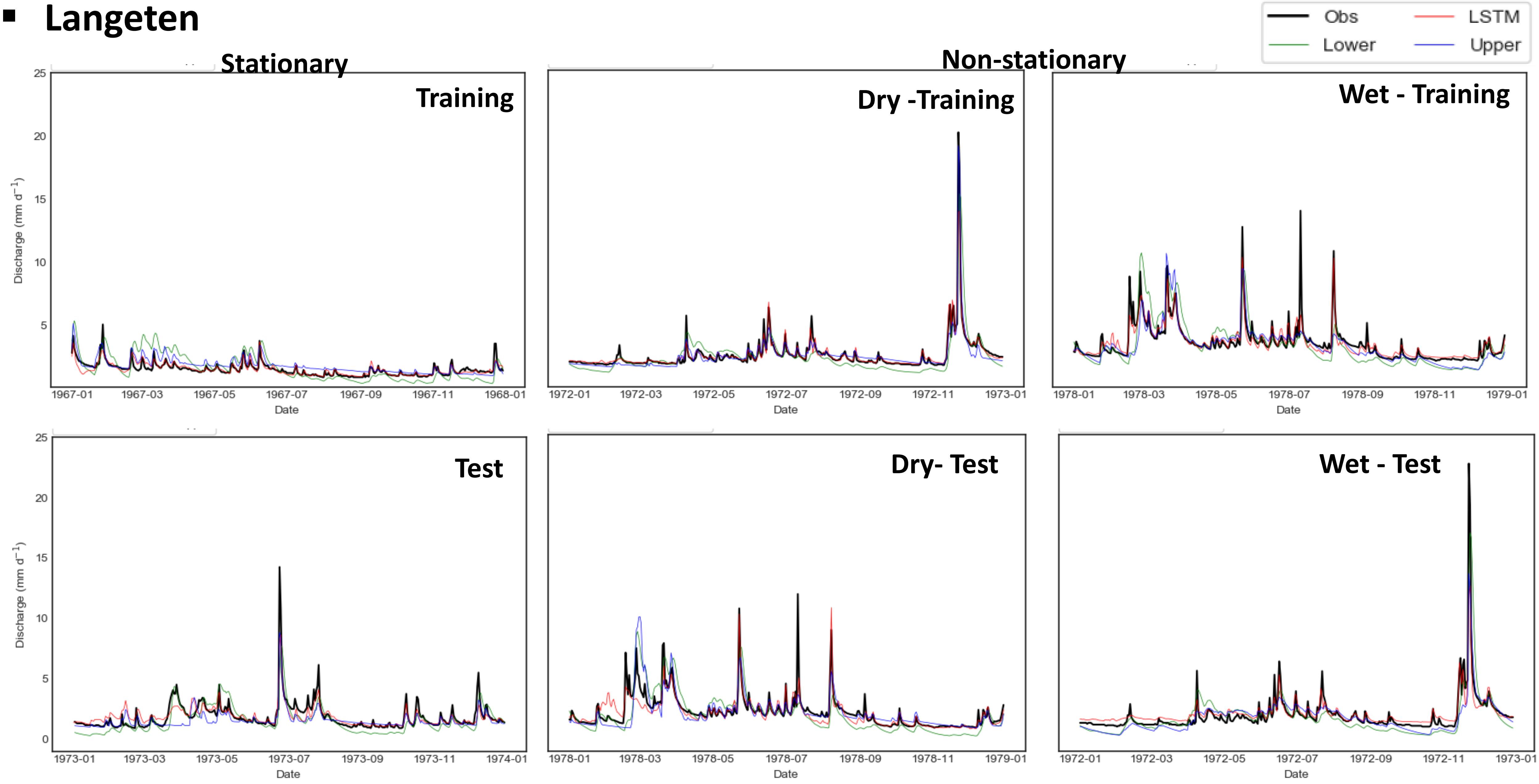
Dry- Test

Wet - Test

- **Ergolz**


Stationary
Non-stationary
Obs
LSTM
Lower
Upper
Training
Dry -Training
Wet - Training
Test
Dry- Test
Wet - Test
Discharge (mm d^{-1})
Date
0
2
4
6
8
10
12
14
16
1962-01
1962-03
1962-05
1962-07
1962-09
1962-11
1963-01
1972-01
1972-03
1972-05
1972-07
1972-09
1972-11
1973-01
1978-01
1978-03
1978-05
1978-07
1978-09
1978-11
1979-01
1968-01
1968-03
1968-05
1968-07
1968-09
1968-11
1969-01

- **Langeten**



Stationary

Non-stationary

Training

Dry -Training

Wet - Training

Test

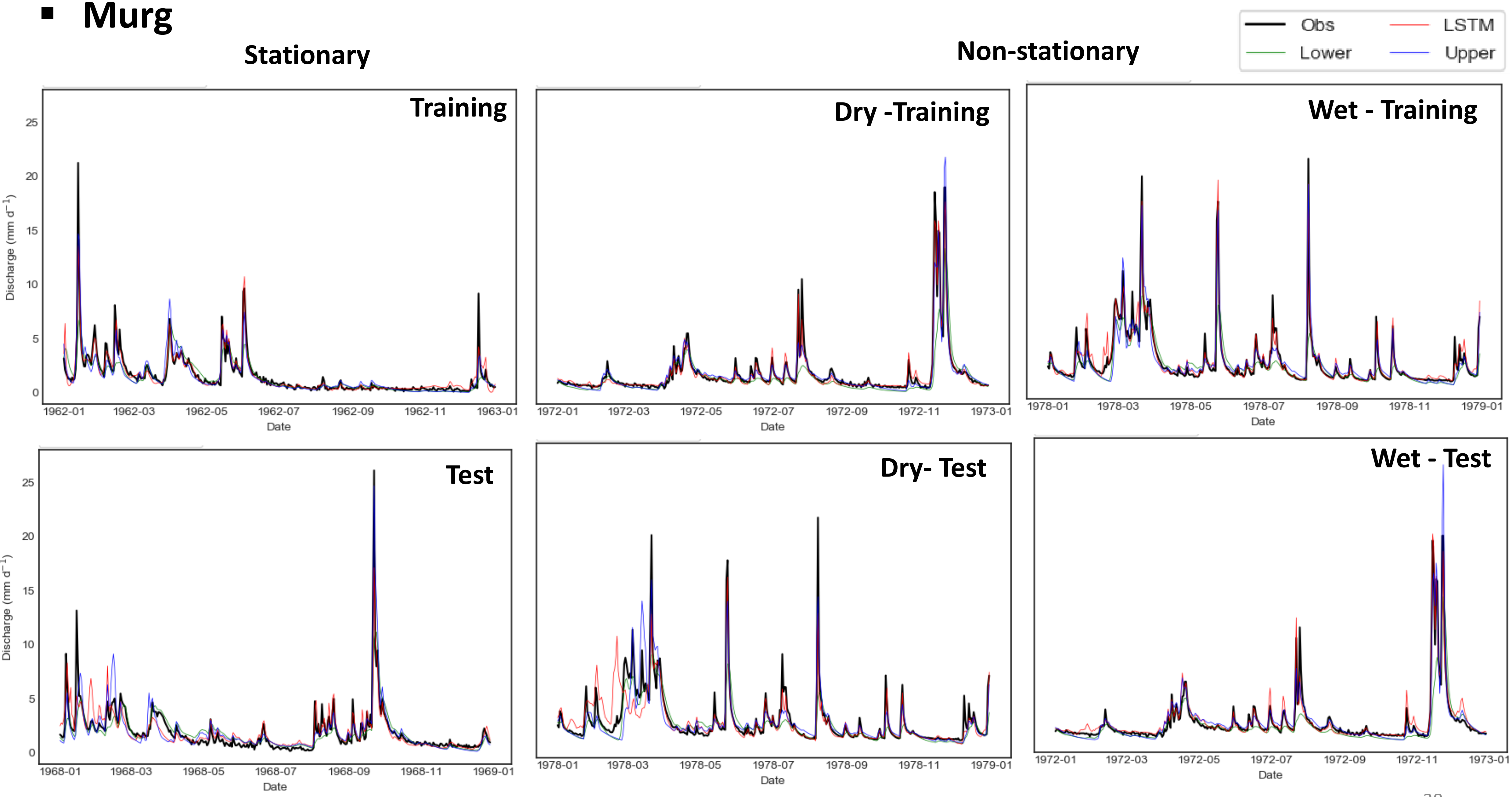
Dry- Test

Wet - Test

▪ Murg

Stationary

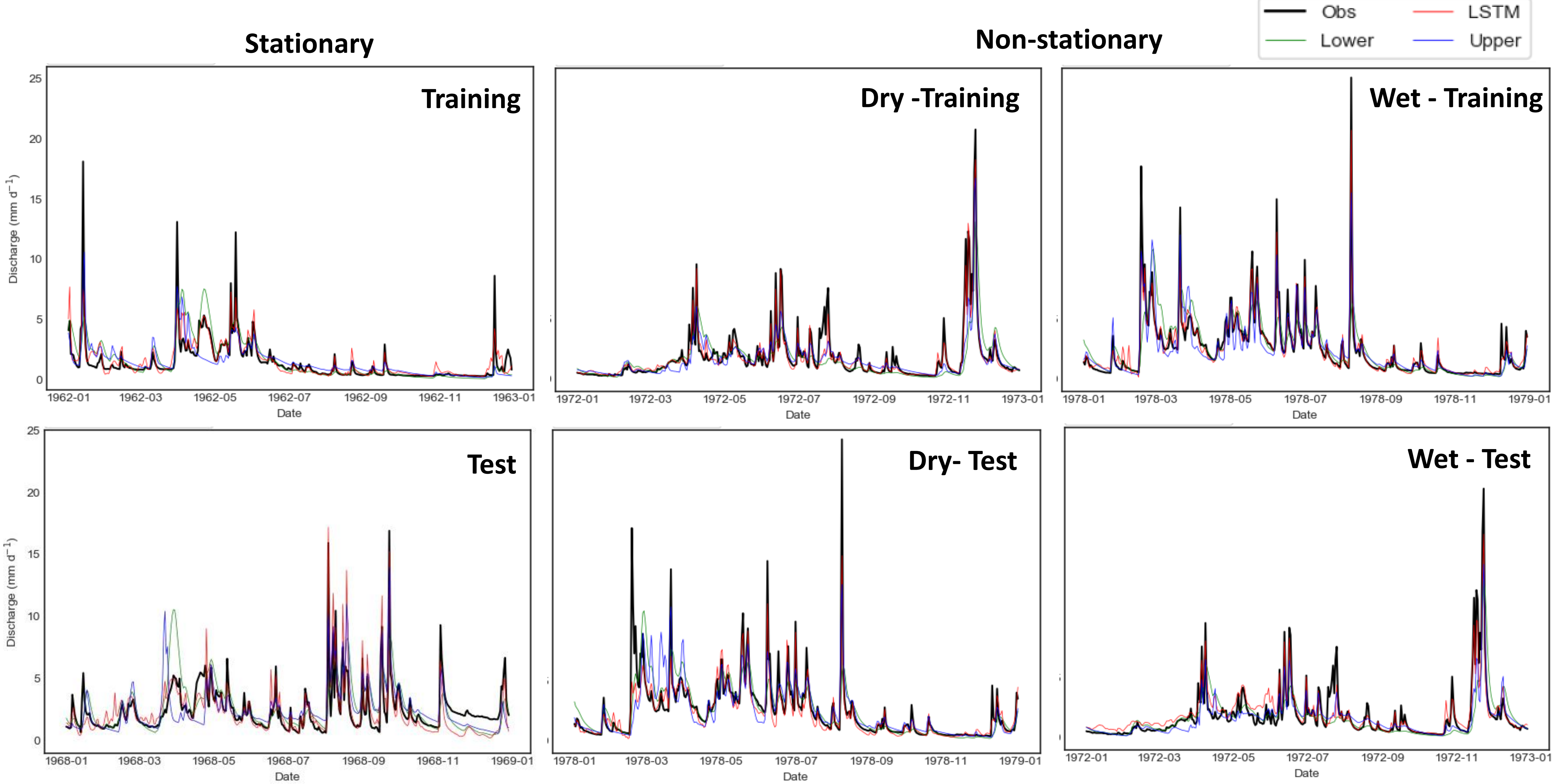
Non-stationary



- **Sense**

Stationary

Non-stationary


Obs
LSTM
Lower
Upper
Training
Dry -Training
Wet - Training
Test
Dry- Test
Wet - Test
Discharge (mm d^{-1})
Date