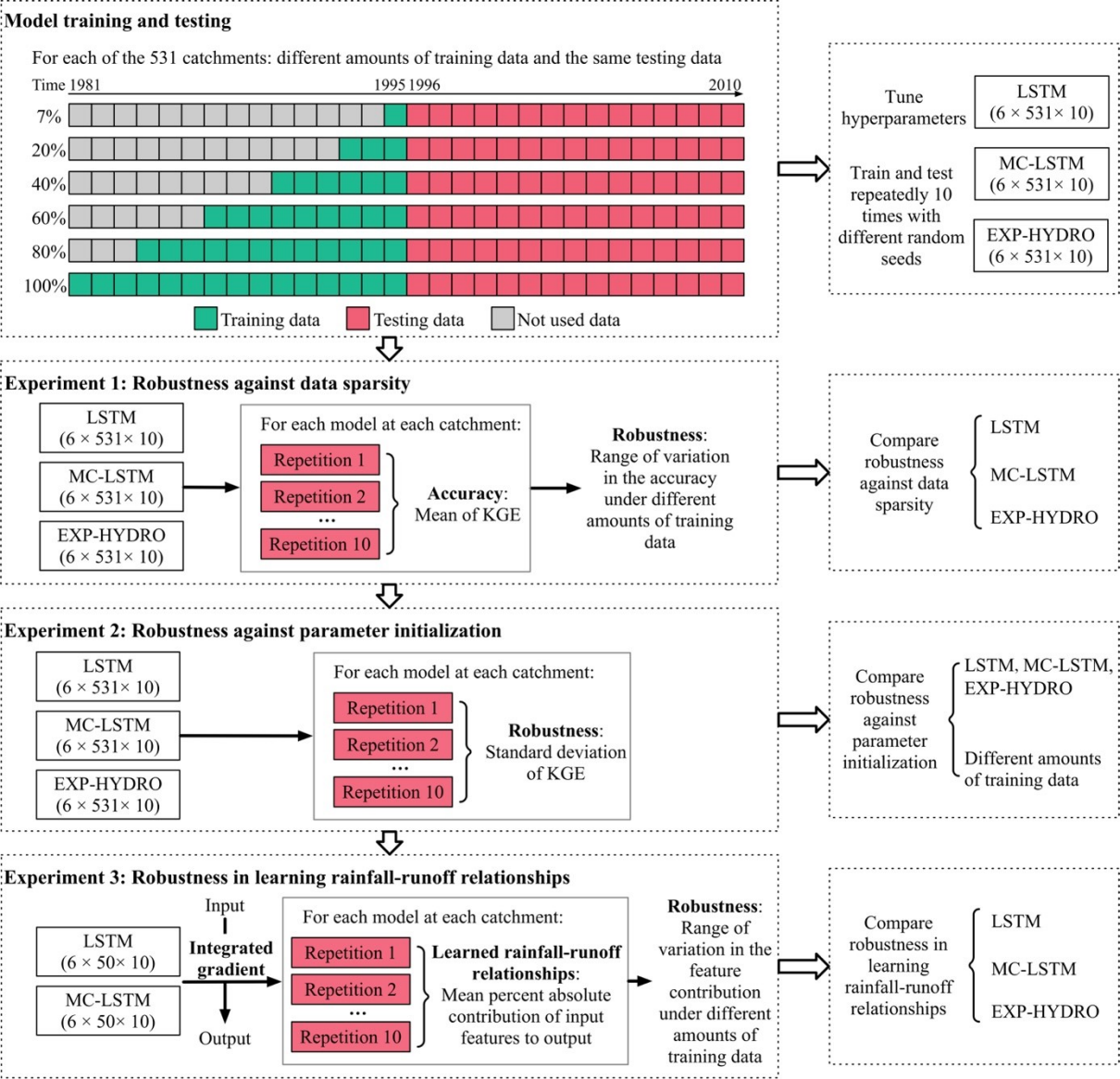
Model training and testing
For each of the 531 catchments: different amounts of training data and the same testing data
Time 1981
1995 1996
2010
7%
20%
40%
60%
80%
100%
Training data
Testing data
Not used data
Tune hyperparameters
LSTM (6 × 531× 10)
Train and test repeatedly 10 times with different random seeds
MC-LSTM (6 × 531× 10)
EXP-HYDRO (6 × 531× 10)
Experiment 1: Robustness against data sparsity
LSTM (6 × 531× 10)
MC-LSTM (6 × 531× 10)
EXP-HYDRO (6 × 531× 10)
For each model at each catchment:
Repetition 1
Repetition 2
...
Repetition 10
Accuracy: Mean of KGE
Robustness: Range of variation in the accuracy under different amounts of training data
Compare robustness against data sparsity
LSTM
MC-LSTM
EXP-HYDRO
Experiment 2: Robustness against parameter initialization
LSTM (6 × 531× 10)
MC-LSTM (6 × 531× 10)
EXP-HYDRO (6 × 531× 10)
For each model at each catchment:
Repetition 1
Repetition 2
...
Repetition 10
Robustness: Standard deviation of KGE
Compare robustness against parameter initialization
LSTM, MC-LSTM, EXP-HYDRO
Different amounts of training data
Experiment 3: Robustness in learning rainfall-runoff relationships
LSTM (6 × 50× 10)
MC-LSTM (6 × 50× 10)
Input
Integrated gradient
Output
For each model at each catchment:
Repetition 1
Repetition 2
...
Repetition 10
Learned rainfall-runoff relationships: Mean percent absolute contribution of input features to output
Robustness: Range of variation in the feature contribution under different amounts of training data
Compare robustness in learning rainfall-runoff relationships
LSTM
MC-LSTM
EXP-HYDRO