IWRA's XVII

WORLD WATER CONGRESS

제 17차 IWRA 세계물총회

29 November – 3 December 2021
EXCO, Daegu, Republic of Korea

50 IWRA | DAEGU METROPOLITAN CITY | KOREA WATER RESOURCES ASSOCIATION | K water Korea Water Resources Corporation | Ministry of Environment

Application of the artificial neural network to regional frequency analysis for estimating rainfall quantiles at ungauged sites

HANBEEN KIM, TAEREEM KIM, **JOOHYUNG LEE**, JUN-HAENG HEO
Hydrology and Water Resources System Lab. Yonsei University

1. Background & Objectives

2. Methodology

3. Application

4. Results

5. Conclusions & Future studies

1. Background & Objectives

2. Methodology

3. Application

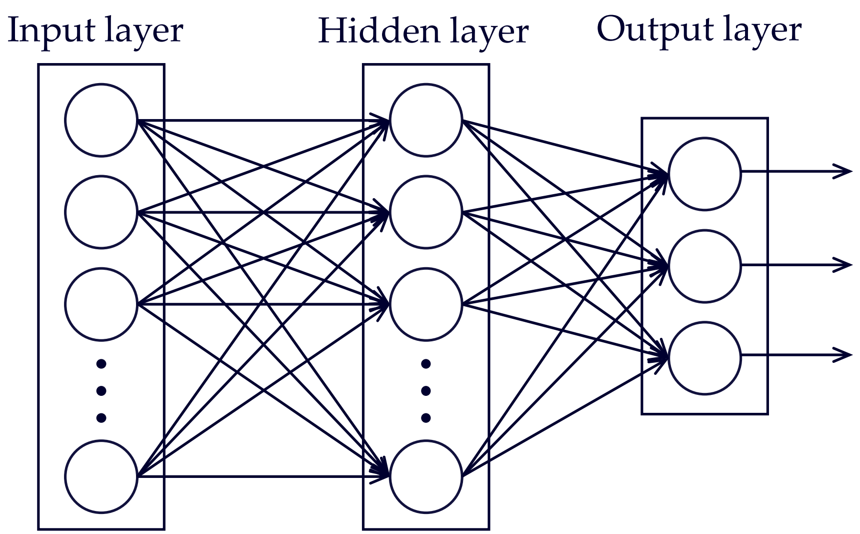
4. Results

5. Conclusions & Future studies

1. Background & Objectives

- Regional frequency analysis
  - To reduce uncertainties due to short data length
  - Can determine more reliable quantiles of the site by using all sites' data in a region
- Widely applied approaches of regional frequency analysis
  - Index flood method
  - Regression technique

- Artificial neural network model (ANN)
  - Suggested by McCulloch and Pitss (1943)
  - Sensitive to model structure and input data

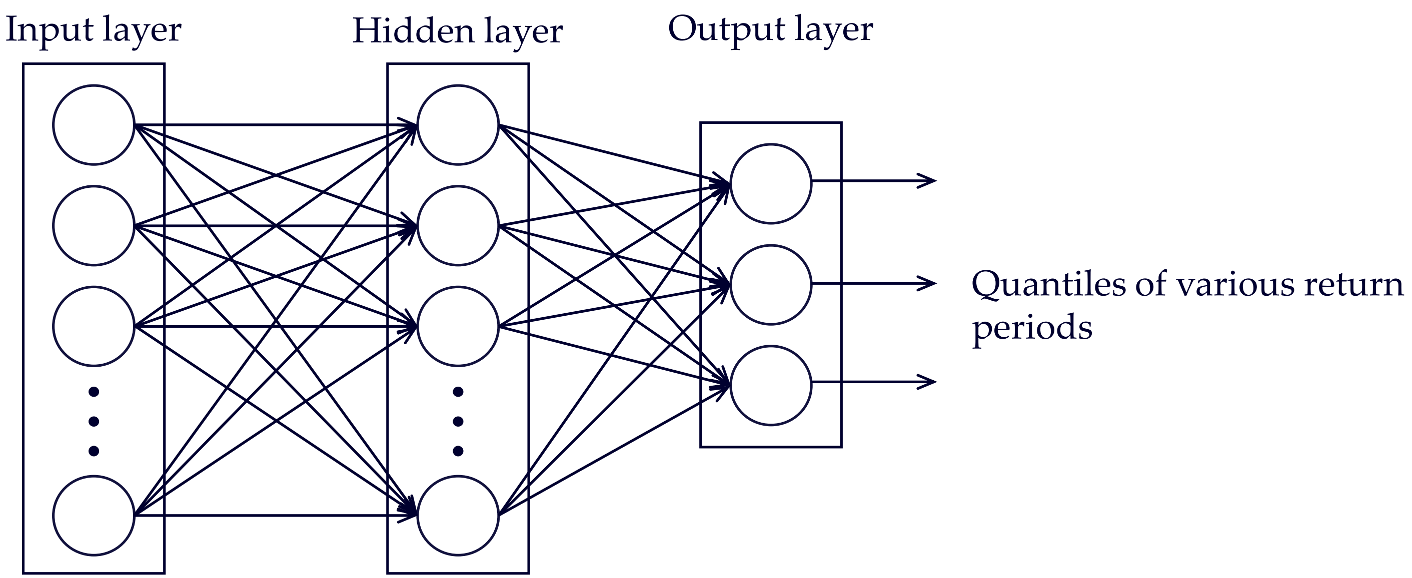


- Objectives
  - Determination of better ANN model to estimate quantiles
  - The assessment of performances for three different regional frequency analysis methods

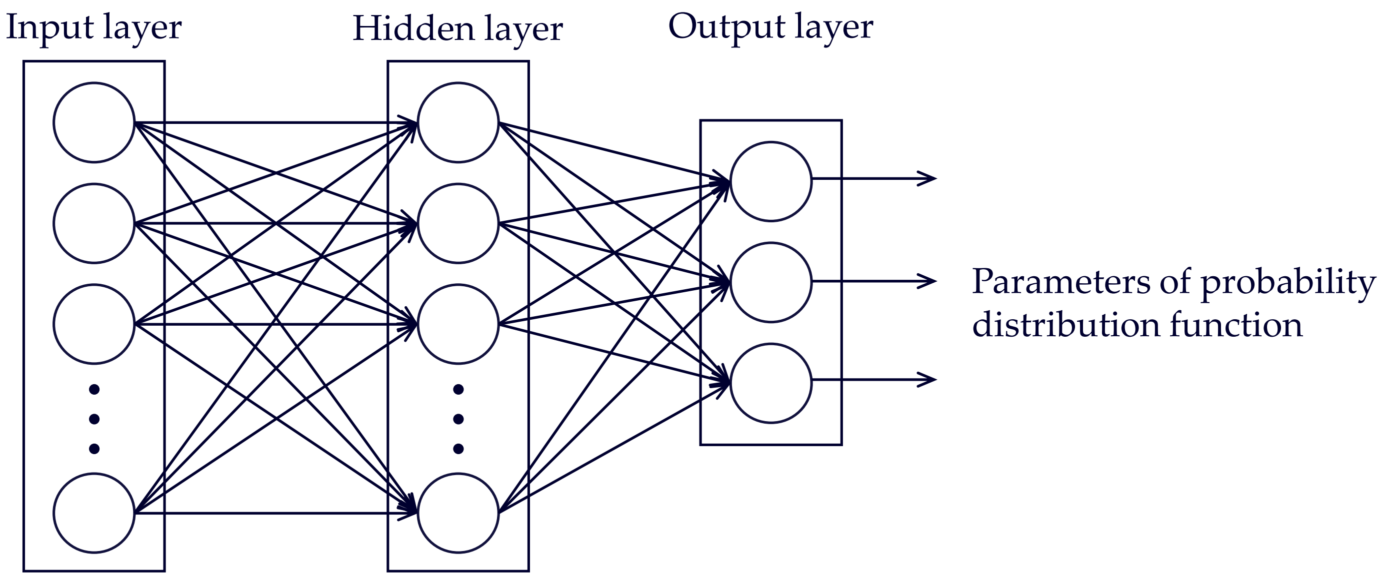
1. Background & Objectives
2. Methodology
3. Application
4. Results
5. Conclusions & Future studies

2. Methodology

- Quantile regression technique (QRT-ANN)
  - Estimates quantiles directly for various return periods
  - Output layer of ANN model is quantiles at each rainfall gauging site



- Parameter regression technique (PRT-ANN)
  - Estimates the parameters of probability distribution function
  - Quantiles are then estimated with the output of model (parameters)



2. Methodology

- Extreme gradient boosting algorithm
  - Proposed by Chen and Guestrin (2016)
  - Winning solution for classification and regression problems
  - Extreme gradient boosting refers to a class of ensemble machine learning algorithms
  - Uses weighted boosting algorithm to make prediction
  - Many hyperparameters to be optimized

2. Methodology

- Evaluation tools
  - Relative root mean square error (RRMSE)
    - $RRMSE_i(F) = \left[\frac{1}{N_{sim}}\sum_{m=1}^{N_{sim}}\left\{\frac{\widehat{Q_i^{[m]}(F)}-Q_i(F)}{Q_i(F)}\right\}^2\right]^{1/2}$
  - Root mean square error (RMSE)
    - $RMSE_i(F) = \left[\frac{1}{N_{sim}}\sum_{m=1}^{N_{sim}}\left\{\widehat{Q_i^{[m]}(F)} - Q_i(F)\right\}^2\right]^{1/2}$

where $\widehat{Q_i^{[m]}(F)}$and $Q_i(F)$ are quantiles for a given non-exceedance probability F in the m^{th} simulation and observation at site i

2. Methodology

- Evaluation tools
  - Bias
    - $b_i(F) = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} \widehat{Q_i^{[m]}}(F) - Q_i(F)$
  - Relative bias
    - $Rb_i(F) = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} \frac{\widehat{Q_i^{[m]}}(F) - Q_i(F)}{Q_i(F)}$

1. Background & Objectives

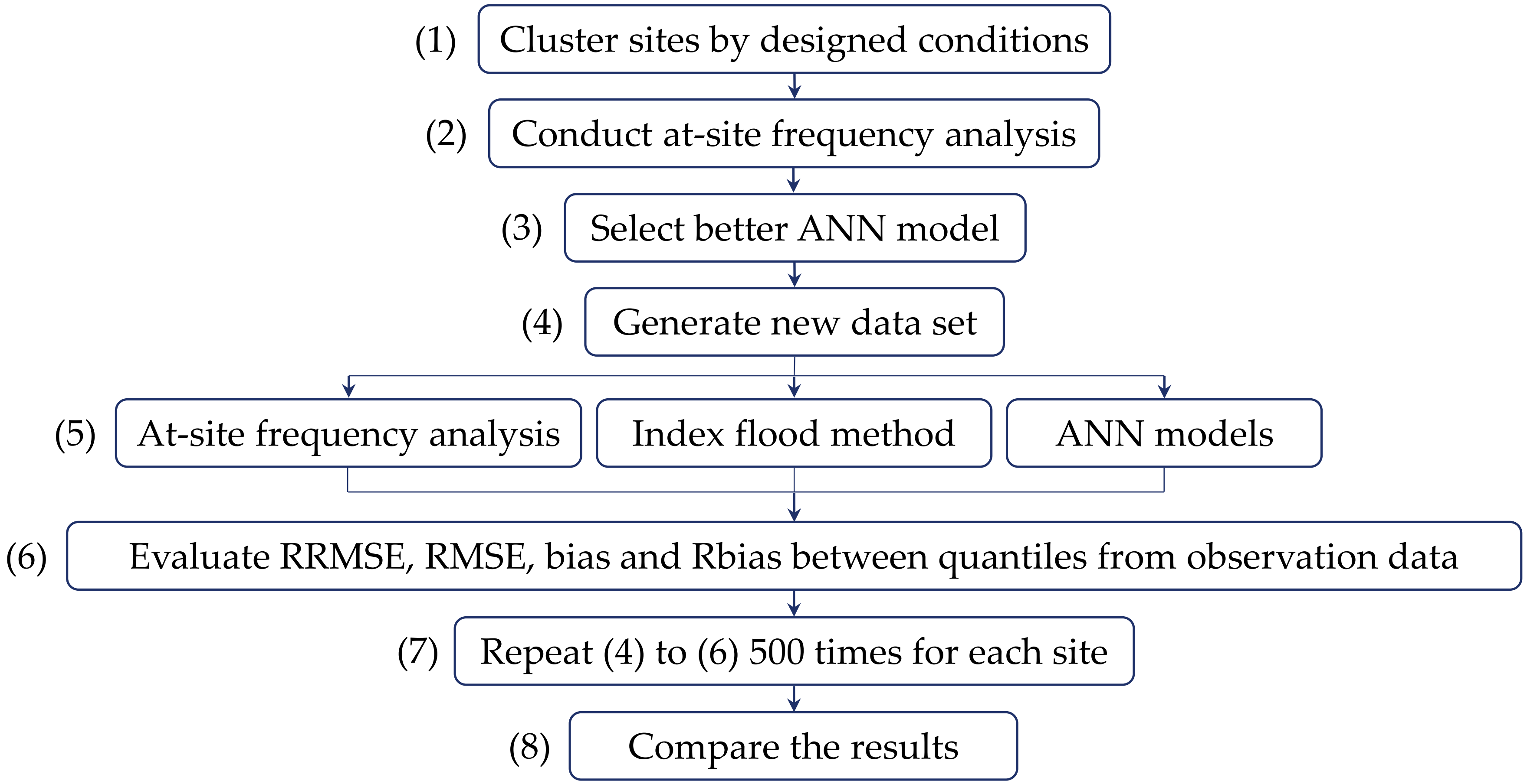
2. Methodology

3. Application

4. Results

5. Conclusions & Future studies

- Flow chart


(1) Cluster sites by designed conditions
(2) Conduct at-site frequency analysis
(3) Select better ANN model
(4) Generate new data set
(5) At-site frequency analysis
Index flood method
ANN models
(6) Evaluate RRMSE, RMSE, bias and Rbias between quantiles from observation data
(7) Repeat (4) to (6) 500 times for each site
(8) Compare the results

- Rainfall gauging site
  - Where daily annual maximum precipitation data fits into generalize extreme value(GEV) distribution examined by at-site frequency analysis
    - shape parameter $\beta < 0, \quad x_0 + {}^{\alpha}/_{\beta} \leq x < \infty$
  - Also has over 30 years of record length
  - Select 113 sites among 615 sites in South Korea

- Input variables
  - Variables that are suitable for Monte Carlo simulation
  - Topographical and hydrological data

Variables	Description
LAT	Latitude (˚)
LONG	Longitude (˚)
ALT	Altitude (m)
AM data	daily annual maximum of recent 30 years

- Designed conditions
  - Factors that affect the accuracy of regional frequency analysis
  - Heterogeneity measure (H)
  - Number of sites in a region (N_{site})

- 9 different Monte Carlo simulations
  - 3 cases in terms of H (H = 1, 2, 3)
  - 3 cases in terms of N_{site} (N_{site} = 5, 10, 15)

3. Application

- Region 1 ($H = 1, N_{site} = 5$)
- Region 2 ($H = 2, N_{site} = 5$)
- Region 3 ($H = 3, N_{site} = 5$)
- Region 4 ($H = 1, N_{site} = 10$)
- Region 5 ($H = 2, N_{site} = 10$)
- Region 6 ($H = 3, N_{site} = 10$)
- Region 7 ($H = 1, N_{site} = 15$)
- Region 8 ($H = 2, N_{site} = 15$)
- Region 9 ($H = 3, N_{site} = 15$)

1. Background & Objectives

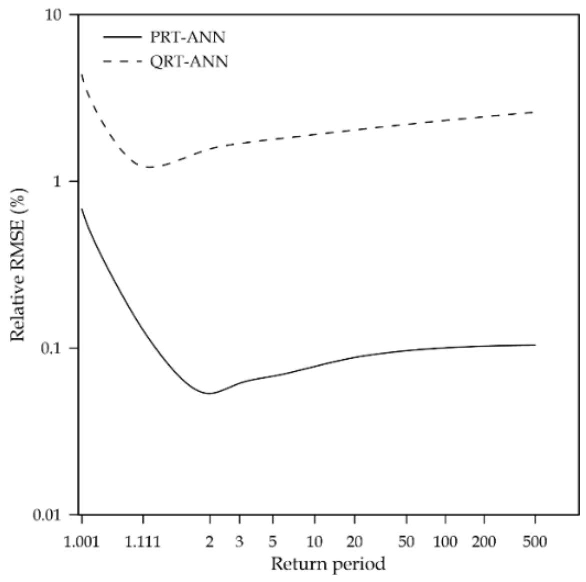
2. Methodology

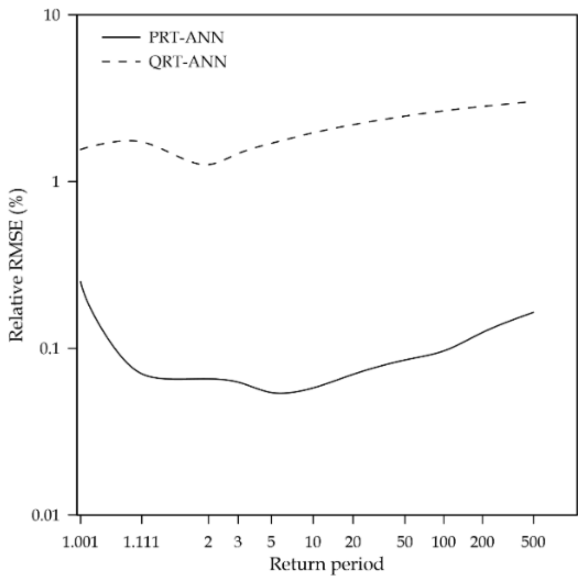
3. Application

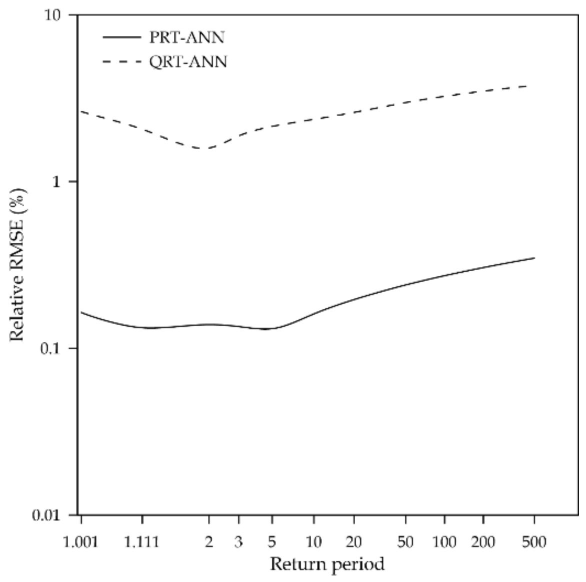
4. Results

5. Conclusions & Future studies

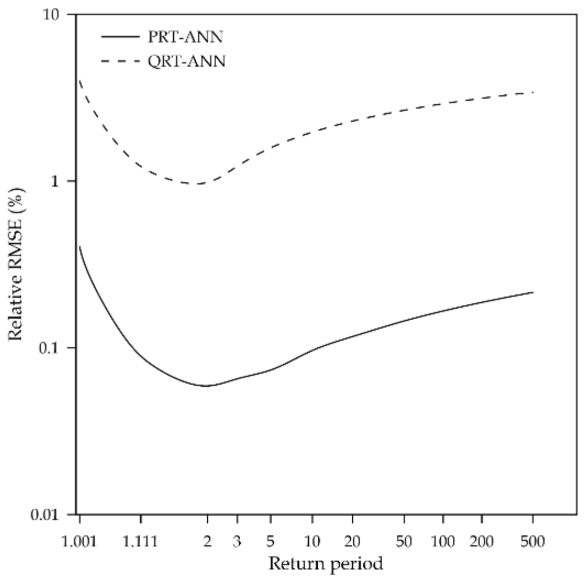
- PRT-ANN versus QRT-ANN
  - Region 1, 2, and 3

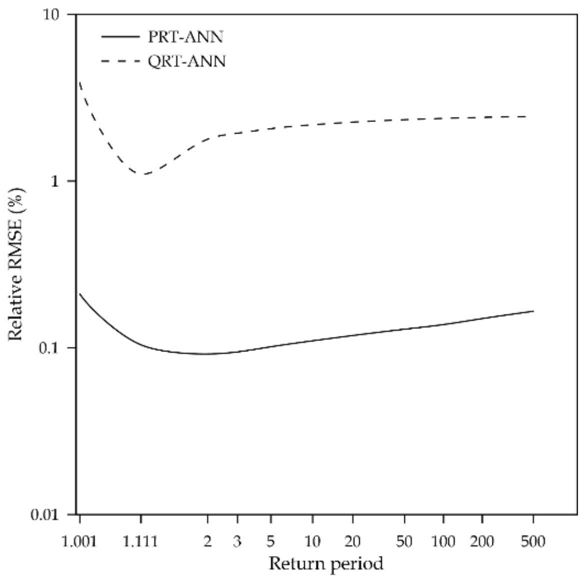


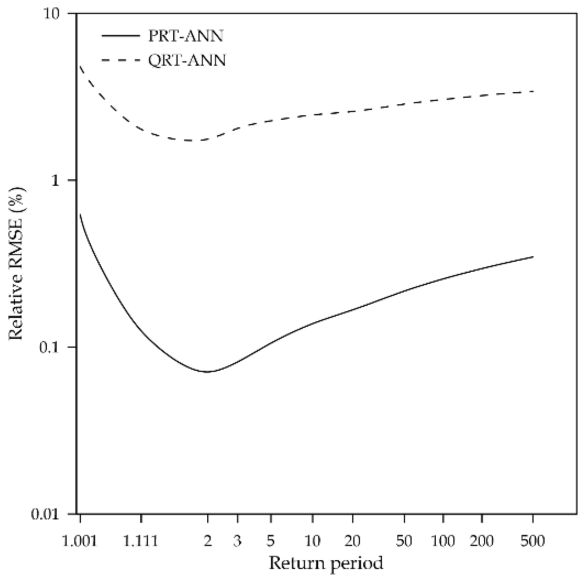




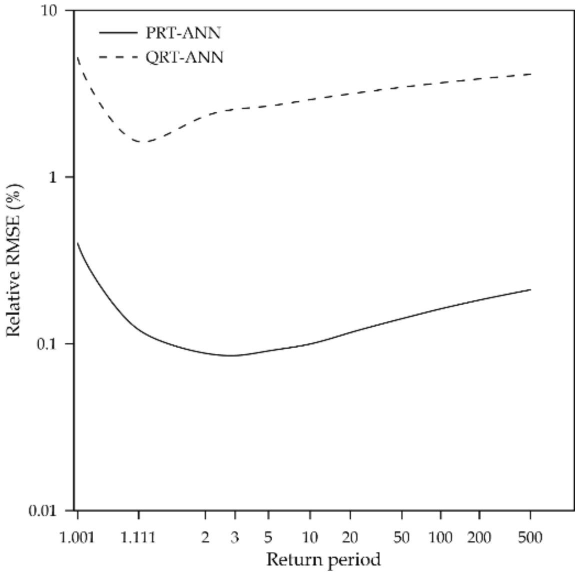
- PRT-ANN versus QRT-ANN
  - Region 4, 5, and 6

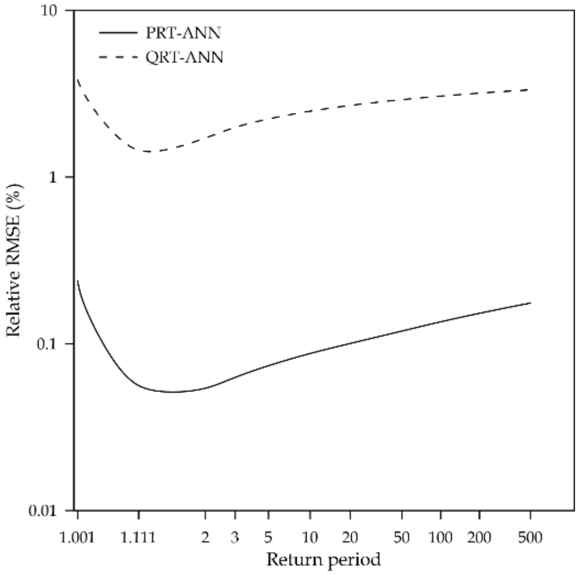


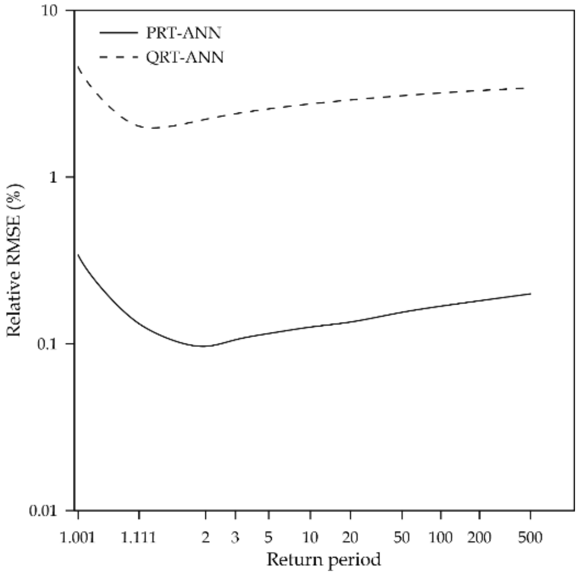




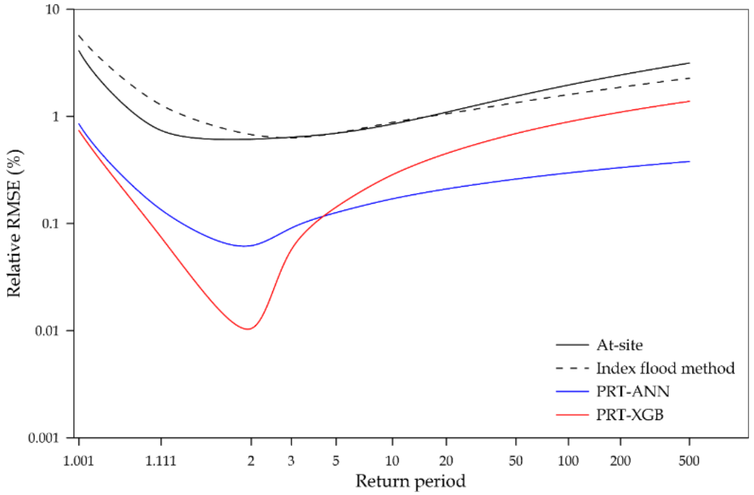
- PRT-ANN versus QRT-ANN
  - Region 7, 8, and 9

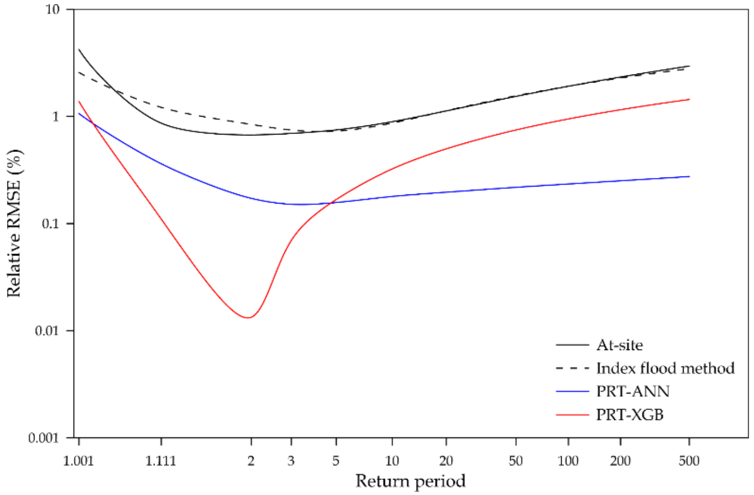


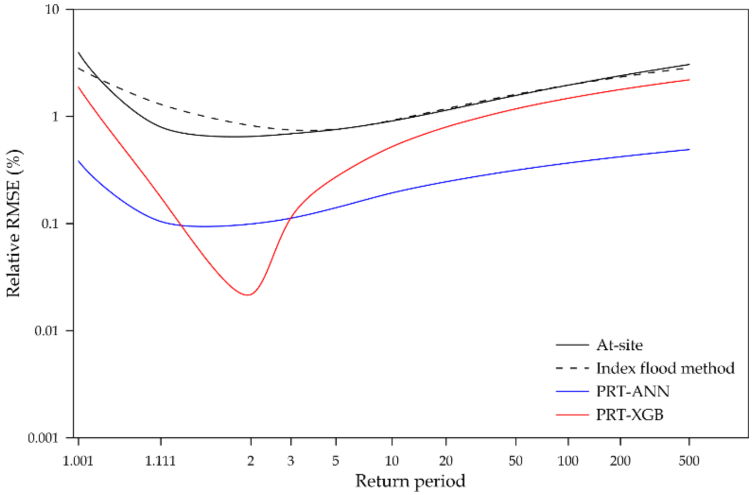




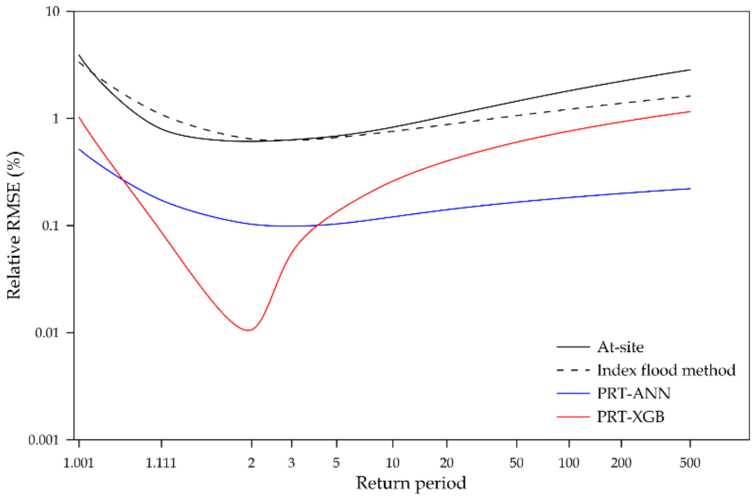
- Monte Carlo simulation 1, 2, and 3

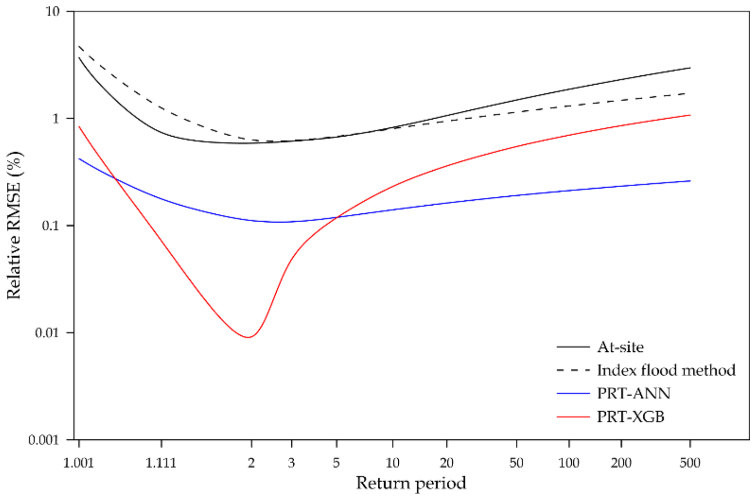


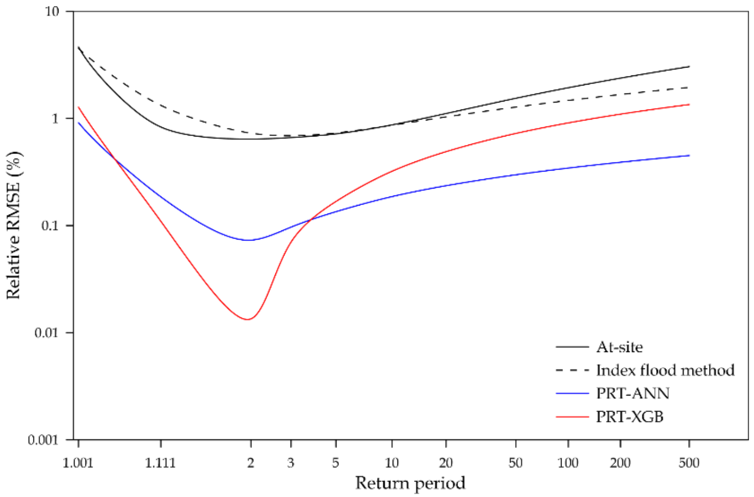




- Monte Carlo simulation 4, 5, and 6

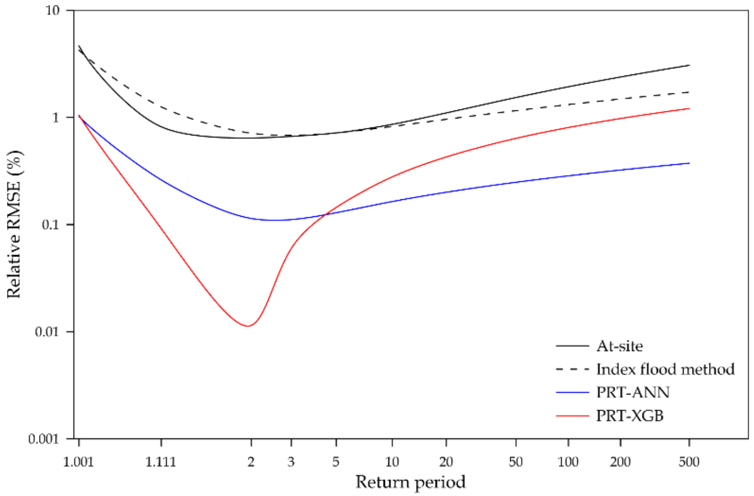


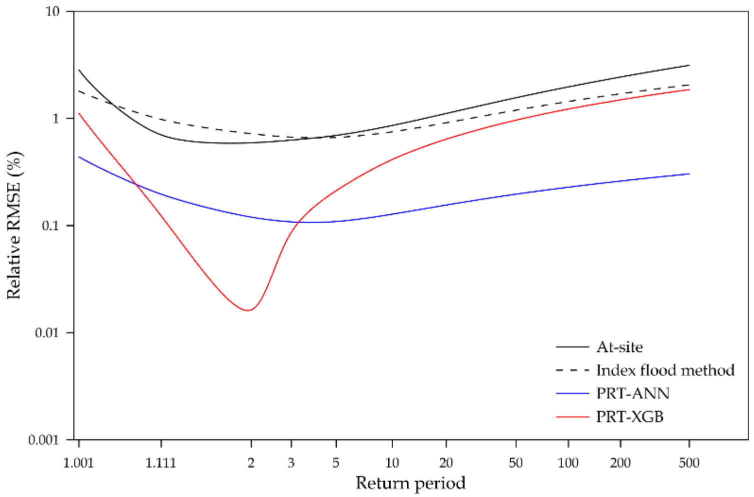


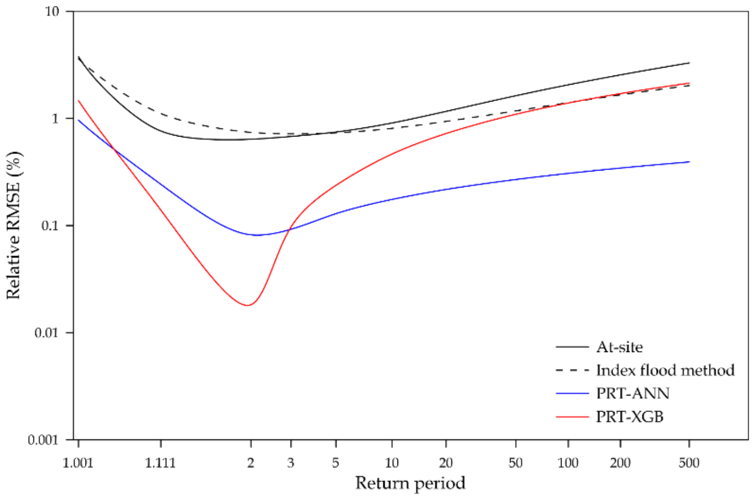


4. Results

- Monte Carlo simulation 7, 8, and 9







1. Background & Objectives

2. Methodology

3. Application

4. Results

5. Conclusions & Future studies

5. Conclusions & Future studies

- PRT-ANN was more accurate than QRT-ANN
  - QRT-ANN predicted 11 output variables while PRT-ANN predicted 3 output variables
  - PRT-ANN showed the less uncertainty in estimating quantiles than QRT-ANN
- Statistical method
  - For $N_{site} = 5$, performance difference between at-site frequency analysis and index flood method is getting close with H increases from 1 to 3

5. Conclusions & Future studies

- Statistical method
  - For $N_{site} = 10$ and 15, at-site frequency analysis shows better performance than index flood method at $T \leq 5$ years and index flood method shows better performance than at-site frequency analysis at $T > 5$ years regardless of H
- Data driven method
  - PRT-ANN shows the better performance than PRT-XGB for higher return period ($T \geq 5$ years)
  - Average performance of PRT-ANN is better than PRT-XGB

5. Conclusions & Future studies

- Statistical method versus data driven method
  - Both machine learning models show better performance than at-site frequency analysis and index flood method

5. Conclusions & Future studies

- Future studies
  - Frequency analysis of climate change scenario
  - Performance analysis of index flood method, population index flood, and ANN models with nonstationary rainfall data

Thank you