**Proposed QR-based approach**

**Quantile regression**

Estimating the conditional quantile of the response: 
$$Q_p(y|x) = x' \beta$$

Minimizing: 
$$\min \sum \rho_p(y_i - x_i' \beta)$$

$\rho_p(u) = \frac{p}{u} (u - p)$ if $u < 0$

$$\rho_p(u) = 0$$ if $u \geq 0$

**General procedure**

1. **Observed data**
   - $Y$: Hydrological variables
   - $X$: Physiographical variables

2. **Linear regression (LR)**
   - Estimated quantiles at ungaged site
   - $\hat{y} = \hat{b}' x$

3. **Quantile regression (QR)**
   - $\hat{y}_p = \hat{b}_p x$

**Quality assessment**

A natural analog of the RMSE in a cross-validation procedure is:
$$MPLF(p) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p} \rho_p(y_i - \hat{y}^{(p)}_i)$$;

$\rho_p(u) = \frac{p}{u} (u - p)$ if $u < 0$

$\rho_p(u) = 0$ if $u \geq 0$

**MPLF**: Mean Picewise Loss Function

**Study design**

- Apply and compare the considered models (QR and LR) using different criteria: the calibration and the application of both models are performed using the entire data.
- Take into account the at-site quantile estimation quality; modify the data used for the calibration step.
- Consider a more suitable case for which the LR performs well and the QR advantages are accounted for; the QR model built and assessed using the entire data / the LR model built using only sites with record length exceeding 30 years and evaluated using the entire data.
- Compare both models using the MPLF criterion; the concept of this criterion permits the model assessment using the entire data set.

### Introduction

Regression-based models are the most widely used tools for estimation purposes in regional frequency analysis (RFA). These latter are built using the estimated at-site quantiles.

**Problems of classical RFA** are mainly related to the use of the at-site estimated quantiles since the quality of these latter is related to the:

- data series record length, which lead to ignore a number of gauged sites with short records,
- model selection,
- parameter estimation...

**How to address this issue?**

Use a regression model based on the observed data instead of the estimated one and allowing accounting for the whole hydrological information in the region.

**Objective:**

Propose a new RFA framework based on quantile regression (QR) model that gives directly the conditional quantile and avoids performing an at-site FA at each gauged site.

### Data and models

**Classical RFA approach**

**Ordinary Linear Regression:**

Estimating the conditional mean of the response: 
$$Y = AX + B + \varepsilon$$

$\varepsilon$ is the model error

Minimizing:
$$\min \sum \varepsilon_i^2$$

**General procedure**

1. **Observed data**
   - $Y$: Hydrological variables
   - $X$: Physiographical variables

2. **Linear regression (LR)**
   - Estimated quantiles at ungaged site
   - $\hat{y} = \hat{b}' x$

3. **Quantile regression (QR)**
   - $\hat{y}_p = \hat{b}_p x$

**Quality assessment**

Use of the RMSE in a cross-validation procedure:
$$RMSE = \sqrt{\frac{1}{n} \sum \varepsilon_i^2}$$

**Case study**

151 hydrometric stations in Quebec, CA;
5 physio-meteorological variables;
3 hydrological variables: $Q_{s10}$, $Q_{s50}$ and $Q_{s100}$;
Historical record of annual maximum flow between 1900 and 2002.

### Results

**Modelling results of LR and QR approaches in terms of BIAS and RMSE**

- Both models are calibrated and evaluated over the entire data set. Points plotted in deep dark designate sites with long records.
- The RMSE and BIAS are insensitive to the data series record lengths at each site, i.e. both short and long records are weighted identically.

RMSE of the regional estimators of $Q_{s10}$ (a) and $Q_{s100}$ (b) according to the length of regional data series. Both models are calibrated using sites with record length exceeding 7 years. Validation is done using the whole data set.

Scatter plots of at-site and regional estimated quantiles using the LR model (first column) and the QR model (the second column) for quantiles $Q_{s10}$, $Q_{s50}$ and $Q_{s100}$.

### Conclusions

Consider observed data directly in the RFA instead of estimated at-site quantiles using the QR model.

Evaluate the estimation performance of the two regional models (LR and QR) through an objective criterion.

The proposed approach is a promising method for the estimation and evaluation of flood quantiles at-sites with short to medium length records.