Statistical ...logy workshop
26 & 27 September 20... Quebec - Canada

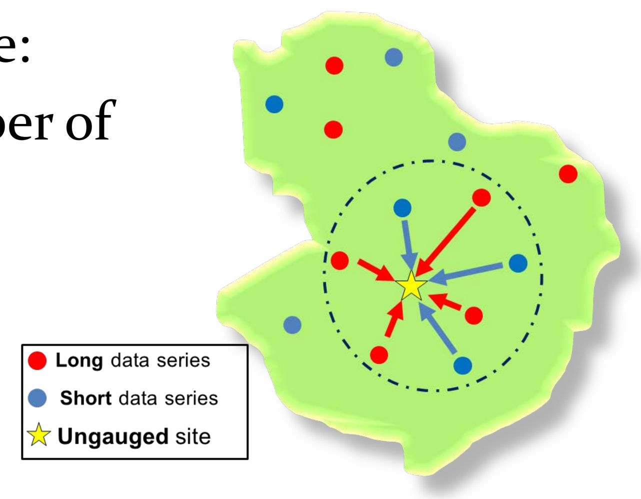
On the use of at-site estimated quantiles in regional frequency analysis

Introduction

Regression-based models are the most widely used tools for estimation purposes in regional frequency analysis (RFA). These latter are build 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

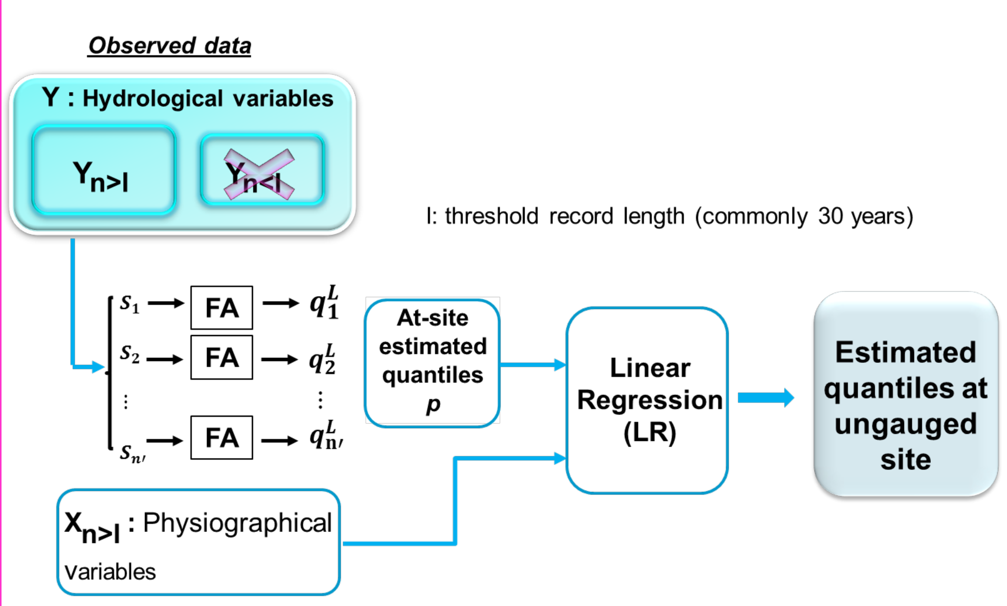
- **Ordinary Linear Regression:**

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

ε is the model error

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

- **General procedure**



- **Quality assessment**

Use of the RMSE in a cross-validation procedure:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(q_L - \hat{q}_i^R\right)^2}$$

Proposed QR-based approach

- **Quantile regression**

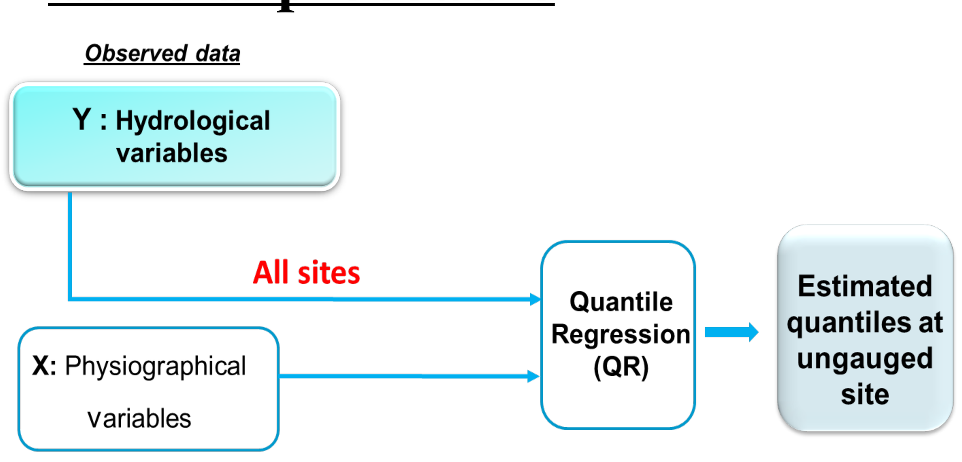
Estimating the conditional quantile of the response : $Q_p(y\,|\,\mathbf{x}) = \mathbf{x}^T \mathbf{b}_p$

Minimizing: $min_\beta \sum_{i=1}^{n} \rho_p\left(y_i - x_i^T \beta\right)$

ρ is the check function defined as:

$$\rho_p(\alpha) = \begin{cases} \alpha(p-1) & if\ \alpha < 0 \\ \alpha p & if\ \alpha \geq 0 \end{cases}$$

- **General procedure**



- **Quality assessment**

A natural analog of the RMSE in a cross-validation procedure is:

$$MPLF(p) = \frac{10^3}{n}\sum_{i=1}^{N}\sum_{j=1}^{n_i}\rho_p\left(y_{ij} - \hat{q}_{ip}^R\right) \quad ; \quad p \in (0,1)$$

MPLF: Mean Picewise Loss Function

- **Study design**

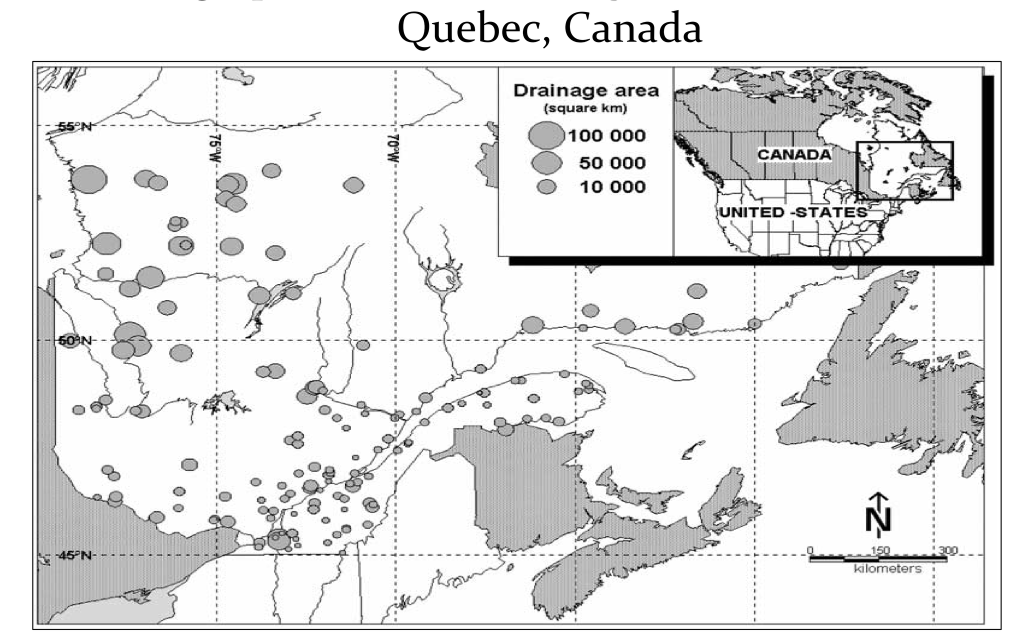
The application of both classical and proposed approaches is performed in several ways:
- Apply and compare the considered models (QR and LR) using different criteria; the calibration and the application of both models are performed using the entire data.
- Take into account the at-site quantile estimation 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.

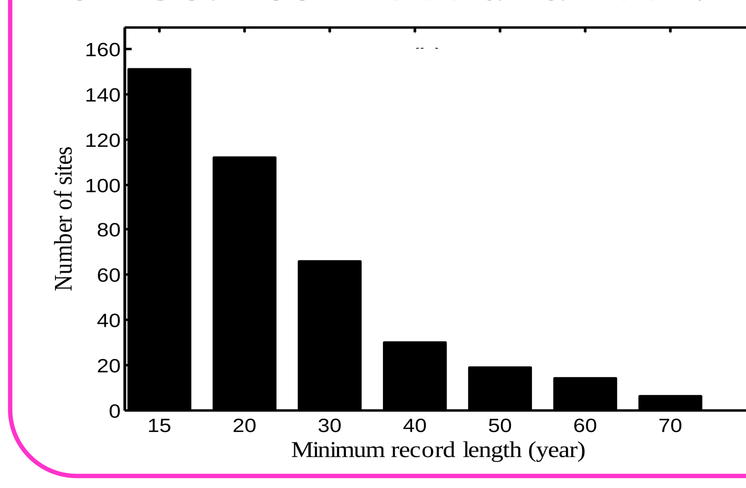
Data and models

- **Case study**

151 hydrometric stations in Quebec, CA;
5 physio-meteorological variables;
3 hydrological variables: Qs10, Qs50 and Qs100;
Historical record of annual maximum flow between 1900 and 2002.

Geographical location of hydrometric stations, Quebec, Canada



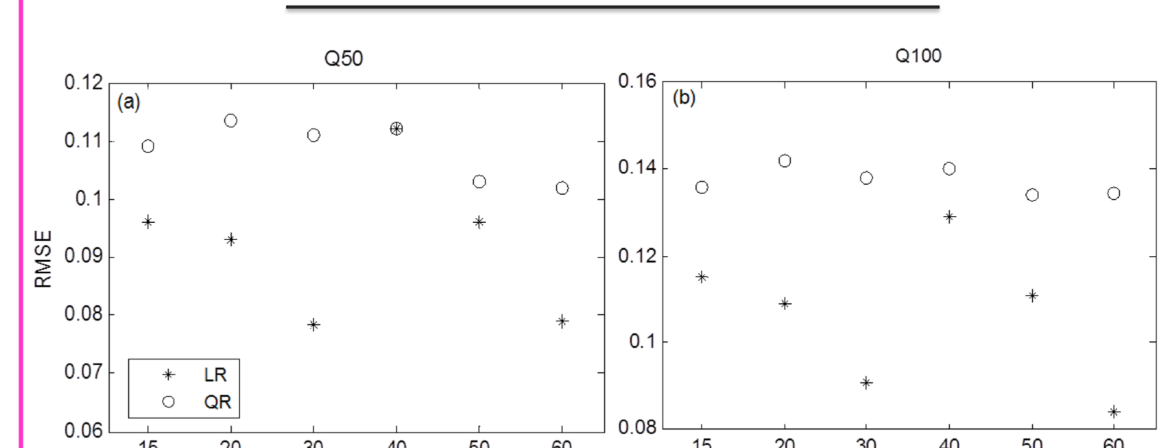


Bar plot of number of stations. Classes are defined to indicate the number of stations with records length exceeding a given minimum.

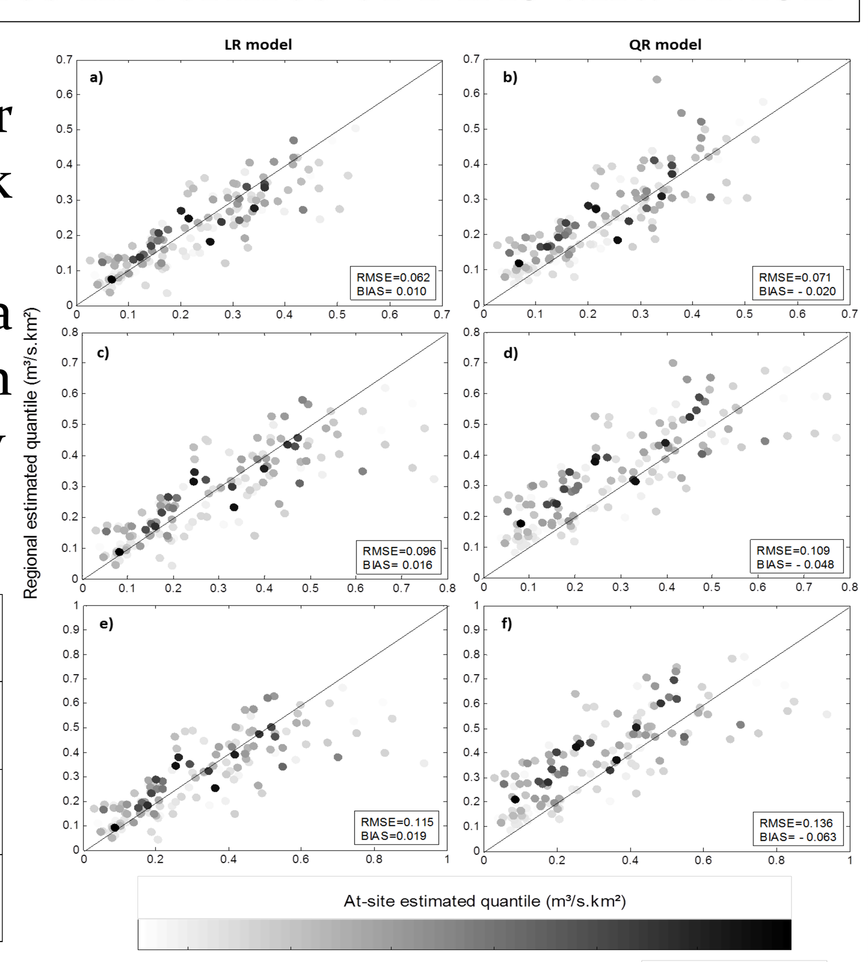
Results

Modelling results of LR and QR approaches in termes 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

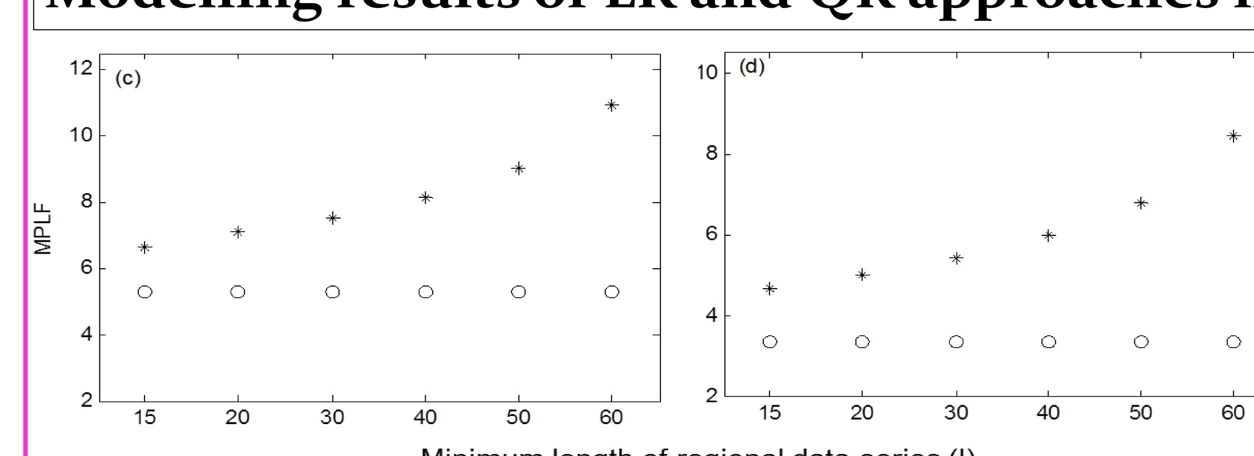


RMSE of the regional estimators of Q_{S50} (a) and Q_{S100} (b) according to the length of regional data series. Both models are calibrated using sites with record length exceeding *l* years. Validation is done using the whole data set.



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}

Modelling results of LR and QR approaches in termes of BIAS and RMSE



- The QR model shows a better performance than the classical model
- For higher record lengths (i.e fewer sites for the calibration step) the LR performance decreases.

MPLF of the regional estimators of Q_{S50} (c) and Q_{S100} (d) according to the length of regional data series. QR was calibrated using the whole data set; Validation of both models is done using the whole data set.

MPLF values associated to QR and LR approaches

	Q_{S10}		Q_{S50}		Q_{S100}	
	LR	QR	LR	QR	LR	QR
MPLF ($m^3/s.km^2$)	16.07	**15.43**	6.62	**5.30**	4.65	**3.43**

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