STAHY 2014
Statistical Hydrology Workshop

INRS Université d'avant-garde

Masdar INSTITUTE

Non-linearity in regional frequency analysis

D. Ouali[1], F. Chebana[1], T.B.M.J. Ouarda[1,2]

1: Institut National de la Recherche Scientifique, Centre Eau Terre et Environnement, Canada.
2: Institute Centre for Water Advanced Technology and Environmental Research, Abu Dhabi, UAE

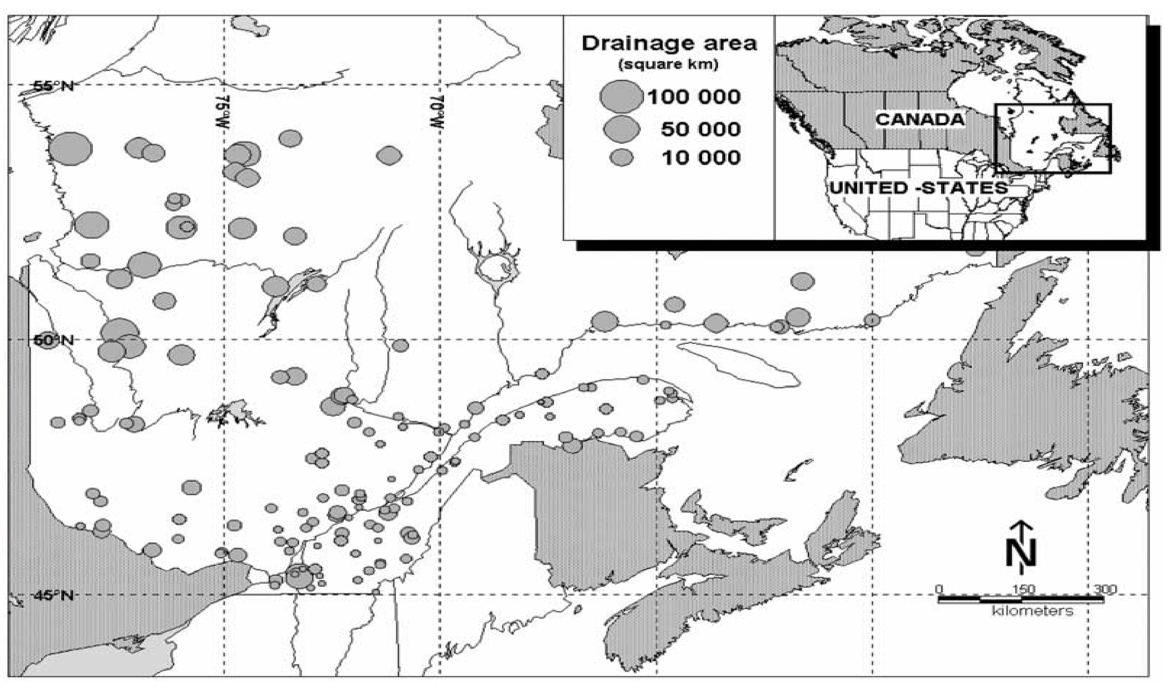
1) Introduction:

- Regional frequency analysis (RFA) is a technique that aim to estimate extreme hydrological events at ungagged sites. The delineation of homogeneous regions (DHR) and the regional estimation (RE) are the two main steps of RFA. A number of methods are commonly used for each step mainly the canonical correlation analysis (CCA) for DHR and the linear regression for the RE.
- Hydrological processes are very complex phenomena which require developed tools to be described and modeled. Despite this high complexity, non linear (NL) approaches have been considered only in the RE and have not been used yet in the DHR step. The considered approach is based on a NL-CCA using neural networks (CCA-NN), coupled to a log-linear regression model for quantile estimation.

2) Objective

: To deal with the issue of non-linearity in RFA by introducing NL-CCA in the DHR strep in order to improve its performance and representativeness.

3) Data:

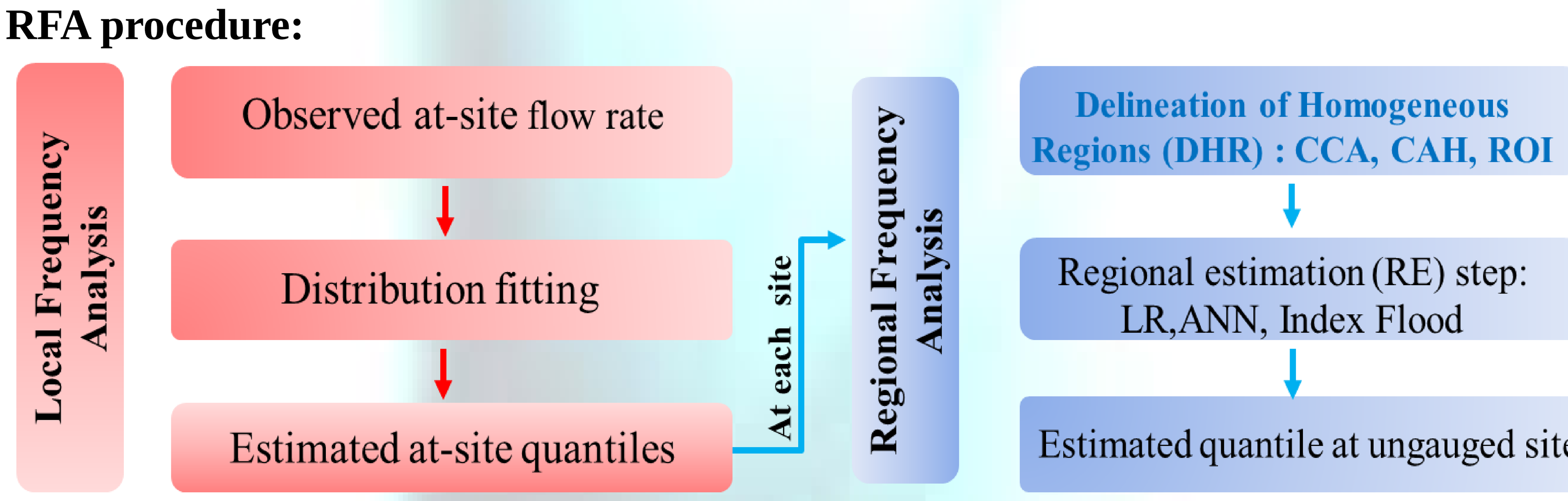


Geographical location of hydrometric stations, Quebec, Canada

Tab.1. Hydrological and physiographical variables- Quebec, Canada

Physiographical variables	Hydrological variables
The mean basin slope (PMBV)	
The basin area (BV)	
The proportion of the basin area covered with lakes (PLAC)	At-site flood quantiles standardized by basin area: QS10, QS100
The annual mean total precipitation (PTMA)	
The annual mean degree-days (DJBZ)	

- **RFA procedure:**



4) Methodology:

- **Canonical Correlation Analysis:**

Let $X_1, X_2, ..., X_q$ and $Y_1, Y_2, ..., Y_q$ denote respectively physiographical and hydrological variables, then canonical variables U_i and V_i are obtained thru linear combinations of original variables:

$$U_i = a_{i1}X_1 + a_{i2}X_2 + ... + a_{iq}X_q \qquad V_i = b_{i1}Y_1 + b_{i2}Y_2 + ...b_{ir}Y_r$$

where i=1,..p with p=min(r,q).
The canonical space is built under constraints of unit variance and maximum correlation between pairs of canonical variables.

- **Nonlinear CCA using a Neural Network (NN) approach (CCA-NN)**:

CCA-NN is an artificial neural network based method. It consists on establishing non-linear combinations between groups of variables (X and Y) and the canonical variables (U and V) via a transfer function *f*. Canonical variables U and V are determined from a linear combination of respective neurons h(x) and h(y):

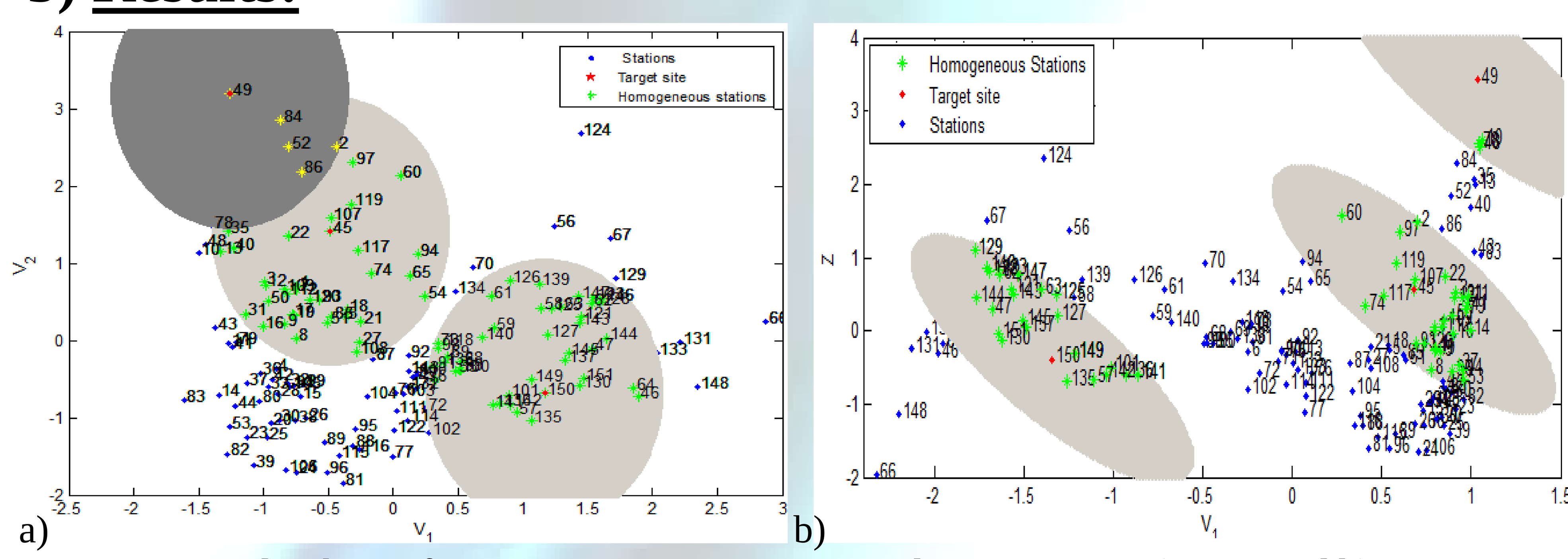
$$U = w^{(x)}h^{(x)} + \overline{b}^{(x)} \qquad V = w^{(y)}h^{(y)} + \overline{b}^{(y)}$$

- **Regional estimation**

The multivariate log-linear regression model is adopted to estimate quantiles at ungauged sites. The relationship between flood quantiles (Y) and the physiographical characteristics (X) is described by a power product model. With a log-transformation, the log-linear model is obtained :

$$\log(Y) = \beta \log(X) + \varepsilon$$

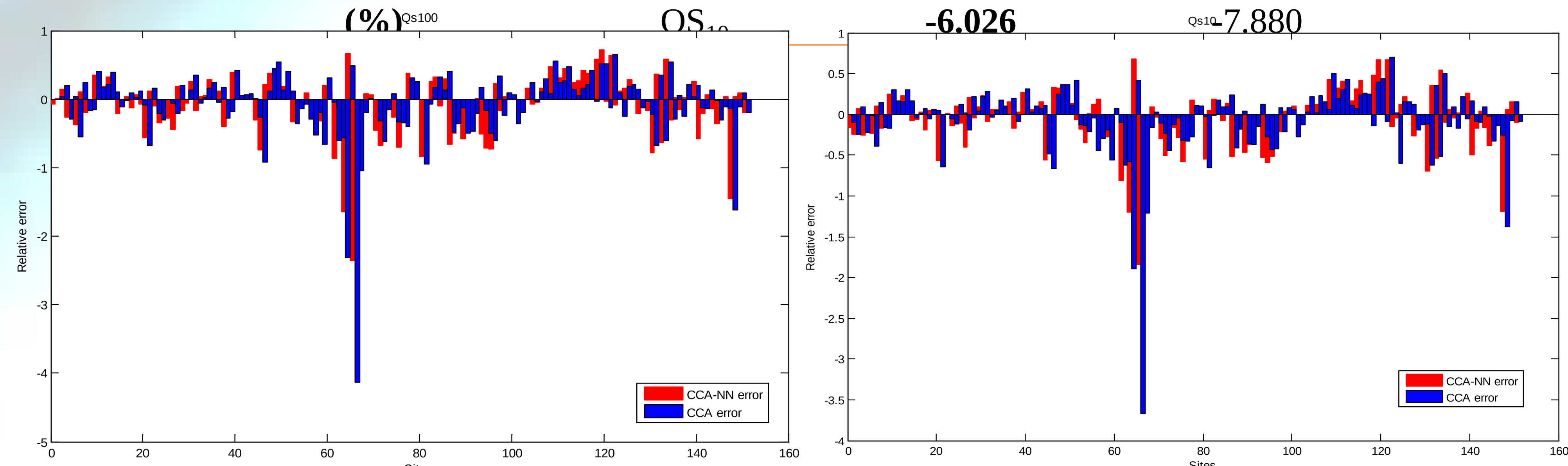
5) Results:



DHR results shown for stations 030340, 030420 and 02717 using: a) CCA and b) CCA-NN approaches, n=45, 49 and 150 respectively.

Tab.2. Jackknife validation results-Quebec

	Variables	CCA-NN& LR	CCA & LR
NASH	QS_{100}	**0.710**	0.700
	QS_{10}	**0.793**	0.790
RMSE	QS_{100}	**0.107**	0.109
(m³/s.km²)	QS_{10}	**0.058**	**0.057**
RMSEr	QS_{100}	**41.400**	51.030
(%)	QS_{10}	**33.903**	44.870
BIAS	QS_{100}	**0.010**	0.017
(m³/s.km²)	QS_{10}	**0.002**	0.005
BIASr	QS_{100}	**-7.747**	-8.390
(%)	QS_{10}	**-6.026**	-7.880



Estimation error resulting from the CCA & LR and CCA-NN& LR models

6) Conclusions:

- The CCA-NN can be adopted to represent the non-linear behavior of hydrological process
- It provides a more accurate and flexible delineation of homogeneous neighborhoods leading to a better regional estimation.
- Using two other databases, namely Arkansas and Texas, the proposed approach outperformed the linear approach which confirm its superiority and robustness.

References:

Hsieh, W. W. (2000). "Nonlinear canonical correlation analysis by neural networks." Neural Networks 13: 1095 -1105.
Hsieh, W. W. (2001). "Nonlinear Canonical Correlation Analysis of the Tropical Pacific Climate Variability Using a Neural Network Approach." Journal of climate_14: 2528-2539.
Ouarda, T. B. M. J. (2013). "Hydrological Frequency Analysis, Regional." Encyclopedia of Environmetrics: DOI:10.1002/9780470057339. vnn9780470057043.
Cannon, A. J. (2008). Multivariate statistical models for seasonal climate prediction and climate downscaling. Atmospheric Science, British Columbia. PhD: 141.