Non-linearity in regional frequency analysis

D. Ouali1, F. Chebana1, T.B.M.J. Ouarda1,2

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 step in order to improve its performance and representativeness.

3) Data:
- RFA procedure:
  - Observed at-site flow rate
  - Distribution fitting
  - Estimated at-site quantiles

4) Methodology:
- Canonical Correlation Analysis:
  Let $X_1, X_2, ..., X_r$ 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:
  $$U_i = a_{i1}X_1 + a_{i2}X_2 + ... + a_{ir}X_r$$
  $$V_i = b_{i1}Y_1 + b_{i2}Y_2 + ... + b_{iq}Y_q$$
  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 $h(y)$:
  $$U = w^{(x)}h^{(x)} + b^{(x)}$$
  $$V = w^{(y)}h^{(y)} + b^{(y)}$$
- Regional estimation
  The multivariate log-linear regression model is adopted to estimate quantiles at ungagged 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:

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: