Improving prediction at ungauged basins with Spatial Copula Framework

INRS Université d'avant-garde

CRM-CANSSI Workshop 2014: New horizons in copula modeling

M. Durocher(1), F. Chebana(1) and T.B.M.J. Ouarda(2)
(1) Institut National de Recherche Scientifique,
490 rue de la Couronne, Québec, Canada (martin.durocher@ete.inrs.ca)
(2) Masdar Institue of Science and Technology
P.O. Box 54224, Abu Dhabi, UAE

1. Introduction

Context

- When dealing with natural hazards, proper estimation of the risk of occurrence is crucial for balancing the safety and the design cost of human settlements.
- A risk is usually quantified as the quantile of an statistical distribution and its severity is characterized by a **return period**, which corresponds to the time separating two events of specific magnitude.
- Collecting valuable observations for studying **annual flood peaks** of rivers discharge requires times and resources. Consequently, to respond to the needs of information at new locations, **Prediction at Ungauged Basins** (PUB) are required.
- The physiographical properties of a basin determines its run-off and two contiguous basins may possess very different physiographical properties. Consequently, contiguous basins may present different hydrological behaviours that should be account in PUB.
- PUB can be carried out by the same methodology initially developed for **geostatistics**, such as kriging, but where the dependance structure is determined instead by the **physiographical proximity** defines as the distance between basin characteristics.

Problematic

- Usually, flood quantiles share log-log relationships with basin characteristics. In traditional kriging context, the **transformation** necessary to recover the normality assumption **creates bias and leads to suboptimal predictions**.
- Another limitation of the traditional kriging techniques is their incapacity to account for **heteroscedasticity**.
- The problem with traditional kriging techniques can be resolved by **Spatial Copula**, an extension of traditional geostatistical framework where the spatial dependance is characterized by a copula.

4. Physiographical space

- At-site analysis provide flood quantile estimates Z_i at gauged basins $i = 1, \ldots, n$. PUB is used to transfer this information at ungauged basins.
- Typically, spatial methods in PUB consider meaningful basin characteristics that characterizes the physiographical proximity. In practice, the basin characteristics are usually correlated, hence a **physiographical space** of lower dimension is built (e.g. $r = 2$) from multivariate techniques.
- The basin characteristics X_i becomes coordinates in the physiographical space

$$S_i = AX_i$$

where A is a transition matrix.

- **Canonical correlation analysis** (CCA) has been shown to provide more appropriate physiographical spaces than principal component analysis [3].
- Let Y be r.v of hydrological variables at a gauged site with basin characteristics X, CCA provides canonical pairs

$$\mathbf{s}_k = \mathbf{a}_k X \quad \text{and} \quad \mathbf{u}_k = \mathbf{b}_k Y$$

that sequentially optimizes $\mathrm{cor}(\mathbf{s_k}, \mathbf{u}_k)$. Therefore, $A' = (\mathbf{a}_1, \ldots, \mathbf{a}_r)$ also implies hydrological proximity.

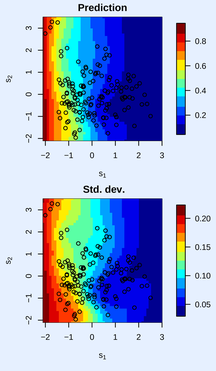

Fig 2 : Predictions in physiographical space

References

[1] Bárdossy, A. (2006) Copula-Based Geostatistical Models for Groundwater Quality Parameters. Water ressour res, 42(11).

[2] Bárdossy, A., and Li, J. (2008) Geostatistical Interpolation Using Copulas. Water ressour res, 44(7).

[3] Chokmani K, Ouarda TBMJ (2004) Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. Water ressour res. 40(12).

2. Spatial copula

- A multivarite distribution G can be expressed as

$$G(\mathbf{x}) = C\left[(F_1(x_1), \ldots, F_n(x_n)\right]$$

where $\{F_i\}_{i=1}^n$ are margins for $\mathbf{x}' = (x_1, \ldots, x_n)$ and C is a copula

- Spatial copula must allows for **strong dependance**

$$C_h \to M^n \quad \text{when} \quad h \to 0$$

where M^n is the copula upper bound and perfect **independence**

$$C_h \to \Pi^n \quad \text{when} \quad h \to \infty$$

- With copulas, margins are treated separately from the dependence. Hence, a model includes 2 set of parameters : the **marginal part** η and the **copula part** θ.
- Estimation can be performed by maximum likelihood or alternatively by optimizing a pairwise likelihood function

$$L(\mathbf{z} \mid \eta, \theta) = \prod_{i<j} f(z_i, z_j \mid \eta, \theta)$$

where $\mathbf{z}' = (z_1, \ldots, z_n)$ are spatial observation and f is the bivariate density of two sites i and j.

- For known parameters $(\hat{\eta}, \hat{\theta})$, **the plug-in predictive distribution** (PPD) at ungauged location is the product of the marginal density and the conditional copula [2] :

$$p(z \mid \mathbf{z}, \theta) = f_{\hat{\eta}}(z) \times c_{\hat{\theta}}\left[F_{\hat{\eta}}^{-1}(z) \mid \mathbf{w}\right]$$

where $\mathbf{w}' = (w_1, \ldots, w_n)$ and $w_i = F_{\hat{\eta}}^{-1}(z_i)$

- Predictors can be calculated from the mean or the median of the PPD. For instance, the median is the quantity $F_{\hat{\eta}}^{-1}(w^*)$ for which

$$1/2 = \int_0^{w^*} c_{\hat{\theta}}(u \mid \mathbf{w}) du$$

5. Model

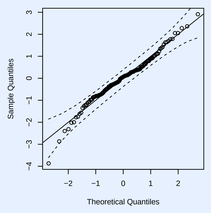

Fig. 3 : Normalized QQ-plot

Marginal part(η)

- **Regional distribution** of a flood quantiles Z_i is **log-normal**.

$$log(Z_i) \to N\left[\mu(S_i), \sigma^2(S_i)\right]$$

- By construction, **a linear trend** must be added to account for the trend resulting from the strong correlation of first canonical coordinates $S_{i,1}$ and the flood quantiles:

$$\mu(S_i) = \beta_{\mu,0} + \beta_{\mu,1}\, S_{i,1}$$
$$\sigma(S_i) = \beta_{\sigma,0} + \beta_{\sigma,1}\, S_{i,1}$$

Copula part (θ)

- The dependance is characterized by a **Gaussian copula with pairwise correlation**

$$\rho(S_i, S_j \mid \lambda, \tau) = (1 - \tau)\ exp\left[-3\frac{d(S_i, S_j)}{\lambda}\right]$$

where $\lambda > 0$ (practical range) controls the correlation as $d(S_i, S_j) \to \infty$ and τ is a local measurement error (nugget effect)

- A **correlogram** in respect of the physio. distance is estimated from binned observations.
- A **Goodness-of-fit** on bivariate copulas [1] validates the Gaussian copula for each bins (p-values > 20%).

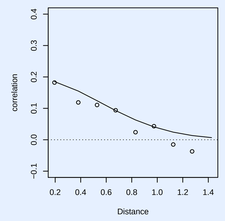

Fig 4 : Correlogram QS100

Acknowledgments

Financial support was provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

3. Case study

Hydrological data

- Response variable : Flood quantiles with 100 years return period
- Predictions from 5 basin characteristics
- To reduce the dominant scale effect of the drainage area, the flood quantiles are standardized.

At-site analysis

- 151 gauged site in Quebec, Canada
- Minimum record length: 15 years
- Rivers with natural flow regime
- Individual time series tested for independence, stationarity

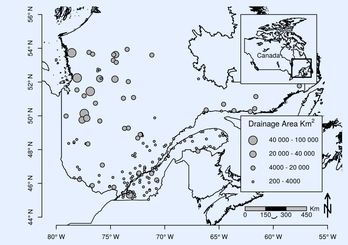

Fig 1 : Map of the stations

6. Results

- **Leave-one-out cross-validation** is used to assess the predictive performance of the model. In turn each gauged basin is considered as ungauged and a predicted value is obtained as the **median of the PPD**.
- The analysis of the residuals shows the presence of large relative discrepancies (Fig. 5-Left) corresponding to **problematic stations** previously identified for this database [3].
- Absolute residuals at logarithm scale (Fig. 5-Right) show **heteroscedasticity**

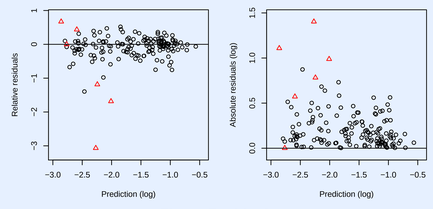

Fig 5 : Residuals

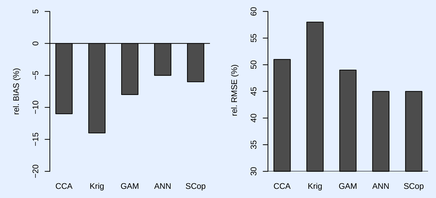

Fig 6 : Performance criteria

- **The prediction power** of the spatial copula approach (SCop) is compared to other methods on the same database:
  - Multiple regression with CCA-delineation (CCA)
  - Residual drift kriging (Krig)
  - Generalized additive model (GAM)
  - Artificial neural networks (ANN)
- In comparison with traditional kriging (Krig), the results of spatial copula (Scop) is associated with an important **reduction of the relative bias**.
- Overall, Scop and ANN have the **best** *rel.RMSE* from the methods considered here.

7. Conclusion

- The spatial copula framework has **competitive performance with the best methods**. In particular, it improves over traditional kriging.
- The important **relative bias** associated to simple transformation **is reduced** greatly with the spatial copula approach.
- The spatial copula framework offers a full probabilistic model that **account for heteroscedasticity**.
- The spatial copula framework appears more **appropriate in presence of problematic stations** in comparison with traditional kriging.