

Analyse des pluies extrêmes annuelles sur la région de Minas Gerais (Brésil): modèle de régionalisation TCEV

BRUNO RABELO VERSIANI

Departamento de Engenharia Hidráulica e Recursos Hídricos, Escola de Engenharia da Universidade Federal de Minas Gerais, Av. do Contorno, 842, 30110-060, Belo Horizonte, Brésil

e-mail: versiani@ehr.ufmg.br

EBER JOSÉ DE ANDRADE PINTO

CPRM, Serviço Geológico do Brasil, Av. Brasil, 1731, 30140-002, Belo Horizonte, Brésil

PHILIPPE BOIS

Laboratoire d'Etude des Transferts en Hydrologie et Environnement (LTHE), Domaine Universitaire, BP 53, F-38041 Grenoble Cedex, France

Résumé Cet article présente la méthode de régionalisation des pluies maximales basée sur la fonction de distribution TCEV (Two Component Extreme Value), proposée par Rossi *et al.* (1984) et son application à des séries de pluies journalières maximales annuelles dans la région de Minas Gerais (Brésil). Ce modèle statistique est composé du produit de deux exponentielles, chacune représentant un processus de Poisson: le premier correspond aux pluies maximales générées plus fréquemment et l'autre aux pluies maximales plus rares, les horsains ("outliers"). Les aspects théoriques du modèle et de la méthode de régionalisation sont brièvement présentés. La région d'étude est un bassin de 50 000 km² situé dans une zone représentative du climat tropical de la région sud-est du Brésil. La procédure utilisée permet de faire une analyse des horsains et d'identifier deux régions homogènes.

INTRODUCTION

La fonction de distribution des Valeurs Extrêmes à deux Composantes (TCEV—Two Component Extreme Value) fait partie d'une famille de modèles statistiques qui cherche à décrire les valeurs extrêmes, par exemple les débits maxima annuels ou les pluies maxima annuelles. Le modèle, proposé par Rossi *et al.* (1984), a été étudié par plusieurs auteurs et a été appliqué en Italie et en Angleterre comme un outil pour l'étude de régionalisation des précipitations et de débits extrêmes.

D'après Rossi *et al.* (1984, 1986), le principe de base du modèle considère que les valeurs extrêmes (débits de pointe, précipitations maximales) ne résultent pas de la même fonction de distribution. Le modèle fait ressortir la présence des horsains (outliers), qui sont liés aux coefficients d'asymétrie élevés dans les échantillons de maxima.

LE MODÈLE TCEV

Pour la description statistique des événements maximaux, le processus composé de Poisson est considéré par les auteurs comme un modèle valable pour la série de valeurs

de K variables Z_{ij} ($j = 1, \dots, K$) indépendantes et identiquement distribuées, supérieures à un seuil de base z_0 , compte tenu que z_0 soit suffisamment grand.

Le modèle poissonien considère les valeurs maximales:

$$X = \max Z_{i,j}, \quad 1 \leq j \leq K \quad (1)$$

étant donné que la variable aléatoire K suit la distribution de Poisson. La fonction de distribution (FD) pour le processus X , comme étant les maxima peut être décrite par Todorovic & Zelenhasic (1970):

$$F_X(x) = P[X \leq x] = \exp\{-\Lambda[1 - F_Z(x)]\} \quad (2)$$

où $\Lambda = E\{K\}$ est le paramètre du processus de Poisson, $F(\cdot)$ est la FD commune pour tous les Z_{ij} .

Il est intéressant de noter que si Z est une variable aléatoire exponentielle, telle que:

$$F_Z(z) = 1 - \exp(-z/\theta) \quad z \geq 0 \quad (3)$$

avec $\theta = E\{Z\}$ d'après l'équation (2) nous avons:

$$F_X(x) = \exp[-\Lambda \exp(-x/\theta)] \quad (4)$$

L'équation (4) est la FD de valeurs extrêmes du type I, soit la FD de Gumbel, qui peut s'écrire:

$$F_X(x) = \exp\{-\exp[-(x - x_0)/\theta]\} \quad (5)$$

avec

$$x_0 = \theta \ln \Lambda \quad (6)$$

où θ est le paramètre d'échelle (= gradex), x_0 est le paramètre de location (= mode).

Le paramètre gradex est un paramètre régional qui est le point de départ pour la Méthode du Gradex, proposée par Guillot & Duband (1967) pour l'estimation des débits de pointe à partir des précipitations extrêmes.

Si on considère qu'il y a deux séquences indépendantes de variables indépendantes et identiquement distribués Z_{1ij} , ($j = 1, \dots, K_1$) et Z_{2ij} ($j = 1, \dots, K_2$), générées par deux types de processus de Poisson, avec les paramètres donnés par:

$$\Lambda_1 = E\{K_1\}, \quad \Lambda_2 = E\{K_2\} \quad \Lambda_1 > \Lambda_2$$

en définissant X_1 et X_2 comme les maxima annuels de Z_{1ij} et de Z_{2ij} , respectivement, et si Z_{1ij} et Z_{2ij} sont des variables aléatoires exponentielles avec des paramètres $\theta_1 = E\{Z_1\}$ et $\theta_2 = E\{Z_2\}$, la fonction de distribution qui combine les processus X_1 et X_2 peut s'écrire:

$$F_X(x) = \exp[-\Lambda_1 \exp(-x/\theta_1) - \Lambda_2 \exp(-x/\theta_2)] \quad x \geq 0 \quad (7)$$

L'équation (7) définit la fonction de distribution TCEV (Two Component Extreme Value). Elle est définie avec les conditions

$$x \geq 0, \quad \theta_2 \geq \theta_1 \geq 0, \quad \Lambda_1 \geq \Lambda_2 \geq 0$$

Les paramètres θ_1 et Λ_1 sont ceux qui décrivent les événements les plus fréquents (la série basique), et θ_2 et Λ_2 décrivent la distribution des événements les plus rares (la série des horsains). Il s'agit d'une FD à quatre paramètres (Λ_1 , Λ_2 , θ_1 , θ_2) qui est équivalente au produit de deux FD de Gumbel:

$$F_x(x) = \exp\{-\exp[-(x - x_{01})/\theta_1]\} \cdot \exp\{-\exp[-(x - x_{02})/\theta_2]\} \quad (8)$$

avec $x_{0j} = \theta_j \ln \Lambda_j$, ($j = 1, 2$).

Les paramètres θ_1 et θ_2 sont respectivement le gradex de la série basique et le gradex de la série horsain.

Une description détaillée des propriétés statistiques de la TCEV, avec la définition de ses moments et aussi des probabilités inhérentes à la composante des horsains est donnée par Beran *et al.* (1986).

APPLICATION DU MODELE TCEV AUX PRECIPITATIONS EXTREMES SUR LA REGION DU BASSIN DE ALTO SÃO FRANCISCO

La région d'étude se trouve dans le sud-est du Brésil et comprend le bassin de Alto São Francisco, caractéristique du climat tropical de l'Etat de Minas Gerais (Fig. 1). Cette région présente un réseau de 92 pluviomètres répartis sur 50 870 km². Les séries de pluies maxima annuelles disponibles, fournis par la CPRM—Serviço Geológico do Brasil, sont de bonne qualité, mais souvent courtes (57% inférieures à 20 ans).

Dans la première partie de cette étude, on a fait l'estimation des paramètres du modèle TCEV pour chaque station isolée et ensuite on a étudié la variabilité spatiale des paramètres dans la région. L'estimation des quatre paramètres peut être faite par la méthode du maximum de vraisemblance (ML), d'après Rossi *et al.* (1984). Cette

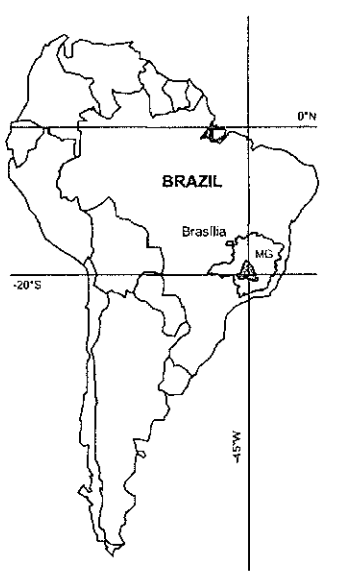


Fig. 1 Localisation du bassin du Alto São Francisco (Minas Gerais, Brésil).

Tableau 1 Caractéristiques statistiques et paramètres de la TCEV, stations isolées.

(1) Station	(2) iter	(3) n	(4) moy	(5) e. type	(6) min	(7) max	(8) g _n	(9) Λ ₁	(10) Λ ₂	(11) θ ₁	(12) θ ₂	(13) θ*	(14) Λ*
1844001	141	22	74.4	28.0	43.3	167.4	2.1303	444.91	1.6362	9.4458	38.238	4.0482	0.3628
1943010	87	37	87.4	28.1	47.2	210.2	2.2206	66.194	0.1699	17.726	67.829	3.8265	0.0568
2043016	57	10	74.4	17.2	48.0	97.0	-0.266	76.639	0.0951	18.229	15.236	1.0005	0.0012
1943012	58	20	78.4	22.6	40.5	137.5	0.7146	34.883	0.1463	19.112	19.200	1.0004	0.0042

procédure d'estimation a été adoptée aussi par Jolley (1987), dans l'application de la TCEV pour la régionalisation des débits dans les régions centrales et du sud du Chili, où le modèle a présenté des bons résultats. Dans notre cas, pour les stations isolées, le programme utilisé a été adapté à partir de Jolley (1987). Quelques exemples de résultats sont montrés dans le Tableau 1.

Ce tableau montre dans les colonnes 2–8 quelques caractéristiques comme le nombre d'itérations, le nombre de données, la moyenne, l'écart type, les valeurs maxima et minima et le coefficient d'asymétrie pour quatre exemples significatifs (les stations de code 1844001, 1943010, 2043016 et 1943012, données de pluie maximale annuelle en millimètres, pour le pas de temps journalier). Les colonnes 9–12 montrent les estimateurs de Λ_1 , Λ_2 , θ_1 , et θ_2 . Les colonnes 13 et 14 montrent les estimateurs des paramètres θ^* et Λ^* , qui sont utilisés dans la régionalisation (deuxième partie de cette étude), et qui sont définis par:

$$\theta^* = \frac{\theta_2}{\theta_1} \quad (9)$$

$$\Lambda^* = \frac{\Lambda_2}{\Lambda_1^{(1/\theta^*)}} \quad (10)$$

Les ajustements de la TCEV ont été comparés avec la FD de Gumbel, celle-ci ajustée par la méthode des moments. Les stations 1844001 et 1943010 sont des exemples où on a eu des ajustements TCEV meilleurs que Gumbel. Pour ces cas, les rapports entre les gradex de la composante horsain et basique (θ^*) est nettement supérieure à 1.0, ce qui indique la présence des horsains. Au contraire, pour les stations 2043016 et 1943012, la valeur de θ^* est voisine de 1.0, ce qui montre que la TCEV tend vers la loi de Gumbel. Cependant, nous avons trouvé que pour 37% des 92 stations, l'algorithme d'identification des paramètres de la TCEV ne converge pas. Pour les cas où on a eu la convergence, 62% des stations on présenté la valeur θ^* voisine de 1.0.

Finalement, pour étudier la variabilité spatiale des paramètres de la TCEV on a obtenu les cartes d'isolignes des valeurs du coefficient d'asymétrie (g_n), du coefficient de variation (c_n) et de θ^* pour les 92 points de mesure. La Fig. 2 montre la carte d'isolignes des valeurs de g_n .

METHODE DE REGIONALISATION PAR LA TCEV

Le modèle statistique TCEV peut être utilisé pour les stations isolées, comme on a vu précédemment. Cependant, pour plusieurs stations nous avons trouvé des incertitudes

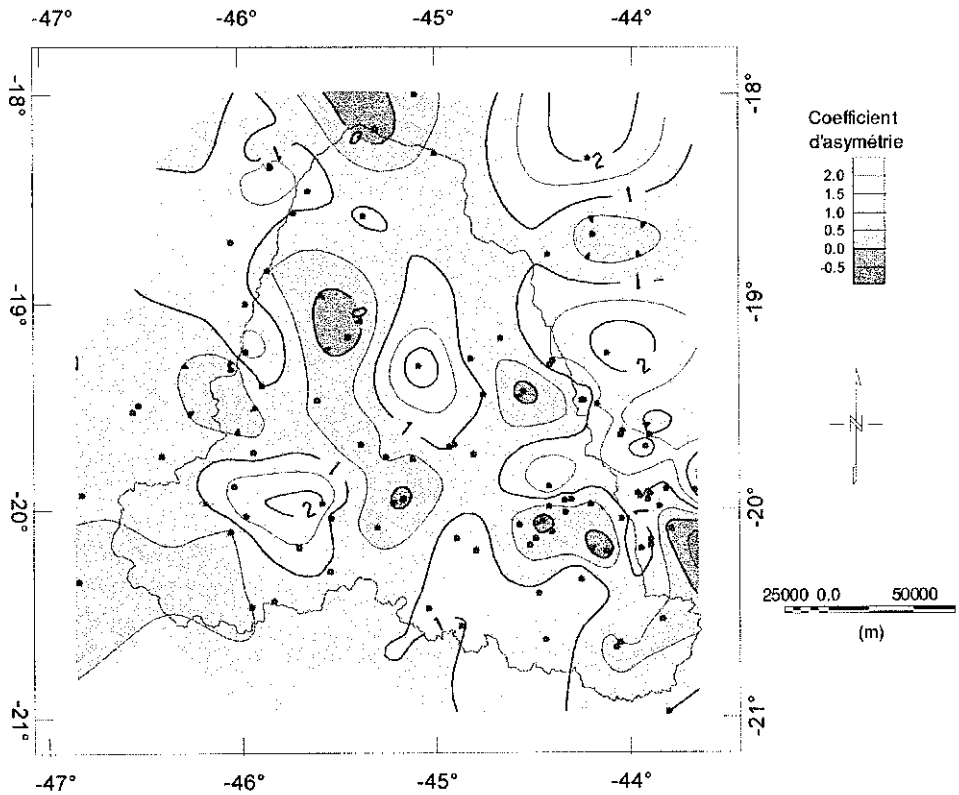


Fig. 2 Isoignes du coefficient d'asymétrie, contour du bassin et réseau pluviométrique.

dans les estimateurs des quatre paramètres, en utilisant l'algorithme basé sur la méthode ML. Ces types de problèmes aussi ont été trouvés par Arnell & Gabriele (1988) et Fiorentino *et al.* (1987a). Pour cela, la procédure peut être considérée comme le niveau zéro d'une méthodologie de régionalisation, qu'on détaillera ensuite. Ce modèle de régionalisation a été étudié et appliqué aux régions d'Italie et d'Angleterre par les auteurs Arnell & Gabriele (1988), Cannarozzo *et al.* (1995), entre autres.

Rappels sur la méthode de régionalisation hiérarchique

Une synthèse des travaux de recherche sur la méthode de régionalisation basée sur l'utilisation de la TCEV est donnée par Fiorentino *et al.* (1987b). Dans notre cas, il s'agit d'étudier son application à des pluies maxima annuelles sur la région du bassin du Alto São Francisco. En accord avec les auteurs, pour ce cas la méthode dite hiérarchique est développée selon deux niveaux. Le premier niveau consiste à chercher les régions homogènes, qui sont caractérisées par les coefficients d'asymétrie peu variables, de tel façon qu'ils peuvent être considérés comme constants.

Si on a I séries de maxima dans la région considérée homogène, on définit la variable standardisée:

$$Y = Y_{ij} = \frac{X_{ij}}{\theta_{1i}} - \ln \Lambda_{1i} \quad i = 1, \dots, I; \quad j = 1, \dots, n_i \quad (11)$$

avec I = nombre de stations (ou de séries), n_i = nombre de données de la série i et θ_{1i} , Λ_{1i} = paramètres de la composante basique de la série i . La variable Y est une variable aléatoire indépendante et identiquement distribuée selon la loi TCEV donnée par:

$$F_Y(y) = P(Y \leq y) = \exp\left[-e^{-y} - \Lambda^* \cdot e^{-y/\theta^*}\right] \quad (12)$$

avec les paramètres θ^* et Λ^* donnés par les expressions (9) et (10).

Les auteurs ont montré que, si la variable Y est distribuée selon (12), le coefficient d'asymétrie théorique est unique dans toute la région et que les rapports entre les paramètres θ_{2i} et θ_{1i} (entre la deuxième et de la première composantes) sont constants pour toutes les stations qui appartiennent à cette région. Le deuxième niveau de régionalisation consiste à chercher les sous-régions homogènes, soit où les coefficients de variation sont peu variables, ce qui implique la constance des paramètres Λ_1 de la TCEV.

Application de la méthode de régionalisation sur la région de Minas Gerais

La méthode d'estimation des paramètres régionaux du modèle TCEV adoptée est celle proposée par Gabriele & Iiritano (1994). Il s'agit d'un algorithme d'estimation conjointe de Λ^* , θ^* et de Λ_{1s} , $s = 1, \dots, M$, M étant le nombre de sous-régions, qui combine le Maximum de Vraisemblance et la méthode de Newton-Raphson.

Le point de départ pour la régionalisation a été les cartes d'isovaleurs du coefficient d'asymétrie (g_n) et du coefficient de variation (c_v). La procédure adoptée a été l'analyse des cartes d'isovaleurs et le choix des régions homogènes, soit où les valeurs de g_n varient peu et ensuite le choix des sous-régions homogènes, soit où les valeurs de c_v sont peu variables. D'après la Fig. 2 (isolignes de g_n) on a pu identifier deux régions du bassin séparées par une ligne qui coïncide à peu près avec le parallèle 19°20': la région nord où les valeurs de g_n sont plus fortes et la région sud où les valeurs sont plus faibles. Le Tableau 2 montre les résultats obtenus pour l'estimation des paramètres régionaux de la TCEV, pour les pluies maxima annuelles, en millimètres.

Tableau 2 Paramètres régionaux de la TCEV, pas de temps journalier.

Région	θ^*	Λ^*	No. stations	No. données	No. sous-régions	$\Lambda_{1,1}$	$\Lambda_{1,2}$
Nord	1.7570	0.6337	32	534	1	114.479	-
Sud	1.3877	0.8710	60	1384	2	53.7623	75.8464

CONCLUSION

L'analyse des Tableaux 1 et 2 nous permet de faire les commentaires suivants:

- (a) Pour les stations isolées, la plupart des séries disponibles est courte (moins de 20 ans d'observations), ce qui pose des problèmes pour l'identification des

paramètres d'une fonction de distribution à quatre paramètres. On voit donc l'intérêt de la régionalisation, où on étudie des échantillons plus fournis, avec toutes les stations d'une région homogène mélangées dans une seule série station-année.

- (b) Pour les stations isolées et dans les cas où on a eu la convergence dans l'algorithme d'identification des paramètres, les valeurs de θ^* forts (38% des cas) sont correspondants à des ajustements TCEV avec seulement un ou deux horsains, ce qui peut être dû à la dispersion d'échantillonnage.
- (c) Les valeurs régionales montrent la division du bassin en deux régions : la région nord ($\theta^* = 1.7570$) et la région sud ($\theta^* = 1.3877$). Néanmoins, pour la région nord on n'a pas pu identifier les sous-régions. Ces résultats partiels nous suggèrent la continuation des études de définition des sous-régions homogènes.

REFERENCES

- Arnell, N. W. & Gabriele, S. (1988) The performance of the Two Component Extreme Value distribution in regional flood frequency analysis. *Wat. Resour. Res.* **24**(6), 879-887.
- Beran, M., Hosking, J. R. M. & Arnell, N. (1986) Comment on "Two Component Extreme Value distribution for flood frequency analysis". *Wat. Resour. Res.* **22**(2), 263-266.
- Cannarozzo, M., D'Assaro, F. & Ferro, V. (1995) Regional rainfall and flood frequency analysis for Sicily using the Two Component Extreme Value distribution. *Hydrol. Sci. J.* **40**(1), 19-42.
- Fiorentino, M., Gabriele, S., Rossi, F. & Versace, P. (1987a) Hierarchical approach for regional flood frequency analysis. In: *Regional Flood Frequency Analysis* (ed. by V. P. Singh), 35-49. Reidel, Dordrecht, The Netherlands.
- Fiorentino, M., Gabriele, S., Rossi, F. & Versace, P. (1987b) Regional flood frequency analysis using the Two Component Extreme Value. A key reference abstract. *Excerpta* **2**, 39-50.
- Gabriele, S. & Iritano, G. (1994). Alcuni aspetti teorici ed applicativi nella regionalizzazione delle piogge con il modello TCEV (Some theoretical and practical aspects in the regionalization of precipitations using the TCEV model, in Italian). In: *Prediction and Prevention of Extreme Hydrological Events and their Mitigation*. Publ. no. 1089, GNDCI—CNR/IRPI, Rende, Italy.
- Guillot, P. & Duband, D. (1967) La méthode du gradex pour le calcul de la probabilité des crues à partir des pluies. In: *Floods and their Computation* (Proc. Leningrad Symp.), 560-569. IAHS Publ. no. 84.
- Jolley, T. J. (1987) The application of a regional Two Component Extreme Value distribution to Chilean flood data. MSc Thesis, Dept of Civil Engng, Imperial College of Science and Technology, London, UK.
- Rossi, F., Fiorentino, M. & Versace, P. (1984) Two Component Extreme Value distribution for flood frequency analysis. *Wat. Resour. Res.* **20**(7), 847-856.
- Rossi, F., Fiorentino, M. & Versace, P. (1986) Reply to "Comment on Two Component Extreme Value distribution for flood frequency analysis". *Wat. Resour. Res.* **22**(2), 267-269.
- Todorovic, P. & Zelenhasic, E. (1970). A stochastic model for flood analysis. *Wat. Resour. Res.* **6**(6), 1641-1648.