

HydRoStat: Guide de démarrage rapide

1 Introduction

Les codes *HydRoStat* sont constitués d'un ensemble de scripts R qui définissent des fonctions pour estimer une distribution (ainsi que les incertitudes associées) à partir d'un échantillon de valeurs observées.

Pour utiliser *HydRoStat*, il faut que R (<https://www.r-project.org/>) soit installé sur votre ordinateur. Vous devez également télécharger et installer les packages suivants :

1. `evd` (<https://cran.r-project.org/web/packages/evd/index.html>)
2. `mvtnorm` (<https://cran.r-project.org/web/packages/mvtnorm/index.html>)
3. `numDeriv` (<https://cran.r-project.org/web/packages/numDeriv/index.html>)

En pratique, les deux fonctions décrites ci-dessous sont suffisantes pour commencer à utiliser *HydRoStat*. Nous vous recommandons d'ouvrir le fichier *Exemples.R* pour comprendre comment ces fonctions sont utilisées. Tout ce que vous avez à faire est de modifier la ligne 2 (dossier où les codes *HydRoStat* sont stockés sur votre ordinateur), puis de lire et d'exécuter le script ligne par ligne.

```
h3 <- Hydro3_Estimation(y,dist,[Emeth],[Umeth],[options])
```

La fonction `Hydro3_Estimation` a 2 entrées obligatoires et 3 entrées optionnelles :

1. `y` [obligatoire] : vecteur des observations (sans valeurs manquantes)
2. `dist` [obligatoire] : chaîne de caractères décrivant la distribution à estimer. La liste des distributions disponibles est donnée en section 2.
3. `Emeth` [optionnel] : méthode d'estimation. L-Moments ("LMOM", défaut, recommandé), moments ("MOM"), maximum de vraisemblance ("ML") ou bayésien ("BAY").
4. `Umeth` [optionnel] : méthode d'estimation des incertitudes. Bootstrap paramétrique ("PBOOT", défaut, recommandé), bootstrap ("BOOT"), maximum de vraisemblance ("ML"), bayésien ("BAY") ou aucune ("NONE").
5. `options` [optionnel] : liste contenant les options d'estimation, cf. section 4.

La sortie de la fonction `Hydro3_Estimation` est un objet *Hydro3*, qui contient tous les résultats de la procédure d'estimation (paramètres estimés, quantiles, incertitudes, résultats des tests statistiques, etc.). Voir la section 3 pour plus de détails.

```
Hydro3_Plot(y,h3)
```

La fonction `Hydro3_Plot` a 2 entrées obligatoires :

1. `y` [obligatoire] : vecteur des observations (sans valeurs manquantes)
2. `h3` [obligatoire] : objet *Hydro3*, résultant d'un appel de la fonction `Hydro3_Estimation` ci-dessus.

La fonction `Hydro3_Plot` produit simplement une figure résumant la procédure d'estimation (paramètres estimés, données, fonction de répartition, densité de probabilité et fonction quantile).

2 Distributions disponibles

Distribution	ID	# paramètres	Utilisation typique *
Normale (ou Gaussienne)	"Normal"	2	QA
Log-normale	"LogNormal"	2	QA, QN
Gumbel	"Gumbel"	2	QX
Généralisée des valeurs extrêmes	"GEV"	3	QX
Pearson III	"PearsonIII"	3	QX
Log-Pearson III	"LogPearsonIII"	3	QX
Exponentielle	"Exponential2"	2	QS
Pareto généralisée	"GPD3"	3	QS
Gumbel pour les minima	"Gumbel_min"	2	QN
Généralisée des valeurs extrêmes pour les minima	"GEV_min"	3	QN
Exponentielle avec un seuil nul	"Exponential1"	1	QS
Pareto généralisée avec un seuil nul	"GPD2"	2	QS
Poisson	"Poisson"	1	N

* QA = Débit annuel, QN = Débit minimum, QX = Débit maximum, QS = Débit SUP-SEUIL, N = comptage.

3 Description d'un objet Hydro3

Un objet Hydro3 `h3` est une liste contenant tous les résultats utiles du processus d'estimation. Cette liste contient les éléments suivants (ceux marqués en rouge devraient être les plus utiles):

1. `h3$dist`: la distribution estimée.
2. `h3$empirical`: estimations empiriques. Un tableau de données (R data frame) avec les colonnes suivantes :
 - a. `y`: données triées
 - b. `freq`: fréquence au non-dépassement
 - c. `T`: période de retour
 - d. `u`: variable réduite
3. `h3$pcdf`: Fonction de répartition et densité de probabilité estimées (en anglais : pdf & cdf). Un tableau de données (R data frame) avec les colonnes suivantes :
 - a. `x`: valeur
 - b. `pdf`: densité de probabilité $f(x)$
 - c. `cdf`: fonction de répartition $F(x)$
4. `h3$quantile`: quantiles estimés. Un tableau de données (R data frame) avec les colonnes suivantes :

- a. **T**: période de retour
 - b. **p**: probabilité au non-dépassement
 - c. **u**: variable réduite
 - d. **q**: quantile estimé
 - e. **IC.low**: borne inférieure de l'intervalle d'incertitude
 - f. **IC.high**: borne supérieure de l'intervalle d'incertitude
5. **h3\$par**: paramètres estimés. Un tableau de données (R data frame) avec les colonnes suivantes :
- a. **index**: index du paramètre
 - b. **name**: nom du paramètre
 - c. **estimate**: paramètre estimé
 - d. **IC.low**: borne inférieure de l'intervalle d'incertitude
 - e. **IC.high**: borne supérieure de l'intervalle d'incertitude
 - f. **mean**: moyenne de la distribution d'échantillonnage (cf. **h3\$u**)
 - g. **median**: médiane de la distribution d'échantillonnage (cf. **h3\$u**)
 - h. **sdev**: écart-type de la distribution d'échantillonnage (cf. **h3\$u**)
6. **h3\$KS**: Résultat du test d'adéquation de Kolmogorov-Smirnov. Une liste avec les éléments suivants :
- a. **pval**: p-valeur du test
 - b. **stat**: statistique de test
 - c. **xtra**: non utilisé
7. **h3\$MK**: Résultat du test de tendance de Mann-Kendall. Une liste avec les éléments suivants :
- a. **pval**: p-valeur du test
 - b. **stat**: statistique de test
 - c. **xtra**: non utilisé
8. **h3\$Pettitt**: Résultat du test de rupture de Pettitt. Une liste avec les éléments suivants :
- a. **pval**: p-valeur du test
 - b. **stat**: statistique de test
 - c. **xtra**: position estimée de la rupture
9. **h3\$u**: Propriétés de la distribution d'échantillonnage de l'estimateur des paramètres, explorée avec des simulations Monte-Carlo :
- a. **cov**: matrice de covariance de l'estimateur des paramètres
 - b. **sim**: paramètres simulés, représentant la distribution d'échantillonnage
 - c. **ok**: indicateur logique renseignant sur le bon déroulement des simulations
 - d. **error**: code d'erreur (entier, 0 = pas d'erreur)
 - e. **message**: message d'erreur éventuel
10. **h3\$ok**: indicateur logique renseignant sur le bon déroulement des calculs
11. **h3\$error**: code d'erreur (entier, 0 = pas d'erreur)
12. **h3\$message**: message d'erreur éventuel

4 Options d'estimation

Un objet `option` est une liste contenant toutes les options disponibles pour l'estimation. En pratique, les plus utiles sont : (i) option `invertT`, qui devrait être vraie (`TRUE`) si les grandes périodes de retour correspondent aux petites valeurs (exemple typique : étiages caractérisés avec les débits minimums annuels); (ii) option `splitZeros` qui devrait être vraie (`TRUE`) si des débits

nuls sont présents dans l'échantillon et devraient être traités à part (exemple typique : débits minimums annuels pour un cours d'eau intermittent).

1. `o$FreqFormula`: formule utilisée pour les fréquences au non-dépassement (défaut: Hazen $(i-0.5)/n$)
2. `o$pgrid`: grille de probabilités définissant les valeurs où fonction de répartition et densité sont évalués.
3. `o$Tgrid`: grille de périodes de retour où les quantiles sont évalués
4. `o$IClevel`: niveau de l'intervalle d'incertitude (défaut: 0.9)
5. `o$p2T`: facteur de conversion entre période de retour et probabilité au non-dépassement, égal au nombre moyen de valeurs par (défaut : 1 ; utile pour l'échantillonnage SUP-SEUIL)
6. `o$invertT`: FALSE si les grandes périodes de retour correspondent aux grandes valeurs, TRUE sinon (défaut : FALSE)
7. `o$splitZeros`: Les valeurs nulles devraient-elle être traitées séparément? (défaut : FALSE)
8. `o$lang`: langage utilisé dans les figures (défaut : français)
9. `o$nsim`: nombre de simulations pour explorer la distribution d'échantillonnage (défaut : 1000)