

Analyse critique d'un article sur le masting selon lequel la synchronie serait plus forte dans la disette que dans l'abondance

Pr Jean R. LOBRY


Le logiciel  est très puissant pour faire des simulations aidant à l'interprétation des résultats mis en avant dans une publication. On l'illustre ici avec un article prétendant que la synchronie entre des séries temporelles de masting serait plus intense entre les années de disette qu'entre les années d'abondance. Le choix d'un critère de classification conduisant à des effectifs très déséquilibrés entre les deux classes induit un impact des points de *nuisance* très contrasté. Les points de *nuisance* sont les années où la classification n'est pas cohérente entre les deux séries. Quelques exercices pour poursuivre l'analyse sont proposés.

Table des matières

1	Introduction	2
2	Simulations	3
2.1	Trois cas dégénérés	3
2.2	Étude par simulation	6
3	Exercices	9
3.1	9
3.2	10
3.3	10
3.4	10
4	Annexes	10
4.1	rsynchr()	10
4.2	plotsynchr()	12
	Références	14

1 Introduction

ON s'intéresse ici à l'article [11] intitulé « *Geography of masting creates greater synchrony in seed scarcity than in seed abundance.* » Les données analysées sont des séries temporelles, \mathbf{x} , de suivi pendant $n = 36$ années de la production de fruits par des arbres¹. Les éléments de \mathbf{x} sont à valeur dans \mathbb{R}_+ , autrement dit nous avons des séries non négatives :

$$\forall i \in \{1, \dots, n\} : x_i \geq 0 \quad (1)$$

UNE première caractéristique importante des séries de masting est leur forte variabilité : pour reprendre l'échelle verbale de KVÅLSETH [3] les trois quart des séries de la base de données MASTREE+ [1] sont dans la gamme haute et très haute de variabilité [6]. Les distributions des valeurs des séries de masting sont donc très fortement asymétriques à droite, ce qui n'est pas sans conséquence sur le choix fait par les auteurs [11] pour distinguer les années de disette et d'abondance. Soit M_i la variable aléatoire de BERNOULLI indicatrice de l'évènement : « la i ème observation x_i correspond à une année de masting ». La règle de décision utilisée est :

$$\begin{aligned} M : \mathbf{x} &\rightarrow \{0, 1\}^n \\ x_i &\mapsto M_i = M(x_i) = \begin{cases} 1 & \text{si } x_i > \frac{\max \mathbf{x}}{2} \\ 0 & \text{sinon} \end{cases} \end{aligned} \quad (2)$$

AUTREMENT dit, on déclare années d'abondance toutes celles dont la production de fruits est strictement supérieure à la valeur maximale observée dans la série divisée par deux. C'est un choix qui est discutable, il y aurait bien des alternatives [4, 5], mais tel n'est pas l'objet ici puisque nous voulons reproduire la méthodologie utilisée [11]. La conséquence immédiate de ce choix est qu'il va y avoir un fort *déséquilibre* dans la partition définie par l'équation 2 : beaucoup d'années de disette et peu d'abondance, et ce déséquilibre ira croissant avec le niveau de variabilité de la série.

UNE seconde caractéristique importante des séries de masting est leur synchronie. Soient deux séries \mathbf{x}_1 et \mathbf{x}_2 issues de deux populations distantes, parfois de plusieurs centaines de kilomètres comme dans [11], on observe souvent qu'elles sont relativement synchrones avec des années d'abondance et de disette souvent identiques entre les deux. Pour quantifier le degré de synchronie entre deux séries on utilise des statistiques de corrélation comme le r de PEARSON [7], le ρ de SPEARMAN [9] ou le τ de KENDALL [2]. Nous utiliserons comme dans [11] le ρ de SPEARMAN qui n'est rien d'autre que le r de PEARSON calculé sur les rangs des séries \mathbf{x}_1 et \mathbf{x}_2 .

L'APPROCHE utilisée dans [11] est de partitionner les données selon l'équation 2 et de calculer un coefficient de corrélation dit « partiel » pour les années d'abondance et de disette. Attention, cette dénomination peut porter à confusion parce que l'on parle également de corrélation partielle pour la corrélation entre deux variables conditionnellement à une autre, on prendra soin de bien toujours mettre des guillemets. Mais il y a un petit problème : que fait-on

¹Sapin argenté *silver fir*, hêtre *European beech*, mélèze *European larch*, épicéa *Norway spruce*, pin sylvestre *Scots pine*, chêne sessile *sessile oak* et chêne pédonculé *pedunculate oak*.

par exemple si $M_i = 1$ pour \mathbf{x}_1 et $M_i = 0$ pour \mathbf{x}_2 ? Voici ce qui est dit [11, p. 7] :


In cases when the annual value of seed production for the two sites falls into opposite tails, that value was included when calculating the partial Spearman correlation in both tails [10].

Et en retournant à la source [10, p. 5] :

In cases when the annual value of seed production for the two sites falls into opposite tails, that value was included when calculating partial Spearman correlation in both tails. Thus, if one site experienced a mast peak and the other a year of seed scarcity in the same year, synchrony was reduced in both tails.


LE diable est dans les détails et c'est ici que le bât blesse. Autant l'équation 2 définit bien une partition du jeu de données, autant on ne peut pas considérer que les corrélations « partielle » d'abondance et de pénurie sont une décomposition de la corrélation globale puisqu'elles peuvent être calculées *avec des points en commun*. Nous appellerons ces points en commun des points de nuisance. Pour un couple de deux séries $(\mathbf{x}_1, \mathbf{x}_2)$ on définit la variable aléatoire de BERNOULLI indicatrice de l'évènement : « la i ème année du couple (x_{1_i}, x_{2_i}) correspond à un point de nuisance » par :

$$\begin{aligned} N : (\mathbf{x}_1, \mathbf{x}_2) &\rightarrow \{0, 1\}^n \\ (x_{1_i}, x_{2_i}) &\mapsto N_i = N(x_{1_i}, x_{2_i}) = \begin{cases} 1 & \text{si } M(x_{1_i}) \neq M(x_{2_i}) \\ 0 & \text{sinon} \end{cases} \end{aligned} \quad (3)$$

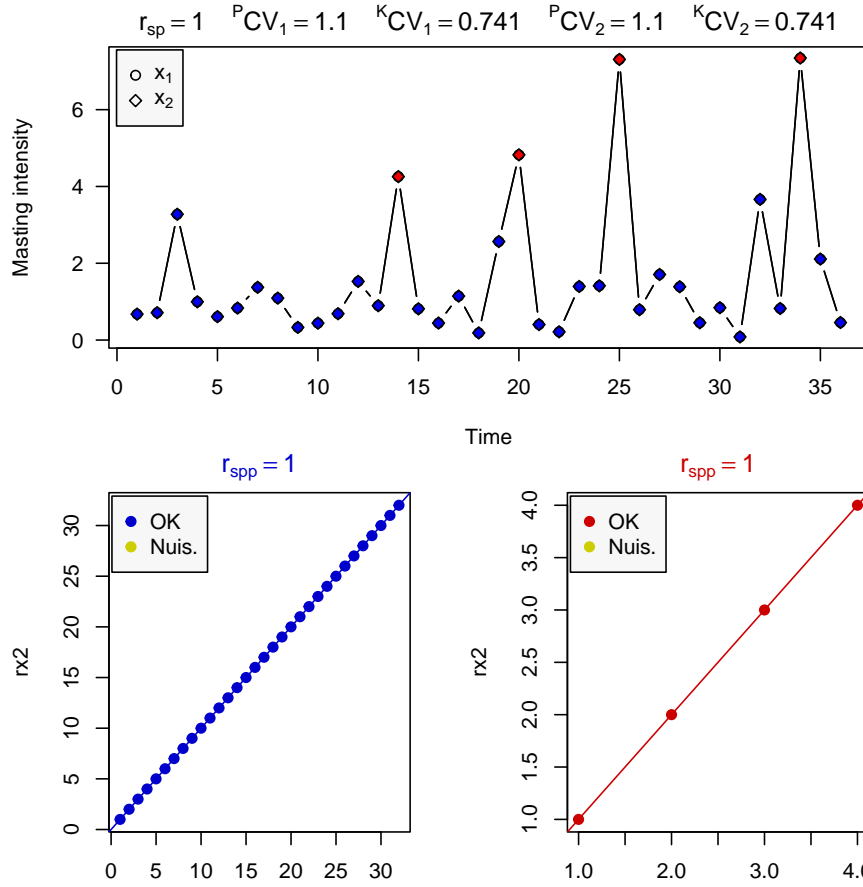
NOTONS que ces points de nuisance seront par construction communs aux sous-ensembles utilisés pour calculer la synchronie des années d'abondance et de disette. Il est un peu abusif de dire que l'on a décomposé la synchronie entre les années de pénurie et d'abondance en s'appuyant sur des statistiques calculées sur des données pouvant se chevaucher. Mais à quel point cela pose-t-il problème en pratique ? C'est ici que  est un outil puissant.

2 Simulations

2.1 Trois cas dégénérés

LE code  utilisé ici est donné en annexe (section 4 page 10). On commence par le cas dégénéré de deux séries parfaitement synchrones pour expliquer les sorties graphiques.

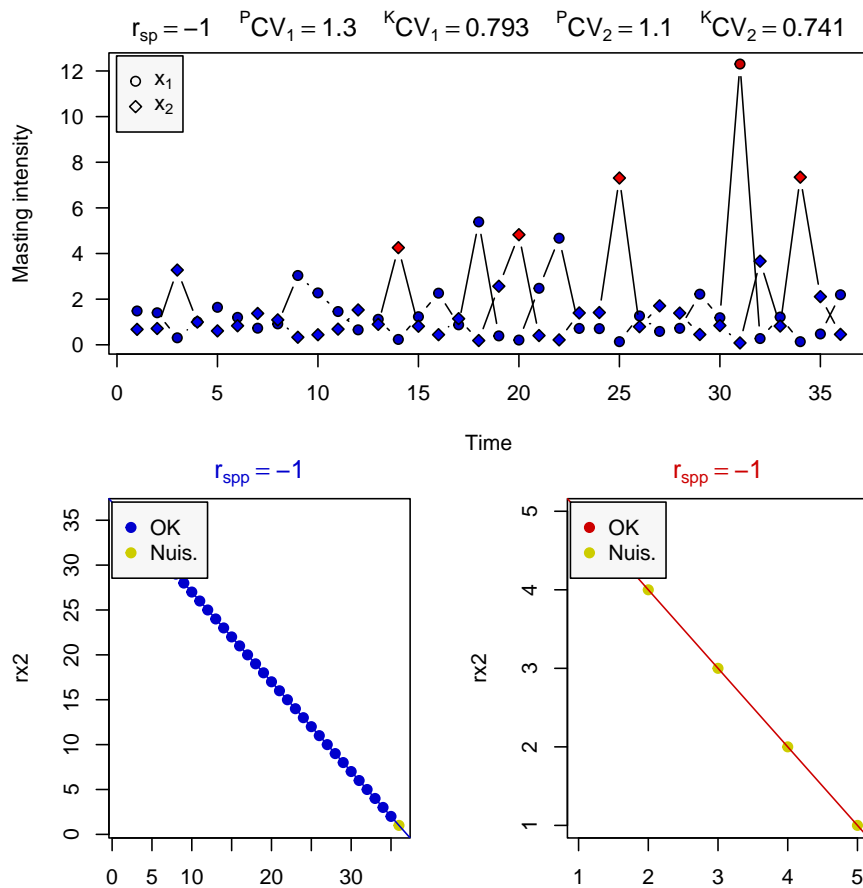
```
set.seed(1) ; plotsynchr(r = 1)
```



SUR le graphique du haut nous avons la représentation de nos deux séries x_1 et x_2 qui dans ce cas particulier sont parfaitement superposées. Les années d'abondance sont en rouge et celles de disette en bleu. La valeur du coefficient de corrélation de SPEARMAN est donnée en titre du graphique ainsi que les coefficients de variation de PEARSON [8], $^P CV$, et celui de KVÅLSETH [3], $^K CV$. Par défaut on fait une série de 36 ans comme dans [11] mais on peut modifier ça avec le paramètre `ny`.

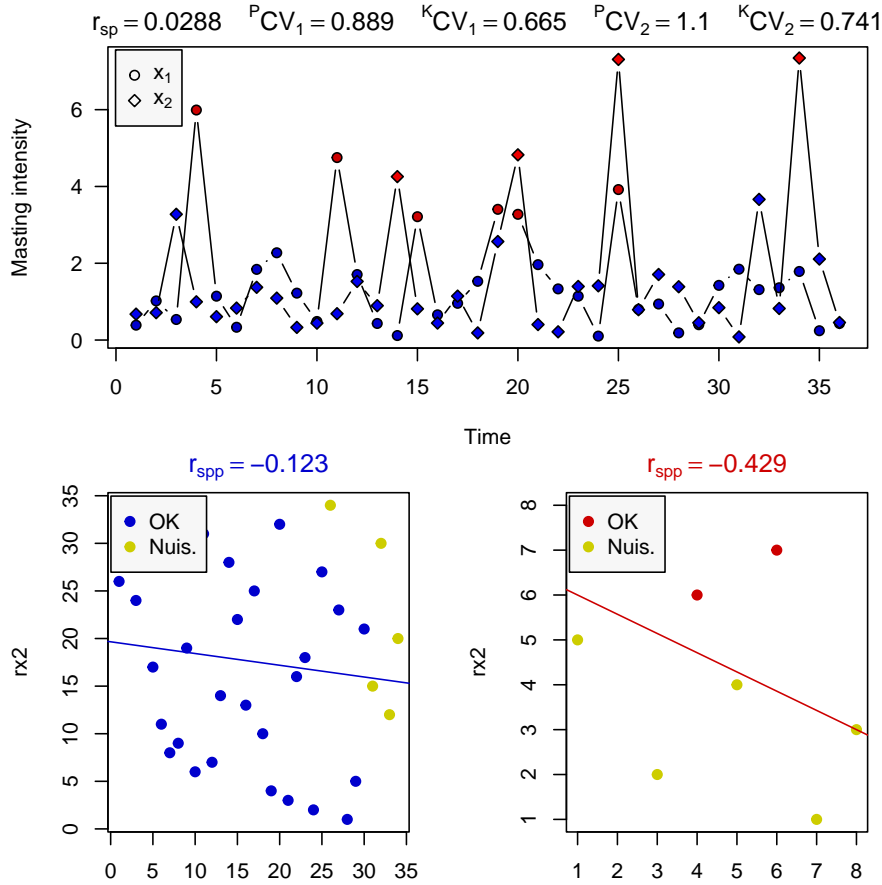
DANS les graphiques du bas on a séparé les années de disette à gauche des années d'abondance à droite. Les titres donnent les ρ « partiels » de SPEARMAN. Les points de nuisance (équation 3) sont en jaune lavasse, il n'y en a pas ici puisque les séries sont identiques. Ce sont les rangs des valeurs qui sont représentées pour avoir le pendant graphique du ρ de SPEARMAN. Pour mieux comprendre la coloration des points prenons le cas dégénéré de deux séries en parfaite opposition de phase.

```
set.seed(1) ; plotsynchr(r = -1)
```



CETTE fois nous avons des points de nuisance en jaune lavasse. La synchronie « partielle » pour les années d'abondance n'est calculée qu'avec des points de nuisance, on voit la limite du sens que l'on peut donner à cette statistique. Voyons maintenant le cas dégénéré de deux séries complètement dé-corrélées.

```
set.seed(1) ; plotsynchr(r = 0)
```



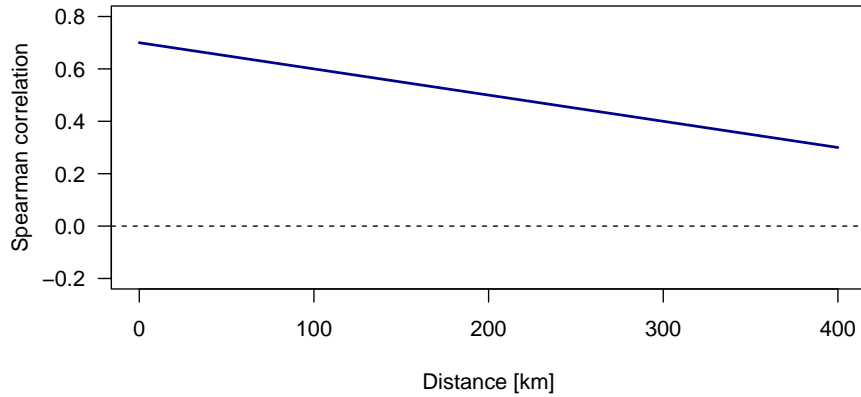
C'EST là que ça devient intéressant sur le graphique en bas à droite. Comme les années d'abondance sont des événements rares il est peu probable d'observer simultanément une année d'abondance dans les deux séries, on se retrouve avec beaucoup de points de nuisance qui ont une valeur élevée dans une série et une valeur basse dans l'autre. On induit ainsi une corrélation *négative* dans la statistique des années d'abondance. On le fait également dans la statistique des années de disette, mais l'effet est moins prononcé car noyé dans plus de valeurs. Maintenant que nous avons compris l'origine du gag, nous pouvons procéder par simulation pour étudier son impact.

2.2 Étude par simulation

ON part de la figure [11, fig. 1] pour le hêtre (*beech*), on voit un ρ de SPEARMAN qui décroît à peu près linéairement de 0.7 à 0.3 entre une distance 0 à 400 km entre les sites. On résume donc ça par le modèle suivant :

```
taupred <- function(distance) -0.001*distance + 0.7
xx <- seq(0, 400, le = 256)
plot(xx, taupred(xx), type = "l", lwd = 2, xlab = "Distance [km]", las = 1,
     main = "Synchronie globale pour le hêtre", ylab = "Spearman correlation",
     ylim = c(-0.2, 0.8), col = "darkblue")
abline(h = 0, lty = 2)
```

Synchronie globale pour le hêtre



ON commence par initialiser avec des valeurs manquantes les tables `GlobalSyn`, `LowTailSyn` et `UppTailSyn` pour stocker les résultats de nos simulations pour les années de disette et d'abondance, respectivement. La variable `ndist` contrôle le nombre de distances entre 0 et 400 km que l'on veut échantillonner et la variable `nsites` le nombre de sites souhaité à chaque point kilométrique. Dans les tables de résultat les sites sont en ligne et les points kilométriques en colonnes.

```
nsites <- 256 ; ndist <- 11 ; seqd <- as.integer(seq(0, 400, le = ndist))
LowTailSyn <- as.data.frame(matrix(NA, nrow = nsites, ncol = ndist))
colnames(LowTailSyn) <- paste0("D", seqd)
GlobalSyn <- UppTailSyn <- LowTailSyn
```

ON définit la fonction `getSyn()` pour tirer au hasard *un* couple de séries temporelles (x_1, x_2) et calculer les corrélations « partielles » pour les années d'abondance (`Hrsp`) et les années de disette (`Lrsp`) ainsi que la synchronie globale (`Grsp`). Elle fait appel à la fonction `rsynch()` expliquée en annexe (section 4 page 10) pour avoir un couple de séries avec un niveau de synchronie donné. L'argument `nuisance` contrôle si on souhaite incorporer les points de nuisance ou non. Comme tout est vecteur dans **R** le code est compact et élégant.

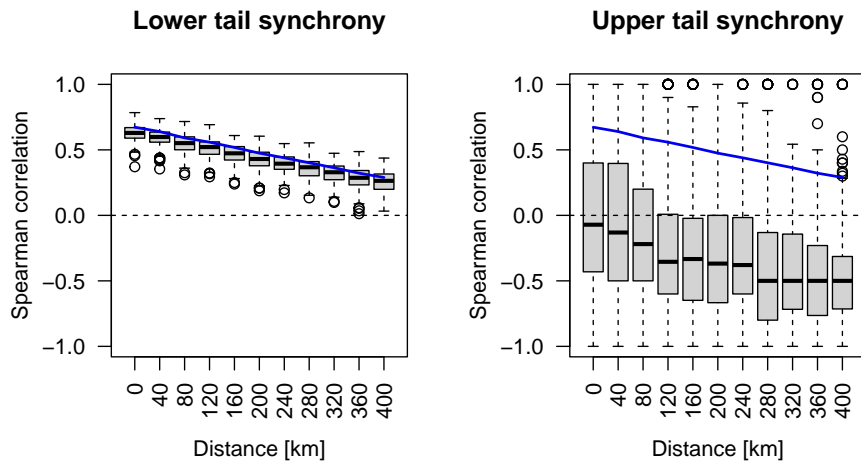
```
getSyn <- function(nuisance = TRUE, method = "spearman", ...){
  z <- rsynch(...); x1 <- z[, 1]; x2 <- z[, 2]
  Hx1 <- x1 > max(x1)/2.0 ; Lx1 <- !Hx1
  Hx2 <- x2 > max(x2)/2.0 ; Lx2 <- !Hx2
  if(nuisance){
    Hx12 <- Hx1 | Hx2
    Lx12 <- Lx1 | Lx2
  } else {
    Hx12 <- Hx1 & Hx2
    Lx12 <- Lx1 & Lx2
  }
  Grsp <- cor(x1, x2, method = method)
  Hrsp <- cor(x1[Hx12], x2[Hx12], method = method)
  Lrsp <- cor(x1[Lx12], x2[Lx12], method = method)
  return(c(Grsp = Grsp, Hrsp = Hrsp, Lrsp = Lrsp))
}
```

IL suffit maintenant de faire rouler notre fonction avec `replicate()` pour peupler les colonnes des tables de résultat `UppTailSyn`, `LowTailSyn` et `GlobalSyn`.

```
r <- taupred(seqd)
for(j in seq_len(ndist)){
  manip <- as.data.frame(t(replicate(nsites, getSyn(r = r[j]))))
  GlobalSyn[, j] <- manip$Grsp
  UppTailSyn[, j] <- manip$Hrsp
  LowTailSyn[, j] <- manip$Lrsp
}
```

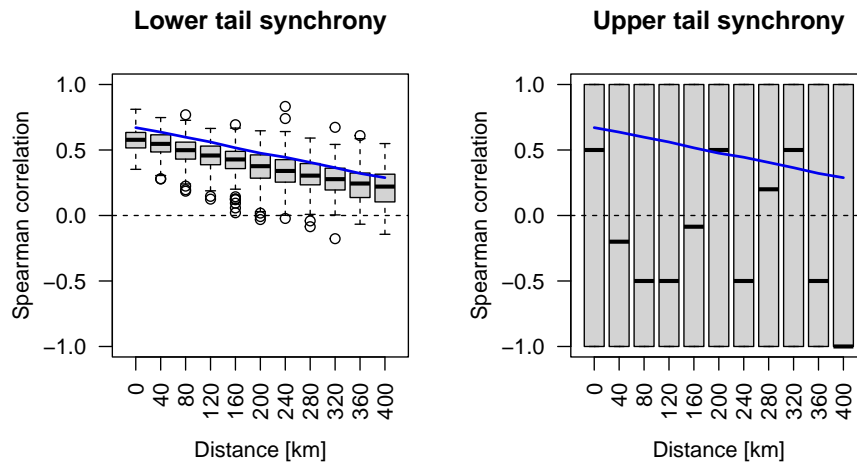
IL ne reste plus qu'à représenter la distribution des valeurs pour chaque point kilométrique. On superposera en bleu la synchronie globale.

```
plotUppLow <- function(ylim = c(-1, 1), ...){
  par(mfrow = c(1, 2))
  mybxplt <- function(x, ...){
    boxplot(x, ylim = ylim, las = 2, names = substr(colnames(x), 2, 4),
      xlab = "Distance [km]", ylab = "Spearman correlation", ...)
    abline(h = 0, lty = 2)
    points(1:ndist, apply(GlobalSyn, 2, median, na.rm = TRUE), col = "blue2", type = "l", lwd = 2)
  }
  mybxplt(LowTailSyn, main = "Lower tail synchrony")
  mybxplt(UppTailSyn, main = "Upper tail synchrony")
}
plotUppLow()
```



LE moins que l'on puisse dire c'est qu'il n'y a pas le même impact des points de nuisance sur les années de disette et d'abondance. À gauche, pour les années de disette, la médiane est assez proche de la valeur globale, on note quelques *outliers* du coté des valeurs faibles, sans doute quand le nombre de points de nuisance est élevé, tirant la corrélation vers le bas. À droite, pour les années d'abondance, l'impact est dramatique puisque la médiane est toujours négative. Les points de nuisance plombent complètement la corrélation. Regardons maintenant ce qui se passe si on n'incorpore pas les points de nuisance.

```
GlobalSyn[, ] <- UppTailSyn[, ] <- LowTailSyn[, ] <- NA
for(j in seq_len(ndist)){
  manip <- as.data.frame(t(replicate(nsites, getSyn(r = r[j], nuisance = FALSE))))
  GlobalSyn[, j] <- manip$Grsp
  UppTailSyn[, j] <- manip$Hrsp
  LowTailSyn[, j] <- manip$Lrsp
}
plotUppLow()
```

À gauche pour les années de disette, cela ne change pas grand chose, on a une distribution plus symétrique avec des *outliers* des deux cotés de la distribution. À droite, pour les années d'abondance, on a quelque chose d'assez moche qui s'explique par le faible nombre de points qui satisfont le critère (eq. 2) simultanément dans les deux séries. On peut s'en convaincre en comptant le nombre de valeurs manquantes (on ne peut pas calculer le coefficient de corrélation quand il y a zéro ou un point) dans la table :

```
apply(UppTailSyn, 2, \(x) sum(is.na(x)))
D0  D40  D80  D120 D160 D200 D240 D280 D320 D360 D400
158 161 167 196 187 193 198 210 218 223 217
```

IL y a de 60 % à 90 % de valeurs manquantes selon le niveau de synchronie. On va aussi se retrouver avec beaucoup de cas avec seulement deux points, soit $\rho = \pm 1$, ce qui explique l'étendue des boîtes de -1 à +1 sur le graphique. Il n'y a en fait pas beaucoup de valeurs distinctes possible pour ρ avec si peu de points :

```
apply(UppTailSyn, 2, \(x) length(unique(zapsmall(x))))
D0  D40  D80  D120 D160 D200 D240 D280 D320 D360 D400
12  12  11   8   9  11   7  10   7   5   4
```

Au vu de ces résultats on pourrait être tenté d'interpréter l'introduction des points de nuisance comme une volonté de pouvoir calculer un coefficient de corrélation « partiel » *quoi qu'il en coûte*, comme disait « l'autre ». Les pessimistes y verront une illustration du triste *publish or perish*, les optimistes une invitation à la très chère *disputatio* de l'Université.

3 Exercices

3.1

PENSEZ-VOUS que les résultats des simulations seront changés si on utilise le coefficient de PEARSON au lieu de celui de SPEARMAN? Faites ensuite la simulation.

3.2

POUR augmenter le nombre de points disponibles (hors points de nuisance) pour calculer le coefficient de corrélation « partiel » des années d'abondance on pourrait être tenté d'augmenter la longueur des séries en jouant sur le paramètre `ny`. Essayez. Ça ne marche pas. Pourquoi ?

3.3

EN jouant sur le niveau de variabilité des séries avec le paramètre `varxy` essayez de vous rapprocher le plus possible de la figure [11, fig. 1]. Faites alors quelques simulations avec la fonction `plotsynchr()`. Ces séries vous semblent-elles bien représenter le phénomène de *masting* ?

3.4

LA fonction utilisée ici pour simuler des séries de *masting* n'est pas complètement satisfaisante dans le sens où elle ne comporte pas suffisamment de valeurs nulles. Transformez-la pour que toutes les valeurs en dessous d'un seuil donné soient forcées à zéro. Refaire alors les analyses avec cette nouvelle version.

4 Annexes

4.1 `rsynchr()`

LA fonction `rsynchr()` qui permet de tirer deux séries x et y avec un niveau de synchronie donné dérive directement de la fonction `mvrnorm()` du paquet MASS [12] qui permet de tirer un échantillon pseudo-aléatoire de n points dans une distribution normale bivariée.

```
rsynchr <- function(ny = 36, mu = c(0, 0), varxy = 1, r = 0.7, empirical = TRUE, ...){  
  require(MASS)  
  Sigma <- matrix(c(varxy, r*varxy, r*varxy, varxy), nrow = 2, ncol = 2)  
  return(exp(mvrnorm(ny, mu, Sigma, empirical = empirical)))  
}
```

LE paramètre `Sigma` de la fonction `mvrnorm()` est la matrice de variance-covariance ς :

$$\varsigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \quad (4)$$

POUR se simplifier la vie on va supposer, ce qui n'est pas déraisonnable pour des études de *masting* intra-spécifiques, qu'il y a homoscédasticité (c'est à dire $\sigma_x^2 = \sigma_y^2 = \sigma^2$) et que la matrice ς se simplifie en :

$$\varsigma = \begin{pmatrix} \sigma^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma^2 \end{pmatrix} \quad (5)$$

Le coefficient de corrélation linéaire r_{xy} de PEARSON [7] s'écrit alors comme :

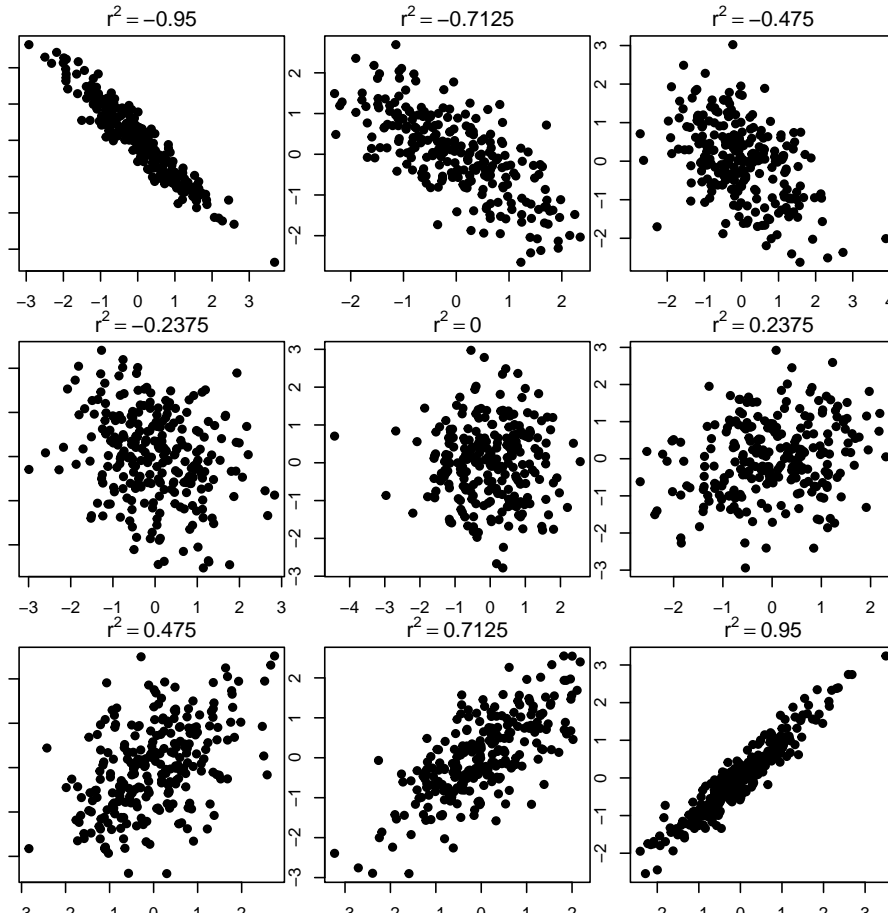
$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma^2} \quad (6)$$

Il suffit donc d'utiliser une matrices de la forme :

$$\varsigma = \begin{pmatrix} \sigma^2 & r_{xy}\sigma^2 \\ r_{xy}\sigma^2 & \sigma^2 \end{pmatrix} \quad (7)$$

VOICI un exemple d'utilisation pour illustrer la signification du coefficient de corrélation linéaire. On utilise ici l'option `empirical = TRUE` pour que la valeur du coefficient de variation soit exactement celle demandée.

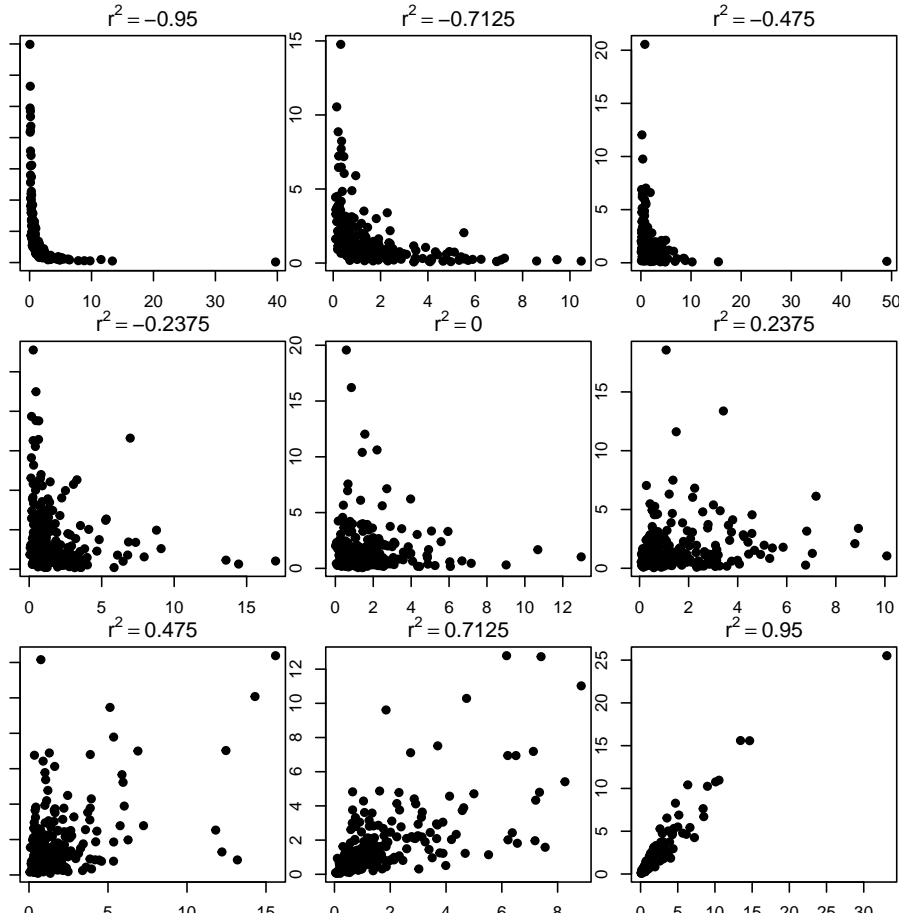
```
set.seed(1) ; par(mfrow = c(3, 3), mar = c(1.5, 1, 2, 1))
for(r in seq(-0.95, +0.95, le = 9))
  plot(mvrnorm(256, c(0,0), matrix(c(1, r, r, 1), 2, 2), empirical = TRUE),
       main = bquote(r^2 == .(r)), pch = 19)
```



POUR avoir une distribution de type *masting* j'ai simplement pris l'exponentielle des valeurs. Cela permet de générer des séries non-négatives asymétriques à droite, et ce d'autant plus que la valeur du paramètre `varxy` est élevée. Avec les valeurs par défaut on obtient des séries comme celles présentées dans

la section 2.1 page 3. La valeur du ρ de SPEARMAN [9] est inchangée par cette transformation monotone mais pas celle du r de PEARSON [7].

```
set.seed(1) ; par(mfrow = c(3, 3), mar = c(1.5, 1, 2, 1))
for(r in seq(-0.95, +0.95, le = 9))
  plot(rsynch(ny = 256, r = r),
       main = bquote(r^2 == .(r)), pch = 19)
```



4.2 plotsynchr()

Le code ci-après est repris de [6, supp. mat.] pour calculer le coefficient de variation de PEARSON [8], ^PCV, et celui de KVALSETH [3], ^KCV. L'URL donnée dans l'article² est obsolète mais devrait rediriger vers la nouvelle³ automatiquement.

```
sdn <- function(x, i, ...){
  n <- sum(!is.na(x)) # number of non missing values
  return(sqrt((n - 1)/n)*sd(x[i], ...))
}
PCV <- function(x, i, ...){
  barx <- mean(x[i], ...)
  if(isTRUE(all.equal(barx, 0))) return(0)
}
```

²<https://pbil.univ-lyon1.fr/R/donnees/CVisDead.zip>

³<https://esb.univ-lyon1.fr/donnees/CVisDead.zip>

```
} return(sdn(x[i], ...)/barx)
}  
KCV <- function(x, i, ...){  
  PCV2 <- PCV(x, i, ...)~2  
  return(sqrt(PCV2/(1 + PCV2)))  
}
```

Le code ci-après est utilisé pour produire les graphiques de la section 2.1 page 3, se reporter à cette dernière pour leur explication.

```
plotsynchr <- function(ny = 36, ...){  
  layout(matrix(c(1, 1, 2, 3), 2, 2, byrow = TRUE))  
  z <- rsynchr(ny = ny, ...) ; x1 <- z[, 1] ; x2 <- z[, 2]  
  rsp <- signif(cor(x1, x2, method = "spearman"), 3)  
  PCV1 <- signif(PCV(x1, na.rm = TRUE), 3)  
  PCV2 <- signif(PCV(x2, na.rm = TRUE), 3)  
  KCV1 <- signif(KCV(x1, na.rm = TRUE), 3)  
  KCV2 <- signif(KCV(x2, na.rm = TRUE), 3)  
  main <- bquote(r[sp] == .(rsp)~  
    phantom(0)~P*CV[1] == .(PCV1)~  
    phantom(0)~K*CV[1] == .(KCV1)~  
    phantom(0)~P*CV[2] == .(PCV2)~  
    phantom(0)~K*CV[2] == .(KCV2))  
  Hx1 <- x1 > max(x1)/2 ; Lx1 <- !Hx1  
  Hx2 <- x2 > max(x2)/2 ; Lx2 <- !Hx2  
  bg1 <- ifelse(Hx1, "red3", "blue3")  
  
  # Time series  
  par(mar = c(4, 4, 2, 2) + 0.1)  
  plot(1:ny, x1, type = "b", pch = 21, ylim = range(z), main = "", bg = bg1,  
    xlab = "Time", ylab = "Masting intensity", las = 1)  
  title(main = main, line = 1)  
  bg2 <- ifelse(Hx2, "red2", "blue2")  
  points(1:ny, z[,2], type = "b", pch = 23, bg = bg2)  
  legend("topleft", inset = 0.01,  
    legend = c(bquote(x[1]), bquote(x[2])), pch = c(21, 23), bg = grey(0.97))  
  Hx12 <- Hx1 | Hx2  
  Hrsp <- signif(cor(x1[Hx12], x2[Hx12], method = "spearman"), 3)  
  Lx12 <- Lx1 | Lx2  
  Lrsp <- signif(cor(x1[Lx12], x2[Lx12], method = "spearman"), 3)  
  par(mar = c(3, 4, 1.5, 2) + 0.1)  
  
  # scarcity correlation plot  
  col <- ifelse(Lx1[Lx12] & Lx2[Lx12], "blue3", "yellow3")  
  rx1 <- rank(x1[Lx12]) ; rx2 <- rank(x2[Lx12])  
  plot(rx1, rx2, pch = 19, col = col)  
  abline(lm(rx2~rx1), col = "blue3")  
  title(main = bquote(r[spp] == .(Lrsp)), line = 1, col.main = "blue3")  
  legend("topleft", inset = 0.01, legend = c("OK", "Nuis."),  
    col = c("blue3", "yellow3"), pch = 19, bg = grey(0.97))  
  
  # abundance correlation plot  
  col <- ifelse(Hx1[Hx12] & Hx2[Hx12], "red3", "yellow3")  
  rx1 <- rank(x1[Hx12]) ; rx2 <- rank(x2[Hx12])  
  plot(rx1, rx2, pch = 19, col = col)  
  title(main = bquote(r[spp] == .(Hrsp)), line = 1, col.main = "red3")  
  abline(lm(rx2~rx1), col = "red3")  
  legend("topleft", inset = 0.01, legend = c("OK", "Nuis."),  
    col = c("red3", "yellow3"), pch = 19, bg = grey(0.97))  
}
```

Références

- [1] A. Hacket-Pain, J.J. Foest, I.S. Pearse, J.M. LaMontagne, W.D. Koenig, G. Vacchiano, M. Bogdziewicz, T. Caignard, P. Celebias, J. van Dormolen, M. Fernández-Martínez, J.V. Moris, C. Palaghianu, M. Pesendorfer, A. Satake, E. Schermer, A.J. Tanentzap, P.A. Thomas, D. Vecchio, A.P. Wion, T. Wohlgemuth, T. Xue, K. Abernethy, M.-C. Aravena A., M.D. Barrera, J.H. Barton, S. Boutin, E.R. Bush, S.D. Calderón, F.S. Carevic, C.V. de Castilho, J.M. Cellini, C.A. Chapman, H. Chapman, F. Chianucci, P. da Costa, L. Croisé, A. Cutini, B. Dantzer, R.J. DeRose, J.-T. Dikangadissi, E. Dimoto, F.L. da Fonseca, L. Gallo, G. Gratzer, D.F. Greene, M.A. Hadad, A.H. Herrera, K.J. Jeffery, J.F. Johnstone, U. Kalbitzer, W. Kantorowicz, C.A. Klimas, J.G.A. Lageard, J. Lane, K. Lapin, M. Ledwoń, A.C. Leeper, M.V. Lencinas, A.C. Lira-Guedes, M.C. Lordon, P. Marchelli, S. Marino, H. Schmidt Van Marle, A.G. McAdam, L.R. Momont, M. Nicolas, L.H. de Oliveira Wadt, P. Panahi, G. Martínez Pastur, T. Patterson, P. Luis Peri, L. Piechnik, M. Pourhashemi, C. Espinoza Quezada, F.A. Roig, K. Peña Rojas, Y. Micaela Rosas, S. Schueler, B. Seget, R. Soler, M.A. Steele, M. Toro-Manríquez, C.E.G. Tutin, T. Ukizintambara, L. White, B. Yadok, J.L. Willis, A. Zolles, M. Żywiec, and D. Ascoli. MASTREE+ : time-series of plant reproductive effort from six continents. *Global Change Biology*, 00 :1–17, 2022.
- [2] M.G. Kendall. A new measure of rank correlation. *Biometrika*, 30 :81–93, 1938.
- [3] T.O. Kvålseth. Coefficient of variation : the second-order alternative. *Journal of Applied Statistics*, 44 :402–415, 2017.
- [4] J.M. LaMontagne. *Spatial and temporal variability in white spruce (Picea glauca) cone production : individual and population responses of North American red squirrels (Tamiasciurus hudsonicus)*. PhD thesis, University of Alberta Edmonton, Alberta, Canada, 2007.
- [5] J.M. Lamontagne and S. Boutin. Local-scale synchrony and variability in mast seed production patterns of *Picea glauca*. *Journal of Ecology*, 95(5) :991–1000, 2007.
- [6] J.R. Lobry, M.-C. Bel-Venner, M. Bogdziewicz, A. Hacket-Pain, and S. Venner. The CV is dead, long live the CV! *Methods in Ecology and Evolution*, 14 :2780–2786, 2023.
- [7] K. Pearson. VII. note on regression and inheritance in the case of two parents. *Proceedings of the royal society of London*, 58 :240–242, 1895.
- [8] K. Pearson. VII. mathematical contributions to the theory of evolution.—III. regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A*, (187) :253–318, 1896.
- [9] C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15 :72–101, 1904.

- [10] J. Szymkowiak, J. Foest, A. Hacket-Pain, V. Journé, D. Ascoli, and M. Bogdziewicz. Tail-dependence of masting synchrony results in continent-wide seed scarcity. *Ecology letters*, 27(7) :e14474, 2024.
- [11] J. Szymkowiak, J.J. Foest, M.K. Dyderski, S. Jastrzębowski, M.K. Barczyk, A. Hacket-Pain, and M. Bogdziewicz. Geography of masting creates greater synchrony in seed scarcity than in seed abundance. *Proceedings of the National Academy of Sciences*, 122(51) :e2518918122, 2025.
- [12] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002.