

## **Modélisations des pleines et basses mers à Hell-Ville en utilisant la méthode Census X11 et l'approche de Box et Jenkins**

**Alimanana Nainamahefa RASOLOMANAMPAMONJY\***, Adolphe Andriamanga RATIARISON  
et Nirry Arinavalona RAKOTOVAO

*Université d'Antananarivo, Laboratoire Dynamique de l'Atmosphère, du Climat et des Océans (DyACO),  
Madagascar*

---

\* Correspondance, courriel : [nainamahef@gmail.com](mailto:nainamahef@gmail.com)

### **Résumé**

Ce travail est une initiation à la modélisation Census X11-ARIMA-GARCH des séries chronologiques périodiques avec tendance. Avant de mettre en œuvre la méthode Census X11, nous avons montré après l'application d'une transformation logarithmique que nos chroniques sur les pleines mers et sur les basses mers possèdent une saisonnalité commune de 29 Lags (14,765 jours) et des tendances puisque les taux de Kendall de -9,98 % pour les pleines mers et de 11,80 % pour les basses mers sont tous les deux différents de 0. L'analyse de la variance à un facteur de nos deux séries chronologiques nous a permis de découvrir sept cycles compris entre 6 ans et demi et 9 ans à l'intérieur desquels la saisonnalité de 29 lags restait stable. Par conséquent, nous avons décidé de construire sept modèles correspondant à chaque cycle. Pour obtenir les sept modèles, nous avons commencé à appliquer la méthode Census X11 aux données correspondant à chaque cycle. Ensuite, nous avons procédé à la modélisation ARIMA-GARCH des séries corrigées des variations saisonnières obtenues après l'application de Census X11. Finalement, nous avons pu construire sept modèles CensusX11-ARIMA-GARCH avec des résidus non corrélés d'ordre supérieur ou égal à 1 au test de Ljung-Box. Dans cette étude, nous avons utilisé le modèle ARCH(1) ou EGARCH (0,1) pour prendre en compte l'hétéroscédasticité des résidus. Pour valider nos résultats, nous avons comparé nos erreurs avec celles fournies par le logiciel de prédictions des marées du Service Hydrographique et Océanographique de la Marine française (SHOM) de l'édition 2014 - 2015. Les indicateurs de THEIL de nos modèles sont tous inférieurs à 1 en prenant la méthode harmonique du SHOM comme méthode alternative sauf pour la prévision des pleines mers. Nous pouvons donc affirmer qu'en général, nos modèles offrent une meilleure précision par rapport à la méthode harmonique. Dans le cas de la prévision des pleines mers ou l'indicateur de THEIL est supérieur à 1, nous avons remarqué que la moyenne des erreurs restait inférieure à 5 %. Par conséquent, nous pouvons dire que nous avons de très bons modèles pour effectuer la prévision des pleines et basses mers à Hell-Ville. Les données des marées que nous avons utilisées dans ce travail proviennent des tables des marées du Centre National de Recherches Océanographiques de Hell-Ville entre 1990 et 2014. La zone d'étude est repérée au point de « latitude 13° 24' Sud et de longitude : 48° 18' Est ». A titre indicatif, le « Lag » sur les figures correspond à la moitié d'une journée lunaire synodique ou à l'équivalent de 12 heures 25 minutes.

**Mots-clés :** *Census X11, ARIMA, GARCH, stratégie de Dickey Fuller Augmenté, test de Ljung Box, critères d'information.*

## **Abstract**

### **Modelling high and low tides from Hell-ville using Census X11 and Box and Jenkins methodology**

This work is an initiation for modeling series with trend and seasonality by Census X11-ARIMA-GARCH. After logarithmic transformation, we showed data have 29 Lags seasonality (14.765 days). Particularly high tides have a decreasing trend and low tides an increasing trend because Kendall's tau have respectively 9.98 % and de11.80 % values. We detect seven local seasonality after applying Analysis of variance one factor located between 6.5 and 9 years. Consequently we decided to establish seven Census X11-ARIMA-GARCH models. Heteroscedasticity of residues was estimated by ARCH(1) or EGARCH(0,1). Finally we established seven models with white noise uncorrelated for order 1 or more after applying Ljung-Box test. To validate our results, we compared our errors with result by the software predictions of tides using by hydrographic and oceanographic service of the french navy on 2014 - 2015. If alternative method results from software of french navy, our THEIL's ratio were less than 1, except forecasting high tides. Nevertheless the mean absolute percentage error for high tide model was smaller than 5 %. We can conclude that we have excellent model to forecast high and low tides from Hell-ville. Data of tides we used in this work result from tables of tides for the national oceanographic researches center from Hell-ville between 1990 to 2014. The area of study is located in "latitude 13 ° 24 ' South and in longitude : 48 ° 18 ' " East. One "Lag" on figures corresponds in the middle of lunar synodic day or 12 hours and 25 minutes.

**Keywords :** *Census X11, ARIMA, GARCH, Dickey Fuller Augmented test, Ljung Box test, information criterion.*

## **1. Introduction**

Les prévisions des pleines et basses mers utilisées à Madagascar actuellement proviennent essentiellement des modèles fournis par le service Hydrographique et Océanographique de la Marine Française [1]. Nous avons donc décidé d'effectuer cette étude pour montrer que cette prévision peut être réalisée directement sur place à Madagascar à l'aide d'outils simples et fiables comme Matlab et Excel. Par rapport aux méthodes harmoniques utilisant l'équation du potentiel générateur des marées [2] et les méthodes numériques basés sur l'approximation par éléments finis, différences finis ou volumes finis [3], la méthode que nous proposons qui combine la méthode de dessaisonalisation Census X11 [4] et la méthode ARIMA-GARCH [5, 6] permet d'étudier en détail les différentes composantes du phénomène des marées telles que la saisonnalité, la tendance et cycles et les fluctuations résiduelles. En bref, notre méthode devrait nous permettre d'établir les tables des marées à Hell-ville et d'étudier en même temps l'évolution de différentes composantes des marées au court du temps. Nous aimerons également fournir au lecteur de cette publication une démarche rationnelle pour effectuer la modélisation Census X11 combinée avec la modélisation ARIMA-GARCH.

## **2. Matériel et méthodes**

### **2-1. Matériel**

Nous avons utilisé les matériels du laboratoire DyaCo, Matlab 2013a et l'utilitaire de Microsoft Office Excel 2007 pour établir et valider nos modèles. L'échantillon est composé de 35 286 individus : 17 643 données sur les pleines mers et 17 643 données sur les basses mers.

**2-2. Méthodes**

- *Représentation graphique des données initiales*

Elle permet de repérer graphiquement l'existence d'une saisonnalité, d'une tendance et d'une rupture de pente au niveau des chroniques [7].

- *Histogramme*

L'histogramme renseigne sur l'hétérogénéité des données. Pour une donnée hétérogène, nous devrions observer sur l'histogramme plusieurs modes [8].

- *Exposant de Hurst H* [9]

L'exposant de Hurst informe sur le degré de persistance des données. Selon Hurst, l'étendue  $R$ , l'écart type  $\sigma$ , et  $H$  sont reliés par la relation :

$$\frac{R(\tau)}{\sigma(\tau)} = \alpha \tau^H \tag{1}$$

La pente de la régression linéaire de  $\log\left(\frac{R}{\sigma}\right)$  sur  $\log \tau$  constitue donc une bonne approximation de  $H$

$$\bar{x} = \frac{1}{\tau} \sum_{i=1}^{\tau} x_i \tag{2}$$

$$y_j = x_j - \bar{x} \quad j \text{ variant de } 1 \text{ à } \tau \tag{3}$$

$$z_j = \sum_{i=1}^j y_i \quad j \text{ variant de } 1 \text{ à } \tau \tag{4}$$

$$R(\tau) = \max(z_j) - \min(z_j) \tag{5}$$

$$1 \leq j \leq \tau \quad 1 \leq j \leq \tau$$

$$\sigma(\tau) = \sqrt{\frac{1}{\tau} \sum_{i=1}^{\tau} (x_i - \bar{x})^2} \tag{6}$$

$\tau$  : Taille de l'échantillon prise égale à  $N, N/2, N/4$ ;  $N$  : Taille de l'échantillon total.

- Si  $0,5 < H$ , les données possèdent une mémoire courte ;
- Si  $0,5 > H$ , les données possèdent une mémoire longue ;
- Si  $0,5 = H$ , les données n'ont pas de mémoire.

- *Préparation des données*

Nous avons utilisé la méthode de la boîte à moustache pour détecter les points aberrants. Ce sont les points situés en dehors de l'intervalle  $[a, b]$  défini par [10] :

$$[a = Q1 - 1,5 * (Q3 - Q1), b = Q3 + 1,5 * (Q3 - Q1)] \tag{7}$$

$Q1$  : Premier quartile ;  $Q3$  : Troisième quartile.

Nous avons corrigé nos données en appliquant des moyennes mobiles pondérées définies par les relations suivantes :

$$X(t) = \sum_{i=t-1}^{i=t+1} c(i) * x(i) \quad \text{pour les données à l'intérieur de la chronique} \quad (8)$$

$$X(t) = \sum_{i=t}^{i=t+2} c(i) * x(i) \quad \text{pour les données situées au début de la chronique} \quad (9)$$

$$X(t) = \sum_{i=t-2}^{i=t} c(i) * x(i) \quad \text{pour les données se trouvant à la fin de la chronique} \quad (10)$$

$x(i)$  : Donnée à l'instant « i » ;  $c(i)$  : Coefficient de pondération égal à  $\frac{|b-x(i)|}{b-a}$  pour les points aberrants et égal à 1 pour les autres points ;  $a$  et  $b$  sont les bornes inférieures et supérieures limitant la boîte à moustache. Nous avons éliminé certains points à la fin de notre chronique puisque nous ne possédons pas les informations nécessaires pour les corriger.

- *Saisonnalité S*

Nous avons recherché la saisonnalité S en utilisant la méthode du corrélogramme et la méthode du périodogramme avec une fonction d'apodisation rectangulaire.

- *Méthode du Corrélogramme ou méthode de BLACKMAN-TUCKEY* [11]

Cette méthode permet d'estimer la densité spectrale en utilisant les formulations suivantes [12, 13] :

$$DSP(f) = \text{fft} [\hat{R}_{xx}(k)] \quad (11)$$

$$\hat{R}_{xx}(k) = \frac{1}{N-k} \sum_{t=0}^{N-k-1} x_{t+k} x_t^* \quad (12)$$

$$\text{fft}(k) = T_e \sum_{t=1}^N x_t e^{-2i\pi(t-1) \frac{(k-1)}{L}} \quad (13)$$

*fft* : Transformée de Fourier Rapide ; *DSP* : Densité Spectrale ; *N* : Nombre de points temporels ; *t* : Variable temporelle variant de 1 à N ;  $x_t$  : Données ; *L* : Nombre de points fréquentiels ; *k* : Variable fréquentielle variant de 1 à L ;  $\hat{R}_{xx}$  : Estimation de la fonction d'autocorrelation  $R_{xx}$  par sa forme non biaisée ;  $T_e$  : Période de l'échantillon ;  $F_e$  : Fréquence de l'échantillon.

- *Méthode du Périodogramme moyenné* [13]

Dans cette méthode, la DSP est estimée par la relation :

$$\widehat{DSP}(f) = \frac{1}{M} \sum_{i=1}^M |\text{fft}(k)|^2 \quad (14)$$

*M* : Nombre d'échantillons égal à  $L / T_e$

- *Périodogramme par la Méthode de l'Entropie Maximale* [14 - 18]

C'est une variante de la méthode du corrélogramme. Elle consiste à maximiser l'entropie ou la probabilité d'obtenir le maximum d'information sur les données. Selon [19], on peut obtenir une estimation de la DSP en utilisant la fonction d'autocorrélation et la fonction de l'entropie définies par :

$$E(x_{t+k} x_t^*) = \hat{R}_{xx}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} DSP(\omega) e^{jk\omega} d\omega \quad (15)$$

$$H = \int_{-\pi}^{\pi} \ln DSP(\omega) d\omega \tag{16}$$

$\omega$  : pulsation ;  $j$  : nombre complexe,  $j^2 = -1$ .

Pour un bruit blanc gaussien centré, en maximisant H, Burg arrivait à formuler la DSP de la manière suivante en utilisant un modèle autorégressif [20] :

$$DSP(\omega) = \frac{\sigma_q^2}{|1 + \sum_{i=1}^q \alpha_i e^{j i \omega}|^2} \tag{17}$$

$\sigma_q^2$  : variance du bruit blanc issu du model autoregressive d'ordre 'q' ;  $\alpha_i$  : Coefficients du model autoregressive d'ordre 'q'.

- *Saisonnalité S en terme journalier*

Notons que la saisonnalité S trouvée par les trois méthodes précédentes est fonction des temps de prélèvement. Pour le convertir en unité journalière, il suffit de la multiplier par la moyenne arithmétique des différences des temps de prélèvement successifs en unité journalière sur les pleines et les basses mers. Cette correspondance en terme journalier d'un Lags sur les figures serait définie par la relation suivante.

$$1 \text{ Lag} = \frac{\sum_{i=1}^N D_i}{N} \tag{18}$$

$D_i$  : Différence en jour entre deux temps de prélèvement successifs

Nous pourrons ensuite comparer la valeur des saisonnalités S trouvées par les trois méthodes à celles fournies par la théorie statique de Newton en utilisant le « test t de Student » qui consiste à vérifier l'égalité de la moyenne d'une série à une valeur donnée que nous détaillerons à la fin de la méthodologie quand nous testerons l'égalité de la moyenne des résidus normalisés à la valeur nulle.

- *Choix de la Saisonnalité S*

Comme la méthode du corrélogramme et celle du périodogramme donnent en général des valeurs non entières, il nous faudrait maintenant trouver la meilleure estimation en valeur entière de cette saisonnalité S pour pouvoir mettre en œuvre la méthode Census X-11. Nous avons donc eu recours à :

- ✓ Un test sur les résidus et à
- ✓ Une analyse de la variance à un facteur pour trouver S

- *Test sur les résidus [21]*

Cette méthode consiste à choisir la saisonnalité S créant moins de fluctuation résiduelle. (Propriété des moyennes mobiles)

- *Analyse de la Variance à un facteur*

Cette technique est utilisée si les données sont indépendantes, homoscedastes et possède une distribution normale ou approximativement normale. [22, 23]. Pour tester la normalité, nous avons utilisé le test de Jarque et Bera et deux autres tests basés sur le coefficient d'aplatissement et le coefficient d'asymétrie [24]. La distribution de la variable  $h$  est normale :

$$\text{Si } JB = \frac{N}{6} S_t^2 + \frac{N}{24} K_t^2 \leq 5,99 \text{ ou} \quad (19)$$

$$\text{Si } |S| = \sqrt{\frac{N}{6}} \cdot S_t \leq 1,96 \text{ ou} \quad (20)$$

$$\text{Si } |K| = \sqrt{\frac{N}{24}} \cdot K_t \leq 1,96 \quad (21)$$

$$S_t = \frac{\frac{1}{N} \sum_{i=1}^N (h_i - \bar{h}_i)^3}{\left[ \frac{1}{N} \sum_{i=1}^N (h_i - \bar{h}_i)^2 \right]^{3/2}} \quad (22)$$

$$K_t = \frac{\frac{1}{N} \sum_{i=1}^N (h_i - \bar{h}_i)^3}{\left[ \frac{1}{N} \sum_{i=1}^N (h_i - \bar{h}_i)^2 \right]^2} - 3 \quad (23)$$

Pour tester si une distribution s'approche de la loi normale, nous avons utilisé les relations entre l'écart type, l'écart moyen et l'écart probable d'une distribution normale définies par [25] :

$$ad = \frac{\sum_{i=1}^N |x(i) - \bar{x}|}{N} = 0,7979 * \sigma \quad (24)$$

$$pd = \frac{Q3 - Q1}{2} = 0,6745 * \sigma \quad (25)$$

$\sigma$  : Ecart type,  $ad$  : Ecart moyen,  $pd$  : Écart probable.

Si l'écart en valeur absolue entre le premier terme et le second terme des **Équations 24 et 25** reste inférieur à 5 %, nous avons décidé d'accepter la vraisemblance avec la loi normale. L'homoscedasticité a été testée par le test de Bartlett et par le test de Levene [26] :

Test de Bartlett : Si  $\frac{B}{C} < \chi^2 (l - 1)$ , nous concluons que les données sont homoscedastes

$$B = (N - l) \ln(s_p^2) - \sum_{i=1}^l (N_i - 1) \ln(s_i^2) \quad (26)$$

$$s_p^2 = \frac{\sum_{i=1}^l (N_i - 1) s_i^2}{n - l} \quad (27)$$

$$C = 1 + \frac{1}{3(l-1)} \left( \sum_{i=1}^l \frac{1}{N_i - 1} - \frac{1}{N - l} \right) \quad (28)$$

Test de Levene : Si  $D < Fisher (l - 1; n - l)$ , nous concluons que les données sont homoscedastes

$$D = \frac{(n-l) \sum_{i=1}^l N_i (\bar{E}_i - \bar{E})^2}{(l-1) \sum_{i=1}^l \sum_{j=1}^{N_i} (E_{i,j} - \bar{E}_i)^2} \quad (29)$$

$$E_{i,j} = |x_{i,j} - \bar{X}_l| \quad (30)$$

$$\bar{E}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} E_{i,j} \tag{31}$$

$$\bar{E} = \frac{1}{N} \sum_{i=1}^l N_i \bar{E}_i \tag{32}$$

$N$  : Taille totale égale à  $N_i * l$  ;  $l$  : Nombre de groupes égal à  $S$  ;  $N_i$  : Taille de chaque groupe ;  $S_i^2$  : Variance de chaque groupe correspondant au colonne «  $i$  » ;  $x_{i,j}$  : Variable correspondant au colonne «  $i$  », ligne «  $j$  » ;  $\bar{X}_i$  : Moyenne du groupe «  $i$  » ; Fisher  $(l - 1; n - l)$  : Valeur de  $F$  sur la table de Fisher pour «  $(l - 1; n - l)$  » degré de liberté ;  $\chi^2 (l - 1)$  : Valeur de Khi-2 sur table avec «  $l-1$  » degré de liberté

Pour le choix de  $S$  non entière, nous avons maximisé la fonction  $(1/f, DSP)$ . Le domaine de définition de  $f$  a été estimé en tenant compte de la parité de la fonction  $DSP$  et du mode d'itération sous Matlab qui commence toujours par 1. Nous avons alors défini ce domaine par l'intervalle [27] :

$$f = \frac{F_{Nyquist}}{\frac{L}{2}} * \left[1, \frac{L}{2}\right] = \frac{F_e}{\frac{L}{2}} * \left[1, \frac{L}{2}\right] \tag{33}$$

$F_{Nyquist}$  : Fréquence de Nyquist.

• **Analyse de la tendance [28]**

L'analyse de l'existence d'une tendance monotonique dans notre travail a été basée sur la mise en œuvre du test de corrélation des rangs de Kendall. Nous avons retenu la définition du coefficient de Kendall  $\tau_b$  suivante :

$$\tau_b = \frac{S''}{\sqrt{\frac{N(N-1)}{2} - E_x} \sqrt{\frac{N(N-1)}{2} - E_y}} \tag{34}$$

$$E_x = \sum_{x=1}^{m1} \frac{N_x (N_x - 1)}{2} \text{ et } E_y = \sum_{y=1}^{m2} \frac{N_y (N_y - 1)}{2} \tag{35}$$

$m1$  : nombre de groupes d'ex-aequos sur les rangs  $R_x$  du grandeur  $X$  ;  $N_x$  : nombre d'ex-aequos contenu dans le  $x$ -ième groupe ;  $m2$  : nombre de groupes d'ex-aequos sur les rangs  $R_y$  du grandeur  $Y$  ;  $N_y$  : nombre d'ex-aequos contenu dans le  $y$ -ième groupe.

Pour le terme  $S''$  correspondant au changement de signe de la différence des rangs, nous avons utilisé la formulation suivante :

$$S'' = \sum_{i < j}^N \text{signe}(X_i - X_j) \text{signe}(Y_i - Y_j) \text{ avec } \text{signe}(G) = \begin{cases} +1 & \text{si } G > 0 \\ 0 & \text{si } G = 0 \\ -1 & \text{si } G < 0 \end{cases} \tag{36}$$

Sous l'hypothèse  $H_0$  : « il n'y a pas de corrélation entre  $X$  et  $Y$  », la moyenne et la variance de  $\tau_b$  peuvent être estimées par les relations suivante :

$$E(\tau_b) = 0 \text{ et } \tag{37}$$

$$\text{Var}(\tau_b) = \frac{2(2N+5)}{9N(N-1)} \quad (38)$$

En particulier, pour un grand échantillon, la règle de décision est : Si la valeur absolue de  $Z_1 = \frac{\tau_b - E(\tau_b)}{\sqrt{\text{Var}(\tau_b)}}$  est supérieure à la valeur critique  $U_{0,1,0,05}$  issue de la loi normale centrée- réduite  $N(0,1)$ , nous rejetons l'hypothèse  $H_0$  en faveur de l'hypothèse  $H_1$  : « il existe une relation monotone entre les deux grandeurs  $X$  et  $Y$  ».

- *Schéma de prévision* [29]

En général, il existe deux schémas de prévision :

Le Schéma additif :  $x_t = t_t + s_t + r_t$  (39)

Le Schéma multiplicatif :  $x_t = t_t * s_t * r_t$  (40)

$t_t$  : *Tendance* ;  $s_t$  : *Composante saisonnière* ;  $r_t$  : *Résidus ou innovation*.

Nous avons testé les trois méthodes suivantes pour reconnaître le schéma adéquat [30] :

- Méthode du profil ;
- Méthode de la bande ;
- Méthode analytique.

➤ *Méthode du profil*

Elle consiste à tracer les graphiques des données pour chaque période.

- Si les courbes sont confondues ou parallèles, on choisit le schéma multiplicatif ;
- Si les courbes s'entrecroisent, on adopte un schéma additif.

➤ *Méthode de la bande*

La méthode de la bande consiste à construire pour chaque période, d'une part la courbe des minima et d'autre part la courbe des maxima.

- Si les deux courbes sont parallèles, on opte pour le schéma multiplicatif ;
- Si les deux courbes s'entrecroisent, on choisit un schéma additif.

➤ *Méthode analytique*

La méthode analytique consiste à régresser pour chaque période l'écart type  $\sigma_i$  par rapport à la moyenne  $\bar{x}_i$ . Si le terme  $\mu$  dans l'équation  $\sigma_i = \mu \bar{x}_i + g$  est significatif, c'est à dire si la valeur absolue de la statistique de Student correspondant à  $\mu$  après regression est supérieure à la valeur critique dans la table de Student, on choisit le schéma multiplicative.

- *Méthode Census X11* [4]

Les résultats de l'analyse de la saisonnalité et de la tendance nous informent sur les différentes composantes de notre schéma de prévision. Avant la mise en œuvre la méthode Census X-11, nous avons dû stabiliser notre série en appliquant une transformation logarithmique de base 10 qui a également permis de transformer notre schéma multiplicatif en schéma additif.

➤ *Tendance  $t_{t,1}$  par moyenne mobile centrée d'ordre S [29]*

La tendance a été estimée en premier lieu en utilisant une moyenne mobile centrée d'ordre S. Si S est paire, c'est-à-dire de la forme  $2m$ , on utilise la relation :

$$t_{t,1} = \frac{1}{2m} \left( \frac{1}{2}x_{t-m} + x_{t-m+1} + \dots + x_{t+m-1} + \frac{1}{2}x_{t+m} \right) = M_{2m}(x_t) \quad (41)$$

Si S est impaire, c'est-à-dire de la forme  $2m + 1$ , on utilise la relation :

$$t_{t,1} = \frac{1}{2m+1} (x_{t-m} + x_{t-m+1} + \dots + x_{t+m-1} + x_{t+m}) = M_{2m+1}(x_t) \quad (42)$$

*m : un entier positif*

➤ *Saisonnalité-Innovation :  $S_{t,1} + r_{t,1}$  [4]*

Le terme Saisonnalité-Innovation est obtenu par différence entre les données initiales et la tendance

$$S_{t,1} + r_{t,1} = x_{t,1} - t_{t,1} \quad (43)$$

La saisonnalité  $S_{t,1}$  se déduit en appliquant une moyenne mobile 5 termes « 3\*3 » à cette valeur de la Saisonnalité-Innovation puis en soustrayant au résultat de  $M_{3*3}(x_{t,1} - t_{t,1})$  la valeur correspondant à sa moyenne mobile centrée d'ordre S.  $S_{t,1}$  s'obtient alors en utilisant la **Formule** :

$$S_{t,1} = [M_{3*3}(x_{t,1} - t_{t,1})] - M_{2m+1}[M_{3*3}(x_{t,1} - t_{t,1})] \quad (44)$$

$M_{3*3}$  : moyenne mobile centrée à 5 termes de coefficient de pondération égal à  $\left\{ \frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{2}{9}, \frac{1}{9} \right\}$

➤ *Série corrigée des variations saisonnières  $SCVS_{t,1}$  [4]*

La  $SCVS_{t,1}$  se déduit en appliquant la relation :

$$SCVS_{t,1} = x_{t,1} - S_{t,1} \quad (45)$$

➤ *Tendance  $t_{t,2}$  par moyenne mobile de Henderson [4]*

La tendance est ici estimée en utilisant une moyenne mobile de Henderson à 9 ou 13 termes. Donc nous pouvons écrire :

$$t_{t,2} = H_{13 \text{ ou } 9}(SCVS_{t,1}) \quad (46)$$

Le principe est simple.

- Si le rapport  $\frac{I}{C}$  est supérieur à 1, nous appliquons une moyenne mobile de Henderson à 13 termes ;
- Si le rapport  $\frac{I}{C}$  est inférieur à 1, nous appliquons une moyenne mobile de Henderson à 9 termes.

I et C sont donnés par les **Formules** suivantes :

$$C = \frac{1}{n-1} \sum_{i=2}^n (t_{i,2} - t_{i-1,2}) \quad (47)$$

$$I = \frac{1}{n-1} \sum_{i=2}^n (scvs_{i,1} - t_{i,2}) - (scvs_{i-1,1} - t_{i-1,2}) \quad (48)$$

➤ *Nouvelle expression de la Saisonnalité-Innovation :  $s_{t,2} + r_{t,2}$  [4]*

Après connaissance de l'estimation de la tendance  $t_{t,2}$ , nous avons réestimé la Saisonnalité-Innovation en utilisant la relation :

$$s_{t,2} + r_{t,2} = x_{t,1} - t_{t,2} \quad (49)$$

Et le terme  $s_{t,2}$  s'obtient en utilisant la relation :

$$s_{t,2} = [M_{3*5}(x_{t,1} - t_{t,2})] - M_{2m+1}[M_{3*5}(x_{t,1} - t_{t,2})] \quad (50)$$

$M_{3*5}$  : moyenne mobile centrée à 7 termes de coefficient de pondération égal à  $\left\{ \frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{3}{15}, \frac{3}{15}, \frac{2}{15}, \frac{1}{15} \right\}$

➤ *Nouvelle expression de la série corrigée des variations saisonnière  $scvs_{t,2}$  [4]*

Finalement, nous obtenons la  $scvs_{t,2}$  en utilisant la relation :

$$scvs_{t,2} = x_{t,1} - s_{t,2} \quad (51)$$

Dans cette étude, l'intervalle de la boîte à moustache a été estimé en utilisant toutes les données.

Par contre le logiciel Censur X11 utilise en général :

- Les données appartenant à 2 saisonnalités antérieures au point considéré ;
- Les données appartenant à la saisonnalité du point considéré et ;
- Les données appartenant à 2 saisonnalités futures au point considéré.

• *Méthodologie de Box et Jenkins*

La méthodologie expliquée dans cette étude concerne la modélisation ARIMA (p + 1, d, q) de Box et Jenkins. Dans ce modèle « q » correspond à l'ordre de la partie à Moyenne Mobile, « d » à l'ordre d'Intégration et « p + 1 » à l'ordre de la partie Auto-Régressive. Le problème se ramène donc à la recherche de p, d et q

➤ *Identification de d, q et p*

Box et Jenkins utilisaient la formulation suivante pour éliminer la tendance et stationnariser la série [31]

$$r_t' = (I - B)^d scvs_{t,2} \quad (52)$$

$B$  : Opérateur de retard ;  $r_t'$  : Processus stationnaire.

Nous avons fixé un ordre assez élevé de retard « p » avant de mettre en œuvre la stratégie de Dickey Fuller Augmenté (DFA) en variant « d entre 0 et 2 ». Si le test de Dickey Fuller Augmenté confirme la stationnarité de la série différenciée alors le « d et p » choisis correspondent à l'ordre d'intégration adéquat  $d_{adéquat}$  et à un

nombre de retard maximal acceptable «  $p_{max}$  ». Si le test de Dickey Fuller Augmenté rejette l'hypothèse de stationnarité alors nous devrions augmenter le nombre de retard d'une ou de plusieurs unités et répéter l'opération qui cette fois ci devrait confirmer la stationnarité au sens de Dickey Fuller Augmenté. Cette nouvelle valeur de «  $p$  » correspondra alors à l'ordre maximal «  $p_{max}$  » du modèle DFA. Une fois  $p_{max}$  connu, on détermine le nombre de retard optimal «  $p_{optimal}$  » des modèles DFA, en fixant «  $d$  égal à  $d_{adéquat}$  et  $p$  égal à  $p_{max}$  », puis en variant progressivement le «  $p$  » dans chaque modèle entre 0 et  $p_{max}$  et enfin en minimisant le critère de Schwarz défini par la relation [32] :

$$SC(p + q) = N * \log(\sigma_{\epsilon t}^2) + (p + q + 3) * \log(N) \tag{53}$$

$\sigma_{\epsilon t}^2$  : *Variance des résidus.*

Avec les valeurs de «  $d_{adéquat}$  et de  $p_{optimal}$  » précédents, nous pourrions maintenant rechercher les valeurs de  $q_{max}$  et de  $q_{optimal}$  en augmentant progressivement  $q$  d'une unité. Le  $q_{max}$  correspond à la valeur maximale de  $q$  rendant non corrélés les résidus au test de Ljung-Box et le  $q_{optimal}$  n'est autre que la valeur de  $q$  compris entre 0 et ce  $q_{max}$  minimisant le critère d'Information d'Akaike défini par la relation [32] :

$$AIC(p + q) = N * \log(\sigma_{\epsilon t}^2) + 2(p + q) \tag{54}$$

Le test de Ljung-Box consiste à comparer la p-Value de la statistique de Ljung-Box à la valeur 0,05 [31]

$$\chi_m^2 = N(N + 2) \sum_{k=1}^m \left[ \frac{\alpha_k^2(r_t)}{N-k} \right] \tag{55}$$

$\alpha_k$  : *Estimation de l'autocorrelation*

Si la p-Value ou  $Prob [\chi^2(m - a) > \chi_m^2]$  est supérieure ou égale à 0,05 pour un ordre «  $k$  » fixé, la série est stationnaire. Finalement, connaissant  $p_{optimal}$ ,  $d_{adéquat}$  et  $q_{optimal}$ , nous pouvons estimer par maximum de vraisemblance les modèles Arima ( $p_{optimal} + 1, d_{adéquat}, q_{optimal}$ ) équivalents à nos modèles DFA et procéder en même temps à l'élimination des coefficients non significatifs. Le test de Dickey Fuller Augmenté consiste à appliquer la stratégie suivante [32]:

$$\text{Modèle DFA 1 : } r_t' = \Phi r'_{t-1} + \sum_{j=1}^p \Phi_j \Delta r'_{t-j} + \epsilon_t \tag{56}$$

$$\text{Modèle DFA 2 : } r_t' = \Phi r'_{t-1} + a + \sum_{j=1}^p \Phi_j \Delta r'_{t-j} + \epsilon_t \tag{57}$$

$$\text{Modèle DFA 3 : } r_t' = \Phi r'_{t-1} + bt + a + \sum_{j=1}^p \Phi_j \Delta r'_{t-j} + \epsilon_t \tag{58}$$

La règle suggère de toujours commencer par tester le modèle DFA 3.

✓ Hypothèse 1 :

Ho :  $\Phi = 1$ , il existe une racine unitaire

H1 :  $|\Phi| < 1$ , il n'y a pas de racine unitaire, la série est stationnaire

✓ Décision :

- Si la statistique de la loi t de Student pour  $\Phi$  est inférieur à la statistique de Dickey-Fuller sur table, on rejette Ho ;
- Si la statistique de la loi t de Student pour  $\Phi$  est supérieur à la statistique de Dickey-Fuller sur table, on accepte Ho.

Si la série est stationnaire au sens de Dickey Fuller Augmenté, on teste la significativité de la tendance « b » et de la constante « a » en utilisant la loi t de Student défini par :

✓ Hypothèse 2 :

Ho : « b » = 0, Pas de tendance

H1 : « b » ≠ 0, Existence de tendance

✓ Décision :

- Si la statistique de la loi t de Student pour « b » est supérieure à 1,96, on rejette Ho ;
- Si la statistique de la loi t de Student pour « b » est inférieure à 1,96, on accepte Ho.

La significativité de la constante « a » se confirme de la même manière que pour la tendance avec comme hypothèse :

Ho : « a » = 0, Pas de constante

H1 : « a » ≠ 0, Existence du terme constant dans le modèle

➤ *Analyse des Résidus*

L'hétéroscédasticité a été testé en utilisant le test ARCH sous Matlab 2013a défini par [12]

$$r_t^2 = \pi_0 + \sum_{k=1}^W \pi_k r_{t-k}^2 + e(t) \quad (59)$$

$r_t$ : Vecteur sur les résidus

Elle est confirmée si la p-value issue du test ARCH est inférieure à 0,05. En cas d'hétéroscédasticité, les résidus seront modélisés par un modèle de type ARCH ou EGARCH.

➤ *ARCH model*[12] :

$$r_t = \eta_t \sigma_t \quad (60)$$

$$\sigma_t^2 = K + \sum_{j=1}^Q \alpha_j r_{t-j}^2 \quad (61)$$

➤ *EGARCH model*[8] :

$$r_t = \eta_t \sigma_t$$

$$\log \sigma_t^2 = K + \sum_{i=1}^P \gamma_i \sigma_{t-i}^2 + \sum_{j=1}^Q \alpha_j \left[ \frac{|r_{t-j}|}{\sigma_{t-j}} - E \left\{ \frac{|r_{t-j}|}{\sigma_{t-j}} \right\} \right] + \sum_{j=1}^Q \xi_j \left( \frac{r_{t-j}}{\sigma_{t-j}} \right) \quad (62)$$

$$E \left\{ \frac{|r_{t-j}|}{\sigma_{t-j}} \right\} = E \{ |\eta_{t-j}| \} = \sqrt{\frac{2}{\pi}} \text{ pour une distribution gaussienne des résidus}$$

$$E \left\{ \frac{|r_{t-j}|}{\sigma_{t-j}} \right\} = E \{ |\eta_{t-j}| \} = \sqrt{\frac{\nu-2}{\pi}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \text{ pour une distribution de Student avec } \nu \geq 2 \text{ degré de liberté}$$

$K$  : constante du modèle ;  $\gamma_i$  : coefficients GARCH ;  $\alpha_j$  : coefficients ARCH ;  $\xi_j$  : Leverage.

La **normalité** de la distribution des résidus standardisés sera confirmée ou rejetée en appliquant l'une des définitions que nous avons utilisées pour confirmer la normalité dans l'analyse de la variance à un facteur. Pour vérifier que la distribution des résidus standardisés est centrée et réduite, il suffit de tester l'égalité entre les paramètres  $(\bar{r}_t, \sigma^2)$  et  $(0, 1)$ . Pour vérifier l'égalité entre la moyenne des résidus normalisés  $\bar{r}_t$  et la valeur zéro (0), nous avons appliqué le test d'égalité de la moyenne d'une série à une valeur donnée ou test de Student. Cette technique permet de déterminer l'intervalle de confiance symétrique à 95 % pour zéro [25]. En notant  $[c, d]$  cet intervalle la règle de décision est la suivante :

- Si  $0 \in [c, d] = [\bar{r}_t - t_{1-\alpha/2} \frac{s}{\sqrt{N}}, \bar{r}_t + t_{1-\alpha/2} \frac{s}{\sqrt{N}}]$ , alors la moyenne  $\bar{r}_t$  est non significativement différente de zéro ;
- Si  $0 < \bar{r}_t - t_{1-\alpha/2} \frac{s}{\sqrt{N}}$  ou  $\bar{r}_t + t_{1-\alpha/2} \frac{s}{\sqrt{N}} > 0$  alors la moyenne  $\bar{r}$  est significativement différente de zéro.

$\bar{r}_t$  : Moyenne des résidus normalisés ;  $\sigma$  : Écart type biaisé de l'échantillon ;  $s$  : Ecart type non biaisé égal à  $\sigma * \frac{N}{N-1}$  ;  $N$  : Taille de l'échantillon ;  $t_{1-\alpha/2}$  : Variable de Student à  $N-1$  degré de liberté pour un risque de première espèce  $\alpha$ .

De la même manière, on teste l'égalité entre la variance des résidus normalisés et un (1) [25] : La règle de décision est la suivante

- Si  $1 \in [s^2 - t_{1-\alpha/2} s^2 \sqrt{\frac{2}{N}}, s^2 + t_{1-\alpha/2} s^2 \sqrt{\frac{2}{N}}]$ , alors la variance  $s^2$  est non significativement différente de un ;
- Si  $1 < s^2 - t_{1-\alpha/2} s^2 \sqrt{\frac{2}{N}}$  ou  $s^2 + t_{1-\alpha/2} s^2 \sqrt{\frac{2}{N}} > 1$  alors la variance  $s^2$  est significativement différente de un.

➤ *Vérification de la qualité de la prédiction et de la prévision* [33, 34]

Cette étape consiste à comparer les erreurs de prédiction sur les données initiales et les erreurs de prévision sur les données futures avec celles d'une méthode alternative en utilisant : L'indicateur de THEIL et l'Erreur Moyenne Absolue en Pourcentage (EMAP) sont définis par les **Formules** suivantes :

$$I_{THEIL} = \frac{RMSE_{modèle}}{RMSE_{méthode\ alternative}} \tag{63}$$

$$RMSE = \sqrt{\frac{1}{N_t} \sum_{t=1}^{N_t} (x_t - x_{modèle})^2} \tag{64}$$

$$EMAP = \frac{\sum_{t=1}^{N_t} \left| \frac{x_t - x_{modèle}}{x_t} \right| * 100}{N_t} \tag{65}$$

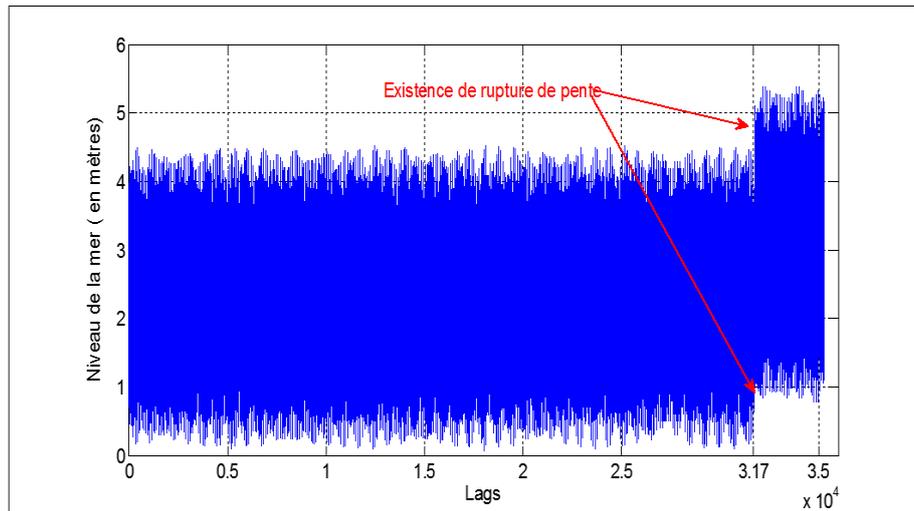
$N_t$  : Taille de l'échantillon ;  $x_t$  : Donnée initiale à l'instant  $t$  ;  $x_{modèle}$  : Valeur prédite de  $x_t$  à l'instant  $t$   
 $N_t$  : Taille de l'échantillon sur laquelle on effectue la projection.

Si  $I_{THEIL}$  est inférieur à 1, le modèle offre une meilleure précision par rapport à la méthode alternative pendant la durée de comparaison. Le modèle est bon si l'EMAP est inférieure à 10 % et il devient excellent pour les valeurs en dessous de 5 %.

### 3. Résultats et discussion

#### 3-1. Représentation graphique des données initiales

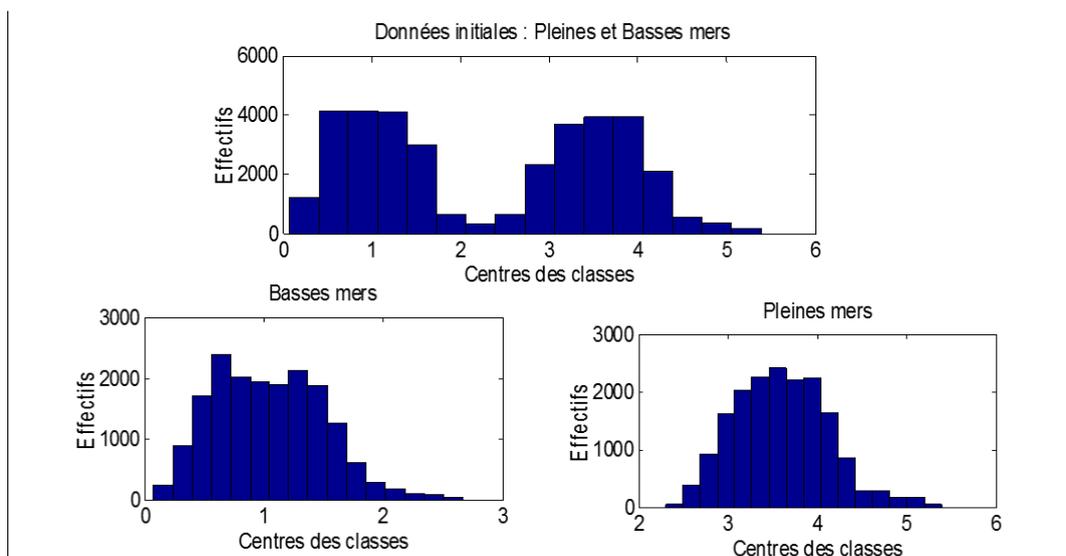
Sur la **Figure 1**, nous observons une rupture de pente autour du 31 700 Lags. Soit les prélèvements sont erronés, soit un facteur climatique à engendré cette évolution. Mais comme nous n'avons pas les informations sur le protocole de prélèvement des données au Centre National Océanographique de Hell-ville, nous avons préféré enlever ces données aberrantes de nos chroniques pour l'établissement des modèles.



**Figure 1 : Données initiales**

#### 3-2. Histogramme

Sur la **Figure 2**, l'histogramme des données initiales est bimodal et ceux des données séparées (pleines mers d'une part et basses mers d'autre part) sont unimodaux. Il est donc préférable de travailler sur les données séparées pour contourner le problème d'hétérogénéité entre les prélèvements issus des pleines mers et ceux issus des basses mers.

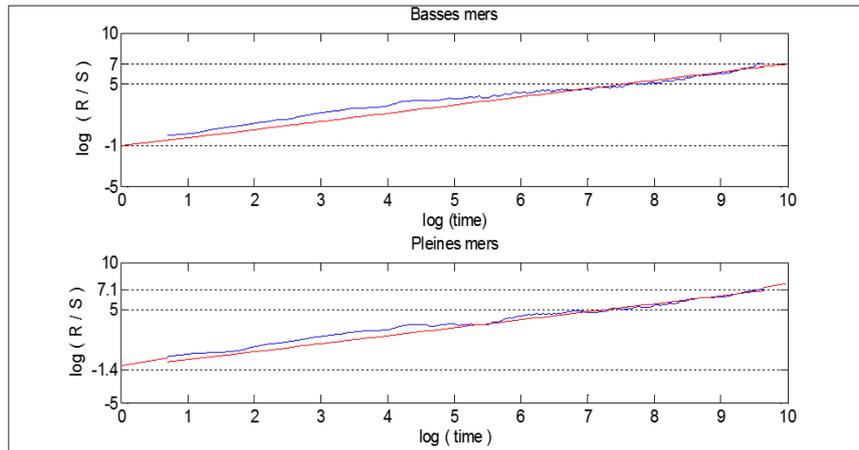


**Figure 2 : Histogrammes**

Par conséquent, nous avons décidé d'effectuer séparément la modélisation des pleines mers et la modélisation des basses mers.

### 3-3. Exposant de Hurst

Sur la **Figure 3**, nous avons trouvé pour la série corrigée des variations saisonnières un exposant de Hurst  $H$  égal à 0,8020 pour les basses mers et  $H$  égal à 0,8510 pour les pleines mers.



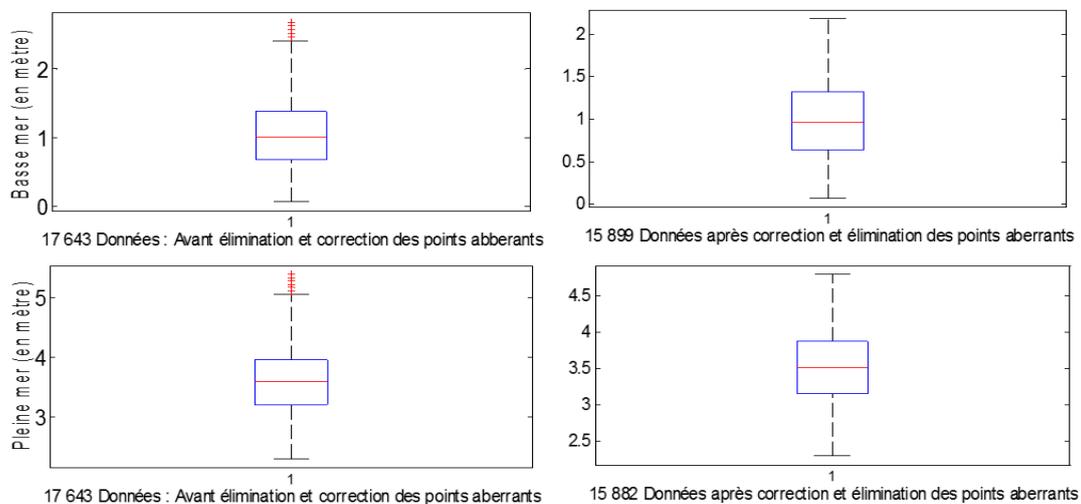
**Figure 3 : Exposants de Hurst**

Avant d'appliquer directement une modélisation FARIMA propre aux série à mémoire longue et qui est plus compliquée à mettre en œuvre, nous allons tester si une modélisation ARIMA ( $p + 1, d, q$ ) marche et suffit à prévoir les pleines et les basses mers dans notre zone d'étude.

### 3-4. Préparation des données

La **Figure 4** montre que nous avons dû :

- Eliminer 1 744 données sur les basses mers pour blanchir notre série des points aberrants et ;
- Eliminer 1 761 données sur les pleines mers pour blanchir notre chronique des points aberrants.

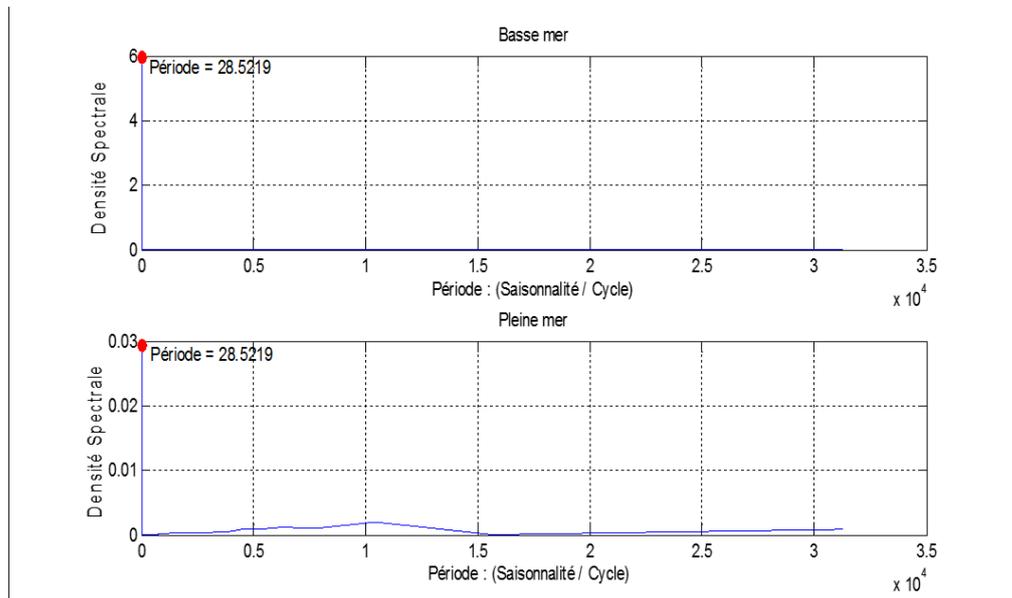


**Figure 4 : Boites à moustache avant et après correction des points aberrants**

**3-5. Saisonnalité S**

➤ *Corrélogramme ou méthode de BLACKMAN-TUCKEY*

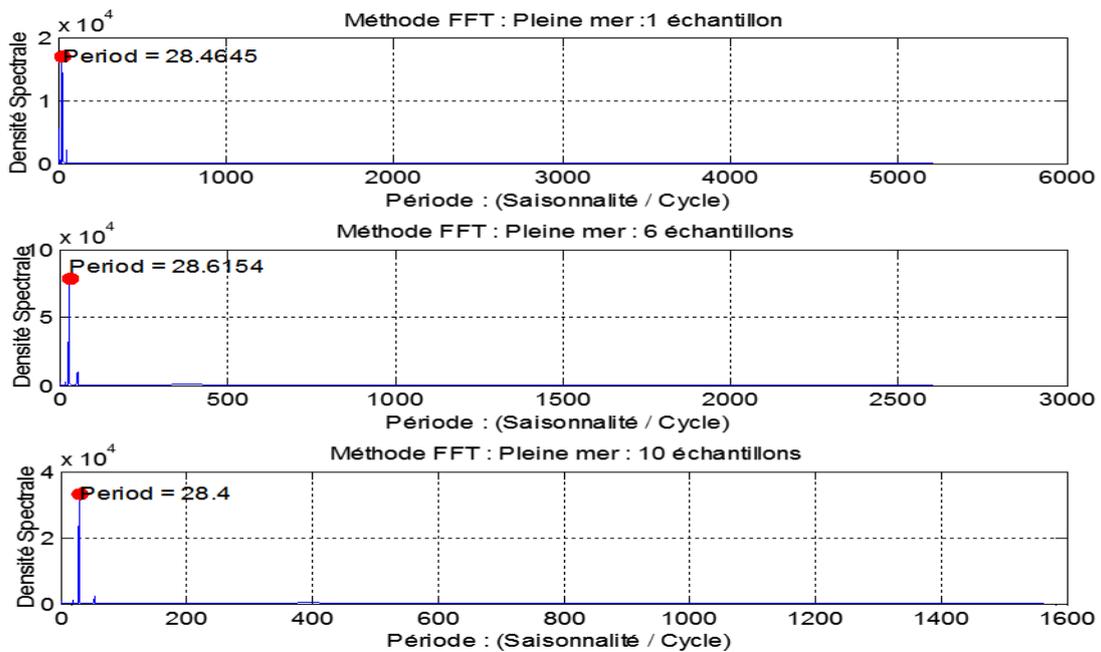
La méthode du corrélogramme permet de déduire sur la **Figure 5** la même saisonnalité S qui est égale à 28.5219 Lags pour les pleines mers et pour les basses mers.



**Figure 5 : Corrélogramme des séries stabilisées**

➤ *Périodogramme moyenné*

Sur la **Figure 6 et 7**, les périodogrammes moyennés des pleines et basses mers donnent la même saisonnalité S comprise entre 28.4 et 28.6154 Lags



**Figure 6 : Périodogramme moyenné : pleines mers**

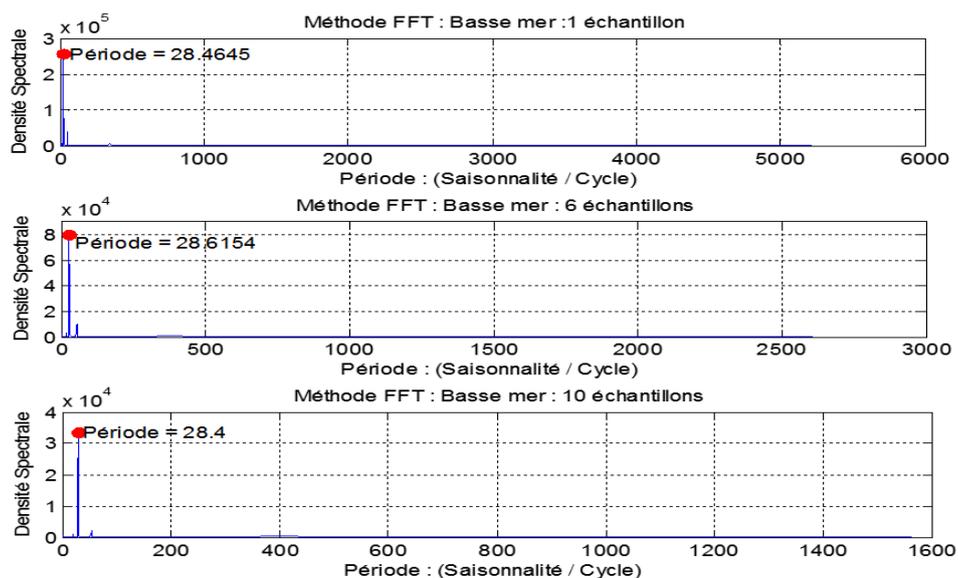


Figure 7 : Périodogramme moyenné : basses mers

➤ Périodogramme par la Méthode de l'Entropie Maximale

Sur la Figure 8, nous trouvons par la méthode de l'entropie maximale que la saisonnalité  $S$  égale à 28.4 Lags pour les basses mers et pour les pleines mers.

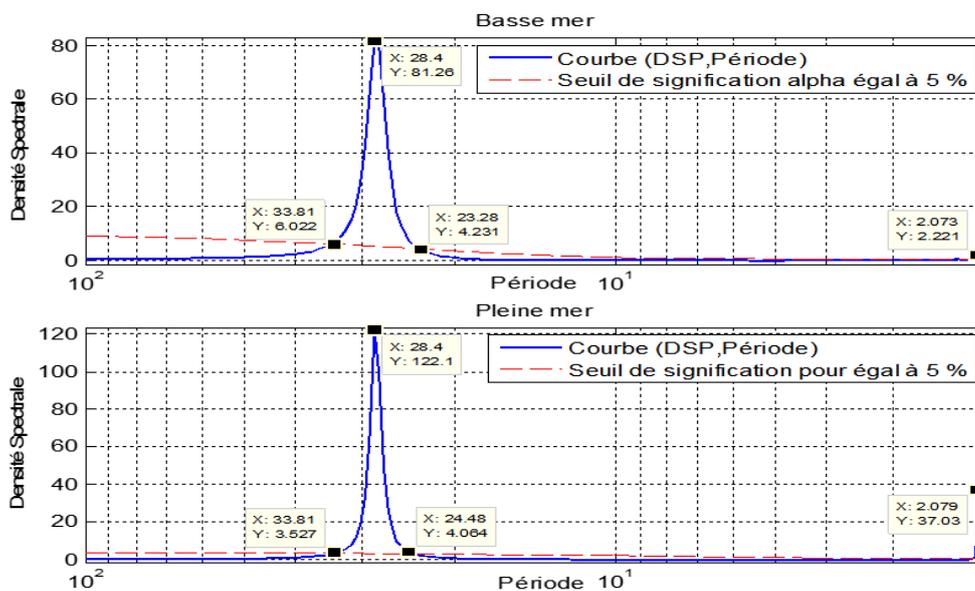


Figure 8 : Périodogramme par la méthode de l'entropie maximale : pleines mers et basses mers

Sur les 10 valeurs que nous avons trouvées pour  $S$ , 9 résultats nous suggèrent de prendre comme meilleure estimation de la Saisonnalité  $S$  la valeur entière 28 ou 29 Lags. Nous admettrons donc que l'estimation de la saisonnalité est soit 28 Lags ou 29 Lags.

3-6. Saisonnalité  $S$  en terme journalier

En passant à la moyenne de la différence en jour entre deux temps de prélèvement consécutifs, nous avons trouvé qu'1 Lag correspond à 0,5175313 jour pour les pleines mers et à 0,51753154 jour pour

les basses mers. En portant ces valeurs aux 10 saisonnalités  $S$  en Lags que nous avons trouvées pour les pleines et basses mers et en mettant en œuvre le test  $t$  de Student, nous avons pu dresser le **Tableau 1**.

**Tableau 1 : Test d'égalité entre la saisonnalité obtenue par les méthodes traditionnelles et par nos méthodes**

	Ancienne Méthode	Nos méthodes	Intervalle de confiance et paramètres	Décision
<b>Saisonnalité : Morte eau, vive eau (longue période)</b>				
Pleine mer	$S = 14,765$ jours	$SN_1 = \begin{cases} S = 14,76097 \text{ jours} \\ S = 14,69789 \text{ jours} \\ S = 14,73127 \text{ jours} \\ S = 14,80936 \text{ jours} \end{cases}$	Moyenne $SN_1 = 14,74988$ Ecart type $SN_1 = 0,05461$ $t_{1-\frac{\alpha}{2}} = 3,182$ $S \text{ et } SN_1 \in [14,6629, 14,8367]$	Il y a concordance parfaite entre $S$ et les valeurs de $SN_1$ au seuil $\alpha = 0,05$
Basse mer	$S = 14,765$ jours	$SN_2 = \begin{cases} S = 14,76098 \text{ jours} \\ S = 14,69789 \text{ jours} \\ S = 14,73127 \text{ jours} \\ S = 14,80937 \text{ jours} \end{cases}$	Moyenne $SN_2 = 14,74988$ Ecart type $SN_2 = 0,05461$ $t_{1-\frac{\alpha}{2}} = 3,182$ $S \text{ et } SN_2 \in [14,6629, 14,8367]$	Il y a concordance parfaite entre $S$ et les valeurs de $SN_2$ au seuil $\alpha = 0,05$
<b>Saisonnalité : Diurne</b>				
Pleine mer	$S = 1,034$ jours	$SN_3 = \begin{cases} 1,06184 \text{ jours} \\ 1,07594 \text{ jours} \end{cases}$	Moyenne $SN_3 = 1,06889$ Ecart type $SN_3 = 0,00997$ $t_{1-\frac{\alpha}{2}} = 12,706$ $S \text{ et } SN_3 \in [0,94219, 1,19560]$	Il y a concordance parfaite entre $S$ et les valeurs de $SN_3$ au seuil $\alpha = 0,05$
Basse mer	$S = 1,034$ jours	$SN_4 = \begin{cases} 1,06443 \text{ jours} \\ 1,07284 \text{ jours} \end{cases}$	Moyenne $SN_4 = 1,06863$ Ecart type $SN_4 = 0,00594$ $t_{1-\frac{\alpha}{2}} = 12,706$ $S \text{ et } SN_4 \in [0,99307, 1,14419]$	Il y a concordance parfaite entre $S$ et les valeurs de $SN_4$ au seuil $\alpha = 0,05$

Nous avons pris la valeur de la première ligne du vecteur  $SN_3$  et du vecteur  $SN_4$  égale à la moyenne des saisonnalités significatives trouvées après l'application de la méthode de l'entropie maximale. Les résultats du **Tableau 1** montre qu'il existe une concordance parfaite entre les saisonnalités  $S$  longue période et diurne obtenues par nos méthodes et celles calculées par les méthodes traditionnelles, en particulier par la théorie statique de Newton. Nous pouvons affirmer également que seule la Méthode de l'Entropie Maximale (MEM) permet de détecter la saisonnalité diurne. Ce qui semble évident d'après le théorème de Nyquist-Shannon [27] qui stipule que pour des échantillons prélevées tous les  $S$  égale à 12 heures 25 minutes, nous ne pourrions à détecter que les saisonnalités  $S$  de valeur supérieure ou égalé à 24 heures 50 minutes. Nous avons aussi remarqué qu'il suffit d'interpoler la valeur de la saisonnalité diurne pour obtenir la période semi-diurne.

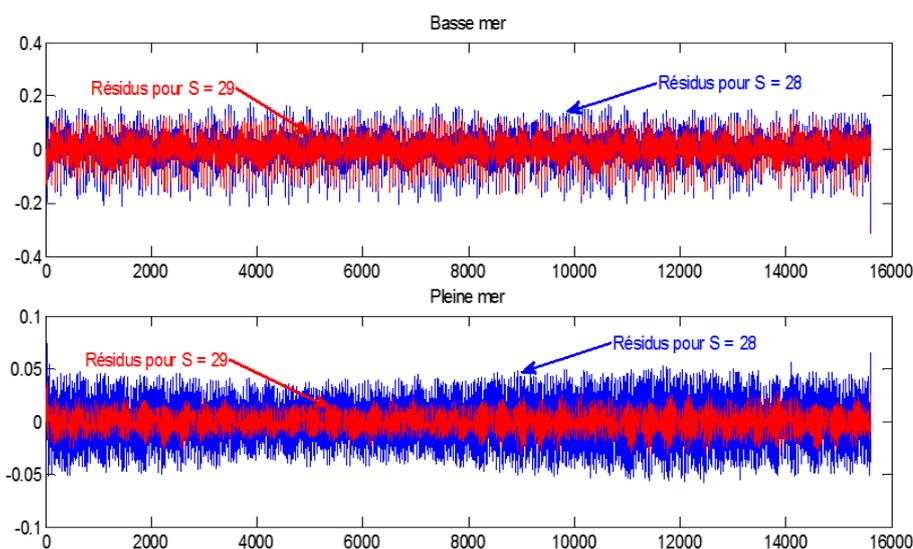
### 3-7. Choix de la Saisonnalité $S$

Pour pouvoir mettre en œuvre le test sur les résidus et le test de Fisher, nous avons dû transformer nos données vectorielles sur les pleines mers et sur les basses mers en matrice :

- B à 558 lignes \* 28 colonnes et en matrice ;
- A à 539 lignes \* 29 colonnes.

➤ *Test sur les Résidus*

Les tests sur les résidus que nous pouvons observer sur la **Figure 9** recommandent aussi bien pour les pleines mers que pour les basses mers de choisir la saisonnalité  $S$  égale à 29.



**Figure 9 : Courbes des résidus pour  $S = 28$  et  $S = 29$**

Cette méthode semble suffisante pour trancher sur le choix de la meilleure saisonnalité  $S$  à prendre pour nos prévisions.

➤ *Analyse de la Variance à un Facteur*

Les résultats du **Tableau 2** confirment la possibilité d'utilisation du test de Fisher pour étudier la stabilité de nos chroniques :

**Tableau 2 : Indépendance, homoscedasticité et normalité**

	Basse mer		Pleine mer	
L'indépendance des données résultent de l'indépendance des prélèvements des niveaux de la mer				
Homoscedasticité acceptée	Bartlett	$\frac{B}{C} = 5,85 < \chi^2(28) = 41,34$	Bartlett	$\frac{B}{C} = 9,158 < \chi^2(28) = 41,34$
	Levene	$D = 0,37 < F(28, 538) = 1,5$	Levene	$D = 3,198 < F(28, 538) = 1,5$
Normalité	Rejet de la normalité	$1,96 \leq \min_{1 \leq i \leq 29} (JB_i) = 6$ $1,96 \leq \min_{1 \leq i \leq 29}  S_i  = 8,8$ $1,96 \leq \min_{1 \leq i \leq 29}  K_i  = 2,25$	Rejet de la normalité	$1,96 \leq \min_{1 \leq i \leq 29} (JB_i) = 6,84$ $ S_i _{1 \leq i \leq 29} \leq 1,96$ pour 13 « i » sur 29 $ K_i _{1 \leq i \leq 29} \leq 1,96$ pour 9 « i » sur 29
		Vraisemblance à la loi normale : oui		$0,0018 \leq  ad - 0,7979\sigma  \leq 0,01$ $0,002 \leq  pd - 0,6745\sigma  \leq 0,04$

Nous apercevons sur la **Figure 10**, que les statistiques de Fisher sont significatives :

- Pour les basses mers :

- ⇒ Entre la ligne 1 et la ligne 142 pour  $S = 28$  et ;
- ⇒ Entre la ligne 1 et la ligne 161 pour  $S = 29$ .

- Pour les pleines mers :

- ⇒ Entre la ligne 1 et la ligne 140 pour  $S = 28$  et ;
- ⇒ Entre la ligne 1 et la ligne 156 pour  $S = 29$ .

Ce qui nous recommande toujours de choisir  $S$  égale à 29.

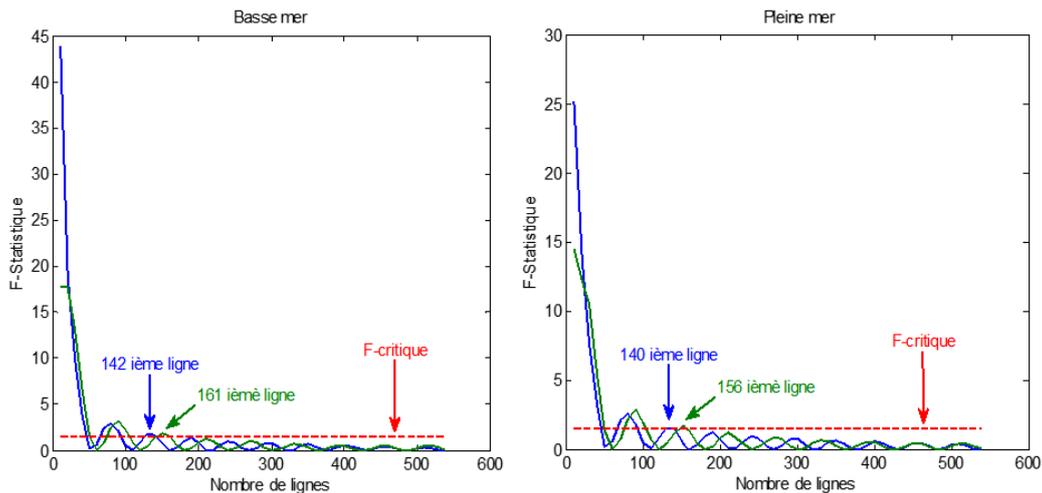


Figure 10 : Test de Stabilité

En répétant la même opération à partir de la ligne 162 pour les basses mers et de la ligne 156 pour les pleines mers et ainsi de suite, nous arrivons à établir le **Tableau 3** qui montre l'existence d'une stabilité locale au niveau de nos chroniques.

Tableau 3 : Analyse de la stabilité pour  $S = 29$

Basses mers			Pleines mers		
Nombre de Lignes	F-Statistique	F-Critique	F-Statistique	F-Critique	Nombre de Lignes
Lignes 1 à 161, 4669 données	1.503	1.478	1,490	1,478	Lignes 1 à 156, 4524 données
Lignes 162 à 323, 6351 données	1.488	1.478	1,533	1,478	Lignes 157 à 314, 4582 données
Lignes 381 à 539, 4611 données	1.527	1.478	1,534	1,478	Lignes 315 à 468, 4466 données
			1,031	1,482	Lignes 469 à 539, 2059 données

Finalement, nous constatons que le test sur les résidus et le test de Fisher arrivent à la même conclusion : il faut prendre la saisonnalité  $S = 29$  Lags pour effectuer la prévision de nos 03 modèles sur les basses mers et de nos 04 modèles sur les pleines mers.

### 3-8. Analyse de la tendance

La mise en œuvre du test de corrélation des rangs de Kendall a permis de dresser le **Tableau 4**:

**Tableau 4 : Test de corrélation de rang de Kendall**

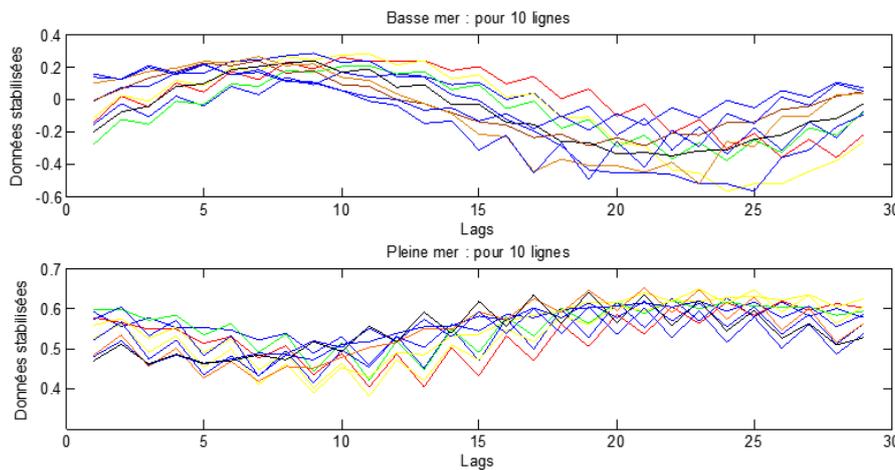
	$\tau_b$	$E(\tau_b)$	$Var(\tau_b)$	$Z_1$	$U_{0,1,0,05}$
Pleine mer	-9,98 %	0	$2,8440.10^{-05}$	-18,7082	1,96
Basse mer	11,80 %	0	$2,8440.10^{-05}$	22,1207	1,96

Les pleines mers ont donc une faible tendance décroissante en fonction du temps de prélèvement et les basses mers faible tendance croissante en fonction du temps de prélèvement.

### 3-9. Schéma de prévision

➤ *Méthode du profil*

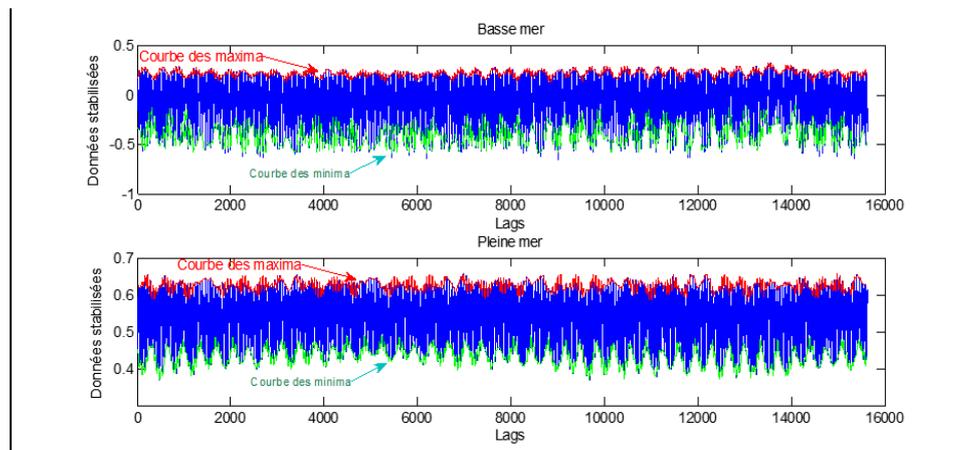
D'après la **Figure 11**, cette méthode ne nous permet pas de trancher sur le choix du schéma puisque plusieurs courbes s'entrecroisent et d'autres sont plus ou moins confondues.



**Figure 11 : Méthode du profil**

➤ *Méthode de la bande*

Nous pouvons observer sur la **Figure 12** que la courbe des minima et celle des maxima ne sont pas parallèles pour les deux chroniques. Cela nous conduit à opter pour un schéma multiplicatif.



**Figure 12 : Méthode de la bande**

➤ *Méthode analytique*

Les Résultats du **Tableau 5** nous recommandent aussi d'appliquer le schéma multiplicatif puisque -28.11 (t-statistique basse mer) et 2,30 (t-statistique pleine mer) sont supérieurs à 1,96 en valeur absolue.

**Tableau 5 : Régression entre la moyenne et l'écart type**

	Basse mer			Pleine mer		
	Coefficients	t -Statistique	Valeur critique	Coefficients	t -Statistique	Valeur critique
g	0,15126582	72,335243	1,96	0,00400041	0,17726521	1,96
$\mu$	-0,84464676	-28,1101052	1,96	0,09609609	2,307198454	1,96

Finalement, sur les trois méthodes testées, deux nous suggèrent de choisir le schéma multiplicatif que nous avons adopté.

### 3-10. Méthode Census X11

La transformation logarithmique a permis :

- Pour les basses mers de réduire l'écart type de 0,4125 à 0,2160 et ;
- Pour les pleines mers de réduire l'écart type de 0,4494 à 0,0566.

Nous avons appliqué la méthode Census X11 aux matrices  $A = 539$  lignes \*29 colonnes et nous avons trouvé la tendance en appliquant la moyenne mobile de Henderson à 13 termes puisque le rapport  $\frac{I}{C}$  reste toujours supérieur à 1 dans les deux cas (1,48 pour les basses mers et 2,24 à pour les pleines mers).

### 3-11. Méthodologie de Box et Jenkins

➤ *Identification des ordres d, q et p + 1 du modèle ARIMA(p + 1, d, q)*

Nous avons choisi un nombre de retard  $p_{max} = 14$ . En appliquant au vecteur  $SCVS_{t,2}$ , la stratégie de Dickey Fuller Augmenté, nous avons trouvé d'après les résultats du **Tableau 6** :

- Pour la basse mer « d » égal à 2 et le modèle ne possède ni tendance, ni constante ;
- Pour la pleine mer « d » égal à 2 pour les deux premiers modèles et « d » égal à 1 pour les deux derniers modèles. Les modèles sont toujours dénuer de tendance et de constante.

**Tableau 6 : Test DFA pour un nombre de retard maximal  $p_{max} = 14$  [32]**

	Basse mer, nombre de retard p = 14									Pleine mer, nombre de retard p = 14									Valeur critique	
	Statistique t pour d = 0			Statistique t pour d = 1			Statistique t pour d = 2			Statistique t pour d = 0			Statistique t pour d = 1			Statistique t pour d = 2			$\phi$	b et a
	$\phi$	b	a	$\phi$	b	a	$\phi$	b	a	$\phi$	b	a	$\phi$	b	a	$\phi$	b	a		
1	238	-2,10	-6,13	13,4	0,06	-0,03	-28,2	0,093	-0,06	133,5	3,36	16,54	1,1	-0,02	-0,006	-24,6	-0,04	0,06	-3,41	1,96
2	266	3,70	-7,23	13,4	0,04	-0,05	-33,0	0,014	-0,02	164,7	-2,74	16,99	4,8	-0,10	0,106	-23,9	0,05	-0,05	-3,41	1,96
3	195	-0,47	-0,40	4,2	-0,04	0,04	-29,8	-0,027	0,03	106,7	-0,47	15,88	-3,6	-0,02	0,022	-24,9	0,07	-0,07	-3,41	1,96
4										64,7	2,56	8,96	-3,8	0,07	-0,069	-17,8	-0,02	0,02	-3,41	1,96

1, 2, 3, 4 représentent le numéro des modèles dans chaque chronique.

**3-12. Modélisation**

Pour  $p_{max}$  égal 14, les résidus des modèles de Dickey Fuller Augmenté sont encore corrélés malgré la confirmation de la stationnarité par le test sauf pour les modèles 3 et 4 pleines mers. Pour blanchir les résidus nous avons dû augmenter  $p_{max}$  de trois unités pour les basses mers et de deux unités pour les modèles 1 et 2 sur les pleines mers avant de procéder à la comparaison du Critère de Schwarz du modèle de Dickey Fuller Augmenté en diminuant progressivement  $p_{max}$  jusqu'à zéro. Ce qui nous a permis de dresser le **Tableau 7** et de déduire l'ordre optimal du modèle de Dickey Fuller Augmenté que nous avons coloré en bleu.

**Tableau 7 : Comparaison des Critères de Schwarz**

Nombre de retard p	Pleine mers				Basse mer		
	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 1	Modèle 2	Modèle 3
17					-16 607,75	-22 256,25	-15 141,27
$p_{optimal} = 16$	-20 432,19	-21 711,17			-16 610,95	-22 258,79	-15 142,25
$p_{optimal} = 15$	-20 435,43	-21 714,38			-16 604,43	-22 249,28	-15 136,79
14	-20 423,08	-21 678,83	-19 440,52	-8 753,77	-16 600,87	-22 240,28	-15 126,71
13	-20 420,65	-21 673,17	-19 443,90	-8 756,88	-16 589,40	-22 225,75	-15 114,69
12	-20 405,63	-21 670,71	-19 441,45	-8 758,22	-16 549,82	-22 169,71	-15 066,35
$p_{optimal} = 11$	-20 404,36	-21 658,82	-19 444,97	-8 761,37	-16 513,18	-22 112,30	-14 996,08
10	-20 388,57	-21 642,55	-19 435,46	-8 755,99	-16 464,27	-22 058,05	-14 956,59
9	-20 388,62	-21 640,43	-19 439,11	-8 759,27	-16 428,94	-21 990,95	-14 927,39
8	-20 362,13	-21 621,24	-19 433,37	-8 760,45	-16 390,92	-21 954,77	-14 925,19
7	-20 363,75	-21 621,63	-19 436,62	-8 763,09	-16 378,58	-21 933,61	-14 912,30
6	-20 320,15	-21 590,05	-19 437,71	-8 762,29	-16 374,73	-21 933,94	-14 908,39
$p_{optimal} = 5$	-20 282,91	-21 585,44	-19 441,36	-8 765,59	-16 375,89	-21 927,64	-14 862,33
4	-20 244,37	-21 500,67	-19 422,66	-8 734,53	-16 341,63	-21 882,97	-14 827,82
3	-20 244,84	-21 485,66	-19 406,68	-8 729,20	-16 116,79	-21 569,25	-14 797,98
2	-20 229,12	-21 484,83	-19 245,32	-8 721,02	-16 092,88	-21 548,27	-14 793,98
1	-20 219,49	-21 331,22	-19 132,94	-8 679,18	-15 481,52	-20 800,01	-14 574,88
0	-19 962,01	-21 019,41	-18 380,12	-8 443,81	-15 436,34	-20 732,82	-14 532,27

Nous avons mis en œuvre le test de Ljung Box et la méthode contrôle ex-post de l'absence d'autocorrélation des innovations » [32] pour obtenir les valeurs de  $q_{optimal}$  que nous avons consignés dans le **Tableau 8 et 9**.

**Tableau 8 : Comparaison des critères d'Akaike et du test de Ljung-box pour les pleines mers**

Ordre q	Pleine mer											
	Modèle 1			Modèle 2			Modèle 3			Modèle 4		
	Pr	k	AIC	Pr	k	AIC	Pr	k	AIC	Pr	k	AIC
12							0,29	3	-19493,10			
11				0,09	5	-21798,22	0,08	15	-19448,15			
10	0,05	4	-20596,66	0,09	15	-21727,05	0,11	14	-19449,93	0,07	4	-8771
9	0,07	11	-20532,64	0,10	15	-21725,79	0,58	13	-19447,09	0,06	12	-8759
8	0,07	11	-20534,64	0,11	15	-21727,67	0,58	13	-19448,95	0,05	12	-8760
7	0,43	9	-20530,24	0,07	15	-21727,88	0,40	13	-19449,70	0,37	11	-8760
6	0,43	9	-20532,11	0,06	15	-21729,54	0,40	13	-19451,65	0,3	11	-8761
5	0,21	9	-20532,16	0,20	14	-21730,42	0,36	13	-19453,30	0,27	11	-8763
4	0,21	9	-20534,03	0,20	14	-21732,19	0,36	13	-19455,30	0,26	11	-8765
3	0,18	9	-20535,54	0,19	14	-21733,52	0,36	13	-19457,16	0,24	11	-8766
2	0,18	9	-20537,40	0,20	14	-21735,41	0,36	13	-19459,16	0,25	11	-8768
1	0,17	9	-20539,20	0,20	14	-21737,29	0,36	13	-19461,12	0,25	11	-8770
0	0,18	9	-20541,20	0,21	14	-21739,29	0,36	13	-19463,12	0,24	11	-8772

**Tableau 9 : Comparaison des critères d'Akaike et du test de Ljung-box pour les basses mers**

Ordre q	Basse mer								
	Modèle 1			Modèle 2			Modèle 3		
	Pr	k	AIC	Pr	k	AIC	Pr	k	AIC
13	0,12	6	-16483,09	0,06	6	-22361,13			
12	0,07	17	-16407,79	0,46	16	-22266,61			
11	0,07	17	-16409,71	0,64	16	-22267,35			
10	0,07	17	-16411,71	0,67	16	-22269,35	0,11	3	-15458,93
9	0,96	16	-16413,56	0,68	16	-22271,35	0,12	16	-15405,67
8	0,96	16	-16415,56	0,69	16	-22273,26	0,15	16	-15406,51
7	0,96	16	-16417,56	0,69	16	-22275,08	0,14	16	-15407,35
6	0,96	16	-16419,48	0,69	16	-22276,90	0,14	16	-15408,15
5	0,96	16	-16421,41	0,69	16	-22278,54	0,15	16	-15408,95
4	0,96	16	-16423,33	0,69	16	-22280,17	0,15	16	-15409,92
3	0,96	16	-16425,10	0,67	16	-22281,90	0,16	16	-15411,30
2	0,05	17	-16427,10	0,67	16	-22283,81	0,17	16	-15413,13
1	0,05	17	-16429,02	0,66	16	-22285,63	0,19	16	-15415,01
0	0,05	17	-16431,02	0,67	16	-22287,63	0,19	16	-15417,01

Pr représente la p-value à l'ordre k au test de Ljung-box, AIC le Critère d'Information d'Akaike et les parties colorées en marrons et en bleue définissent respectivement  $q_{\max}$  et  $q_{\text{optimal}}$  pour chaque modèle. Les paramètres  $d_{\text{adéquat}}$ ,  $p_{\text{optimal}}$  et  $q_{\text{optimal}}$  que nous avons obtenus à partir des **Tableaux 6 à 9** nous permettent de définir maintenant les modèles Arima ( $p_{\text{optimal}} + 1, d_{\text{adéquat}}, q_{\text{optimal}}$ ) pour nos deux chroniques.

➤ *En particulier pour la pleine mer :*

*Le modèle 1* est un ARIMA (16,2,0) sans constante ni tendance :

$$\left\{ \begin{array}{l} \text{AR : } \{-1.32244 \ -0.460963 \ -0.22449 \ -0.266043 \ -0.339991 \ -0.520838 \ -0.496018 \ -0.37686 \ - \\ 0.33032 \ -0.311032 \ -0.251778 \ -0.237872 \ -0.219805 \ -0.212848 \ -0.185309 \ -0.0985826\} \text{ at Lags} \\ [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13 \ 14 \ 15 \ 16] \\ \text{Variance : } 2,83263 \cdot 10^{-05}, \text{ la p-value à l'ordre } k = 9 \text{ au test de Ljung box est égale à } 0,1876 \end{array} \right.$$

*Le modèle 2* est un ARIMA (16,2,0) sans constante :

$$\left\{ \begin{array}{l} \text{AR : } \{-1.30148 \ -0.207144 \ 0.071424 \ -0.307404 \ -0.469021 \ -0.439674 \ -0.36106 \ -0.325447 \ - \\ 0.283225 \ -0.275951 \ -0.26728 \ -0.253469 \ -0.157726 \ -0.169205 \ -0.249312 \ -0.140137\} \text{ at Lags} \\ [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13 \ 14 \ 15 \ 16] \\ \text{Variance : } 1,77285 \cdot 10^{-05}, \text{ la p-value à l'ordre } k = 14 \text{ au test de Ljung box est égale à } 0.2068 \end{array} \right.$$

*Le modèle 3* est un ARIMA (12,1,0) sans constant :

$$\left\{ \begin{array}{l} \text{AR : } \{-0.445972 \ 0.801033 \ 0.199155 \ -0.248153 \ -0.145722 \ -0.102132 \ -0.0501472 \ -0.0709159\} \\ \text{at Lags } [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 9 \ 12] \\ \text{Variance : } 4,33812 \cdot 10^{-05}, \text{ la p-value à l'ordre } k = 13 \text{ au test de Ljung box est égale à } 0.0867 \end{array} \right.$$

*Le modèle 4* est un ARIMA (6,1,0) sans constante

$$\left\{ \begin{array}{l} \text{AR : } \{-0.509639 \ 0.610855 \ 0.280004 \ -0.191357 \ -0.210269\} \text{ at Lags } [1 \ 2 \ 3 \ 5 \ 6] \\ \text{Variance : } 4,99936 \cdot 10^{-05}, \text{ la p-value à l'ordre } k = 11 \text{ au test de Ljung box est égale à } 0.1593 \\ \text{Résidus : normalité acceptée car } |S| = 0,2959 \leq 1,96 \text{ et ad} - 0,7979\sigma = \\ 0,0048; \text{pd} - 0,6745\sigma = -0,0014 \end{array} \right.$$

➤ **Pour la basse mer,**

Le modèle 1 est un ARIMA (17,2,0) sans constante

$$\left. \begin{aligned} \text{AR : } & \{-1.14728 \ 0.235627 \ 0.391886 \ -0.46953 \ -0.601879 \ -0.336974 \ -0.216538 \ -0.224985 \ - \\ & 0.268324 \ -0.28613 \ -0.280508 \ -0.294943 \ -0.281346 \ -0.197207 \ -0.142584 \ -0.146011 \ - \\ & 0.0711499\} \text{ at Lags [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17]} \\ \text{Variance : } & 2.60092 \cdot 10^{-05}, \text{ la p-value à l'ordre } k = 16 \text{ au test de Ljung box est égale à } 0,9520 \end{aligned} \right\}$$

Le modèle 2 est un ARIMA (17,2,0) sans constante

$$\left. \begin{aligned} \text{AR : } & \{-1.15684 \ 0.195224 \ 0.371394 \ -0.448749 \ -0.60393 \ -0.371812 \ -0.238813 \ -0.216011 \ - \\ & 0.263704 \ -0.296908 \ -0.291201 \ -0.297713 \ -0.293599 \ -0.196896 \ -0.145571 \ -0.148846 \ - \\ & 0.0697773\} \text{ at Lags [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17]} \\ \text{Variance : } & 0.000305774, \text{ la p-value à l'ordre } k = 16 \text{ au test de Ljung box est égale à } 0,6461 \end{aligned} \right\}$$

Le modèle 3 est un ARIMA (17,2,0) sans constante

$$\left. \begin{aligned} \text{AR : } & \{-1.15933 \ -0.102458 \ 0.0580946 \ -0.30043 \ -0.410597 \ -0.473176 \ -0.424736 \ -0.336947 \ - \\ & 0.278822 \ -0.261219 \ -0.321135 \ -0.383823 \ -0.381879 \ -0.262808 \ -0.191327 \ -0.162117 \ - \\ & 0.0678325\} \text{ at Lags [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17]} \\ \text{Variance : } & 0,000490815, \text{ la p-value à l'ordre } k = 16 \text{ au test de Ljung box est égale à } 0,2154 \end{aligned} \right\}$$

➤ **Analyse des résidus standardisés**

Le test ARCH indique la présence d'hétéroscédasticité au niveau des résidus car après exécution de la fonction « archtest » sous Matlab nous avons obtenu la valeur 1 pour tous les modèles. Nous avons donc modélisé la variance en utilisant soit la fonction garch(0,1) (respectivement (egarch(0,1))), soit la fonction "garchset" sous Matlab. Les résultats du **Tableau 10** montrent que tous les résidus standardisés  $\eta_t$  restent non corrélés à l'ordre 1 ou plus et convergent vers la distribution normale centrée réduite. En particulier, pour le modèle 3 des données sur les pleines mers,  $\eta_t$  reste non corrélé jusqu'à l'ordre 9.

**Tableau 10 : Paramètres des résidus standardisés**

	N°	ARCH (1) ou EGARCH (0,1)	p-value	Test sur l'Intervalle de confiance	
Modèles Pleine mer	1	$K = 1,6869 \cdot 10^{-05}, \alpha_1 = 0,4289$	$p = 0,4256$ et $k = 1$	$0 \in [-0,0185, 0,0398]$ Moyenne : 0,0106	$1 \in [0,9796, 1,0208]$ Variance : 1,0004
	2	$K = 1,1685 \cdot 10^{-05}, \alpha_1 = 0,3345$	$p = 0,2103$ et $k = 1$	$0 \in [-0,0161, 0,0418]$ Moyenne : 0,0128	$1 \in [0,9595, 1,0415]$ Variance : 1,0005
	3	$K = -10,0638 \ \alpha_1 = 0,3478,$ Leverage = 0,2586	$p = 0,0750$ et $k = 9$	$0 \in [-0,0207, 0,0378]$ Moyenne : 0,0086	$1 \in [0,9526, 1,0351]$ Variance : 0,9938
	4	$K = 3,4297 \cdot 10^{-05}, \alpha_1 = 0,3603$	$p = 0,2059$ et $k = 1$	$0 \in [-0,0626, 0,0238]$ Moyenne : -0,0194	$1 \in [0,9391, 1,0613]$ Variance : 1,0002
Modèles Basse mer	1	$K = 9,4464 \cdot 10^{-05}, \alpha_1 = 0,6566$	$p = 0,9948$ et $k = 1$	$0 \in [-0,0094, 0,0480]$ Moyenne : 0,0193	$1 \in [0,9593, 1,0404]$ Variance : 0,9999
	2	$K = -8,3557 \ \alpha_1 = 0,5480,$ Leverage = 0,1134	$p = 0,5970$ et $k = 1$	$0 \in [-0,0385, 0,0106]$ Moyenne : -0,0140	$1 \in [0,9613, 1,0306]$ Variance : 0,9960
	3	$K = 6,8272 \cdot 10^{-04} \ \alpha_1 = 0,7038$	$p = 0,2518$ et $k = 1$	$0 \in [-0,0322, 0,0256]$ Moyenne : -0,0033	$1 \in [0,9594, 1,0411]$ Variance : 1,0002

➤ *Vérification de la qualité de la prédiction et de la prévision*

Nous avons comparé les erreurs issues de nos modèles avec les résidus provenant du logiciel de prédiction du Service Hydrographique Océanographique de la Marine française qui utilise un modèle harmonique [1]. Pour la comparaison, nous avons pris les coefficients saisonniers de l'année 2011 pour effectuer la prévision de 251 points sur les pleines mers et de 268 points sur les basses mers de l'année 2012 ainsi que la prédiction des 58 avant derniers points de chaque chronique. Les calculs de l'indicateur de THEIL et de l'EMAP aboutissaient aux résultats suivants :

**Tableau 11 : Comparaison des erreurs et de l'Indicateur de THEIL**

Indicateurs	Pleine mer		Basse mer		Prévision avec terme moyenne mobile
	Prédiction	Prévision	Prédiction	Prévision	
THEIL	0,651	2,28	0,22	0,98	1,2
EMAP (modèle)	0,87 %	2,76	2,5 %	16,28 %	18,50 %
EMAP (méthode harmonique)	1,61 %	1,31	17,53 %	18,68 %	18,98 %

En se référant aux indicateurs de THEIL, nous pouvons affirmer que nos modèles offrent de meilleure précision par rapport à la méthode harmonique sauf pour la prévision des pleines mers. Cependant, le modèle reste excellent car l'EMAP est inférieure à 5 %. Nous pouvons également dire que tous nos modèles sont excellents puisque leurs EMAP sont inférieurs à 5 % sauf pour la prévision des basses mers qui entre dans le classement des modèles à moyenne précision. Malgré cela, ce modèle reste meilleur par rapport à la méthode harmonique car son EMAP est inférieur à celle de la méthode harmonique. Pour confirmer si nos modèles s'ajustent bien aux niveaux des pleines mers et basses mers prélevés à Hell-ville, nous avons passé à la simulation qui montre sur la **Figure 13** qu'il est préférable d'ajouter un terme moyenne mobile au lieu d'utiliser un modèle purement autorégressif pour prévoir les basses mers. Nous allons donc remplacer *le modèle 3* pour la prévision des basses mers précédent par un ARIMA (17,2,8) sans constante défini par :

AR :  $\{-1.20115 -0.155344 0.33759 0.398387 -0.549408 -0.786393 -0.135724 0.384957 -0.273516 -0.253587 -0.27176 -0.143191 0.0691767 0.0410379\}$  at Lags [1 2 3 4 6 7 8 9 11 12 13 14 16 17]  
 MA :  $\{-0.0712812 -0.33589 -0.446318 0.330744 -0.457627\}$  at Lags [2 3 4 7 8]  
 Variance : 0,000454712, la p-value à l'ordre k = 19 au test de Ljung box est égale à 0,5288  
 $K = 1,9294 \cdot 10^{-04}$   $\alpha_1 = 0,6798$  ; p = 0,0891 et k=1 ; Moyenne : - 0,0109 ; Variance : 1,0001  
 Résidus : normalité grossière acceptée car ad - 0,7979 $\sigma = -0,0065$ ; pd - 0,6745 $\sigma = -0,0039$

Nos modèles et la méthode harmonique ont détecté également un changement structurel anormal (différence d'environ 1 mètre) sur les prélèvements à partir du mois de Juillet 2012. Notre crainte d'utiliser les données à partir de cette date est donc bien fondée.

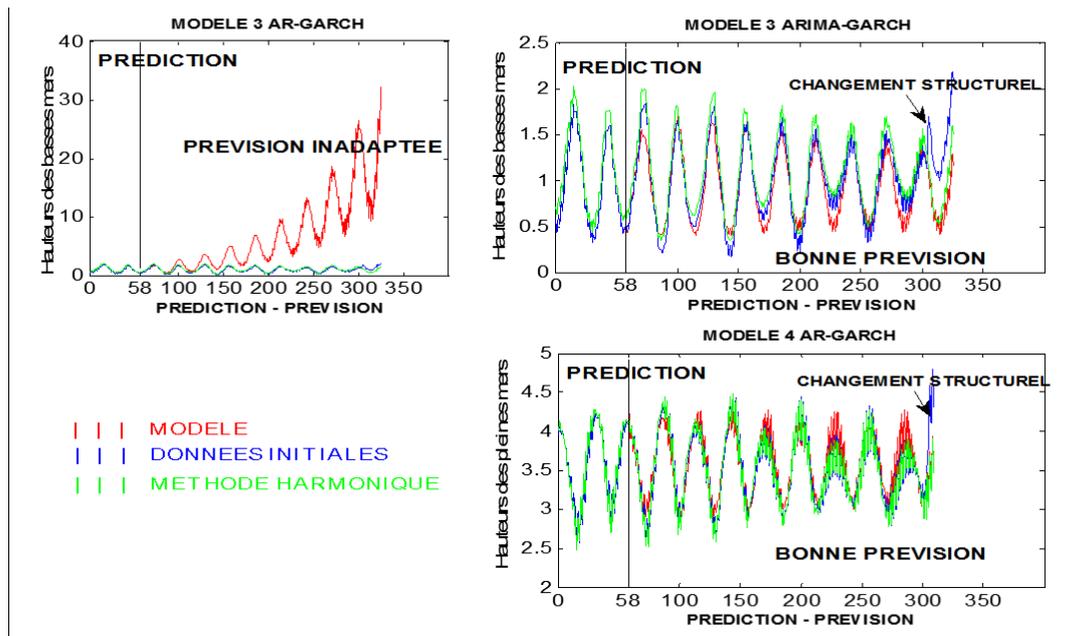


Figure 13 : Préviation et prédiction

#### 4. Conclusion

Nous avons construit 7 modèles stationnaires par différenciation (modèles DS) dont 4 pour les pleines mers et 3 pour les basses mers. Malgré la stabilité locale, la variation saisonnière n'est pas rigide mais souple. Nous avons donc utilisé les coefficients saisonniers de l'année N pour effectuer la prévision de l'année N + 1 qui semblent fournir de bons résultats. Les modèles que nous avons obtenus au départ sont purement autorégressifs, mais après simulation, nous avons constaté qu'il faudrait inclure des termes moyennes mobiles pour obtenir une bonne prévision des basses mers et que le centre de Hell-ville devrait revoir l'anormalité de leurs données à partir de Juillet 2012. En particulier, nous avons trouvé :

- Pour le modèle 1 basse mer un Arima (17,2,0) correspondant aux données 1 à 4669 ;
- Pour le modèle 2 basse mer un Arima (17,2,0) correspondant aux données 4670 à 11020 ;
- Pour le modèle 3 basse mer un Arima (17,2,8) correspondant aux données 11021 à 15631 ;
- Pour le modèle 1 pleine mer un Arima (16,2,0) correspondant aux données 1 à 4524 ;
- Pour le modèle 2 pleine mer un Arima (16,2,0) correspondant aux données 4525 à 9106 ;
- Pour le modèle 3 pleine mer un Arima (12,1,0) correspondant aux données 9107 à 13572 ;
- Pour le modèle 4 pleine mer un Arima (6,1,0) correspondant aux données 13573 à 15631.

Nous avons montré que ces modèles possèdent des résidus normalisés centrés et réduits non corrélés à l'ordre 1 ou plus et que leurs variances résiduelles peuvent être modélisées par ARCH (1) ou EGARCH (0,1). Le test de Fisher nous informe sur la présence d'un cycle à l'intérieur duquel les saisonnalités restent stables. Pour la pleine mer, ce cycle varie entre 4466 (environ 6 ans et 4 mois) et 4582 Lags (environ 6 ans et demi) et pour la basse mer il est moins précis et varie entre 4669 (environ 6 ans et 7 mois) et 6351 Lags (9 ans). En réalité, nous avons 5 chance sur 6 que le cycle des pleines et basses se situe entre 6,25 et 6,75 ans. Nous avons aussi remarqué que la tendance est non linéaire et que nos méthodes aboutissaient aux mêmes résultats que les saisonnalités de 14,765 jours, de 1,034 jours et de 0,517 jours obtenues par la théorie statique de Newton. Nous avons également constaté que la meilleure approche pour choisir le bon modèle

consiste à minimiser le Critère d'Information d'Akaike en variant le  $q_{\max}$  dans le modèle ARIMA ( $p_{\text{optimal}} + 1, d_{\text{adéqua}}, q_{\max}$ ) et non le  $q_{\text{optimal}}$  dans le modèle ARIMA( $p_{\text{optimal}} + 1, d_{\text{adéqua}}, q_{\text{optimal}}$ ) obtenu à partir de la stratégie de Dickey Fuller Augmenté. Finalement, nous avons trouvé des modèles AR-GARCH et des modèles ARIMA-GARCH pour prévoir les pleines mers et les basses mers à Hell-ville et qui nous semblent satisfaisants. Il nous paraît donc inutile de procéder à une modélisation FARIMA.

## Références

- [1] - Annuaire des marées, Ports d'outre-mer Tome 2, Service Hydrographique et Océanographique de la Marine, (2015) 127 - 150
- [2] - Simon BERNARD, La marée : la marée océanique cotière, référence 942-MOG, Collection « Synthèse », Institut océanographique, Fondation Albert I, prince de Monaco, (2007) 69 - 236
- [3] - Maher ABDALLAH, Thèse de doctorat ; sur les méthodes de discrétisation numérique de problèmes hyperboliques non linéaires appliquées aux équations de Barré de Saint Venant pour la modélisation de l'hydraulique en réseau d'assainissement, Ecole Doctorale sciences pour l'ingénieur, Université Louis Pasteur, Strasbourg I, soutenue le 03 Mai 2005
- [4] - Dominique LADIRAY, Benoît QUENNEVILLE, Désaisonnaliser avec la méthode X-11, (2000)
- [5] - L. de MONTERA, C. MALLET et L. BARTHES, Prédiction multi-step de la volatilité : le modèle ARIMA-GARCH appliqué aux séries temporelles d'affaiblissement par la pluie sur les liaisons Terre-Satellite, (2008)
- [6] - MISULTANE Hélène, Econométrie des séries temporelles, (2002) Consulté le 31 mars 2017 à l'adresse <http://halshs.archives-ouvertes.fr/cel-01261174>
- [7] - Jairo CULGIARI, Séries chronologiques saisonnières et prévision, L1 IDEA-S 2014-2015, Université Lumière Lyon 2, version compilée le 05 février 2015
- [8] - Henri IMMEDIATO, Cours de statistique - MASS UE11- Chapitre 2- NTE Lyon 1, Consulté le 16 Novembre 2016 à l'adresse, <http://nte-serveur.univ-lyon1.fr/nte/immediato/math2002/mass11/cours/chapitr2b.htm>
- [9] - [https://en.wikipedia.org/wiki/Hurst\\_exponent](https://en.wikipedia.org/wiki/Hurst_exponent)
- [10] - Florence NICOLAU, Etudes statistiques 2 : Traitement des valeurs manquantes et des valeurs aberrantes, (2005 - 2006)
- [11] - Olivier BESSON ; Analyse spectrale paramétrique, 3ème année ENSICA
- [12] - Matlab 2013a, Matlab fonction
- [13] - P. GRANJON, J. L LACOUME, Introduction à l'analyse spectrale
- [14] - W. H. PRESS, S. A. TENKOLSKY, W. T. VETTERLING, B. P. Flannery, Numerical Recipes in C. s.l., 2 (1992)
- [15] - Vincent GOUDARD, Percussion aérienne : amélioration de la détection des coups dans le jeu d'un percussionniste. Université de la Méditerranée, Aix-Marseille II, (2004)
- [16] - <http://www.labri.fr/perso/mraspaul/publications/dafx03.pdf>
- [17] - [http://www.dcsc.tudelft.nl/Research/Publication Files/publication-5756.pdf](http://www.dcsc.tudelft.nl/Research/Publication%20Files/publication-5756.pdf)
- [18] - N. KAYSER, J. ROBST, D. C. YVES TARD, *Propagation des ondes de sécheresse et d'humidité à travers le monde*. C.R.Acad.Sci.Paris. Série II Hydrologie/hydrology, (1990) 757 - 763
- [19] - J. P. BURG, Maximum entropy spectral analysis. Oklahoma City : in Proc. of the 37th meeting of the society of exploration geophysicists, October (1967)
- [20] - <http://www.enee.umd.edu/~afshin/adsp3/report.pdf>

- [21] - Grais BERNARD, Inspecteur général de l'INSEE, Professeur à l'INTEC, Expert auprès de l'Office statistique des Communautés européennes, *Techniques statistiques 2 ; Méthodes statistiques, troisième édition, nouvelle présentation, DUNOD, (1992)*
- [22] - Fanny MEYER, Morgane CADRAN, Margaux GAILLARD, *Analyse de la variance, M2 statistiques et économétrie*
- [23] - Ricco RAKOTOMALALA, *Comparaison de populations, Tests paramétriques, Version 1,2, Université Lumière Lyon 2, (2013)*
- [24] - Chapitre 4, *Estimation, Tests de validation et prévision des processus ARMA*
- [25] - A. VESSEREAU, Inspecteur Général des Manufactures de l'Etat, *Méthodes statistiques en biologie et en agronomie, (1998) 55 - 158*
- [26] - A. JEBRANE, *Test d'égalité des variances, Université de Bourgogne, Master 1 psychologie, 28 septembre 2016*
- [27] - Nicolas THOME ; *Bases du traitement des images : Transformée de Fourier Avancée - Numériastion, 4 octobre 2016*
- [28] - M. G. KENDALL, *Rank Correlation Methods*. Griffin, London, (1975)
- [29] - Agnès LAGNOUX, *Séries Chronologiques, ISMAG, MASTER1- MI00141X, Université de Toulouse Le MIRAIL*
- [30] - Thibault GIBERT, *Etude statistique, Dossier analyse d'une série temporelle, Agriculture - Production laitière bovine en France, Master II Ingénierie de projets, (2010 - 2011)*
- [31] - Michel TENENHAUS, *La méthodologie de Box-Jenkins*
- [32] - Chapitre 2, *Tests de non stationnarité et processus aléatoires non stationnaires*
- [33] - Karine BOUTHEVILLAIN, Alexandre MATHIS, *Prévision : mesures, erreurs et principaux résultants. In : Economie et statistique, N°285 - 286 (1995) Doi:10.3406/estat.1995.5982 consulté le 17 février 2017 à l'adresse, [http://www.persee.fr/web/revues/home/prescript/article/estat\\_0336-1454\\_1995\\_num\\_285\\_1\\_5982](http://www.persee.fr/web/revues/home/prescript/article/estat_0336-1454_1995_num_285_1_5982)*
- [34] - Amar RAMUDHIN, Ingénieur.Ph.D, *Prévision de la demande, 38p, consulté le 17 Février 2017 à l'adresse : <https://samos.univ-paris1.fr/archives/membres/perraudin/ST/resumechap2.pdf>*