

Statistiques et Traitement de Données : analyse harmonique de deux séries de données provenant du marégraphe de Saint John

Josselin Fatah-Roux

Université de Toulon
Master 2 Physique Science de l'Ingénieur
Spécialité Physique Surveillance de l'Environnement
vufic@outlook.com

14 octobre 2017

Résumé

Il est possible de reconstruire le signal de marée en utilisant la décomposition en harmonique. Néanmoins il faut connaître l'amplitude et la phase des composantes de marée les plus importantes. Deux méthodes existent pour cela : les moindres carrés et la transformée de Fourier. On s'appliquera ici à deux séries de données du marégraphe de Saint John.

Mots-clés :

Analyse harmonique ; Moindres carrés ; Transformée de Fourier discrète rapide ; Marée ; Marégraphe de Saint John

1 Introduction

Introduite par Fourier, l'analyse harmonique est par définition une branche des mathématiques qui s'occupe de la représentation des fonctions ou des signaux comme superposition d'ondes de base également appelées harmoniques.

C'est au 18^e siècle que de nombreux mathématiciens et physiciens se posèrent la question de l'expression des solutions des premières équations aux dérivées partielles de l'époque : l'équation des cordes vibrantes et l'équation de la chaleur. Bien que beaucoup eurent l'idée d'une décomposition en ondes de base, ce n'est qu'un peu plus tard au 19^e siècle que l'un d'eux, nommé Fourier, apporta de nombreux éléments pour expliciter cette décomposition et la systématiser : l'analyse harmonique était née.

2 Matériels et Méthodes

2.1 Analyse harmonique

Le système s'écrit comme suit :

$$h(t) = \bar{h} + \sum_{q=1}^M C_q \cos(2\pi f_q t - \Phi_q) + h_r(t)$$

\bar{h} : moyenne

C_q : amplitude

f_q : fréquence

Φ_q : phase

$h_r(t)$: résiduelle

La résolution par la méthode des moindres carrées donne :

$$h(t) = \bar{h} + \sum_{q=1}^M C_q \cos(2\pi f_q t - \Phi_q) + h_r(t)$$

$$\text{Or : } \cos(a + b) = \cos(a) \cos(b) - \sin(a) \sin(b)$$

$$\text{D'où : } h(t) = \bar{h} + \sum_{q=1}^M A_q \cos(2\pi f_q t) + B_q \sin(2\pi f_q t) + h_r(t)$$

$$\text{Avec : } A_q = C_q \cos(\Phi_q) \text{ et } B_q = C_q \sin(\Phi_q)$$

$$\text{Soit : } C_q = \sqrt{A_q^2 + B_q^2} \text{ et } \Phi_q = \arctan\left(\frac{B_q}{A_q}\right)$$

$$\Leftrightarrow h(t) = A_t \theta + h_r(t)$$

$$\text{Avec : } A_t = \left(1 \quad \sum_{q=1}^M \cos(2\pi f_q t) \quad \sum_{q=1}^M \sin(2\pi f_q t) \right) \text{ et } \theta = \begin{pmatrix} \bar{h} \\ A_q \\ B_q \end{pmatrix}$$

$$\text{Or : } \hat{\theta} = \left(\sum_{t=1}^N A_t^T A_t \right)^{-1} \left(\sum_{t=1}^N h(t) A_t^T \right)$$

$$\text{On a : } \sum_{t=1}^N A_t^T A_t = \sum_{t=1}^N \begin{pmatrix} 1 \\ \sum_{q=1}^M \cos(2\pi f_q t) \\ \sum_{q=1}^M \sin(2\pi f_q t) \end{pmatrix} \begin{pmatrix} 1 & \sum_{j=1}^M \cos(2\pi f_j t) & \sum_{j=1}^M \sin(2\pi f_j t) \end{pmatrix}$$

D'où :

$$\sum_{t=1}^N A_t^T A_t = \begin{pmatrix} \sum_{t=1}^N 1 & \sum_{j=1}^M \left(\sum_{t=1}^N \cos(2\pi f_j t) \right) & \sum_{j=1}^M \left(\sum_{t=1}^N \sin(2\pi f_j t) \right) \\ \sum_{j=1}^M \left(\sum_{t=1}^N \cos(2\pi f_q t) \right) & \sum_{q=1}^M \sum_{j=1}^M \left(\sum_{t=1}^N \cos(2\pi f_q t) \cos(2\pi f_j t) \right) & \sum_{q=1}^M \sum_{j=1}^M \left(\sum_{t=1}^N \cos(2\pi f_q t) \sin(2\pi f_j t) \right) \\ \sum_{q=1}^M \left(\sum_{t=1}^N \sin(2\pi f_q t) \right) & \sum_{q=1}^M \sum_{j=1}^M \left(\sum_{t=1}^N \cos(2\pi f_j t) \sin(2\pi f_q t) \right) & \sum_{q=1}^M \sum_{j=1}^M \left(\sum_{t=1}^N \sin(2\pi f_q t) \sin(2\pi f_j t) \right) \end{pmatrix}$$

$$\Leftrightarrow \left(\sum_{t=1}^N A_t^T A_t \right)^{-1} = \begin{pmatrix} N & \sum_{j=1}^M C_j & \sum_{j=1}^M S_j \\ \sum_{q=1}^M C_q & \sum_{q=1}^M \sum_{j=1}^M C_q C_j & \sum_{q=1}^M \sum_{j=1}^M C_q S_j \\ \sum_{q=1}^M S_q & \sum_{q=1}^M \sum_{j=1}^M C_j S_q & \sum_{q=1}^M \sum_{j=1}^M S_q S_j \end{pmatrix}^{-1}$$

$$C_q = \sum_{t=1}^N \cos(2\pi f_q t)$$

$$S_q = \sum_{t=1}^N \sin(2\pi f_q t)$$

$$CC_{qj} = \sum_{t=1}^N \cos(2\pi f_q t) \cos(2\pi f_j t)$$

$$CS_{qj} = \sum_{t=1}^N \cos(2\pi f_q t) \sin(2\pi f_j t)$$

$$SS_{qj} = \sum_{t=1}^N \sin(2\pi f_q t) \sin(2\pi f_j t)$$

$$\text{De plus : } \sum_{t=1}^N h(t) A_t^T = \begin{pmatrix} \sum_{t=1}^N h(t) \\ \sum_{q=1}^M \left(\sum_{t=1}^N h(t) \cos(2\pi f_q t) \right) \\ \sum_{q=1}^M \left(\sum_{t=1}^N h(t) \sin(2\pi f_q t) \right) \end{pmatrix}$$

$$\Leftrightarrow \sum_{t=1}^N h(t) A_t^T = \begin{pmatrix} XC_0 \\ \sum_{q=1}^M XC_q \\ \sum_{q=1}^M XS_q \end{pmatrix}$$

$$XC_0 = \sum_{t=1}^N h(t)$$

$$XC_q = \sum_{t=1}^N h(t) \cos(2\pi f_q t)$$

$$XS_q = \sum_{t=1}^N h(t) \sin(2\pi f_q t)$$

$$\text{Donc : } \hat{\theta} = \begin{pmatrix} \hat{h} \\ \hat{A}_q \\ \hat{B}_q \end{pmatrix} = \begin{pmatrix} N & \sum_{j=1}^M C_j & \sum_{j=1}^M S_j \\ \sum_{q=1}^M C_q & \sum_{q=1}^M \sum_{j=1}^M C_q C_j & \sum_{q=1}^M \sum_{j=1}^M C_q S_j \\ \sum_{q=1}^M S_q & \sum_{q=1}^M \sum_{j=1}^M C_j S_q & \sum_{q=1}^M \sum_{j=1}^M S_q S_j \end{pmatrix}^{-1} \begin{pmatrix} XC_0 \\ \sum_{q=1}^M XC_q \\ \sum_{q=1}^M XS_q \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} N & C_1 & C_2 & \dots & C_M & S_1 & S_2 & \dots & S_M \\ C_1 & CC_{11} & CC_{12} & \dots & CC_{1M} & CS_{11} & CS_{12} & \dots & CS_{1M} \\ C_2 & CC_{21} & CC_{22} & \dots & CC_{2M} & CS_{21} & CS_{22} & \dots & CS_{2M} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ C_M & CC_{M1} & CC_{M2} & \dots & CC_{MM} & CS_{M1} & CS_{M2} & \dots & CS_{MM} \\ S_1 & SS_{11} & SS_{12} & \dots & SS_{1M} & SC_{11} & SC_{12} & \dots & SC_{1M} \\ S_2 & SS_{21} & SS_{22} & \dots & SS_{2M} & SC_{21} & SC_{22} & \dots & SC_{2M} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_M & SS_{M1} & SS_{M2} & \dots & SS_{MM} & SC_{M1} & SC_{M2} & \dots & SC_{MM} \end{pmatrix} \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ \dots \\ A_M \\ B_1 \\ B_2 \\ \dots \\ B_M \end{pmatrix} =$$

$$\begin{pmatrix} XC_0 \\ XC_1 \\ XC_2 \\ \dots \\ XC_M \\ XS_1 \\ XS_2 \\ \dots \\ XS_M \end{pmatrix} \Leftrightarrow Dz = y \text{ soit } z = D^{-1}y$$

Une autre méthode est possible en minimisant la variance résiduelle :

$$e^2 = \sum_{t=1}^N h_r^2(t) = \sum_{t=1}^N [h(t) - (\bar{h} + \sum_{q=1}^M A_q \cos(2\pi f_q t) + B_q \sin(2\pi f_q t))]^2$$

$$\Leftrightarrow e^2 = \sum_{t=1}^N [h(t)^2 - 2h(t)(\bar{h} + \sum_{q=1}^M A_q \cos(2\pi f_q t) + B_q \sin(2\pi f_q t)) + (\bar{h} + \sum_{q=1}^M A_q \cos(2\pi f_q t) + B_q \sin(2\pi f_q t))^2]$$

$$\Leftrightarrow e^2 = \sum_{t=1}^N [h(t)^2 - 2h(t)\bar{h} - 2h(t) \sum_{q=1}^M A_q \cos(2\pi f_q t) + B_q \sin(2\pi f_q t) + \bar{h}^2 + 2\bar{h} \sum_{q=1}^M A_q \cos(2\pi f_q t) + B_q \sin(2\pi f_q t) + (\sum_{q=1}^M A_q \cos(2\pi f_q t) + B_q \sin(2\pi f_q t))^2]$$

$$\Leftrightarrow e^2 = \sum_{t=1}^N [h(t)^2 - 2h(t)\bar{h} - 2h(t) \sum_{q=1}^M A_q \cos(2\pi f_q t) - 2h(t) \sum_{q=1}^M B_q \sin(2\pi f_q t) + \bar{h}^2 + 2\bar{h} \sum_{q=1}^M A_q \cos(2\pi f_q t) + 2\bar{h} \sum_{q=1}^M B_q \sin(2\pi f_q t) + (\sum_{q=1}^M A_q \cos(2\pi f_q t) + B_q \sin(2\pi f_q t))^2]$$

$$\Leftrightarrow e^2 = \sum_{t=1}^N [h(t)^2 - 2h(t)\bar{h} - 2h(t) \sum_{q=1}^M A_q \cos(2\pi f_q t) - 2h(t) \sum_{q=1}^M B_q \sin(2\pi f_q t) + \bar{h}^2 + 2\bar{h} \sum_{q=1}^M A_q \cos(2\pi f_q t) + 2\bar{h} \sum_{q=1}^M B_q \sin(2\pi f_q t) + \sum_{q=1}^M \sum_{j=1}^M A_q^2 \cos(2\pi f_q t) \cos(2\pi f_j t) + \sum_{q=1}^M \sum_{j=1}^M B_q^2 \sin(2\pi f_q t) \sin(2\pi f_j t) + 2 \sum_{q=1}^M \sum_{j=1}^M A_q B_q \sin(2\pi f_q t) \cos(2\pi f_j t)]$$

Pour minimiser : $\frac{\partial e^2}{\partial A_q} = \frac{\partial e^2}{\partial B_q} = 0$

$$\frac{\partial e^2}{\partial A_q} = \sum_{t=1}^N [-h(t) \sum_{q=1}^M \cos(2\pi f_q t) + \bar{h} \sum_{q=1}^M \cos(2\pi f_q t) + \sum_{q=1}^M \sum_{j=1}^M A_q \cos(2\pi f_q t) \cos(2\pi f_j t) + \sum_{q=1}^M \sum_{j=1}^M B_q \sin(2\pi f_q t) \cos(2\pi f_j t)]$$

$$\frac{\partial e^2}{\partial B_q} = \sum_{t=1}^N [-h(t) \sum_{q=1}^M \sin(2\pi f_q t) + \bar{h} \sum_{q=1}^M \sin(2\pi f_q t) + \sum_{q=1}^M \sum_{j=1}^M B_q \sin(2\pi f_q t) \sin(2\pi f_j t) + \sum_{q=1}^M \sum_{j=1}^M A_q \sin(2\pi f_q t) \cos(2\pi f_j t)]$$

En réalisant l'égalité on obtient à nouveau le $Dz = y$ trouvé au-dessus.

2.2 Application sur MATLAB

Pour être sur que la résolution faite précédemment soit bien retranscrit sur MATLAB il est primordial de créer un premier script pour une seule harmonique avec un signal non bruité dont la valeur de ϕ , de \bar{h} et de f sont choisis délibérément. Ainsi lors de son exécution les valeurs attendues doivent être les mêmes que celles données. La même méthode est ensuite appliquée pour plusieurs harmoniques.

ANNEXE 1 et 2

2.3 Marégraphe de Saint John

À notre disposition deux séries de données (g223b0493965 et g223b0493928) reçues du marégraphe Saint John (47.5666°N, 52.7166°W) situé sur la côte Atlantique au Canada, qui a enregistré le niveau de la mer toutes les heures.

En adaptant le script sur plusieurs harmoniques créé précédemment aux données du marégraphe de Saint John, on peut alors faire une décomposition en harmoniques et calculer via la méthode des moindres carrés les amplitudes et les phases de ses composantes. Cependant cette méthode exige de connaître les fréquences. Par conséquent on suppose que les composantes principales de la marée suivent les cycles :

- semi-diurne : 12heures
- semi-diurne lunaire : 12.5 heures
- diurne solaire : 24 heures
- diurne lunaire : 25 heures

Les fréquences à insérer sont alors les suivantes : $1/12$ $1/12.5$ $1/24$ et $1/25$.

À la suite du programme il est intéressant d'ajouter la transformée de Fourier permettant une comparaison des résultats. Il est également utile de calculer la résiduelle pour discerner les composantes non-principales. Celle-ci se calculant comme suit :

$$h_r(t) = h(t) - \bar{h} + \sum_{q=1}^M C_q \cos(2\pi f_q t - \Phi_q)$$

L'ensemble des résultats obtenus doivent converger.

ANNEXE 3 et 4

3 Résultats

3.1 Première série de données

Le premier signal comporte exactement 8634 mesures :

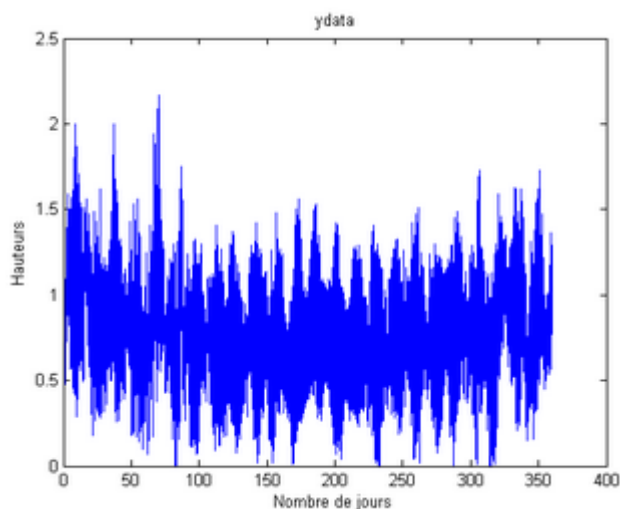


FIGURE 1 – Signal long

Résultats des moindres carrés :

$$\begin{aligned}h &= (6.4259e - 005) \\phase &= (-1.8088 \quad -0.6537 \quad -1.3214 \quad 0.1626) \\Amplitude &= (0.1377 \quad 0.0246 \quad 0.0072 \quad 0.0057)\end{aligned}$$

Résultats de la transformée de Fourier :

- 4 pics correspondant aux fréquences : $1/12$ $1/12.5$ $1/24$ et $1/25$.
- dont les amplitudes valent : 0.08354 0.3368 0.06491 et 0.03503.
- dont les phases valent : no data

Résultats de la résiduelle :

En soustrayant la résiduelle au signal d'origine on observe des paquets d'onde parfaitement lisse avec une amplitude d'environ 0.14.

Si le fichier de données est découpé en plusieurs morceaux alors les résultats sont les mêmes que ceux de la deuxième série de données.

ANNEXE 5

3.2 Deuxième série de données

Le second signal comporte exactement 480 mesures :

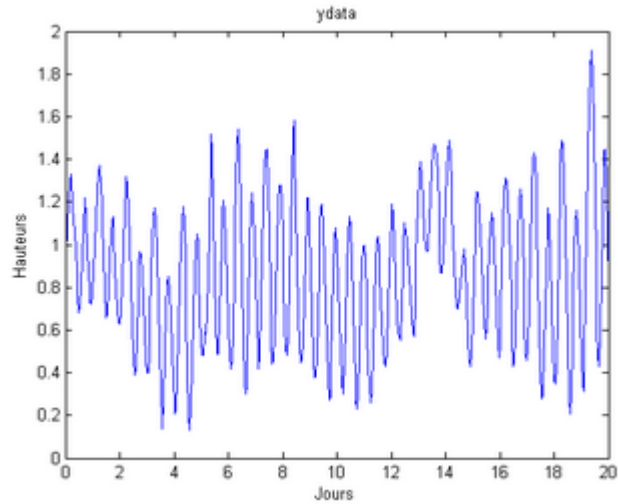


FIGURE 2 – Signal court

Résultats des moindres carrés :

$$\begin{aligned}h &= (-0.0029) \\ \text{phase} &= (-1.6586 \quad 1.1383 \quad 1.7953 \quad 1.8236) \\ \text{Amplitude} &= (0.0955 \quad 0.3614 \quad 0.1131 \quad 0.0488)\end{aligned}$$

Résultats de la transformée de Fourier :

- 2 pics correspondant aux fréquences : $1/12.5$ et $1/24$.
- dont les amplitudes valent : 0.2811 et 0.1094.
- dont les phases valent : no data

Résultats de la résiduelle :

En soustrayant la résiduelle au signal d'origine on observe des paquets d'onde parfaitement lisse avec une amplitude d'environ 0.5.

ANNEXE 6

4 Discussion

4.1 Interprétation des résultats

On voit que pour un signal long la FFT affiche bien les quatre fréquences principales mais les amplitudes calculées via les moindres carrés sont plutôt mauvaises.

Par contre pour un signal court la FFT n'affiche que deux pics sur quatre néanmoins les amplitudes calculées via les moindres carrés sont plutôt bonnes cette fois-ci.

C'est pourquoi il est important de réaliser une série de mesures sur une longue durée (minimum plus de 6 mois) et éventuellement par la suite découpé ce fichier en plusieurs sections.

4.2 Commentaire sur MATLAB

MATLAB ne semble pas réussir à interpréter correctement les données d'un fichier trop long ce qui expliquerait les erreurs sur les amplitudes calculées via les moindres carrés.

4.3 Commentaire sur les moindres carrés et la FFT

Malgré qu'il faille beaucoup de donnée pour réaliser une FFT très précise elle possède un avantage non négligeable par rapport à la méthode des moindres carrés, celui de ne pas exiger la connaissance des fréquences principales.

5 Conclusion

L'analyse harmonique permet, comme on a pu le voir dans ce rapport, une meilleure compréhension des phénomènes de marée. Toutefois son utilisation ne s'arrête pas à l'océanographie et s'applique également dans bien d'autres domaines scientifiques (mécanique quantique, neurosciences, stratigraphie ...). Elle est et reste encore aujourd'hui une méthode incontournable dont le perfectionnement se continue.