



## Etude de cas 1 - Prédiction des marées

Véronique Demianenko

29 Avril 2024

### 1 Introduction

Nous possédons de nombreux fichiers texte contenant des enregistrements de la hauteur de l'eau dans le port de l'île d'Yeu obtenues grâce à un marégraphe. On retrouve notamment des informations concernant la date et l'heure de l'enregistrement (un enregistrement se fait toutes les six minutes), selon les fichiers on a une moyenne de la hauteur ou 10 mesures de la hauteur en cm prises par le marégraphe, la température en degrés Celsius et la pression atmosphérique en hectopascals. Le matériel n'étant pas idéal, les données sont parfois erronées, ce qui fausse les valeurs finales.

Il faudra donc dans le cadre de cette étude de cas tout d'abord nettoyer les données et ensuite créer un prédicteur de la hauteur de l'eau pour une date donnée, ainsi qu'évaluer l'erreur qui existe entre les hauteurs prédites et mesurées. Finalement, il s'agira de créer un outil d'aide à la décision sur Excel, pour prévoir dans le futur la hauteur, dont la formule aura été déterminée précédemment.

## 2 Assemblage et nettoyage des données obtenues par le marégraphe

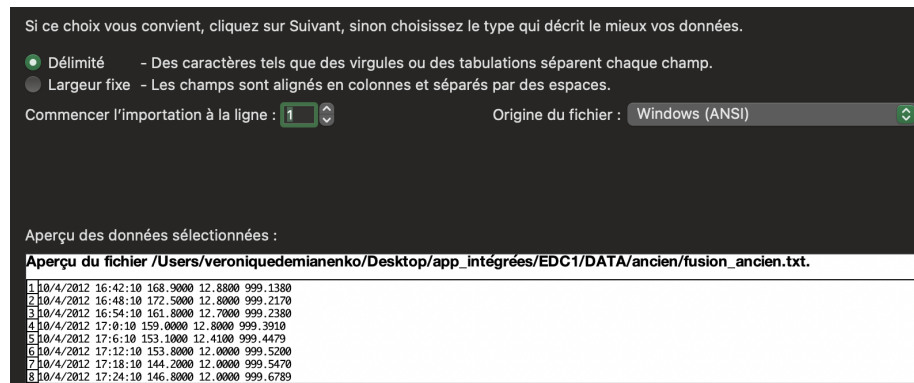
### 2.1 Importation des données

En observant les différentes données fournies, on remarque que deux types de fichiers textes existent : l'un des deux présente la moyenne de la hauteur d'eau à une date et heure donnée, alors que l'autre présente 10 valeurs dont il faudra récupérer la moyenne.

Pour implémenter ces données sur Excel, il a fallu tout d'abord fusionner les fichiers. Pour cela, j'ai rentré la ligne de commande suivante dans mon terminal :

```
cat *.txt > données_anciennes.txt  
cat *.txt > données_nouvelles.txt
```

Ensuite, puisque j'utilise un mac, je n'arrive pas à relier mon Module VBA à un quelconque fichier de mon ordinateur par son chemin. J'ai donc utilisé PowerQuery, sur lequel il est possible de délimiter manuellement et facilement les différentes colonnes :



Ici je peux choisir le type de délimitation de mes données. Vu qu'il y a deux espaces entre la date et l'heure, je choisis l'option "Délimité" et non pas "Largeur fixe".

Cette étape vous permet de choisir les séparateurs contenus dans vos données.

Délimiteurs

☒ Tabulation ☒ Interpréter des séparateurs identiques consécutifs comme uniques

☐ Point-virgule

☐ Virgule

☒ Espace

☐ Autre :

Identificateur de texte :

Aperçu des données sélectionnées :

10/4/2012	16:42:10	168.9000	12.8800	999.1380
10/4/2012	16:48:10	172.5000	12.8000	999.2170
10/4/2012	16:54:10	161.8000	12.7000	999.2380
10/4/2012	17:0:10	159.0000	12.8000	999.3910
10/4/2012	17:6:10	153.1000	12.4100	999.4470
10/4/2012	17:12:10	153.8000	12.0000	999.5200
10/4/2012	17:18:10	144.2000	12.0000	999.5470
10/4/2012	17:24:10	146.8000	12.0000	999.6789

Ici je sélectionne les délimiteurs tabulation et espace, qui me permettent d'avoir la date et l'heure dans deux colonnes séparées.

Cette étape vous permet de sélectionner chaque colonne et de définir le format des données.

Format des données en colonne

☐ Général

☐ Texte

☒ Date :

☐ Colonne non distribuée

Avancé...

Aperçu des données sélectionnées :

JMA	Standard	Standard	Standard	Standard
11/4/2012	9:18:10	594.0000	8.9000	1000.8890
11/4/2012	9:24:10	594.2000	8.9000	1000.9710
11/4/2012	9:30:10	593.4000	8.9000	1001.0550
11/4/2012	9:36:10	595.7000	9.0000	1001.0220
11/4/2012	9:42:10	589.4000	9.1800	1000.9100
11/4/2012	9:48:10	590.9000	9.3000	1000.8859
11/4/2012	9:54:10	599.8000	9.5000	1000.9040
11/4/2012	10:0:10	595.5000	9.5000	1000.8040

Enfin, je peux choisir le type de données de chaque colonne : pour ne pas avoir de problèmes avec la date, je sélectionne le type JMA dans l'onglet "Date" pour la première colonne. Suite à cela, les données sont importées dans la feuille de mon choix sur mon tableur excel.

J'effectue les mêmes opérations pour le deuxième type de fichiers textes.

## 2.2 Nettoyage des données

À présent, en observant les données, on remarque de nombreux défauts qu'il faudra corriger à l'aide d'un Module en VBA :

- certaines hauteurs sont trop petites (inférieures à 34cm),
- d'autres sont au contraire trop grande, on remarque la présence aléatoire de la hauteur 701 qui revient souvent, signe d'une erreur qui provient du marégraphe,
- certaines lignes sont vides,
- certaines dates ont mal été enregistrées (elles commencent par 165...), dans ce cas il faut supprimer la ligne dans son intégralité.

Pour les données dites nouvelles (c'est-à-dire celles où on a 10 valeurs à une

seule date), on calcule la moyenne, puis on supprime les lignes qui n'en ont pas (cela signifie que lors du nettoyage, toutes les hauteurs ont été considérées comme aberrantes).

Voici le code VBA résultant :

```
Dim nv As Worksheet
Dim anc As Worksheet
Dim nblin_nv As Integer
Dim nblin_anc As Integer
Dim somme As Double
Dim cpt As Integer
Dim moy As Double

Set nv = ThisWorkbook.Sheets("DonnéesNv")
Set anc = ThisWorkbook.Sheets("DonnéesAnc")

' Il y a 58491 lignes dans DonnéesBrutesNv

With ThisWorkbook.Worksheets("DonnéesNv")
    For i = 1 To 100
        For j = 3 To 12
            If .Cells(i, j).Value = 701 Or .Cells(i, j).Value < 35 Then
                .Cells(i, j).Value = " "
            End If
        Next j
    Next i

    For i = 1 To 1 Step -1
        If Left(nv.Cells(i, 1).Value, 1) = "#" Then
            .Rows(i).Delete
        End If
    Next i

    For i = 1 To 100
        somme = 0
        cpt = 0
        moy = 0
        For j = 3 To 12
            If IsNumeric(.Cells(i, j).Value) And Not .Cells(i, j).Value = " " Then
                somme = somme + .Cells(i, j).Value
                cpt = cpt + 1
            End If
        Next j

        If Not cpt = 0 Then
            moy = somme / cpt
        End If
    Next i
End With
```

```

        .Cells(i, 15).Value = moy
    End If

Next i

' Il faut supprimer les lignes ou la moyenne n'existe pas :

For i = 3 To 2 Step -1
    If .Cells(i, 15).Value = "" Then
        .Rows(i).Delete
    End If
Next i

End With

With ThisWorkbook.Worksheets("DonnéesAnc")

For i = 25000 To 1 Step -1
    If .Cells(i, 1).Value = "" Then
        .Rows(i).Delete
    End If
Next i

End With

```

Enfin, on fusionne les feuilles "DonnéesAnc" et "DonnéesNv" en ne conservant que la date, l'heure, la hauteur moyenne, la température et la pression atmosphérique.

### 3 Prédiction

Pour prédire les hauteurs, il faut utiliser les programmes Matlab fournis. Pour cela, j'exporte les données obtenues précédemment dans un fichier texte : ye.txt, en ayant fait attention à fusionner les deux colonnes de date et heure. Attention toutefois à avoir des chiffres dont la décimale est indiquée par un point et non une virgule. Je supprime les secondes du format de la date dans matlab car elles ne se sont pas copiées dans mon fichier texte (dans tous les cas, les enregistrements sont faits exactement toutes les 6 minutes donc les secondes importent peu). J'obtiens ainsi tous les paramètres qui nous sont nécessaires :

- $z0 = 351.5025$
- les  $A_i, f_i, w_i, (v0 + u)_i, g_i$ .

On récupère la date et l'heure associées aux enregistrements, que l'on convertit en format numérique et auquel on soustrait 1 (cela est dû au fait que sur

Matlab l'origine des temps n'est pas la même que sur Excel), puis on multiplie par 24 pour avoir un temps en heure et non en jours.

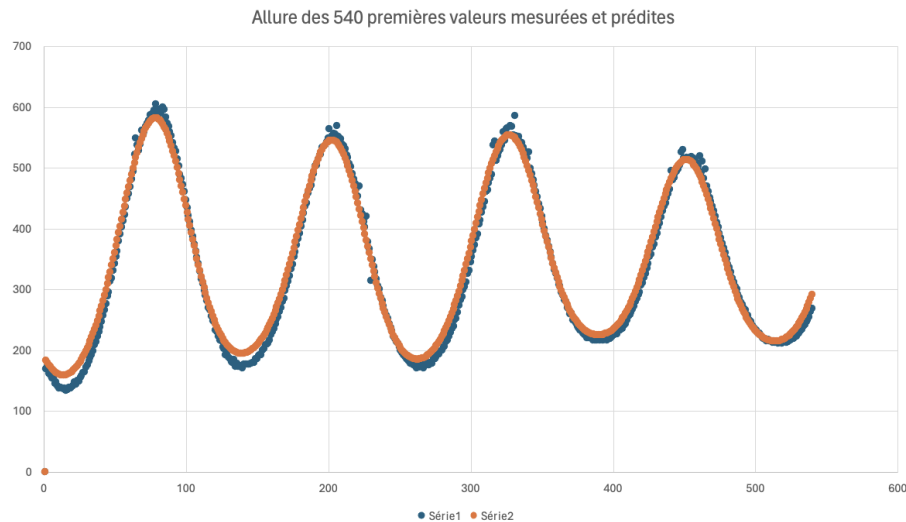
De plus, il faut faire attention à la cohérence en termes de représentation de la décimale : sur Matlab la décimale se distingue par un point, alors que sur Excel, pour le système français, la décimale se distingue par une virgule. De plus, Matlab transforme la date à l'anglaise, nous ne sommes plus dans le format DD-MM-YYYY mais MM-DD-YYYY. Il faut donc faire attention aux dates !

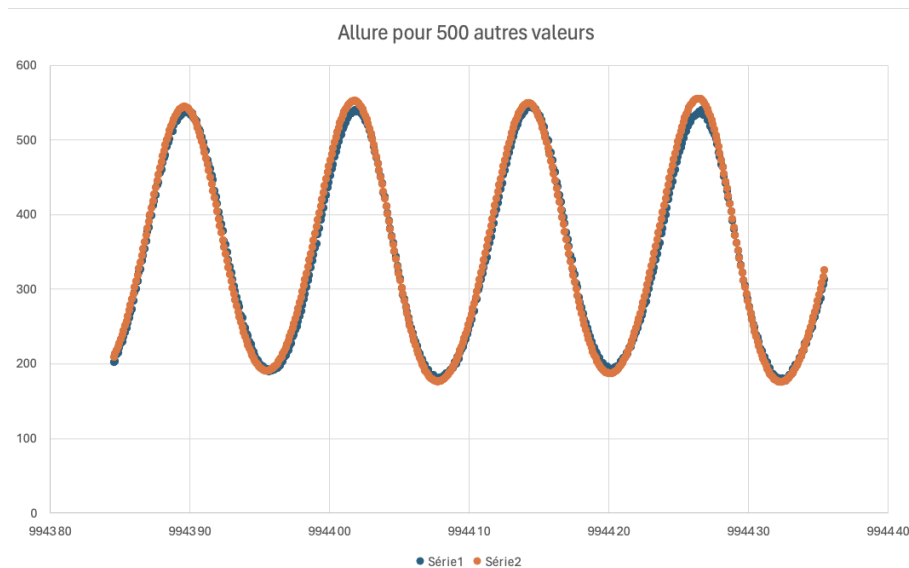
J'ai suivi deux méthodes pour obtenir les hauteurs prédites : une avec VBA, où j'ai pu la calculer par le produit matriciel  $M * X = h_{predit}$  grâce à la formule suivante :  $h(t) = z_0 + \sum_{i=1}^n a_i(t)x_i + b_i(t)y_i$   
Sur 500 valeurs, l'erreur relative mesurée est de 9%.

L'autre méthode consiste à ne pas passer par VBA et tout faire sur Excel directement : les calculs seront plus rapides à réaliser. On importe les données qui nous intéressent dans la feuille "h prédit formule complète", et on convertit les  $w_i$ ,  $(v_0+u)_i$ ,  $g_i$  en gradient.

La formule à suivre est la suivante :  $h(t) = z_0 + \sum_{i=1}^n A_i f_i \cos(w_i t + (v_0+u)_i - g_i)$

On obtient, pour plusieurs échantillons de données, les allures suivantes pour les valeurs mesurées (nuage de point bleu) et les valeurs prédites (nuage de point orange)





On remarque qu'il n'y a pas de déphasage apparent, et que les amplitudes sont proches des amplitudes réelles.

Je calcule l'erreur relative pour chaque mesure, et j'obtiens en moyenne sur toutes les mesures une valeur de 6.19%, ce qui est acceptable. Il semble y avoir moins d'erreurs de prédiction pour cette méthode-ci; cela certainement dû au fait que les données sont moins traitées par Matlab et VBA et davantage par Excel lui-même, il y a donc moins d'interpolations ou d'approximations, ce qui réduit les erreurs potentielles. De plus la méthode sans le produit matriciel est largement préférable car la matrice  $M$  est très lourde (c'est une matrice de taille  $81000 \times 75$ ), les calculs sont donc très longs.

## 4 Aide à la décision

Grâce à la partie II, nous avons pu déterminer une fonction  $h$  ainsi que tous ses paramètres pour pouvoir calculer la hauteur à n'importe quel moment  $t$  dans l'avenir. Pour cette partie III, il s'agira de créer un UserForm dans lequel un marin pourra entrer une certaine date, et qui renverra les heures limites d'entrée et de sortie du port pour son navire. Il faut afficher les Basses et Pleines Mer dans l'ordre d'occurrence dans les 24h suivant l'heure saisie par le marin.

Tout d'abord, il faut rentrer tous les coefficients dans la Feuille Excel "coeffs". Il est possible de procéder de deux façons : soit avec un programme VBA, que voici :

```
Sub LectureFichierCoefficients()
    Dim iRow As Long
```

```

Dim iFile As Integer
Dim Fichier()
Dim donnee() As String
Dim chaine As String

Fichier() = Application.GetOpenFilename(, , , , True)

iRow = 1
iFile = FreeFile

With ThisWorkbook.Worksheets("coeffs")
    Open Fichier(1) For Input As #iFile
    While (Not EOF(iFile))
        Line Input #iFile, chaine
        donnee = Split(chaine, vbTab)
        For i = 1 To UBound(donnee)
            .Cells(iRow, i + 1).Value = donnee(i)
        Next i
        iRow = iRow + 1
    Wend
    Close #iFile
End With
End Sub

```

... Soit passer par PowerQuery (qui est la méthode que j'ai utilisée, puisque j'utilise un Mac et que mon ordinateur n'arrive pas à trouver mon fichier, toutefois le code ci-dessus devrait bien fonctionner sur Windows).

Une fois cela fait, il faut déterminer les maxima et minima les 24h suivant la date entrée par le marin. Pour cela, on recopie les données de la feuille "coeffs" dans des listes sur VBA (cela permettra à l'ordinateur d'éviter les "allers-retours" entre les deux plateformes et donc avoir une complexité moins élevée) :

```

Dim fi(1 To 37) As Double
Dim Ai(1 To 37) As Double
Dim wi(1 To 37) As Double
Dim vou(1 To 37) As Double
Dim gi(1 To 37) As Double
Dim t As Date
Dim ttemp As Double
Dim h As Double
Dim somme As Double
Dim z0 As Double
Dim i As Integer

```



```

Dim tlim As Date
Dim booleen As Integer
Dim havt As Double
Dim hmax As Double
Dim hmin As Double
Dim ligne As Integer
Dim colonne As Integer

```

```

With ThisWorkbook.Worksheets("coeffs")

```

```

    z0 = .Cells(1, 2).Value
    For i = 3 To 39
        fi(i - 2) = .Cells(i, 1).Value
        Ai(i - 2) = .Cells(i, 2).Value
        wi(i - 2) = .Cells(i, 3).Value
        vou(i - 2) = .Cells(i, 4).Value
        gi(i - 2) = .Cells(i, 5).Value
    Next i

```

```

End With

```

Ensuite, toutes les 6 minutes, entre t (la date entrée par le marin) et tlim (qui est le jour après la date t), on calcule la hauteur dans le Sub Calcul\_BM\_PM, puis on compare les valeurs deux à deux les unes après les autres.

Dans le cas des Pleines Mers :

Si la hauteur à l'instant t est supérieure à celle en t-1, cette hauteur devient le nouveau maximum. Il faut également ajouter d'autres conditions : il faut vérifier que la valeur en t+1 est inférieure à celle en t, faire attention que la valeur de la hauteur maximale a bien été modifiée au moins une fois (c'est pour cela que je prends 0 pour valeur aberrante). On vérifie que cette hauteur à cette date précise n'a pas été inscrite grâce au booléen. Voici la partie du code associée au calcul les hautes mers (pour les basses mers le principe est le même) :

```

t = saisieDH_UserForm.getEnteredDate
tlim = t + 1 'On garde en jours
hmax = 0
hmin = 1000
ligne = 6

```

```

With ThisWorkbook.Worksheets("BM-PM")

```

```

    While t < tlim
        ttemp = (t - 1) * 24 ' On convertit en heure pour la fonction h

        somme = 0
        For i = 1 To 37
            somme = somme + fi(i) * Ai(i) * Cos(wi(i) * ttemp + vou(i) - gi(i))
        Next i
    End While

```

```

Next i

h = z0 + somme
' Premier cas : on calcule les hauteurs de Pleine Mer

If (h > havt) Then ' Si la valeur est supérieure à la précédente,
c'est le maximum
    hmax = h

ElseIf (h < havt) And (booleen = 0) And hmax <> 0 Then 'Si la valeur est
inférieure à la précédente, la précédente est le maximum. On vérifie que
cette hauteur à cette date précise n'a pas été inscrite grâce au Booléen
    booleen = 1

ElseIf (h < havt) And (booleen = 1) And hmax <> 0 Then 'Si au cran d'après
on a toujours hmax qui n'est pas modifié, on inscrit hmax
    colonne = 6

    .Cells(ligne, colonne).Value = "Pleine Mer"
    colonne = colonne + 1
    .Cells(ligne, colonne).Value = t
    colonne = colonne + 1
    .Cells(ligne, colonne).Value = hmax
    ligne = ligne + 1

    hmax = 0
    booleen = 0

End If

```

Avec ce code, on peut appuyer sur l'UserForm "saisie date/heure de sortie" et obtenir les valeurs des PM et BM (qui sont toutes les 6h approximativement) :

	B	C	D	E	F	G	H	I
1								
2					date/heure envisagées pour la sortie:			
3					29/04/2024 12:50			
4								
5								
6					Pleine Mer	29/04/2024 13:14	514.492469	
7					Basse Mer	29/04/2024 19:02	212.344019	
8					Pleine Mer	30/04/2024 01:38	525.809407	
9					Basse Mer	30/04/2024 07:26	239.900303	
10								
11								
12								

## 5 Conclusion

Il est désormais possible d'obtenir une prédiction avec une erreur assez faible sur la hauteur de l'eau, et également d'obtenir les horaires des basses mers et hautes mers suivant les 24h inscrites par le marin. Pour aller plus loin, il serait possible de davantage nettoyer les données afin d'avoir une erreur plus faible, en comparant par exemple les valeurs des hauteurs d'une ligne à l'autre ou au sein d'une même ligne, dans le cas des nouvelles données : si la différence est trop grande et qu'une telle variation est impossible, il faudrait supprimer la valeur aberrante.