# Étude de comblement de lacunes : Cas des séries pluviométriques observées du réseau de l'ONM

KERTALI Fouzia [1]*

**Abstract**
Les données manquantes sont fréquentes dans les séries météorologiques et peuvent entraîner d'importantes erreurs dans l'éstimation statistique des paramètres. Ce travail expose les différentes méthodes théoriques de traitement des séries lacunaires. La méthode de double masse est choisie dans cette étude pour rattraper des données manquantes d'une série de données pluviométriques. L'efficacité de cette méthode est mise en exergue après un second test appliqué sur un autre jeu de données.

**Keywords**
comblement de lacune, méthode de double masse, donnée pluviométrique manquante, homogénéité.

[1] *Office national de la météorologie, Dar El Beida, Alger*
***Correspondant**: fouzia.khadidja@yahoo.com

## Contents

**Introduction** 1
**1 Méthode de comblement des données manquantes** 1
**2 Cas d'étude et choix de méthode** 2
2.1 Description des données utilisées . . . . . 2
2.2 Méthode et discussion . . . . . 2
Test de la médiane : appliquée à la série de référence (Dar El Beida) • Méthode des doubles masses
2.3 Calcul du gain . . . . . 5
2.4 Extension de la série $Y$ : . . . . . 5
**3 Test de validation de la méthode** 6
**4 Conclusion** 9
**References** 9

La climatologie est la science du climat. Mais son domaine d'application n'est pas restreint à ce dernier. Il s'agit d'une discipline beaucoup plus vaste. Elle emprunte à d'autres sciences des notions ou des résultats dont elle a besoin en faisant appel à des moyens techniques de plus en plus sophistiqués. On peut citer quelques unes : toutes les sciences concernant l'atmosphère comme la physique, la biologie, l'agronomie, l'hydrologie, l'économie, l'informatique et surtout les statistiques pour le traitement et l'utilisation rationnelle des données. [1] Pourquoi corriger la série climatique ? Parmi les questions qui agitent les scientifiques à propos du changement climatique revient continuellement celle de la comparaison de la variabilité climatique actuelle avec celle du passé. A l'échelle du dernier millénaire, les mesures directes ne sont pas disponibles, il faut donc les reconstruire. Cependant, depuis le XIXème siècle, nous disposons de nombreuses longues séries d'observation instrumentales. Mais leur qualité doit être étudiée au préalable. Par définition, une série d'observation météorologique[2] est dite homogène lorsque les conditions de mesure n'ont pas varié au cours du temps. L'homogénéité ou plutôt l'absence d'homogénéité des longues séries instrumentales en climatologie est un problème connu depuis longtemps, le déplacement des postes climatologique au cours du temps, la modification des sites de mesure, de l'instrumentation, des méthodes de calcul des paramètres météorologiques[3] et les changements d'observateurs, parmi tant d'autres, vont se traduire par autant de sauts dans les séries de données. Or, ces ruptures artificielles peuvent être du même ordre de grandeur que les phénomènes qu'on cherche à mettre en évidence dans les séries climatiques, comme les tendances, les cycles, etc. Leur correction est donc indispensable avant toute étude climatique sérieuse.

## 1. Méthode de comblement des données manquantes

En statistique, on parle de valeur manquante lorsqu'on ne dispose pas d'observation pour une variable donnée ou pour un individu donné. Le problème de la gestion des données manquantes est un vaste sujet. Les données manquantes ne peuvent pas être ignorées lors d'une analyse statistique. Selon leur proportion et leur type, des solutions différentes vont être choisies. On pourra soit tirer les variables ou les individus présentant des données manquantes ou imputer des valeurs aux données manquantes ou encore développer des méthodes (ou algorithme) qui permettent de mener les analyses en présence de données manquantes. En statistique, on définit trois type de données manquantes :

- Données manquantes complètement aléatoire (MCAR)
- Données manquantes aléatoires (MAR).
- Données manquantes non aléatoires (NMAR).

Les modèles mathématiques les plus connues pour combler les données manquantes sont :

a. **Les méthodes d'imputation les plus simples** : consistent à remplacer les données manquantes par leur moyenne ou leur médiane.

b. **Imputation par tirage conditionnel** : on peut améliorer l'idée de l'imputation par la moyenne en réalisant de l'imputation par tirage conditionnel. Le principe est d'utiliser l'information apportée par les variables renseignées. Plusieurs approches sont possibles :

b.1. Estimer la loi jointe et générer conditionnellement une réalisation pseudo-aléatoire de cette loi. Mais il est généralement difficile d'estimer une loi jointe au-delà de deux ou trois variables. Une alternative intéressante et plus facile à mettre en œuvre et qui consiste à utiliser une méthode des plus proches voisins.

b.2. Réaliser une classification à partir des variables complètement renseignées et estimer la moyenne conditionnelle par classe. On peut voir cette approche comme une sorte de généralisation de la méthode des plus proches voisins.

b.3. Construire un modèle de régression à partir des individus complètement renseignés et l'utiliser pour prédire les données correspondantes aux données manquantes.

c. **Imputation par analyse factorielle** : Considérons le cas de données issues de variables quantitatives. L'analyse en composante factorielle permet de reconstruire des données par projection dans un espace de dimension réduite. Cette caractéristique peut être exploitée pour remplacer des données manquantes. L'approche la plus naïve consiste à estimer la matrice de covariance à partir des individus renseignés puis d'estimer les paramètres de l'analyse en composantes principales et enfin à construire les données manquantes.

## 2. Cas d'étude et choix de méthode

Soient deux stations pluviométriques X et Y situées à quelques kilomètres l'une de l'autre. Ces stations ayant fonctionné sur une période connue. En supposant que les séries pluviométriques des précipitations annuelles de la station X est la station de référence et que l'erreur recherchée se trouve au niveau de la station pluviométrique Y (série à étudier). Pour résoudre le problème qui se pose nous avons choisi la méthode de construction de modèle par régression à partir des individus complètement renseignées, on demande donc de :

- Vérifier l'homogénéité de la série de la station de référence X en appliquant le test de la médiane.
- Détecter l'erreur systématique de la station étudiée et faire la correction par la méthode des doubles masses s'il y'à des erreurs.
- Projeter sur un diagramme de dispersion les séries avant et après correction.
- Vérifier l'homogénéité de la série Y.
- Donner la droite de régression de Y en X.
- Calculer le gain obtenu pour l'extension.
- Faire l'extension de la série courte à la série longue.

### Description des données utilisées

Les données utilisées pour cette étude sont les données des précipitations annuelles obtenues de l'Office National de la Météorologie (ONM) et relevées à la station de Dar El Beida qui est une station de référence (ou de base) et la station d'Alger Port qui est supposée être la station à étudier, et ce, sur une période 24 ans allant de 1983 jusqu'au 2018. La série d'Alger port présente des lacunes entre 1995 et 2001.

### Méthode et discussion

L'homogénéisation des données consiste à identifier les séries pluviométriques et vérifier s'il n'existe pas d'erreurs systématiques qu'il convient de rechercher et de corriger s'il y'à lieu pour la fiabilité de l'information et de tester enfin la série de référence utilisée pour d'autres séries.

#### Test de la médiane : appliquée à la série de référence (Dar El Beida)

- $N_s$ : nombre total de série de + ou de – dans la série tels que ; + Pour les $X_i > m$ (médiane) Pour les $X_i < m$
- $T_s$ : taille de la plus grande série de + ou de – au-dessus de la médiane
- $N_s$ suit une loi normale et $T_s$ suit une loi binomiale.

Pour un seuil de signification compris entre 91% et 95 %, les conditions du test sont les suivantes :

$$N_s > \frac{1}{2}(N+1-U_{1-\frac{\alpha}{2}}\sqrt{N+1})$$

$$T_s < 3.3log_{10}(N+1)$$

Où $N$ : représente le nombre total de valeurs de la série de référence Si les conditions du test sont vérifiées, on conclut que la série étudiée est homogène au seuil $1-\alpha$.

**Résultat du test sur la série de Dar El Beida :**

- $N = 36$, $m = 594$, $U_1 - \frac{\alpha}{2} = 1.96$ (lu sur la table de gausse pour un seuil de signification $1-\alpha = 95\%$).
- $N_s = 18$(determiné sur la série)> 12.53(calculé).
- $T_s = 3$ (determiné sur la série) < 8.43 (calculé).

| Année | Dar El Beida | Alger port | Année | Dar El Beida | Alger port |
|---|---|---|---|---|---|
| 1983 | 335 | 363 | 2001 | 443 | - |
| 1984 | 886 | 577 | 2002 | 478 | 528 |
| 1985 | 709 | 379 | 2003 | 733 | 739 |
| 1986 | 748 | 592 | 2004 | 706 | 845 |
| 1987 | 489 | 498 | 2005 | 539 | 739 |
| 1988 | 578 | 458 | 2006 | 609 | 681 |
| 1989 | 320 | 295 | 2007 | 804 | 793 |
| 1990 | 457 | 330 | 2008 | 530 | 707 |
| 1991 | 433 | 368 | 2009 | 621 | 581 |
| 1992 | 758 | 540 | 2010 | 649 | 810 |
| 1993 | 491 | 396 | 2011 | 673 | 838 |
| 1994 | 458 | 479 | 2012 | 853 | 873 |
| 1995 | 553 | - | 2013 | 883 | 908 |
| 1996 | 803 | - | 2014 | 555 | 561 |
| 1997 | 548 | - | 2015 | 439 | 612 |
| 1998 | 611 | - | 2016 | 567 | 575 |
| 1998 | 807 | - | 2017 | 680 | 609 |
| 1998 | 283 | - | 2018 | 630 | 586 |

**Table 1.** Données pluviométriques utilisées dans cette étude

La série de référence est donc homogène. L'homogénéité de la série de référence étant vérifiée, cette dernière servira de base pour la détection des erreurs systématiques dans la série à étudier.
Cependant, les stations pluviométriques auxquelles les séries sont concernées doivent appartenir aux mêmes conditions climatiques.
Il est important d'identifier la station de base ou de référence pour pouvoir détecter et corriger les erreurs de la station à étudier.

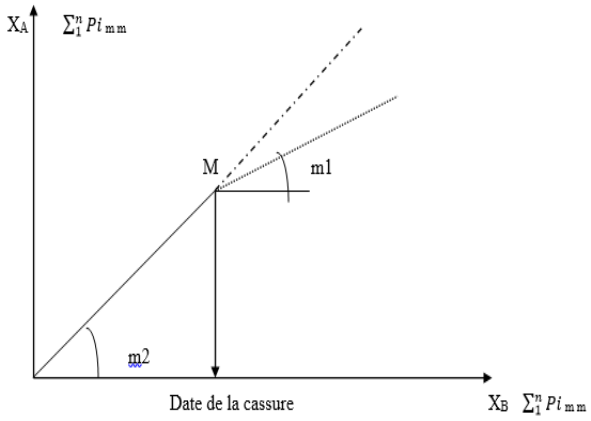


**Figure 1.** Méthode des doubles masses et son principe

#### Méthode des doubles masses

La station à étudier est celle d'Alger Port (Y) qu'il convient de vérifier et de corriger en cas d'erreurs et d'étendre pour son utilisation future. Dans cette étude, la méthode des doubles masses est utilisée. Les valeurs correspondantes à la même période d'observation sont reportées en coordonnées rectangulaires, obtenant une courbe dite courbe de double cumul. Si les données de la station contrôlée sont homogènes par rapport à celles de la station de base, la courbe des doubles cumuls prend une forme de droite (Fig. 1). Si elle possède une cassure à partir d'un point M, les observations à partir de ce point sont considérées hétérogènes. Dans le cas où l'hétérogénéité est détectée, la correction s'effectue par modification de la pente de la droite de double cumul des données antérieures ou postérieures à la date de la cassure.

On considère les données observées en multipliant le rapport de pente $\frac{m_1}{m_2}$ par la valeur erronée dans la série observée.
**Procédé :**
Le tableau 2 représente les valeurs initiales et cumulées des précipitations annuelles au niveau des deux stations pluviométriques. La méthode de la double masse appliquée aux cumuls annuels des deux stations a permis de confirmer l'homogénéité de la série des pluies annuelles de la station (Y) comme le montre la figure 2.

| Année | Précipitations annuelles en mm relevées à la station de base (X) | Précipitations annuelles en mm relevées à la station à étudier (Y) | Cumul à la station (X) en (mm) | Cumul à la station (Y) en (mm) |
|---|---|---|---|---|
| 2002 | 478 | 528 | 478 | 528 |
| 2003 | 733 | 739 | 1211 | 1267 |
| 2004 | 706 | 845 | 1917 | 2111 |
| 2005 | 539 | 739 | 2457 | 2850 |
| 2006 | 609 | 681 | 3066 | 3531 |
| 2007 | 804 | 793 | 3870 | 4324 |
| 2008 | 530 | 707 | 4400 | 5031 |
| 2009 | 621 | 581 | 5021 | 5611 |
| 2010 | 649 | 810 | 5670 | 6421 |
| 2011 | 673 | 838 | 6343 | 7259 |
| 2012 | 853 | 873 | 7196 | 8131 |
| 2013 | 883 | 908 | 8079 | 9039 |
| 2014 | 555 | 561 | 8634 | 9600 |
| 2015 | 439 | 612 | 9073 | 10211 |
| 2016 | 567 | 575 | 9640 | 10786 |
| 2017 | 680 | 609 | 10319 | 11395 |
| 2018 | 630 | 586 | 10950 | 11981 |

**Table 2.** Cumul annuel des précipitations au niveau des 2 stations.

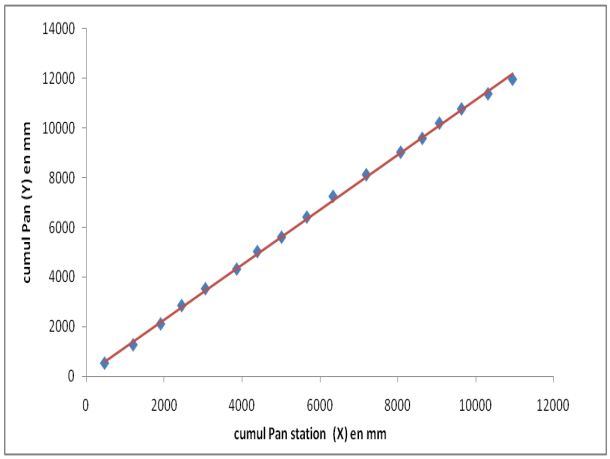


**Figure 2.** Représentation graphique des données

La droite de régression régissant Y en X :

$$Y = 0,733X + 232,6 \quad (1)$$

avec un coefficient de détermination $R^2 = 0,55$ ce qui donne un coefficient de corrélation $r = 0.74$.

**Test de corrélation :**

Le coefficient de corrélation est un indice statistique qui exprime l'intensité et le sens (positif ou négatif) de la relation linéaire entre deux variables. C'est une mesure de la liaison linéaire. C'est-à-dire de la capacité de prédire une variable $X$ par une autre $Y$ à l'aide d'un modèle linéaire.

**Matrice de corrélation (Pearson (n)) :**

| Variables | Alg.DEB | Alg.PORT |
|---|---|---|
| DEB | 1.000 | 0.741 |
| APORT | 0.741 | 1.000 |

Les valeurs en gras sont différentes de 0 à un niveau de signification alpha=0,05

**Test de Sphéricité de Bartlett :**

| | |
|---|---|
| $\chi^2$ (valeur observée) | 11,526 |
| $\chi^2$ (valeur critique) | 3,841 |
| DDL | 1 |
| P-value | < 0,0001 |
| $\alpha$ | 0,05 |

**Interprétation du test :**

$H_0$ : Il n'y a pas de corrélation significativement différente de 0 entre les variables.

$H_a$ : Au moins l'une des corrélations entre les variables est significativement différente de 0.

Étant donné que la p-value calculée est inférieure au niveau de signification
$alpha = 0,05$, on doit rejeter l'hypothèse nulle $H_0$, et retenir l'hypothèse alternative $H_a$.

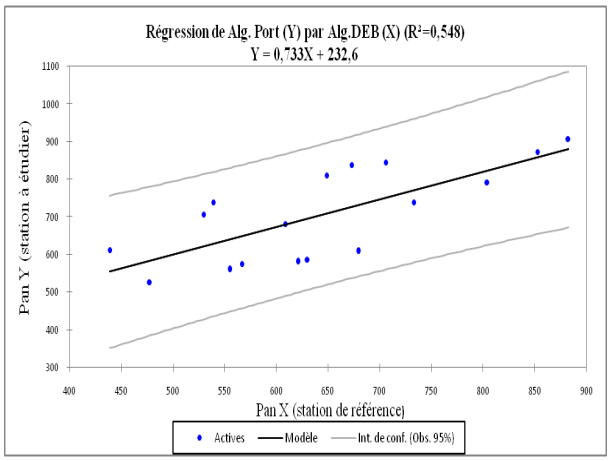


**Figure 3.** Relation entre la station de référence (Alger DEB) matérialisée par (X) et celle d'Alger port par (Y)

Cette droite permettra de faire l'extension de la série Y, c'est-à-dire combler les lacunes de la série des pluies relative à Alger port.

## Calcul du gain

Avant de combler les lacunes,[4] il convient de chercher la taille de la nouvelle série, c'est-à-dire, en d'autres termes, cela revient à calculer le gain. Les séries étant de tailles différentes, il est difficile d'étendre la série courte à la série longue sans préalablement connaître jusqu'à combien de valeurs peut on étendre la série courte.

**Rappel :**

Le bénéfice de l'extension de la série Y à l'aide de la série X pour la connaissance de la série Y est d'autant plus grand que le coefficient de corrélation est élevé.

$$E = 1 + (1 - \frac{N}{K})[\frac{1-(K-2)r^2}{K-3}] \quad (2)$$

$E$: Efficacité relative de $\overline{y_K}$ et de $\overline{\hat{y}}$ définie par le rapport de la variance de $\overline{\hat{y}}$ et celle de $\overline{y_K}$.

Ce bénéfice est traduit, en utilisant E sous forme d'un gain réel d'information que l'on exprime à l'aide du nombre d'années efficaces ou fictives $N'$ à laquelle correspond l'échantillon $Y$ étendu.

$N'$ varie de $K$ à $N$ (gain maximum, liaison fonctionnelle entre $X$ et $Y$ et $r = 1$)

$$N' = \frac{K}{E}$$

avec $K > 3$

Pour notre cas :

$K = 17$ ,$R^2 = 0.55$ d'où $r = 0.74$

$N = 36$, $N' \approx 22$ , $E = 0.78$

Ceci représente une série de vraies valeurs observées dans laquelle on pourrait avoir la même confiance que dans les 17 observations et 7 valeurs reconstituées.

## Extension de la série $Y$ :

On commence d'abord de combler les lacunes par les années les plus récentes en utilisant l'équation 1 régissant la relation de $Y$ en $X$ . Les résultats sont présentés dans le tableau 3 :

Après comblement des lacunes de la station d'Alger Port, le coefficient de corrélation est significatif. Il est passé de 0.74 à 0.83(Fig. 4). La droite de régression régissant la relation Y en X est :

$$Pan(Y) = 0.7333 Pan(X) + 232.6 \quad (3)$$

Avec Pan représente les précipitations annuelles.

**Test de corrélation :**

**Matrice de corrélation :**

| Variables | Alg.DEB (X) | Alg.Port (Y) |
|---|---|---|
| Alg. DEB (X) | 1.000 | 0.837 |
| Alg. Port (Y) | 0.837 | 1.000 |

Les valeurs en gras sont différentes de 0 à un niveau de signification $\alpha = 0.05$

| | |
|---|---|
| $\chi^2$ (valeur observée) | 25,968 |
| $\chi^2$ (valeur critique) | 3,841 |
| DDL | 1 |
| P-value | < 0,0001 |
| $\alpha$ | 0,05 |

**Interprétation du test :**

$H_0$ : Il n'y a pas de corrélation significativement différente de 0 entre les variables.

$H_a$ : Au moins l'une des corrélations entre les variables est significativement différente de 0.

Etant donné que la p-value calculée est inférieure au niveau de signification $\alpha = 0.05$, on doit rejeter l'hypothèse nulle H0, et retenir l'hypothèse alternative Ha.

**L'erreur quadratique moyenne :**

Pour tester la méthode appliquer on calcule l'erreur relatives quadratique moyenne donnée par l'expression suivante :

$$ER = \frac{\sum_1^n (V_O - V_R)}{\sum_1^n (V_O - V_R)^2}$$

$V_O$ : Valeur observée

$V_R$ : Valeur rattrapée

$ER$ calculé = 0.02, Où $n$ est le nombre de valeur rattrapée.

| Année | Précipitations annuelles en mm relevées à la station de base (X) | Précipitations annuelles en mm relevées à la station à étudier (Y) |
|---|---|---|
| 1995 | 553 | 638 |
| 1996 | 803 | 821 |
| 1997 | 548 | 634 |
| 1998 | 611 | 680 |
| 1999 | 807 | 824 |
| 2000 | 283 | 440 |
| 2001 | 443 | 558 |
| 2002 | 478 | 528 |
| 2003 | 733 | 739 |
| 2004 | 706 | 845 |
| 2005 | 539 | 739 |
| 2006 | 609 | 681 |
| 2007 | 804 | 793 |
| 2008 | 530 | 707 |
| 2009 | 621 | 581 |
| 2010 | 649 | 810 |
| 2011 | 673 | 838 |
| 2012 | 853 | 873 |
| 2013 | 883 | 908 |
| 2014 | 555 | 561 |
| 2015 | 439 | 612 |
| 2016 | 567 | 575 |
| 2017 | 680 | 609 |
| 2018 | 630 | 586 |

**Table 3.** Comblement des données manquantes (valeurs en rouge) avec la méthode de double masses.

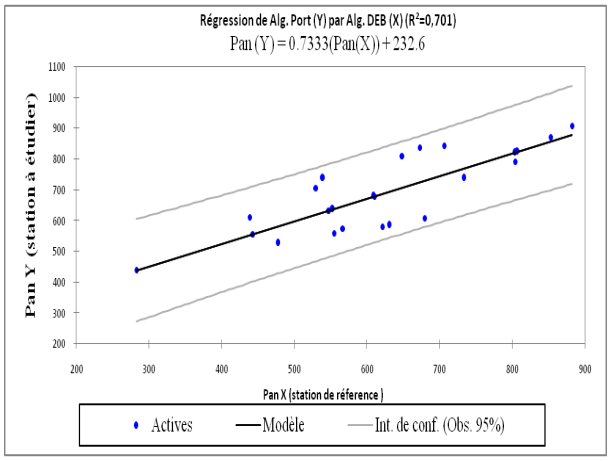


**Figure 4.** Relation de Y en X après comblement de la série lacunaire Y (Alger Port)

## 3. Test de validation de la méthode

Pour valider la méthode, on procédera à une vérification à travers un test sur la série d'Oran Es Senia qui est la station de référence et la station d'Oran port qui sera considérée comme une station à tester. Les données des deux stations étant homogènes ne possèdent pas de lacunes. La méthode consiste à enlever volontairement quelques valeurs de la série d'Oran port et d'essayer d'estimer ces valeurs en appliquant la méthode de double masse utilisée précédemment. On procédera par la suite à la comparaison des résultats obtenus avec les données réelles. On suppose que la série d'Oran port est lacunaire de 1987 jusqu'à 1991 et de 2015 à 2018, les données sont présentées dans le tableau suivant :

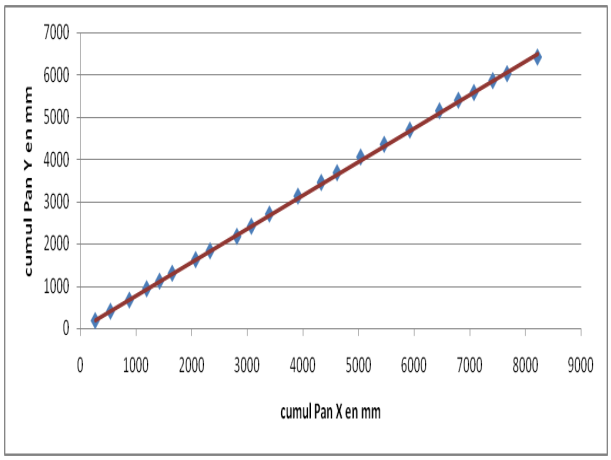


**Figure 5.** Représentation graphique des données du cumul des précipitations issues du tableau N:5.

| Année | Précipitations annuelles en mm relevées à la station de base (X) (Essenia) | Précipitations annuelles en mm relevées à la station à étudier (Y) (Oran port) |
|---|---|---|
| 1988 | 264 | 171 |
| 1989 | 274 | 222 |
| 1990 | 458 | 409 |
| 1991 | 352 | 381 |
| 1992 | 340 | 265 |
| 1993 | 311 | 268 |
| 1994 | 232 | 179 |
| 1995 | 405 | 235 |
| 1996 | 347 | 271 |
| 1997 | 286 | 199 |
| 1998 | 227 | 188 |
| 1999 | 421 | 323 |
| 2000 | 260 | 209 |
| 2001 | 482 | 341 |
| 2002 | 260 | 239 |
| 2003 | 391 | 398 |
| 2004 | 381 | 312 |
| 2005 | 272 | 208 |
| 2006 | 325 | 290 |
| 2007 | 513 | 424 |
| 2008 | 418 | 325 |
| 2009 | 286 | 230 |
| 2010 | 423 | 373 |
| 2011 | 423 | 299 |
| 2012 | 462 | 338 |
| 2013 | 533 | 458 |
| 2014 | 339 | 248 |
| 2015 | 280 | 184 |
| 2016 | 338 | 281 |
| 2017 | 258 | 164 |
| 2018 | 546 | 394 |

**Table 4.** Données de précipitations au niveau des stations d'Oran Es Senia et d'Oran Port. Les valeurs en rouge sont les valeurs qui seront supposées manquantes.

La droite de régression régissant $Y$ en $X$ :

$$Y = 0.778X + 0.917 \quad (4)$$

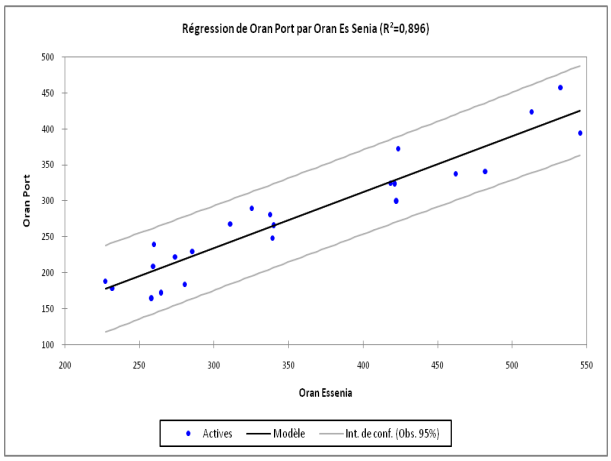


**Figure 6.** Relation entre la station de référence (Es Senia) matérialisée par (X) et celle d'Oran port par (Y)

avec un coefficient de détermination $R^2 = 0,55$ ce qui donne un coefficient de corrélation $r = 0.94$.

| Année | Précipitations annuelles en mm relevées à la station de base Oran Es Senia(X) | Précipitations annuelles en mm relevées à la station à étudier Oran port (Y) | Cumul à la station Oran Es Senia (X)en (mm) | Cumul à la station Oran port(Y) en (mm) |
|---|---|---|---|---|
| 1988 | 264 | 171 | 264 | 171 |
| 1989 | 274 | 222 | 539 | 393 |
| 1992 | 340 | 265 | 879 | 659 |
| 1993 | 311 | 268 | 1190 | 927 |
| 1994 | 232 | 179 | 1422 | 1106 |
| 1998 | 227 | 188 | 1649 | 1294 |
| 1999 | 421 | 323 | 2070 | 1617 |
| 2000 | 260 | 209 | 2330 | 1826 |
| 2001 | 482 | 341 | 2812 | 2167 |
| 2002 | 260 | 239 | 3071 | 2406 |
| 2006 | 325 | 290 | 3396 | 2696 |
| 2007 | 513 | 424 | 3909 | 3120 |
| 2008 | 418 | 325 | 4328 | 3445 |
| 2009 | 286 | 230 | 4613 | 3675 |
| 2010 | 423 | 373 | 5037 | 4048 |
| 2011 | 423 | 299 | 5459 | 4347 |
| 2012 | 462 | 338 | 5921 | 4685 |
| 2013 | 533 | 458 | 6454 | 5143 |
| 2014 | 339 | 248 | 6793 | 5391 |
| 2015 | 280 | 184 | 7073 | 5575 |
| 2016 | 338 | 281 | 7411 | 5856 |
| 2017 | 258 | 164 | 7669 | 6020 |
| 2018 | 546 | 394 | 8214 | 6414 |

**Table 5.** Données de précipitations annuelles relevées au niveau des deux stations (Oran Es Senia et Oran Port) ainsi que Cumul annuel des précipitations au niveau des 2 stations

**Calcul du gain :**

$K = 23$ ; $R^2 = 0.896$ d'où $r = 0.94$

$N = 31$ ; $N' = 29.4$; $E = 0.78$

**Extension de la série Y :**

On commence d'abord de combler les lacunes par les années les plus récentes en utilisant l'équation 4 régissant la relation de $Y$ en $X$. Les résultats sont présentés dans le tableau 6:

Après comblement des lacunes de la station d'Oran Port le coefficient de corrélation est significatif.

Il est passé de $r = 0.88$(Fig. 7) à 0.96 (Fig. 8).

La droite de régression régissant la relation Y en $X$ est :

$$Pan(Y) = 0.792 Pan(X) – 5.348 \quad (5)$$

avec Pan est le cumul des précipitations annuelles.

**L'erreur quadratique moyenne :**

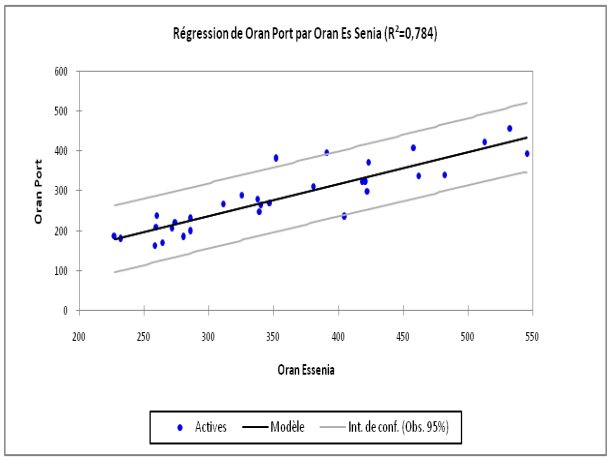


**Figure 7.** Relation entre la station de référence (Oran Es Senia) et celle d'Oran port

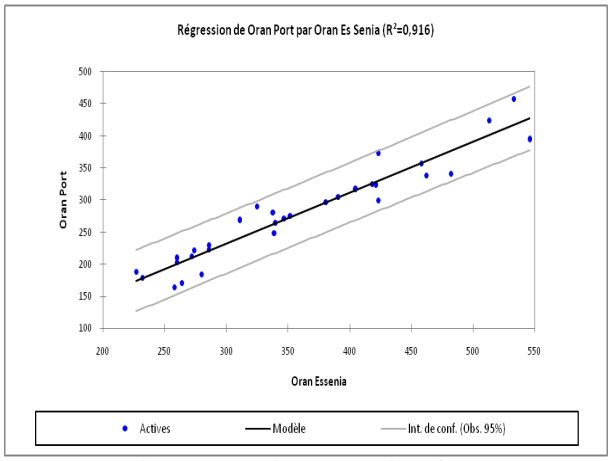


**Figure 8.** Relation entre la station de référence (Oran Es Senia) et celle d'Oran port après comblement

| Année | Précipitations annuelles en mm relevées à la station de base Oran Es Senia(X) | Précipitations annuelles en mm relevées à la station à étudier Oran port (Y) | Précipitation annuelles en mm estimées à la station (Y) à étudier (Oran port) |
|---|---|---|---|
| 1988 | 264 | 171 | - |
| 1989 | 274 | 222 | - |
| 1990 | 458 | 409 | 357 |
| 1991 | 352 | 381 | 275 |
| 1992 | 340 | 265 | - |
| 1993 | 311 | 268 | - |
| 1994 | 232 | 179 | - |
| 1995 | 405 | 235 | 316 |
| 1996 | 347 | 271 | 271 |
| 1997 | 286 | 199 | 223 |
| 1998 | 227 | 188 | - |
| 1999 | 421 | 323 | - |
| 2000 | 260 | 209 | - |
| 2001 | 482 | 341 | - |
| 2002 | 260 | 239 | 203 |
| 2003 | 391 | 398 | 305 |
| 2004 | 381 | 312 | 297 |
| 2005 | 272 | 208 | 212 |
| 2006 | 325 | 290 | - |
| 2007 | 513 | 424 | - |
| 2008 | 418 | 325 | - |
| 2009 | 286 | 230 | - |
| 2010 | 423 | 373 | - |
| 2011 | 423 | 299 | - |
| 2012 | 462 | 338 | - |
| 2013 | 533 | 458 | - |
| 2014 | 339 | 248 | - |
| 2015 | 280 | 184 | - |
| 2016 | 338 | 281 | - |
| 2017 | 258 | 164 | - |
| 2018 | 546 | 394 | - |

**Table 6.** Comparaison entre les valeurs réelles et les valeurs estimées.

$ER = 0.04$ De ce fait, on peut conclure que la méthode appliquée a pu en gros restituer les valeurs réelles malgré quelques différences, ce qui nous permettra de dire que cette méthode pourra constituer une alternative dans le rattrapage des données manquantes.

## 4. Conclusion

Les résultats obtenus dans cette modeste étude nous permettent d'ouvrir un chantier portant sur la reconstruction des longues séries des principaux paramètres climatologiques du réseau d'observation de l'ONM dans notre pays. A travers cette méthode d'imputation, notre choix s'est porté sur la méthode des doubles masses qui reste une méthode dépendante de la disponibilité des données et du type de paramètres étudié. Un cas d'étude effectuée sur la série d'Oran port nous a permis de tester l'application et d'apprécier cette méthode sur des données réelles. Les données rattrapées malgré qu'elles soient différentes en termes de valeurs, cette méthode a pu mettre en exergue une tendance proche de la réalité des quantités de pluie observées et l'erreur quadratique moyenne calculée entre les deux séries réelles et estimées reste relativement faible. Nous recommandons l'emploi d'autres méthodes de comblement de lacunes afin de tirer définitivement les meilleures approches à appliquer qui dépendent fortement de la nature du paramètre météorologique concerné ainsi que la longueur de la période lacunaire.

## References

[1] Yves Brunet-Moret. Etude de l'homogénéité de séries chronologiques de précipitations annuelles par la méthode des doubles masses. *Cah ORSTOM, ser Hydrol*, 8(4):3–31, 1971.

[2] Présidente Mme LAGHA KARIMA Mr ADJLOUA, AZROU Mebrouk, Examinatrice Mme BARACHE AICHA, Rapporteur Mr AISSANI DJAMIL, and Rapporteur Mr AKDIM ABDELGHANI. Analyse statistique du couple pluie-température du bassin versant de la soummam. 2015.

[3] Karima SOLTANI and Mahmoud HAOUARI. Reconstitution des séries mensuelles de températures maximales et minimales sur l'ouest algérien. *JAMA*, 1:83–87, 2017.

[4] l'Institut des sciences mathématiques. Mathématiques de la planète Terre. `http://www.breves-de-maths.fr/pourquoi-corriger-les-series-climatiques/`, 2018. [en ligne; accées 2018].