



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ DES SCIENCES ET DE LA TECHNOLOGIE
HOUARI BOUMEDIÈNE (USTHB)

Laboratoire d'Environnement, Géotechnique et HYDraulique
Faculté de Génie Civil



Journées Nationales sur les Sciences de l'Eau
JONASE'2017

ATTESTATION DE PARTICIPATION

Le Comité d'Organisation des Journées Nationales sur les Sciences de l'Eau
(JONASE'2017) qui se sont tenues à l'Université des Sciences et Technologie Houari
Boumediène les 12 & 13 septembre 2017, atteste que :

DJERBOUAI Salim

A participé par une communication intitulée : « Etude comparative des méthodes
d'estimation des données manquantes dans les enregistrements des précipitations à différents pas
de temps. »

La présente attestation est émise pour servir et valoir ce que de droit.

Alger, le 13 septembre 2017
pour le Comité d'Organisation,
La présidente de JONASE'2017
Pr. Doudja SOUAG



Etude comparative des méthodes d'estimation des données Manquantes dans les enregistrements des précipitations à différents pas de temps.

Salim DJERBOUAI Doctorant USTHB. Laboratoire LEGHYD, USTHB djersalim@yahoo.fr
Doudja SOUAG Maître de conférences Laboratoire LEGHYD, USTHB souagd@yahoo.fr
Saïd ZAMOUM Doctorant USTHB. Laboratoire LEGHYD, USTHB said.zamoum1985@gmail.com

Résumé :

Les enregistrements de précipitations sont généralement liés au problème des données manquantes dû aux défauts de fonctionnement dans les stations pluviométriques. L'évaluation des valeurs manquantes en hydrologie est un problème difficile du fait de la variabilité spatiotemporelle des précipitations et la complexité des processus physiques impliqués. Dans le but de juger les méthodes qui permettent de mieux estimer les données manquantes des précipitations dans le bassin versant de l'algérois au Nord de l'Algérie, nous avons mené une étude comparative entre les méthodes d'estimation: des méthodes classiques et des méthodes basées sur les algorithmes génétiques. Ces méthodes ont été testées sur deux pas de temps, journalier et mensuel, en utilisant les critères de comparaison les plus recommandés. Toutes les méthodes utilisées ont donné de très bons résultats d'estimation excepté la méthode IEWM. Les modèles basés sur les algorithmes génétiques ont donné des résultats plus performants que toutes les autres méthodes.

Mots clés :

Précipitations, algorithmes génétiques, corrélations, méthodes de pondération, ACP.

1 INTRODUCTION

Les précipitations journalières sont probablement les données d'entrée les plus importantes pour la modélisation et la prévision des débits, l'apport des réservoirs et l'étude de qualité de l'eau, ainsi que les études de détection de changements climatiques et de leurs impacts sur les systèmes des ressources en eau sur et l'agriculture. Cependant, dans la pratique, les enregistrements de précipitations sont liés au problème des données manquantes dû aux défauts de fonctionnement de la station d'une part (longues périodes d'arrêt) et de l'équipement d'autre part. Le traitement des données pluviométriques avec observations manquantes est un problème concret et toujours embarrassant, et si cela n'a pas de conséquences pratiques lorsqu'on dispose de données très nombreuses, cela peut supprimer tout intérêt à l'étude si le nombre de données restantes est trop faible. L'estimation des données manquantes des précipitations se fait généralement par : les méthodes de pondération classiques (Smith, 1993), méthodes de pondération basées sur la distance (Simanton et Osborn, 1980 ; Wei et McGuinness, 1973), méthodes déterministes d'interpolation non-linéaires et stochastique. Des études comparatives ont été menées par plusieurs auteurs afin d'évaluer la performance des méthodes classiques d'estimation des valeurs manquantes et de tester de nouvelles méthodes telles que celles basées sur les régressions (régressions spatiales, modèle de régression généralisé (GLM) (Sullivan and Unwin, 2003 ; Pechlivanidis et al., 2005)., les méthodes d'interpolation comme le krigeage et le co-krigeage (Dingman, 2002; Vieux, 2001; Ashraf et al., 1997; Teegavarapu, 2007 ; Seo et al., 1990a, b et Seo, 1996), et dernièrement les méthodes basées sur l'intelligence artificielle et les algorithmes génétiques (Teegavarapua et Chandramouli, 2005 ; French et al., 1992; Tufail and Ormsbee, 2006).

En Algérie, généralement les services hydrologiques reconstituent les observations de précipitations manquantes par les méthodes simples (moyenne, moyennes pondérées par la tendance annuelle), l'analyse en composantes principales ou par l'analyse de régression, qui ne sont pas

toujours applicables, surtout pour les précipitations journalières. L'objectif escompté par ce travail est d'analyser les performances de différentes méthodes, incluant des méthodes classiques et des méthodes basées les algorithmes génétiques, pour l'estimation des données manquantes à deux pas de temps journalier et mensuel sur le bassin versant du côtier algérois.

2 METHODES UTILISEES

2.1 Méthodes classiques

Les méthodes classiques utilisées sont les méthodes basées sur les techniques d'interpolation spatiale telle que la méthode de pondération par la distance inverse (IDWM) (Wei and McGuinness, 1973) et ses différentes variantes, la version modifiée (MIDWM) et la méthode de pondération exponentielle inverse (IEWM) (Teegavarapu et Chandramouli, 2005), la méthode de pondération par le coefficient de corrélation ainsi que la méthode basée sur l'analyse en composantes principales (PCAM) (Labordes, 2002).

2.2 Méthodes basées sur les algorithmes génétiques

Dans ce travail, nous avons utilisé une méthode basée sur les algorithmes génétiques, récemment proposée par Tufail et Ormsbee (2006) : Fixed Function Set Genitic Algorithm Method (FFSGAM) Cette méthode est basée sur le développement d'un modèle inductif, en utilisant les algorithmes génétiques et l'optimisation afin d'obtenir une fonction optimale d'estimation des données manquantes de précipitations. Le processus de recherche de la fonction optimale en utilisant la méthode FFSGAM se fait en deux étapes : i. Rechercher les fonctions optimales des variables de décision (ou modèle inputs) en utilisant les algorithmes génétiques (AG) ; ii. Evaluer les coefficients de la fonction optimale choisit en utilisant l'optimisation.

Le modèle FFSGAM d'estimation des données manquantes de précipitations, est établi en utilisant comme variables de décision: La distance d_{mi} entre chaque station i et celle de base m (dont on veut compléter l'information) ; le coefficient de corrélation R_{mi} entre chaque station i et la station de base. Une fonction prédéfinie des variables de décision à optimiser donnée par :

$$\text{Fonction prédéfinie} = \{C_1[\text{fonction-1}(R_{mi})\text{opérateur-1fonction-2}(R_{mi})]\}\text{opérateur-3} \{C_2[\text{fonction-3}(d_{mi})\text{opérateur-1fonction-4}(d_{mi})]\} \quad (1)$$

où les coefficients C_1 et C_2 sont des nombres réels ; et en fonction de l'opérateur utilisé on peut déduire plusieurs fonctions prédéfinies. La précipitation manquante à la station de base m peut être estimée par :

$$P_m = \frac{\sum_{i=1}^n P_i * (\text{fonction FFSGAM})_i}{\sum_{i=1}^n (\text{fonction FFSGAM})_i} \quad (2)$$

Quatre fonctions FFSGAM ont été utilisées dans cette étude et sont données dans le tableau 1.

Les performances des différentes méthodes d'estimations sont comparées en utilisant deux types de critères d'évaluation : graphiques et numériques. Les critères numériques utilisés sont : la racine de l'erreur moyenne quadratique (RMSE), l'erreur moyenne absolue (MAE) et le coefficient de détermination (R^2).

Tableau 1 Formulations des différentes variantes du modèle FFSGAM utilisées.
Table 1 formulations of the various FFSGAM models used

	Modèle d'estimation
FFSGAM ¹	$P_m = \frac{\sum_{i=1}^n P_i C_i [R_{mi} \log_{10} \left(\frac{1}{R_{mi}}\right) - \left(\frac{1}{R_{mi}}\right) \log_{10} R_{mi}] [\log_{10} \left(\frac{1}{d_{mi}}\right) / \log_{10}(d_{mi})]}{\sum_{i=1}^n C_i [R_{mi} \log_{10} \left(\frac{1}{R_{mi}}\right) - \left(\frac{1}{R_{mi}}\right) \log_{10} R_{mi}] [-\log_{10} \left(\frac{1}{d_{mi}}\right) / \log_{10}(d_{mi})]}$
FFSGAM ²	$P_m = \frac{\sum_{i=1}^n P_i C_i \left[\frac{\exp(R_{mi})}{\left(\frac{1}{R_{mi}}\right) \log_{10} \left(\frac{1}{R_{mi}}\right)} + [\sqrt{d_{mi}} \log_{10} \left(\frac{1}{d_{mi}}\right)] \right]}{\sum_{i=1}^n C_i \left[\frac{\exp(R_{mi})}{\left(\frac{1}{R_{mi}}\right) \log_{10} \left(\frac{1}{R_{mi}}\right)} + [\sqrt{d_{mi}} \log_{10} \left(\frac{1}{d_{mi}}\right)] \right]}$
FFSGAM ³	$P_m = \frac{\sum_{i=1}^n P_i C_i \left[\frac{(\log_{10} \left(\frac{1}{R_{mi}}\right))^2}{R_{mi}} [\log_{10} \left(\frac{1}{R_{mi}}\right)]^2 \right]}{\sum_{i=1}^n C_i \left[\frac{(\log_{10} \left(\frac{1}{R_{mi}}\right))^2}{R_{mi}} [\log_{10} \left(\frac{1}{R_{mi}}\right)]^2 \right]}$
FFSGAM ⁴	$P_m = \frac{\sum_{i=1}^n P_i C_i \left[\frac{R_{mi} \log_{10} \left(\frac{1}{R_{mi}}\right) \log_{10}(R_{mi})}{\left(\frac{1}{R_{mi}}\right) \ln(R_{mi}) / \left(\frac{1}{d_{mi}}\right) \ln(d_{mi})} \right]}{\sum_{i=1}^n C_i \left[\frac{R_{mi} \log_{10} \left(\frac{1}{R_{mi}}\right) \log_{10}(R_{mi})}{\left(\frac{1}{R_{mi}}\right) \ln(R_{mi}) / \left(\frac{1}{d_{mi}}\right) \ln(d_{mi})} \right]}$

3 DONNEES UTILISEES

Nous disposons des données pluviométriques de cinq stations pluviométriques, qui nous ont été aimablement fournies par l'agence nationale des ressources hydraulique (A.N.R.H), Ces stations sont situées dans le bassin versant Côtier Algérois (voir tableau 2).

Les périodes de fonctionnement des stations varient entre 9 et 37 ans. Cependant, nous ne disposons d'observations communes que pour les années 1976, 1980, 1981, 1985, sur lesquelles nous avons fait l'application des différentes méthodes citées précédemment.

Tableau 2 Caractéristiques des cinq stations pluviométriques.
Table 2 characteristics of the rain gauges used

Noms des stations	Codes ANRH	Coordonnées			Années d'observation	Nombre d'années complètes
		X	Y	Z		
Pont CW 7 Hadjout	02 03 25	475.85	358.85	59	1967-2008	09 ans
Bordj Ghobrini	02 03 01	460.5	367.05	15	1967-2008	31 ans
Sidi Amar	02 03 10	464.75	362.75	45	1967-2008	13 ans
Gouraya	02 03 15	429.3	363.8	34	1967-2008	37 ans

4 RESULTATS ET DISCUSSION

Toutes les méthodes sont utilisées pour estimer les données de la station de base m de Sidi Amar, supposées manquantes dans le but de tester l'estimation. Les données journalières et mensuelles des cinq stations sur les quatre années : 1975, 1980, 1981, 1985 soit 75 % des données (1095 jours) sont utilisées pour le calage et 25% (365 jours, 1985) sont utilisées pour la validation.

Pour les méthodes simples basées, les coefficients de pondération ont d'abord été estimés. Pour les méthodes IDWM, IEWM et MIDWM les facteurs de pondération sont des distances d_{mi} entre la station de base m et les stations voisines i , et pour la méthode CCWM ce sont des coefficients de corrélation R_{mi} . Pour les coefficients des fonctions issues des algorithmes génétiques, ils ont été optimisés en utilisant la fonction d'optimisation Solveur de l'Excel.

Les résultats obtenus pour l'estimation des valeurs journalières sont présentés par le tableau 3.

Tableau 3 Critères de comparaison des estimations des valeurs journalières

Table 3 Performace criteria of daily estimated values

Méthode Critère	CCWM	IDWM	MIDWM	IEWM	ACP	FFSGAM ¹	FFSGAM ²	FFSGAM ³	FFSGAM ⁹
RMSE	1,566	1,189	1,186	2,534	1,543	1,033	1,068	1,068	1,068
MAE	0,478	0,400	0,399	0,546	0,492	0,363	0,375	0,375	0,375
R ²	0,893	0,915	0,915	0,691	0,899	0,936	0,933	0,933	0,933

Le tableau 1 montre que toutes les méthodes utilisées ont donné de très bons résultats pour le pas de temps journalier excepté la méthode IEWM et avec une légère supériorité les méthodes issues des algorithmes génétiques. Contrairement à la variante IEWM de l'IDWM, la variante MIDWM a donné des résultats légèrement améliorés par rapport à l'IDWM, ceci nous permet de conclure que dans ce cas les distances d_{mi}^0 obtenues par les polygones de Thiessen, utilisées dans la méthode MIDWM, constituent un meilleur facteur de pondération.

Les méthodes, CCWM et ACP ont donné pratiquement les mêmes résultats d'estimation, ceci peut être expliqué par le fait que les deux méthodes utilisent le principe de la corrélation. La comparaison graphique des différentes méthodes a montré que FFSGAM¹ est celle qui reproduit le mieux les valeurs maximales. La méthode basée sur l'ACP et la CCWM donnent des critères de performance proches de ceux des autres méthodes mais sous-estiment les valeurs maximales. Il faut noter aussi que l'estimation des précipitations journalières mois par mois améliore la performance des méthodes FFSGAM (où le R^2 passe de 0.933 à 0.979 et la RMSE diminue de 1.068 à 0.711) alors que toutes les autres méthodes ne sont pas affectées.

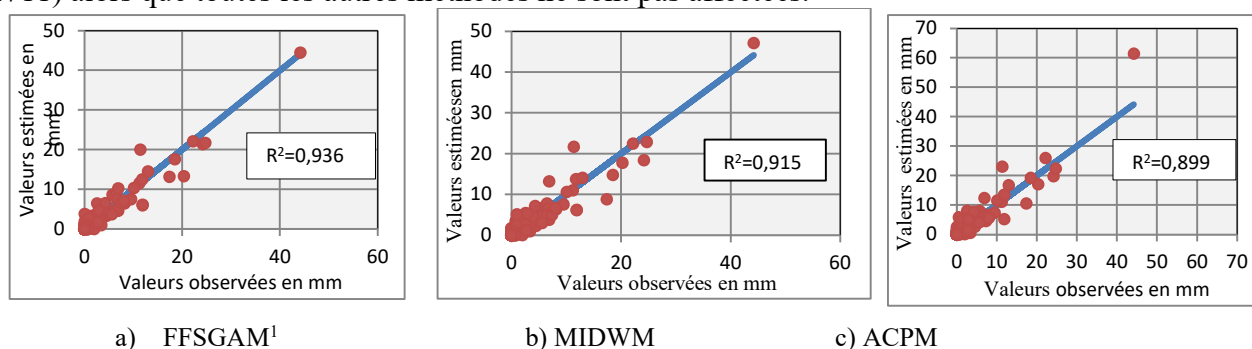


Figure 1 Comparaison entre les valeurs journalières estimées et les valeurs observées

Figure 1 Comparison between observed and estimated daily values

Pour le pas de temps mensuel toutes les méthodes ont conduit à de bons résultats d'estimation avec une supériorité des méthodes FFSGAM comme le montrent les critères de performance présentés dans le tableau 4. En outre, on a comparé les valeurs mensuelles estimées directement et les valeurs mensuelles cumulées obtenues par l'estimation des valeurs journalières et on a trouvé que pour les méthodes FFSGAM et CCWM, les valeurs mensuelles obtenues par le cumul des valeurs journalières estimées sont plus proches des valeurs observées que celles estimées directement pour le pas de temps mensuel ; alors que pour l'ACP il vaut mieux l'utiliser directement pour l'estimation des valeurs mensuelles.

Tableau 4 Critères de comparaison des estimations des précipitations mensuelles.
Table 4 Performance criteria of monthly estimated values

Méthode Critère	CCWM	IDWM	MIDWM	IEWM	ACP	FFSGAM ¹	FFSGAM ²	FFSGAM ³	FFSGAM ⁴
RMSE	8,896	7,116	7,090	14,924	9,386	3,905	3,905	5,361	5,361
MAE	5,258	5,113	5,109	9,839	5,304	3,413	3,413	4,040	4,040
R ²	0,938	0,946	9,947	0,825	0,946	0,984	0,984	0,974	0,974

5 CONCLUSION

Dans ce travail, nous avons mené une étude comparative entre les différentes méthodes d'estimation des données manquantes dans les enregistrements de précipitation. L'étude a été faite sur deux pas de temps, journalier et mensuel. Sur la base des résultats obtenus nous pouvons conclure que, pour les deux pas de temps journalier et mensuel, toutes les méthodes utilisées ont donné de très bons résultats d'estimations excepté la méthode IEWM. La méthode FFSGAM a donné des résultats plus performants que toutes les autres méthodes, cela peut être justifié par le fait que ces modèles utilisent comme facteur de pondération non seulement un seul paramètre comme la distance pour la méthode IDWM et ses variantes, ou comme le coefficient de corrélation dans le cas de la méthode CCWM mais une combinaison de deux paramètres : la distance et le coefficient de corrélation ce qui conduit à mieux estimer les valeurs de précipitations manquantes.

Enfin, les deux méthodes IDWM et MIDWM ont bien reproduit les valeurs maximales elles peuvent donc être recommandées pour l'estimation des valeurs journalières pour leur simplicité mais pour des résultats plus performants, les méthodes FFSGAM sont recommandées.

Références

- Ashraf M, Loftis JC, Hubbard KG. (1997) *Application of geostatistics to evaluate partial weather station network*. Agr Forest Meteorol; 84: 255–271.
- Dingman SL. (2002) *Physical Hydrology*. Prentice Hall, NJ.
- French MN, Krajewski WF, Cuykendal RR. (1992) Rainfall forecasting in space and time using a neural network. J Hydrol; 137: 1–37.
- Laborde JP. (2003). *Hydrologie de surface*. Presses de Dar El-Houda Ain M'lila Algerie.
- Pechlivanidis IG, Segond ML, McIntyre N, Wheater HS. (2005) *Optimal functional forms for estimation of missing precipitation data*. <http://3imperial.ac.uk/pls/porfallive/doc/1/35540/pdf>.
- Seo DJ. (1996) *Nonlinear estimation of spatial distribution of rainfall – an indicator cokriging approach*. Stoch Hydrol Hydraulics; 10: 7–150.
- Seo DJ, Krajewski WF, Bowles DS. (1990a) *Stochastic interpolation of rainfall data from rain gauges and radar using cokriging – 1.Design of experiments*. Water Resour Res; 26 (3): 469–477.
- Seo DJ, Krajewski WF, Bowles DS. (1990) *Stochastic interpolation of rainfall data from rain gauges and radar using cokriging – 2.Results*. Water Resour Res 26 (5): 915–924.
- Smith JA. (1993) *Precipitation*. In: Maidment, D.R. (Ed.), *Handbook of Hydrology*, vol. 3. McGraw Hill, New York.
- Sullivan DO, Unwin DJ. (2003) *Geographical Information Analysis*. John Wiley & Sons, Inc., NJ.
- Teegavarapu RSV. (2007) *Use of universal function approximation in variance dependent interpolation technique: An application in hydrology*. J Hydrol; 332: 16–29.
- Teegavarapu RSV, Chandramouli V. (2005) *Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records*. J Hydrol; 312: 191–206.
- Tufail M, Ormsbee LE. (2006) *A fixed functional set genetic algorithm (FFSGAM) approach for functional approximation*. IWA J Hydroinformatics; 3: 193–206.
- Vieux BE. (2001) *Distributed Hydrologic Modeling using GIS*, Water Science and Technology Library. Kluwer Academic Publishers.

*Journées Nationales
sur les Sciences de l'Eau
JONASE'2017
12 et 13 Septembre 2017*

Président du Comité Scientifique

Pr. BOUHAFEF Malek

Comité Scientifique

**ABBAS Leila
ACHOUR Bachir
AKCHICHE Mustapha
BENMAMAR Saadia
BENDAHGANE Mohamed
BERRAMA Mohamed
BERMAD Abdelmalek
BOUZELHA Karima
CHERRARED Merzouk
DEBIANE Mohamed
DEBIECHE Messaouda
HADJ RABIA Nacima
KETTAB Ahmed
KHELLAF Mohamed Cherif
LARIBI Boualem
MEDDI Mohamed
MEDDOUR Mohamed Salah Eddine
MEDJERAB Abderrahmane
MOUZAI Liatim
REMINI Boualem
SAFRI Abdelhamid
SOUAG Doudja
TOUAIBIA Bénina
ZEGADI Rabah**

Président du comité d'organisation

Pr. SOUAG Doudja

Comité d'organisation

**BENBACHIR Mohamed
BOURAS Fatma Zohra
DEBIECHE Messaouda
DJERBOUAI Salim
KHELLAF Mohamed Cherif
KOUIDER Khaled
LARABI Abdelkader
MAALIOU Aziz
MOUSSOUNI Abderzak
MOULAI Leila
SAFRI Abdelhamid
TOUHAMI Houssam Eddine
YOUNSI Abdelkader**

*Journées Nationales sur
les Sciences de l'Eau
JONASE'2017
12-13 Septembre 2017*



Organisées par

**Université des Sciences
et Technologie Houari
Boumediene**

**Laboratoire d'Environnement
Géotechnique et Hydraulique
LEGHYD**

Présentation et Objectifs

La maîtrise des ressources en eau est la clé de toute politique visant un développement durable. L'Algérie, durant les dernières décennies, a consacré des moyens très importants pour la réalisation d'un programme ambitieux de mobilisation des ressources en eau superficielles sous forme de retenues collinaires, de barrages réservoirs et de transferts dans le but de répondre à la satisfaction de besoins de plus en plus grands associés à une croissance démographique rapide et des aléas climatiques difficilement maîtrisables.

Les journées sur l'Hydraulique ont pour objectif de réunir les chercheurs et les gestionnaires dans le domaine des ressources en eau. Cette rencontre sera une occasion pour débattre de la problématique de l'eau en Algérie en abordant différents aspects qui concernent les systèmes de captage, de stockage, de transfert et de distribution et gestion de l'eau. Ce sera également une occasion pour échanger les

multiples expériences et résultats de recherche entre les universitaires algériens et les différents auteurs. C'est également l'occasion de promouvoir la coopération entre l'Université et les autres structures et établissements travaillant dans le domaine de l'hydraulique.



Thèmes

Thème 1: Hydraulique

Ce thème est consacré à l'étude et la modélisation des eaux de surface et souterraines, la gestion des eaux pluviales et usées ainsi que la géotechnique et ouvrages hydrauliques, etc..

Thème 2: Hydrologie

Ce thème englobe tout ce qui a trait au changement climatique, les inondations, les risques naturels et les aménagements des cours d'eau, la modélisation hydrologique, la télédétection et SIG, l'érosion et le transport solide, etc..

Thème 3: Eau et Environnement

Gestion des ressources en eau, qualité des eaux, pollution et protection des eaux, traitement et épuration des eaux, etc..



Format

Les articles doivent être soumis sous format Word selon le Modèle téléchargeable sur le site www.hydrofgc.webnode.fr



Dates Importantes

14 Juillet 2017: Date limite de réception du texte des communications
05 Aout 2017: Notification d'acceptation
15 Juillet 2017 :

Secrétariat

Contacts: jonase2017@usthb.dz
secret_jonase17@uasthb.dz

Site web: <http://leghyd.usthb.dz>

Frais de participation

Etudiant: **2000 DA**
Non Etudiant: **4000 DA**
A noter que les frais couvrent l'acte des conférences, les pauses café et les déjeuners.