Débruitage des Séries Temporelles de Coordonnées de Stations de Géodésie Spatiale : Application aux Stations DORIS

S. KHELIFA , B. GHEZALI
Centre National des Techniques Spatiales (CNTS), Division de Géodésie
PB 13, Arzew -31200- Algérie, Email : khelifa_sofiane@yahoo.fr

Résumé : Le présent article a pour objectif de présenter une méthodologie de débruitage des séries temporelles de coordonnées de stations de géodésie spatiale. La technique de débruitage adoptée est basée sur le seuillage des coefficients d'ondelettes obtenus à partir d'une décomposition discrète décimée du signal en utilisant deux méthodes : la méthode du seuil universel (cas d'un bruit blanc) et la méthode de seuillage dépendant du niveau (cas d'un bruit coloré). La méthodologie développée repose sur l'étude de la stationnarité des séries en traçant leur fonction d'autocorrélation, la détermination du type de bruit des séries par la variance d'Allan et le débruitage des séries par la transformée en ondelettes.
Les tests effectués ont porté sur l'analyse de séries hebdomadaires de jeux de coordonnées résiduelles de deux stations DORIS issues du centre d'analyse IGN/JPL et exprimées dans le repère géodésique local après enlèvement du modèle ITRF2000 de positions et vitesses.
Les résultats obtenus ont montré que la plupart des séries sont affectées par un bruit blanc et que la méthode optimale de débruitage basée sur le critère de minimisation de l'écart type est celle du seuillage doux.

Mots clés : Série temporelle, Débruitage, Ondelettes, Variance d'Allan, Fonction d'autocorrélation, DORIS.

Abstract : The objective of this paper consists in proposing a methodology of denoising of space geodesy stations coordinates time series. The denoising technique is based on the shrinkage wavelet transform coefficients obtained by a decimated discrete decomposition of the signal. We use VisuShrink method when the series is affected by a white noise and level-dependent thresholding method in the case of coloured noise.
The tests effectuated concern the analysis of weekly series of residual co-ordinates sets of two DORIS stations provided by IGN/JPL analysis centre and expressed in the local geodetic reference frame after removal of ITRF2000 model of positions and velocities.
The results obtained show that majority of the studied series are affected by a white noise and the optimal denoising method is the soft thresholding one.

Key words : Time series, Denoising, Wavelet, Allan variance, Autocorrelation function, DORIS.

1. Introduction

La richesse des mesures collectées par les systèmes spatiaux de géodésie spatiale permet aujourd'hui de représenter le déplacement des stations au sol sous forme de séries temporelles de coordonnées. Au-delà de la modélisation déterministe habituelle des phénomènes géophysiques, un nouveau domaine de recherche en géodésie s'est ouvert il y a quelques années, visant à la validation et à l'interprétation de ces séries temporelles.
L'objectif principal de cet article consiste à évaluer la qualité de ces séries temporelles de coordonnées de stations en faisant arbitrairement la distinction entre le "signal bruit" et le "signal utile" , sans se préoccuper de la valorisation des résultats par leur interprétation.
La méthodologie d'analyse adoptée comporte trois (03) étapes :

1. Étude de la stationnarité ou de la non stationnarité des séries en traçant leur fonction d'autocorrélation.
2. Détermination du type de bruit des séries par la variance d'Allan.
3. Débruitage des séries par la transformée en ondelettes.

Les données utilisées lors du traitement sont constituées de séries hebdomadaires de jeux de coordonnées résiduelles de deux stations DORIS issues du centre d'analyse IGN/JPL conduites entre 1993 et 2003. Ces séries de jeux de coordonnées résiduelles sont exprimées dans le repère géodésique local après enlèvement du modèle ITRF2000 de positions et vitesses (mouvement tectonique des plaques).

2. Débruitage des séries : transformée en ondelettes

Depuis quelques années, les ondelettes sont apparues comme un nouvel outil de débruitage notamment grâce aux travaux de Donoho et Johnstone [Donoho & Johnstone, 94 et 95].
La majorité de ces algorithmes utilisent une décomposition discrète décimée du signal (décomposition par bancs de filtres) [Mallat, 89]. Cette décomposition a la particularité d'être orthogonale et de concentrer l'information en quelques grands coefficients d'ondelettes. L'idée de débruitage est de conserver que les plus grands coefficients (signal utile) et de mettre les autres (bruit) à zéro avant de reconstruire le signal.

2.1 Étapes de débruitage

Supposons que l'on ait des observations bruitées X_t d'un signal original informatif s_t :

$$X_t = s_t + r_t \quad \text{avec :} \quad t = 1, 2, \ldots, N \qquad (1)$$

Où r_t est un bruit blanc gaussien centré indépendant et identiquement distribué de variance σ^2.

Soient $\mathcal{W}(\cdot)$ et $\mathcal{W}^{-1}(\cdot)$ respectivement les opérateurs direct et inverse de la transformée en ondelettes discrète et $D(\cdot,\lambda)$ l'opérateur de débruitage associé au seuil λ.

La procédure de débruitage pour estimer le signal informatif s se traduit en trois étapes :

- La décomposition : le signal est décomposé dans une base d'ondelettes ; $W = \mathcal{W}(X)$
- Le seuillage des coefficients d'ondelettes (coefficients de détails) ; $Z = D(W, \lambda)$
- La reconstruction : à partir des coefficients seuillés, on reconstruit le signal ; $\hat{s} = \mathcal{W}^{-1}(Z)$

2.2 Stratégies de seuillage

Nous proposons dans cet article, deux types de fonctions de seuillage (notée T_λ avec λ un seuil) [Donoho, 95] :

- *Seuillage dur* : soit W les coefficients d'ondelettes, et λ la valeur du seuil (λ>0). Si la valeur absolue du W est supérieure à λ, on garde ce coefficient tel quel, sinon on le met à zéro.

$$T_\lambda^{Dur}(W) = \begin{cases} W \; si \; |W| > \lambda \\ 0 \; si \; |W| \le \lambda \end{cases}$$

- *Seuillage doux* : opère, en plus de l'annulation, une soustraction du seuil λ aux valeurs restantes, supérieures au seuil afin de s'assurer d'avoir enlevé l'effet du bruit même pour les forts coefficients.

$$T_\lambda^{Doux}(W) = \begin{cases} W - \lambda & si \; W \ge \lambda \\ W + \lambda & si \; W \le -\lambda \\ 0 & si \; |W| \le \lambda \end{cases}$$

2.3 Détermination du seuil

Les méthodes de seuillage proposées dans ce travail sont en l'occurrence celle de *VisuShrink* appliquée dans le cas d'un bruit blanc, et celle de seuillage dépendant du niveau dans le cas d'un bruit coloré. Elles sont basées sur la minimisation de l'erreur quadratique moyenne.

2.3.1 Seuil universel (VisuShrink)

Cette méthode propose un seuil universel λ qui ne dépend que du nombre de mesures N et de la variance du bruit σ^2. Elle est utilisée quand le bruit est indépendant et identiquement distribué suivant une loi normale centrée (bruit blanc) ; Sa valeur est donnée par [Donoho&Johnstone,94] :

$$\lambda = \hat{\sigma}\sqrt{2\log(N)} \qquad (2)$$

La variance du bruit σ^2 n'est pas toujours connue. On peut en calculer un estimateur robuste à partir de la médiane (Med) des valeurs absolues des coefficients de détail à l'échelle la plus fine comme suit [Donoho & Johnstone, 94] :

$$\hat{\sigma} = \frac{\text{Med}}{0.6745} \qquad (3)$$

Le facteur 0.6745 est choisi suite à une calibration avec une distribution Gaussienne.

2.3.2 Seuil dépendant du niveau

Un problème posé par le seuillage universel est la valeur unique du seuil pour toutes les bandes de fréquence de la décomposition en ondelettes ou, autrement dit, sa difficulté à prendre en compte un bruit coloré. Pour remédier à ce problème, nous proposons une méthode basée sur un seuil dépendant du niveau, donné par [Johnstone & Silverman, 97] :

$$\lambda_l = \sigma_l \sqrt{2\log\left(\frac{N}{2^l}\right)} \qquad (4)$$

Avec : σ_l=Med($|W_l|$)/0.6745 et W_l représente les coefficients d'ondelettes à l'échelle l du signal bruité (original).

3. Détermination du type de bruit des séries : variance d'Allan

Les bruits observés peuvent être de deux natures différentes [Agnew, 92] :

- *Le bruit blanc* : c'est un bruit non corrélé avec le temps dont l'amplitude peut être facilement diminuée en augmentant le nombre de mesures ;
- *Le bruit coloré* (bruit de scintillation ou bruit de marche aléatoire) : c'est un bruit corrélé avec le temps.

Ces bruits sont bien connus dans le domaine spectral où on les détermine à partir de la fonction de densité spectrale de puissance. Le spectre de puissance S_X d'un grand nombre de phénomènes géophysiques est bien approximé en fonction de la fréquence f par la relation suivante [Agnew, 92] :

$$S_X(f) = P_0 f^{\alpha}$$

Où α représente l'indice spectral. P_0 est une constante.

Les bruits sont définis en fonction de leur *indice spectral* :

- *le bruit blanc* a un indice spectral égale à 0 ;
- *les bruits colorés* ont des indices spectraux différents de 0 :

- indice spectral = –1 : *bruit de scintillation (bruit de flicker)*,
- indice spectral = –2 : *marche aléatoire (random walk)*.

Habituellement, on obtient une estimation des différents types de bruits présents dans un signal par l'utilisation d'une variance. Les variances sont des outils statistiques définis dans le domaine temporel par leur séquence de calcul et dans le domaine des fréquences par leur fonction de transfert. En faisant varier le temps d'intégration t de la variance, on obtient une courbe caractérisant le type de bruit présent dans le signal.

La variance utilisée dans cet article est la *variance d'Allan* ou *variance à deux échantillons* qui a été développée dans le domaine temps/fréquence [Allan, 66, 87 ; Rutman, 78 ; Azoubib, 74]. Elle est définie comme le recouvrement de la variance à différentes espaces de temps d'une série $(X_j)_{j=1,N}$ régulière sur un intervalle de temps constant de τ_0 et stationnaire (propriétés statistiques invariantes au cours du temps).

Pour un temps d'échantillonnage donnée τ (τ multiple de τ_0 tel que $\tau = M\tau_0$), découpons l'échantillon initial $(X_j)_{j=1,N}$ en sous échantillons de longueur τ (figure 1) que l'on note :

$$(X_K)_{K=\{i,i+M-1\}},\ i \in \{1, N-M+1\}$$

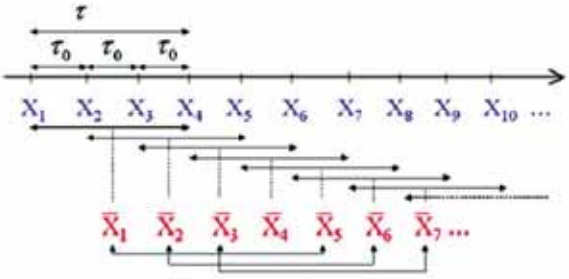

Fig. 1 Principe de sous échantillonnage par la variance d'Allan.

La moyenne de chacun de ces sous échantillons est :

$$\overline{X}_{K,M} = \frac{1}{M}\sum_{j=K}^{K+M-1} X_j,\ k \in \{1, N-M+1\}$$

Avec : $M = \dfrac{\tau}{\tau_0}$

La variance d'Allan pour le temps d'échantillonnage τ est alors définie par :

$$\sigma_X^2 = \left(\frac{1}{2}E\left[\left(\overline{X}_{K+M,M} - \overline{X}_{K,M}\right)^2\right]\right. \qquad (5)$$

L'estimation de la variance d'Allan est [Le Bail, 04]:

$$\hat{\sigma}_X^2(\tau) = \frac{1}{2(N-2M+1)}\sum_{K=1}^{N-2M+1}\left(\overline{X}_{K+M,M} - \overline{X}_{K,M}\right)^2 \qquad (6)$$

La représentation de la variance d'Allan (figure 2) permet de déterminer le type de bruit de la série en fonction de la pente μ du graphe [Le Bail, 04]:

$$\log(\hat{\sigma}_X^2(\tau)) = \mu \log(\tau)\ ,\ \text{pour}\ \ \tau = \tau_0,\ 2\tau_0, 4\tau_0, \ldots$$

- Si μ = - 1 : c'est un bruit blanc (densité spectrale indépendante de la fréquence f) ;
- Si μ = 0 : c'est un bruit de scintillation (densité spectrale variant en 1/f) ;
- Si μ = 1 : c'est un bruit de marche aléatoire (densité spectrale variant en $1/f^2$).

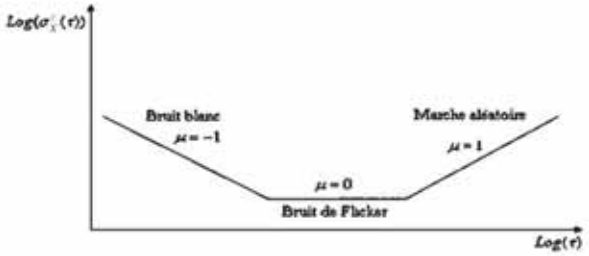

Fig. 2 Représentation de la variance d'Allan en échelle Log/Log.

L'ordonnée représente la variance d'Allan et l'abscisse est le temps d'échantillonnage τ.

4. Application

4.1 Données utilisées

Les données utilisées dans le traitement sont constituées de séries hebdomadaires de jeux de coordonnées résiduelles (dN : composante Nord, dE: composante Est et dH : composante Verticale) de deux (02) stations DORIS (NOUA et SYPB) issues du centre d'analyse IGN/JPL (tableau 1). Ces séries de jeux de coordonnées résiduelles sont exprimées dans le repère géodésique local après enlèvement du modèle ITRF2000 de positions et vitesses (mouvement tectonique d'ensemble).

Tableau 1. Les deux stations étudiées.

Nom station	Pays	Date début et fin	Durée (ans)	Nbre d'obs.
NOUA	Nouvelle-Calédonie	1993-2000.5	7.5	347
SYPB	Antarctique	1999.2-2003.6	4.4	225

4.2 Méthodologie de débruitage

La méthodologie adoptée pour le débruitage des séries temporelles DORIS est composée de trois (03) étapes [S. Khelifa, 2007] :

1ère étape : Identification de la stationnarité (ou de la non stationnarité) des séries.
La non stationnarité d'une série temporelle est caractérisée par la très lente décroissance de sa fonction d'autocorrélation.

2ème étape : Détermination du type de bruit des séries.
Cette étape porte sur la détermination du type de bruit d'une série temporelle régulière et stationnaire par la variance d'Allan. Si elle n'est pas stationnaire, il faut la rendre stationnaire par différentiation : à partir d'une série X, on définit une nouvelle série Y qui à l'instant *t* fait correspondre la différence entre la valeur de la série observée à l'instant t et celle observée à l'instant *t*-1 : $Y_t = X_t - X_{t-1}$.

3ème étape : Débruitage des séries.
Cette étape consiste à séparer le bruit du signal utile. La technique de débruitage utilisée est basée sur le seuillage des coefficients d'ondelettes obtenus à partir d'une décomposition discrète décimée du signal (décomposition par bancs de filtres) en utilisant la méthode de VisuShrink (cas d'un bruit blanc) et la méthode de seuillage dépendant du niveau (cas d'un bruit coloré). L'ondelette utilisée dans cette analyse est celle de Daubechies à 6 moments nuls jusqu'à la 3ème décomposition

4.3 Résultats et analyse

La figure 3 représente les séries originales de coordonnées résiduelles des stations NOUA et SYPB suivant les trois composantes Nord (dN), Est (dE) et Verticale (dH).

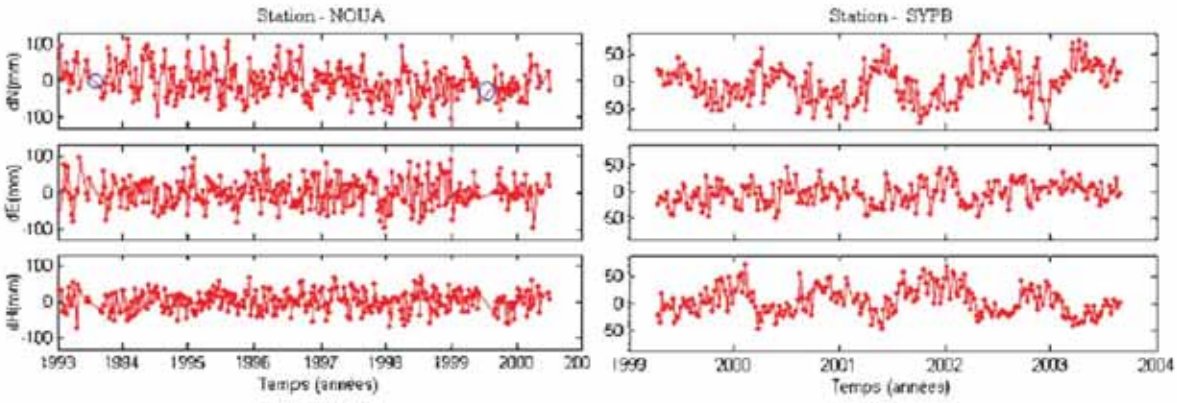

Fig. 3 Résidus sur les composantes (dN, dE et dH) des stations NOUA et SYPB.

D'après la figure 3, nous remarquons que :

- La durée d'observation pour la station NOUA (≈ 8 ans) est plus longue que celle pour la station SYPB (≈5 ans).

- Les trois composantes (dN, dE et dH) de la station NOUA ne semblent pas avoir de tendance ou d'un effet saisonnier. Par contre, les composantes (dN et dH) de la station SYPB semblent périodiques mais sans tendance apparente.

- Au cours des périodes 1993-1994 et 1999-2000 il y a des trous d'observations pour la station NOUA (cercle bleu).

La figure 4 représente la fonction d'autocorrélation (corrélogramme) pour les trois composantes (dN, dE et dH) des stations NOUA et SYPB.

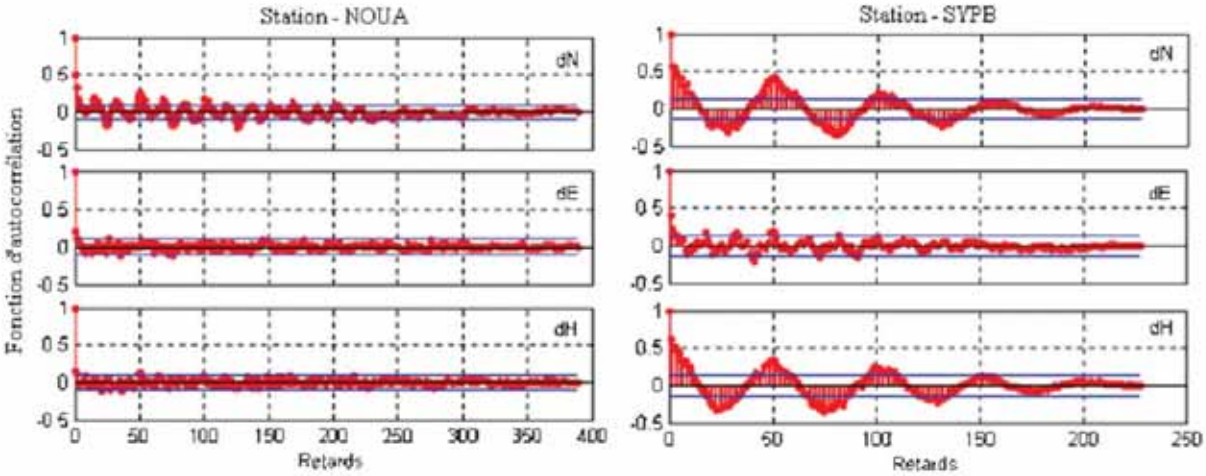

Fig. 4 Corrélogrammes des composantes (dN, dE et dH) des stations NOUA et SYPB (lignes en bleu représentent l'intervalle de confiance à 95%).

On constate clairement (figure 4) que les séries (dE) et (dH) de la station NOUA et la série (dE) de la station SYPB sont stationnaires. Par contre, la série (dN) de la station NOUA et les séries (dN) et (dH) de la station SYPB semblent être stationnaires, expliqué par l'existence d'une légère tendance (évolution à long terme). En effet, on observe un phénomène périodique de période de 52 semaines avec diminution progressive au cours du temps.

Les graphes suivants représentent la variance d'Allan des composantes (dN, dE et dH) des stations NOUA et SYPB. On constate clairement que le bruit affectant les séries (dE) et (dH) de la station NOUA (figure 5) et la série (dE) de la station SYPB (figure 6) est un bruit blanc. La figure 5 révèle que le bruit affectant la série (dN) de la station NOUA est une combinaison d'un bruit blanc dominant et d'un bruit de marche aléatoire faible. La figure 6 montre que le bruit affectant les séries (dN) et (dH) de la station SYPB est un bruit de scintillation.

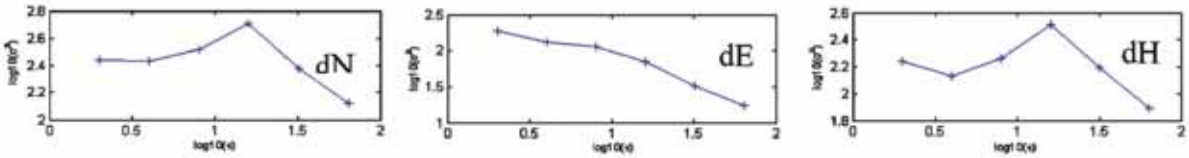

Fig. 5 Variances d'Allan des composantes (dN, dE et dH) de la station NOUA.

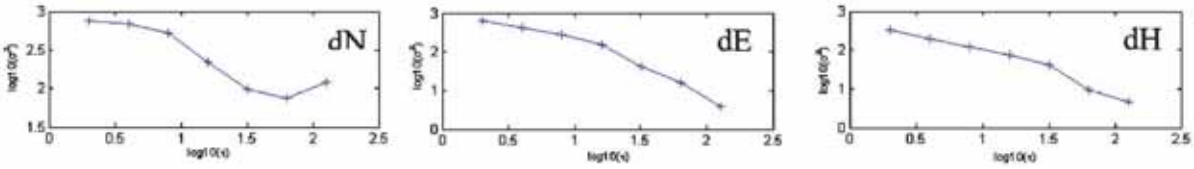

Fig. 6 Variances d'Allan des composantes (dN, dE et dH) de la station SYPB.

La figure 7 représente les séries originales (dN, dE et dH) des stations NOUA et SYPB et leurs séries débruitées par les deux stratégies de seuillage doux et dur. Les séries ont été débruitées en fonction du type de bruit affectant chaque série. Cependant, on a utilisé la méthode de seuillage universel pour les séries (dN, dE et dH) de la station NOUA et la série (dE) de la station SYPB, et la méthode de seuillage dépendant du niveau pour les séries (dN et dH) de la station SYPB.

Nous remarquons que les séries reconstruites (débruitées) suivent effectivement les variations de leurs séries originales. En outre, nous constatons des irrégularités (oscillations parasites) des séries reconstruites par le seuillage dur en quelques points ; ce qui est expliqué par le caractère discontinu de la fonction de seuillage dur. Tandis que pour les séries débruitées par un seuillage doux, nous remarquons qu'elles sont régulières; ceci est lié au caractère continu de la fonction de seuillage doux.

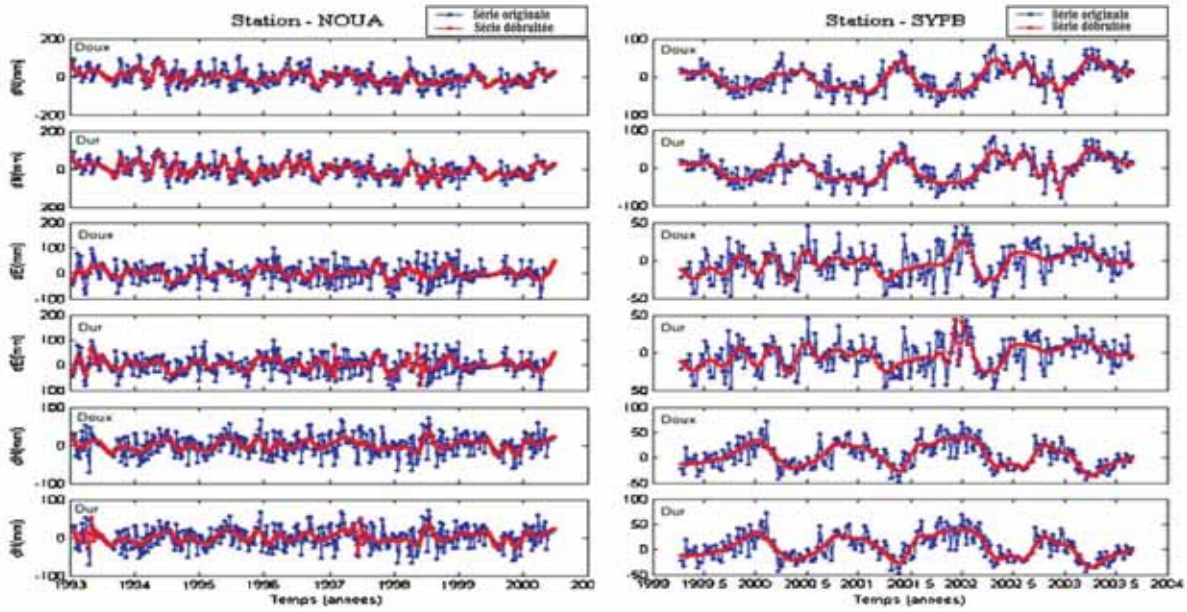

Fig. 7 Débruitage des composantes (dN, dE et dH) de la station NOUA par seuillage universel en utilisant les deux stratégies de seuillage (doux et dur).

Les tableaux 2 et 3 fournissent les statistiques sur les composantes (dN, dE et dH) des station NOUA et SYPB, avant et après débruitage. Ces tableaux montrent que la réduction du bruit est assez considérable après le débruitage par les deux méthodes de seuillage, expliquée par la diminution considérable de l'écart-type. Par exemple, l'écart-type de la série débruitée (dE) de la station NOUA est de 17.02 mm pour le seuillage doux et de 20.01 mm pour le seuillage dur, à comparer au 36.33 mm de la série originale (tableau 2).

En outre, on constate que les statistiques de débruitage des séries reconstruites de la station SYPB par les deux stratégies de seuillage (tableau 3) sont quasiment identiques; ce qui explique que la station SYPB est moins bruitée que la station NOUA. En effet, le tableau 4 qui fournit le rapport signal sur bruit RSB des composantes (dN, dE et dH), exprimé en décibels (dB) par l'équation 7, montre que le RSB de la série (dH) de la station NOUA est de 1.07 dB, à comparer au 4.17 dB de la même série (dH) de la station SYPB.

$$RSB = 10\log_{10}\left(\frac{\sum_{t=1}^{N}|X_t|^2}{\sum_{t=1}^{N}|X_t - \hat{X}_t|^2}\right) \quad (7)$$

Où X_t est le signal original et $\hat{X}_t$ le signal débruité.

Tableau 2. Statistiques sur les composantes (dN, dE et dH) de la station NOUA, avant et après débruitage (mm).

Série		Minimum	Maximum	Moyenne	Écart-type
dN	Série originale	-106	112.70	-3.95	41.91
	Série débruitée par seuillage doux	-60.80	81.40	-3.92	26.06
	Série débruitée par seuillage dur	-78.97	85.31	-3.92	29.80
dE	Série originale	-98.20	100.80	0.45	36.33
	Série débruitée par seuillage doux	-46.50	48.43	0.48	17.02
	Série débruitée par seuillage dur	-81.87	73.25	0.55	20.01
dH	Série originale	-71.90	71	0.84	26.18
	Série débruitée par seuillage doux	-26.23	35.72	0.84	12.02
	Série débruitée par seuillage dur	-40.76	52.40	0.84	13.26

Tableau 3. Statistiques sur les composantes (dN, dE et dH) de la station SYPB, avant et après débruitage (mm).

Série		Minimum	Maximum	Moyenne	Écart-type
dN	Série originale	-79.40	82	-3.99	33.64
	Série débruitée par seuillage doux	-41.85	51.88	-4.04	25.89
	Série débruitée par seuillage dur	-64.53	51.78	-4.04	26.58
dE	Série originale	-47.20	44.90	-4.00	20.37
	Série débruitée par seuillage doux	-27.10	26.22	-4.03	12.02
	Série débruitée par seuillage dur	-27.10	44.36	-4.03	12.59
DH	Série originale	-48.20	72.80	3.94	25.60
	Série débruitée par seuillage doux	-36.60	37.89	3.91	19.98
	Série débruitée par seuillage dur	-36.60	37.89	3.91	19.98

Tableau 4. Rapport signal sur bruit (RSB) des composantes (dN, dE et dH) (dB).

Série	Station	Méthode de débruitage	Stratégie de seuillage	RSB
(dN)	NOUA	seuillage universel	Doux	2.36
(dE)	NOUA	seuillage universel	Doux	1.24
(dH)	NOUA	seuillage universel	Doux	1.07
(dN)	SYPB	seuillage dépendant du niveau	Doux	4.12
(dE)	SYPB	seuillage universel	Doux	2.06
(dH)	SYPB	seuillage dépendant du niveau	Doux	4.17

5. Conclusion

Le choix de la méthode de débruitage adéquate exige au préalable la connaissance du type de bruit affectant le signal. Dans le cadre de cet article, c'est la variance d'Allan qui a été employée pour déterminer le type de bruit.

Dans le cas d'un bruit blanc, c'est la méthode VisuShrink qui a été utilisée et dans le cas d'un bruit coloré nous avons choisi la méthode de seuillage dépendant du niveau.

En revanche, la variance d'Allan exige que la série soit régulière et stationnaire. A cet effet, nous avons adopté l'interpolation linéaire pour régulariser les séries (combler les trous) et la fonction d'auto-corrélation (ACF) pour étudier la stationnarité des séries. Si la série est non stationnaire, il faut la rendre stationnaire par différentiation.

L'application de la méthodologie adoptée aux séries temporelles des coordonnées des stations NOUA et SYPB a permis de ressortir les conclusions suivantes :

- Le bruit déterminé pour les deux composantes (dE) et (dH) de la station NOUA est un bruit blanc et pour la composante (dN) est une combinaison d'un bruit blanc dominant et d'un bruit de marche aléatoire faible. Par contre, pour la station SYPB, le bruit déterminé pour les composantes (dN) et (dH) est un bruit de scintillation et pour la composante (dE) est un bruit blanc.

- Le rapport signal sur bruit (RSB) des trois composantes (dN, dE et dH) de la station NOUA est faible, ce qui explique que ces trois composantes sont trop perturbées. En revanche, les composantes (dN, dE et dH) de la station SYPB sont moins bruitées.

- La méthode optimale de débruitage reposant sur le critère de minimisation de l'écart-type est celle du seuillage doux.

Références

D. Donoho and I. Johnstone, 1994, "Ideal spatial adaptation via wavelet shrinkage", Biometrika, 81: 425-455.

D.C. Agnew, 1992, "The time-domain behaviour of power-law noises", in Journal of Geophysical Research, vol. 19, n°4, pp.333-336.

D.W. Allan, 1966, "Statistics of Atomic Frequency Standard", Proceedings of the IEEE, 54, N°2, 221-231.

D.W. Allan, 1987, "Time and frequency characterisation, estimation, and prediction of precision clocks and oscillators", IEEE Trans UFFC, vol 34, n.6.

I.Johnstone and B.Silverman,1997, "Wavelet threshold estimators for data with correlated noise", J.Roy. Statist.

J. Azoubib, 1974, "Caractérisation des fluctuations aléatoires d'un processus temporel par la variance d'Allan", Rapport BIPM-74/7, exposé présenté au BIPM le 26 novembre.

J. Rutman, 1978, "Characterization of phase and frequency instabilities in precision frequency sources : Fifteen years of progress", Proceedings of the IEEE 66, 1048–1075.

K. Lebail, 2004, "Étude statistique de la stabilité des stations de géodésie spatiale - Application à DORIS-", IGN, LAREG, thèse de doctorat Soutenue à l'Observatoire de Paris.

L. David Donoho, 1995, "De-noising by soft-thresholding", IEEE Transactions on Information Theory, vol. 41(3) : pp. 613–627.

S. Khelifa, 2007, "Analyse de séries temporelles de coordonnées de stations de géodésie spatiale", Mémoire de Magister en Techniques Spatiales et Applications (Option : Géodésie), Arzew.

S. Mallat, 1989, "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation", Proc. IEEE Trans on Pattern Anal. and Math. intel., Vol. 11, No 7.