

B.12 Évaluation de la qualité des séries temporelles de mesures : signal et bruit

Karine Le Bail

Introduction

En géodésie, on collecte toutes sortes de mesures. Par exemple, des mesures d'angles (théodolite) et de distances (distancemètre) en géodésie terrestre ou bien des mesures satellitaires en géodésie spatiale, mesures de retard du signal pour les satellites GPS, de distance station-satellite pour les satellites LASER tels que Lageos ou Starlette, de l'effet DOPPLER pour le système spatial DORIS (Topex/Poséidon, Jason). Ces mesures sont collectées, prétraitées et affectées d'une date de mesure. On en forme ainsi des séries temporelles.

L'étude de telles séries de mesures se développe depuis quelques années, par exemple : [MAO et al 99]. On en trouve notamment dans le domaine des systèmes de référence (séries de coordonnées de stations, de paramètres de transformation et d'orientation de la Terre). De telles séries portent des informations en géophysique (étude du mouvement local de stations) mais elles sont dépendantes des différentes techniques de calcul (choix des modèles utilisés pour l'estimation).

Il est important d'avoir une vue d'ensemble de chaque série étudiée afin de recueillir le maximum d'informations exploitables. Nous faisons arbitrairement la distinction entre les deux domaines d'étude suivants : le domaine "bruit" qui permet de qualifier le type du processus et le domaine "signal" qui permet de déterminer et de quantifier les signaux systématiques tels que périodicité et tendance.

Cet exposé vise à comparer différentes méthodes pour l'étude des séries temporelles et de dégager les plus appropriées à chaque cas. Nous les appliquerons à trois exemples très différents. Par ailleurs, dans les méthodes de caractérisation du processus, nous verrons la fonction d'auto-corrélation et la densité spectrale de puissance, tirée de la théorie statistique principalement utilisée en économétrie, puis la variance d'Allan largement utilisée dans la métrologie du temps et des fréquences. Ensuite, pour la détermination du signal, nous exploiterons les séries avec le spectre de Fourier et la Singular Spectral Analysis (SSA) développée dans le cadre d'études météorologiques.

Les données

Nous utilisons trois séries temporelles de résidus (provenant de l'analyse par moindres carrés) de la coordonnée verticale de stations DORIS.

Les deux premières sont des séries mises en référence avec le logiciel CATREF (Combination and Analysis of Terrestrial Reference Frames) à partir de fichiers de coordonnées au format *Sinex* du JPL-IGN (disponibles sur le serveur : <ftp://www.cddisa.nasa.gov>). *Sinex* est un format d'échange de la géodésie spatiale qui a la propriété de porter non seulement les résultats échangés mais aussi leur matrice de variance-covariance.

Le repère de référence est l'ITRF2000. On a les résidus hebdomadaires des stations :

- Nouméa entre 1993 et 2001 (NOUA) ;
- Syowa entre 1999 et 2004 (SYPB).

La troisième série est donnée sur le site de l'International DORIS Service - IDS : <http://www.ids.cls.fr>.

Ce sont les résidus mensuels par rapport à un référencement propre au réseau DORIS considéré : on fixe la station Toulouse et les paramètres d'orientation de la Terre (EOP) et on estime ainsi les résidus des coordonnées des autres stations du réseau. La série choisie est celle de :

- Yellowknife entre 1993 et 2002 (YELA).

Ces trois séries ont été choisies parce qu'elles ont des caractéristiques différentes visibles sur leurs courbes respectives (figure B.12.1, page suivante) :

- NOUA : son niveau de bruit est élevé (5 cm) mais elle ne semble pas avoir de signal périodique ou de pente ;
- SYPB : un signal périodique se détache clairement ;
- YELA : le niveau de bruit est faible mais on distingue nettement une pente.

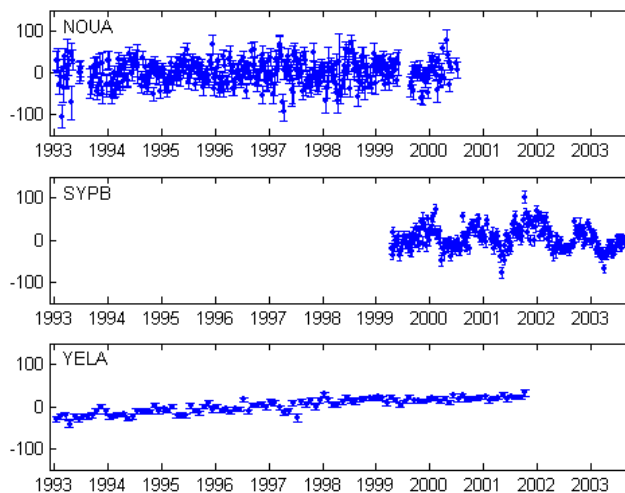


figure B.12.1 : résidus (en millimètres) sur la coordonnée verticale des stations NOUA, SYPB et YELA

Regard panoramique sur une série temporelle : le logiciel Crono_Vue

Le logiciel Crono_Vue est un outil qui permet de porter un regard panoramique sur une série chronologique. Il rassemble plusieurs méthodes mathématiques pour l'estimation d'une pente, l'extraction de composantes (tendance et termes cycliques) par la méthode Census X-11 (méthode de filtrage numérique développée pour les analyses économiques), en donne les périodogrammes, les variances d'Allan et plusieurs indicateurs statistiques usuels (écart-type, χ^2).

Les résultats se présentent sous forme d'une planche de plusieurs graphiques. Le tiers supérieur de la feuille concerne le signal lui-même, celui du milieu les termes périodiques et celui du bas le terme irrégulier. En décomposant la planche sur la largeur, la bande gauche concerne le contenu spectral, la bande centrale donne la série chronologique et ses composantes et la bande de droite l'analyse par la variance d'Allan.

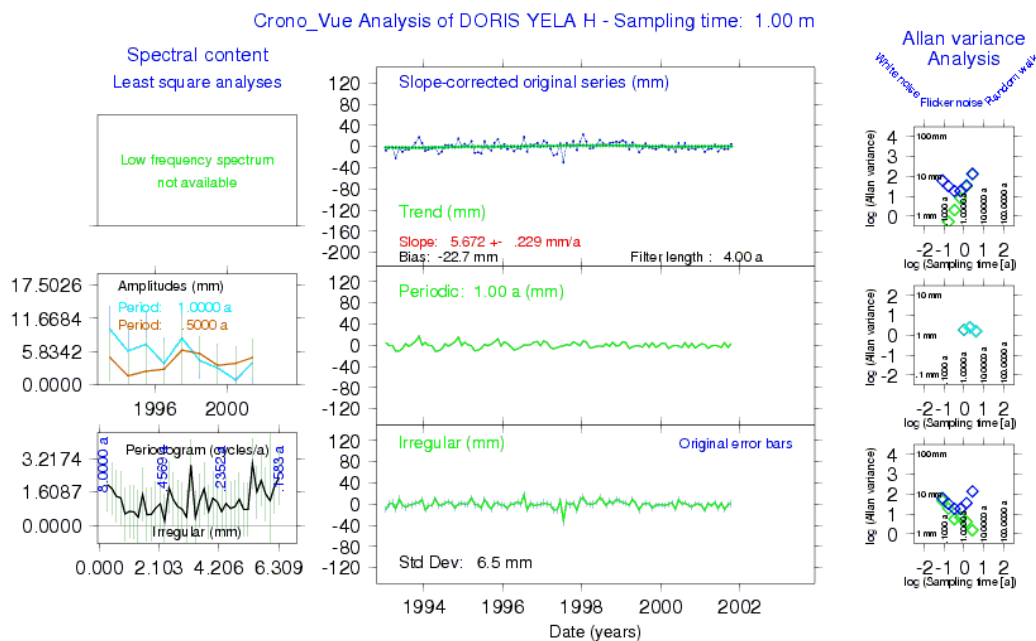


figure B.12.2 : exemple de planche graphique obtenue avec Crono_Vue : la série YELA

On applique Crono_Vue à nos trois séries en lui demandant d'investiguer un terme périodique à un an. On en tire plusieurs renseignements :

- NOUA : la dérive linéaire est nulle et l'analyse par la variance d'Allan indique un bruit blanc pour la série et sa composante irrégulière. Par contre, on ne perçoit pas de périodicité à un an ;
- SYPB : la variance d'Allan indique un bruit de scintillation. Mais la série est trop courte pour la recherche d'une périodicité d'un an. En effet, il faudrait que la durée de la série soit cinq fois plus grande que la période. Aussi, rien ne ressort pour cette série dans ce cadre ;

- YELA (figure B.12.2) : la pente a un coefficient élevé (5,7 mm/an) et la série est une marche aléatoire. Un léger terme périodique à un an se distingue sur le graphique du milieu.

Le logiciel Crono_Vue a été développé au laboratoire LAREG par M. Feissel-Vernier, voir : [FEISSEL-VERNIER 02]. Il peut être téléchargé à l'adresse : <http://lareg.ensg.ign.fr/IDS/software.html>.

Toute série chronologique peut être considérée comme un processus stochastique (processus aléatoire indexé par le temps), $x(t)$, où $t \in I$, I intervalle borné et fini.

Dans la suite, on note $x(t)$ les points de nos séries au temps t . Les notations choisies sont celles de l'ouvrage de C. Gouriéroux et A. Montfort : [GOURIÉROUX, MONFORT 95].

Étude du bruit d'une série chronologique

Un premier volet de l'étude de $(x(t))_{t \in I}$ concerne son bruit. Sa connaissance permet de qualifier la stabilité de la série.

Stationnarité d'un processus : fonction d'auto-corrélation

Un processus est dit stationnaire si son espérance ($E(x(t)) = \frac{1}{\text{card}(I)} \sum_{t \in I} x(t)$) ne dépend pas du temps et que, pour un τ donné, $\text{cov}(x(t), x(t+\tau)) = E(x(t))E(x(t+\tau)) - E(x(t), x(t+\tau))$ ne dépend que de τ . ne dépend que de τ .

Soit R la fonction d'auto-covariance de x définie par :

$$R(\tau, t) = E(x(t)x(t+\tau)).$$

On appelle τ le retard.

La fonction d'auto-corrélation est alors :

$$\Gamma(\tau) = \frac{R(\tau)}{R(0)}.$$

Si elle décroît très lentement, cela signifie que le processus considéré est non stationnaire.

En traçant cette fonction pour nos trois exemples (figure B.12.3, page suivante), on observe que pour le premier (NOUA), il n'y a pas d'ambiguïté, le processus est effectivement stationnaire, alors que ce n'est pas le cas pour le troisième exemple (YELA), la fonction d'auto-corrélation indique que le processus est non stationnaire.

Détermination du type de bruit

On peut déterminer le type de bruit d'une série chronologique si le processus est stationnaire, comme je l'ai déjà explicité. Si c'est le cas, il existe plusieurs méthodes pour caractériser le type de bruit. Nous en exposons deux ici. La méthode par décomposition ARMA (Auto Régressif à Moyenne Mobile) ne sera pas développée puisque cela a été préalablement fait dans l'article de P. Sillard [SILLARD et al 01].

Densité spectrale

La densité spectrale de puissance (DSP) est la transformée de Fourier de l'auto-covariance :

$$F(\omega) = \int_{-\infty}^{+\infty} R(\tau) e^{-j\omega\tau} d\tau.$$

Le processus est caractérisé par le comportement de F en fonction de la fréquence ω . Le processus est :

- un bruit blanc si $F(\omega) \sim 1$;
- un bruit de scintillation si $F(\omega) \sim \frac{1}{\omega}$;
- une marche aléatoire si $F(\omega) \sim \frac{1}{\omega^2}$.

La DSP des trois séries étudiées est montrée sur la figure B.12.4, page suivante. Pour la série NOUA, la DSP indique que le processus est un bruit blanc (puisque le processus est stationnaire). Pour SYPB, elle indique que le processus est un bruit de scintillation. On ne s'engagera pas pour YELA puisque le processus n'est pas nettement stationnaire.

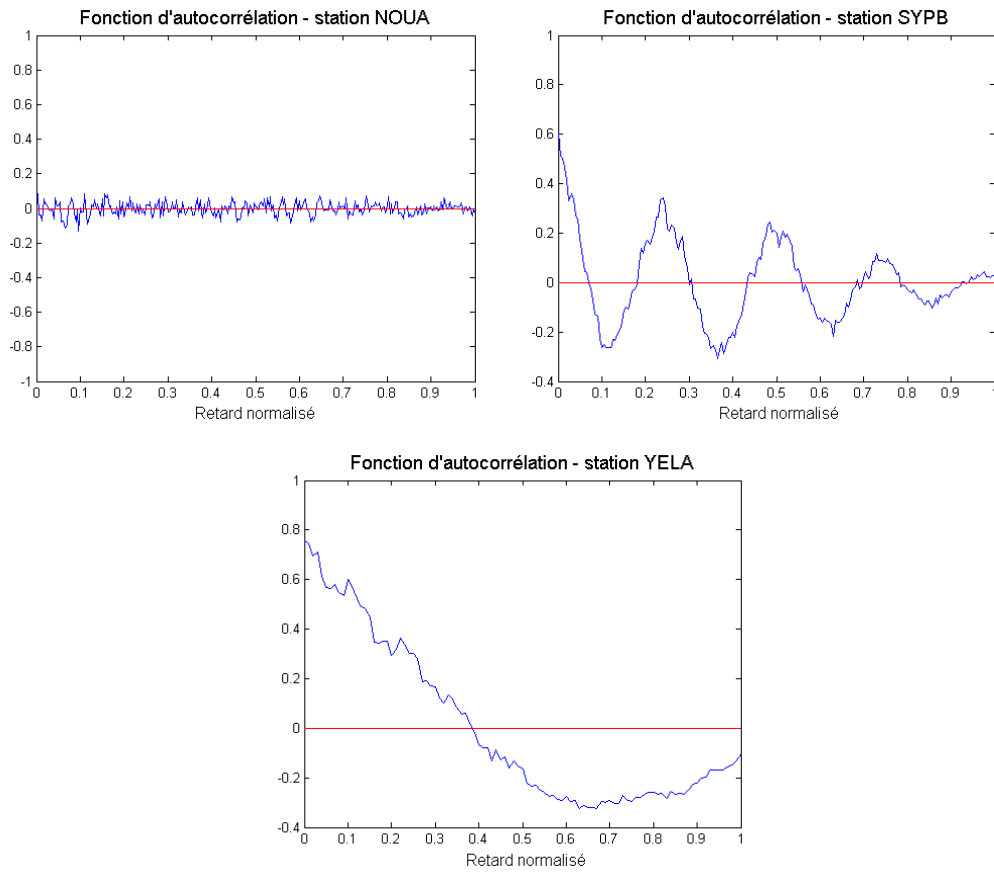


figure B.12.3 : fonctions d'auto-corrélation pour les séries NOUA, SYPB et YELA

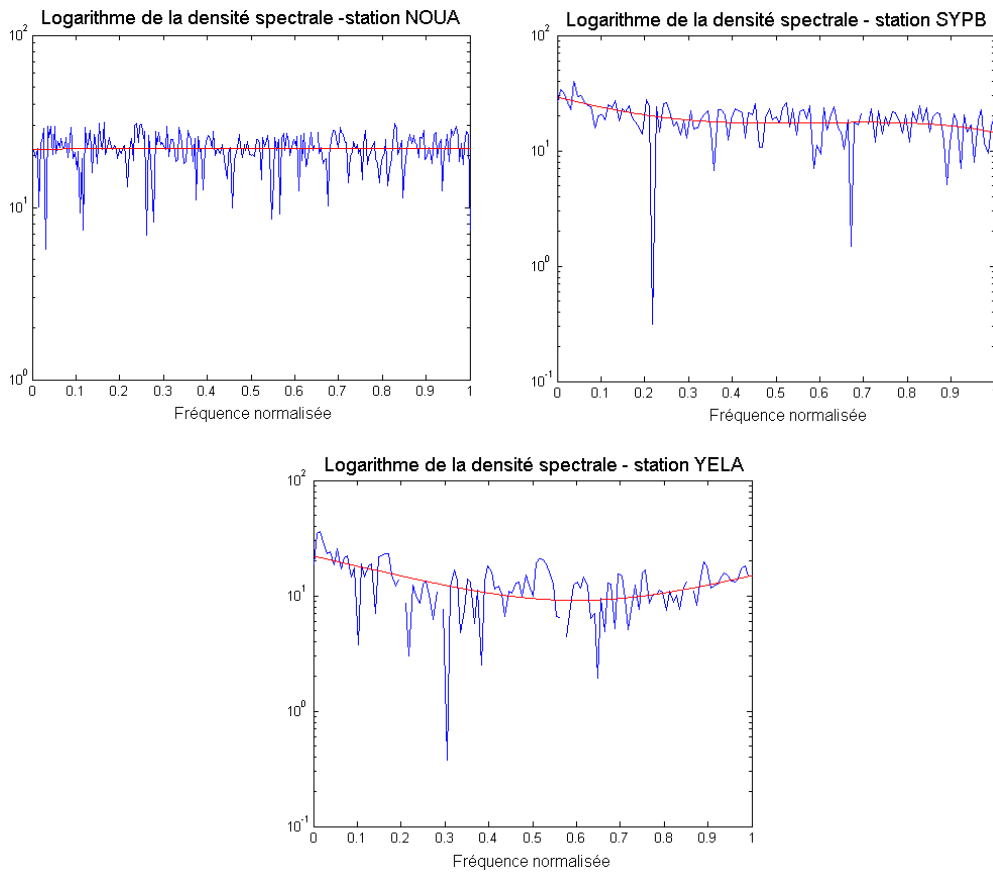


figure B.12.4 : densité spectrale de puissance des séries NOUA, SYPB et YELA

Variance d'Allan

La variance d'Allan est un outil très utilisé dans le domaine du temps et des fréquences, notamment dans l'étude de la stabilité des horloges atomiques au Bureau International des Poids et Mesures [RUTMAN 78]. On l'applique ici à nos séries temporelles.

La formule de la variance d'Allan est ([ALLAN 66]) :

$$\sigma_x^2(\tau) = \frac{1}{2} \langle (\bar{x}_{k+1} - \bar{x}_k)^2 \rangle$$

où τ est le temps d'échantillonnage, et $\bar{x}_k = \frac{1}{\tau} \int_{t_k}^{t_{k+\tau}} x(t) dt$ et $\langle . \rangle$ représente la moyenne.

Si le processus est stationnaire, l'évolution de la variance d'Allan en fonction du temps d'échantillonnage τ est reliée à la densité spectrale du processus ($S_x(\omega) = h_\alpha \omega^\alpha$) par la relation suivante :

$$\sigma_x^2(\tau) = \int_0^{+\infty} S_x(\omega_i) \frac{2 \sin^4(\pi \tau \omega_i)}{(\pi \tau \omega_i)^2} d\omega$$

Ainsi, lorsqu'on représente la variance d'Allan dans un graphe tel que : $\log \sigma_x^2 = \mu \log \tau$, le spectre de bruit du processus peut être déterminé d'après la pente du graphe, comme suit :

- si $\mu = -1$, bruit blanc (densité spectrale indépendante de la fréquence ω) ;
- si $\mu = 0$, bruit de scintillation (densité spectrale variant en $\frac{1}{\omega}$) ;
- si $\mu = 1$, marche aléatoire (densité spectrale variant en $\frac{1}{\omega^2}$).

Cependant, la variance d'Allan permet aussi de déterminer un signal périodique puisque la théorie dit qu'une perturbation périodique se traduit sur la variance d'Allan par un biais. Sur la courbe log/log, cela se traduit par des rebonds, dont l'espacement dépend de la période du signal périodique [VERMOTTE 91].

De plus, s'il existe une pente importante dans une série, la variance d'Allan en est très perturbée. Si on simule une série avec un bruit blanc puis on lui ajoute une pente importante, la variance d'Allan indique que le processus est une marche aléatoire. Ajouter une pente importante à la série revient à corrélérer ses points, ce qui fait dégénérer le bruit.

Les résultats (figure B.12.5) nous donnent que :

- pour NOUA, la caractérisation du processus est en accord avec la DSP puisqu'on observe un bruit blanc ;
- pour SYPB, on se retrouve dans le même cas : bruit de scintillation confirmée. De plus, on observe un rebond au point qui correspond à un temps d'échantillonnage d'un an et qui indique une périodicité d'un an dans la série ;
- pour YELA, le processus semble être un bruit blanc jusqu'à un taux d'échantillonnage d'un an et une marche aléatoire par la suite mais le processus n'est pas nettement stationnaire. Ceci est expliqué par la présence de la pente avec un fort coefficient par rapport aux résidus (plus de 5 mm/an pour des résidus de quelques millimètres) : la série est auto-corrélée.

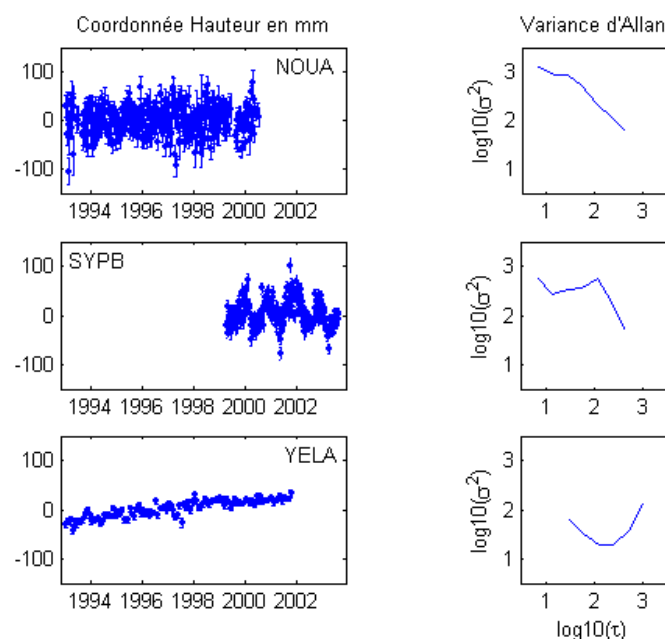


figure B.12.5 : les séries NOUA, SYPB et YELA et leurs variances d'Allan

Recherche d'un signal périodique ou pseudo-périodique dans une série chronologique

Lorsque l'on regarde attentivement une série chronologique, la recherche de signatures récurrentes dans le signal est intéressante. En effet, on a vu précédemment qu'une simple pente fait dégénérer le type de bruit de la série, aussi, il faut pouvoir les détecter. De plus, dans notre cas, chercher la tendance et les termes périodiques donne des indications sur le mouvement non modélisé de la station et les erreurs dues à la stratégie de calcul.

Spectre de Fourier

La technique la plus utilisée pour cette étude est le calcul des coefficients de la transformée de Fourier de la série considérée. Ces coefficients sont donnés par la formule suivante :

$$X(k) = \sum_{j \in I} x(j) e^{-\frac{2\pi i}{\text{card}(I)}(j-1)(k-1)}$$

où $x(j)$ sont les termes de la série et I l'intervalle sur lequel s'étend la série.

La transformée de Fourier est généralement représentée sur un graphique avec la période en abscisse (ici, des semaines ou des mois). Les pics représentent les périodes présentes dans la série et leur amplitude. Appliquée à nos trois séries, cette méthode donne les résultats de la figure B.12.6 (ci-dessous).

On voit apparaître systématiquement un pic à un an pour les trois séries, beaucoup plus important pour la série SYPB, et un pic à huit semaines (deux mois) pour la série NOUA. On ne connaît pas le sens de la périodicité à un an dans notre cas. Il semblerait que cela est un systématisme dans le calcul dû à l'estimation des paramètres d'orientation de la terre (EOP).

Singular Spectral Analysis (SSA)

Cette méthode a été développée pour l'étude temporelle de champs de paramètres météorologiques couvrant la planète à une douzaine de couches en altitude. Elle est basée sur le calcul des vecteurs propres de la matrice d'auto-covariance, voir [GHIL et al 02]. On décompose son procédé en trois étapes.

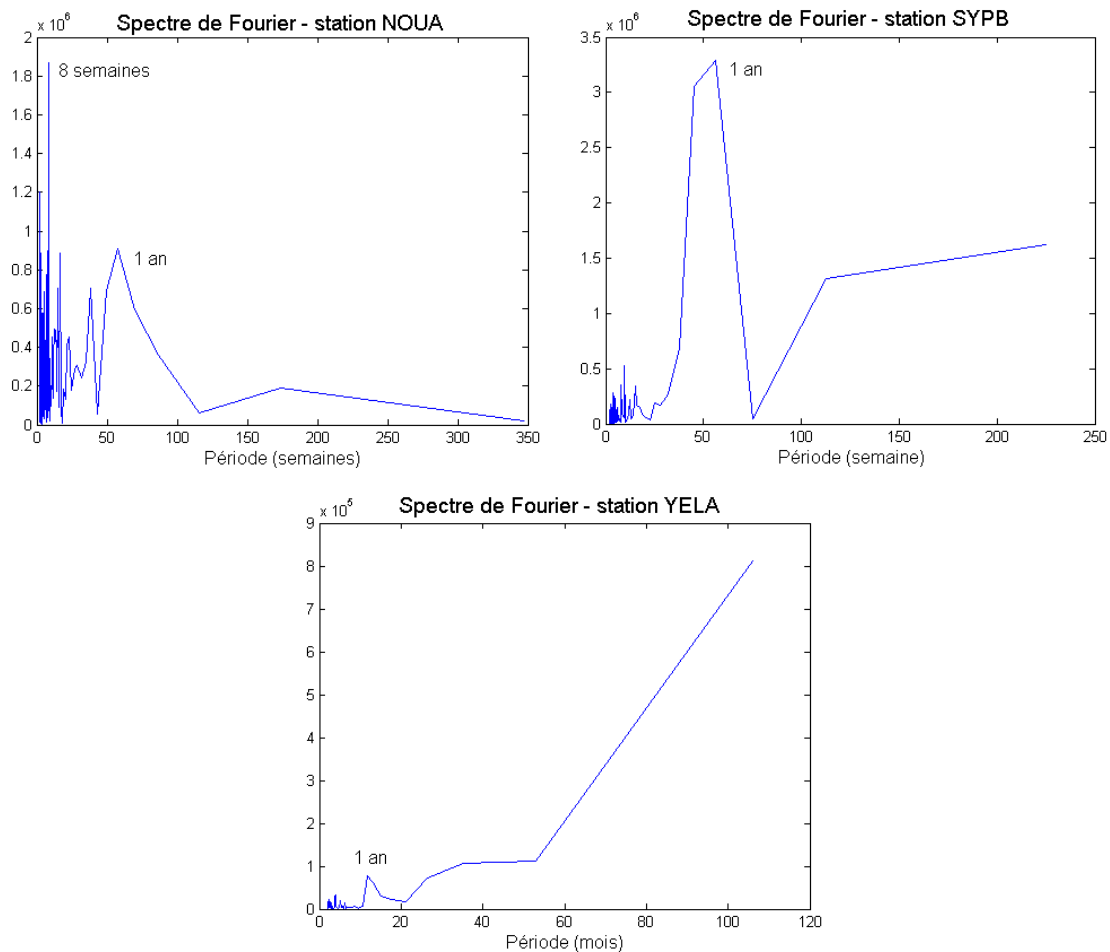


figure B.12.6 : spectres de Fourier des séries NOUA, SYPB et YELA

Choix d'un retard de covariance M

C'est l'étape la plus délicate puisque cela revient en fait à tester une certaine périodicité et donc à bien connaître les effets liés à la série étudiée. Une fois M choisie, on découpe le signal en morceaux de longueur M de sorte que :

$$D = \begin{pmatrix} x(1) & x(2) & \cdots & x(M) \\ x(2) & x(3) & \cdots & x(M+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-M+1) & x(N-M+2) & \cdots & x(N) \end{pmatrix}$$

La matrice d'auto-covariance C est de taille (M, M) et s'écrit : $C = D^t D$.

Étude des valeurs propres de la matrice C

On détermine les M valeurs propres de la matrice C . En les représentant dans un diagramme à barres en ordre décroissant, on identifie immédiatement les valeurs propres isolées et les valeurs propres proches.

La théorie de la SSA donne :

- valeur propre isolée = tendance ;
- deux valeurs propres proches = signal périodique.

Les valeurs propres petites constituent le bruit du signal.

Projection de la série originale sur les vecteurs propres et reconstruction des composantes.

Appliquée à nos séries et en donnant comme retard de covariance les périodes obtenues précédemment, on obtient les résultats suivants.

NOUA

Pour un retard de covariance de un an ($M = 52$ semaines), deux valeurs propres proches se détachent des 50 restantes, ce qui indique une composante cyclique. Après reconstruction de cette composante, on superpose le signal avec la composante dans le graphe B.12.7, ci-dessous. Cela représente sensiblement l'allure des variations de la courbe mais pas de manière nette. Il faut souligner que les deux valeurs propres proches n'expliquent à peine que 10% du signal. Il n'en ressort pas clairement un signal périodique. Il en est de même en prenant un retard de covariance de deux mois ($M = 8$ semaines).

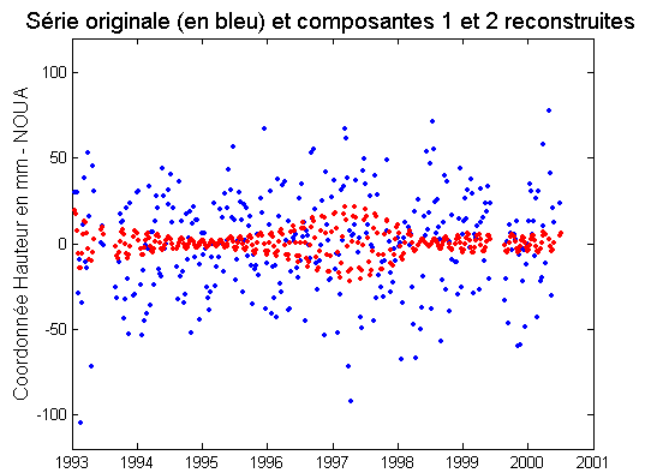
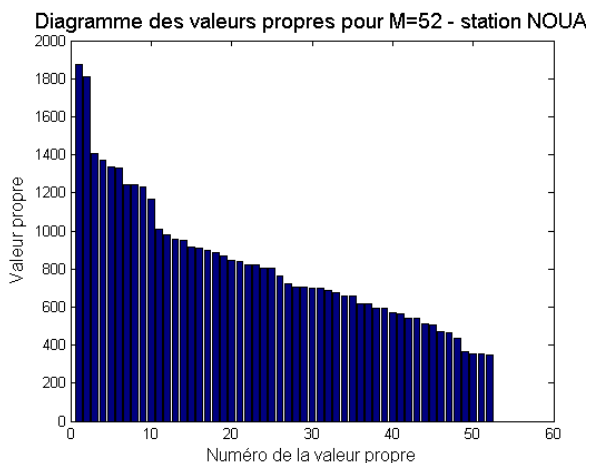


figure B.12.7 : série NOUA.

Diagramme en barres des 52 valeurs propres étudiées et superposition de la série avec sa composante périodique.

SYPB

En prenant $M = 52$ semaines, soit un an, il ressort bien du diagramme des valeurs propres une composante périodique (les deux premières valeurs propres) et une tendance (la troisième valeur propre). Cependant, il reste 50% du signal inexpliqué. Dans le graphique représentant le signal et ses deux principales composantes, on voit que le signal périodique suit effectivement les variations du signal mais que reste un bruit important. La tendance semble indiquer un point de la fin de l'année 2002.

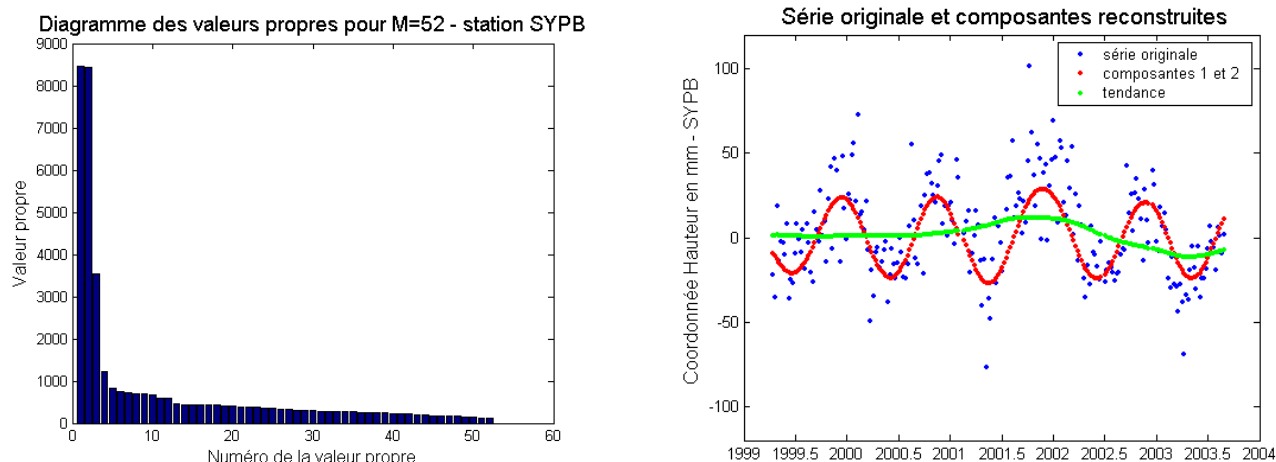


figure B.12.8 : série SYPB.

Diagramme en barres des cinquante-deux valeurs propres étudiées et superposition de la série avec la composante périodique et la tendance reconstruites

YELA

Puisque cette série est mensuelle, on découpe le signal en douze points (une année). Il en ressort une composante qui explique la moitié du signal. Nous voyons clairement cette tendance sur la courbe de la station et elle ressort avec Crono_Vue. Par contre, le spectre de Fourier la noie puisque ce n'est pas un signal strictement périodique.

Ces deux méthodes sont complémentaires puisque certains éléments ressortent mieux pour l'une que pour l'autre. Il existe aussi la méthode par décomposition en ondelettes du signal qui permet de faire ressortir les termes périodiques. Son avantage est qu'elle peut détecter un signal transitoire dans la série, contrairement aux méthodes du spectre de Fourier et de la SSA. Nous ne l'exposerons pas ici mais on en trouvera un développement dans l'exposé de A. Rivoldini & F. Renaud [RIVOLDINI 03].

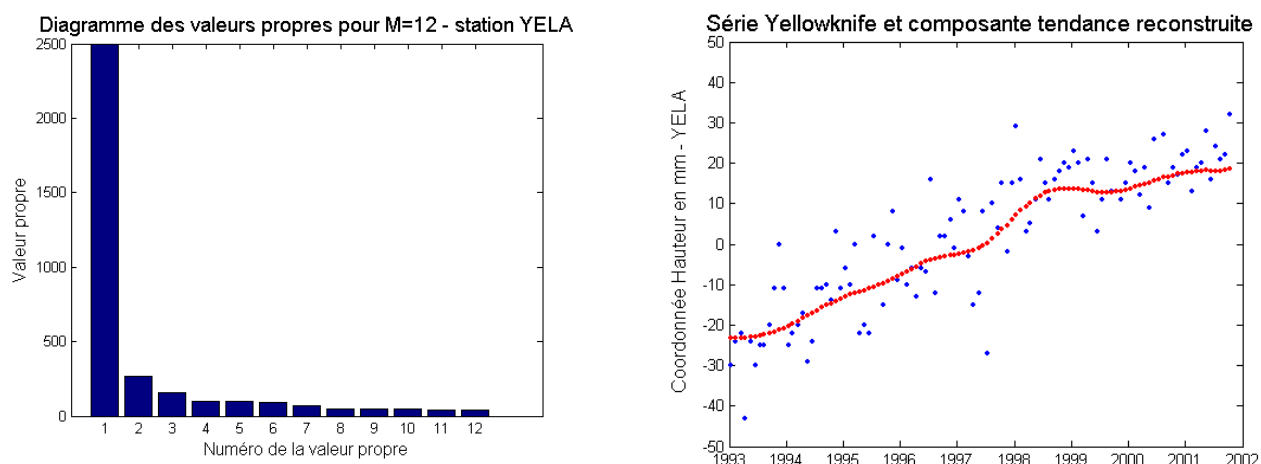


figure B.12.9 : série YELA.

Diagramme en barres des douze valeurs propres étudiées et superposition de la série avec sa tendance reconstruite.

Conclusion

Les méthodes explicitées dans cet exposé permettent de faire une étude complète d'une série chronologique en distinguant les deux domaines définis a priori comme "signal", modélisé analytiquement et "bruit", modélisé comme processus statistique, figure B.12.10.

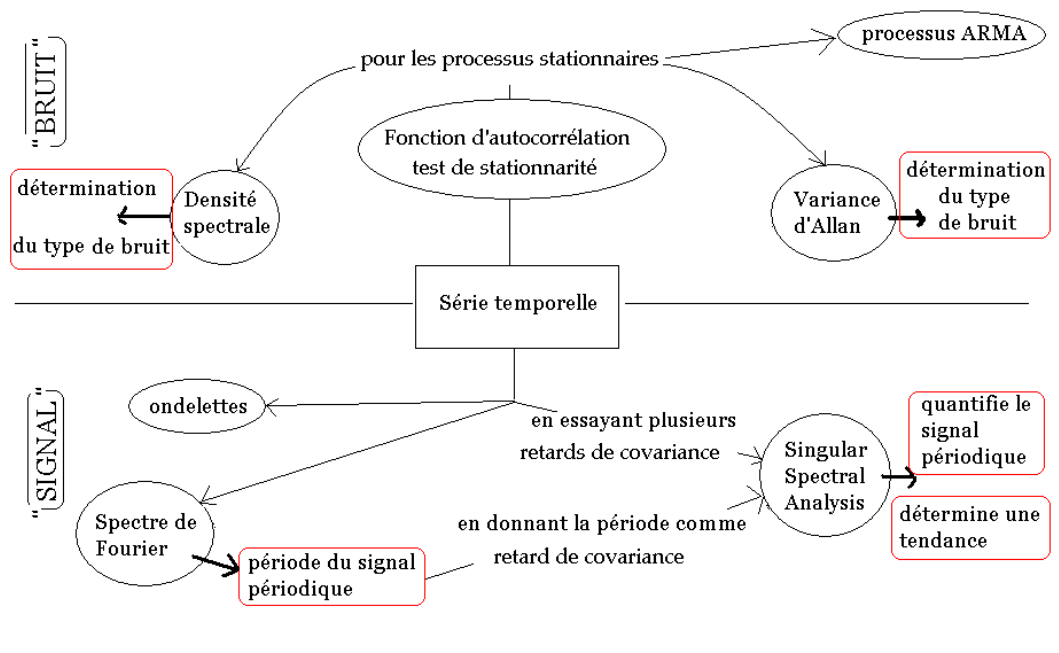


figure B.12.10 : schéma d'étude "signal / bruit" d'une série chronologique

Si on récapitule les résultats pour chaque série, on a, pour les stations suivantes :

NOUA

Le processus est clairement stationnaire et le bruit déterminé est le bruit blanc. Ce qui signifie que chaque point de la série est indépendant. Dans le domaine du signal, on observe deux petites composantes périodiques à un an et deux mois mais la signification doit être reliée à la stratégie de calcul.

SYPB

Le processus semble stationnaire et on obtient du bruit de scintillation. Cependant, on observe clairement un signal périodique à un an et la tendance semble prêter attention à un point, fin 2002. Il faudrait vérifier l'existence d'un phénomène soit géophysique, soit du à un changement d'antenne (mais la tendance semble reprendre sa place ensuite), soit à une mauvaise connaissance d'un des paramètres qui sert à l'estimation des résidus de coordonnées.

YELA

Le processus ne semble pas stationnaire et les indicateurs de bruit donnent une marche aléatoire. Il faudrait refaire l'étude pour Yellowknife en retirant une pente pour déterminer le type de bruit résultant. Concernant le signal, une petite composante périodique à un an est décelée par le spectre de Fourier (et la variance d'Allan) mais la composante la plus marquée est la tendance qui confirme ce qui a été dit précédemment.

En conclusion générale, chaque série temporelle a des caractéristiques différentes qui ne peuvent pas toutes être mises en évidence par une seule méthode. Certaines de ces méthodes sont plus sensibles aux pentes par exemple (variance d'Allan) ou à la stationnarité du processus, d'autres identifient les périodes ou bien déterminent les composantes du signal. D'où l'avantage de répertorier les méthodes existantes afin de tirer le maximum d'informations sur chaque série étudiée.

Références

[ALLAN 66] D.W. Allan : *Statistics of atomic frequency standards*, in Proceedings IEEE 54, 221, 1966.

[GHIL et al 02] M. Ghil, M.R. Allen, M.D. Dettinger, K. Ide, D. Kondrashov, M.E. Mann, A.W. Robertson, A. Saunders, Y. Tian, F. Varadi and P. Yiou : *Advanced spectral method for climatic time series*, in Reviews of Geophysics, 40(1), pp. 3.1-3.41, 10.1029/2000GR000092, 2002.

[GOURIÉROUX, MONFORT 95] C. Gouriéroux & A. Montfort : *Séries temporelles et modèles dynamiques*, 2^{ème} édition, in *Economica*, 1995.

[MAO et al 99] A. Mao, C.G.A. Harrison and T.H. Dixon : *Noise in GPS coordinate time series*, in *Journal of Geophysical Research*, vol. 104, n° B2, pp. 2797-2816, 1999.

[RIVOLDINI 03] A. Rivoldini & F. Renaud : *Le signal du marégraphe de Brest vu par les ondelettes*, voir sur le site : http://lareg.ensg.ign.fr/AGRET/coll_03/rivoldini.pdf, communication au GdR G2, 13 novembre 2003.

[RUTMAN 78] J. Rutman : *Characterization of Phase and Frequency Instabilities in Precision Frequency Sources : Fifteen Years of progress*, in *Proceedings IEEE* 66, 1048, 1978.

[VERNOTTE 91] F. Vernotte : *Stabilité temporelle des oscillateurs : nouvelles variances, leurs propriétés, leurs applications*, thèse de doctorat de l'Université de Besançon, 1991.

Publications

[FEISSEL-VERNIER 02] M. Feissel-Vernier : *Crono_Vue : regards croisés sur les séries chronologiques*, <http://lareg.ensg.ign.fr/IDS/software.html>, laboratoire LAREG, IGN-ENSG, Marne-la-Vallée, 2002.

[SILLARD et al 01] P. Sillard, S. Botton & D. Bouteloup : *Un exemple de traitement de séries temporelles de positions d'instrument géodésique*, in *Bulletin d'information de l'IGN n°72*, pp. 1-14, 2001, laboratoire LAREG, IGN-ENSG, Marne-la-Vallée, 2001.