

# Extraction en ondelettes des fluctuations turbulentes cohérentes : application au plasma de bord du tokamak Tore-Supra

## Wavelet-based extraction of turbulent coherent fluctuations : application to the Tore-Supra tokamak edge plasma

Romain Nguyen van yen, Marie Farge

*Laboratoire de Météorologie Dynamique-CNRS, École Normale Supérieure, Paris, France*

Kai Schneider

*Laboratoire de Mécanique, Modélisation et Procédés Propres-CNRS et Centre de Mathématiques et d'Informatique, Université d'Aix-Marseille, Marseille, France*

---

### Abstract

Using an orthogonal wavelet basis turbulent fluctuations are decomposed into two components (coherent and incoherent) which can thus be studied independently. This method, called CFE (Coherent Fluctuations Extraction), is applied to density signals measured in the edge plasma of Tore-Supra tokamak (CEA-Euratom, Cadarache). Coherent fluctuations retain most of the variance and are non-Gaussian and intermittent. In contrast, incoherent fluctuations are non-intermittent and Gaussian. From these results one conjectures that the former drive turbulent transport, while the latter have only a diffusive role.

**Keywords :** plasma, wavelets, turbulence

### Résumé

Grâce aux ondelettes orthogonales on décompose les fluctuations turbulentes en deux composantes (cohérente et incohérente) qui peuvent ainsi être étudiées indépendamment. Cette méthode, appelée EFC (Extraction des Fluctuations Cohérentes), est appliquée aux signaux de densité mesurés dans le plasma de bord du tokamak Tore-Supra (CEA-Euratom, Cadarache). Les fluctuations cohérentes retiennent l'essentiel de la variance, sont intermittentes et non-Gaussiennes. Par contre les fluctuations incohérentes résiduelles sont Gaussiennes et non intermittentes. Ces résultats conduisent à conjecturer que les premières sont responsables du transport turbulent, alors que les secondes n'ont qu'un rôle diffusif.

**Mots clés :** plasma, wavelets, turbulence

---

### Introduction

Dans les écoulements turbulents pleinement développés, aussi bien pour les fluides que pour les plasmas, on voit apparaître des structures cohérentes qui se forment au sein des fluctuations turbulentes sous l'effet des interactions non linéaires. Nous les appellerons donc 'fluctuations cohérentes'. Nous proposons une méthode pour les extraire car nous partons de l'hypothèse que ces structures, une fois formées, jouent un rôle dynamique majeur car ce sont leurs interactions non linéaires qui gouvernent l'évolution de l'écoulement et ce sont elles qui sont responsables du transport turbulent. Comme nous n'avons pas à ce jour une description précise en terme de structure de ces fluctuations turbulentes nous ne pouvons pas utiliser des formes-type pour les extraire. Nous avons alors décidé d'employer une approche apophétique : plutôt que de définir ce qu'elles sont nous définirons ce qu'elles ne sont pas. Nous partons de l'hypothèse que les structures cohérentes qui se forment au sein des écoulements turbulents pleinement développés ne sont pas du bruit mais des 'formes' organisées. Pour les extraire il suffira alors d'éliminer le bruit en faisant une hypothèse, non plus sur la forme des structures mais sur

la nature du bruit à enlever. Dans ces conditions le problème de l'extraction des fluctuations cohérentes devient un problème de débruitage. Pour le résoudre nous utiliserons la représentation en ondelettes orthogonales qui présente des propriétés optimales, que nous expliquerons, en terme de débruitage quand le bruit est additif et Gaussien. Nous présenterons deux types d'algorithmes de débruitage ou nous supposerons soit que le bruit est blanc, soit que nous ne connaissons pas a priori sa corrélation. À titre d'exemple nous appliquerons ces deux techniques pour extraire les fluctuations cohérentes de plusieurs signaux de densité mesurés dans le plasma de bord du tokamak Tore-Supra (CEA-Euratom, Cadarache).

## 1. Transformée en ondelettes

Nous introduisons dans cette section plusieurs types de transformées en ondelettes dont nous aurons besoin par la suite. Une introduction plus générale aux applications des ondelettes pour l'étude de la turbulence peut-être trouvée dans la référence [7].

### 1.1. Transformée en ondelettes continue

On considère un signal dépendant d'un seul paramètre, par exemple le temps, modélisé comme un élément de l'espace de Hilbert  $L^2(\mathbb{R})$  muni du produit scalaire usuel. On cherche à analyser ce signal avec une famille de fonctions engendrés par dilatation et translation d'une fonction de référence, l'ondelette mère  $\psi$ . Étant donné un paramètre d'échelle  $a$  réel strictement positif et un paramètre de position réel  $b$ , on définit donc

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)$$

la transformée intégrale de  $f$  relativement à cette famille étant alors définie par

$$\tilde{f}(a,b) = \langle f | \psi_{a,b} \rangle \quad (1)$$

Mais pour pouvoir dire que l'on a extrait toute l'information contenue dans  $f$ , on souhaite également pouvoir reconstruire  $f$  à partir de sa transformée  $\tilde{f}$ . Un résultat important, découvert par Grossmann et Morlet en 1984, est que cela est possible si et seulement si  $\psi$  vérifie la condition d'admissibilité

$$C_\psi = \int_0^\infty |\hat{\psi}|^2 \frac{dk}{k} < \infty \quad (2)$$

Si  $\psi$  est intégrable ( $\psi \in L^1 \cap L^2(\mathbb{R})$ ), (2) signifie simplement que  $\psi$  doit être de moyenne nulle. Or pour une fonction intégrable, la seule façon d'être de moyenne nulle est d'osciller, c'est pourquoi on appelle  $\psi$  "ondelette" (figure 1a). La formule de reconstruction est alors

$$f(x) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^\infty \tilde{f}(a,b) \psi_{a,b}(x) \frac{dadb}{a^2}$$

$\tilde{f}$  s'appelle transformée en ondelettes continue de  $f$  par l'ondelette mère  $\psi$ . Chaque coefficient  $\tilde{f}(a,b)$  apporte des informations sur le comportement de  $f$  au voisinage du point  $b$  et à l'échelle  $a$ . Cependant, la nature précise de ces informations dépend de l'ondelette mère choisie, dont le choix ne doit donc en aucun cas être arbitraire. Les trois propriétés les plus importantes sont :

- la rapidité de la décroissance de  $\psi$ , qui assure que chaque coefficient  $\tilde{f}(a,b)$  apporte bien des informations sur le comportement de  $f$  au voisinage immédiat du point  $b$
- la rapidité de la décroissance de  $\hat{\psi}$ , qui assure que chaque coefficient  $\tilde{f}(a,b)$  apporte bien des informations sur le comportement de  $f$  au voisinage immédiat de l'échelle  $a$ . La décroissance à l'infini de  $\hat{\psi}$  est reliée par une propriété bien connue de la transformée de Fourier à la régularité de l'ondelette  $\psi$
- le nombre de moments nuls, c'est à dire le plus grand entier  $m$  tel que  $\int_{-\infty}^\infty x^{m-1} \psi(x) dx = 0$ .

La réalisation simultanée des deux premières propriétés est limitée par l'inégalité dite de Heisenberg, qui affirme que le produit des écarts types de  $\psi$  et  $\hat{\psi}$  ne peut pas être inférieur à  $\frac{1}{2}$ . La troisième propriété assure que plus  $f$  est régulière au voisinage du point  $b$ , plus  $\tilde{f}(a,b)$  tendra vite vers zéro.

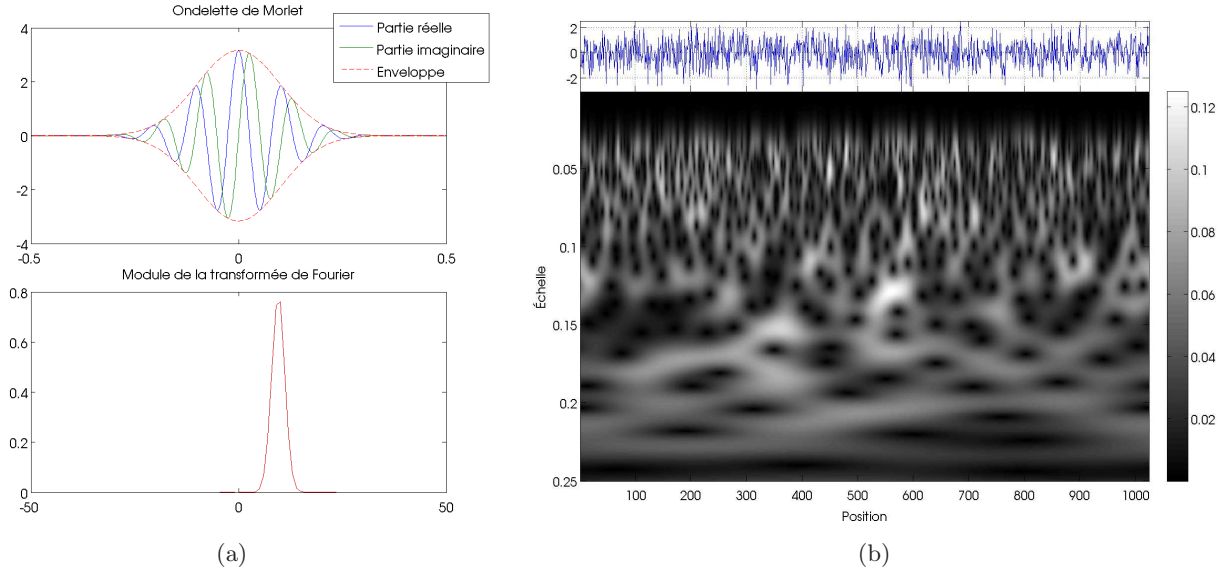


FIG. 1: L'ondelette de Morlet (a) est bien localisée dans l'espace physique et dans l'espace de Fourier. Cependant, les modules des coefficients d'ondelettes d'un bruit blanc Gaussien (b) ne sont pas indépendants, du fait de la redondance de la transformée en ondelettes continues (paragraphe 1.2).

Pour préciser les choses, nous nous placerons dans le cadre de la régularité Lipschitzienne. On dit qu'une fonction  $f$  est uniformément Lipschitzienne d'ordre  $r$  sur  $[x_1, x_2]$  si

$$\exists K > 0, \forall x \in [x_1, x_2], \exists P_x \in \mathbb{R}_n[x], \forall y \in \mathbb{R}, |f(y) - P_x(y)| \leq K|x - y|^r \quad (3)$$

où on a noté  $d$  la partie entière de  $r$ . On a alors le théorème suivant ([15] page 167) :

Si  $f$  est uniformément Lipschitzienne d'ordre  $r \leq m$  sur  $[x_1, x_2]$ , alors il existe  $A > 0$  tel que

$$\forall (a, b) \in [x_1, x_2] \times \mathbb{R}^+, |\tilde{f}(a, b)| \leq Aa^{r+\frac{1}{2}} \quad (4)$$

Intuitivement, cela signifie que les fonctions qui sont presque partout lisses auront presque partout des coefficients d'ondelettes proches de zéro aux échelles fines. Les grands coefficients d'ondelettes seront ainsi concentrés uniquement au voisinage des singularités. Nous verrons l'importance de ce fait lorsque nous aborderons les applications pratiques.

### 1.2. Transformée en ondelettes discrète

L'information fournie par  $\tilde{f}$  est redondante, car les coefficients d'ondelettes voisins ne sont pas indépendants. On le voit bien sur la transformée d'un bruit blanc Gaussien (figure 1). Cette redondance de la transformée en ondelettes continue est un handicap pour les applications à la compression de signaux, mais aussi pour tout type de manipulation qu'on souhaiterait faire subir aux coefficients. En effet, cela n'a pas de sens de modifier un des coefficients  $\tilde{f}(a, b)$  indépendamment de ses voisins en espace et en échelle. La portée de l'interaction est donnée par le noyau reproduisant de l'ondelette choisie [6].

Une première étape pour réduire la redondance est de ne considérer qu'une sous-famille discrète des  $\psi_{a,b}$ ,  $(a, b) \in \Lambda$ . On souhaite cependant que les coefficients restants caractérisent encore complètement la fonction  $f$ , il faut donc choisir soigneusement ceux que l'on garde. Éliminer complètement la redondance signifierait que les ondelettes choisies forment une base de Hilbert de  $L^2(\mathbb{R})$ , auquel cas on aurait pour toute  $f$

$$f = \sum_{\lambda \in \Lambda} \langle \psi_\lambda | f \rangle \psi_\lambda \quad (5)$$

Intuitivement, on sent qu'il est difficile que deux fonctions  $\psi_{a,b}$  et  $\psi_{a',b'}$  pour  $(a', b')$  et  $(a, b)$  voisins soient deux à deux orthogonales, car leurs supports se recouvrent, de même que les supports de leurs transformées de Fourier. Une solution connue depuis 1901 est l'ondelette de Haar, échantillonnée sur la grille dyadique

$$(a, b) \in \Lambda = \{(2^{-j}, 2^{-j}k) \mid j \in \mathbb{Z}, k \in \mathbb{Z}\} \quad (6)$$

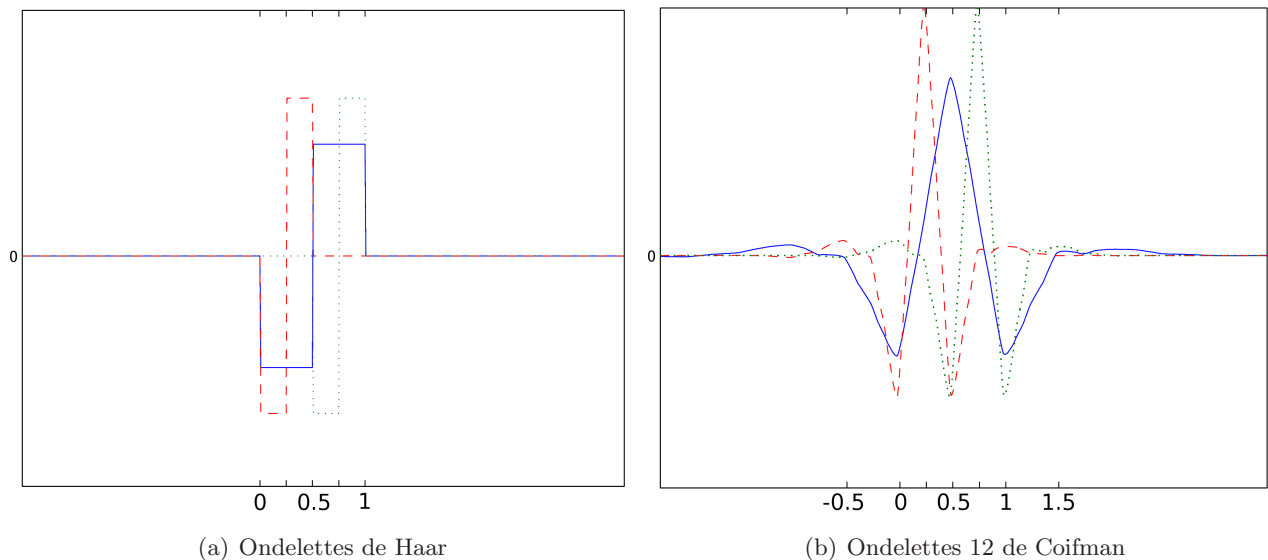


FIG. 2: Pour la famille d'ondelettes la plus simple (a), dite de Haar, qui a été introduite bien avant la théorie des ondelettes elle-même, on voit aisément que l'ondelette contractée d'un facteur 2 (en tirets rouges) est orthogonale à l'ondelette initiale (en bleu). Il en est de même pour sa voisine sur la grille dyadique (en pointillés verts). L'ondelette 12 de Coifman (b) possède également ces propriétés, mais a en plus 4 moments nuls. Sa construction s'effectue dans le formalisme de l'analyse multirésolution (paragraphe 1.3).

pour laquelle on visualise facilement (figure 2) l'orthogonalité entre les échelles, mais qui n'est même pas continue et n'a qu'un seul moment nul. La construction de bases orthonormales d'ondelettes ayant plus de moments nuls est un problème difficile. Nous en verrons un exemple un peu plus loin. Pour certaines des applications que nous considérerons, il est cependant crucial de ne pas se limiter à la stricte orthogonalité. Cela permet en effet, au prix d'une légère redondance de la transformée, d'obtenir d'autres propriétés. Une condition introduite dans [4] est que l'énergie des coefficients d'ondelette soit, à une constante près, la même que celle de la fonction analysée :

$$\|f\|_2^2 = A \sum_{\lambda \in \Lambda} |\tilde{f}(\lambda)|^2$$

auquel cas on aura presque la même formule de reconstruction que pour une base orthonormale (5) :

$$f = \frac{1}{A} \sum_{\lambda \in \Lambda} \langle \psi_\lambda | f \rangle \psi_\lambda \quad (7)$$

On dit que la famille  $(\psi_\lambda)_{\lambda \in \Lambda}$  constitue une *tight frame* de  $L^2(\mathbb{R})$ . Nous donnerons un exemple de *tight frame* dans le paragraphe 1.4. Pour les *frames*, il n'est en principe pas nécessaire de se limiter à une grille dyadique, mais c'est pourtant le seul cas que nous aurons besoin de considérer. Notons que pour les *frames* comme pour les bases orthonormales, le choix de la grille (6) implique *a priori* la perte de l'invariance par translation propre à la transformée en ondelettes continue.

### 1.3. Analyse multirésolution et transformée en ondelettes rapide

Nous allons introduire très brièvement le formalisme de l'analyse multirésolution, proposé par Mallat en 1988 et qui a permis par la suite la construction des premières bases d'ondelettes orthonormales à support compact par Daubechies. C'est aussi le cadre adapté pour établir l'algorithme de calcul rapide des coefficients d'ondelettes utilisé dans la pratique.

Nous noterons  $\psi_{j,i}$  (resp.  $\tilde{f}_{j,i}$ ) l'ondelette (resp. le coefficient) correspondant à  $\lambda = (2^{-j}, 2^{-j}i)$  sur la grille dyadique. Par abus de langage, nous dirons que ce coefficient est à l'échelle  $j$  (au lieu de  $2^{-j}$ ). Soit maintenant  $J$  un entier fixé. Nous appellerons approximation de la fonction  $f$  à la résolution  $2^J$  la somme partielle de ses coefficients d'ondelettes d'échelles  $j < J$  :

$$f_j = \sum_{j < J} \sum_{i \in \mathbb{Z}} \tilde{f}_{j,i} \psi_{j,i}$$

Ainsi, les coefficients d'ondelettes  $\tilde{f}_{J,i}$  contiennent toute l'information nécessaire pour passer à une approximation de  $f$  à la résolution  $2^{J+1}$ . Pour l'ondelette de Haar, remarquons que  $f_J$  est une approximation de  $f$  par une fonction constante sur chaque intervalle  $[\frac{i}{2^J}, \frac{i+1}{2^J}]$ . Dans le cas général, on montre qu'il existe une fonction  $\phi$ , appelée fonction d'échelle, telle que

$$(x \rightarrow \phi_{J,i}(x) = 2^{-\frac{J}{2}} \phi(2^{-J}x - i) \mid i \in \mathbb{Z})$$

constitue une base orthonormée du sous-espace des approximations de toutes les fonctions de  $L^2$  à la résolution  $2^J$ . Pour l'ondelette de Haar,  $\phi$  est simplement l'indicatrice de l'intervalle  $[0, 1]$ . Plus généralement, tout comme  $\psi$ , la fonction d'échelle  $\phi$  est une fonction test pour l'analyse de  $f$ , mais

- les coefficients de fonctions d'échelle contiennent l'information aux échelles  $j < J$ . Ils seront notés  $\bar{f}_{J,i}$ .
- les coefficients d'ondelettes contiennent l'information nécessaire pour passer de la résolution  $J$  aux résolutions  $2^{J+1}, 2^{J+2}, \dots$

On peut toujours reconstruire  $f$  à partir de ces deux familles de coefficients :

$$f = \sum_{i=-\infty}^{\infty} \bar{f}_{J,i} \phi_{J,i} + \sum_{j=J}^{\infty} \sum_{i \in \mathbb{Z}} \tilde{f}_{j,i} \psi_{j,i} \quad (8)$$

où le premier terme n'est autre que l'approximation à l'échelle  $J$ ,  $f_J$ , et le deuxième contient l'ensemble des détails supplémentaires nécessaires pour reconstruire entièrement  $f$ . Mallat a montré l'existence de filtres  $h$  et  $g$  indépendants de  $J$  tels que :

$$\begin{aligned} \phi_{J-1,i} &= \sum_{n \in \mathbb{Z}} h_{n-2i} \phi_{J,n} \\ \psi_{J-1,i} &= \sum_{n \in \mathbb{Z}} g_{n-2i} \phi_{J,n} \end{aligned}$$

où on a en fait  $g_n = (-1)^n h_n$ . Ces relations, combinées avec (8), permettent de calculer les coefficients d'ondelettes et de fonctions d'échelle de proche en proche :

$$\tilde{f}_{J-1,i} = \sum_{n \in \mathbb{Z}} h_{n-2i} \tilde{f}_{J,n} \quad (9)$$

$$\bar{f}_{J-1,i} = \sum_{n \in \mathbb{Z}} g_{n-2i} \bar{f}_{J,n} \quad (10)$$

En appliquant les relations (9-10) autant de fois que nécessaire, on peut donc calculer des approximations de  $f$  à des résolutions de plus en plus grossières, ainsi que tous les coefficients d'ondelettes qui contiennent l'information nécessaire pour revenir à la résolution de départ, d'où le nom d'analyse multirésolution. Remarquons que (9-10) sont en fait plus fondamentales qu'il n'y paraît, puisque c'est en explicitant les filtres  $g$  et  $h$  qu'on construit les bases d'ondelettes orthonormales. On cherche en général les filtres les plus courts satisfaisant un certain nombre de contraintes *a priori*. Sur la figure 2.b, on donne l'exemple de la Coiflet 12, qui est l'ondelette au support le plus court ayant 4 moments nuls, et telle que la fonction d'échelle associée a 3 moments nuls. C'est celle que nous utiliserons pour tous les exemples et applications que nous aborderons bientôt.

Supposons qu'on dispose d'un échantillonnage de  $f$  à la résolution  $2^J$ , issu par exemple d'une mesure physique :

$$f[k] = f\left(\frac{k}{2^J}\right), \quad k \in \mathbb{Z} \quad (11)$$

En pratique, on dispose d'un échantillon fini  $N$  de valeurs de la fonction  $f$ , qu'on supposera régulièrement espacés comme dans (11). Sans perte de généralité, on peut se placer sur l'intervalle  $[0, 1]$ , ce qui revient à prendre  $N = 2^J$ . Sous certaines conditions, les coefficients de fonction d'échelle à cette résolution sont bien approximatés par les valeurs de  $f$  aux points d'échantillonnage. Afin d'appliquer les relations (9-10) pour les points situés près des bords de l'intervalle  $[0, 1]$ , on aurait besoin de connaître les valeurs que prend  $f$  en dehors de cette intervalle. Nous adopterons la solution la plus simple pour contourner cet écueil, qui est de prolonger  $f$  en une fonction 1-périodique sur  $\mathbb{R}$ , comme on le fait

couramment pour calculer la transformée de Fourier discrète. Dans le cas des ondelettes, il existe une façon plus élégante de procéder, mais nous la laisserons de côté.

Partant de  $N$  coefficients de fonctions d'échelle à l'échelle  $J$ , on calcule  $\frac{N}{2}$  coefficients d'ondelettes à l'échelle  $J - 1$  par convolution avec  $h$ . Si le filtre  $h$  a pour longueur  $K$ , on a donc effectué  $\frac{N}{2} \times K$  opérations. Après  $\frac{N}{2} \times K$  opérations supplémentaires, on a calculé les coefficients de fonctions d'échelle à l'échelle  $J - 1$ . On procède récursivement jusqu'à l'échelle maximale souhaitée, que l'on notera toujours  $L$ . Le nombre d'opérations est divisé par deux à chaque fois qu'on passe à une échelle plus grossière, donc au total on aura effectué  $K \times (2^{J-1} + 2^{J-2} + \dots + 2^L)$  opérations élémentaires. La complexité de l'algorithme est donc majorée par  $K \times 2^J$ . C'est à comparer avec la transformée de Fourier rapide, dont la complexité dans une situation équivalente est en  $O(J \times 2^J)$ .

Pour reconstruire le signal discret à partir de ses coefficients d'ondelettes, il est nécessaire d'inverser le système linéaire (9-10) à chaque échelle, les inconnues étant maintenant les  $\bar{f}_{J,i}$ . Grâce aux propriétés des ondelettes orthonormales, on montre que  $\bar{f}_{J,k}$  s'obtient par une somme de convolutions de  $\bar{f}_{J-1,i}$  et  $\bar{f}_{J-1,i}$  avec des filtres indépendants de  $J$ , ce qui fournit un algorithme d'inversion exact de la transformée en ondelettes orthonormales rigoureusement symétrique de l'algorithme d'analyse, et de complexité équivalente.

#### 1.4. "Frames" d'ondelettes à valeurs complexes

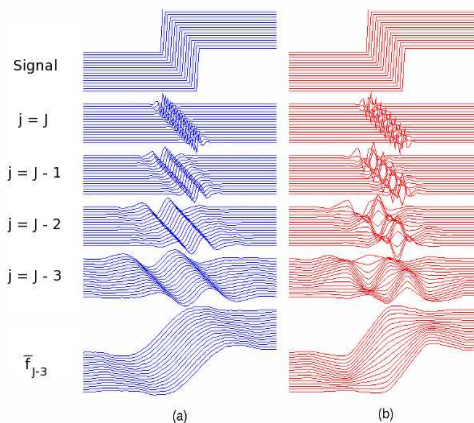


FIG. 3: Comparons les reconstructions de 16 créneaux superposables à partir de leurs coefficients d'ondelettes à différentes échelles, suivant qu'on considère une base orthonormale d'ondelettes à valeurs réelles (b) comme au paragraphe 1.3, ou une frame d'ondelettes complexes quasi-invariantes par translation (a), comme au paragraphe 1.4. Suivant le décalage entre les signaux de départ, les reconstructions de (b) présentent des oscillations différentes tandis que celles de (a) sont simplement translattées.

Comme nous l'avons déjà souligné, l'échantillonnage sur une grille dyadique, s'il permet d'éliminer la redondance, entraîne aussi la perte de l'invariance par translation. Or cela pose des problèmes pour les applications au traitement du signal. Considérons en effet un ensemble de signaux strictement superposables mais décalés dans le temps (figure 3.b). L'agencement des points de grille imposé artificiellement par l'élimination de la redondance entraîne une répartition très différente de l'information entre les échelles. Ainsi, la transformée en ondelettes orthonormales perd la trace du fait que pour la plupart des applications pratiques, deux signaux superposables portent la même signification.

Kingsbury a construit à partir de 1998 des familles d'ondelettes à valeurs complexes [13, 14], qui ne sont pas des bases orthonormales mais seulement des *tight frames* de  $L^2(\mathbb{R})$ . Au prix d'un facteur 2 de redondance (c'est à dire que la transformée d'un signal échantillonné en  $N$  points comporte  $2N$  coefficients), ces familles permettent de maintenir l'invariance par translation avec une très bonne approximation (figure 3.a). L'idée est de mettre en commun l'information fournie par deux bases orthogonales d'ondelettes, construites de façon à ce que la dépendance en translation de l'une compense presque exactement celle de l'autre. Pour des raisons que nous ne détaillerons pas, les coefficients d'un signal relativement à l'une et l'autre base se comprennent alors comme les parties réelle et imaginaire de coefficients complexes échantillonnés sur la grille dyadique.



L'intérêt de cette construction est de pouvoir utiliser tous les résultats connus sur les ondelettes orthonormales, en particulier l'algorithme rapide, pour calculer indépendamment la partie réelle et la partie imaginaire de chaque coefficient. C'est seulement au moment de l'analyse des coefficients qu'on se place dans la *frame* d'ondelettes complexes afin de retrouver la propriété d'invariance par translation. Ainsi, le coût en temps de calcul est multiplié par deux par rapport à une transformée réelle. Il en est de même pour l'espace mémoire occupé par les coefficients. Pour chaque application il s'agit de déterminer si la quasi-invariance par translation justifie ce coût additionnel.

## 2. Extraction des fluctuations cohérentes

### 2.1. Principe

#### 2.1.1. Définition des fluctuations cohérentes

Nous voulons extraire les structures cohérentes qui apparaissent au sein des fluctuations turbulentes pour les écoulements des fluides ou des plasmas quand on est en régime pleinement développé. Pour cela nous utilisons les propriétés de localisation de la représentation en base d'ondelettes à la fois en espace (ou en temps) et en échelles. Le signal que l'on étudie est tout d'abord représenté en ondelettes, plutôt que dans l'espace physique ou en modes de Fourier, comme on le fait classiquement en turbulence. Étant donné que les chercheurs travaillant dans le domaine de la turbulence ne sont pas encore arrivés à se mettre d'accord sur une définition précise et opérationnelle des structures cohérentes (que nous préférons appeler cohérentes), nous proposons la définition minimale suivante : "les fluctuations cohérentes ne sont pas du bruit" [9]. Si on admet cette hypothèse, le problème de leur extraction se ramène à un problème de débruitage. Nous n'avons plus besoin dans ce cas d'hypothèses sur les structures elles-mêmes, mais seulement sur le bruit que nous cherchons à éliminer pour les extraire. Comme première hypothèse nous choisirons la plus simple, à savoir nous supposons que le bruit est additif, Gaussien et blanc (donc décorrélé). Cette hypothèse a déjà été appliquée à des signaux plasmas dans [8]. Nous étudierons également ici une seconde hypothèse moins restrictive : nous ne supposons plus que le bruit est décorrélé (blanc), et nous proposerons un algorithme qui effectue le débruitage échelle par échelle ce qui permet de trouver la corrélation du bruit.

#### 2.1.2. Processus aléatoires

Dans cette partie et les suivantes, nous introduisons les notations nécessaires à une définition mathématique rigoureuse du bruit qui, selon notre définition, modélise la partie incohérente de l'écoulement. Nous avons supposé que le bruit est un processus Gaussien. Par souci de simplicité nous le supposons également stationnaire et de moyenne nulle. Il est alors entièrement défini par sa fonction de covariance :

$$c(\tau) = \langle X(t)X(t + \tau) \rangle$$

où  $\langle . \rangle$  désigne l'opération de moyenne d'ensemble (ou espérance mathématique). De plus, toute transformation linéaire d'un processus Gaussien reste un processus Gaussien. En particulier, les coefficients de Fourier d'un processus Gaussien stationnaire sont Gaussiens, et on montre qu'ils sont même deux à deux indépendants : ils forment une décomposition de Karhunen-Loeve du processus. Comme les phases des coefficients de Fourier sont indépendantes, le processus est toujours incohérent, indépendamment de sa fonction de corrélation. Il est donc très important de distinguer corrélation et cohérence.

Nous introduirons successivement un modèle de corrélations à longue portée et un modèle de corrélations à courte portée.

On appelle temps de corrélation du processus l'intégrale

$$\tau_C = \frac{1}{c(0)} \int_0^\infty c(\tau) d\tau \quad (12)$$

Dans beaucoup de cas, on peut s'attendre à ce que la fonction  $c(\tau)$  tende rapidement vers 0 à l'infini, c'est à dire que deux valeurs du processus en des points très éloignés sont quasiment indépendantes. Si le temps de corrélation  $\tau_C$  est fini (c'est-à-dire, en pratique, beaucoup plus court que la durée de l'expérience), on parle de corrélation à courte portée. On va s'intéresser à présent au cas  $\tau_C = \infty$ , représentatif de la situation où il y a des corrélations d'un bout à l'autre du temps d'observation.

Sur le plan mathématique, cela signifie qu'on ne peut pas considérer deux valeurs de  $X$ , même en des points très éloignés, comme des variables aléatoires indépendantes. On parle aussi de processus à longue mémoire. Cela se produit en particulier lorsque :

$$c(\tau) \asymp \tau^{-\gamma} \quad (\gamma < 1) \quad (13)$$

Expérimentalement, on n'a jamais accès à la limite  $\tau \rightarrow \infty$ , mais on a souvent observé la loi d'échelle (13) lorsque  $\tau$  devient de l'ordre de la durée de l'expérience (ou de la taille du système dans le cas d'une variable spatiale). C'est le cas en particulier des incréments des cours de certains actifs boursiers, mais la première description détaillée en a été faite par Hurst dans le cadre d'une étude hydrologique sur les niveaux d'eaux de certains barrages.

On appelle puissance spectrale du processus stationnaire  $X$  la transformée de Fourier  $\hat{c}(\nu)$  de  $c(\tau)$ . Si on regarde à quoi correspond (13) en termes de puissance spectrale, on trouve une loi de puissance du type :

$$\hat{c}(\nu) \asymp \nu^{-\alpha} \quad (14)$$

(où  $\alpha = 1 - \gamma$ ). On a donc cherché à construire des processus vérifiant exactement (14) sur toute gamme spectrale, sous le nom générique de "processus en  $\frac{1}{f}$ " (où  $f$  désigne ici la fréquence, que nous avons choisi de noter  $\nu$ ). Ce sont donc des versions idéalisées de processus à longue mémoire, dont nous allons brièvement rappeler la définition.

### 2.1.3. Corrélations à portée infinie

Revenons d'abord au mouvement Brownien standard. Comme il n'est pas stationnaire, on ne peut pas définir directement la puissance spectrale comme transformée de Fourier de  $c(\tau)$ . En revanche on peut fixer un intervalle  $[0, t_{max}]$ , et calculer la puissance spectrale moyenne des réalisations limitées à cet intervalle. Elle vérifie alors (14) avec  $\alpha = 2$ . On sait simuler facilement des réalisations indépendantes d'un mouvement Brownien standard. Pour cela, il suffit de générer ses incréments, c'est-à-dire une suite de variables aléatoires Gaussiennes indépendantes (par exemple par la méthode classique de Box-Müller). Il est donc naturel de partir de là pour construire d'autres processus à longue mémoire. Mathématiquement, la première construction explicite revient à Mandelbrot et van Ness en 1968 [16] qui ont introduit les mouvements Browniens fractionnaires (FBM). Ils sont caractérisés par les trois propriétés suivantes :

- $B_H(t)$  est un processus Gaussien sur  $\mathbb{R}$
- les incréments du processus sont stationnaires et auto-similaires :  $\forall a, t, \tau, B_H(t + a\tau) - B_H(t) \stackrel{\Delta}{=} a^H (B_H(t + \tau) - B_H(t))$  pour un certain  $H \in ]0, 1[$
- le processus est continu en moyenne quadratique

Le cas  $H = \frac{1}{2}$  correspond au mouvement Brownien standard. Le terme "fractionnaire" fait référence à la méthode de construction mathématique de  $B_H$  au moyen d'une dérivée fractionnaire. Le paramètre  $H$  contrôle la régularité des trajectoires (au sens de Lipschitz, cf (3)), comme on peut le voir sur les exemples de la figure 4. Comme modèle de processus Gaussiens stationnaires à longue mémoire, nous prendrons les incréments de mouvements Browniens fractionnaires, qu'on appelle bruits Gaussiens fractionnaires.

$$X_H[n] = B_H(n + 1) - B_H(n) \quad (n \in \mathbb{N}) \quad (15)$$

La fonction de covariance de  $X_H$  est donnée par :

$$c[n] = \mathbb{E}(X_H[0]X_H[n]) = \frac{\sigma^2}{2} (|n + 1|^{2H} + |n - 1|^{2H} - 2|n|^{2H}) \quad (16)$$

Elle vérifie en particulier (13) pour  $\gamma = 2 - 2H$ . Sa puissance spectrale vérifie (14) avec  $\alpha = 2H - 1$ , sur une large bande de fréquence. On voit que la fonction de covariance est toujours négative pour  $0 < H < \frac{1}{2}$  (anticorrélation), et toujours positive pour  $\frac{1}{2} < H < 1$  (corrélation). Pour  $H = \frac{1}{2}$ , on retrouve le cas du bruit blanc Gaussien ( $c[n] = 0$  sauf pour  $n = 0$ ). Trois exemples sont tracés sur la figure 4. Pour les simulations numériques, on a utilisé la méthode de la matrice circulante introduite dans [3], qui ne sera pas détaillée ici.



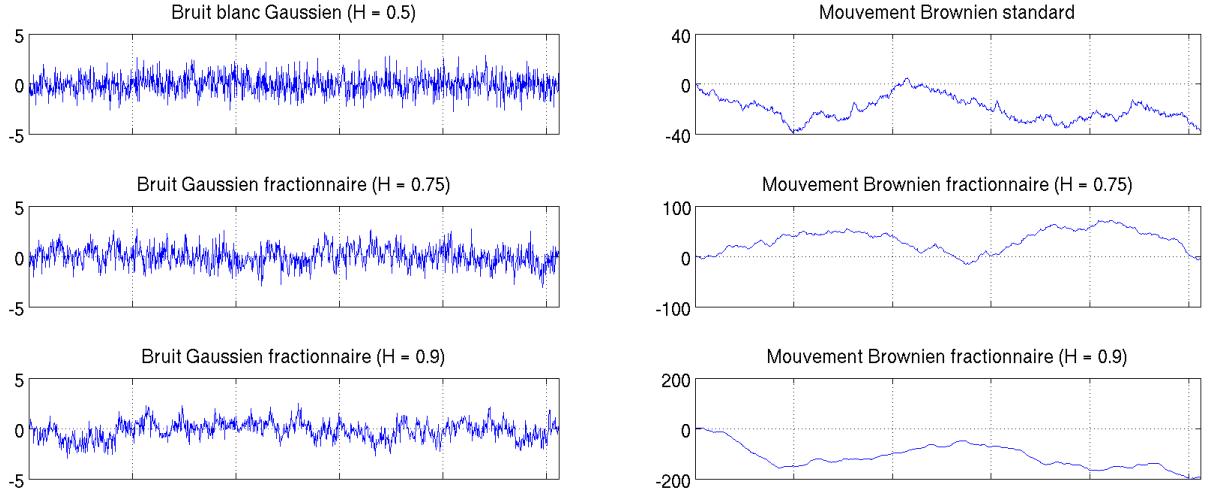


FIG. 4: Exemples de trajectoires de bruits Gaussiens fractionnaires (colonne de gauche), et des mouvements Browniens fractionnaires dont ils sont les incréments (colonnes de droite). On a utilisé trois valeurs différentes de l'exposant de Hurst,  $H$ . La résolution est fixée à  $N = 1024$ .

Il existe d'autres façons de modéliser les processus corrélés à longue portée, que nous n'aborderons pas ici. En particulier, certaines méthodes utilisent les ondelettes. On peut également partir directement du modèle discret (16) et ajouter des complications, par exemple en mélangeant deux fonctions de corrélations avec des exposants  $H_1$  et  $H_2$  différents. L'avantage essentiel des mouvements Browniens est la rigueur et la simplicité de leur construction mathématique.

#### 2.1.4. Corrélations à portée finie

Pour modéliser un processus aléatoire présentant des corrélations à courte portée, nous supposons que sa fonction de covariance décroît exponentiellement :

$$\mathbb{E}(X(t)(X(t+n))) = c[n] = \sigma^2 \exp\left(-\frac{|n|}{\tau_C}\right) \quad (17)$$

C'est en effet une façon naturelle d'introduire un temps de corrélation fini  $\tau_C$ . De plus, ce type de fonctions de corrélation intervient pour de nombreux systèmes physiques que nous ne détaillerons pas ici. Afin de simuler numériquement un bruit de ce type, on utilise un schéma intégral décrit dans [10]. Quelques exemples de trajectoires sont tracés sur la figure 5.

À partir de (17), on calcule la puissance spectrale de  $X(t)$  dont on aura besoin par la suite :

$$\hat{c}(\nu) = \frac{4\tau_C\sigma^2}{(1 + (\nu\tau_C)^2)} \quad (18)$$

Remarquons que pour  $\nu \gg \tau_C^{-1}$ ,  $\hat{c}(\nu) \propto \nu^{-2}$ .

#### 2.1.5. Propriétés élémentaires en base d'ondelettes

On admettra que les coefficients d'ondelettes de tous les processus aléatoires introduits jusqu'ici vérifient les propriétés suivantes [1] :

- ils sont presque deux à deux décorrés. On ne tiendra jamais compte des corrélations résiduelles
- leur variance ne dépend presque pas de la position, mais seulement de l'échelle. Cette propriété est liée à la stationnarité des processus. Le "presque" est lié à des effets de bords que nous négligerons.

On montre que si  $\hat{c}$  a un comportement en loi de puissance alors la variance à chaque échelle, notée  $\sigma_j^2$ , vue comme une fonction de  $2^j$ , suit le même comportement :

$$\hat{c}(\nu = 2^j) \asymp \sigma_j^2 \quad (19)$$

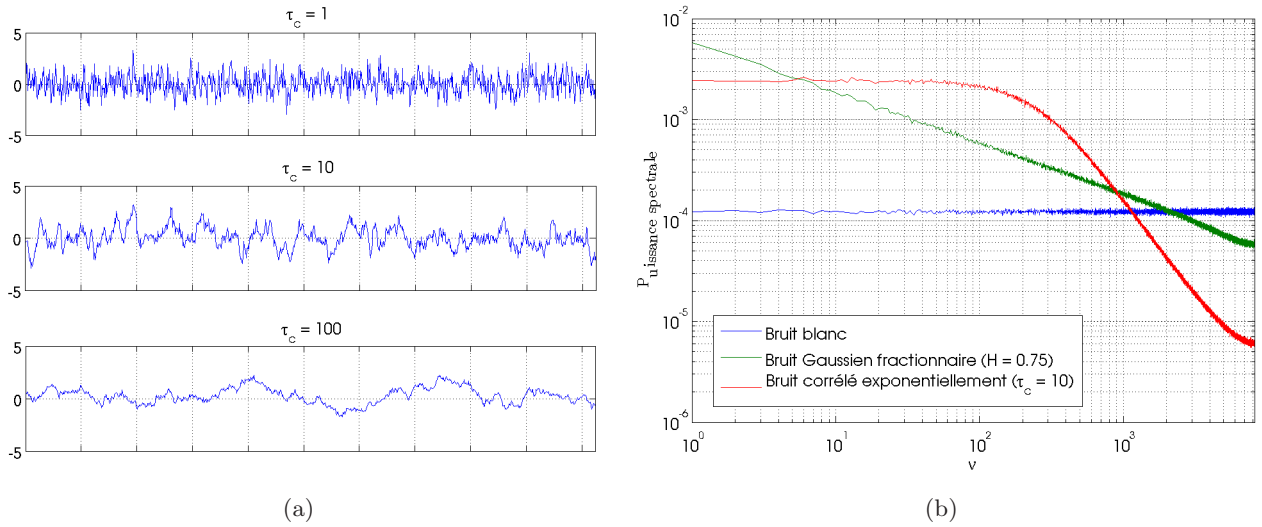


FIG. 5: (a) Exemples de trajectoires pour des bruits Gaussiens dont la fonction de covariance décroît exponentiellement. (b) Spectres moyens pour 1000 réalisations de trois types de bruits de variances identiques, échantillonnés sur  $N = 2^{14}$  points.

Ainsi pour les bruits Gaussiens fractionnaires, on a :

$$\sigma_j^2 \propto 2^{-j(1-2H)}$$

ce qui est équivalent à la relation suivante :

$$2H - 1 = \log_2(\sigma_{j-1}^2) - \log_2(\sigma_j^2) \quad (20)$$

Pour les processus à corrélation exponentielle, la décroissance de  $\sigma_j^2$  est plus rapide mais ne se produit qu'aux échelles suffisamment fines, où on a :

$$\sigma_j^2 \asymp 2^{-2j}$$

## 2.2. Extraction globale

### 2.2.1. Notations

On se place dans le cadre d'un modèle de bruit additif, où le signal observé s'écrit :

$$X[n] = F[n] + W[n], \quad n = 1..N \quad (21)$$

avec  $W$  un processus Gaussien de moyenne nulle, et  $F$  un signal déterministe donné par échantillonnage d'une certaine fonction définie sur  $[0, 1]$  :

$$F[n] = F\left(\frac{n}{N}\right)$$

On cherche à estimer  $F$  à partir d'une seule observation des données bruitées. Pour quantifier l'efficacité du débruitage on introduit le risque  $L^2$  (ou erreur quadratique moyenne) associé à un estimateur  $\mathcal{E}(X)$  :

$$R(\mathcal{E}, F) = \langle \|F - \mathcal{E}(X)\|_2^2 \rangle = \sum_{n=1}^N \langle |F[n] - \mathcal{E}(X)[n]|^2 \rangle$$

Une quantité équivalente que nous utiliserons souvent est le rapport signal sur bruit (SNR) mesuré en décibels, défini par

$$SNR = 10 \log_{10} \left( \frac{\|F\|_2^2}{R(\mathcal{E}, F)} \right)$$

Le cas échéant, il sera intéressant de considérer le risque relatif  $\frac{R(\mathcal{E}, F)}{E_N}$ , où l'énergie du bruit  $E_N$  est définie par :

$$E_N = \sum_{n=1}^N \langle |W[n]|^2 \rangle$$

On souhaite construire un estimateur  $\mathcal{E}$  qui vérifie

$$\frac{R(\mathcal{E}, F)}{E_N} \ll 1$$

pour toutes les fonctions  $F$  susceptibles de représenter nos signaux. Afin de formaliser la connaissance *a priori* que l'on a du signal, on suppose qu'il appartient à un certain ensemble de fonctions  $\Theta$ , et on souhaite alors que l'estimateur  $\mathcal{E}$  minimise le risque maximum :

$$R(\mathcal{E}, \Theta) = \sup_{F \in \Theta} R(\mathcal{E}, F) \quad (22)$$

c'est-à-dire que le risque associé à  $\mathcal{E}$  soit égal au risque minimax non linéaire :

$$R_n(\Theta) = \inf_{\mathcal{E}} R(\mathcal{E}, \Theta)$$

où la borne inférieure est prise sur tous les estimateurs  $\mathcal{E}$  possibles. Dans le formalisme minimax, deux paramètres peuvent donc influencer la qualité de l'estimation optimale :

- l'espace  $\Theta$  dans lequel on se place (ce point est détaillé dans [15]),
- le type de bruit que l'on considère

On utilisera désormais la notation  $\check{\cdot}$  pour désigner les estimateurs statistiques :  $\mathcal{E}(X) = \check{F}$

### 2.2.2. Seuillage des coefficients d'ondelettes

Donoho et Johnstone ont introduit [5] la méthode de seuillage des coefficients d'ondelette pour le débruitage de signaux non-stationnaires. Cette méthode, que nous allons maintenant brièvement exposer, permet de se rapprocher du risque minimax non-linéaire pour une classe très générale d'ensembles  $\Theta$ , c'est pourquoi on parle de méthode adaptative. La propriété essentielle qui doit être vérifiée par les éléments de  $\Theta$  pour que la méthode fonctionne est la régularité locale. Intuitivement, on peut concevoir un signal localement régulier comme présentant seulement quelques singularités isolées.

On se fixe d'abord une base orthonormale d'ondelettes, dans laquelle on calcule les coefficients du signal bruité ( $\tilde{X}_{j,k}$ ). Par transformation linéaire de l'équation (21), on obtient une équation identique pour les coefficients d'ondelettes et de fonctions d'échelle du signal :

$$\tilde{X}_{j,k} = \tilde{F}_{j,k} + \tilde{W}_{j,k} \quad (23)$$

$$\tilde{X}_{L,k} = \tilde{F}_{L,k} + \tilde{W}_{L,k} \quad (24)$$

Nous supposons désormais que  $W$  est un bruit blanc Gaussien. Grâce aux propriétés de la loi Gaussienne, les coefficients d'ondelettes du bruit sont donc Gaussiens et indépendants. L'intérêt de (23-24) par rapport à (21) est que pour les signaux localement réguliers que nous considérons, l'énergie de  $F$  est concentrée sur un faible pourcentage des coefficients d'ondelettes  $\tilde{F}_{j,k}$ , les autres étant très proches de zéro. On dit que ces signaux ont une représentation sporadique en base d'ondelettes. Cette propriété est liée au nombre de moments nuls de l'ondelette analysante comme nous l'avons évoqué à la fin du paragraphe 1.1.

L'énergie du bruit en revanche est distribuée équitablement entre tous les coefficients  $\tilde{W}_{j,k}$ . Or pour  $\alpha$  suffisamment petit, ces coefficients ont une forte probabilité d'être inférieurs en valeur absolue à  $\sigma c_\alpha$ . On estime donc que les coefficients d'ondelettes contenant la majorité de l'énergie du signal sont les coefficients  $\alpha$ -aberrants du point de vue du bruit. Les autres coefficients d'ondelettes du signal sont perdus sous le niveau du bruit, mais ils ne représentent qu'une minorité de l'énergie du signal. On les estime donc par la valeur 0 :

$$\begin{aligned} \check{\tilde{F}}_{j,k} &= \tilde{Y}_{j,k} & \text{si } |\tilde{Y}_{j,k}| > \sigma c_\alpha \\ \check{\tilde{F}}_{j,k} &= 0 & \text{sinon} \end{aligned}$$

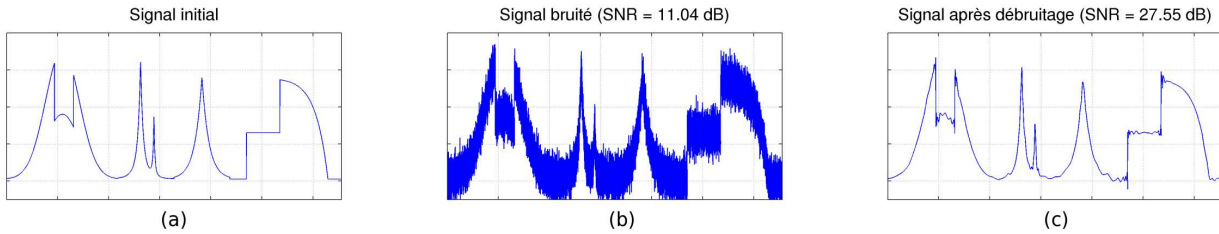


FIG. 6: Débruitage d'un signal régulier par morceaux, par seuillage de ses coefficients d'ondelettes. Le seuil est déterminé par la procédure itérative décrite au paragraphe 2.2.3.

On dit qu'on a effectué un seuillage des coefficients d'ondelettes. Considérant que l'énergie à grande échelle provient essentiellement du signal et non du bruit, on n'applique par contre pas de seuillage aux coefficients de fonctions d'échelle :

$$\bar{F}_{L,k} = \bar{Y}_{L,k}$$

On peut ensuite estimer  $F$  par transformée en ondelettes inverse des  $(\bar{F}_{L,k}, \tilde{F}_{j,k})$ . Donoho et Johnstone ont démontré plusieurs résultats mathématiques qui garantissent l'efficacité de cette méthode. En particulier, ils ont montré que lorsque le nombre de points  $N$  auxquels on échantillonne le signal tend vers l'infini, le risque associé à cet estimateur est proche du risque minimax non-linéaire  $R_n(\Theta)$ , pour toute une variété d'ensembles  $\Theta$  que nous ne détaillerons pas ici. Nous verrons plus loin des exemples qui illustrent l'efficacité pratique de cette méthode.

Notons cependant que pour l'appliquer il est nécessaire de connaître la variance  $\sigma^2$  du bruit. En pratique, on doit estimer cette variance à partir du signal bruité, ce que nous allons faire dès à présent.

### 2.2.3. Estimation de la variance

Du fait de la régularité locale du signal, la meilleure échelle pour estimer la variance du bruit est la plus petite échelle. Cependant, le signal n'est pas partout régulier au sens de Lipschitz, et la variance empirique des coefficients d'ondelettes à la plus petite échelle, qui est trop sensible aux quelques grands coefficients d'ondelettes, ne peut pas être utilisée directement. On doit utiliser un estimateur de la variance qui est robuste, c'est à dire insensible aux valeurs les plus grandes. Dans le domaine du débruitage, on utilise couramment la déviation médiane, qui est l'estimateur le plus robuste. Cependant, pour les signaux qui nous intéresseront, on préférera utiliser la méthode itérative suivante, introduite dans [2] :

1. on initialise :  $l = 0$  et  $\Lambda_{(0)}$  est l'ensemble de tous les coefficients d'ondelettes de  $X$
2. on calcule la variance empirique des éléments de  $\Lambda_{(l)}$

$$\sigma_{(l)}^2 = \frac{1}{\#\Lambda_{(l)}} \sum_{\Lambda_{(l)}} |\tilde{X}_{j,k}|^2$$

3. on extrait les indices  $\tilde{X}_{j,k}$  qui sont  $\alpha$ -aberrants par rapport à la variance  $\sigma_{(0)}^2$
4.  $\Lambda_{(l+1)}$  est défini comme l'ensemble des coefficients d'ondelettes restants
5. si  $\Lambda_{(l)} = \Lambda_{(l+1)}$  l'algorithme se termine, sinon on retourne à l'étape 2 avec  $l \rightarrow l + 1$

La convergence en un nombre fini d'itérations de cet algorithme sous des formes légèrement différentes a pu être démontrée [2, 11]. Nous admettrons ici qu'elle a lieu, ce que nous avons observé dans toutes les situations pratiques.  $\sigma_{(l)}^2$  est ensuite retenu comme estimation robuste de la variance du bruit. En effet, par définition, l'ensemble  $\Lambda_{(l)}$  ne contient plus aucun point  $\alpha$ -aberrant, tandis que son complémentaire ne contient que des points  $\alpha$ -aberrants. La figure (6) montre un exemple de débruitage d'un signal régulier par morceaux (a) détérioré par un bruit blanc Gaussien additif de variance  $\sigma^2 = 25$  (b). L'algorithme itératif a donné 25.6 comme estimation de  $\sigma^2$  ce qui a permis d'appliquer le seuillage (2.2.2) aux coefficients d'ondelettes et d'obtenir une bonne estimation (c) du signal de départ.

#### 2.2.4. Utilisation des ondelettes à valeurs complexes

Dans tout ce qui précède, nous n'avons pas utilisé le fait que les coefficients d'ondelettes étaient à valeurs réelles. Nous avons juste utilisé la définition des points  $\alpha$ -aberrants et de la covariance. Or nous avons vu dans la partie 2 que ces définitions avaient leur équivalent pour les processus à valeurs complexes. La même méthode de débruitage peut donc être appliquée dans une *frame* d'ondelettes à valeurs complexes, pour les bruits blancs comme pour les bruits corrélés, comme cela a été fait dans [17] dans un tout autre contexte.

#### 2.2.5. Choix du paramètre $c_\alpha$

Supposons qu'on dispose de  $M$  réalisations indépendantes d'une variable aléatoire Gaussienne standard, par exemple les coefficients d'ondelettes d'un signal à une échelle donnée. Pour un certain  $\alpha$  fixé, la probabilité que le module d'une de ces réalisations soit plus grand que  $c_\alpha$  est égale à  $\alpha$ . Le nombre de réalisations de module plus grand que  $c_\alpha$  est donc de l'ordre de  $\alpha M$ . Pour que ce nombre reste borné, il faut faire dépendre  $\alpha$  de  $M$  :

$$\alpha < \frac{1}{M}$$

On montre que le choix  $c_\alpha = \sqrt{2 \log M}$  pour les ondelettes réelles (et donc  $c'_\alpha = \sqrt{\log N}$  pour les ondelettes complexes) permet de remplir cette condition. C'est donc la valeur que nous avons adoptée. Par exemple, pour estimer la variance à l'échelle  $j$  d'un signal 1D, nous disposons de  $2^j$  coefficients d'ondelettes réels, et nous prendrons donc  $c_\alpha = \sqrt{2 \log(2^j)}$ . Notons que ce choix n'a rien d'universel.

### 2.3. Extraction échelle par échelle

Comme cela a été proposé à maintes reprises [19, 18, 12], on peut généraliser la méthode de débruitage en ondelettes à des bruits corrélés en appliquant un seuil différent à chaque échelle. Tout au long de cette étude, nous ne chercherons pas à établir de résultats mathématiques concernant le comportement asymptotique des méthodes de débruitage proposées lorsque la résolution d'échantillonnage tend vers l'infini. Nous raisonnerons à résolution fixée, en faisant varier la corrélation du bruit. Dans un premier temps, nous mettrons en évidence deux obstacles qui empêchent de réaliser un débruitage efficace pour des corrélations quelconques du bruit. Puis nous introduirons des hypothèses supplémentaires qui permettent de débruiter une classe de bruits dont la régularité est suffisamment faible. Enfin, nous verrons comment réaliser ce débruitage dans la pratique.

#### 2.3.1. L'obstacle de la régularité

Nous avons déjà souligné l'importance de la notion de régularité pour le débruitage. C'est la régularité locale qui contrôle la sporadicité de la transformée en ondelettes du signal, et donc la possibilité de l'extraire d'un bruit blanc. Si dans (21)  $W$  n'est plus un bruit blanc mais un processus Gaussien stationnaire quelconque, il se peut que ses trajectoires soient plus régulières que celles d'un bruit blanc, c'est-à-dire que ses coefficients d'ondelette tendent vers zéro aux petites échelles. Nous continuerons cependant à appeler  $W$  un bruit car il perturbe l'observation du signal  $F$ . Mais l'énergie de ce bruit, au lieu d'être également répartie entre tous les coefficients d'ondelettes, est variable suivant les échelles. Par analogie avec le cas du bruit blanc, on voit qu'il faut maintenant appliquer un seuil différent à chaque échelle :

$$\begin{aligned} \check{F}_{j,k} &= \check{Y}_{j,k} & \text{si } |\check{Y}_{j,k}| < \sigma_j c_\alpha \\ \check{F}_{j,k} &= 0 & \text{sinon} \end{aligned}$$

où  $\sigma_j$  est la variance du bruit à l'échelle  $j$ . Voyons comment les performances du débruitage sont affectées par la présence de corrélations dans le bruit, pour les deux modèles de corrélations que nous avons introduits. Afin de s'affranchir du problème d'estimation de la variance, nous nous placerons dans le cas idéal où on connaît à l'avance le signal  $F$  à débruiter, auquel cas on peut calculer la valeur de  $c_\alpha$  qui minimise le risque à chaque échelle (figure 7). On dit qu'on réalise un débruitage par seuillage avec oracle.

Dans chaque cas, on met en évidence une limite forte au débruitage par seuillage des coefficients d'ondelettes :

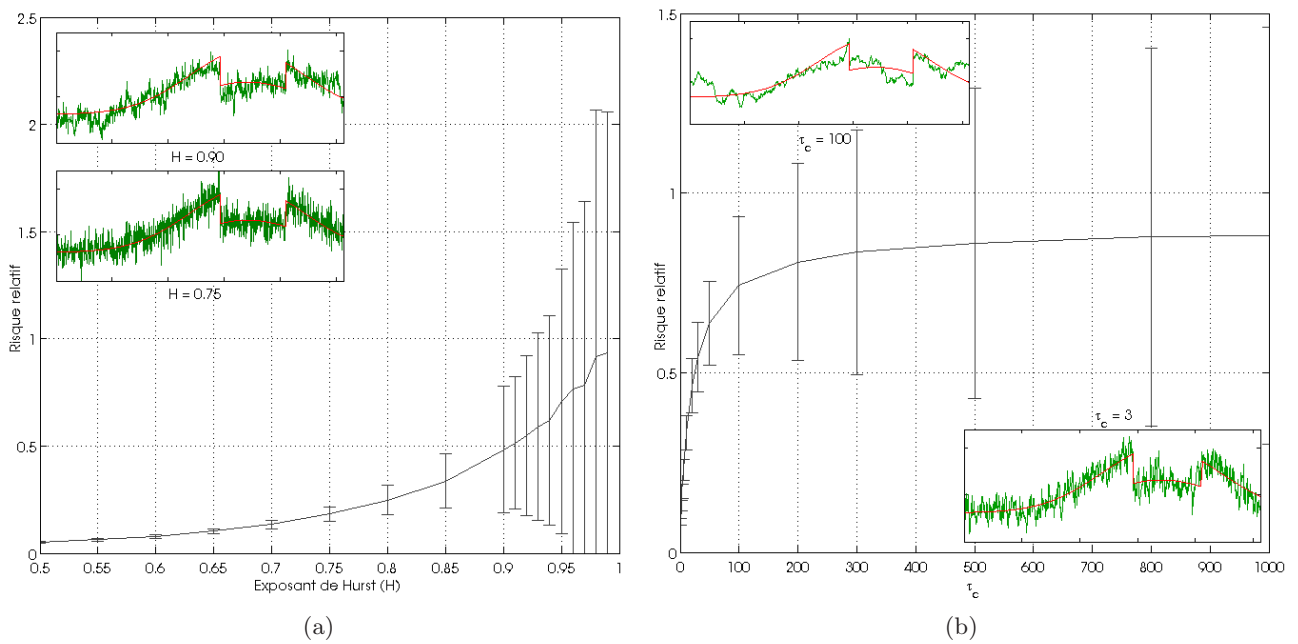


FIG. 7: Influence des corrélations du bruit sur les performances du débruitage. Pour le signal régulier par morceau (figure 6) contaminé par un bruit corrélé, on a tracé le risque relatif de l'estimateur par seuillage optimal échelle par échelle (moyenné sur 1000 réalisations du bruit). On a considéré les deux modèles de bruit développés au paragraphe 2.1.2, soit (a) un bruit Gaussien fractionnaire où l'on fait varier l'exposant de Hurst, soit (b) un bruit à corrélations exponentielles, où l'on fait varier le temps de corrélation. Le rapport signal sur bruit est fixé à 5 dB et la résolution à  $N = 2^{12} = 4096$ . Chaque encart montre un zoom sur les 1024 premiers points du signal original (rouge) ou bruité (vert). Les barres d'erreur donnent une indication sur la variabilité du risque au sein des 1000 réalisations.

- pour le bruit Gaussien fractionnaire, lorsque l'exposant de Hurst tend vers 1
- pour le bruit corrélé exponentiellement, lorsque le temps de corrélation dépasse quelques pourcents de la taille du signal

Dans chaque cas, plus le bruit devient organisé à grande échelle, moins le débruitage est efficace. Ce résultat restera qualitativement vrai quels que soient les bruits et les signaux considérés, bien que les seuils critiques puissent bien entendu varier. Étant donné le résultat obtenu sur le bruit Gaussien fractionnaire, nous proposons comme mesure du degré d'organisation du bruit la régularité de ses trajectoires.

Mais remarquons que pour le bruit corrélé exponentiellement, la décroissance du spectre reste en  $\nu^{-2}$  quel que soit le temps de corrélation, ce qui implique que la régularité des trajectoires, au sens mathématique, ne change pas. Pourtant, augmenter le temps de corrélation revient à déplacer l'énergie vers les basses fréquences (figure 5), ce qui augmente le degré d'organisation du bruit jusqu'au point où il n'est plus possible de le distinguer du signal. Nous verrons dans la partie 4 un cas concret où cela se produit. Malheureusement, nous n'avons pas trouvé de solution pour résoudre ce problème.

### 2.3.2. L'obstacle de l'estimation de la covariance

Dans le paragraphe précédent, nous avons utilisé le seuil optimal échelle par échelle calculé en connaissant à l'avance le signal à débruiter. En pratique, on n'a jamais accès à ce seuil optimal, et on doit choisir un seuil en estimant la variance de chaque coefficient d'ondelettes du bruit. Si on ne faisait aucune hypothèse sur la fonction de covariance  $\hat{c}$ , il faudrait l'estimer en  $N$  points à partir d'un échantillon de  $N$  valeurs de  $X$ , ce qui est évidemment irréaliste. On se placera donc uniquement dans le cas de bruits vérifiant les hypothèses du paragraphe 2.1.5.

On doit donc estimer la variance  $\sigma_j^2$  des coefficients d'ondelettes du bruit à chaque échelle. Deux problèmes bien distincts vont limiter la qualité de cette estimation :

1. Une limitation d'ordre statistique, car aux grandes échelles on dispose de très peu de coefficients d'ondelettes pour réaliser cette estimation. D'après le théorème de la limite centrale, on sait que l'erreur sur l'estimation d'une moyenne statistique à partir de  $2^j$  réalisations décroît en  $2^{-j/2}$ . Plus précisément, on montre que l'écart type de l'estimation de  $\sigma_j^2$  par la variance empirique



pour un bruit Gaussien est égal à  $\sqrt{\frac{2}{2^j-1}}\sigma_j$ . Afin d'avoir une erreur relative inférieure à 10% avec une probabilité supérieure à 0.95, on ne doit donc procéder à cette estimation que si  $j \geq 8$ .

2. Il ne suffit plus que la représentation de  $F$  en base d'ondelettes soit globalement sporadique, il faut maintenant qu'à chaque échelle où l'on va estimer le bruit, suffisamment peu de coefficients d'ondelettes soient affectés par  $F$ . En général, on s'attend à ce que  $F$  soit sporadique seulement aux échelles les plus fines. On définit donc  $j_{sp}$  comme l'échelle la plus grossière où il est possible d'estimer la variance du bruit. Dans [12], les auteurs proposent d'utiliser les connaissances que l'on a sur le signal pour fixer *a priori* l'échelle  $j_{sp}$ . Nous verrons dans le paragraphe suivant une façon d'estimer  $j_{sp}$  en faisant seulement une hypothèse supplémentaire sur le bruit.

### 2.3.3. L'algorithme itératif échelle par échelle

D'après ce que nous avons vu au paragraphe 2.3.1, on peut conjecturer que l'hypothèse qui donnerait les meilleurs résultats est que la variance du bruit ne soit pas trop concentrée aux grandes échelles. En effet, cela engloberait à la fois le problème des corrélations à portée finie et des corrélations à portée infinie. Cependant il n'est pas possible en pratique de vérifier si cette hypothèse est réalisée, car il faudrait estimer la variance du bruit à toutes les échelles, or on n'a en général accès qu'à la variance du bruit aux échelles les plus fines. On choisit donc comme pis-aller de mesurer la régularité des trajectoires du bruit. Pour cela, on introduit la quantité

$$q_j = \log_2(\sigma_{j-1}^2) - \log_2(\sigma_j^2)$$

Supposons que

$$\forall j, q_j < q_{max} \quad (25)$$

alors en sommant ces inégalités :

$$\log_2(\sigma_L^2) - \log_2(\sigma_j^2) < (J - L)q_{max}$$

ce qui est équivalent à

$$\forall j, \sigma_j^2 > 2^{(j-L)q_{max}} \sigma_L^2$$

D'après le théorème (4) c'est une condition suffisante pour que la régularité Lipschitzienne des trajectoires soit inférieure à  $\frac{q_{max}-1}{2}$ . Cette condition conduit naturellement à l'algorithme suivant pour l'estimation de la variance du bruit à chaque échelle :

1. initialiser  $j = J$ , et calculer  $\sigma_j^2$  par la méthode itérative du paragraphe 2.2.3 appliquée aux seuls coefficients d'ondelettes à l'échelle  $j$ ,  $\{\tilde{X}_{j,k}\}$ .
2. estimer  $\sigma_{j-1}^2$  de la même façon
3. calculer  $q_j$
4. si  $q_j \leq q_{max}$ , faire  $j \rightarrow j + 1$ , retourner à l'étape 2
5. si  $q_j > q_{max}$ , l'estimation  $\sigma_{j-1}^2$  n'est pas compatible avec notre hypothèse sur le bruit. Notre méthode ne permet pas d'estimer de façon consistante la variance du bruit aux échelles plus grossières que  $j = j_{sp}$  (où l'indice *sp* signifie sporadique) et nous devons procéder d'une autre manière.

Pour débruiter le signal aux échelles  $j < j_{sp}$ , nous avons envisagé deux méthodes :

1. Extrapolation : si on dispose d'un modèle paramétrique pour le bruit, l'information obtenue aux échelles les plus fines peut suffire à déterminer  $\sigma_j$  à toutes les échelles. Par exemple, un bruit Gaussien fractionnaire est entièrement caractérisé par deux paramètres,  $\sigma^2$  et  $H$ . On peut estimer simplement  $H$  à partir de  $\sigma_J$  et  $\sigma_{J-1}$  grâce à (20). On obtient ensuite  $\sigma_j$  pour tout  $j$  par extrapolation.
2. Seuil conservatif : on estime la variance aux échelles non sporadiques par  $\sigma_{j_{sp}}^2$ , c'est-à-dire la variance à la dernière échelle sporadique.

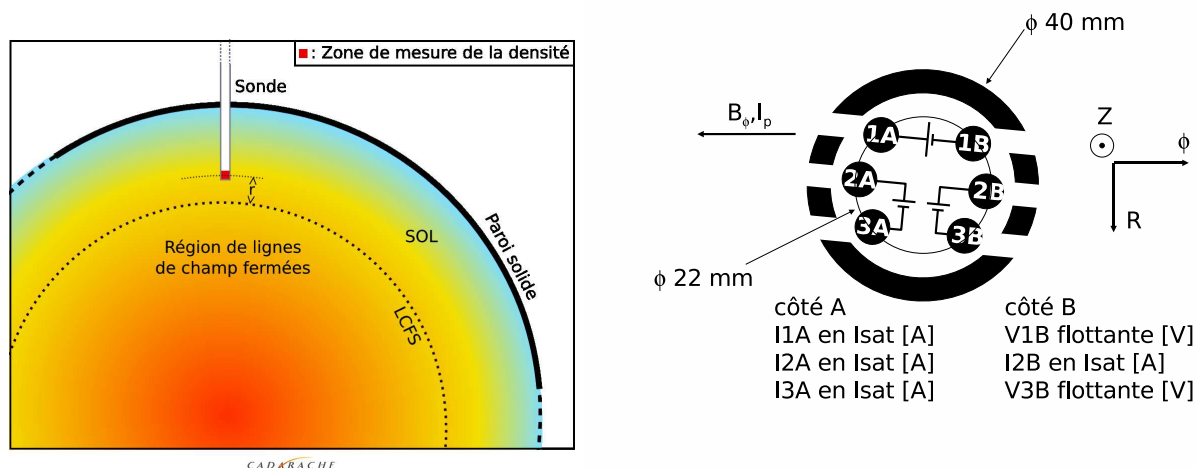


FIG. 8: À gauche : Section verticale du tokamak Tore Supra indiquant la distance séparant la sonde de la séparatrice. La sonde est mobile verticalement et ne doit rester plongée dans la SOL que pendant le bref instant nécessaire à la mesure, faute de quoi elle serait détériorée. À droite : Vue du dessus schématique de la sonde

### 3. Applications aux signaux du tokamak Tore-Supra

#### 3.1. Protocole expérimental

Dans le tokamak Tore Supra, un plasma chaud est confiné par un champ magnétique intense dans des tubes de flux de forme torique emboîtés les uns dans les autres. Comme le confinement n'est pas parfait, une partie des électrons et des ions qui composent le plasma s'échappent de ces tores et vont collisionner sur des éléments solides prévus à cet effet, les limiteurs. Entre la dernière surface de flux fermée (séparatrice) et les limiteurs se trouve une région de forts gradients de température et de densité, appelée SOL (scrape-off layer).

On dispose de mesures, effectuées par Nicolas Fedorczak, Pascale Monier-Garbet et Jamie Gunn (CEA Cadarache) du courant de saturation ionique arrivant sur une des électrodes d'une sonde de Langmuir placée dans la SOL, successivement à quatre valeurs différentes de la distance  $r$  à la séparatrice (figure 8). Pour des raisons que nous ne détaillerons pas ici, ce courant de saturation est proportionnel à la densité  $n(t)$  d'ions au voisinage de la sonde. En analysant des séries temporelles enregistrées pendant  $16ms$  à une fréquence de  $1MHz$ , on cherche à caractériser les bouffées intermittentes qui jouent un rôle important pour le transport de particules et de chaleur à travers la SOL.

On dispose également de mesures du potentiel flottant qui, sous certaines hypothèses, dépend linéairement du potentiel plasma. Les isolignes du potentiel plasma dans le plan poloidal sont quant-à elles reliées aux lignes de champ de la vitesse de dérive ionique. Dans la suite, nous parlerons du potentiel sans nous soucier de la signification physique de ces mesures. Nous oublierons également les unités associées aux grandeurs physiques car bien que les algorithmes que nous utilisons soient non linéaires, leurs résultats ne sont pas affectés si on multiplie le signal d'entrée par une constante.

Quatre pinoches de la sonde (figure 8), notées  $I_{A1}$ ,  $I_{A2}$ ,  $I_{A3}$  et  $I_{B2}$  mesurent le courant de saturation. Les deux pinoches restantes ( $V_{B1}$  et  $V_{B3}$ ) mesurent le potentiel flottant. Pour chaque pinoche nous disposons de 5 séries temporelles indépendantes. La première a été acquise en l'absence de plasma, et les quatre autres correspondent à quatre distances  $r$  différentes entre la sonde et la séparatrice :  $r = 15, 20, 35$  et  $70mm$ . Pour une distance  $r$  donnée, les mesures des six pinoches sont effectuées simultanément, et il existe des corrélations entre les séries temporelles correspondantes. Notons que pour pouvoir effectuer les transformations en ondelettes, nous avons dû augmenter la taille des signaux jusqu'à  $N = 2^{14} = 16384$ , en complétant par la valeur moyenne entre  $t = 16001$  et  $t = 16384$ . En revanche, toute l'analyse ultérieure (spectres, calcul des moments, etc) est faite uniquement sur l'intervalle de temps  $[1, 16000]$ .

La figure 9 illustre les propriétés les plus immédiates des mesures du courant de saturation en prenant l'exemple de la pinoche  $I_{B2}$ . C'est une quantité positive dont la moyenne et la variance décroissent à mesure que la sonde s'éloigne du plasma. À première vue, seuls les signaux correspondant à  $r = 35mm$  et  $r = 70mm$  présentent des excursions piquées jusqu'à des valeurs de l'ordre de 3 fois

supérieures à la médiane. Les graphes quantile-quantile montrent que les distributions de ces signaux sont toutes asymétriques et sous-Gaussiennes du côté gauche. Elles sont de plus sur-Gaussiennes du côté droit lorsque  $r \geq 20\text{mm}$ , et sous-Gaussienne du côté droit lorsque  $r = 15\text{mm}$ . Ce type d'asymétrie se retrouve souvent pour les distributions supportées sur  $[0, +\infty[$ , comme par exemple les lois Gamma.

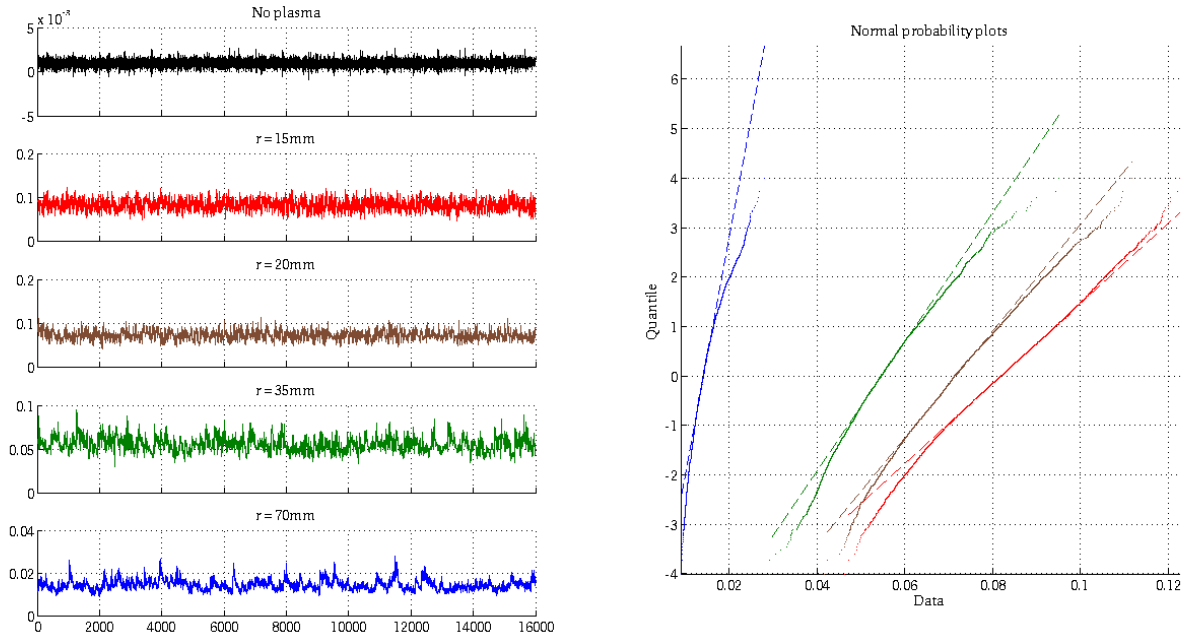


FIG. 9: Signaux d'intensité  $IB2$  et leurs graphes quantile-quantile Gaussiens

Considérons maintenant la pinoche  $V_{B1}$  mesurant le potentiel (figure 10). Les signaux correspondants sont de moyenne nulle et leur variance décroît avec la distance au plasma. Pour  $r = 35\text{mm}$  et  $r = 70\text{mm}$  leurs distributions sont symétriques mais sur-Gaussiennes, tandis que pour  $r = 15\text{mm}$  et  $r = 20\text{mm}$  elles sont légèrement asymétriques en direction des valeurs négatives du potentiel.

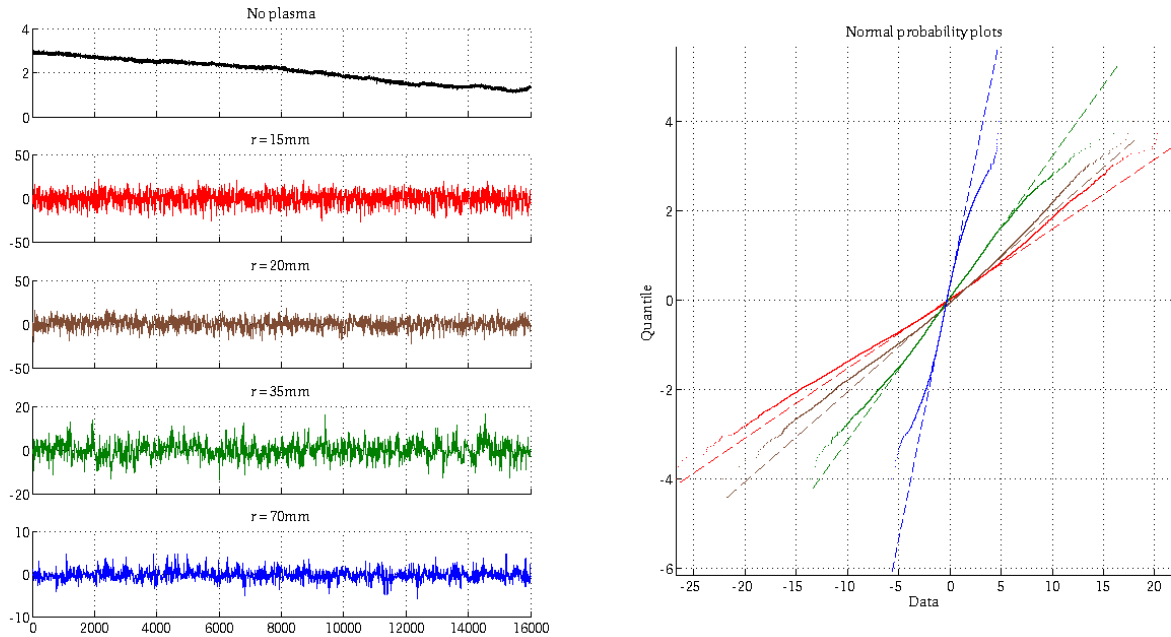


FIG. 10: Signaux d'intensité  $VB1$  et leurs graphes quantile-quantile Gaussiens

Dans la suite, nous considérerons uniquement la différence

$$U = V_{B3} - V_{B1}$$

### 3.2. Du débruitage à l'extraction des structures cohérentes

#### 3.2.1. Extraction dans les signaux de courant de saturation et de potentiel

En principe, on peut utiliser le modèle de bruit additif (21) et donc les résultats de la partie précédente pour décrire tout signal provenant d'une mesure physique. Dans cette partie, nous appliquons la méthode aux signaux présentés précédemment. Conformément au vocabulaire introduit dans le contexte des écoulements fluides, nous appellerons partie cohérente d'un signal le résultat obtenu après application d'un algorithme de débruitage à ce signal. La différence entre le signal de départ et la partie cohérente sera quant-à elle appelée partie incohérente. Ainsi :

$$\begin{aligned} I(t) &= I^c(t) + I_{incoherent}^i(t) \\ U(t) &= U^c(t) + U^i(t) \end{aligned}$$

Nous pensons que cette séparation apporte des informations sur les caractéristiques de la turbulence qui règne dans le plasma dont sont issus les signaux. Par exemple, nous avons observé dans les fluides qu'une forte variance de la partie incohérente est liée à un fort taux de turbulence. Les coefficients associés à la partie cohérente sont liés à la présence de structures qui, dans le cas des fluides, sont attachées aux maxima de la vorticit   c'est-  -dire aux tourbillons. Comme nous ne disposons ici que de signaux unidimensionnels l'interpr  tation de ces structures est beaucoup plus difficile. Nous verrons d'ailleurs que suivant l'algorithme utilis   le r  sultat obtenu peut   tre qualitativement diff  rent.

Afin de tester l'ad  quation des signaux exp  rimentaux avec le mod  le sous-jacent que nous avons pr  sent   dans la partie 3, nous serons   galement amen  s   tudier tout au long de cette partie les propri  t  s statistiques de la partie incoh  rente, en particulier sa Gaussianit  .

Le code couleur sera toujours : **signal total**, **partie coh  rente**, **partie incoh  rente**.

#### 3.2.2. Extraction dans le signal de flux

Comme le flux  $\Gamma$  est une quantit   du second ordre, il requiert un traitement diff  rent. Nous avons les s  parations :

$$\begin{aligned} I_{B2}(t) &= I_{B2}^c(t) + I_{B2}^i(t) \\ U(t) &= U^c(t) + U^i(t) \end{aligned}$$

Il est donc naturel de d  finir les quatres contributions suivantes au flux :

$$\begin{aligned} \Gamma^{cc}(t) &= I_{B2}^c(t) \times U^c(t) \\ \Gamma^{ci}(t) &= I_{B2}^c(t) \times U^i(t) \\ \Gamma^{ic}(t) &= I_{B2}^i(t) \times U^c(t) \\ \Gamma^{ii}(t) &= I_{B2}^i(t) \times U^i(t) \end{aligned}$$

qui m  nent    la s  paration suivante :

$$\Gamma(t) = \Gamma^{cc}(t) + \Gamma^{ci}(t) + \Gamma^{ic}(t) + \Gamma^{ii}(t)$$

qui s'applique de la m  me fa  on    la moyenne temporelle de  $\Gamma(t)$ . Notons qu'afin d'interpr  ter physiquement chacun de ces quatre termes, il faudrait revenir    la relation pr  cise entre le courant de saturation et la densit   du plasma (respectivement entre le potentiel flottant et la vitesse de d  rive ionique). Ici, nous nous bornerons    remarquer que le flux se s  pare en un flux purement coh  rent  $\Gamma^{cc}$ , et trois flux incoh  rents de natures diff  rentes.

#### 3.2.3. Comportement des algorithmes sur les signaux mesur  s en l'absence de plasma

Nous ne consid  rerons dans ce paragraphe que les signaux de courant de saturation. Une s  rie temporelle typique est repr  sent  e sur la figure 11, o   l'on remarque que les quelques coefficients d'ondelettes (de l'ordre de 0.1%) retenus dans la partie coh  rente semblent r  partis    une fr  quence de l'ordre de 1.2 kHz, et se situent    l'  chelle la plus fine. Cependant, le pourcentage de coefficients coh  rents est syst  matiquement inf  rieur    1% (tableau 1). Ainsi, pour toutes les analyses en pr  sence

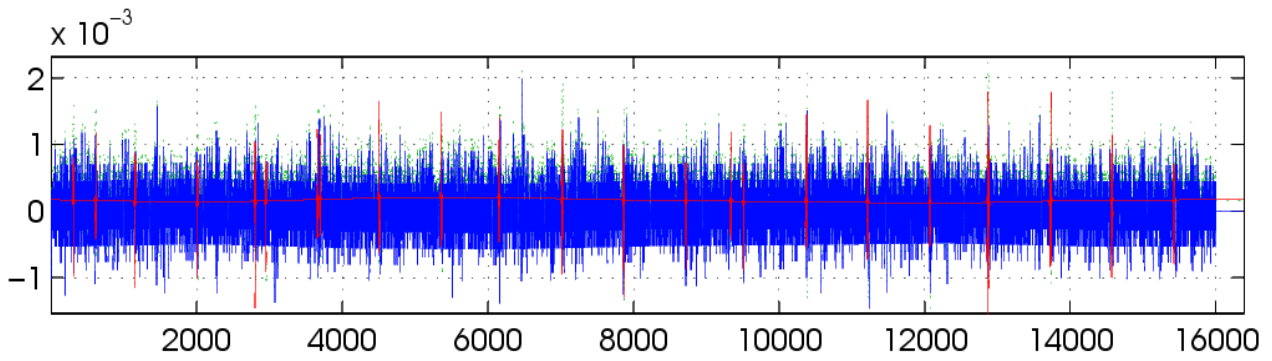


FIG. 11: Résultat typique de l'extraction des structures cohérentes dans un signal mesuré en l'absence de plasma (pinoche  $IA_1$ , seuillage global réel)

		Réel	Complexe	Réel échelle par échelle	Complexe échelle par échelle
<b>Variance cohérente (% du total)</b>	$IA_1$	0.9	0.8	2.3	7.8
	$IA_2$	5.8	6.6	0.5	5.4
	$IA_3$	1.1	1.1	1.4	7.0
	$IB_2$	10.2	10.4	7.8	9.6
<b>Coefficients cohérents (% du total)</b>	$IA_1$	0.1	0.3	0.1	0.6
	$IA_2$	0.2	0.6	0.1	0.4
	$IA_3$	0.1	0.3	0.1	0.5
	$IB_2$	0.3	0.8	0.2	0.7

TAB. 1: Comportement des quatre algorithmes sur les signaux d'intensité mesurés en l'absence de plasma

de plasma, il est probable que le bruit dû à l'appareillage se retrouve dans la partie incohérente, sans peut-être (et il faudra y prendre garde) certains pics à l'échelle la plus fine. Cette situation où un "bruit" au sens expérimental est reconnu comme un signal cohérent par nos algorithmes a malheureusement déjà été rencontrée dans d'autres situations, et il n'y a pas de remède miracle sous les hypothèses que nous nous sommes fixées. En effet, ce comportement est lié au caractère non Gaussien de beaucoup de bruits expérimentaux.

### 3.3. Extraction globale

#### 3.3.1. Gaussianité de la partie incohérente

L'assymétrie de la partie incohérente (tableau 2a) est systématiquement inférieure (en valeur absolue) à celle du total. En revanche, son coefficient d'aplatissement (tableau 2b) est souvent proche de voire supérieur à celui du total. Ainsi, l'adéquation des signaux avec le modèle de bruit additif blanc Gaussien est loin d'être idéale. En particulier, la figure 12 montre que des corrélations de phase subsistent dans les parties incohérentes, et que celles-ci sont responsables du caractère non Gaussien. Par la suite, nous verrons que l'adéquation à un modèle de bruit Gaussien corrélé est souvent meilleure.

#### 3.3.2. Corrélations croisées entre mesures simultanées

Pour une distance  $r$  à la séparatrice fixée, les signaux correspondant aux six pinoches sont enregistrés simultanément. Les corrélations croisées entre ces signaux nous renseignent donc sur les corrélations spatiales du plasma. Dans le cadre de l'extraction des structures cohérentes, les questions suivantes nous intéressent particulièrement :

1. Peut-on interpréter les corrélations croisées entre pinoches comme étant dues au passage de structures cohérentes ?
2. Si oui, peut-on utiliser la corrélation croisée entre les parties cohérentes de deux pinoches pour en déduire des informations quantitatives sur la propagation de ces structures ?

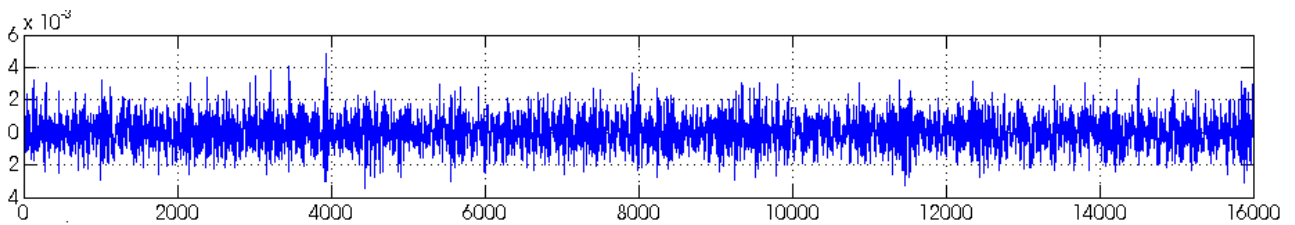
	$r = 15mm$	$r = 20mm$	$r = 35mm$	$r = 70mm$
$IA2$	0.61 0.05	0.59 0.18	0.21 0.11	0.51 -0.11
$IB2$	0.12 0.01	0.26 0.07	0.41 0.31	1.03 0.13
$U$	-0.04 -0.18	0.02 0.03	-0.21 -0.11	-0.13 -0.12

(a)

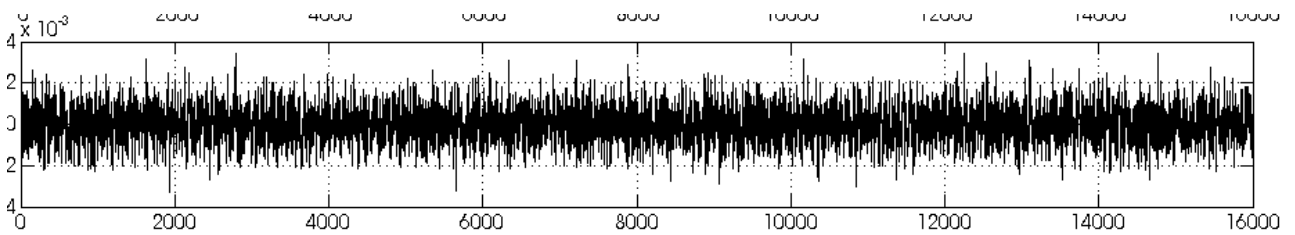
	$r = 15mm$	$r = 20mm$	$r = 35mm$	$r = 70mm$
$IA2$	4.5 4.3	3.9 5.0	3.4 5.2	4.0 3.6
$IB2$	2.7 3.0	3.1 3.5	3.2 4.2	4.9 3.7
$U$	3.0 3.2	3.0 3.3	3.7 3.5	4.2 3.6

(b)

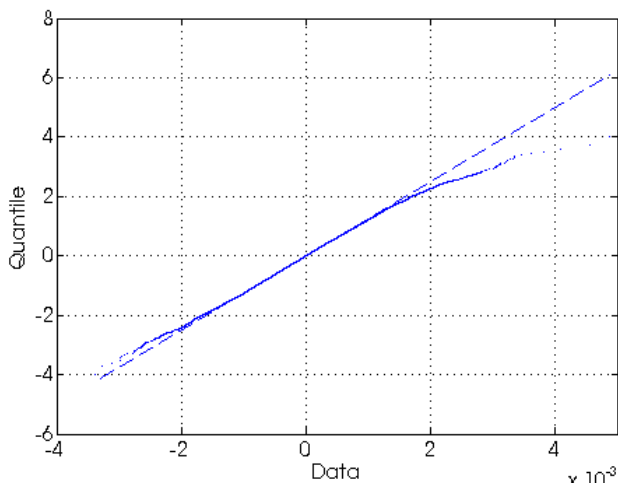
TAB. 2: Coefficients d'asymétrie (a) et d'aplatissement (b) des signaux Tore Supra ainsi que de la partie incohérente obtenue après seuillage global en ondelettes réelles orthogonales (algorithme standard).



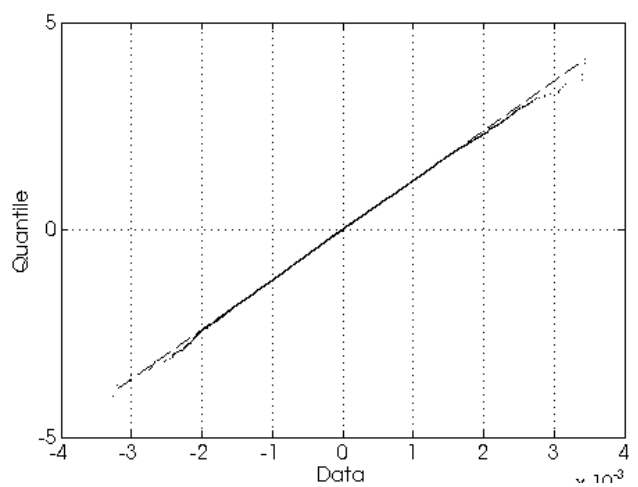
(a)



(b)



(c)



(d)

FIG. 12: (a) Série temporelle de la partie incohérente associée à la pinoche  $IB1$  pour  $r = 70mm$  par l'algorithme standard. (b) Série temporelle obtenue à partir de (a) par randomisation de la phase des coefficients de Fourier. (c) et (d) Graphes quantile-quantile Gaussien respectifs de (a) et (b).



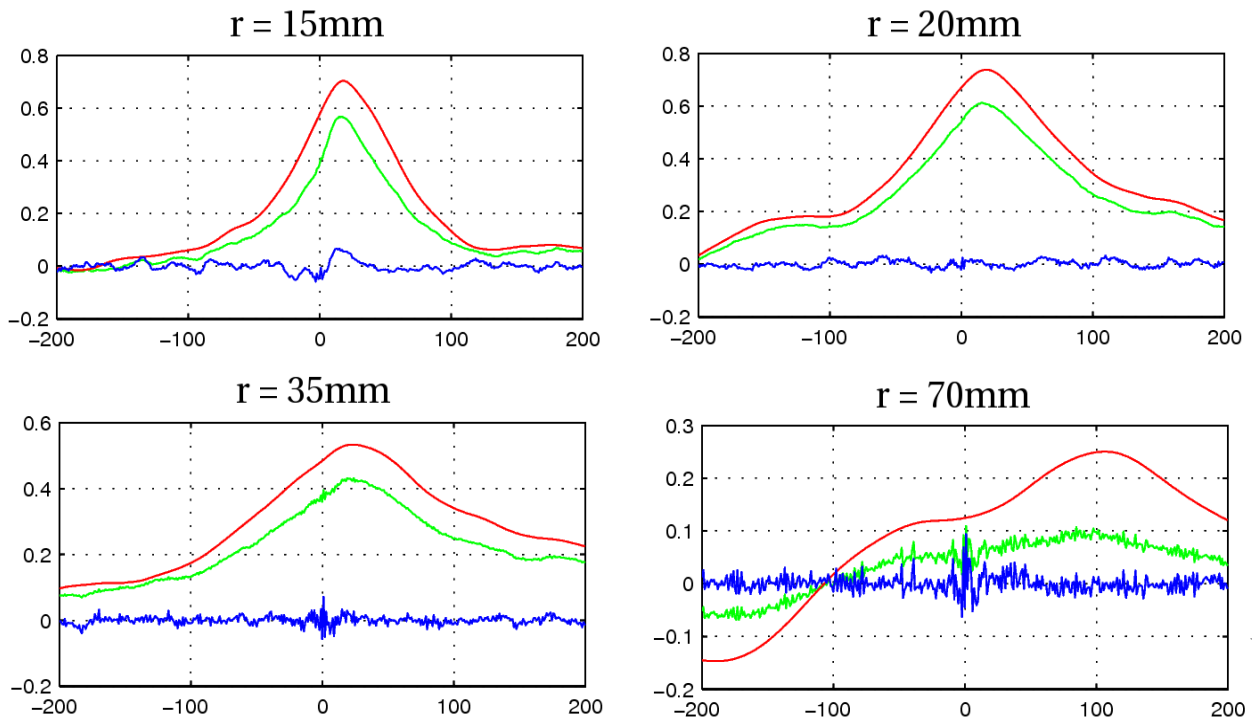


FIG. 13: Fonctions de corrélation croisée entre les signaux totaux et leurs parties cohérentes et incohérentes pour les courants de saturation  $I_{A1}$  et  $I_{A2}$  (algorithme standard)

Nous ne proposerons pas dans ce qui suit de réponse définitive à ces questions. Afin d'illustrer toutefois la problématique liée à la première question nous avons représenté sur les figures 13 et 14 les fonctions de corrélation entre  $I_{A1}$  et  $I_{A2}$  d'une part, et  $I_{A2}$  et  $I_{B2}$  d'autre part.

Pour le premier couple de sondes, on observe que la corrélation entre les parties cohérentes est supérieure à la corrélation entre les signaux totaux, tandis que la corrélation entre les parties incohérentes est proche de zéro et non significative. Ainsi, la réponse de l'algorithme standard à la question 1. pour ce couple de sondes est positive : l'intégralité de la corrélation entre  $I_{A2}$  et  $I_{B2}$  est dûe au passage de structures.

Pour les sondes  $I_{A2}$  et  $I_{B2}$  la situation est différente, puisqu'on observe un fort pic de corrélation entre les parties incohérentes. Pour  $r \leq 35mm$  le maximum de ce pic est même largement supérieur au maximum de la fonction de corrélation entre les parties cohérentes. Le cas le plus intéressant est  $r = 70mm$ . On remarque en effet sur la fonction de corrélation entre les signaux totaux un maximum très localisé en 0. Or ce petit pic n'apparaît plus sur la fonction de corrélation entre les parties cohérentes. Il semble qu'il soit au contraire dû aux contributions incohérentes. À ce stade, nous ne savons pas si ce pic est lié à l'appareil de mesure ou correspond à une propriété du plasma. Ainsi pour  $r = 70mm$ , une partie seulement des corrélations entre  $I_{A2}$  et  $I_{B2}$  est dûe au passage de structures, l'origine de l'autre partie restant indéterminée pour le moment.

### 3.3.3. Influence de la distance entre la sonde et la séparatrice

La figure 15 montre que la variance de la partie incohérente décroît systématiquement lorsque la distance  $r$  à la séparatrice augmentent. Les graphes en coordonnées log-log semblent montrer que cette quantité suit une loi d'échelle en fonction de  $r$ . En revanche, la variance de la partie incohérente exprimée en pourcentage de la variance totale ne varie pas de façon monotone en fonction de  $r$  (tableau 3).

### 3.3.4. Flux

Les quatre contributions au flux moyen sont données dans le tableau 4 en fonction de la distance à la séparatrice. On peut remarquer que le flux total diminue en valeur absolue lorsque la sonde s'éloigne du plasma, mais que son signe change entre  $r = 20mm$  et  $r = 35mm$ . Dans tous les cas, la contribution cohérente  $\Gamma^{cc}$  est de même signe et très proche du flux total. Par contre les contributions incohérentes sont presque toutes positives et faibles devant la valeur absolue du flux total.

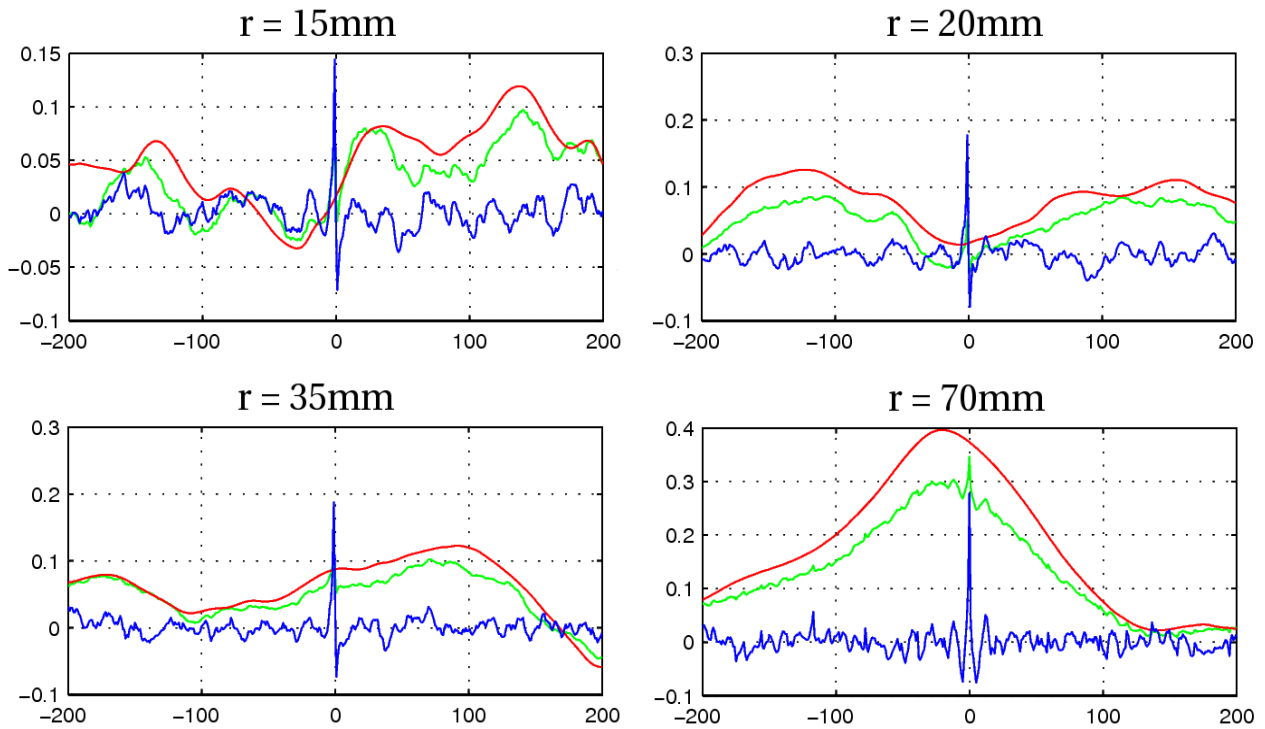


FIG. 14: Fonctions de corrélation croisée entre les signaux totaux et leurs parties cohérentes et incohérentes des courants de saturation  $I_{A2}$  et  $I_{B2}$  (algorithme standard)

Algorithme	$r = 15mm$	$r = 20mm$	$r = 35mm$	$r = 70mm$
Réel	36%	23%	26%	39%
Réel échelle par échelle	90%	74%	68%	68%
Complexe échelle par échelle	90%	49%	50%	48%

TAB. 3: Comparaison des pourcentages de la variance provenant de la contribution incohérente en fonction de la distance à la séparatrice, pour les trois algorithmes.

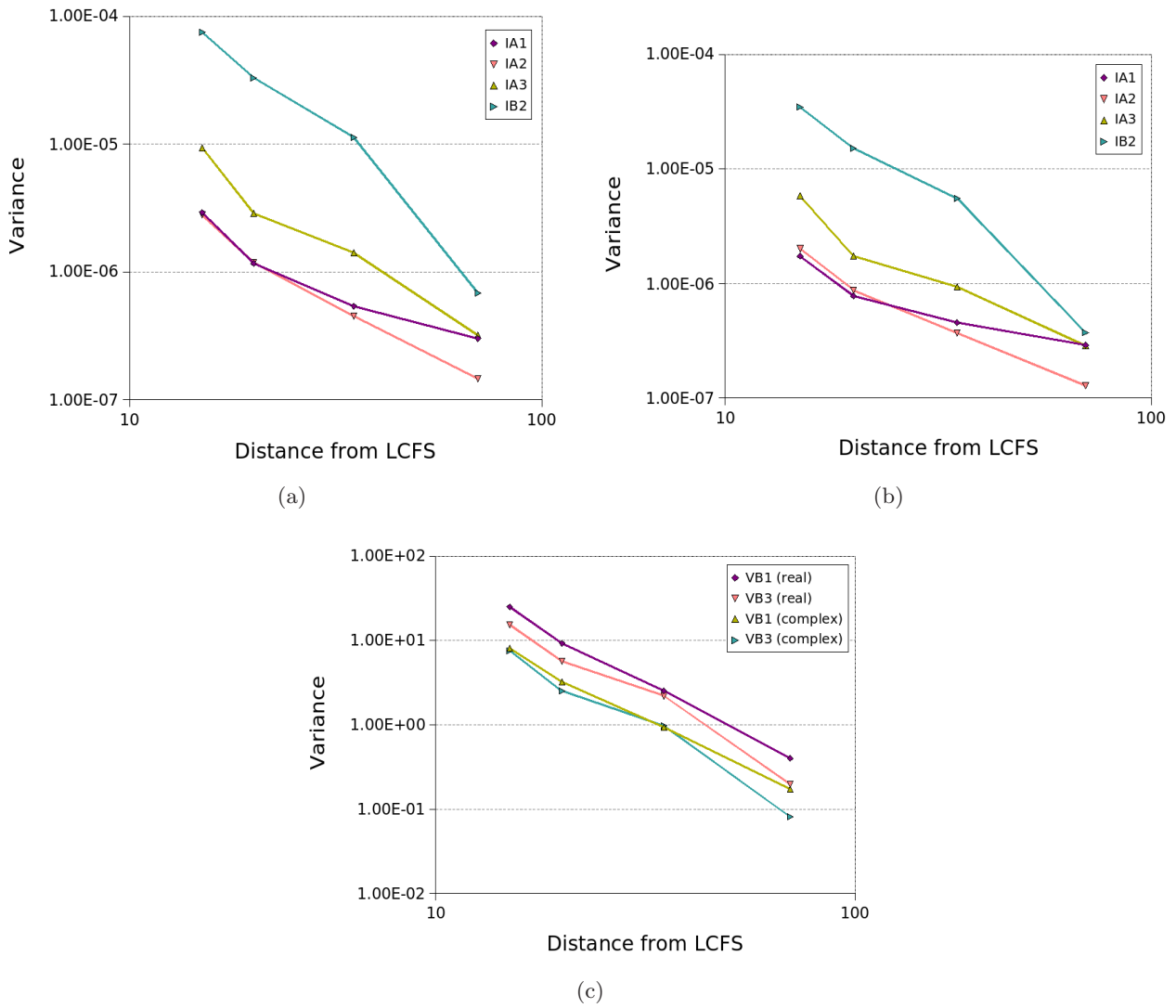


FIG. 15: Variance de la partie incohérente en fonction de la distance entre la sonde et la séparatrice. (a) Sondes d'intensité, seuillage global par ondelettes réelles. (b) Sondes d'intensité, seuillage global par ondelettes complexes. (c) Sondes de potentiel

	$\Gamma_{CC}$	$\Gamma_{CI}$	$\Gamma_{IC}$	$\Gamma_{II}$	$\Gamma$
$r = 15mm$	$3.10.10^{-1}$	$4.7.10^{-3}$	$3.18.10^{-3}$	$1.57.10^{-3}$	$3.2.10^{-1}$
$r = 20mm$	$3.24.10^{-2}$	$1.81.10^{-3}$	$1.94.10^{-3}$	$1.24.10^{-3}$	$3.74.10^{-2}$
$r = 35mm$	$-3.43.10^{-2}$	$4.16.10^{-4}$	$3.91.10^{-4}$	$8.90.10^{-5}$	$-3.34.10^{-2}$
$r = 70mm$	$5.67.10^{-3}$	$2.71.10^{-5}$	$-9.19.10^{-6}$	$3.5.10^{-6}$	$5.69.10^{-3}$

TAB. 4: Valeurs moyennes du flux et de ses quatre contributions pour la séparation en ondelettes réelles, en fonction de la distance à la séparatrice

En admettant que la quantité ainsi décomposée est une bonne estimation du flux radial de particules au voisinage de la sonde, nous avons donc montré que les structures cohérentes se propagent vers l'extérieur du tokamak.

### 3.3.5. Utilisation des ondelettes complexes de Kingsbury

Les ondelettes complexes permettent dans certains cas d'obtenir des parties incohérentes dont les corrélations croisées entre sondes sont plus réduites que pour l'algorithme utilisant les ondelettes orthogonales (figure 16). Cependant, leur principal avantage apparaîtra dans le cadre du seuillage échelle par échelle au paragraphe suivant. D'autre part le tableau 5 montre bien qu'en passant des ondelettes réelles orthogonales aux ondelettes complexes, le nombre de coefficients conservés par l'algorithme augmente d'un facteur de l'ordre de 4. Si l'on considère la taille mémoire occupée par la partie cohérente (dans la perspective d'une simulation numérique ou de la compression de gros jeux de données), il faut ajouter un facteur 2 parce que les coefficients sont des nombres complexes.

	$IB2$	$IB2$	$VB1$	$VB1$
	$r = 15mm$	$r = 70mm$	$r = 15mm$	$r = 70mm$
Ondelettes réelles orthogonales	1.2%	2.0%	1.0%	1.7%
Ondelettes complexes de Kingsbury	4.4%	4.8%	5.4%	5.0%

TAB. 5: Comparaison des taux de compression du nombre de coefficients

### 3.4. Extraction échelle par échelle

Comme nous l'avons montré sur les signaux académiques, le succès du seuillage échelle par échelle repose sur des contraintes plus fortes que celui du seuillage global. Or il semble que ces conditions ne soient pas remplies par les signaux de potentiel  $V_{B1}$  et  $V_{B3}$ . En observant qualitativement les résultats des extractions pour ces signaux, on observe en effet un comportement "tout ou rien" :

- soit le seuil est estimé uniquement pour l'échelle la plus fine et le résultat est proche de celui du seuillage global
- soit la quasi totalité des coefficients sont considérés comme incohérents

Dans la suite de ce paragraphe nous ne considérerons donc que les signaux d'intensité. Nous comparerons les résultats obtenus par les seuillages réel global, réel échelle par échelle et complexe échelle par échelle. Nous prendrons comme exemple la pinoche  $I_{A2}$  pour  $r = 70mm$ . Les séries temporelles correspondant aux différents algorithmes sont représentées sur la figure 17.

Dans la plupart des cas l'algorithme d'extraction échelle par échelle permet de mettre en évidence une partie incohérente dont les propriétés statistiques s'approchent de celles d'un bruit Gaussien corrélé. Pour le vérifier, on évalue la Gaussianité qualitativement au moyen des graphes quantile-quantile Gaussiens (figure 18). Bien que la partie incohérente contienne plus de coefficients d'ondelettes dans le cas du seuillage échelle par échelle (tableau 6 (a)), son facteur d'aplatissement est moindre que dans le cas du seuillage global. Ce résultat est surprenant au premier abord. En effet, on pourrait s'attendre à ce que la non-Gaussianité du signal de départ se répercute d'autant plus sur la partie incohérente que celle-ci contient de coefficients d'ondelettes. Mais ce serait oublier que, si la transformée en ondelettes conserve l'énergie, elle ne conserve pas les moments d'ordre 4.

Un argument un peu technique permet d'interpréter la réduction de l'aplatissement liée à l'extraction échelle par échelle : comme les ondelettes sont normalisées en énergie, ce sont celles des échelles les plus fines qui ont les extrema les plus grands : on dira qu'elles sont "piquées" tandis que les ondelettes

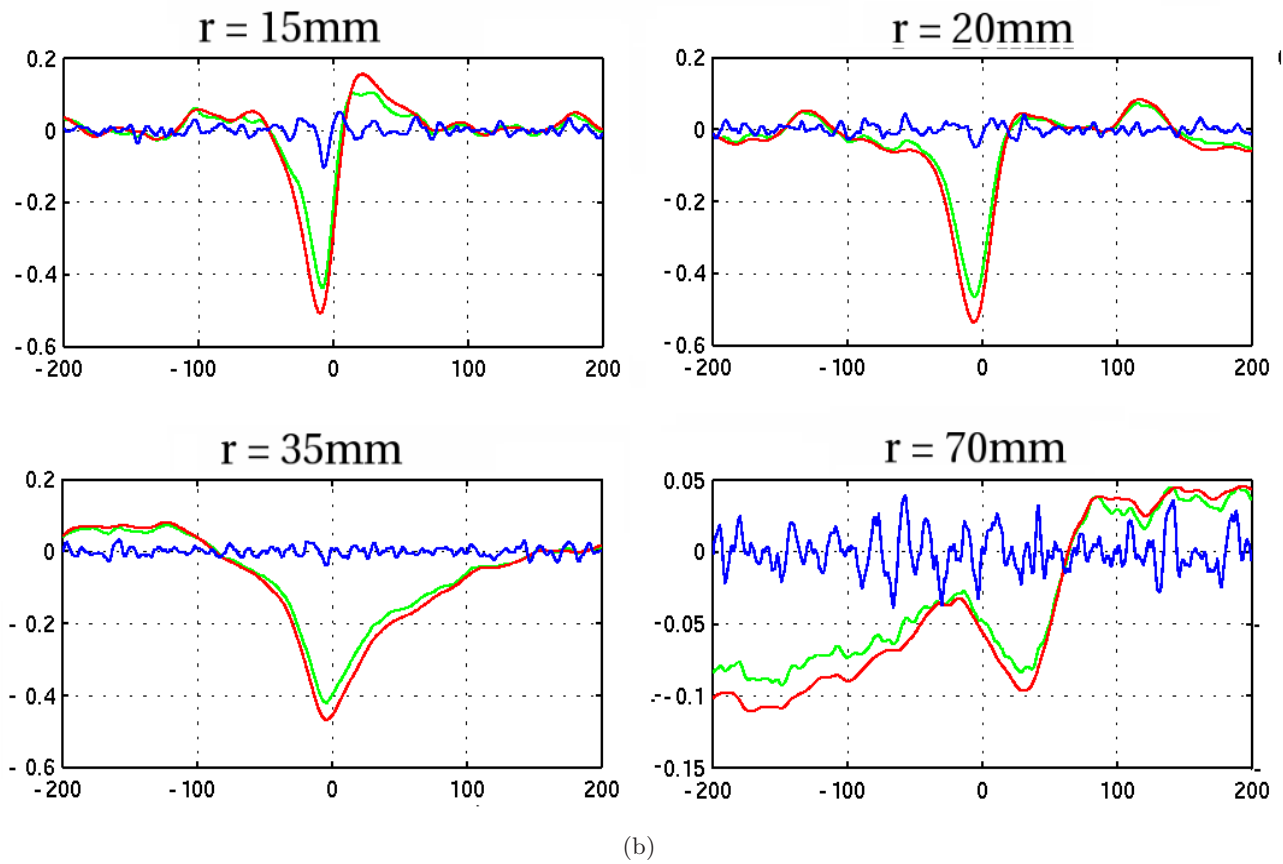
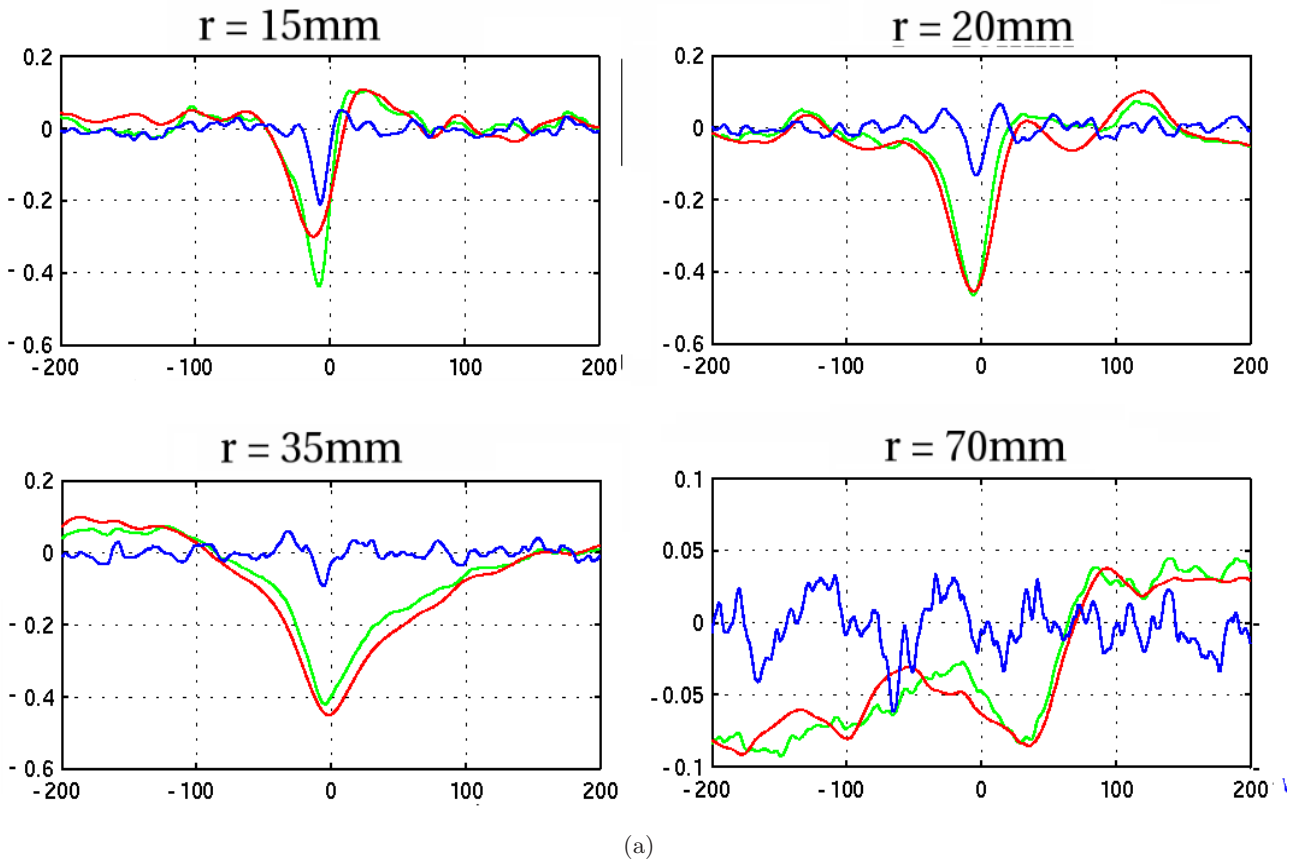


FIG. 16: Corrélations croisées entre les parties **cohérentes** et **incohérentes** du potentiel  $VB1$  et du courant de saturation  $IB2$ , (a) ondelettes réelles orthogonales (b) ondelettes complexes de Kingsbury. Les parties **incohérentes** associées à la transformée en ondelettes complexes présentent moins de corrélations croisées, probablement grâce à l'invariance par translation.

grossières sont plus "plates". Or les coefficients d'ondelettes aux échelles les plus fines sont en moyenne plus faibles que ceux des échelles grossières. Ainsi, l'algorithme échelle par échelle a tendance à mettre dans la partie incohérente plus de grands coefficients, associés à des ondelettes plates, et moins de petits coefficients, associés à des ondelettes piquées. Pour une variance donnée, la partie incohérente obtenue comportera donc moins de pics, d'où son facteur d'aplatissement plus faible.

#### 4. Conclusion

Notre objectif à l'abord de ce projet était d'analyser les structures cohérentes présentes dans des signaux expérimentaux de courant de saturation ionique, mesurés dans un plasma de bord du tokamak Tore Supra au CEA Cadarache. Nous avons pour cela analysé des signaux acquis lors de la campagne de mesure de 2005 par l'équipe de Tore Supra. Nous savons que des structures organisées spatialement, souvent appelées *blobs* ont été observées expérimentalement dans plusieurs tokamaks. Ces structures naissent probablement près de la dernière surface magnétique fermée, aussi appelée séparatrice, et se propagent ensuite à travers la *Scrape-Off Layer* (SOL) tout en interagissant avec les limiteurs. Cependant, nous ne disposons pas de modèle physique, ni pour décrire la propagation de ces structures, ni pour prévoir la trace qu'elles laisseraient sur la sonde de Langmuir, agissant comme un limiteur.

Nous avons repris et poursuivi le développement de modèles mathématiques permettant d'extraire les fluctuations cohérentes en séparant la partie cohérente d'un bruit résiduel supposé additif et Gaussien. Dans le cadre de ces modèles, notre méthode d'extraction des structures cohérentes nous a fourni deux types de résultats que nous allons maintenant présenter.

En premier lieu, nous avons obtenu la décomposition des signaux en partie cohérente et incohérente, que nous avons ensuite analysée au moyen de techniques statistiques classiques indépendantes des ondelettes.

Nous avons montré que les corrélations croisées entre les différentes pinoches de la sonde sont essentiellement dues aux contributions cohérentes. En particulier, nous avons trouvé que les structures cohérentes, à l'échelle spatiale correspondant à l'écartement entre deux pinoches, sont responsable de la majeure partie du flux radial de particules. La contribution incohérente au flux radial a par contre une amplitude très réduite, à savoir moins 15% du total. Pourtant, la variance de la partie incohérente de tous les signaux est supérieure à 20% de la variance totale. Ainsi, bien que la partie incohérente que nous avons extraite contribue aux fluctuations, elle ne contribue que très peu au transport.

Nous avons également analysé la partie incohérente afin de vérifier l'adéquation des signaux expérimentaux à chacun de nos modèles. Au terme de cette étude, nous concluons que si on souhaite maintenir l'hypothèse de Gaussianité de la partie incohérente il est nécessaire de prendre en compte les corrélations du bruit et l'extraction doit dans ce cas être réalisée séparément à chaque échelle des coefficients d'ondelettes.

En second lieu, nous avons calculé plusieurs quantités agrégées susceptibles de caractériser la décomposition en partie cohérente et incohérente. Pour les signaux que nous avons analysés, nous pensons que les deux diagnostics les plus porteurs d'information sont :

- la variance de la partie incohérente, pour laquelle nous avons pu mettre en évidence une décroissance en loi de puissance en fonction de la distance à la séparatrice.
- le pourcentage de coefficients d'ondelettes correspondants à la partie cohérente, qui nous permet d'estimer le taux de turbulence.

À l'avenir, la procédure d'analyse que nous avons développée permettra de traiter de façon systématique une grande quantité de signaux expérimentaux, par exemple ceux de la campagne 2007. Nous aurons deux objectifs :

- analyser des signaux acquis avec un échantillonnage beaucoup plus fin de la distance à la séparatrice, pour des paramètres du plasma fixés. Cela devrait permettre de caractériser les propriétés du plasma tout au long du profil de densité dans la SOL.
- explorer un plus grand nombre de régimes de turbulence de bord, obtenus expérimentalement par l'équipe de Tore Supra au CEA Cadarache en faisant varier les paramètres du plasma, afin comprendre le lien entre les propriétés des structures cohérentes observées dans la SOL et celles de la turbulence de bord.



	$r = 15mm$	$r = 20mm$	$r = 35mm$	$r = 70mm$
Réel	1.5	1.6	1.5	1.0
Réel échelle par échelle	0.4	0.7	0.5	0.2
Complexe échelle par échelle	1.1	2.5	1.9	0.9

(a)

	$r = 15mm$	$r = 20mm$	$r = 35mm$	$r = 70mm$
Signal total	4.5	3.9	3.4	4.0
Réel	4.3	5.0	5.2	3.6
Réel échelle par échelle	3.9	3.6	3.0	3.5
Complexe échelle par échelle	4.2	3.2	3.0	3.3

(b)

TAB. 6: Pour le signal  $IA2$  en fonction de l'algorithme utilisé et de la distance à la séparatrice : (a) pourcentage de coefficients d'ondelettes de la partie cohérente, (b) aplatissement du signal total et de sa partie incohérente.

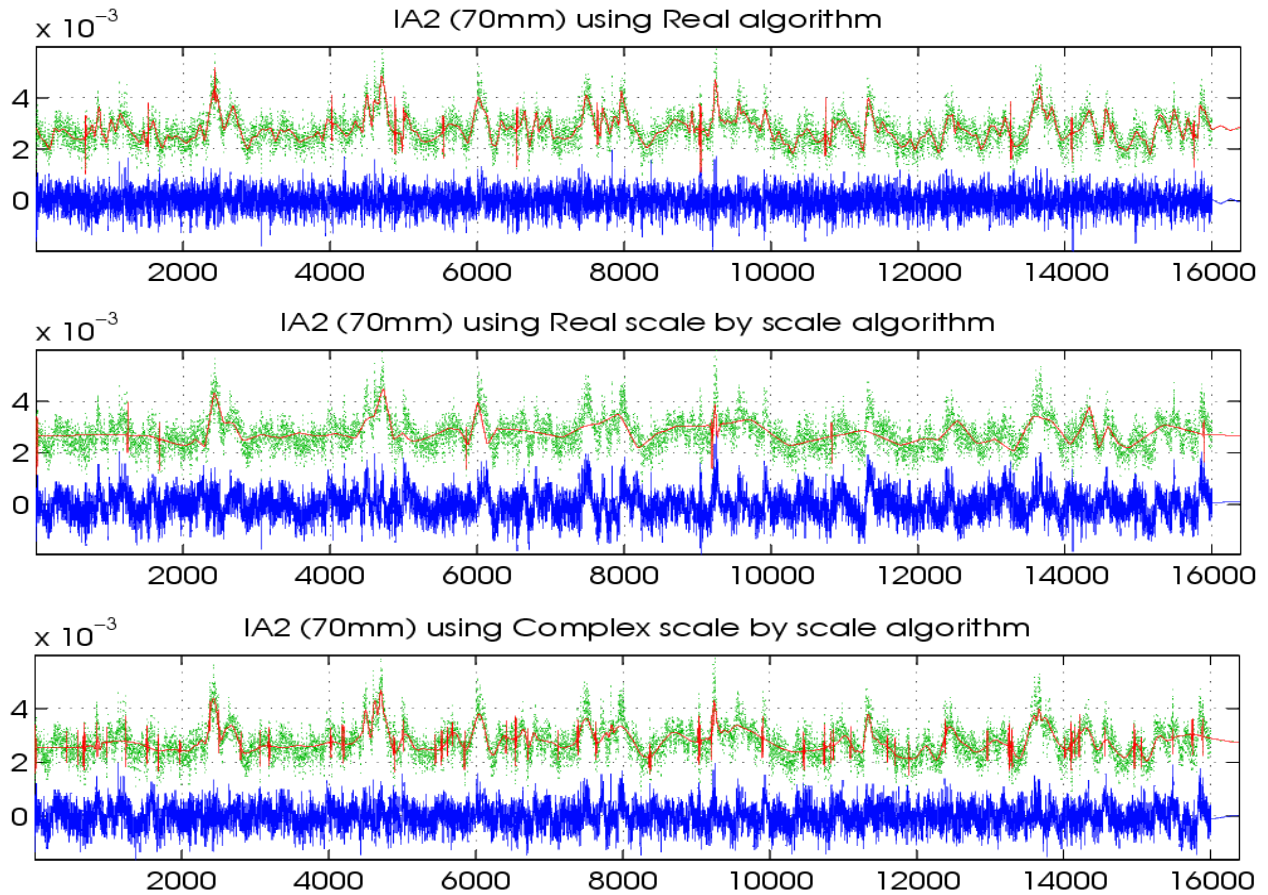


FIG. 17: Comparaison des résultats de l'extraction globale (en haut), de l'extraction échelle par échelle en ondelettes réelles (au milieu) et de l'extraction échelle par échelle en ondelettes complexes (en bas) pour le courant de saturation  $IA2$  à  $r = 70mm$ . Séries temporelles des parties cohérente, incohérente et du total.

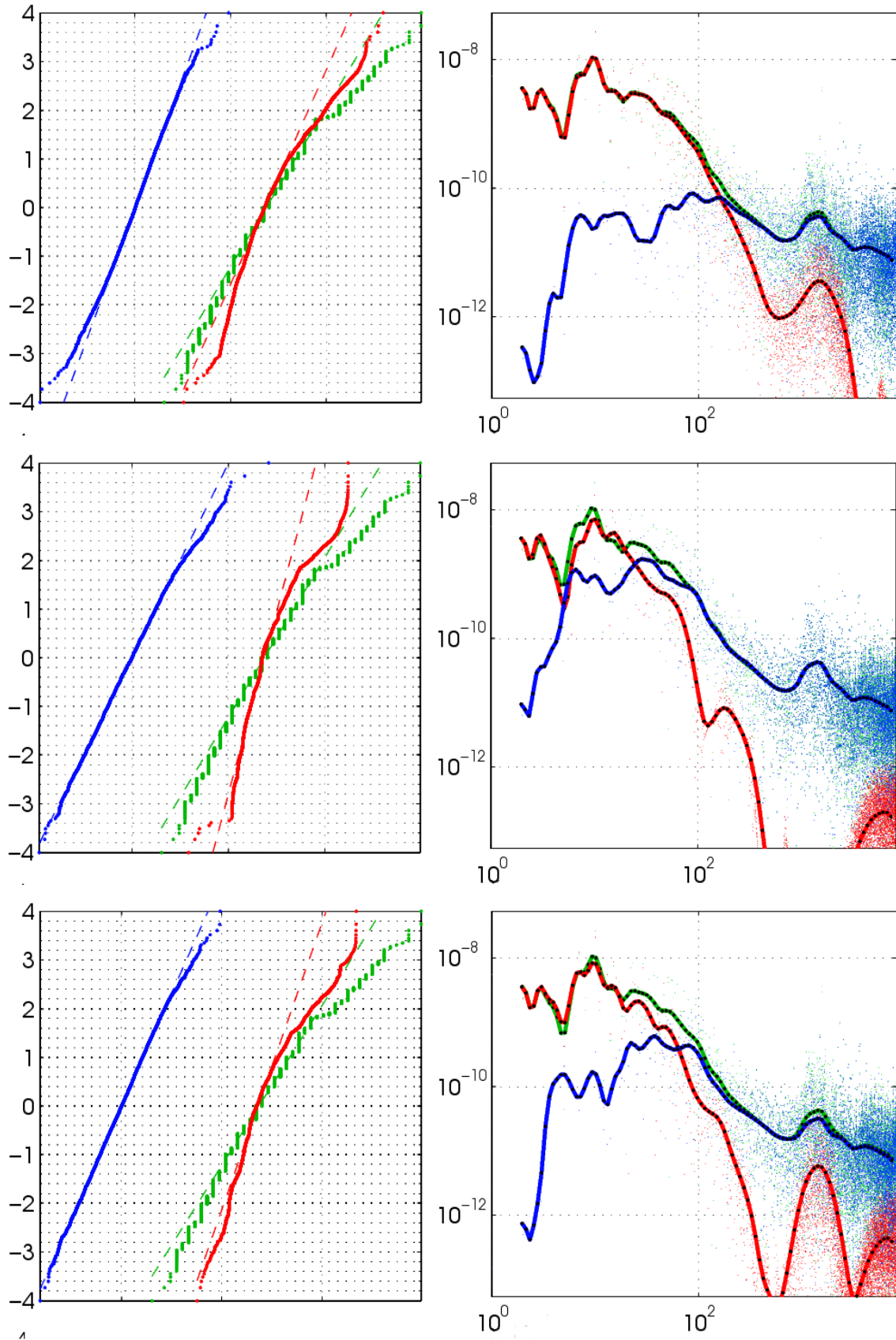


FIG. 18: Comparaison des résultats de l'extraction globale (en haut), de l'extraction échelle par échelle en ondelettes réelles (au milieu) et de l'extraction échelle par échelle en ondelettes complexes (en bas) pour le courant de saturation  $IA2$  à  $r = 70mm$ . Graphes quantile-quantile gaussiens (à gauche) et spectres en ondelettes de Morlet (à droite en lignes pleines).

Nous n'avons pas, à ce jour, pu mettre en évidence le lien entre les structures cohérentes présentes dans les séries temporelles que nous avons analysées et les structures spatialement organisées appelées *blobs*. Cela reste donc une direction de recherche importante. Nous pensons que l'analyse des images de la caméra rapide maintenant installée sur Tore Supra permettra de progresser dans cette direction.

## Remerciements

Nous remercions Nicolas Fedorczak, Pascale Monier-Garbet et Jamie Gunn de nous avoir donné accès à leurs signaux, et de leur aide dans leur interprétation. The authors acknowledge financial support from the Euratom-CEA association under contract V.3258.006. This work, supported by the European Communities under the contract of Association between EURATOM, CEA and the French Research Federation for Fusion Studies, was carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

## Références

- [1] P. Abry, P. Goncalves, and P. Flandrin. Wavelet-based spectral analysis of  $1/|e^{i\lambda f}|/e^{i\lambda}$  processes. In P. Goncalves, editor, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-93*, volume 3, pages 237–240 vol.3, 1993.
- [2] A. Azzalini, M. Farge, and K. Schneider. Nonlinear wavelet thresholding : a recursive method to determine the optimal denoising threshold. *Appl. Comput. Harmon. Anal.*, 18(2) :177–185, Oct. 2004.
- [3] M. S. Crouse and R. G. Baraniuk. Fast, Exact synthesis of Gaussian and nonGaussian Long-Range-Dependent Processes. *IEEE T. Inform. Theory.*, Aug. 1999.
- [4] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27(5), May 1986.
- [5] D. Donoho and I. Jonhstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81 :425–455, 1994.
- [6] M. Farge. Wavelet transforms and their applications to turbulence. *Ann. Rev. Fluid Mech.*, 24 :395–457, 1992.
- [7] M. Farge, E. Goirand, Y. Meyer, F. Pascal, and M. V. Wickerhauser. Improved predictability of two-dimensional turbulent flows using wavelet packet compression. *Fluid Dynam. Res.*, 10(4-6) :229 – 250, 1992.
- [8] M. Farge, K. Schneider, and P. Devynck. Extraction of coherent bursts from turbulent edge plasma in magnetic fusion devices using orthogonal wavelets. *Phys. Plasmas*, 13 :042304, 2006.
- [9] M. Farge, K. Schneider, and N. Kevlahan. Non-gaussianity and coherent vortex simulation for two-dimensional turbulence using an adaptative orthogonal wavelet basis. *Phys. Fluids*, 11(8) :2187–2201, 1999.
- [10] R. F. Fox, I. R. Gatland, R. Roy, and G. Vemuri. Fast, accurate algorithm for numerical simulation of exponentially correlated noise. *Physical review A*, 38(11) :5938, 1 Dec. 1988.
- [11] C. Hennig. Clusters, outliers, and regression : Fixed point clusters. *Journal of Multivariate Analysis*, 86 :183–212, 2003.
- [12] I. Johnstone and B. Silverman. Wavelet treshold estimators for data with correlated noise. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 59 :319–351, 1997.
- [13] N. Kingsbury. The dual-tree complex wavelet transform : A new technique for shift invariance and directional filters. In *Proc. 8th IEEE DSP Workshop*, 1998.
- [14] N. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Appl. Comput. Harm. Anal.*, 10(3) :234–253, May 2001.
- [15] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1999.
- [16] B. B. Mandelbrot and J. W. V. Ness. Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Rev.*, 10(4) :422–437, Oct. 1968.
- [17] R. Nguyen van yen, M. Farge, D. Kolomenskiy, K. Schneider, and N. Kingsbury. Wavelets meet Burgulence : CVS-filtered Burgers equation. *Physica D*, 237(14–17) :2151–2155, 2008.
- [18] B. W. Silverman. Wavelets in statistics : beyond the standard assumptions. *Phil. Trans. R. Soc. Lond. A*, 1999.
- [19] Y. Wang. Function estimation via wavelet shrinkage for long memory data. *Ann. Statist.*, 24(1) :466–484, 1996.