

Dynamique intermittente du plancton : analyse de la dynamique multi-échelle en utilisant la décomposition modale empirique

François G. Schmitt¹, Orlane Anneville² & Sami Souissi¹

¹ CNRS, Laboratoire d’Océanologie et de Géosciences, UMR 8187, F-62930 Wimereux

² INRA, UMR CARTELE, F-74203 Thonon les Bains

francois.schmitt@univ-lille1.fr; www.fg-schmitt.fr

Résumé. Les écosystèmes sont des systèmes complexes et leur variabilité dans le temps et dans l’espace, est caractérisée par de fortes fluctuations à toutes les échelles. Ceci est vérifié en particulier pour l’abondance de populations, où l’on peut rencontrer typiquement des fluctuations très importantes. Des méthodes d’analyse adéquates doivent être trouvées pour analyser et caractériser ces fluctuations. Dans cet article méthodologique, appliqué à des données d’abondance, nous considérons deux aspects. En premier lieu nous considérons la densité de probabilité de l’abondance. Nous considérons aussi la dynamique à de multiples échelles de cette série extrêmement variable, et la caractérisons dans le cadre de la décomposition modale empirique (EMD), pour mettre en évidence certaines échelles spécifiques. La décomposition modale empirique est une nouvelle technique d’analyse de séries temporelles, particulièrement mise en avant pour les séries non-linéaires et non stationnaires. Cette approche peut être appliquée à des séries même relativement courtes, ayant un échantillonnage irrégulier. L’approche EMD est appliquée ici au logarithme des données d’abondance puis exponentiée.

Les analyses sont appliquées à une série d’abondance du phytoplancton provenant du lac Léman. Nous montrons qu’en utilisant quelques modes, les données complexes, explosives, peuvent être progressivement reconstruites. Chaque mode a une échelle caractéristique, et cette approche peut également être utilisée pour identifier ou enlever une tendance, ou pour lisser des données. Il s’agit ici de la première application de la méthode EMD à des séries temporelles en biodiversité et en écologie aquatique.

Abstract. Ecosystems are complex systems and ecosystem variability are characterized, in time and space, by large fluctuations at all scales. This is especially true of abundance studies, where one typically finds huge fluctuations. New methods are needed to analyze and characterize these fluctuations. Here this methodological paper applied to a biodiversity situation, considers two aspects. First, we consider extreme abundance probabilities. We also consider the multiscale dynamics of such highly variable series and characterize it using empirical mode decomposition (EMD) to find specific scales of variation. Empirical mode decomposition is a new time series analysis, especially useful for nonlinear and non stationary data. It can be applied even to short time series, with irregular sampling. The EMD approach is applied here on the logarithm of the time series and then exponentiated.

The analyses are performed on a phytoplankton abundance time series from Lake Geneva. We show that, using a few modes, the complex and bursty abundance series is reconstructed. Each mode has a characteristic scale, and this approach can also be used for detrending, or for smoothing applications. This is the first application of the EMD method to biodiversity and aquatic ecology abundance time series.

1 Introduction

Les écosystèmes, avec leurs forçages multiples, à de nombreuses échelles, et des interactions entre espèces et entre individus, sont des exemple-types de systèmes complexes avec de fortes non-linéarités, dans l’espace et dans le temps. Dans ce domaine, la forte variabilité constatée est souvent considérée comme étant une conséquence de forçages physiques ou anthropiques sur la biosphère. D’autres études ont plutôt insisté sur le fait que les écosystèmes, en tant que systèmes complexes, pouvaient s’auto-organiser et générer, de façon intrinsèque (sans forcément faire appel à un forçage extérieur mais celui-ci étant possible), des fortes fluctuations spatio-temporelles au niveau d’une espèce ou au niveau de la biodiversité des communautés [1, 2]. De nombreuses études s’intéressent à décrire et expliquer cette forte variabilité dans la dynamique de population et l’abondance des espèces [3, 4] en considérant en

particulier la dynamique des extrêmes [5–7]. Ici, nous considérons une série temporelle d’abondance du phytoplancton, recueillie en point fixe au lac Léman, et nous en considérons une modélisation à l’aide de la méthode de décomposition modale empirique, appliquée au logarithme de l’abondance.

2 Présentation des données et leur densité de probabilité

Le Léman est un grand et profond lac péri-alpin (surface 580 km², volume 89 km³, profondeur maximale 309 m, profondeur moyenne 152 m) situé à une altitude de 372 m à la frontière entre la France et la Suisse. Un suivi écologique du lac a été mis en place depuis le milieu des années 1970. Ceci inclut des paramètres physiques, chimiques et biologiques, mesurées chaque mois de 1974 à 1980, puis tous les 15 jours pendant le printemps, l’été et l’automne. L’échantillonnage a été réalisé dans une station située dans le bassin principal, en un point de profondeur maximale. Des échantillons d’eau sont prélevés sur les 10 premiers mètres à l’aide d’une bouteille intégratrice, et servent ensuite à l’identification et au dénombrement en laboratoire des différentes espèces phytoplanctoniques (plancton végétal) pour en estimer l’abondance. L’identification du phytoplancton et le comptage des cellules sont effectués dans une chambre de sédimentation utilisant un microscope inversé [8] selon une méthode standardisée [9]. Le comptage des cellules a été ensuite transformé en des données de biovolume (en $\mu\text{m}^3/\text{mL}$), à l’aide d’une estimation géométrique pour chaque espèce.

Nous avons analysé ici un sous-ensemble (1974–2000) de la base de données disponible, en nous focalisant sur une période d’échantillonnage régulier, avec une méthode identique et un point fixe. Nous avons une base de données comportant 491 différents échantillons mensuels, avec 184 différentes espèces de phytoplancton recensées. La figure 1a représente une vue en perspective de l’ensemble de la base de données : en abscisse les espèces et en ordonnées le temps. On visualise ainsi la très forte variabilité de l’abondance des espèces, et la complexité de la caractérisation et de la compréhension de cette dynamique. De plus, il existe typiquement pour ce type de base de données un grand nombre de zéros : ici, 90%.

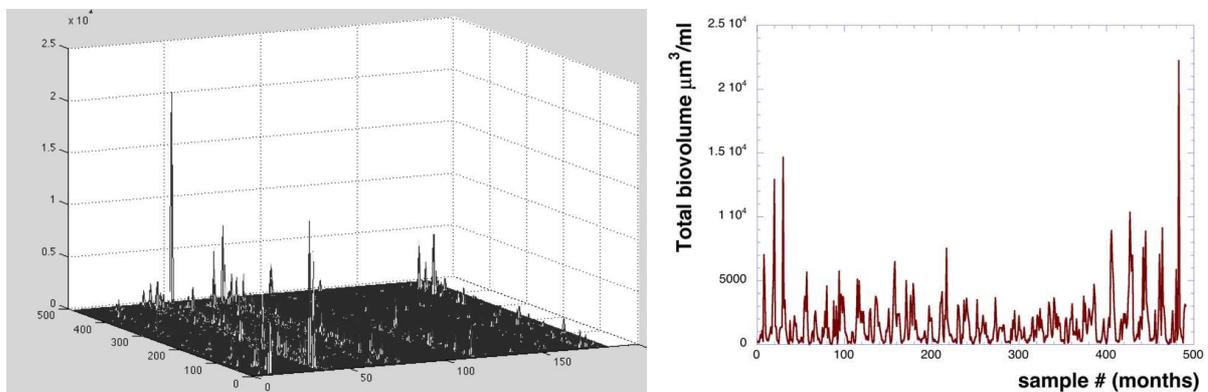


Figure 1. (a) : À gauche, une vue en perspective de l’ensemble de la base de données : en abscisse les espèces et en ordonnées le temps (biovolume en $\mu\text{m}^3/\text{mL}$). (b) : À droite, le biovolume total, c’est-à-dire la projection de la matrice de la figure 1a sur l’axe des y .

Comme dans la suite nous considérons le logarithme des données, pour éviter d’avoir des valeurs nulles, et pour simplifier l’analyse, nous considérons le biovolume total, c’est-à-dire la projection de la matrice de la figure 1a sur l’axe des y . Ceci est représenté en figure 1b. Dans la suite, c’est cette dernière série temporelle qui est analysée. Le logarithme des données est représenté en figure 2a ; on constate que l’allure des données est plus stationnaire. Leur densité de probabilité est représentée en figure 2b, avec un ajustement lognormal. Comme souvent proposé en écologie, l’abondance est ici approximativement lognormale, aussi bien pour les faibles valeurs que pour les extrêmes. Un test statistique est négatif,

montrant que les données sont pas strictement lognormales, mais leur écart à la lognormalité est ici visuellement assez faible. Ceci est en accord avec d'autres travaux en écologie [3, 6].

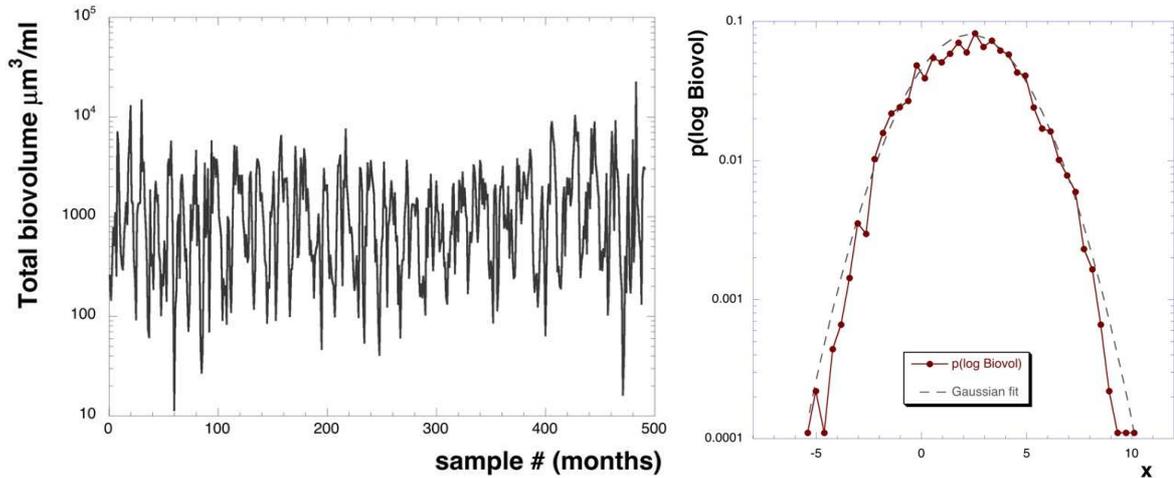


Figure 2. (a) : À gauche, le logarithme des données, montrant une allure plus stationnaire. (b) : À droite, la densité de probabilité des données avec un ajustement normal (donc une pdf lognormale pour les données d'origine).

3 Décomposition modale empirique

Norden Huang a proposé en 1998 une méthode intitulée décomposition modale empirique (EMD, pour *Empirical Mode Decomposition*), pour analyser des séries temporelles non-linéaires et non stationnaires [10]. Cette méthode permet de décomposer une série temporelle quelconque en une somme de plusieurs séries temporelles, appelées « modes » (IMF, pour *Intrinsic Mode Function*). Chaque mode possède une fréquence caractéristique. Pour un bruit blanc ou un bruit Gaussien fractionnaire, la répartition de ces fréquences caractéristiques est telle que la décomposition correspond à un banc de filtre dyadique [11, 12]. Cette nouvelle méthode d'analyse a rencontré un grand succès dans de nombreux domaines des sciences de la nature et de la société, avec plus de 2000 citations de l'article original. Elle peut être utilisée pour étudier par exemple l'intermittence invariante d'échelle en turbulence [13–15].

L'idée de la décomposition est de considérer un signal en tant que superposition de différentes oscillations monocomposantes, appelées fonctions modales intrinsèques. Pour cela chaque mode doit satisfaire à deux conditions : (i) la différence entre le nombre de minima locaux et le nombre de traversée du zéro doit être zéro ou un ; (ii) la moyenne glissante de la somme de l'enveloppe définie par les maxima locaux et l'enveloppe définie par les minima locaux est nulle. La décomposition est faite de façon itérative, en considérant à chaque étape le signal comme étant la somme d'une composante basse fréquence et d'une composante haute fréquence. A la fin de la décomposition, la méthode exprime une série temporelle $x(t)$ comme la somme d'un nombre fini de modes $C_i(t)$ et un résidu final $r_n(t)$ [10] :

$$x(t) = \sum_{i=1}^N C_i(t) + r_n(t) \quad (1)$$

Le nombre de modes nécessaires pour une série temporelle de longueur N est de l'ordre de $\log_2 N$.

Pour une série temporelle strictement positive comme celle qui nous occupe ici, la méthode EMD ne semble pas fonctionner de façon réaliste puisque chaque mode oscille autour de 0 et aura donc des

valeurs négatives. Pour cette raison, nous avons ici plutôt appliqué la méthode EMD au log-abondance (figure 2a). Nous avons ainsi obtenu 7 modes, chacun ayant une échelle caractéristique de plus en plus grande, et un résidu. Dans notre cas, la série temporelle d'origine sera donc obtenue comme le produit de l'exponentielle des modes successifs. La figure 3 représente les six premiers modes obtenus à partir de l'analyse EMD de la série des log-abondances, ou plus précisément l'exponentielle des modes successifs. La série à modéliser est donc obtenue à partir du produit des séries temporelles représentées en Figure 3.

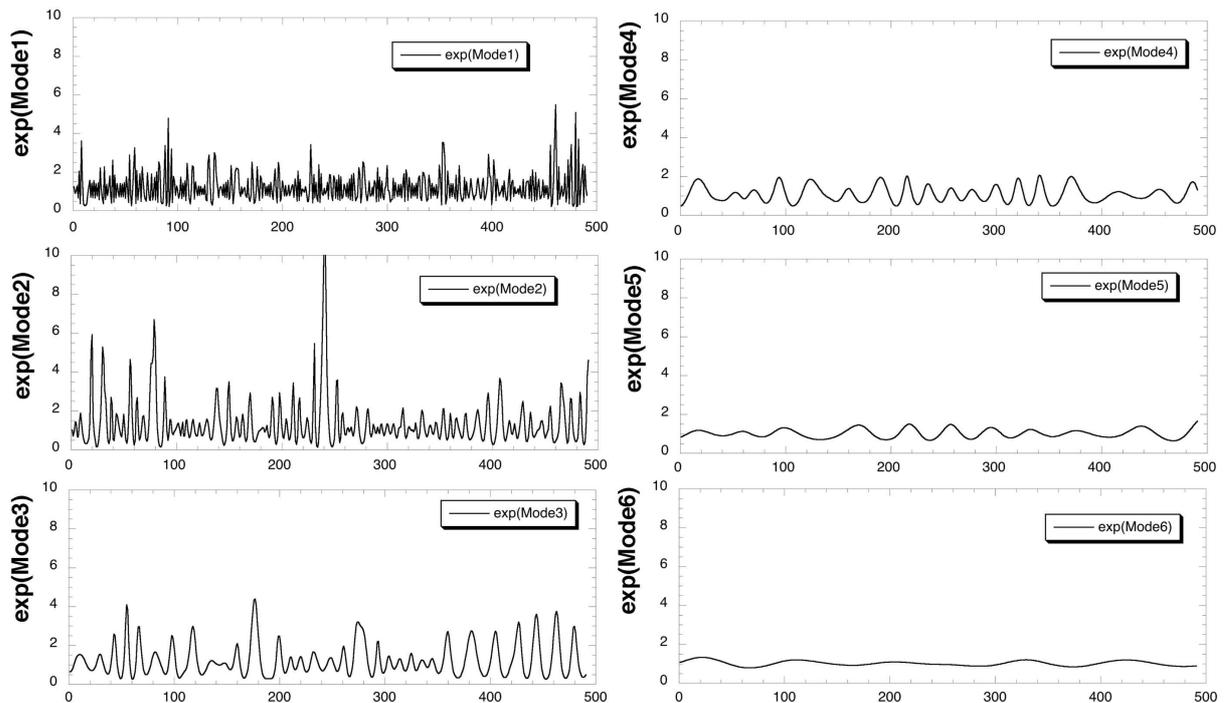


Figure 3. Représentation de 6 modes obtenus à partir de l'analyse EMD de la série des log-abondance. Ici on représente l'exponentielle des modes. La série résultante est donc modélisée par le produit des différentes séries représentées ici.

Cette approche permet de reconstruire la dynamique très fluctuante de la série d'abondance, à l'aide d'un processus stochastique lognormal comportant un nombre réduit de modes. La figure 4 représente la reconstruction du log-biovolume à l'aide du résidu et des P premiers modes ($P = 2$ à 5 de haut en bas), comparée aux données expérimentales. On constate que les deux premiers modes ne capturent pas toute la dynamique, mais qu'en augmentant progressivement le nombre de modes, la superposition devient bien meilleure, et finalement les 4 premiers modes permettent déjà une très bonne superposition des données. Ceci fournit une modélisation empirique de la dynamique de l'abondance en écologie, permettant de représenter ces fortes et intermittentes fluctuations comme une cascade multiplicative, obtenue à partir de la multiplication d'un nombre assez limité de modes, vus comme autant d'instabilités. Une continuation possible de ces travaux pourra être de considérer différents niveaux trophiques (par exemple ajouter le zooplancton) et les conditions environnementales (température, teneur en nutriments...) et d'analyser leurs corrélations mode à mode.

Nous avons ici appliqué directement la méthode EMD au logarithme de nos données d'origine, permettant d'utiliser cette approche pour des données positives (mais non nulles), et ensuite reconstruisant la série d'origine de façon multiplicative. Il s'agit ici à notre connaissance de la première application de la méthode EMD en écologie et en biodiversité.

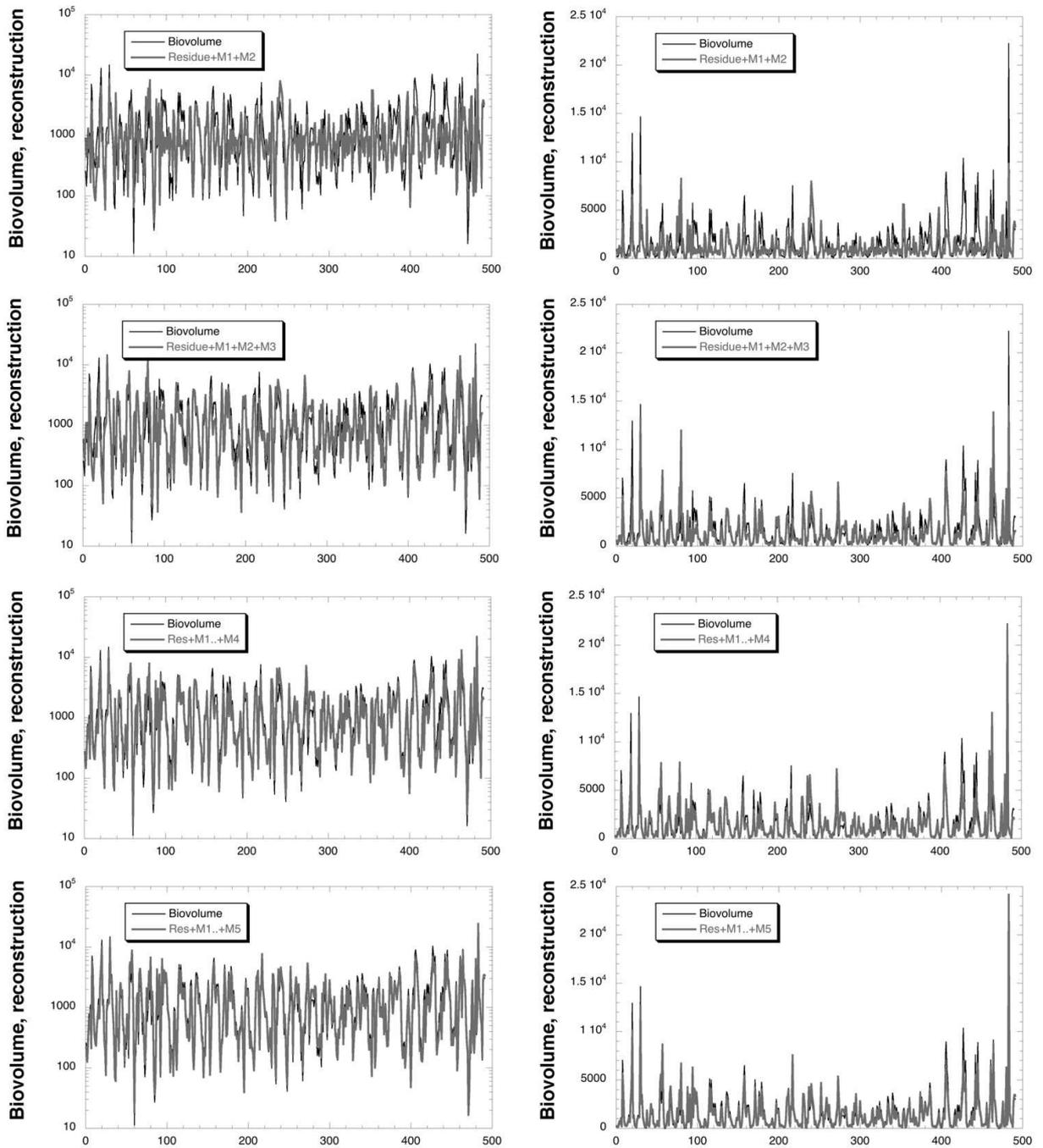


Figure 4. Colonnes de gauche : reconstruction du log-biovolume à l'aide du résidu et des P premiers modes ($P = 2$ à 5 de haut en bas), comparée aux données expérimentales. Colonnes de droite : la même reconstruction pour le biovolume lui-même.

Remerciements. Les données sont issues de la base de données SOERE-INRA de Thonon-les-bains et ont été obtenues dans le cadre du suivi écologique du Léman dirigé et financé par la CIPEL (<http://www.cipel.org/>). Nous remercions toutes les personnes de l'UMR-CARRETEL ayant travaillé à la collecte et au dépouillage des échantillons. Le code EMD en Matlab utilisé ici a été écrit par Dr Gabriel Rilling et Pr. Patrick Flandrin du laboratoire de Physique, CNRS & Ens Lyon, et est disponible ici : <http://perso.ens-lyon.fr/patrick.flandrin/emd.html>. Il a été adapté par Yongxiang Huang de l'Université de Shanghai. Cette étude fait partie du projet Bio-Asie FASCICLE (<http://fascicle.cnrs.fr/accueil.html>).

Références

1. R. V. SOLE & J. BASCOMPTE, *Self-organization in complex ecosystems*, Princeton University Press, Princeton (2006).
2. D. STORCH, P. A. MARQUET & J. H. BROWN, *Scaling biodiversity*, Cambridge University Press, Cambridge (2007).
3. S. ENGEN & R. LANDE, Population dynamic models generating the lognormal species abundance distribution, *Mathematical biosciences*, **132**, 169–183 (1996).
4. B. J. MCGILL *et al.*, Species abundance distributions : moving beyond single prediction theories to integration within an ecological framework, *Ecology Letters*, **10**, 995–1015 (2007).
5. A. BLARER & M. DOEBELI, Resonance effects and outbreaks in ecological time series, *Ecology Letters*, **2**, 167–177 (1999).
6. J. HALLEY & P. INCHAUSTI, Lognormality in ecological time series, *OIKOS*, **99**, 518–530 (2002).
7. J. D. REYNOLDS AND R. P. FRECKLETON, Population dynamics : growing to extremes, *Science*, **309**, 567–568 (2005).
8. H. UTERMÖHL, Zur Vervollkommung der quantitative Phytoplankton Methodik, *Mitt. Internat. Ver. Theor. Angew. Limnol.*, **9**, 1–38 (1958).
9. EN 15204, *Water Quality - Guidance Standard on the Enumeration of Phytoplankton using Inverted Microscopy (Utermohl Technique)*, European Committee for Standardization, Brussels (2006).
10. N. E. HUANG *et al.*, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. London*, **454**, 903–995 (1998).
11. P. FLANDRIN, G. RILLING & P. GONÇALVÈS, Empirical mode decomposition as a filter bank, *Signal Processing Letters, IEEE*, **11**, 112–114 (2004).
12. Z. WU & N. E. HUANG, A study of the characteristics of white noise using the empirical mode decomposition method, *Proc. R. Soc. London*, **460**, 1597–1611 (2004).
13. Y. HUANG, F. G. SCHMITT, Z. LU, Z. & Y. LIU, An amplitude-frequency study of turbulent scaling intermittency using Hilbert spectral analysis, *Europhys. Lett.*, **84**, 40010 (2008).
14. F. G. SCHMITT, Y. HUANG, Z. LU, Y. LIU & N. FERNANDEZ, Analysis of turbulent fluctuations and their intermittency properties in the surf zone using empirical mode decomposition, *Journal of Marine Systems*, **77**, 473–481 (2009).
15. Y. HUANG, F. G. SCHMITT, Z. LU, Z. & Y. LIU, Arbitrary order Hilbert spectral analysis for time series possessing scaling statistics : a comparison study with detrended fluctuation analysis and wavelet leaders, *Phys. Rev. E*, **84**, 016208 (2011).