HAL
open science

On optimal wavelet bases for the realization of microcanonical cascade processes

Oriol Pont, Antonio Turiel, Conrad J. Perez-Vicente

To cite this version:

Oriol Pont, Antonio Turiel, Conrad J. Perez-Vicente. On optimal wavelet bases for the realization of microcanonical cascade processes. International Journal of Wavelets, Multiresolution and Information Processing, 2011, 9 (1), pp.35-61. 10.1142/S0219691311003943 . inria-00582372

HAL Id: inria-00582372

https://inria.hal.science/inria-00582372v1

Submitted on 1 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On optimal wavelet bases for the realization of microcanonical cascade processes

ORIOL PONT
INRIA Bordeaux Sud-Ouest. 351 Cours de la Libération.
33405 Talence, France
Tel. +33 524574109
oriol.pont@inria.fr oriol@ffn.ub.es

ANTONIO TURIEL
Institut de Ciències del Mar - CSIC. Passeig Marítim de la Barceloneta, 37-49.
08003 Barcelona, Catalunya, Spain
turiel@icm.csic.es

CONRAD J. PÉREZ-VICENTE
Departament de Física Fonamental. Universitat de Barcelona. Diagonal, 647.
08028 Barcelona, Catalunya, Spain
conrad@ffn.ub.es

Received 23 june 2009
Revised 10 October 2010

Multiplicative cascades are often used to represent the structure of turbulence. Under the action of a multiplicative cascade, the relevant variables of the system can be understood as the result of a successive transfer of information in cascade from large to small scales. However, to make this cascade transfer explicit (i.e, being able to decompose each variable as the product of larger scale contributions) is only achieved when signals are represented in an optimal wavelet basis. Finding such a basis is a data-demanding, highly-complex task. In this paper we propose a formalism that allows to find the optimal wavelet of a signal in an efficient, little data-demanding way. We confirm the appropriateness of this approach by analyzing the results on synthetic signals constructed with prescribed optimal bases. We show the validity of our approach constrained to given families of wavelets, though it can be generalized for a continuous unconstrained search scheme.

Keywords: optimal wavelets; multiplicative cascades; multifractals; multiscale signal processing

AMS Subject Classification: 28A80, 42C40, 60G18, 82C80, 46N55

1. Introduction

Multiplicative processes giving rise to cascades are quite ubiquitous in Nature. Either as a real mechanism or as an effective one, cascades spontaneously develop in many scale-free systems. For instance, in a three-dimensional flow under fully

developed turbulence, energy is transferred from large to small scales (where it is utterly dissipated) through a well-defined cascade process. In other systems with cascade structure, the transfer of information between different scales gives rise to a hierarchical arrangement.[14] In particular the distribution of scaling exponents can be determined from the study of the statistical properties of the cascade process.[9,52,29] By mean of appropriate techniques taking cascades into account information about the structure of the system can be gained; examples include cloud organization,[33,35] rain distribution,[36,44] phytoplankton patchiness,[18] vegetation patterns,[37] traffic distribution,[32] ocean circulation[46,53] and even stock market analysis.[23,25,26,38,50,51]

Apart from some models,[36,3,17,7,19,8,16] there have been few attempts to characterize the structure of particular signals in terms of local cascade descriptors. The advantage of such an approach is that one can extract geometrical information about the signal in contrast to standard methods where only global statistical information is available. Given a signal (or dataset), the key point is to find a representation basis where the cascade process can be expressed in a *microcanonical* form, in other words, to find an appropriate transformation in which the representation variables are precisely these local cascade descriptors.

Wavelets are a standard analysis tool in signal processing[21]: wavelets separate the relevant details of a signal at different scale levels, and hence they are appropriate to analyze the multiscale behavior of cascade processes and to represent them. Most of the standard wavelets are able to accurately estimate the distribution of energy (or equivalent quantity) at each stage of the cascade, something that is very useful as a global descriptor. In addition, for a given system, there is a particular wavelet called *optimal wavelet* that also characterizes the dynamics at a local level, as it corresponds to the proper representation basis for cascade variables. The main advantage of optimal wavelet projections is that they can be expressed as products of successive cascade variables chosen along a branch of a dyadic tree. This representation is minimally redundant, as cascade variables are independent between consecutive cascade stages, and it defines a local effective dynamics that opens the way to new theoretical developments and practical applications such as improved multifractal models,[39] data compression,[4] reconstruction of data gaps[31] and time-series forecast.[30]

An attempt to find the optimal wavelet of natural images from a sample dataset has been reported in Ref. 49. The methodologies presented there are quite limited, as they exploit particular symmetries of natural images, and the uncertainty in the so-obtained empirical optimal wavelet is rather large to allow fine developments. In this paper we will present an improvement of the methodology presented in Ref. 49 in order to derive the optimal wavelet of more general types of data with more precision.

The paper has the following arrangement: The next section explains the concept of multiplicative cascade and how it is identified in real signals. In section 3

we mathematically formalize canonical and microcanonical cascades through the use of wavelet projections, and we also introduce the concept of optimal wavelet. In section 4 we introduce a quantifier of the optimality degree and discuss about optimization strategies. Then, we generate synthetic cascades and check their optimality, showing the results in section 5. Finally, in section 6 we give our conclusions.

2. Persistence in scale invariant signals

Multiplicative cascades are present in many different systems, but they are not usually recognized as such. Usually, their presence is reported by means of indirect evidence about its effects on the properties of signals. One of the most commonly reported effect of multiplicative cascades is the persistence of feature detection across scales. The importance of persistence is that the detection of a feature at a coarse scale allows to infer the presence of the same feature at finer scales. This phenomenon is well known since the introduction of wavelet representation of signals, and it is first described by Mallat and co-workers.[20,22]

Feature persistence is a strong, relevant property of physical signals, as it implies that the signal is highly redundant. It is precisely by means of the wavelet representation that this redundancy becomes evident. Persistence implies that we can predict to some extent what is going to happen at the next resolution level from the wavelet coefficients of a given level. Some authors[4,55] have exploited this redundancy to devise algorithms for image compression. Particularly, Simoncelli and co-workers have noticed that the mutual dependence between consecutive scales can be better highlighted using conditional histograms.[15,4,55,40] The histograms of fine-scale (also called "child") coefficients conditioned by the value of the coarse-scale (also called "parent") coefficient at the same location have a clear tie-bow shape for any wavelet[4,40] (we also observe the same behavior in Figure 1 top). This shape implies that the dispersion of the child increases with the absolute value of the parent coefficient. This suggests that the child coefficient depends on its parent coefficient in a multiplicative fashion. For that reason, the distribution of the logarithm of the child coefficient conditioned by a value of the logarithm of the parent coefficient exhibits a linear dependence[4,15] (see also our Figure 1 bottom). The authors found that, depending on the wavelet, the range of validity of this linear dependence can be larger or smaller.

More recently, Pottier et al.[31] studied satellite images of surface chlorophyll concentration and found them to be persistent across scales. Although they used very different wavelet bases, for none of them the histogram of the logarithm of the child conditioned by the logarithm of the parent have a full linear range. Pottier et al. proposed a particular model to describe the child-parent dependency, similar to the one introduced in Ref. 55 and valid for many different wavelets that are not the optimal one but are not too far from it anyway. We will call this model the linear model, and it reads as:

$$\alpha_{\mathrm{C}} = \eta_0 \, \alpha_{\mathrm{P}} + \alpha_0 \tag{2.1}$$

where α_{C} stands for the child wavelet coefficient and α_{P} stands for its parent (i.e.,

it is obtained at the immediately coarser scaler and at the same position). η_0 and α_0 are random variables mutually independent, also independent from $\alpha_{\rm P}$. The authors observed that this model fits reasonably well the conditioned histograms for many different wavelet bases, although depending on the particular basis the amplitude of the variable α_0 varies; for smaller α_0 the linear range in the conditioned histogram is larger and the converse. Now a question reasonably raises: is there any particular choice of wavelet for which the amplitude of α_0 vanishes?

3. Towards an optimal representation of data

3.1. *Canonical cascades*

The paradigm of systems in which multiplicative cascades develop are scale-invariant systems with one or many fractal interfaces. In them, conveniently designed intensive variables put in evidence a complex interplay between different scales.

Let $s(\vec{x})$ be a physical variable representing the signal under study. To study the scale relations of the system, we will need a properly defined, intensive, scale-dependent functional $\mathbf{T}$ applied to the signal, $\mathbf{T}[s](r,\vec{x})$. This variable depends on the point $\vec{x}$ and a scope or scale parameter r that characterizes the range of influence of the functional. Typical examples of such a functional include the derivative at radius r, nonlinear measures based on the derivative or wavelet projections.

The canonical approach to multiplicative cascades is a statistical approach. The name "canonical" comes from its analogy in Statistical Physics with the Canonical ensemble, in which system variables are characterized by their average values over all the states compatible with the thermal constraint. In the canonical approach to multiplicative cascades, the object under study is the distribution of the variable $\mathbf{T}[s](r,\bullet)$ for different values of the scale parameter r only, disregarding the localization $\vec{x}$, i.e., considering all the points as statistically equivalent. That is why we will simply denote this variable as $\mathbf{T}_r$. The analysis of its distribution is achieved through its order-p moments; studying the moments is enough to completely define the distribution provided they do not diverge too fast with p.[5]

A multiscaling (also called multifractal) signal s is characterized by the power-law scaling in the order-p moments of the related variable $\mathbf{T}_r$, in the way:

$$\langle \mathbf{T}_r^p \rangle \;=\; A_p^{\mathbf{T}}\, r^{\tau_p} \;+\; o(r^{\tau_p}). \tag{3.1}$$

Recall that the symbol $o(r^{\tau_p})$ means a contribution that is negligible compared to r^{τ_p} when r goes to zero. In fractal signals, the exponent τ_p is directly proportional to the moment order p and the proportionality constant is called singularity exponent or Hurst exponent. In multifractal signals,[14] the dependence of τ_p on p is nonlinear, a fact known as *anomalous scaling*. In Appendix Appendix A, the connection between geometry and statistics of multifractal signals is discussed in greater detail.

In order to separate the part of the statistics that has to do with changes in

scale, two different scales r, L with $r < L$ can be compared, so:

$$\langle \mathbf{T}_r^p \rangle = \left(\frac{r}{L}\right)^{\tau_p} \langle \mathbf{T}_L^p \rangle \tag{3.2}$$

which is valid at lowest order in the limit of small r and L. For some particular τ_p, this relation implies the existence of a variable η_κ such that:

$$\langle \eta_\kappa^p \rangle = \kappa^{\tau_p} \tag{3.3}$$

where $\kappa = r/L < 1$. Notice that one of the conditions for the existence of this variable is the validity of the expansion above, which in turn depends on taking a scale ratio parameter κ smaller than 1; for this reason we have taken the ratio of the smaller scale by the larger scale. Notice also that there is no general proof on the existence of η_κ for an arbitrary τ_p; it can however be assumed to exist if τ_p defines infinitely divisible processes.[27,42,43,41] These cases cover many situations of interest, such as log-normal, log-Lévy or log-Poisson processes.

With the aid of the variable η_κ we can express eq. (3.2) in a more elegant way, making the cascade relation explicit:

$$\mathbf{T}_r \doteq \eta_{r/L}\, \mathbf{T}_L \tag{3.4}$$

with $\eta_{r/L}$ and $\mathbf{T}_L$ being mutually independent. Here the symbol $\doteq$ means that the equality holds distributionally, i.e., $\rho(\mathbf{T}_r) = \rho(\eta_{r/L}\, \mathbf{T}_L)$. However, this relation does not necessarily hold pointwise, as we will explain in the following subsection.

The introduction of eq. (3.4) now allows to split the statistics of the scaling variable $\mathbf{T}_r$ in two parts: one part, given by $\eta_{r/L}$, accounts for the properties of transformation under changes in scale, while the other part, given by $\mathbf{T}_L$, takes into account the behavior at a given reference scale L. Taking L as the largest possible scale in the system, the distribution of all the variables $\mathbf{T}_r$ at any arbitrary scale r can be referred to the fixed level $\mathbf{T}_L$ once the process of change in scale, $\eta_{r/L}$, is known.

We will call the $\eta_{r/L}$ *cascade variables*. Their distributions do not depend on the particular scales r and L they connect but only on the scale ratio $\kappa = r/L$. If we now consider three scales $r < r' < L$ and we apply eq. (3.4) to the three possible scale pairs it follows:

$$\eta_{r/L} \doteq \eta_{r/r'}\, \eta_{r'/L} \tag{3.5}$$

from which the name "cascade variable" becomes evident: the variable relating scales r and L is equivalent to the product of the variables relating any two intermediate scales. If any intermediate scale is allowed, it follows that the cascade variables must have an infinitely divisible distribution.[27,12,6] Another important characteristic of the distribution of the cascade variables is that it is a property of the signal and does not depend on the particular functional $\mathbf{T}$ used to obtain them, i.e., any functional capable to resolve the scaling exponents τ_p of the signal in eq. (3.1) leads to exactly the same distribution of cascade variables $\eta_{r/L}$.[14]

3.2. *Microcanonical cascade*

Equation (3.4) makes sense only as a distributional equality and does not imply that the functional of scale r at some point $\vec{x}$ is related to the functional of scale L at the same point through an independent multiplicative factor. In general, $\mathbf{T}[s](r,\vec{x})$ and $\mathbf{T}[s](L,\vec{x})$ are not related by a variable $\eta_{r/L}(\vec{x})$ that is statistically independent of $\mathbf{T}[s](L,\vec{x})$. Of course, we can always define $\tilde{\eta}_{r/L}(\vec{x})$ as the ratio of these two variables,

$$\tilde{\eta}_{r/L}(\vec{x}) \;=\; \frac{\mathbf{T}[s](r,\vec{x})}{\mathbf{T}[s](L,\vec{x})} \tag{3.6}$$

but for most of the possible functionals $\mathbf{T}$, the variables $\tilde{\eta}_{r/L}(\vec{x})$ are not independent of $\mathbf{T}[s](L,\vec{x})$ and thus they cannot be considered cascade variables, as they do not verify eq. (3.5). It is convenient to deal with cascade variables, as they are independent of the starting scale and only depend on the ratio of scales; this implies that they serve both to characterize the global properties of the system and to compactly codify its dynamics.

In many multifractal systems, the cascade process governs their dynamics as a local effective mechanism, what implies that there is a local variable $\eta_{r/L}(\vec{x})$ transferring energy, matter or information (depending on the system) from coarser to finer scales. Therefore, there may exist a system variable s and a scale-tunable functional $\mathbf{T}$ for which eq. (3.4) makes sense not only distributionally but also at any point $\vec{x}$ of the system. That is what we call *microcanonical cascade*. The name "microcanonical", once again, comes from the analogy with microcanonical ensembles from Statistical Physics; in microcanonical ensemble all the microscopic states compatible with dynamic constraints are known and given the same weight. Analogously, in microcanonical states all the details of the cascading process at each point are known, and not only as an averaged quantity. The notions of canonical and microcanonical cascades fit well with the frameworks of Canonical and Microcanonical formalisms.[54]

Among the functionals that are most commonly used to analyze the scaling properties of multifractal systems, wavelets occupy a prominent position. In many different multifractal systems, wavelet projections have been used to characterize their scaling properties with success.[48,50] In addition, wavelet projections can be inverted to retrieve the original signal.[10] That is why wavelet projections are good candidates to realize the microcanonical cascade.

Let $s(\vec{x})$ be a multifractal signal and let $\Psi(\vec{x})$ be a wavelet. We define the wavelet projection of s on Ψ at the position $\vec{x}$ and the resolution scale r as:

$$\mathbf{T}_{\Psi}[s](r,\vec{x}) \;\equiv\; \int \mathrm{d}\vec{y}\, s(\vec{y})\, \Psi\!\left(\frac{\vec{x}-\vec{y}}{r}\right). \tag{3.7}$$

In terms of wavelet projections, a microcanonical cascade has the following form:

$$\mathbf{T}_{\Psi}[s](r,\vec{x}) \;=\; \eta_{r/L}(\vec{x})\; \mathbf{T}_{\Psi}[s](L,\vec{x}). \tag{3.8}$$

Notice that the key point is that $\eta_{r/L}(\vec{x})$ has to be both a cascade variable –in the sense of eq. (3.5)– and independent from $\mathbf{T}_{\Psi}[s](L,\vec{x})$. We can thus define the

optimality of a wavelet as the degree of independence of $\tilde{\eta}_{r/L}(\vec{x})$ vs. $\mathbf{T}_\Psi[s](L,\vec{x})$; we will discuss this possibility in depth in Section 4. There are evidences that such an optimal wavelet exists in natural images[49] and in marine turbulence[31] for the specific case of wavelet dyadic representations.

In terms of a dyadic representation the cascade takes a simple form. Let us consider, by simplicity, a 1D signal $s(x)$ and Ψ an appropriate wavelet representation basis capable. Then, the signal can be expanded as a series of wavelet terms

$$s(x) \;=\; \sum_{j=-\infty}^{\infty} \sum_{k} \alpha_{j,k}\, \Psi_{j,k}(x) \tag{3.9}$$

where the coefficients $\alpha_{j,k}$, called *wavelet coefficients*, can be obtained as projections on the wavelet basis, namely:

$$\alpha_{j,k} \;=\; \langle \Psi_{j,k} \| s \rangle \tag{3.10}$$

provided the wavelet basis is orthonormal (for non-orthonormal bases, the extension is rather straightforward using the dual basis $\langle \tilde{\Psi}_{j,k} \|)$. Under these conditions, the canonical cascade relation, eq. (3.4), takes the following form:

$$\alpha_{j,k} \;\doteq\; \eta_{\frac{1}{2}}\, \alpha_{j+1,\lfloor k/2 \rfloor} \tag{3.11}$$

where the notation $\lfloor k/2 \rfloor$ means *the integer part of* $k/2$. Here we have written the cascade relation mimicking eqs. (3.4) and (3.8), although that the wavelet coefficients $\alpha_{j,k}$ are not intensive variables as the wavelet projections are as defined in eq. (3.7) (while wavelet projections are ∞-norm normalized, wavelet coefficients are 2-norm normalized, which is highly convenient in the derivations to follow, especially in section 4). This means that the η-like variables written hereafter will differ from those appearing in Equations (3.3) to (3.8) in a constant normalization factor of $\sqrt{r/L} = \frac{1}{\sqrt{2}}$.

Notice that $\alpha_{j,k}$ is the wavelet projection at the scale $r_{\rm C} = 2^j$ and position $x_{\rm C} = 2^j\, k$, while $\alpha_{j+1,\lfloor k/2 \rfloor}$ is the wavelet projection at the coarser scale $r_{\rm P} = 2^{j+1}$ and position $x_{\rm P} = 2^{j+1}\, \lfloor k/2 \rfloor$; the positions $x_{\rm C}$ and $x_{\rm P}$ differ at most by $r_{\rm C}$, which is the spatial uncertainty at the scale $r_{\rm P}$, so at the scale $r_{\rm P}$ we can consider that $x_{\rm C}$ and $x_{\rm P}$ refer to the same position. To alleviate the notation, for given fixed scale index j and position index k, $\alpha_{\rm P} \equiv \alpha_{j+1,\lfloor k/2 \rfloor}$ is known as the Parent coefficient, $\alpha_{\rm C} \equiv \alpha_{j,k}$ is the Child coefficient and the cascade variable is $\eta \equiv \eta_{\frac{1}{2}}$, and we just write the canonical cascade relation above as:

$$\alpha_{\rm C} \;\doteq\; \eta\, \alpha_{\rm P}. \tag{3.12}$$

A dyadic wavelet basis is said to be optimal if the associated wavelet coefficients verify the microcanonical cascade relation, namely:

$$\alpha_{\rm C} \;=\; \eta\, \alpha_{\rm P} \tag{3.13}$$

where η is independent of the parent wavelet coefficient $\alpha_{\rm P}$ and is thus a cascade variable with associated scale ratio $\frac{1}{2}$.

It has been shown that if the optimal wavelet exists, there is a constructive formula to unambiguously obtain it from a large enough dataset.[49] This formula proves uniqueness of the wavelet, but it is rather unstable, specially in estimating the tails of the wavelet, and it is not very useful unless a large amount of data is available as learning set.[45,11] We will next analyze alternative strategies for a more stable determination of the optimal wavelet.

4. Optimization from suboptimal representations

4.1. *Quadrature Mirror Filters*

Dyadic wavelet expansions can be used to describe the cascade with a discrete set of parameters. A particularly, widely used way to implement dyadic representation is in terms of Quadrature Mirror Filters (QMF), which are very robust in practical applications. The main advantage of QMFs is that they are discrete filters so both obtaining the wavelet coefficients from a signal and reconstructing the signal from its wavelet coefficients are numerically exact operations (apart from round-off errors). In the following, we will summarize the most relevant facts about QMFs; the interested reader can consult some wavelet textbooks.[10,21]

When a function Ψ defines a wavelet basis, it is possible to find another function Φ, called unity function, that is orthogonal to the wavelet but, contrarily to it, has non-zero mean and that can be used to represent the approximation of the signal at a given scale. The approximation of a signal $s(x)$ at a scale indexed as j_0 is given by an expansion of functions $\Phi_{j_0,k}$ whose coefficients are called *approximation coefficients*. The signal can hence be expanded as follows:

$$\begin{aligned} s(x) &= \sum_{j=-\infty}^{\infty} \sum_k \alpha_{j,k}\, \Psi_{j,k}(x) \qquad (4.1) \\ &= \sum_{j=-\infty}^{j_0} \sum_k \alpha_{j,k}\, \Psi_{j,k}(x) \;+\; \underbrace{\sum_k \beta_{j_0,k}\, \Phi_{j_0,k}(x)}_{A_{j_0}(x)} . \end{aligned}$$

Taking into account that the approximations $A_{j_0}(x)$ are numerable sums, we can define the QMF in the usual way, with two numerable filters, denoted by $\{g_n\}$ and $\{h_n\}$, that can be used to obtain the wavelet coefficients at any level in an accurate, fast way, provided that we know the approximation at the finest level, i.e., the signal at its discretization level. These filters can be obtained[10,21] by expanding the scaling function itself Φ (i.e., $\Phi_{0,0}$) up to the next coarser scale, $j_0 = 1$:

$$\Phi(x) = \sum_n g_n\, \Psi_{1,n}(x) + \sum_n h_n\, \Phi_{1,n}(x). \qquad (4.2)$$

In practice we will always analyze discretized signals, defined by a collection of values s_k. We will identify this collection with the approximation coefficients at the highest resolution $\beta_{0,k} = s_k$. Notice that since $r = 2^j$, the highest resolution at level

$j = 0$ implies to express r in units of pixels. The wavelet coefficients at the next coarser level $j = 1$ can be found by applying the filter $\vec{g}$. Let $\vec{\alpha}_1$ be the vector of these wavelet coefficients, then we have:

$$\alpha_{1,k} \;=\; \sum_n g_{n-2k}\, \beta_{0,n}. \tag{4.3}$$

that is, the filter $\vec{g}$ acts by convolution on the vector $\vec{\beta}$. For later convenience, let us introduce the matrix $\mathbb{G}$ that represents the action of $\vec{g}$ by convolution, i.e., $\mathbb{G}_{nn'} = g_{n'-2n}$. We can now elegantly express eq. (4.3) in vectorial form as:

$$\vec{\alpha}_1 \;=\; \mathbb{G} \cdot \vec{\beta}_0. \tag{4.4}$$

Notice that the expression above can be used to relate the approximation and the wavelet coefficients of any two consecutive resolution levels, i.e.,

$$\vec{\alpha}_{j+1} \;=\; \mathbb{G} \cdot \vec{\beta}_j \tag{4.5}$$

but in order to obtain the wavelet coefficients at any other resolution we need an expression to obtain the coarser approximations derived from the highest resolved one. This can be done by means of the filter $\vec{h}$. Analogously to what has been derived previously, we have that two consecutive approximation levels can be related by the filter $\vec{h}$ as follows:

$$\vec{\beta}_{j+1} \;=\; \mathbb{H} \cdot \vec{\beta}_j \tag{4.6}$$

where $\mathbb{H}_{nn'} = h_{n'-2n}$. We already have the essentials to perform a perturbative analysis on the wavelet.

4.2. *Perturbative analysis*

In general, most of the wavelet bases applied to the analysis of given data are not optimal. This means that the cascade does not hold in the microcanonical sense and so eq. (3.13) cannot be used. In the following we will show that when the wavelet basis is relatively close to the optimal basis, the linear model proposed by Pottier et al., eq. (2.1), is verified. Our proof is based on the QMF representation introduced in the previous subsection and it is focused on 1D signals for simplicity. The generalization of higher dimensions is straightforward.

First, let the optimal QMF be denoted by $(\vec{g}, \vec{h})$. At the discretization level $j = 0$, the signal corresponds to the vector $\vec{\beta}_0^{\rm opt} = (\ldots, s_{-1}, s_0, s_1, \ldots)$. Let us consider now the Child and the Parent scale levels as the two next coarser dyadic levels, namely $j_{\rm C} = 1$, $r_{\rm C} = 2$ pixels and $j_{\rm P} = 2$, $r_{\rm P} = 4$ pixels (notice that the wavelet coefficients at levels $j \leq 0$ are all zero as discrete signals cannot vary inside their pixels, i.e., at levels finer than the discretization scale). This way, eq. (4.4) is notated as:

$$\vec{\alpha}_C^{\rm opt} \;=\; \mathbb{G} \cdot \vec{\beta}_0^{\rm opt}. \tag{4.7}$$

The approximation to the next level is given by:

$$\vec{\beta}_1^{\rm opt} \;=\; \mathbb{H} \cdot \vec{\beta}_0^{\rm opt} \tag{4.8}$$

from which the details at the coarser resolution (parent coefficients) can be deduced:

$$\vec{\alpha}_P^{\rm opt} = \mathbb{G} \cdot \vec{\beta}_1^{\rm opt} = \mathbb{G} \cdot \mathbb{H} \cdot \vec{\beta}_0^{\rm opt}. \tag{4.9}$$

Owing to the fact that the QMF is optimal, at each location k we can find an independent cascade variable η_k such that:

$$\alpha_{C,\,k}^{\rm opt} \;=\; \eta_k \, \alpha_{P,\lfloor k/2 \rfloor}^{\rm opt}. \tag{4.10}$$

If we define now the matrix $\mathbb{N}$ formed by these cascade variables disposed on the diagonal, namely:

$$\mathbb{N}_{kk'} \;=\; \eta_k \, \delta_{\lfloor k/2 \rfloor k'} \tag{4.11}$$

we have that the cascade relation between children and parent coefficients can be written for the child and parent detail vectors as follows:

$$\vec{\alpha}_{\rm C}^{\rm opt} \;=\; \mathbb{N} \cdot \vec{\alpha}_{\rm P}^{\rm opt}. \tag{4.12}$$

Let us now introduce a small perturbation on the optimal QMF; we will define a new, suboptimal QMF $(\vec{g}', \vec{h}') = (\vec{g} + \delta\vec{g}, \vec{h} + \delta\vec{h})$ for small $\delta\vec{g}$ and $\delta\vec{h}$. The new child detail vector will be given by:

$$\vec{\alpha}_{\rm C} = (\mathbb{G} + \delta\mathbb{G}) \cdot \vec{\beta}_0 = \vec{\alpha}_{\rm C}^{\rm opt} + \delta\mathbb{G} \cdot \vec{\beta}_0 = \mathbb{N} \cdot \vec{\alpha}_{\rm P}^{\rm opt} + \delta\mathbb{G} \cdot \vec{\beta}_0. \tag{4.13}$$

Notice that we have made the assumption $\vec{\beta}_0 = \vec{\beta}_0^{\rm opt}$ as both are identified with the signal itself at its discretization scale. The next coarser approximation vector is:

$$\vec{\beta}_1 = (\mathbb{H} + \delta\mathbb{H}) \cdot \vec{\beta}_0 = \vec{\beta}_1^{\rm opt} + \delta\mathbb{H} \cdot \vec{\beta}_0. \tag{4.14}$$

Finally, the details at the next coarser resolution up to the first perturbation order are given by the following vector:

$$\vec{\alpha}_P = (\mathbb{G} + \delta\mathbb{G}) \cdot \vec{\beta}_1^{\rm opt} + \mathbb{G} \cdot \delta\mathbb{H} \cdot \vec{\beta}_0 = \vec{\alpha}_{\rm P}^{\rm opt} + (\delta\mathbb{G} \cdot \mathbb{H} + \mathbb{G} \cdot \delta\mathbb{H}) \cdot \vec{\beta}_0. \tag{4.15}$$

Combining eq. (4.13) and eq. (4.15) we obtain:

$$\vec{\alpha}_C \;=\; \mathbb{N} \cdot \vec{\alpha}_{\rm P} \;+\; [\delta\mathbb{G} - \mathbb{N} \cdot (\delta\mathbb{G} \cdot \mathbb{H} + \mathbb{G} \cdot \delta\mathbb{H})] \cdot \vec{\beta}_0. \tag{4.16}$$

Defining now $\vec{\alpha}_0$ as:

$$\vec{\alpha}_0 \;\equiv\; [\delta\mathbb{G} - \mathbb{N} \cdot (\delta\mathbb{G} \cdot \mathbb{H} + \mathbb{G} \cdot \delta\mathbb{H})] \cdot \vec{\beta}_0 \tag{4.17}$$

when substituted in eq. (4.16) we obtain the vector version of the linear model, eq. (2.1), introduced in Ref. 31, namely:

$$\vec{\alpha_{\rm C}} \;=\; \mathbb{N} \cdot \vec{\alpha}_{\rm P} \;+\; \vec{\alpha}_0. \tag{4.18}$$

According to our derivation we can now make some remarks about the variables η_0 and α_0 appearing in the linear model. First, the variable η_0 is an actual cascade variable, distributed according to the same statistics, and up to the first order it is independent from the parent coefficient in the suboptimal basis. Second, the variable α_0 is much smaller than the term $\eta_0 \, \alpha_{\rm P}$ and is only relevant for small values of $\alpha_{\rm P}$. We cannot say much about the statistical distribution of α_0, not even whether it is independent or not from the other term. However, it is reasonable to think that this variable is governed by the fluctuations due to the mixing of the different terms in

the definition of α_0 (see eq. (4.17)) and the arbitrary character of the perturbations $\delta\mathbb{G}$ and $\delta\mathbb{H}$. This fact allows to consider this variable independent from $\eta_0\,\alpha_{\rm P}$, as the experiences in Ref. 31 confirm.

4.3. *Optimization strategies*

The results in the previous subsection show that the amplitude of α_0 (the optimality degree) varies continuously under perturbations on the wavelet. Hence, an optimization strategy based on successive corrections of the wavelet would lead to the actual optimal wavelet, provided that the initial guess is not too far away from the optimality.

As seen in Section 3 all cascade variables η are equally distributed, independent of the wavelet basis from which they are derived, and their moments can be retrieved from τ_p. In addition, the expectation value of $|\eta|$ is fixed due to translational invariance[48,49]: $\langle|\eta|\rangle = 2^{-d/2}$ in an arbitrary dimension d; $\langle|\eta|\rangle = \frac{1}{\sqrt{2}}$ for 1D signals. According to the linear model, eq. (2.1), the expectation value of $|\tilde{\eta}|$ is:

$$\langle|\tilde{\eta}|\rangle = \langle|\eta_0 + \alpha_0\,\alpha_{\rm P}^{-1}|\rangle. \tag{4.19}$$

Let us explore the two asymptotic limits. If the wavelet is optimal then $\alpha_0 = 0$ so:

$$\langle|\tilde{\eta}|\rangle = \langle|\eta_0|\rangle = \langle|\eta|\rangle. \tag{4.20}$$

In the opposite case, for a highly non-optimal wavelet we will have that $\alpha_0/\alpha_{\rm P} \gg \eta_0$ and taking α_0 independent of $\alpha_{\rm P}$ we would obtain that:

$$\langle|\tilde{\eta}|\rangle = \langle|\alpha_0|\rangle\langle|\alpha_{\rm P}|^{-1}\rangle = q\langle|\eta|\rangle \tag{4.21}$$

where $q = \langle|\alpha_{\rm P}|\rangle\langle|\alpha_{\rm P}|^{-1}\rangle$, which by Jensen's inequality[34] is greater than one: $q > 1$, for any random variable $\alpha_{\rm P}$. For an intermediate case, the preceding two regimes are combined. If p is the proportion of the range of values of $\alpha_{\rm P}$ for which $\eta_0 > \alpha_0/\alpha_{\rm P}$ and $(1-p)$ is its complementary, we roughly have that:

$$\langle|\tilde{\eta}|\rangle \;\approx\; p\,\langle|\eta|\rangle \;+\; (1-p)\,q\,\langle|\eta|\rangle \tag{4.22}$$

Hence, in any instance $\langle|\tilde{\eta}|\rangle \geq \langle|\eta|\rangle$ and $\langle|\tilde{\eta}|\rangle = \langle|\eta|\rangle$ for the optimal wavelet only. We normalize this quantity to define the optimality degree Q as:

$$Q = \frac{\langle|\tilde{\eta}|\rangle}{\langle|\eta|\rangle} \tag{4.23}$$

which is $Q \geq 1$, and $Q = 1$ for the optimal wavelet only. Q is a monotonic function of the amplitude of α_0 (which in fact measures the deviation from the optimal case), so that Q not only evidences the optimal wavelet case (when $Q = 1$) but it actually ranks suboptimal wavelets by their respective deviation from optimality.

An alternative approach would consist in analyzing the degree of independence between $\tilde{\eta}$ and $\alpha_{\rm P}$. As stated in Section 3.2, independence between these variables is an indicator of the optimality of the wavelet. This can be expected, as having $Q > 1$ implies correlation between $\tilde{\eta}$ and $\alpha_{\rm P}$, and correlation implies statistical dependence.

In this case decorrelation ($Q = 1$) implies independence also, as $Q = 1$ implies optimality and optimality implies independence. In fact, $\tilde{\eta}$ and $\alpha_{\rm P}$ are negatively correlated in suboptimal cases ($Q > 1$), and uncorrelated only for the optimal wavelet:

$$Q = \frac{\langle|\tilde{\eta}|\rangle}{\langle|\eta|\rangle} = 1 - \frac{\mathrm{Cov}(|\tilde{\eta}|, |\alpha_{\rm P}|)}{\langle|\alpha_{\rm C}|\rangle}. \tag{4.24}$$

A standard measure of statistical dependence is the mutual information. Therefore, the mutual information between $\tilde{\eta}$ and $\alpha_{\rm P}$, $I = I(\tilde{\eta}, \alpha_{\rm P})$, could also measure the degree of optimality of a wavelet. However, the advantage of using Q instead of I comes from the fact that Q is less numerically sensitive to sampling size than I. The main problem with the practical calculation of the mutual information is that it is very data demanding (see the estimation of uncertainties in Appendix Appendix B and the numerical study in the next section). Hence, when only small and short datasets are available, Q is more convenient as indicator of the optimality degree.

5. Results

Now we want to show in practice the theoretical results given in the previous section, namely the validity of the linear model, eq. (4.18), and the performance of our measures of optimality, Q and I. We have generated synthetic signals according to a given cascade process and with a prefixed optimal wavelet basis. The cascades are generated by first calculating the wavelet coefficients through eq. (3.13) for dyadic scale steps, and then generating the signal from these wavelet coefficients, eq. (3.9), with the chosen wavelet basis. The multiplicative variable η is a random variable following a given cascade distribution without horizontal correlations, i.e., it follows Benzi et al.'s model.[3,1] As distribution for the cascade variable η we have chosen the log-Poisson distribution, which has been proposed in many different physical systems.[12,42,48] Hence, we have chosen a translationally invariant log-Poisson characterized by having a most singular manifold of dimension $D_\infty = 0$ and singularity exponent $h_\infty = -\frac{1}{2}$, which is a realistic choice of parameters.[47,48] See the Appendix Appendix A for a description of the log-Poisson distribution and parameters.

Regarding the linear model, it has been derived by perturbative analysis. In Figure 1 we validate this model in practice, for a very long series of 67 108 864 points. Figure 1 top shows the probability density function of the child coefficient $\alpha_{\rm C}$ conditioned by a given value of the parent coefficient $\alpha_{\rm P}$. In Figure 1 top left the analysis wavelet is the optimal wavelet and in Figure 1 top right the analysis wavelet is a suboptimal wavelet. First, we can observe that for any value of the parent coefficient, the child coefficient is symmetrically distributed $\rho(\alpha_{\rm C}|\alpha_{\rm P}) = \rho(-\alpha_{\rm C}|\alpha_{\rm P})$, what means that $\langle\eta_0\rangle = \langle\alpha_0\rangle = 0$; this also implies $\rho(\alpha_{\rm C}|\alpha_{\rm P}) = \rho(\alpha_{\rm C}| - \alpha_{\rm P})$. We also observe that the standard deviation of the child coefficient conditioned by a value of the parent coefficient depends hyperbolically on it, as predicted by the linear model, namely:

$$\sigma_{\alpha_{\rm C}|\alpha_{\rm P}} = \sqrt{\langle\alpha_{\rm C}^2|\alpha_{\rm P}\rangle} = \sqrt{A\,\alpha_{\rm P}^2 + B} \tag{5.1}$$

where the constants A and B are given by the linear model: $\langle \eta_0^2 \rangle = A$ and $\langle \alpha_0^2 \rangle = B$. For the optimal wavelet, $A = \langle \eta^2 \rangle$ and $B = 0$, so that η_0 coincides with η. Additional evidence is furnished by the conditioned histograms of logarithms of the parent and child coefficients, i.e., the conditional probability of $\ln|\alpha_{\mathrm{C}}|$ for a given value of $\ln|\alpha_{\mathrm{P}}|$, which is shown in Figure 1 bottom. As before, the bottom left histogram corresponds to the optimal case while the bottom right histogram is a suboptimal case. The absolute values fold the top histograms to the first quadrant while the logarithms balance the kurtotic distributions of the wavelet coefficients. When the series is analyzed with its optimal wavelet, the histogram exhibits a perfectly straight maximum-probability line and small dispersion around this line. In contrast, when the series is analyzed with a suboptimal wavelet the histogram bends on the left to a horizontal line. The bending of the maximum-probability line to a horizontal line means that α_{C} becomes independent of α_{P}, because hence for any value of α_{P} the distribution of α_{C} is the same. This effect is in agreement with the linear model, eq. (4.18), as the term α_0 becomes dominant when α_{P} is too small, and it is independent of α_{P}. The two asymptotic limits predicted by the linear model can be easily obtained from eq. (4.18): when the value of the parent coefficient α_{P} is large, in $\ln|\alpha_{\mathrm{C}}| = \ln|\eta_0\,\alpha_{\mathrm{P}}| + \ln\left|1 + \frac{\alpha_0}{\eta_0\,\alpha_{\mathrm{P}}}\right|$ the second term becomes irrelevant, so that $\ln|\alpha_{\mathrm{C}}| \approx \ln|\alpha_{\mathrm{P}}| + \ln|\eta_0|$. When the value of the parent coefficient α_{P} is small, in $\ln|\alpha_{\mathrm{C}}| = \ln|\alpha_0| + \ln\left|1 + \frac{\eta_0\,\alpha_{\mathrm{P}}}{\alpha_0}\right|$ the second term rapidly becomes irrelevant, so that $\ln|\alpha_{\mathrm{C}}| \approx \ln|\alpha_0|$. Not only the asymptotes, but also the central behavior is the one given by the model, as the line of maximum-probability of the histogram fits a shape:

$$\ln|\alpha_{\mathrm{C}}|_{\mathrm{m.p.}} = \ln\left(|\alpha_0|_{\mathrm{m.p.}} + |\eta_0|_{\mathrm{m.p.}}\,\exp\ln|\alpha_{\mathrm{P}}|\right) \tag{5.2}$$

where m.p. stands for *maximum probable*, i.e., these values are the probability maxima of their respective distributions. Notice also that the amplitude of the fluctuations of α_0 is larger than that of η_0, and for that reason dispersion grows when the value of α_{P} decreases.

In a more extensive test, we have used 24 standard wavelets of very different families. These are: Haar, Daubechies (orders 2 to 10), Coiflet (orders 1 to 5), Symmlet (orders 4 to 8) and Battle-Lemarié (spline wavelets) (orders 1, 2, 3 and 6). Each family has some remarkable properties, which can be controlled by an order parameter p. Wavelets from Daubechies family have the shortest possible QMFs compatible with having vanishing moments up to order p. The Symmlet family verifies the same requirement but also requesting wavelets to be as symmetric as possible. The Coiflet family verifies the same requirement than Daubechies family, but in addition the associated unity function also verifies to have vanishing moments up to order p. Finally, Battle-Lemarié family is generated by the orthonormalization of order-p splines. In essence, each family searches to approximate the signal with a basis with prefixed degree of regularity, attending to a number vanishing derivatives, as Daubechies wavelets, or by approximation with a regular enough polynomial, as Battle-Lemarié wavelets. Please refer to Ref. 21 for a detailed description of these

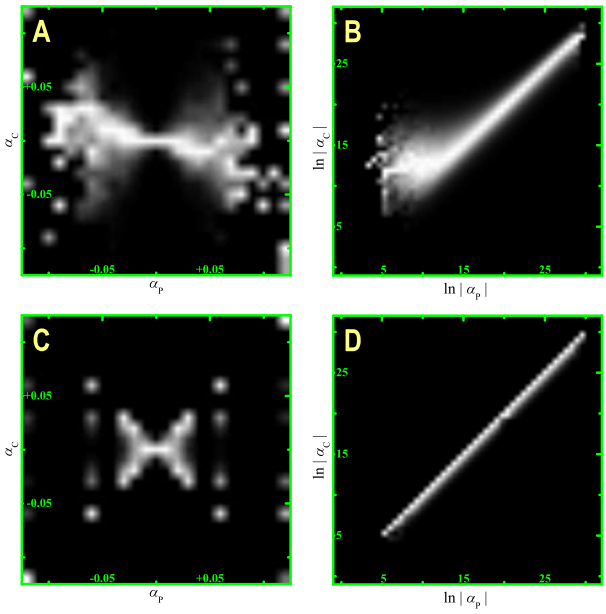

Fig. 1. Example that shows the influence of parent coefficients, α_{P}, over child coefficients, α_{C}, in the non-optimal and optimal cases. We plot the joint histograms of α_{C} vs. α_{P} and $\ln|\alpha_{\mathrm{C}}|$ vs. $\ln|\alpha_{\mathrm{P}}|$ for synthetic cascade data generated with the Coiflet-1 wavelet and analyzed, eq. (3.10), with the Battle-Lemarié-6 wavelet (histograms A and B, a non-optimal case) and the Coiflet-1 wavelet (histograms C and D, the optimal case). In each histogram, each column has been normalized so that vertical slices correspond to the probability distribution function of the vertical-axis variable conditioned to the value in the horizontal-axis. This way, black corresponds to zero probability and white corresponds to maximum probability. For wavelet coefficients (histograms A and C), their values range from -0.125 to 0.125 in both axes (nondimensional) and the histograms are defined by a grid of 25×25 bins. For logarithms of wavelet coefficients (histograms B and D), their values range from -32 to 1 (nondimensional) and the histograms are defined by a grid of 50×50 bins. In all cases, the bins are smoothed with a cubic spline to enhance presentation. The analyzed data are a single series of very high resolution (67 108 864 points). Cascade process is a log-Poisson of parameters $D_\infty = 0$ and $h_\infty = -\frac{1}{2}$ (see Appendix Appendix A for a detailed description of the process).

wavelet bases.

Notice that Haar and Daubechies-1 are the same, and Symmlet 1 to 3 also coincide with Daubechies 1 to 3 respectively, so we have not repeated them. For each wavelet, we have generated 64 series of 4096 points, which is a quite realistic size. Hence, we have generated 24 ensembles of series and each wavelet is optimal in an ensemble. For a given ensemble, we have processed it with the same 24 wavelet bases. That is, for each ensemble we have tried its optimal basis and 23 non-optimal bases. We have hence performed $24 \times 24 = 576$ different tests to check the validity of the linear model.

In Figure 2 we present the joint histograms of $\ln|\alpha_{\mathrm{C}}|$ vs. $\ln|\alpha_{\mathrm{P}}|$, obtained from

the different ensembles when they are analyzed with the 24 bases, arranged in a tabular form. By construction, the histograms on the diagonal of this table correspond to the case in which the ensemble is analyzed with its optimal wavelet, and hence these histograms exhibit the same optimal behavior seen in Figure 1, bottom left. In contrast, when an ensemble is analyzed with a suboptimal wavelet the maximum-probability line bends on the left to a horizontal line, as in Figure 1, bottom right. In all cases, linear model holds, as the histogram only present two asymptotic regimes: linear dependence for great values of α_{P} (right side) and independence for small values of α_{P} (left size). As the optimal and analyzing wavelets become more different, the amplitude of the term α_0 increases and hence the extension of the horizontal line in the joint histogram becomes longer.

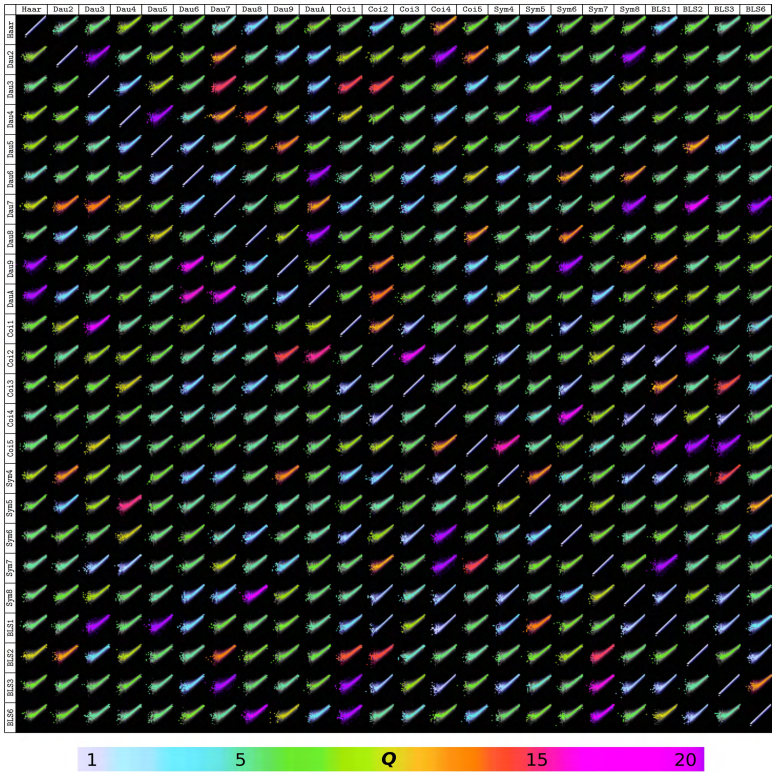

Fig. 2. Joint histograms of $\ln \alpha_{\mathrm{C}}$ vs. $\ln \alpha_{\mathrm{P}}$ for synthetically generated cascade process data. The histograms have been arranged as in Table 1, i.e., generation wavelet (rows) vs. analysis wavelet (columns), so that the main diagonal corresponds to the optimal wavelet cases. Each histogram column has been normalized so that it corresponds to the probability distribution function of $\ln \alpha_{\mathrm{C}}$ conditioned to a given value of $\ln \alpha_{\mathrm{P}}$ in the horizontal axis. Values range -20 to 1 in both axes, the same for all the histograms. The analyzed ensembles correspond to 64 series of 4096 points each and the generating cascade process is a log-Poisson of parameters $D_\infty = 1$ and $h_\infty = -\frac{1}{2}$. Each histogram has 30×30 boxes. Histograms are colored according to its value of Q parameter.

	Haar	Dau2	Dau3	Dau4	Dau5	Dau6	Dau7	Dau8	Dau9	DauA	Coi1	Coi2	Coi3	Coi4	Coi5	Sym4	Sym5	Sym6	Sym7	Sym8	BLS1	BLS2	BLS3	BLS6
Haar	1.00	5.88	6.69	9.19	8.40	7.68	5.63	4.87	7.42	8.77	7.12	5.11	7.30	11.3	8.15	6.11	4.86	7.61	8.18	8.06	5.44	7.50	5.94	7.37
	0.00	0.13	0.12	0.12	0.13	0.12	0.12	0.12	0.13	0.14	0.13	0.12	0.12	0.12	0.12	0.13	0.12	0.12	0.13	0.12	0.12	0.13	0.12	0.13
Dau2	8.81	1.00	22.9	6.26	8.92	8.17	11.8	6.43	5.13	5.10	9.91	8.84	9.66	45.8	12.2	5.95	4.28	7.43	7.22	19.5	8.42	7.47	6.38	7.70
	0.12	0.00	0.11	0.12	0.13	0.13	0.12	0.12	0.12	0.13	0.15	0.12	0.13	0.12	0.12	0.12	0.10	0.13	0.13	0.12	0.12	0.12	0.12	0.13
Dau3	6.53	7.31	1.00	4.69	9.96	6.86	14.6	8.63	7.04	4.80	14.3	14.0	7.42	8.54	5.28	6.63	8.41	8.59	3.83	8.83	8.24	6.91	7.01	6.63
	0.11	0.12	0.00	0.11	0.13	0.14	0.13	0.13	0.12	0.12	0.12	0.13	0.13	0.13	0.13	0.12	0.13	0.13	0.11	0.12	0.13	0.12	0.13	0.12
Dau4	9.10	8.59	4.97	1.00	21.8	5.64	11.4	13.1	9.09	4.95	10.6	8.69	6.70	5.27	6.25	8.49	38.6	6.59	3.35	5.86	8.68	7.10	7.63	7.57
	0.12	0.12	0.11	0.00	0.11	0.13	0.14	0.13	0.13	0.12	0.12	0.12	0.13	0.12	0.12	0.13	0.12	0.12	0.09	0.13	0.12	0.13	0.13	0.13
Dau5	9.04	8.14	5.88	4.02	1.00	3.75	5.86	9.35	12.2	7.94	6.00	5.64	6.95	10.3	7.73	6.85	8.78	9.28	6.85	6.58	6.31	11.7	5.24	6.38
	0.13	0.13	0.13	0.11	0.00	0.11	0.13	0.14	0.13	0.13	0.13	0.12	0.13	0.12	0.13	0.12	0.13	0.13	0.12	0.11	0.12	0.13	0.12	0.13
Dau6	5.57	7.12	6.58	7.60	3.65	1.00	4.02	6.40	8.41	34.0	6.05	7.92	4.41	4.87	10.6	4.67	6.06	11.2	6.51	11.4	6.79	6.53	6.35	5.96
	0.13	0.14	0.13	0.12	0.11	0.00	0.12	0.13	0.13	0.13	0.12	0.12	0.12	0.12	0.13	0.12	0.13	0.12	0.13	0.12	0.12	0.12	0.12	0.13
Dau7	10.2	12.7	12.8	9.84	7.72	3.87	1.00	6.33	8.00	11.9	4.63	5.76	4.62	6.16	6.73	6.05	5.77	5.92	7.55	22.3	6.71	17.4	6.12	21.3
	0.13	0.12	0.14	0.13	0.13	0.11	0.00	0.11	0.13	0.13	0.13	0.13	0.10	0.12	0.13	0.12	0.12	0.11	0.13	0.12	0.12	0.13	0.12	0.12
Dau8	6.64	5.25	6.19	8.13	10.8	7.16	5.71	1.00	9.82	34.6	7.38	7.41	7.18	6.18	11.4	6.86	6.24	12.1	7.85	5.88	7.81	8.14	7.79	9.05
	0.13	0.13	0.13	0.14	0.13	0.13	0.12	0.00	0.12	0.13	0.12	0.13	0.12	0.13	0.13	0.13	0.12	0.11	0.12	0.12	0.13	0.14	0.13	0.13
Dau9	21.3	8.34	8.11	7.16	6.56	16.8	8.50	4.78	1.00	9.04	8.17	12.2	8.51	6.43	5.23	6.89	8.32	23.4	7.06	11.8	11.7	5.95	7.76	8.44
	0.13	0.13	0.12	0.13	0.14	0.13	0.13	0.12	0.00	0.12	0.13	0.14	0.13	0.14	0.13	0.13	0.12	0.12	0.12	0.13	0.14	0.13	0.13	0.14
DauA	29.6	4.95	6.30	8.67	6.26	15.9	16.8	5.85	3.78	1.00	7.84	12.8	8.25	7.09	4.88	9.75	6.84	7.82	4.66	7.99	9.80	9.54	7.59	7.09
	0.13	0.13	0.12	0.12	0.13	0.13	0.13	0.13	0.11	0.00	0.14	0.14	0.13	0.14	0.12	0.14	0.13	0.14	0.12	0.14	0.14	0.13	0.14	0.13
Coi1	7.78	10.1	18.1	6.30	7.18	9.07	4.67	5.30	7.70	9.89	1.00	11.4	2.63	6.98	8.70	6.59	8.44	3.02	8.54	6.14	12.2	8.40	5.73	5.51
	0.13	0.14	0.12	0.12	0.13	0.12	0.12	0.13	0.13	0.13	0.00	0.13	0.09	0.13	0.13	0.13	0.13	0.10	0.12	0.13	0.12	0.12	0.13	0.12
Coi2	8.38	6.25	9.09	9.42	7.50	6.29	5.58	6.38	14.2	15.3	6.73	1.00	17.0	2.05	7.71	2.62	6.99	7.29	10.1	2.23	1.17	44.1	6.49	6.26
	0.12	0.12	0.12	0.12	0.11	0.12	0.13	0.13	0.13	0.13	0.12	0.00	0.13	0.08	0.12	0.09	0.13	0.13	0.12	0.09	0.04	0.12	0.08	0.13
Coi3	7.45	10.0	7.68	10.6	5.84	4.56	6.34	5.12	6.56	7.38	2.81	6.69	1.00	6.75	9.09	7.09	7.04	2.35	7.05	5.88	11.7	6.36	13.9	4.87
	0.13	0.12	0.12	0.12	0.13	0.12	0.09	0.12	0.13	0.14	0.10	0.13	0.00	0.13	0.12	0.13	0.12	0.07	0.13	0.13	0.13	0.11	0.12	0.12
Coi4	6.34	7.85	6.86	7.58	5.84	5.76	5.73	6.36	6.40	7.65	5.73	2.06	6.28	1.00	6.79	3.22	5.68	16.8	9.23	1.86	1.96	9.38	2.06	7.18
	0.12	0.12	0.12	0.12	0.11	0.12	0.13	0.13	0.13	0.13	0.13	0.08	0.13	0.00	0.12	0.10	0.13	0.13	0.12	0.07	0.06	0.13	0.06	0.13
Coi5	6.57	8.36	10.3	6.42	6.72	7.27	8.52	6.09	7.00	6.72	8.84	9.85	7.22	11.9	1.00	15.8	6.47	8.84	5.69	6.80	16.8	20.7	20.8	9.33
	0.13	0.12	0.13	0.13	0.12	0.13	0.13	0.13	0.13	0.13	0.12	0.12	0.12	0.12	0.00	0.12	0.13	0.13	0.12	0.12	0.12	0.13	0.12	0.13
Sym4	9.93	11.9	9.21	6.49	8.22	4.73	5.11	7.21	12.3	7.29	5.25	4.15	7.73	2.84	8.11	1.00	12.0	5.55	7.88	3.46	2.97	6.26	14.2	6.58
	0.12	0.12	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.13	0.13	0.09	0.13	0.10	0.13	0.00	0.13	0.13	0.13	0.08	0.09	0.12	0.10	0.13
Sym5	7.37	4.95	9.19	14.8	7.32	6.19	7.06	6.19	5.83	6.14	7.82	5.82	7.44	6.27	7.47	9.90	1.00	6.00	9.13	6.94	7.32	8.81	6.08	11.1
	0.12	0.11	0.13	0.13	0.13	0.12	0.12	0.12	0.12	0.13	0.12	0.12	0.11	0.13	0.12	0.12	0.00	0.11	0.12	0.12	0.13	0.13	0.13	0.13
Sym6	6.63	7.73	7.17	10.7	6.56	7.94	5.64	4.26	6.91	5.90	3.38	8.96	2.54	39.2	7.32	5.54	5.04	1.00	8.70	6.01	8.11	6.21	8.40	5.38
	0.13	0.13	0.13	0.12	0.13	0.13	0.11	0.11	0.12	0.14	0.09	0.13	0.08	0.13	0.13	0.13	0.11	0.00	0.12	0.13	0.12	0.11	0.13	0.12
Sym7	6.31	6.04	3.72	2.90	6.47	6.91	10.1	6.46	5.49	8.23	7.03	11.5	7.22	21.3	14.2	6.82	8.25	8.60	1.00	8.43	23.7	6.24	7.20	7.05
	0.12	0.13	0.10	0.09	0.13	0.13	0.13	0.12	0.12	0.12	0.13	0.13	0.13	0.12	0.12	0.12	0.13	0.12	0.00	0.13	0.13	0.12	0.13	0.12
Sym8	6.59	7.59	9.59	7.13	6.28	4.72	4.92	18.5	8.84	6.77	6.07	2.53	5.63	3.62	6.34	2.92	6.35	4.82	10.1	1.00	2.80	8.96	2.75	6.62
	0.13	0.12	0.13	0.12	0.12	0.12	0.13	0.13	0.13	0.13	0.12	0.09	0.13	0.08	0.12	0.09	0.13	0.13	0.12	0.00	0.09	0.12	0.09	0.13
BLS1	6.57	6.78	20.6	6.97	21.6	4.78	7.48	6.87	8.41	6.35	5.27	2.43	9.54	1.95	7.07	4.16	12.7	8.60	7.32	2.44	1.00	6.59	2.77	5.18
	0.12	0.12	0.12	0.12	0.11	0.12	0.13	0.13	0.12	0.13	0.12	0.03	0.13	0.08	0.12	0.09	0.13	0.13	0.13	0.09	0.00	0.13	0.09	0.13
BLS2	10.3	11.8	5.41	9.95	7.07	6.45	12.8	8.66	7.67	7.45	13.4	14.2	5.69	6.70	5.96	6.94	8.14	9.08	14.6	6.92	8.05	1.00	8.01	2.68
	0.12	0.12	0.13	0.12	0.13	0.13	0.12	0.13	0.13	0.12	0.12	0.13	0.12	0.13	0.13	0.13	0.12	0.12	0.12	0.13	0.13	0.00	0.13	0.10
BLS3	8.23	6.90	7.10	7.66	6.51	4.91	24.1	7.02	6.35	8.48	34.5	3.58	9.00	2.18	8.06	3.47	5.66	6.72	15.9	3.36	2.45	6.38	1.00	12.0
	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.12	0.09	0.13	0.08	0.12	0.10	0.13	0.13	0.12	0.08	0.10	0.13	0.00	0.12
BLS6	8.40	8.15	6.41	5.87	6.56	8.52	6.47	18.8	10.5	5.29	55.7	6.16	5.51	8.46	5.38	7.17	8.46	7.57	18.6	7.22	10.4	2.72	5.97	1.00
	0.12	0.12	0.13	0.13	0.13	0.13	0.12	0.13	0.13	0.13	0.13	0.13	0.12	0.13	0.13	0.13	0.13	0.12	0.13	0.13	0.13	0.11	0.13	0.00

Q	1.00 - 1.50	1.50 - 3.00	3.00 - 6.00	> 6.00
I	0.00 - 0.04	0.04 - 0.08	0.08 - 0.12	> 0.12

Table 1. Summary of the Q (upper side of the cell) and I (lower side of the cell) optimality measures for synthetic cascade data. Each row corresponds to an ensemble generated with the wavelet written sideways at left (generation wavelet), while each column corresponds to the results obtained while analyzing these ensembles with the wavelet written at top (analysis wavelet). The ensembles correspond to 64 series of 4096 points each and the generating cascade process is a log-Poisson of parameters $D_\infty = 0$ and $h_\infty = -\frac{1}{2}$. Mutual information (I) is expressed in bits. Uncertainties of two sigmas are 0.0025 for Q and 0.03 bits for I.

In Table 1 we present the results of the mutual information I between $\tilde{\eta}$ and α_p, and the Q parameter as defined in eq. (4.23) for the different combinations of ensemble and analysis wavelet. As shown in the table, only when the processing wavelet coincides with the optimal wavelet the values of I and Q drop to 0 and 1, respectively, while for other, non-optimal wavelets these values are always larger. This proves that Q has the same performance as I to assess the optimality of a wavelet basis, but the Q parameter is less statistically demanding.

The Q parameter is obtained by means of the average of $\tilde{\eta}$ and so, according to the Central Limit Theorem, it converges to its theoretical value with a standard deviation that depends on the number of samples N as $N^{-\frac{1}{2}}$, $\sigma_{\langle|\tilde{\eta}|\rangle} = \sigma_{|\tilde{\eta}|}\, N^{-\frac{1}{2}}$ (recall that the average in the denominator of Q, $\langle|\eta|\rangle$, is theoretically fixed to $\frac{1}{\sqrt{2}}$ due to translational invariance). $\sigma_{|\tilde{\eta}|}$ depends on the wavelet and can be analytically calculated for the optimal case only, which in fact is the most interesting case as we want to have the error bar that discriminates optimal from non-optimal wavelets. For the distribution used here, log-Poisson with $D_\infty = 1$ and $h_\infty = -\frac{1}{2}$, it is $\sigma_{|\eta|} = \sqrt{2^{-\frac{3}{4}} - \frac{1}{2}} = 0.31$, and so the standard deviation of Q goes as $0.62\, N^{-\frac{1}{2}}$. As shown in Appendix Appendix B, the estimation of the mutual information I has also a standard deviation depending on $N^{-\frac{1}{2}}$, but the proportionality constant is $\sqrt{\omega}$, which in our log-Poisson distribution is 5.66 bits. In addition, we do not take into account other sampling uncertainties stated in Appendix Appendix B that do not depend on N. The absolute uncertainty for I is 12 times that of Q, although their typical values are more than an order of magnitude smaller. For these reasons, we have analyzed relative large ensembles (64 series of 4096 points each) to show that Q performs equally well as I for large ensembles, but Q has the potential to be useful for smaller ensembles.

6. Conclusions

In this paper we have discussed on the properties of optimal wavelet bases for the representation of multiplicative cascades, which are proper to many physical systems with turbulent-like behaviour. With the aid of optimal wavelet bases, any given signal originated by a cascade can be explicitly represented in terms of that multiplicative cascade. When the wavelet basis used in the analysis is suboptimal, the local cascade variables are poorly described. We have shown that the multiplicative process is perturbed by the inclusion of an additive, noise-like term, with an amplitude depending on the deviation from optimality of the studied basis.

We have then proposed to quantify the degree of optimality of a given basis with a simple descriptor Q, defined as the ratio of the first order moment of estimated cascade variable by the first order moment of the actual cascade variable. As this quantity is obtained from first-order moments, it is not demanding in data, and as any deviation implies an increase in the first order moment of the estimated cascade variable, the optimal wavelet is an absolute minimum of this quantifier. Hence, Q can be used in any minimization strategy to derive the optimal wavelet.

To exemplify the derivations on the behavior of Q, we have used 24 different standard wavelets to analyze synthetic cascades generated with these same 24 wavelets. Our experiences reveal that Q is more accurate than mutual informations in order to determine optimality in reduced datasets.

With the help of the theory settled in this paper, we can undertake a more ambitious program of research. On the technical side, a natural future research line consists on implementing a continuous optimization strategy based on the descriptor Q, so that we could derive optimal wavelets of given databases of real signals without restricting the search to given families of wavelet bases. For each system we could hence prove if there exists such an optimal wavelet basis and even if it does not exist, we will be able to derive the best one. This would improve our knowledge on the dynamics of the studied systems. On the side of applications, optimal cascade representations are useful for instance for the inference of missing data by imposing the cascade as a physical constraint – an example of this application for the inference of data gaps in ocean chlorophyll concentration is shown in Ref. 31. Another applications include data forecast and data compression. Finally, on a more physically sound side, cascades are useful to characterize the arrangement of a system, to show how fluxes (heat, material) take place across scales and to unveil any cyclic or regular structure not evident due to the intermittent character of cascades.

Appendix A. Connection of the microcanonical cascade with the multifractal singularity spectrum

A signal s is said to be multifractal in the microcanonical sense[54] if an intensive functional ϵ_r acting on this signal (see Section 3) can be characterized by local scaling relations of the type:

$$\epsilon_r(\vec{x}) \;=\; \alpha(\vec{x})\, r^{h(\vec{x})} \;+\; o\left(r^{h(\vec{x})}\right) \tag{A.1}$$

where the symbol $o\left(r^{h(\vec{x})}\right)$ means a term that is negligible in comparison with $r^{h(\vec{x})}$. The function that comprises the local properties of changes in scale, $h(\vec{x})$, is called the *singularity exponent* of the signal at the point $\vec{x}$.[48,54] A signal verifying eq. (A.1) is said "multifractal" (in the microcanonical sense) because each value h of singularity exponent is associated to a singularity component $F_h \equiv \{\vec{x} : h(\vec{x}) = h\}$ of fractal character, with Hausdorff dimension $D(h)$. The function $D(h)$ is known as the singularity spectrum of the signal.[13]

An interesting feature of the singularity spectrum is that although it is a geometrical feature of the multifractal, it completely defines the statistical properties of the cascade process. In fact, Parisi and Frisch[28] proved that the knowledge of $D(h)$ granted the knowledge of the distribution of the cascade variables η through the knowledge of the multiscaling exponents τ_p, as expressed by eq. (3.1). In that case, τ_p is related to the singularity spectrum of the multifractal through a Legendre transform:

$$\tau_p = \inf_h \{ph + d - D(h)\} \tag{A.2}$$

which is known as the Parisi-Frisch formula and is the cornerstone of the canonical multifractal formalism. An interesting corollary of eq. (A.2) is that when $D(h)$ is convex the Legendre transform can be inverted and hence $D(h)$ can be expressed as the Legendre transform of the multiscaling exponents τ_p, namely:

$$D_L(h) = \inf_p \{ph + d - \tau_p\} . \tag{A.3}$$

The function $D_L(h)$ is the so-called Legendre singularity spectrum, which is a convex function of h because Legendre transforms are always convex. If $D(h)$ is convex, $D(h) = D_L(h)$; if $D(h)$ is not convex, $D_L(h)$ will be its convex hull.

There is a more direct approach to $D(h)$ that can be used when the cascade variables are accessible and eliminates the necessity of imposing convex spectra $D(h)$. This approach consists in calculating the limit as $\kappa \to 0$ of the distribution of cascade singularity exponents. The cascade singularity exponents are defined as follows:

$$h_\kappa \;=\; \log_\kappa \eta_\kappa \;=\; \frac{\ln \eta_\kappa}{\ln \kappa} \tag{A.4}$$

where η_κ is the multiplicative cascade variable that relates ϵ_r with ϵ_L, $\kappa = r/L$, as in eq. (3.4). The cascade singularity exponents represent the singularity exponents in the same sense of eq. (A.1) when they are obtained at the resolution level,[54] i.e., when the scale ratio κ is the one that compares the largest (whole-domain wide) scale L with the smallest (resolution-level) scale r, meaning that $r << L$ or equivalently $\kappa \to 0$. As the singularity components F_{h_κ} are of fractal character, the distribution of singularity exponents at a given observation scale behaves as[13]:

$$\rho(h_\kappa) \sim \kappa^{d-D(h_\kappa)} \tag{A.5}$$

with, as stated, $\kappa \to 0$. A direct obtaining of the $D(h)$ is hence possible through:

$$\lim_{\kappa\to 0} \frac{\ln \rho(h_\kappa)}{\ln \kappa} = d - D(h) \tag{A.6}$$

where:

$$h \equiv h_0 = \lim_{\kappa\to 0} h_\kappa . \tag{A.7}$$

Lemma: The singularity spectrum derived according eq. (A.6) coincides with the Legendre spectrum, eq. (A.3), when the singularity spectrum is convex

Proof: First, we define a random variable h_κ such that $\eta_\kappa = \kappa^{h_\kappa}$, i.e.,

$$h_\kappa = \frac{\ln \eta_\kappa}{\ln \kappa} . \tag{A.8}$$

As the cascade variable η_κ is derived from a multifractal signal, the limit in eq. (A.6) exists and it is $d - D(h)$ (the Hausdorff spectrum of the signal).[54] Therefore, the distribution of h_κ has a leading order $\kappa^{d-D(h_\kappa)}$ as follows:

$$\rho(h_\kappa) = A_\kappa \kappa^{d-D(h_\kappa)} + o(\kappa^{d-D(h_\kappa)}) \tag{A.9}$$

for small values of κ. Recalling here eq. (3.3) we have:

$$\tau_p = \lim_{\kappa\to 0} \frac{\ln \langle \eta_\kappa^p \rangle}{\ln \kappa} . \tag{A.10}$$

We then expand it to find that:

$$\begin{aligned}\tau_p &= \lim_{\kappa\to 0}\frac{1}{\ln\kappa}\ln\left(\int \mathrm{d}h_\kappa\,\kappa^{h_\kappa p}\rho(h_\kappa)\right) &&= \lim_{\kappa\to 0}\frac{1}{\ln\kappa}\ln\left(\int \mathrm{d}h_\kappa\,\kappa^{h_\kappa p}A_\kappa\kappa^{d-D(h_\kappa)}\right)\\ &= \lim_{\kappa\to 0}\inf_{h_\kappa}\{h_\kappa p + d - D(h_\kappa)\} &&= \inf_h\{hp + d - D(h)\} \qquad \text{(A.11)}\end{aligned}$$

where we used the saddle-point approximation. Notice that eq. (A.11) is analogous to eq. (A.2). Recalling that the inverse of a Legendre transform on convex functions is another Legendre transform, if we obtain now the Legendre spectrum, eq. (A.3), and assuming that $D(h)$ is convex we conclude $D_L(h) = D(h)$, q.e.d.

We will show now two examples of the lemma above, for two commonly used multiplicative processes, namely log-normal and log-Poisson processes. A log-normal process has the following distribution:

$$\rho(\ln\eta_\kappa) = \frac{1}{\sqrt{2\pi\sigma_\kappa^2}}e^{-\frac{1}{2}\left(\frac{\ln\eta_\kappa-\mu_\kappa}{\sigma_\kappa}\right)^2}. \qquad \text{(A.12)}$$

Hence, the τ_p as defined in eq. (3.1) are given by:

$$\tau_p = \frac{\mu_\kappa}{\ln\kappa}\,p + \frac{\sigma_\kappa^2}{2\ln\kappa}\,p^2. \qquad \text{(A.13)}$$

Let $h_m = \mu_\kappa/\ln\kappa$ and $\sigma_h^2 = -2\sigma_\kappa^2/\ln\kappa$ (remember that $\kappa < 1$), so eq. (A.3) leads to the singularity spectrum $D(h)$:

$$D(h) = d - \left(\frac{h-h_m}{\sigma_h}\right)^2. \qquad \text{(A.14)}$$

Let us show now that eq. (A.6) leads to the same expression. Notice that eq. (A.4) means that $\rho(h_\kappa) = -\ln\kappa\,\rho(\ln\eta_\kappa)$. Then, we substitute $\mu_\kappa = h_m\ln\kappa$ and $\sigma_\kappa^2 = -\sigma_h^2\,\frac{\ln\kappa}{2}$ in eq. (A.12) to obtain:

$$\frac{\ln\rho(h_\kappa)}{\ln\kappa} = \left(\frac{h-h_m}{\sigma_h}\right)^2 + \frac{\ln\sqrt{\frac{-\ln\kappa}{\pi\sigma_h^2}}}{\ln\kappa} \qquad \text{(A.15)}$$

and the second term vanishes as $\kappa\to 0$ leading to eq. (A.14). It follows that eq. (A.6) holds.

The log-Poisson case is a little bit more elaborated due to the discrete-to-continuous passage. A log-Poisson process is defined as $\eta_\kappa = \kappa^{h_\infty}\beta^n$ with n being a Poisson variable of parameter λ. Then the distribution of $\ln\eta_\kappa$ is:

$$\rho(\ln\eta_\kappa) = \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}\,\delta(\ln\eta_\kappa - h_\infty\ln\kappa - n\ln\beta) \qquad \text{(A.16)}$$

which is discrete, i.e., it only takes nonzero values for some values of $\ln\eta_\kappa$. The parameter h_∞ is the singularity exponent of the Most Singular Component (MSC),[48,52] while the parameter λ is related to the dimension of the MSC:

$\lambda = (d - D_\infty)(-\ln\kappa)$ (both parentheses are always positive). It is also required that $0 < \beta < 1$. After some simple algebra, it is obtained that τ_p are given by:

$$\tau_p = ph_\infty + (d - D_\infty)(1 - \beta^p) \tag{A.17}$$

and through eq. (A.3) the singularity spectrum is:

$$D(h) = D_\infty + (d - D_\infty)\,\omega(h)\,(1 - \ln\omega(h)) \tag{A.18}$$

with

$$\omega(h) = -\frac{1}{\ln\beta}\frac{h - h_\infty}{d - D_\infty}. \tag{A.19}$$

Let us now apply eq. (A.6). From equations (A.4) and (A.16), the h_κ deviates from the most singular exponent h_∞ in an integer number n of contributions $\log_\kappa \beta$, namely

$$h_\kappa = h_\infty + \underbrace{n\frac{\ln\beta}{\ln\kappa}}_{\Delta h_\kappa} \tag{A.20}$$

which give rise to a continuum of h in the limit $(-\ln\kappa) \to \infty$. Let us now define a convenient auxiliary variable, $\omega(h_\kappa)$, as

$$\omega(h_\kappa) = \frac{n}{\lambda} = \frac{1}{d - D_\infty}\frac{n}{(-\ln\kappa)} = -\frac{1}{\ln\beta}\frac{h_\kappa - h_\infty}{d - D_\infty}. \tag{A.21}$$

Notice that $\omega(h_\kappa)$ is positive and proportional to Δh_κ. We now recall eq. (A.16) to obtain:

$$\frac{\ln\rho(h_\kappa)}{\ln\kappa} = \frac{-\lambda + n\ln\lambda - \ln n!}{\ln\kappa}. \tag{A.22}$$

Hence, according to eq. (A.6), the singularity spectrum is:

$$D(h) = d - (d - D_\infty) + \lim_{\kappa\to 0}\frac{n\ln\lambda - \ln n!}{-\ln\kappa} \tag{A.23}$$

where $h = h_{\kappa\to 0}$ as in eq. (A.7). For any h_κ different from h_∞, i.e., $\Delta h_\kappa \neq 0$, when κ goes to 0, n grows accordingly, because n is proportional to $(-\ln\kappa)$. So the limit $\kappa \to 0$ is equivalent to $n \to \infty$:

$$D(h) = D_\infty + \lim_{n\to\infty}\frac{n\ln\lambda - n\ln n + n - \ln(\sqrt{2\pi n})}{-\ln\kappa} \tag{A.24}$$

where we have used the Stirling approximation to expand $n!$. Recalling $(-\ln\kappa) = n\left((d - D_\infty)\,\omega(h_\kappa)\right)^{-1}$ we have:

$$D(h) \;=\; D_\infty + (d - D_\infty)\lim_{n\to\infty}(\ln\lambda - \ln n + 1)\,\omega(h_\kappa) \tag{A.25}$$

which, as $\omega(h_\kappa) = n/\lambda$, leads to eq. (A.18).

Appendix B. Convergence of the estimates of the mutual information

In this section we calculate the standard deviation of the empirical estimates of the mutual information between two random variables. The derivation presented here is similar to the ones presented in Ref. 24, 2. The mutual information between two variables X and Y is given by the following expression:

$$I(X,Y) \;=\; H(X) + H(Y) - H(X,Y) \tag{B.1}$$

where $H(X)$ and $H(Y)$ stand for the marginal entropies and $H(X,Y)$ stands for the joint entropy. The ideal joint entropy is given by:

$$H(X,Y) \;=\; -\sum_{n,m} p_{nm} \ln p_{nm}. \tag{B.2}$$

Let us suppose that we sample the state space with a histogram of $B_x \times B_y$ boxes. We will assume that the sampling is efficient, so each box contains only one p_{nm} at most. Empty boxes, in case they exist, can directly be discarded as they do not change anything in the calculation, so without loss of generality we can assume that the sampling is perfect, and each box contains one and only one weight p_{nm}, so we can index boxes according to the weight index: B_{nm}. In practical cases, B_x and B_y are finite; we will assume that they are large enough to make the contribution by the uncounted tails negligible.

For a sample of N independent events we estimate p_{nm} with the following statistical:

$$\hat{p}_{nm} \;=\; \frac{N_{nm}}{N} \tag{B.3}$$

where N_{nm} is the number of events happening to lie in box B_{nm}. The joint distribution of the variables N_{nm} is a multinomial of order N with $B_x \times B_y$ variables, each with probability p_{nm}. Let us introduce a convenient representation for $\hat{p}_{nm}$:

$$\hat{p}_{nm} \;=\; p_{nm} \;+\; \delta p_{nm} \tag{B.4}$$

where we can express the deviation δp_{nm} in the following way:

$$\delta p_{nm} \;=\; \sqrt{\frac{p_{nm}(1-p_{nm})}{N}}\,\epsilon_{nm} \tag{B.5}$$

where the random variable ϵ_{nm} is standardized (i.e., it has zero mean and unit variance).

The estimated joint entropy $\hat{H}(X,Y)$ is hence given by:

$$\begin{aligned}\hat{H}(X,Y) &= -\sum_{n,m} \hat{p}_{nm} \ln \hat{p}_{nm} \\ &= H(X,Y) - \sum_{n,m} \delta p_{nm} \ln p_{nm} - \sum_{n,m} (p_{nm} + \delta p_{nm}) \ln\left(1 + \frac{\delta p_{nm}}{p_{nm}}\right)\end{aligned} \tag{B.6}$$

where the third term cancels in the first order expansion in δp_{nm} because $\sum_{n,m} \delta p_{nm} = 0$ (and it is $-\frac{1}{2}\sum_{n,m} \delta p_{nm}^2/p_{nm}$ in the second order). Therefore,

in the first order in δp_{nm}, the deviation between the estimate of the joint entropy and its actual value is given by the following expression:

$$\delta H(X,Y) \equiv \hat{H}(X,Y) - H(X,Y) \approx -\sum_{n,m} \ln p_{nm} \sqrt{\frac{p_{nm}(1-p_{nm})}{N}}\, \epsilon_{nm}. \tag{B.7}$$

To calculate this sum, we need to consider the correlations between the different ϵ_{nm} (which coincide with those of $\hat{p}_{nm}$, i.e., the multinomial). These correlations are $C(nm, n'm') = -\sqrt{p_{nm}/(1-p_{nm})\; p_{n'm'}/(1-p_{n'm'})}$. We have:

$$\delta H(X,Y) \;=\; \sqrt{\frac{\omega_{XY}}{N}}\, \epsilon \tag{B.8}$$

where ϵ is a standardized variable and ω_{XY} is given by:

$$\omega_{XY} \;=\; \sum_{n,m} p_{nm} (\ln p_{nm})^2 - \left(\sum_{n,m} p_{nm}\, \ln p_{nm} \right)^2 . \tag{B.9}$$

The standard deviation of $\hat{H}(X,Y)$ coincides with that of $\delta H(X,Y)$ and it is $\sqrt{\omega_{XY}}/\sqrt{N}$.

Analogous expressions arise for $\delta H(X)$ and $\delta H(Y)$, having their corresponding ω_X and ω_Y respectively. Hence, the deviation of the mutual information estimate $\delta I(X,Y)$ is given by the squared sum of the deviations of the joint and marginal entropies, with a global ω that is the sum of the joint ω_{XY} and marginal ω_X and ω_Y. Thus we can estimate the minimum number of samples N_0 to attain a given accuracy level $\delta I(X,Y)$ according to the following expression:

$$N_0 \;=\; \frac{\omega}{\delta I(X,Y)^2}. \tag{B.10}$$

The dependence on the square of the accuracy level makes entropy estimation very demanding in data. For instance, to attain an accuracy of 0.1 bits ($\approx$ 0.07 nats) we have $N_{0.1} \approx 200\,\omega$; to attain an accuracy of 0.01 bits we need a sample 100 times larger, $N_{0.01} \approx 2 \cdot 10^4\,\omega$. For the case studied in section 5, 4096-point series generated with log-Poisson distribution of parameters $D_\infty = 0$ and $h_\infty = -\frac{1}{2}$, the computed value of ω is around $\omega = 15.4$ nats^2.

As a final remark, notice that we have made important assumptions to derive this formula. The two most significant ones depend on the properties of the sampling using $B_x \times B_y$ boxes. First, we have assumed that we have properly sampled the histogram; second, we considered that the non-sampled tails do not significantly contribute to uncertainty. Concerning the first, we are assuming that the sample of the state space with $B_x \times B_y$ boxes is such that the associated weights $\{p_{nm}\}$ give an accurate idea of the mutual information; for instance, if X and Y are independent then that with a good approximation $p_{nm} = p^x_n\, p^y_m$. Concerning the second, we need to assume that the excluded tails decay fast enough not to significantly alter the value of the entropies. These two contributions will increase the dispersion δH estimated here in a way that does not depend on N, so the mutual information will

never be decreased below a certain level even if N goes to infinity. These sampling effects are absolutely depending on the distribution we are considering and hence no a priori bound can be given here.

Acknowledgments

This work is a contribution to OCEANTECH (PIF 2006 Project) and FIS2006-13321-CO2-01.

References

1. Arneodo, A., Bacry, E., and Muzy, J., Random cascades on wavelet dyadic trees, *Journal of Mathematical Physics* **39** (1998) 4142–4164.
2. Bazso, F., Zalanyi, L., and Petroczi, A., Accuracy of joint entropy and mutual information estimates, in *Proceedings of IEEE International Joint Conference on Neural Networks,*, Vol. 4 (IEEE, 2004), pp. 2843–2846.
3. Benzi, R., Biferale, L., Crisanti, A., Paladin, G., Vergassola, M., and Vulpiani, A., A random process for the construction of multiaffine fields, *Physica D* **65** (1993) 352–358.
4. Buccigrossi, R. W. and Simoncelli, E. P., Image compression via joint statistical characterization in the wavelet domain, *IEEE Transactions in Image Processing* **8** (1999) 1688–1701.
5. Carleman, T., Sur le problème des moments, *Comptes rendus Acad. Sci. Paris* **174** (1922) 1680.
6. Castaing, B., The temperature of turbulent flows, *Journal de Physique II* **6** (1996) 105–114.
7. Chainais, P., Infinitely divisible cascades to model the statistics of natural images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29** (2007) 2105–2119.
8. Chen, D.-R., A construction of convergent cascade algorithms in sobolev spaces, *International Journal of Wavelets, Multiresolution and Information Processing* **5** (2007) 685–697.
9. Chhabra, A., Meneveau, C., Jensen, R., and Sreenivasan, K., Direct determination of the f(*alpha*) singularity spectrum and its application to fully developed turbulence, *Physical Review A* **40** (1989) 5284–5294.
10. Daubechies, I., *Ten lectures on wavelets*, CBMS-NSF Series in App. Math. (Capital City Press, Montpelier, Vermont, 1992).
11. Delgado, J., Turiel, A., and Parga, N., Receptive fields of simple cells from a taxonomic study of natural images and suppresion of scale redundancy, *Neurocomputing* **69** (2006) 1224–1227.
12. Dubrulle, B., Intermittency in fully developed turbulence:log-poisson statistics and generalized scale covariance, *Physical Review Letters* **73** (1994) 959–962.
13. Falconer, K., *Fractal Geometry: Mathematical Foundations andApplications* (John Wiley and sons, Chichester, 1990).
14. Frisch, U., *Turbulence: The legacy of A.N. Kolmogorov* (Cambridge Univ. Press, Cambridge MA, 1995).
15. Huang, J. and Mumford, D., Statistics of natural images and models, in *Proc. CVPR* (1999), pp. 541–547.
16. Jansen, M., Refinement independent wavelets for use in adaptive multiresolution schemes, *International Journal of Wavelets, Multiresolution and Information Processing* **6** (2008) 521–539.

17. Lashermes, B., Abry, P., and Chainais, P., New insights into the estimation of scaling exponents, *International Journal of Wavelets, Multiresolution and Information Processing* **2** (2004) 497–523.
18. Lovejoy, S., Curri, W., Tessier, Y., Claereboudt, M., Bourget, E., Roff, J., and Schertzer, D., Universal multifractals and ocean patchiness: phytoplankton, physical fields and coastal heterogeneity, *J. Plankt. Res.* **23** (2001) 117–141.
19. Mabrouk, A. B., Study of some nonlinear self-similar distributions, *International Journal of Wavelets, Multiresolution and Information Processing* **5** (2007) 907–916.
20. Mallat, S., A theory for multiresolution signal decomposition : the wavelet representation, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **11** (1989) 674–693.
21. Mallat, S., *A Wavelet Tour of Signal Processing* (Academic Press, 2nd Edition, 1999).
22. Mallat, S. and Zhong, S., Wavelet transform maxima and multiscale edges, in *Wavelets and their applications*, ed. et al, M. B. R. (Jones and Bartlett, Boston, 1991).
23. Mandelbrot, B., Fisher, A., and Calvet, L., A multifractal model of asset returns, *Cowles Foundation Discussion Paper No. 1164* (1997).
24. Moddemeijer, R., A statistic to estimate the variance of the histogram-based mutual information estimator based on dependent pairs of observations, *Signal Processing* **75** (1999) 51–63.
25. Muzy, J. F., Delour, J., and Bacry, E., Modelling fluctuations of financial time series: from cascade process to stochastic volatility model., *Euro. Phys. Journal B* **17** (2000) 537–548.
26. Muzy, J.-F., Sornette, D., Delour, J., and Arneodo, A., Multifractal returns and hierarchical portfolio theory, *Quantitative Finance* **1** (2001) 131–148.
27. Novikov, E. A., Infinitely divisible distributions in turbulence, *Physical Reviev E* **50** (1994) R3303.
28. Parisi, G. and Frisch, U., On the singularity structure of fully developedturbulence, in *Turbulence and Predictability in Geophysical Fluid Dynamics. Proc. Intl. School of Physics E. Fermi*, eds. Ghil, M., Benzi, R., and Parisi, G. (North Holland, Amsterdam, 1985), pp. 84–87.
29. Pont, O., Turiel, A., and Perez-Vicente, C., Empirical evidences of a common multifractal signature in economic, biological and physical systems, *Physica A* **388** (2009) 2025–2035.
30. Pont, O., Turiel, A., and Pérez-Vicente, C. J., Description, modeling and forecasting of data with optimal wavelets, *Journal of Economic Interaction and Coordination* **4** (2009) 39–54.
31. Pottier, C., Turiel, A., and Garçon, V., Inferring missing data in satellite chlorophyll maps using turbulent cascading, *Remote Sensing of Environment* **112** (2008) 4242–4260.
32. Riedi, R. H., Crouse, M. S., Ribeiro, V. J., and Baraniuk, R. G., A multifractal wavelet model with application to network traffic, *IEEE Transactions on Information Theory* **45** (1999) 992–1018.
33. Roux, S. G., Arneodo, A., and Decoster, N., A wavelet-based method for multifractal image analysis. iii. applications to high-resolution satellite images of cloud structure, *Eur. Phys. J. B* **15** (2000) 765–786.
34. Rudin, W., *Real and Complex Analysis* (Mc Graw Hill, New York, USA, 1987).
35. Sachs, D., Lovejoy, S., and Schertzer, D., The multifractal scaling of cloud radiances from 1 m to 1 km, *Fractals* **10** (2002) 1–12.
36. Schertzer, D. and Lovejoy, S., Physically based rain and cloud modeling by anisotropic, multiplicative turbulent cascades, *Journal of Geophysical Research* **92** (1987) 9692–

9714.
37. Scheuring, I. and Riedi, R., Application of multifractals to the analysis of vegetation pattern, *Journal of Vegetation Science* **5** (1994) 489–496.
38. Schmitt, F., Schertzer, D., and Lovejoy, S., Multifractal fluctuations in finance, *International Journal of Theoretical and Applied Finance* **3** (2000) 361–364.
39. Schmitt, F. G. and Chainais, P., On causal stochastic equations for log-stable multiplicative cascades, *European Physical Journal B* **58** (2007) 149–158.
40. Schwartz, O. and Simoncelli, E., Natural signal statistics and sensory gain control, *Nature Neuroscience* **4** (2001) 819–825.
41. She, Z., Towards a complex system approach for the study of turbulence, *Chemical Engineer Science* **62** (2007) 3595–3604.
42. She, Z. S. and Leveque, E., Universal scaling laws in fully developed turbulence, *Physical Review Letters* **72** (1994) 336–339.
43. She, Z. S. and Waymire, E. C., Quantized energy cascade and log-poisson statistics in fully developed turbulence, *Physical Review Letters* **74** (1995) 262–265.
44. Tessier, Y., Lovejoy, S., and Schertzer, D., Universal multifractals in rain and clouds: theory and observations, *Journal of Applied Meteorology* **32** (1993) 223–250.
45. Turiel, A., Delgado, J. M., and Parga, N., Learning efficient internal representations from natural image collections, *Neurocomputing* **58-60** (2004) 915–921.
46. Turiel, A., Isern-Fontanet, J., García-Ladona, E., and Font, J., Multifractal method for the instantaneous evaluation of the stream function in geophysical flows, *Physical Review Letters* **95** (2005) 104502.
47. Turiel, A., Mato, G., Parga, N., and Nadal, J. P., The self-similarity properties of natural images resemble those of turbulent flows, *Physical Review Letters* **80** (1998) 1098–1101.
48. Turiel, A. and Parga, N., The multi-fractal structure of contrast changes innatural images: from sharp edges to textures, *Neural Computation* **12** (2000) 763–793.
49. Turiel, A. and Parga, N., Multifractal wavelet filter of natural images, *Physical Review Letters* **85** (2000) 3325–3328.
50. Turiel, A. and Pérez-Vicente, C., Multifractal geometry in stock market time series, *Physica A* **322** (2003) 629–649.
51. Turiel, A. and Pérez-Vicente, C., Role of multifractal sources in the analysis of stock market time series, *Physica A* **355** (2005) 475–496.
52. Turiel, A., Pérez-Vicente, C., and Grazzini, J., Numerical methods for the estimation of multifractal singularity spectra on sampled data: a comparative study, *Journal of Computational Physics* **216** (2006) 362–390.
53. Turiel, A., Solé, J., Nieves, V., Ballabrera-Poy, J., and García-Ladona, E., Tracking oceanic currents by singularity analysis of micro-wave sea surface temperature images, *Remote Sensing of Environment* **112** (2008) 2246–2260.
54. Turiel, A., Yahia, H., and Pérez-Vicente, C., Microcanonical multifractal formalism: a geometrical approach to multifractal systems. Part I: Singularity analysis, *Journal of Physics A* **41** (2008) 015501.
55. Wainwright, M. J., Simoncelli, E. P., and Willsky, A. S., Random cascades on wavelet trees and their use in modeling and analyzing natural images, *Applied Computational and Harmonic Analysis* **11** (2001) 89–123.