

A microcanonical cascade  
formalism for multifractal  
systems and its application  
to data inference and  
forecasting

Oriol Pont





# **A microcanonical cascade formalism for multifractal systems and its application to data inference and forecasting**

**Oriol Pont i Pla**

Departament de Física Fonamental  
Universitat de Barcelona

Supervised by:

Conrad J. Pérez Vicente  
Antonio M. Turiel Martínez

Graduate programme in Advanced Physics 2004-2006



*A l'avi Joan*



# Abstract

Many complex systems in Nature are multifractal, a feature closely related to scale invariance. Multifractality is ubiquitous and so it can be found in systems as diverse as marine turbulence, econometric series, heartbeat dynamics and the solar magnetic field. In recent years, there has been growing interest in modelling the multifractal structure in these systems. This has improved our understanding of certain phenomena and has opened the way for applications such as reduction of coding redundancy, reconstruction of data gaps and forecasting of multifractal variables.

Exhaustive multifractal characterization of experimental data is needed for tuning parameters of the models. The design of appropriate algorithms to achieve this purpose remains a major challenge, since discretization, gaps, noise and long-range correlations require advanced processing, especially since multifractal signals are not smooth: due to scale invariance, they are intrinsically uneven and intermittent.

In the present study, we introduce a formalism for multifractal data based on microcanonical cascades. We show that with appropriate selection of the representation basis, we greatly improve inference capabilities in a robust fashion. In addition, we show two applications of microcanonical cascades: first, forecasting of stock market series; and second, detection of interscale heat transfer in the ocean.





## Acknowledgements

Many people deserve thanks for their contribution, help and support. I am grateful to everyone who has helped in this process, either directly or indirectly. Some of them may have been omitted from the following list: thank you all, I have you all in mind.

First, I want to thank my advisers, coworkers and friends: Conrad and Antonio. They have inspired, proposed, guided, supported and contributed a lot to this work. I have learned a lot from their knowledge and expertise, and this thesis would simply not exist without them. They have decisively helped me start a scientific career, God willing.

I also owe a big thank you to all of the early inspirers of this work, back to the days of high school, degree and early postgraduate: Ferran, Maite, David, Aurora, José María, Eduard, Rosario, M<sup>a</sup>Àngels, Ignasi, Alfred, Josep, Miquel, Jaume, Adolf, and José Ignacio. I have learned a lot from them and their advice has been a great help.

This interdisciplinary work has brought me into contact with world experts in different fields, especially statistical physics, game theory and econophysics. In particular, Esteban, Anxo, José, Raúl, Elka, Pedro, Romualdo, Jordi, Álvaro, Juanma, Miguel Ángel, Maxi, Raúl, Emilio, Mario, Francisco, Juanjo, Carmen, Rosario, Fabrizio, Ingve, Milan, Ryszard and Janusz. Both in congresses and in private communications they have always treated me as an equal, even when their

expertise was greatly superior to mine: thank you for this and for your discussion and advice.

Additionally, I am grateful to all the former and current members of the Complex Systems group at UB and URV, and in particular, to Albert, Àlex, Ramon, Juan, Sergio, Marian, Marian [*sic*], Leon, Jordi, Sergi, Javier and Luce. Also, from ICM and collaborators, I thank Emili, Jordi, Vero, Jose, Manel, Marta, Hussein and Veronique. Working with them in discussions and taking their advice has helped me a lot during these years. Many office colleagues not mentioned above also deserve acknowledgement: Ramon, Miquel, Manel, Majo, Raül, Hicham and Dani, for their good working environment, keeping company and their help in practical situations.

Support from friends has also been of great importance. I thank all of them, be they *amics*, *coneguts o saludats*, but very especially, Arnau, Miquel, Martí, David, Víctor, Josep, Mendeli, Pau, Toni, Mercè, Marga and Gemma.

The help from my family has been of vital importance. Their unique support, encouragement and understanding, especially in the final stages of production, have contributed decisively to make this thesis possible: an enormous thank you to all the aunts and uncles, cousins, brother-in-law and, very especially, grandparents, parents and sister. And last and most of all, thank you to Alexandra for sharing my dreams and hopes.

# Contents

<b>1</b>	<b>An introduction to multiple scales</b>	<b>1</b>
1.1	Research context of multifractal analysis . . . . .	9
1.2	Thesis outline . . . . .	15
<b>2</b>	<b>Multiresolution analysis</b>	<b>18</b>
2.1	Multiresolution decomposition . . . . .	20
2.2	Wavelets: a basis for multiscale functions . . . . .	25
2.2.1	Settings for a dyadic representation . . . . .	27
2.2.2	Norm-1 wavelet transform . . . . .	30
2.2.3	Quadrature mirror filters (QMFs) . . . . .	31
2.2.4	QMF filters from the wavelet coefficients . . . . .	36
<b>3</b>	<b>Microcanonical multiplicative cascades</b>	<b>41</b>
3.1	Persistence in scale invariant signals . . . . .	41
3.1.1	Towards an optimal representation of data . . . . .	48
3.2	Optimization from suboptimal representations . . . . .	60
3.3	A numerical validation . . . . .	72
<b>4</b>	<b>Cascades are multifractal</b>	<b>82</b>
4.1	Singularity scaling . . . . .	82
4.2	Cascade-singularity connection . . . . .	83
4.3	Regularization of diverging measures . . . . .	95
<b>5</b>	<b>Application to stock market series</b>	<b>102</b>

5.1	Description of the data and notation . . . . .	104
5.2	Scale invariant properties of the stock market cascade	105
5.2.1	Optimal wavelet for Spanish stock market series	109
5.3	Cascade-induced distribution of quotation values . . .	114
5.3.1	Settings . . . . .	114
5.3.2	The model . . . . .	117
5.3.3	Conditioned distribution and maximum likelihood forecasting . . . . .	122
5.4	A cascade inference for time series: discussion . . . .	124
<b>6</b>	<b>Application to oceanographic data</b>	<b>126</b>
6.1	Sea Surface Temperature (SST) . . . . .	127
6.2	Wavelets in 2D signals . . . . .	128
6.3	Optimal wavelet for SST . . . . .	131
6.4	Local anomalies . . . . .	133
<b>7</b>	<b>Conclusions</b>	<b>136</b>
7.1	Future work . . . . .	140
<b>A</b>	<b>QMF and biorthogonal constraints</b>	<b>145</b>
A.1	Biorthogonal generalization . . . . .	153
A.2	A possible wavelet dual . . . . .	159
	<b>Bibliography</b>	<b>164</b>

עָתָה בּוֹא כְתֹבָה עַל-לוּחַ אֲפָם וְעַל-  
סֵפֶר חֶקֶה וּתְהִי לְיוֹם אַחֲרוֹן לְעַד עַד-  
עוֹלָם:

*“Now therefore go in and write  
for them upon box, and note it  
diligently in a book, and it shall  
be in the latter days for a testi-  
mony for ever”*

— Book of Isaiah



# 1 An introduction to multiple scales

When gazing at our environment, we see complex phenomena all around us: either natural or artificial, ranging from the tiny vacuum fluctuations and elementary particles to the huge galactic structures. Nonlinear interactions between elementary constituents make complexity emerge. Since the Dawn of Mankind, description of complexity has been a challenge and a stimulus to human creativity. Poets, for example, have made the most of it as a source of inspiration. The role of researchers is to observe and systematically analyze each delimited phenomenon in order to obtain schematic representations of them. Such representations are called *models* – or *theories*, in a more general way – and they help us to understand how things work: they describe their regular behaviour, anticipate changes and even, in some cases, give us the ability to control those changes and take advantage of them. When this final stage of research is reached, the phenomenon is no longer matter for the researcher: it passes to the domain of technological development.

Human knowledge has traditionally been divided into different disciplines: chemistry, biology, geology, economy, sociology... and a long list of *et cæteras*. In the early stage, these areas were evolving as a whole combining empirical and theoretical facts with challenges and findings that allowed progress in them, thus making the borders between them vague. Later on, these fields became more and more specialized and from the middle of the nineteenth century there

was a clear separation (except in special cases) in contents and rate of progress. Nowadays, we are facing a *back-to-the-basics* process. Some systems in these different areas turn out to follow similar laws and equations, which is opening the way to interdisciplinary research in the now fuzzy borders between the various scientific disciplines.

The study of scale-invariant systems is certainly one of these cases. A system is **scale invariant** if it looks similar at different scales. More precisely, scale invariance means that distribution and correlations of the system variables do not vary when the length of observation is expanded (or contracted), i.e., when we zoom in (or out). Scale invariance has been intensively studied during the last decades because it is present in very different systems: physics, oceanography, chemistry, geology, biology, economy, etc. As a consequence of scale invariance, it is possible to define an effective dynamics able to accurately describe the structure and evolution of the system, improving the classical knowledge provided by microscopic dynamics. The common fingerprint in these so varied systems is that certain variables follow power-law distributions. In some cases, the exponents of these power laws coincide, meaning that their effective dynamics are universal [1, 2]. This fact links the study of scale invariance and critical phenomena, which makes this subject so interesting.

The basic idea behind scale invariance is better understood with practical examples. In most cases, a system looks different depending on the scale at which it is observed. One of these cases is, e.g., the human skin: the scale of centimetres (our usual experience) is very different from the scale of micrometers, which we can see in a microscope. In the former, the skin is an even, thin tissue, while in the latter we distinguish layers of different width and we can see pores, wrinkles, furrows, etc. In contrast, the branches of a tree each resem-



ble the whole tree. Particularly, in cypresses a very similar shape is found from a scale of metres up to millimetres. This is a case of scale invariance between the whole-tree scale and the leaf scale. Also ferns have this property; we can see an example in Figure 1.1.

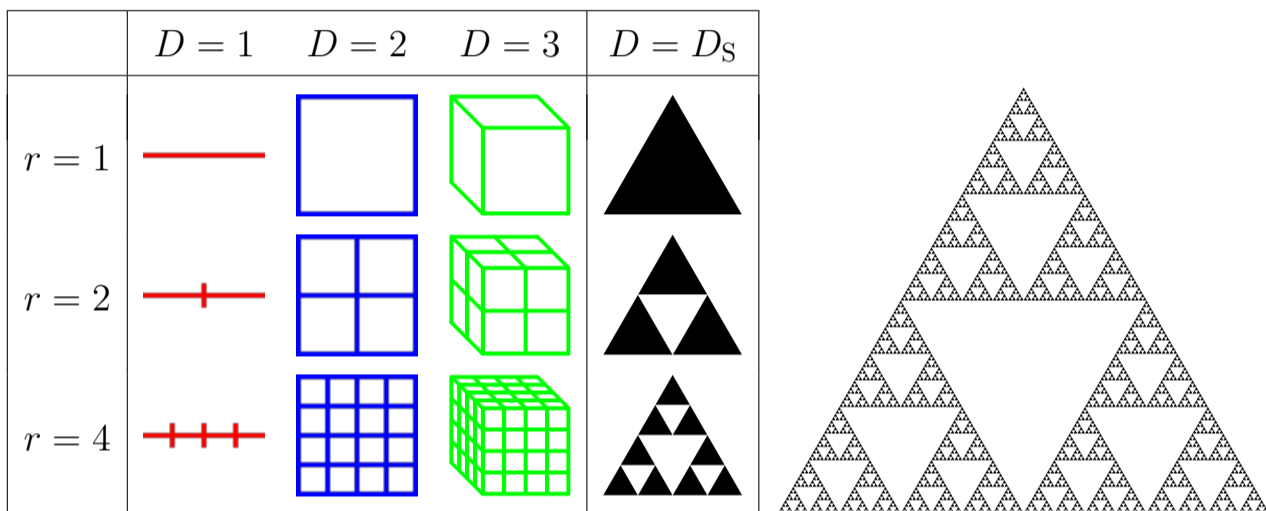


**Figure 1.1:** Triangular fern (left) and computer-generated fern (right). In the computer case, the fractal object is generated by reproducing at each stage a shrunk and symmetrized version of the whole object. Nature mimics a very similar dynamics when plant cells are arranged to form the real fern: blades start forming at both sides of a local symmetry axis and then they grow (by cell division) with new cells being arranged in the same way along new local symmetry axes, and so on in an iterative way. Photograph and computer art both by Rodd Halstead are used with permission of the author.

Schematically, this phenomenon has its archetype in a kind of geometrical figures called **fractals**. Fractals mimic the growth pattern of the fern: the object is formed by parts that reproduce the same shape of the whole object, but are smaller (they reproduce the structure at a smaller scale). As these parts reproduce the whole object, they are formed by subparts that reproduce themselves; iteratively, the same shape is reproduced at successive scales *ad infinitum*.

A more general variant of fractals is the case of *statistical fractals*, in which each downscaled replica is statistically equivalent to the whole, i.e., the structure follows random variables that have the same distribution at any scale. In a real fern, each subpart of the frond mostly resembles the whole. It is not an exact scaling of the same shape, due to randomness in cell growth and formation of the blades

(the lobes of the frond). Nevertheless, such randomness follows a fixed distribution that makes the structures formed statistically similar at all the scales, that is why they look alike. Statistical fractals are still a restrictive case of scale invariance: all the subparts of a given scale have a fixed size, which is proportionally related to that of the preceding scale. This proportionality ratio is governed by a parameter called fractal dimension  $D$ , see Figure 1.2.



**Figure 1.2:** An illustration of the concept of fractal dimension. We show four geometrical objects, namely a segment (1D), a square (2D), a cube (3D) and a fractal, at three different scales (left). This fractal is called Sierpinski gasket (shown at infinitesimal scale in the right side). The number of replicas that divide the original pattern  $N$  evolves with the scale  $r$  as  $N = r^D$  where  $D$  is the dimension. For the first three objects,  $D$  is their Euclidean dimension (1, 2 and 3, respectively). We extend the concept of dimension to include the scaling of the fractal, as it follows the same law  $N = r^D$ : in the case of the Sierpinski gasket, we can see that when the scale  $r$  doubles, there are three triangles ( $N = 3$ ), and when the scale is multiplied by four, there are nine triangles. Therefore, its dimension is  $D_S = \ln 9 / \ln 4 = \ln 3 / \ln 2 \approx 1.6$ , i.e., it is not an integer and its geometrical interpretation is not straightforward (it is sometimes said these fractal dimensions are *more than a curve* but *less than a surface*, as their support is 1D but they can only be embedded in a 2D space).

**Multifractals** are less restrictive than fractals and generalize them so that they cover most cases of scale invariance in Nature. Multifractal objects are formed by a continuum of fractal components, each one scaling with its own fractal dimension. Let  $h$  be a continuous variable indexing the fractal components. The set of fractal dimensions,  $D(h)$ , form a curve called *singularity spectrum* that characterizes all the scaling properties of the multifractal. Mandelbrot

identified the parameter  $h$  with the exponent of a local power law [3], i.e., given a multifractal function  $f(x) \sim x^{h(x)}$ , the exponent of the power law,  $h(x)$ , indicates the order of singularity (or regularity) of the function at the point  $x$ .

The description of multifractals based on singularity spectra is very abstract; we will return to this concept in Chapter 4, with all the formalism involved. For the time being, we will concentrate on the more intuitive framework of **multiplicative cascade processes** [4], often simply called **cascades**, which describe the multifractal structure of most real-world multifractals. In a cascade, the values of a scale-adjustable variable in consecutive scales are multiplicatively related by a random variable called *cascade variable*. The scale-adjustable variable can be a direct observable of the system or a measure or scale-tuned functional of another variable. The intuitive idea is presented in Section 3.1, and especially in Figure 3.1, while we discuss the technical details in Chapter 3. One of the advantages of the cascade representation is that it separates the multiple scales, thus all the representation variables are equivalent since the system is scale invariant.

The classical approach to multifractal analysis was a statistical one, based on the characterization of global quantities. This is called *canonical multifractal formalism*. Only recently, a new approach based on the characterization of local properties has been introduced: the *microcanonical multifractal formalism (MMF)* [5, 6]. By means of MMF, multifractal signals are realized as actual geometrical entities, different fractal components can be isolated and as a consequence new applications emerge. In particular, using an appropriate representation basis, cascades have a geometrical interpretation: these are the **microcanonical cascades**.

In this thesis, we attempt to enhance the capabilities of multifractal analysis applied to real data. We have achieved both theoretical and practical extension of the formalism to functions of non-total support or with diverging measures [7], we have established a connection with the formalism of multiplicative cascades in the microcanonical sense [8], we have found a robust criterion to obtain optimal wavelets [8], we have discovered evidence for universal behaviour in very different multifractal systems [9] and we have performed applications for the characterization and inference of multiscale data [10, 11, 12, 13]. These applications include: anomaly detection in stock trading, study of the effect of sources in financial time-series dynamics, characterization of direct correlations in asset portfolios, study of the effects of additive noise in multiplicative cascades, forecasting of future points distribution in econometric series [10], forecasting of volatility and characterization of tradewind-driven currents and frontogenesis in the ocean [11]. In addition, an open-source program (written in C) able to analyze local singularities and retrieve the singularity spectrum from them has been developed and is freely available.<sup>1</sup>

## 1.1 Research context of multifractal analysis

The state-of-the-art of multiscaling and multifractal models to describe physical systems can be traced back to the studies by Kolmogorov [14, 15], Obukhov [16], Yaglom [17, 18] and Kraichnan [19], applied to the context of fully developed turbulence. In these articles, the authors introduce multiscale descriptions that model the exponents of distribution moments for energy dissipation, velocity increments or gradients of advected scalars under a turbulent flow. Though not described in these terms, the proposed models envisaged

---

<sup>1</sup>[http://www.icm.csic.es/oce/people/turiel/SUPP\\_INFO/MF-analyzer.html](http://www.icm.csic.es/oce/people/turiel/SUPP_INFO/MF-analyzer.html)

a multifractal description. The first models were derived from Navier-Stokes equations and constructed simply with some phenomenological assumptions, but they later incorporated additional complexity to fit empirical facts [15]. In this context, Novikov worked in cascades that are infinitely divisible in the scale domain [20, 21], while Mandelbrot found that these models are explicitly describable as multifractals and studied their multiplicative cascade properties [3].

Those first ideas were additionally developed by various researchers and in different directions. Cascades and multifractals drifted and were bridged continuously, depending on the focus of research. While it is now widely accepted that in turbulence a real tangible cascade is the cause of the multifractal structure [22], in other systems it is still doubtful. In this context, works by Parisi and Frisch have been crucial in bridging the gap between the structure-function formalism and the multifractal-cascade formalism [23]. The Italian school of turbulence brought important developments in the modelling of turbulent flows; notably works by Benzi's and Vulpiani's groups [24, 25, 26, 27, 28, 29, 30]. These included one of the first *microcanonical cascades* [31], which was similar to the microcanonical cascades that we introduce in the present work. Later, further developments focused on the generalization of structure functions, and the use of wavelets allowed the French school to improve the models and the algorithms for the analysis of empirical data. These improvements include the studies by Arnéodo, Muzy, Bacry, Castaing, Dubrulle, Lévêque, Chainais and their groups [32, 33, 34, 35, 36, 37, 38, 39, 30, 40, 41, 42, 43, 44], among others.

In a different way, refinements to cascade formulations allowed Lovejoy, Schertzer, Seuront and their groups to propose multifractal universality classes [45, 46, 47, 48, 49, 50]. In this case, the focus of

research moved from lab turbulence to geophysical turbulent flows, namely the ocean and the atmosphere. In this context, wavelet analysis for characterization of multifractality is of crucial importance; a topic extensively reviewed by Davis *et al.* [51].

Multifractality has also been reported in systems that have turbulent-like behaviour though they have nothing to do with turbulent flows. For instance, Amaral, Goldberger, Ivanov, Struzik and Stanley have studied the multifractality in human gait and heartbeat dynamics [52, 53, 54], and Riedi and his group have studied multifractality in network traffic [55, 56, 57]. Buccigrossi, Simoncelli and their groups have characterized scale invariant properties of natural images from a statistical perspective [58, 59, 60]. Apart from this, financial time series are possibly the non-turbulent system where multifractality has been most intensively studied. Mandelbrot envisaged fractal models that explain the non-gaussianity of returns [61] and later revisited the topic with the *multifractal model of asset returns* with Calvet and Fisher [62, 63, 64]. Stanley and Mantegna also observed multifractality in econometric data [65, 66, 67, 68], while Arnéodo, Bacry, Muzy and their groups have also explored financial dynamics with multifractal models [69, 70, 71]. More recently, we could cite works by Perello based on the same ideas [72, 73]; in these models, multifractality is seen in a continuous-time dynamics and not directly in the signal itself.

As we have said, a multifractal is an ensemble of fractal components, discriminated by a hierarchy parameter  $h$  and each one having its own fractal dimension,  $D(h)$  (the *singularity spectrum*). In a multifractal function  $f(x)$ , the parameter  $h$  is not usually described as an abstract index but as the leading order of a local expansion at each point  $x$ ,  $f(x) \sim x^{h(x)}$ . This fact – the existence of local singularity

exponents,  $h(x)$  – is reflected in most of the multifractal formalisms [74, 75, 76, 77, 78, 79, 80]. However, efficient analysis of multifractal data based on singularity exponents has remained unexplored until recent years. A practical way to obtain these exponents from empirical data was not available until the improvement of analysis algorithms based on wavelet projection of gradient modulus measures, first achieved by Turiel *et al.* [5, 6]. These algorithms are appropriate for empirical data as they filter all the common artefacts that could arise due to discretisation, aliasing, noise, lack of stationarity, correlations, instabilities, and other problems related to the nature of real signals or to the numerical analysis of them. The milestone achieved is that this kind of multifractal characterization avoids the need to pass through the exponents of structure functions and the Parisi-Frisch transform [23], thus giving rise not only to a new analysis technique and a more direct characterization of the  $D(h)$  curve, but also to a new multifractal formalism called the **microcanonical multifractal formalism (MMF)** [5, 7, 81, 6].

Some of the recent applications of MMF to oceanography include works by Turiel and coworkers, e.g., Isern-Fontanet [81], Nieves [82] and Pottier [83]. Notably, these analyses allow them to obtain the ocean velocity field from temperature maps [84] and reconstruct data gaps in chlorophyll concentration maps [83]. There are also remarkable applications to atmospheric flows [85, 86, 87, 88] and vision [89, 90, 91, 92, 93, 94]. Turiel and Pérez-Vicente also applied MMF to characterize econometric time series: they studied the presence of a Markovian most singular component and sift this Markovian dynamics from another slowly changing one driven by information and capital injections [95, 96, 97]. This was the starting point of my thesis.

Before the presented study could be started, there were several theoretical, methodological and applied questions to be answered. In addition, the applications needed some previous theoretical and methodological developments. One of these questions was how the optimal wavelet – a concept developed by Turiel in [91] though already envisaged by Buccigrossi and Simoncelli in [58] – could define a microcanonical cascade compatible with MMF and whether this would provide a direct description of a cascade in a way that avoided the delocalization of the Parisi-Frisch transform. Another need was to improve the singularity analysis of MMF to overcome the problems that it faced when applied to fractal supports and diverging measures. Another methodological quest was to provide a measure of wavelet optimality that could be used to obtain optimal wavelets from empirical data. Regarding applications, it was not known which were the optimal wavelets for stock market series and geophysical variables such as sea surface temperature, and it was unclear whether their description as microcanonical cascades would improve knowledge of the structure and dynamics of these systems, or even provide a forecast of them. As we will see, this thesis has answered all these questions.

## 1.2 Thesis outline

This document is thematically structured and so it presents both the theory and the results of each topic in its respective chapter. Chapter 2 and Chapter 4 are mostly theoretical and they report analytical results, while Chapter 3 is more methodological and it introduces both analytical and numerical results. Finally, Chapter 5 and Chapter 6 develop the applications of the theory and methods to real multifractal systems: namely, stock-market series and ocean turbulence.



The present introductory chapter has explained the context and main motivations of the research, together with the state-of-the-art. Chapter 2 introduces the technical notation and reviews the formal aspects of multiresolution representation. We place the emphasis on wavelet projections, particularly in their expression as quadrature mirror filters (QMF) and a generalized version of them. New theoretical and empirical work is presented in Chapter 3, where the formalism of microcanonical cascades is introduced and related to the concept of optimal wavelet. In this context, we calculate how wavelet optimality affects a microcanonical cascade representation and show that the resulting statistical association between cascade variables corresponds to the association in wavelet projections observed in different natural systems: this feature has been reported in natural images [58] and ocean chlorophyll concentration maps [83], and we have also found it in stock market series (Chapter 5, [10]) and sea surface temperature maps (Chapter 6, [11]). Thus we provide an explanation for these empirical statistical associations, we contextualize them in terms of the microcanonical cascade process and we derive a parameter that determines the degree of optimality of a wavelet for a given dataset. With this parameter, we can calculate the optimal wavelet for these systems. In Chapter 4, we relate the microcanonical cascade formalism with the canonical and the microcanonical multifractal formalisms. We review the main aspects of the microcanonical multifractal formalism and shows characterization of singularity exponents and the singularity spectrum in stock-market series and applications of it. Coincidence of singularity spectra for different natural systems is also shown. Then, in Chapter 5 we present an inference model based on the optimal wavelet and infer different variables for daily stock-market series. A different application of a very similar algorithm is shown in Chapter 6, where multiscale information of ocean structures

is derived from cascade variables of sea surface temperature. Finally, global conclusions of the different aspects of this work are shown in Chapter 7.

## 2 Multiresolution analysis

Space and time are the variables in terms of which we usually describe and model the relevant features, either static or dynamic, of any physical system. The validity of models depends crucially on the scale that they aim to describe, because the physical processes that govern dynamic evolution and structural properties of the system can be completely different from one scale to another.

Linear phenomena are usually described at one relevant scale, related either to some characteristic time in a dynamic approach or to some structural size. Even if more than one scale is involved, phenomena can be described in terms of decoupled equations for each scale. The picture is completely different for nonlinear systems. There, one usually finds that scales are coupled, structures are complex and dynamics is often chaotic. Self-organized patterns emerge spontaneously and collective effects that cannot be understood from the individual behaviour of single isolated constituents determine the macroscopic properties of the system. Then, it is no longer possible to look at single scales as all of them become relevant. Fortunately, in some cases, the system evolves towards a regime in which structure and dynamics become scale-invariant and where physical observables follow power-laws. We enter in the realm of a very special type of nonlinear systems called critical systems [2].

Physics, like other sciences, is essentially an empirical discipline. We obtain data from experiments and from them we create a theoret-

ical corpus aimed at explaining what we observe. The data provided by experiments must be appropriately processed and manipulated in order to create good models which can help us to understand nature. Data processing is not just a minor element of our methodology, but it plays a crucial role in the analysis of nonlinear systems.

The standard representation basis, which is constituted by space and time coordinates, is highly convenient to describe static or inertial systems, while Fourier basis, which is constituted by spatial and temporal frequencies, is appropriate for oscillations and orbits. However, in order to do appropriate analysis of scale-invariant data we need to work on a multiscale basis that is adjustable both in space (time) and spatial (temporal) frequency. This is done with sliding-window measures – where the scale is tuned as the size of the window – or related functionals, especially *wavelet transforms* (see Section 2.2).

## 2.1 Multiresolution decomposition

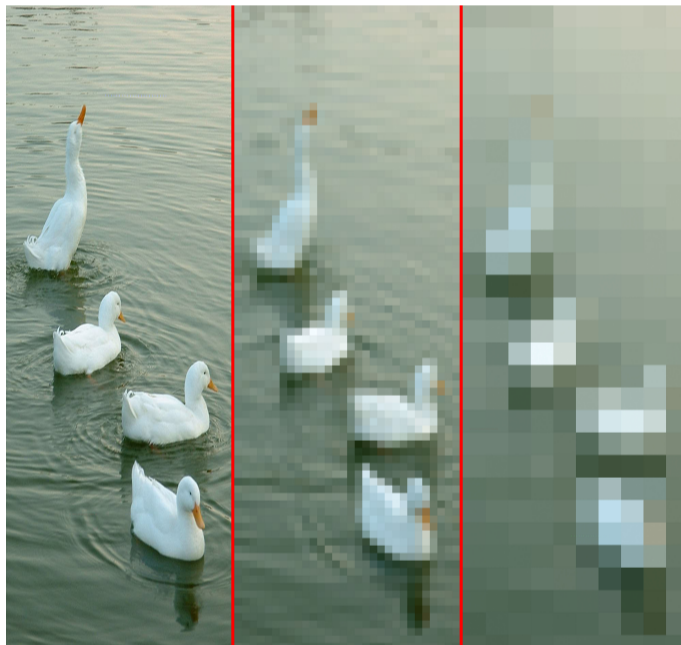
How to decompose a function into its scale components? We have seen why a decomposition into the different scales is convenient to properly analyze a complex signal. Let us see now how is it done. Firstly, we consider approximation components: they are functions that belong to approximation spaces of different resolution. We can see an example of this decomposition in Figure 2.1. Approximation spaces are characterized by the maximum scale (resolution) that they are able to achieve. Therefore any component of a certain approximation space  $\mathbb{A}_j$  can be also represented in the following (more resolute) space  $\mathbb{A}_{j+1}$ , in other words,  $\mathbb{A}_j$  is a subspace of  $\mathbb{A}_{j+1}$ :

$$\mathbb{A}_j \subset \mathbb{A}_{j+1} \quad (2.1)$$

Additionally, each approximation space  $\mathbb{A}_j$  has its projector  $A_j$ , which is an operator that takes any function and approximates it to fit the  $\mathbb{A}_j$  space, i.e.,

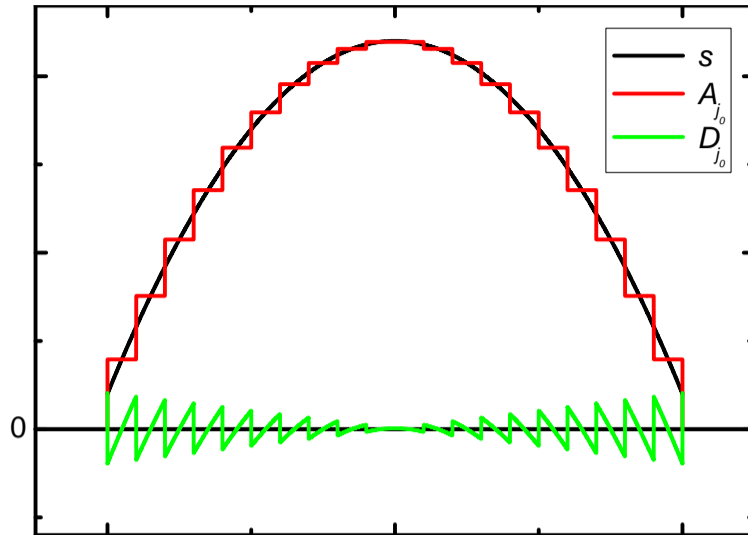
$$A_j(x) \equiv (A_j f)(x) \in \mathbb{A}_j \quad (2.2)$$

where  $A_j(x)$  is called the approximation component of  $f(x)$  at a scale of index  $j$ . An example of this is shown in Figure 2.2. Since  $j$  indicates the maximum attainable scale (resolution), the infinite-resolution approximation component of any function is the function itself:  $A_\infty(x) = f(x)$ , i.e., it is not approximated.



**Figure 2.1:** Original signal (left) and approximation components at a fine scale (centre) and at a coarse scale (right). Large structures are the only ones seen at the coarse-scale approximation. As resolution increases, finer details become distinguishable. Therefore, approximations filter signal structures by their size. Picture by Laura Pont and elaboration by the author.

Approximating to a maximum scale accumulates all the information up to that scale. Rather than this, sometimes we prefer to concentrate in the information contained in a given scale only. This is what detail components do: they tell what has changed from a certain



**Figure 2.2:** A function and its approximation and detail components for a certain scale with index  $j_0$ . The function belongs to the subspace of index  $j_0 + 1$ , i.e.,  $f(x) \in \mathbb{A}_{j_0+1}$ . Approximation component takes most of the large-scale structure of the function and provides a harsh approximation of it. Detail component  $\mathbb{D}_{j_0}$  retains the difference from  $\mathbb{A}_{j_0}$  and it emphasizes the details at the scale of index  $j_0$ : detail component is small in flat areas and large in points of high-contrast (those of steep slopes).

approximation to the following one (see Figure 2.2):

$$D_j(x) = A_{j+1}(x) - A_j(x) \quad (2.3)$$

Detail components lie in detail spaces  $\mathbb{D}_j$ . As a consequence of Eq. (2.3), the detail space  $\mathbb{D}_j$  is the complement of the approximation space  $\mathbb{A}_j$  in the following approximation space  $\mathbb{A}_{j+1}$ . This is formally expressed as:

$$\mathbb{A}_j \oplus \mathbb{D}_j = \mathbb{A}_{j+1} \quad (2.4)$$

Similarly to approximation spaces, detail spaces also have their projectors. The detail projector  $D_j$  takes a function and filters the information contained in the scale of index  $j$ , i.e., it extracts the detail component  $D_j(\vec{x})$ :

$$D_j(x) \equiv (D_j f)(x) \in \mathbb{D}_j \quad (2.5)$$

As intuitively expected, the sum of all the details recovers the function:  $\sum_j D_j(x) = f(x)$ ; the same in terms of their respective spaces:  $f \in \bigoplus^{\forall j \in \mathbb{Z}} \mathbb{D}_j = \mathbb{A}_\infty$ .<sup>1</sup>

For reasons of simplicity, we often desire that the approximation and the detail components of a same scale both contain the same amount of information ( $\mathbb{A}_j$  has the same dimensionality as  $\mathbb{D}_j$ ). This corresponds to the case when the scale step is dyadic, i.e., two consecutive scales are related by a factor two: for any function  $g^{(J)}(x) \in \mathbb{A}_J$  then  $g^{(J)}(2x) \in \mathbb{A}_{J+1}$ . In this case, the physical scale  $r$  changes with a factor two:  $r_j = 2r_{j+1}$  and so it varies exponentially with the index  $j$ :  $r_j \propto 2^{-j}$ .

## 2.2 Wavelets: a basis for multiscale functions

As we have said, for a given signal there are different operators that can be used to unveil its multiscaling properties. For instance, in stock market series increment measures and volatilities are both known to have cascade properties [95, 96], also in fully developed 2D turbulence, energy and enstrophy both follow cascade processes [22]. While these operators are nonlinear, linear operators also work in most cases: even the simplest operator, the increment at distance  $r$  (of the logarithm of the price in stock market series and of the velocity field in turbulence). In other cases, correlations mask the cascade behaviour in increments and convolution with an appropriate kernel is required.

These *appropriate kernels* are functions called *wavelets*. Wavelets are waveforms that decay in the tails. They are capable of filtering long-range correlations out, also provide a smooth interpolation that

---

<sup>1</sup>It should be noticed that in most practical situations, with real-world data, one never reaches infinite resolutions. Beyond the resolution limit, approximations are perfect (the function  $f$  itself) and details are null, which effectively truncate these infinite sums.

reduces the effect of noise. In addition, wavelets can be designed to make a decomposition into scale components (presented in Section 2.1) that is a complete representation able to reconstruct the signal.

The elements of a wavelet basis are dilations and translations of a representation pattern, i.e., they are tuned in both the position and the scale. That is why they have two kinds of coordinates: those of translations and those of scales. The basis for the approximation space of scale index  $j$ ,  $\mathbb{A}_j$ , is composed of functions  $\phi_{j,k}(\vec{x})$  where  $k$  is the translation index. When the scale step is dyadic, the vectors  $\phi_{j,k}(2x)$  for  $k = -\infty, \dots, +\infty$  form a basis of the  $\mathbb{A}_{j+1}$  space.<sup>2</sup> Analogously, we notate the detail space  $\mathbb{D}_j$  basis as  $\psi_{j,k}(x)$ . This way, for all the integer values of scale  $j$  and translation  $k$  indices,  $\psi_{j,k}(x)$  form a basis of the whole space of functions. This is called *wavelet basis*.

### 2.2.1 Settings for a dyadic representation

In the following, we present the analysis and representation of signals with dyadic wavelet bases. We show all the relevant equations to put our notation in context (most of them can also be found in wavelets textbooks [98, 99]). We introduce further details from Section A.2 and thereafter. First of all, we will restrict to  $\mathcal{L}^2$  space of functions,<sup>3</sup> which is needed for the wavelet approach to work. In this context, approximation and detail spaces are subspaces of  $\mathcal{L}^2$ .

$$\dots \subseteq \mathbb{A}_{-1} \subseteq \mathbb{A}_0 \subseteq \mathbb{A}_1 \subseteq \dots \subseteq \mathcal{L}^2 \quad (2.6)$$

Therefore,  $\mathbb{A}_j$  is coarser than  $\mathbb{A}_{j+1}$ . In addition, if  $|s(x)\rangle \in \mathbb{A}_J$  then  $|s(\frac{x}{2})\rangle \in \mathbb{A}_{J-1}$ .

<sup>2</sup> Remember that the scale is  $r_j = 2^{-j}$ .

<sup>3</sup> This means that given a signal  $s(x) \in \mathcal{L}^2$ , it follows that  $\int |s(x)|^2 dx$  is finite



The bases of these spaces are called wavelets. The scaling function  $\phi(x)$  (sometimes also called father wavelet) is a basis of the  $\mathbb{A}_0$  space, while the mother wavelet  $\psi(x)$  is the basis of the  $\mathbb{D}_0$  space. Then, the projectors  $A_j$  and  $D_j$  can be expressed as:

$$A_j|s\rangle = \sum_{k=-\infty}^{+\infty} \underbrace{\langle \overline{\phi_{j,k}} | s \rangle}_{\beta_{j,k}} |\phi_{j,k}\rangle \quad (2.7)$$

$$D_j|s\rangle = \sum_{k=-\infty}^{+\infty} \underbrace{\langle \overline{\psi_{j,k}} | s \rangle}_{\alpha_{j,k}} |\psi_{j,k}\rangle \quad (2.8)$$

where  $\{|\phi_{j,k}\rangle\}_{k \in \mathbb{Z}}$  is the basis of  $\mathbb{A}_j$  and  $\{|\psi_{j,k}\rangle\}_{k \in \mathbb{Z}}$  is the basis of  $\mathbb{D}_j$ . The aforementioned equation assumes that the wavelets are norm-2 normalized. This means that:

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k) \quad \|\phi_{j,k}\|_2 = 1 \quad (2.9)$$

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad \|\psi_{j,k}\|_2 = 1 \quad (2.10)$$

and also the (bi-)orthonormality conditions:

$$\langle \overline{\phi_{j,l}} | \phi_{j,k} \rangle = \delta(k - l) \quad (2.11)$$

$$\langle \overline{\psi_{i,l}} | \psi_{j,k} \rangle = \delta(j - i) \delta(k - l) \quad (2.12)$$

Finally, note that the coefficients  $\beta_{j,k}$  and  $\alpha_{j,k}$  define a wavelet series that can represent the original signal. Provided that  $|s\rangle \in \mathbb{A}_J$

and using Eq. (2.7) and Eq. (2.8),

$$\begin{aligned}
 |s\rangle &= A_J |s\rangle \\
 &= \sum_{j=0}^{J-1} D_j |s\rangle + A_0 |s\rangle \\
 &= \sum_{j=0}^{J-1} \sum_{k=-\infty}^{+\infty} \alpha_{j,k} |\psi_{j,k}\rangle + \sum_{k=-\infty}^{+\infty} \beta_{0,k} |\phi_{0,k}\rangle \quad (2.13)
 \end{aligned}$$

$$= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{+\infty} \alpha_{j,k} |\psi_{j,k}\rangle \quad (2.14)$$

### 2.2.2 Norm-1 wavelet transform

In signal processing, it is rather usual to calculate the detail coefficients with a wavelet basis that is norm-1 normalized (provided that this normalization is finite). The link between the two types of normalization, i.e. norm-1 and norm-2, is straightforward:

$$\begin{aligned}
 |s\rangle &= \sum_{j,k} \beta_{j,k} |\psi_{j,k}\rangle \\
 &= \sum_{j,k} \tilde{\beta}_{j,k} |\tilde{\psi}_{j,k}\rangle \quad (2.15)
 \end{aligned}$$

where  $\|\psi\|_2 = \|\tilde{\psi}\|_1 = 1$  and  $\int |\psi|(t) dt = A$ , i.e.,  $\tilde{\psi} = \frac{\psi}{A}$ . Furthermore,  $\|\tilde{\psi}_{j,k}\|_1 = 1$  which means that:

$$|\psi_{j,k}(t)\rangle = 2^{j/2} |\psi(2^j t - k)\rangle \quad (2.16)$$

$$\begin{aligned}
 |\tilde{\psi}_{j,k}(t)\rangle &= 2^j |\tilde{\psi}(2^j t - k)\rangle \\
 &= \frac{2^{j/2}}{A} |\psi_{j,k}(t)\rangle \quad (2.17)
 \end{aligned}$$

so the detail coefficients are related through:

$$\tilde{\alpha}_{j,k} = A 2^{-j/2} \alpha_{j,k} \quad (2.18)$$

The advantage of the norm-1 normalization is that the convolution with the wavelet does not depend on the scale:

$$\|\tilde{\psi} \otimes s\|_1 \equiv \|\tilde{\alpha}_{j,k}\|_1 \leq \|s\|_1 \quad (2.19)$$

and so, if the signal is scale invariant and translationally invariant, the typical amplitudes of the coefficients  $\tilde{\alpha}_{j,k}$  are similar for all  $j, k$ .

### 2.2.3 Quadrature mirror filters (QMFs)

It is well known that the decomposition of the scaling function and the wavelet in terms of the immediately finer scaling function defines two filters,  $h$  and  $g$  (with associated dual filters in the biorthogonal case), called quadrature mirror filters, that can be recursively applied to produce a fast wavelet decomposition [99]. The filters  $h$  and  $g$  are defined by its coefficients  $h_k$  and  $g_k$ , according to the following expressions:

$$|\phi_{0,0}\rangle = \sum_k \underbrace{\langle \overline{\phi_{1,k}} | \phi_{0,0}\rangle}_{h_k} |\phi_{1,k}\rangle \quad (2.20)$$

$$|\psi_{0,0}\rangle = \sum_k \underbrace{\langle \overline{\phi_{1,k}} | \psi_{0,0}\rangle}_{g_k} |\phi_{1,k}\rangle \quad (2.21)$$

and are also norm-2 normalized:  $\sum_k \bar{h}_k^* h_k = 1$  and  $\sum_k \bar{g}_k^* g_k = 1$ , and mutually orthogonal:  $\sum_k \bar{h}_k^* g_k = \sum_k \bar{g}_k^* h_k = 0$ . To analyze a

signal means going from finer to coarser scales:

$$\beta_{j,k} = \sum_l \bar{h}_l^* \beta_{j+1,2k+l} \quad (2.22)$$

$$\alpha_{j,k} = \sum_l \bar{g}_l^* \beta_{j+1,2k+l} \quad (2.23)$$

while synthesizing or reconstructing means going from coarser to finer scales:

$$\beta_{j+1,k} = \sum_l h_{k-2l} \beta_{j,l} + \sum_l g_{k-2l} \alpha_{j,l} \quad (2.24)$$

and a practical implementation of the expression above is:

$$\beta_{j+1,2k} = \sum_l h_{2l} \beta_{j,k-l} + \sum_l g_{2l} \alpha_{j,k-l} \quad (2.24'a)$$

$$\beta_{j+1,2k+1} = \sum_l h_{2l+1} \beta_{j,k-l} + \sum_l g_{2l+1} \alpha_{j,k-l} \quad (2.24'b)$$

### Proof of analysis and synthesis equations

Let us prove equations (2.22–2.24'b). Eq. (2.20) in a more general way is:

$$|\phi_{j,k}\rangle = \sum_l h_l |\phi_{j+1,2k+l}\rangle \quad (2.25)$$

and its dual counterpart is:

$$\langle \overline{\phi_{j,k}} | = \sum_l \bar{h}_l^* \langle \overline{\phi_{j+1,2k+l}} | \quad (2.26)$$

now put  $|s\rangle$  to both sides and note that these are  $\beta$ 's coefficients. The result is Eq. (2.22). An equivalent deduction applies to Eq. (2.23) from Eq. (2.21).

The synthesis formula, Eq. (2.24), comes from Eq. (2.3):

$$A_{j+1} = A_j + D_j \quad (2.27)$$

and the definitions Eq. (2.7) and Eq. (2.8),

$$A_j|s\rangle = \sum_l \beta_{j,l} |\phi_{j,l}\rangle \quad (2.28)$$

$$D_j|s\rangle = \sum_l \alpha_{j,l} |\psi_{j,l}\rangle \quad (2.29)$$

$$A_{j+1}|s\rangle = \sum_l \beta_{j+1,l} |\phi_{j+1,l}\rangle \quad (2.30)$$

so that inserting  $\langle \overline{\phi_{j+1,k}} |$  in front of both sides of Eq. (2.27) it becomes:

$$\langle \overline{\phi_{j+1,k}} | A_{j+1} |s\rangle = \langle \overline{\phi_{j+1,k}} | A_j |s\rangle + \langle \overline{\phi_{j+1,k}} | D_j |s\rangle \quad (2.31)$$

Substituting definitions of  $A$ 's and  $D$ 's:

$$\sum_l \beta_{j+1,l} \underbrace{\langle \overline{\phi_{j+1,k}} | \phi_{j+1,l} \rangle}_{\delta(k-l)} = \sum_l \beta_{j,l} \langle \overline{\phi_{j+1,k}} | \phi_{j,l} \rangle + \sum_l \alpha_{j,l} \langle \overline{\phi_{j+1,k}} | \psi_{j,l} \rangle \quad (2.32)$$

Now we use Eq. (2.25) and its equivalent for  $g$ :

$$\beta_{j+1,k} = \sum_{l,l'} \beta_{j,l} h_{l'} \underbrace{\langle \overline{\phi_{j+1,k}} | \phi_{j+1,2l+l'} \rangle}_{\delta(k-2l-l')} + \sum_{l,l'} \alpha_{j,l} g_{l'} \underbrace{\langle \overline{\phi_{j+1,k}} | \phi_{j+1,2l+l'} \rangle}_{\delta(k-2l-l')} \quad (2.33)$$

and Eq. (2.24) is directly obtained. Eq. (2.24'a) and Eq. (2.24'b) are obtained changing  $k$  by a new  $k'$ . We have two cases, depending on the parity of  $k$ :  $k = 2k'$  and  $k = 2k' + 1$ . Then, we contract the  $l$

instead of  $l'$  in the equation above:

$$\beta_{j+1,2k'} = \sum_{l'} h_{l'} \beta_{j,k'-\frac{l'}{2}} + \sum_{l'} g_{l'} \alpha_{j,k'-\frac{l'}{2}} \quad (2.24'' \text{ a})$$

$$\beta_{j+1,2k'+1} = \sum_{l'} h_{l'} \beta_{j,k'-\frac{l'-1}{2}} + \sum_{l'} g_{l'} \alpha_{j,k'-\frac{l'-1}{2}} \quad (2.24'' \text{ b})$$

but the cases where the subindex is not integer are not allowed, so we redefine  $l$  so that  $l' = 2l$  in the even case and  $l' = 2l + 1$  in the odd case, directly resulting in Eq. (2.24'a) and Eq. (2.24'b).

#### 2.2.4 QMF filters from the wavelet coefficients

In the following we take the standard elements introduced so far to a more geometrical point of view. For simplicity, we focus on orthonormal and linear-phase filters [98]. Orthonormality implies that the filters equal their duals, while linear-phase implies that:

$$g_l = (-1)^l h_{1-l}^* \quad (2.34)$$

In addition, we will consider only real-valued filters. Hence, only one real-valued, infinite-dimension vector  $\vec{h}$  is needed.

The algorithm that we are presenting compares the wavelet coefficients at two consecutive scales. These *detection* scales should be coarser than the resolution scale of the signal, i.e., for a signal of resolution<sup>4</sup>  $J$ , the detection scales are  $J_0 - 1$  and  $J_0 \leq J$ . Additionally, we will consider the *pixellation* scale  $J_1 \geq J$ , i.e., the scale at which the scaling function covers only individual pixels. Roughly,  $J_1 - J = \log_2(\text{scale of the scaling function in pixels})$ . The wavelet

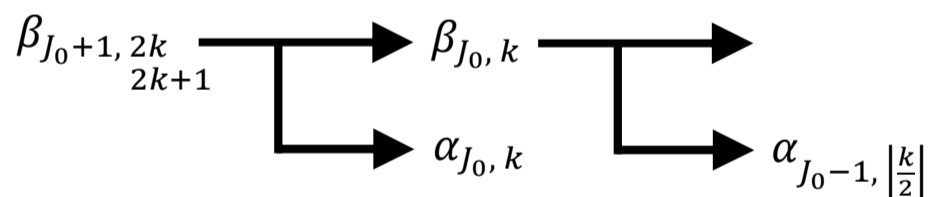
---

<sup>4</sup> With scale  $j$  we mean  $r = 2^{-j}$ .

coefficients at scales  $J_0$  and  $J_0 - 1$  can be expressed as:

$$\begin{aligned}
 \alpha_{J_0, k} &= \sum_l \beta_{J_0+1, 2k+l} g_l \\
 &= \sum_l \underbrace{(-1)^{1-l} \beta_{J_0+1, 2k-l+1}}_{(\vec{\beta}_{\mathbf{k}})_l} h_l \\
 &= \vec{\beta}_{\mathbf{k}} \cdot \vec{h}
 \end{aligned} \tag{2.35}$$

$$\begin{aligned}
 \alpha_{J_0-1, \lfloor \frac{k}{2} \rfloor} &= \sum_{l'} \beta_{J_0, 2\lfloor \frac{k}{2} \rfloor + l'} g_{l'} \\
 &= \sum_{l'} (-1)^{1-l'} \beta_{J_0, 2\lfloor \frac{k}{2} \rfloor - l' + 1} h_{l'} \\
 &= \sum_{l'} \left( \sum_l \beta_{J_0+1, 4\lfloor \frac{k}{2} \rfloor + 2l' + l} h_l \right) g_{l'} \\
 &= \sum_{l, l'} \underbrace{(-1)^{1-l'} \beta_{J_0+1, 4\lfloor \frac{k}{2} \rfloor - 2l' + l + 2} h_l h_{l'}}_{(\mathbb{B}_{\mathbf{k}})_{l, l'}} \\
 &= \vec{h}^t \mathbb{B}_{\mathbf{k}} \vec{h}
 \end{aligned} \tag{2.36}$$



**Figure 2.3:** Analysis algorithm to extract the  $\alpha_{J_0, k}$  and  $\alpha_{J_0-1, \lfloor \frac{k}{2} \rfloor}$  wavelet coefficients.

To obtain the  $\beta_{J_0+1,\bullet}$  we build them from the pixellation scale  $J_1$ . Let  $\Delta J = J_1 - J$ ,

$$\begin{aligned}
 \beta_{J_1,k} &= \langle \phi_{J_1,k} | s \rangle \\
 &= 2^{J_1/2} \sum_{n=0}^{2^J-1} \int_{2^{-J_n}}^{2^{-J(n+1)}} \phi^*(2^{J_1}t - k) s_n dt \\
 &\approx \left( 2^{-J_1/2} \int_{-\infty}^{+\infty} \phi^*(t) dt \right) \sum_{n=0}^{2^J-1} s_n \delta \left( \left\lfloor \frac{k}{2^{\Delta J}} \right\rfloor - n \right) \\
 &\approx 2^{-J_1/2} \left( \int_{-\infty}^{+\infty} \phi^*(t) dt \right) s_{\lfloor \frac{k}{2^{\Delta J}} \rfloor} \quad k = 0, \dots, 2^{J_1} - 1
 \end{aligned} \tag{2.37}$$

with  $|\int \phi(t) dt| = 1$  (see Appendix A), and now we successively apply the  $h$  filter until we reach the  $J_0 + 1$  scale needed in Eq. (2.35) – Eq. (2.36). The main problem with this is that we suppose a smooth signal, not a multifractal, so that choosing a very high  $\Delta J$  will affect the multifractal characterization. The only solution here is to work with narrow scaling functions.

The elements introduced in this chapter will be used in all the following chapters. The concept of QMFs is quite standard nowadays, and we have reproduced its full development as it constitutes a central ingredient, especially in Chapter 3. New developments in this context include the geometrical expression of QMFs, Eq. (2.35) and Eq. (2.36); we will take advantage of it in the next chapter. The algorithm for a biorthogonal scheme presented in the appendix, in Section A.2, is also a new achievement; we have put in the appendix because it is a quite long derivation and we do not use it in the following, though it represents a future line of research that could be used to generalize the wavelet optimization.



## 3 Microcanonical multiplicative cascades

### 3.1 Persistence in scale invariant signals

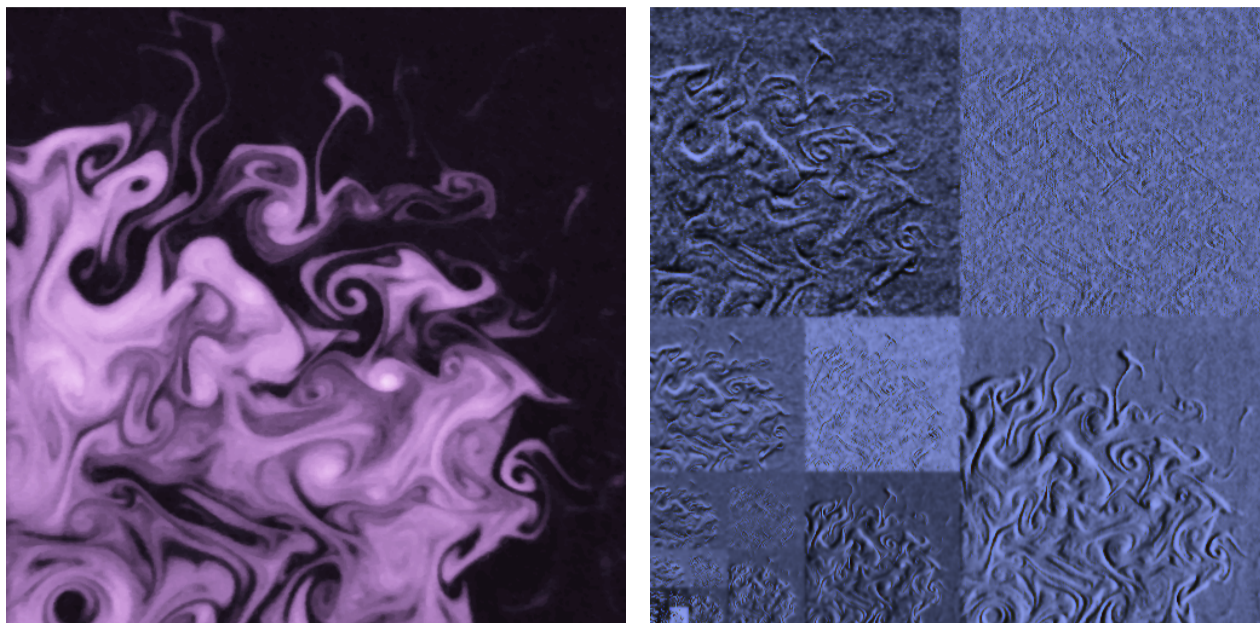
Multiplicative cascades are present in many different systems, but they are not usually recognized as such. Usually, their presence is reported by means of indirect evidence about its effects on the properties of signals. One of the most commonly reported effect of multiplicative cascades is the persistence of feature detection across scales. The importance of persistence is that the detection of a feature at a coarse scale allows inferring the presence of the same feature at finer scales. This phenomenon is well known since the introduction of wavelet representation of signals, and it was first described by Mallat and co-workers twenty years ago [100, 101]. The optimal wavelet is the one that maximizes this inference capability.

To understand what is the role of wavelet processing it is convenient to clarify what a multiresolution decomposition is. In a multiresolution decomposition the signal can be represented as a combination of wavelet coefficients that can be arranged according progressive levels of resolution, from finer to coarser. This representation is just an algebraic change of basis, so the multiresolution decomposition of a signal contain exactly the same information as the original signal, and we can pass from one to the other with a linear transformation and

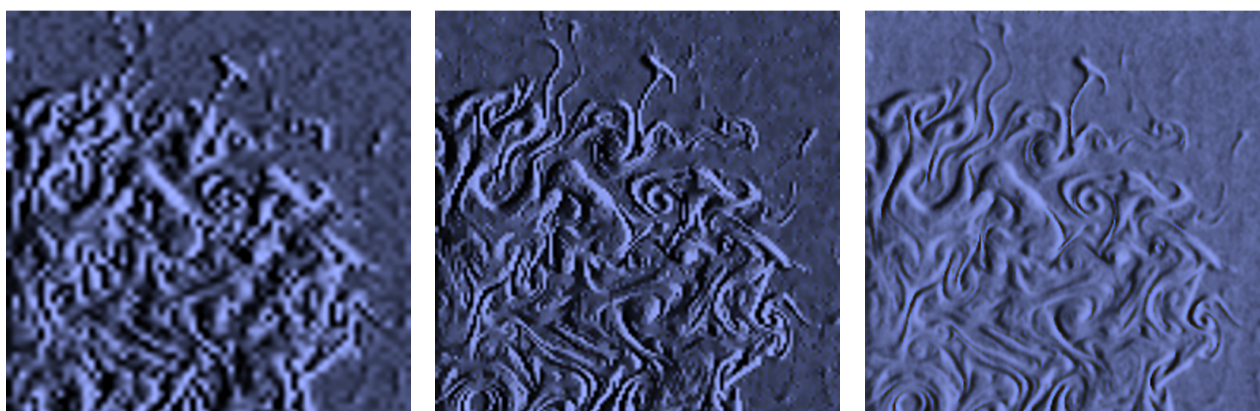
without any loss of information. In the case of 1D signals a single wavelet can be used to fully represent the signal in a dyadic scheme; for 2D signals, we need three different wavelets that will expand three different pyramids of resolution levels [99]. In a dyadic scheme, when we pass from one resolution to the next coarser one the scale changes by a factor two, i.e., the diameter of the wavelet at the coarser scale is exactly twice the diameter of the wavelet at the previous, finer scale. This implies that a wavelet coefficient obtained at the coarser scale affects an area that is twice larger in diameter than that of the finer scale; roughly speaking, a wavelet coefficient at the coarser scale covers the area of two wavelet coefficients at the finer scale in 1D and the area of four wavelet coefficients at the finer scale in 2D. In section 3.1.1 the concepts of wavelet basis and dyadic decomposition will be introduced in greater detail; see also [98, 99].

In Figure 3.1 we show a typical example of interscale persistence. In the left panel we show a snapshot of a passive scalar dispersed by a 2D turbulent flow. This turbulent regime can be described by means of a direct enstrophy cascade. In the right panel we show a multiresolution decomposition of this image, formed by all the wavelet coefficients of the representation arranged in a compact shape.

In Figure 3.2 we present a detail of three consecutive resolutions of vertical coefficients extracted from Figure 3.1. Notice that the multiresolution decomposition is just a change of vectorial basis, so the wavelet coefficients are algebraically independent. It is however obvious from Figure 3.2 that the coefficients do not take arbitrary values: the structures detected at coarser scales persist at the same location but with better resolution at the finer scales. This is the persistence of edges, and it is a consequence of the structure of the signal, which implies that on many real systems edges are multiscale.



**Figure 3.1:** **Left:** Snapshot of dye distribution submitted to the action of 2D turbulence, under direct enstrophy cascade. Brightness is proportional to dye concentration: white corresponds to the maximum and black corresponds to the minimum. The image was obtained in a laboratory experiment; see [102, 103, 104] for details. **Right:** Multiresolution decomposition with Haar wavelet of the image on the left. A separable 2D multiresolution basis requires three wavelets and hence there are three types of wavelet coefficients, which are labelled as horizontal (leftmost squares), vertical (those with a side on the bottom of the panel) and diagonal. Each resolution level and orientation has been independently normalized to enhance details. Both images have been coloured to beautify presentation, without any change in brightness values.



**Figure 3.2:** The three finer resolution levels of vertical wavelet coefficients, extracted from Figure 3.1; going from left to right we go from the coarser to the finest resolution. The three resolution levels are represented at the same size to help comparison.

Edge persistence is a strong, relevant feature 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 [58, 59] 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 [105, 58, 59, 106]. 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 [58, 106]. 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 [58, 105]. 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. [83] 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. As we will see later, a wavelet for which the conditioned histogram is fully linear is called optimal, in the same sense that the one introduced by Turiel and Parga in [91]. Pottier et al. proposed a particular model to describe the child-parent dependency, 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:

$$\alpha_C = \eta_* \alpha_P + \alpha_* \quad (3.1)$$

where  $\alpha_C$  stands for the child wavelet coefficient and  $\alpha_P$  stands for its parent (i.e., it is obtained at the immediately coarser scale and at the same position).  $\eta_*$  and  $\alpha_*$  are random variables mutually independent, also independent from  $\alpha_P$ . 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  $\alpha_*$  varies; for smaller  $\alpha_*$  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  $\alpha_*$  vanishes? This would be the optimal wavelet, in the same sense as in [91, 107].

The importance of finding such an optimal wavelet must be emphasized. First, because with the aid of this wavelet the description of the mutual dependence between parents and children can be simplified; in fact,  $\alpha_* = 0$  implies that the mutual information between  $\alpha_P$  and  $\alpha_C$  is maximized. So, a coding scheme as the one proposed by [58] attains the highest quality and smallest coding cost with the use of this wavelet. Besides, using this wavelet basis the inference of the value of the coefficients is improved, what has an impact on the quality of reconstruction algorithms to fill data gaps (as in [83]) or on forecasting of time series. Finally, optimal wavelets can be used to derive improved models of multifractal systems (for instance, some variables under fully developed turbulence).

### 3.1.1 Towards an optimal representation of data

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. Hence, 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  $p$  [108].

A 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:

$$\langle \mathbf{T}_r^p \rangle = A_p^{\mathbf{T}} r^{\tau_p} + o(r^{\tau_p}) \quad (3.2)$$

Recall that the symbol  $o(r^{\tau_p})$  means a contribution that is negligible compared to  $r^{\tau_p}$  when  $r$  goes to zero. In fractal signals, the exponent  $\tau_p$  is directly proportional to the moment order  $p$  and the proportionality constant is called singularity exponent or Hurst exponent.

In multifractal signals, the dependence of  $\tau_p$  on  $p$  is nonlinear, a fact known as *anomalous scaling* [22]. This name derives from the study of fully developed turbulence and is used to call a scaling that deviates from the fractal one proposed by Kolmogorov in [14, 109] (more precisely, the extension  $\forall p$  of this scaling). In this field, the observed nonlinearity is attributed to intermittency of energy dissipation in turbulent flows [22], though a recent study [110] has shown that intermittency alone does not imply any change in Kolmogorov's scaling below a certain order  $p$  larger than three and above this order the prefactors  $A_p^{\mathbf{T}}$  diverge.

In Section 4.2, 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:

$$\langle \mathbf{T}_r^p \rangle = \left( \frac{r}{L} \right)^{\tau_p} \langle \mathbf{T}_L^p \rangle \quad (3.3)$$

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

$$\langle \eta_\kappa^p \rangle = \kappa^{\tau_p} \quad (3.4)$$

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  $\kappa$  smaller than 1; for this reason we have taken the ratio of the smaller scale by the larger scale.

With the aid of the variable  $\eta_\kappa$  we can express Eq. (3.3) in a more elegant way, making the cascade relation explicit:

$$\mathbf{T}_r \doteq \eta_{r/L} \mathbf{T}_L \quad (3.5)$$

with  $\eta_{r/L}$  and  $\mathbf{T}_L$  being mutually independent. Here the symbol  $\doteq$  means that the equality holds distributionally, i.e.,

$$\rho(\mathbf{T}_r) = \rho(\eta_{r/L} \mathbf{T}_L) \quad (3.6)$$

However, this relation does not necessarily hold pointwise, as we will explain in the following subsection.

The introduction of Eq. (3.5) now allows splitting 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 behaviour 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. (3.5) to the three possible scale pairs it follows:

$$\eta_{r/L} \doteq \eta_{r/r'} \eta_{r'/L} \quad (3.7)$$

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 [21, 35, 38]. 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  $\tau_p$  of the signal in Eq. (3.2) leads to exactly the same distribution of cascade variables  $\eta_{r/L}$  [37].

#### Microcanonical cascade

Eq. (3.5) 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,

$$\tilde{\eta}_{r/L}(\vec{x}) = \frac{\mathbf{T}[s](r, \vec{x})}{\mathbf{T}[s](L, \vec{x})} \quad (3.8)$$

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. (3.4). 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. (3.5) makes sense not only distributionally but also at any point  $\vec{x}$  of the system. That is what we call *microcanonical cascade*.

Among the functionals that are most commonly used to analyze the scaling properties of multifractal systems, wavelets occupy a prominent position. Wavelets have been used to perform local Fourier analysis and to characterize the local singularities of functions [111]. In many different multifractal systems, wavelet projections have been used to characterize their scaling properties with success [5, 95]. Something that is very convenient about wavelet projections is that they can be inverted to retrieve the original signal [98], as we have seen in Section 2.2: wavelet projections do not only analyze the signal, but also constitute a representation of it. That is why wavelet projections are good candidates to realize the microcanonical cascade.

Roughly speaking, a **wavelet** is a function that oscillates in the centre of its domain and decays in its tails; we can think about wavelets as a pulse that decays very fast. Let  $s(\vec{x})$  be a multifractal signal and let  $\psi(\vec{x})$  be a wavelet. We define the wavelet projection of  $s$  on  $\psi$  at the position  $\vec{x}$  and the resolution scale  $r$  as:

$$\mathbf{T}_\psi[s](r, \vec{x}) \equiv \int d\vec{y} s(\vec{y}) \psi\left(\frac{\vec{x} - \vec{y}}{r}\right) \quad (3.9)$$

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

$$\mathbf{T}_\psi[s](r, \vec{x}) = \eta_{r/L}(\vec{x}) \mathbf{T}_\psi[s](L, \vec{x}) \quad (3.10)$$

Notice that the key point is that  $\eta_{r/L}(\vec{x})$  has to be both a cascade variable – in the sense of Eq. (3.7) – 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 3.2. There are evidences that such an optimal wavelet exists in natural images [91] and in marine turbulence [83] for the specific case of wavelet dyadic representations.

#### Dyadic representations of the cascade

Wavelet projections of a signal can be used to characterize the local properties of the signal or to represent it in an efficient scheme [111, 98]. Although a signal can be retrieved from its wavelet projections, such a representation is highly redundant and so a subset of wavelet projections must be retained. A typical way to subsample continuous wavelet projections is to select a dyadic subset, like that that we have introduced in Section 2.2, from which the signal is fully reconstructed. In a dyadic subset, the scale varies by a factor two and at each resolution level the positions are taken as integer amounts of the resolution size. This discretization leads to wavelet coefficients instead of wavelet projection.

To keep formulae simple, hereafter we will limit our attention to one-dimensional systems. Hence, given a 1D signal  $s(x)$  and a wavelet  $\psi$  capable to spawn a dyadic representation basis, the signal can be expanded as a series of wavelet terms:

$$s(x) = \sum_{j=-\infty}^{\infty} \sum_k \alpha_{j,k} \psi_{j,k}(x) \quad (3.11)$$

where

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k) \quad (3.12)$$

and  $j, k$  are integer numbers. The coefficients of this representation basis, the  $\alpha_{j,k}$ , are called *wavelet coefficients*. As  $\psi$  defines a representation basis, there is a unique set of wavelet coefficients  $\{\alpha_{j,k}\}$  such that Eq. (3.11) is valid. The  $2^{j/2}$  normalization factor ensures that the 2-norm is 1,  $\int dx |\psi_{j,k}(x)|^2 = 1$ .

If the wavelet basis is orthonormal, i.e.,

$$\langle \psi_{j,k} || \psi_{j',k'} \rangle \equiv \int dx \psi_{j,k}^*(x) \psi_{j',k'}(x) = \delta_{j,j'} \delta_{k,k'} \quad (3.13)$$

we can obtain the wavelet coefficients as projections on the wavelet basis, namely:

$$\alpha_{j,k} = \langle \psi_{j,k} || s \rangle \quad (3.14)$$

If the wavelet basis is not orthonormal, the extension is rather straightforward: each basis vector  $||\psi_{j,k}\rangle$  has its dual  $\langle\tilde{\psi}_{j,k}||$  so that,

$$\langle\tilde{\psi}_{j,k}||\psi_{j',k'}\rangle = \delta_{j,j'} \delta_{k,k'} \quad (3.15)$$

and the wavelet coefficients can be obtained as  $\alpha_{j,k} = \langle\tilde{\psi}_{j,k}||s\rangle$ .

In terms of a dyadic representation, the cascade takes a relatively simple form. For any wavelet basis, the canonical cascade relation, Eq. (3.5), takes the following form:

$$\alpha_{j,k} \doteq \eta_{\frac{1}{2}} \alpha_{j-1, [k/2]} \quad (3.16)$$

where the notation  $[k/2]$  means *the integer part of  $k/2$* . Here we have written the cascade relation mimicking Eq. (3.5) and Eq. (3.10). Notice that the wavelet coefficients,  $\alpha_{j,k}$ , are not intensive variables as the wavelet projections are, as defined in Eq. (3.9) (while wavelet projections are  $\infty$ -norm normalized, wavelet coefficients are 2-norm normalized, which is highly convenient in the derivations to follow, espe-

cially those in section 3.2). This means that the  $\eta$ -like variables written hereafter will differ from those appearing in Eq. (3.4) to Eq. (3.10) 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  $r_C = 2^{-j}$  and position  $x_C = 2^{-j} k$ , while  $\alpha_{j-1, \lfloor k/2 \rfloor}$  is the wavelet projection at the coarser scale  $r_P = 2^{-j+1}$  and position  $x_P = 2^{-j+1} \lfloor k/2 \rfloor$ ; the positions  $x_C$  and  $x_P$  differ at most by  $r_C$ , which is the spatial uncertainty at the scale  $r_P$ , so at the scale  $r_P$  we can consider that  $x_C$  and  $x_P$  refer to the same position. To alleviate the notation, for given fixed scale index  $j$  and position index  $k$ ,  $\alpha_P \equiv \alpha_{j-1, \lfloor k/2 \rfloor}$  is known as the Parent coefficient,  $\alpha_C \equiv \alpha_{j,k}$  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:

$$\alpha_C \doteq \eta \alpha_P \quad (3.17)$$

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

$$\alpha_C = \eta \alpha_P \quad (3.18)$$

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

### 3.2 Optimization from suboptimal representations

As we have already introduced in the previous chapter, 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 a series of infinite levels  $j$ , as in Eq. (3.11), or approximated at level  $j_0$  and expanded

to the details finer than  $j_0$ . Namely, we can expand the signal  $s(x)$  as follows:

$$s(x) = \sum_{j=-\infty}^{\infty} \sum_k \alpha_{j,k} \psi_{j,k}(x) \quad (3.19)$$

$$= \sum_{j=j_0}^{\infty} \sum_k \alpha_{j,k} \psi_{j,k}(x) + \underbrace{\sum_k \beta_{j_0,k} \phi_{j_0,k}(x)}_{A_{j_0}(x)} \quad (3.20)$$

The approximation  $A_{j_0}(x)$  can be expressed either as a wavelet expansion of all the levels coarser than the approximation level ( $j$  ranging from  $-\infty$  to  $j_0 - 1$ ), or as an expansion on unity functions at the single level  $j_0$ . Hence, it is possible to obtain the wavelet coefficients from the approximation, as the wavelet projections of the approximation coincide with those of the signal at any level coarser than  $j_0$ . It should be noticed that if the signal is discrete, it coincides with its approximation at any level finer than that of the discretisation scale.

The main advantage of this new decomposition is that the approximations  $A_{j_0}(x)$  are countable sums, so we can define two countable filters, denoted by  $\{g_n\}$  and  $\{h_n\}$ , that can be used to obtain the wavelet coefficients at any level provided that we know the approximation at the finest level, i.e., the signal at its discretisation level. Then, applying the conjugate mirror filters  $\{g_n\}$  and  $\{h_n\}$  we can both obtain the wavelet coefficients from the signal or retrieve the signal from their coefficients with very fast algorithms, which are exact over discretized collections of coefficients.

When we expand the scaling function itself  $\phi$  (i.e.,  $\phi_{0,0}$ ) up to the next coarser scale  $j_0 = -1$ , the filter  $\{g_n\}$ , which we will denote by the vector  $\vec{g} = (\dots, g_{-1}, g_0, g_1, g_2, \dots)$ , is given by the wavelet coefficients, while the filter  $\{h_n\}$ , which we will denote by the vector

$\vec{h} = (\dots, h_{-1}, h_0, h_1, h_2, \dots)$ , is given by the approximation coefficients, namely:

$$\phi(x) = \sum_n g_n \psi_{-1,n}(x) + \sum_n h_n \phi_{-1,n}(x) \quad (3.21)$$

where the wavelet coefficients at any level finer than  $j = -1$  are zero because the unity function coincides with itself at level  $j = 0$ .

Let us now suppose that we have a discretized signal  $s_k$  defined by a collection of values, which are naturally identified as the approximation coefficients at the highest resolution  $\beta_{0,k} = s_k$ . We will denote this collection of approximation coefficients by the vector  $\vec{\beta}_0 = (\dots, \beta_{0,-1}, \beta_{0,0}, \beta_{0,1}, \beta_{0,2}, \dots)$ . Since we have previously said that  $r = 2^{-j}$ , having the highest resolution at level  $j = 0$  means that we are expressing  $r$  in units of pixels. To obtain the wavelet coefficients at the next coarser level  $j = -1$  we apply the filter  $\vec{g}$ . Let  $\vec{\alpha}_{-1}$  be the vector of these wavelet coefficients, then we have:

$$\alpha_{-1,k} = \sum_n g_{n-2k} \beta_{0,n} \quad (3.22)$$

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. (3.22) in vectorial form as:

$$\vec{\alpha}_{-1} = \mathbb{G} \cdot \vec{\beta}_0 \quad (3.23)$$

Notice that the expression above can be used to relate the approximation and the wavelet coefficients of any two consecutive resolution

levels, i.e.,

$$\vec{\alpha}_{j-1} = \mathbb{G} \cdot \vec{\beta}_j \quad (3.24)$$

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:

$$\vec{\beta}_{j-1} = \mathbb{H} \cdot \vec{\beta}_j \quad (3.25)$$

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

### Perturbative analysis

In general, a given wavelet basis that is applied to the analysis of given data is not optimal. This means that the cascade does not hold in the microcanonical sense and so Eq. (3.18) 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. (3.1), 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 discretisation level  $j = 0$ , the signal corresponds to the vector  $\vec{\beta}_0^{\text{opt}} = (\dots, s_{-1}, s_0, s_1, \dots)$ . Let us consider now the Child and the Parent scale levels as the two next coarser dyadic levels, namely  $j_C = -1$ ,  $r_C = 2$  pixels and  $j_P = -2$ ,  $r_P = 4$  pixels (notice that the wavelet



coefficients at levels  $j \geq 0$  are all zero as discrete signals cannot vary inside their pixels, i.e., at levels finer than the discretisation scale). This way, Eq. (3.23) is notated as:

$$\vec{\alpha}_C^{\text{opt}} = \mathbb{G} \cdot \vec{\beta}_0^{\text{opt}} \quad (3.26)$$

The approximation to the next level is given by:

$$\vec{\beta}_{-1}^{\text{opt}} = \mathbb{H} \cdot \vec{\beta}_0^{\text{opt}} \quad (3.27)$$

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

$$\begin{aligned} \vec{\alpha}_P^{\text{opt}} &= \mathbb{G} \cdot \vec{\beta}_{-1}^{\text{opt}} \\ &= \mathbb{G} \cdot \mathbb{H} \cdot \vec{\beta}_0^{\text{opt}} \end{aligned} \quad (3.28)$$

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

$$\alpha_{C,k}^{\text{opt}} = \eta_k \alpha_{P,[k/2]}^{\text{opt}} \quad (3.29)$$

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

$$\mathbb{N}_{kk'} = \eta_k \delta_{[k/2]k'} \quad (3.30)$$

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

$$\vec{\alpha}_C^{\text{opt}} = \mathbb{N} \cdot \vec{\alpha}_P^{\text{opt}} \quad (3.31)$$

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:

$$\begin{aligned}\vec{\alpha}_C &= (\mathbb{G} + \delta\mathbb{G}) \cdot \vec{\beta}_0 \\ &= \vec{\alpha}_C^{\text{opt}} + \delta\mathbb{G} \cdot \vec{\beta}_0 \\ &= \mathbb{N} \cdot \vec{\alpha}_P^{\text{opt}} + \delta\mathbb{G} \cdot \vec{\beta}_0\end{aligned}\quad (3.32)$$

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

$$\begin{aligned}\vec{\beta}_{-1} &= (\mathbb{H} + \delta\mathbb{H}) \cdot \vec{\beta}_0 \\ &= \vec{\beta}_{-1}^{\text{opt}} + \delta\mathbb{H} \cdot \vec{\beta}_0\end{aligned}\quad (3.33)$$

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

$$\begin{aligned}\vec{\alpha}_P &= (\mathbb{G} + \delta\mathbb{G}) \cdot \vec{\beta}_{-1}^{\text{opt}} + \mathbb{G} \cdot \delta\mathbb{H} \cdot \vec{\beta}_0 \\ &= \vec{\alpha}_P^{\text{opt}} + (\delta\mathbb{G} \cdot \mathbb{H} + \mathbb{G} \cdot \delta\mathbb{H}) \cdot \vec{\beta}_0\end{aligned}\quad (3.34)$$

Combining Eq. (3.32) and Eq. (3.34) we obtain:

$$\vec{\alpha}_C = \mathbb{N} \cdot \vec{\alpha}_P + [\delta\mathbb{G} - \mathbb{N} \cdot (\delta\mathbb{G} \cdot \mathbb{H} + \mathbb{G} \cdot \delta\mathbb{H})] \cdot \vec{\beta}_0\quad (3.35)$$

Defining now  $\vec{\alpha}_*$  as:

$$\vec{\alpha}_* \equiv [\delta\mathbb{G} - \mathbb{N} \cdot (\delta\mathbb{G} \cdot \mathbb{H} + \mathbb{G} \cdot \delta\mathbb{H})] \cdot \vec{\beta}_0\quad (3.36)$$

when substituted in Eq. (3.35) we obtain the vector version of the linear model, Eq. (3.1), introduced in [83], namely:

$$\vec{\alpha}_C = \mathbb{N} \cdot \vec{\alpha}_P + \vec{\alpha}_* \quad (3.37)$$

According to our derivation we can now make some remarks about the variables  $\eta_*$  and  $\alpha_*$  appearing in the linear model. First, the variable  $\eta_*$  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  $\alpha_*$  is much smaller than the term  $\eta_* \alpha_P$  and is only relevant for small values of  $\alpha_P$ . We cannot say much about the statistical distribution of  $\alpha_*$ , 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  $\alpha_*$  (see Eq. (3.36)) and the arbitrary character of the perturbations  $\delta\mathbb{G}$  and  $\delta\mathbb{H}$ . This fact allows us to consider this variable independent from  $\eta_* \alpha_P$ , as the experiences in [83] confirm.

### Optimization strategies

The results in the previous subsection show that the amplitude of  $\alpha_*$  (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 3.1.1 all cascade variables  $\eta$  are equally distributed, independent of the wavelet basis from which they are derived, and their moments can be retrieved from  $\tau_p$ . In addition, the expectation value of  $|\eta|$  is fixed due to translational invariance [5, 91]:

$\langle |\eta| \rangle = 2^{-d/2}$  in an arbitrary dimension  $d$ ;  $\langle |\eta| \rangle = \frac{1}{\sqrt{2}}$  for 1D signals. According to the linear model, Eq. (3.1), the expectation value of  $|\tilde{\eta}|$  is:

$$\langle |\tilde{\eta}| \rangle = \langle |\eta_* + \alpha_* \alpha_P^{-1}| \rangle \quad (3.38)$$

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

$$\langle |\tilde{\eta}| \rangle = \langle |\eta_0| \rangle = \langle |\eta| \rangle \quad (3.39)$$

In the opposite case, for a highly non-optimal wavelet we will have that  $\alpha_*/\alpha_P \gg \eta_*$  and taking  $\alpha_*$  independent of  $\alpha_P$  we would obtain that:

$$\langle |\tilde{\eta}| \rangle = \langle |\alpha_*| \rangle \langle |\alpha_P|^{-1} \rangle = q \langle |\eta| \rangle \quad (3.40)$$

where  $q = \langle |\alpha_P| \rangle \langle |\alpha_P|^{-1} \rangle$ , which by Jensen's inequality [112] is greater than one:  $q > 1$ , for any random variable  $\alpha_P$ . For an intermediate case, the preceding two regimes are combined. If  $p$  is the proportion of the range of values of  $\alpha_P$  for which  $\eta_* > \alpha_*/\alpha_P$  and  $(1 - p)$  is its complementary, we roughly have that:

$$\langle |\tilde{\eta}| \rangle \approx p \langle |\eta| \rangle + (1 - p) q \langle |\eta| \rangle \quad (3.41)$$

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:

$$Q = \frac{\langle |\tilde{\eta}| \rangle}{\langle |\eta| \rangle} \quad (3.42)$$

which is  $Q \geq 1$ , and  $Q = 1$  for the optimal wavelet only.  $Q$  is a monotonic function of the amplitude of  $\alpha_*$  (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  $\alpha_P$ . As stated in Section 3.1.1, 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  $\alpha_P$ , 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  $\alpha_P$  are negatively correlated in suboptimal cases ( $Q > 1$ ), and uncorrelated only for the optimal wavelet:

$$Q = \frac{\langle |\tilde{\eta}| \rangle}{\langle |\eta| \rangle} = 1 - \frac{\text{Cov}(|\tilde{\eta}|, |\alpha_P|)}{\langle |\alpha_C| \rangle} \quad (3.43)$$

A standard measure of statistical dependence is the mutual information. Therefore, the mutual information between  $\tilde{\eta}$  and  $\alpha_P$ ,  $I = I(\tilde{\eta}, \alpha_P)$ , 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. Hence, when only small and short datasets are available,  $Q$  is more convenient as indicator of the optimality degree.

### 3.3 A numerical validation

Now we want to show in practice the theoretical results given in the previous section, namely the validity of the linear model, Eq. (3.37), 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. (3.18) for dyadic scale steps, and then generating the signal from these wavelet coefficients, Eq. (3.11), with the chosen wavelet basis. The multiplicative variable  $\eta$  is a random variable following a given cascade distribution without horizontal correlations, i.e., it follows Benzi *et al.*'s model [31]. As distribution for the cascade variable  $\eta$  we have chosen the log-Poisson distribution, which has been proposed in many different physical systems [35, 36, 5]. 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 [90, 5]. See Section 4.2 for a description of the log-Poisson distribution and parameters.

Regarding the linear model, it has been derived by perturbative analysis. In Figure 3.3 we validate this model in practice, for a very long series of 67 108 864 points. Figure 3.3 top shows the probability density function of the child coefficient  $\alpha_C$  conditioned by a given value of the parent coefficient  $\alpha_P$ . In Figure 3.3 top left the analysis wavelet is the optimal wavelet and in Figure 3.3 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  $\rho(\alpha_C|\alpha_P) = \rho(-\alpha_C|\alpha_P)$ , what means that  $\langle \eta_* \rangle = \langle \alpha_* \rangle = 0$ ; this also implies  $\rho(\alpha_C|\alpha_P) = \rho(\alpha_C|-\alpha_P)$ . 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:

$$\sigma_{\alpha_C|\alpha_P} = \sqrt{\langle \alpha_C^2 | \alpha_P \rangle} = \sqrt{A \alpha_P^2 + B} \quad (3.44)$$

where the constants  $A$  and  $B$  are given by the linear model:  $\langle \eta_*^2 \rangle = A$  and  $\langle \alpha_*^2 \rangle = B$ . For the optimal wavelet,  $A = \langle \eta^2 \rangle$  and  $B = 0$ , so that  $\eta_*$  coincides with  $\eta$ . Additional evidence is furnished by the conditioned histograms of logarithms of the parent and child coefficients, i.e., the conditional probability of  $\ln |\alpha_C|$  for a given value of  $\ln |\alpha_P|$ , which is shown in Figure 3.3 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. This bending is in agreement with the linear model, Eq. (3.37), as the term  $\alpha_*$  becomes dominant when  $\alpha_P$  is too small. The two asymptotic limits can be easily obtained from Eq. (3.37): when the value of the parent coefficient  $\alpha_P$  is large, in  $\ln |\alpha_C| = \ln |\eta_* \alpha_P| + \ln \left| 1 + \frac{\alpha_*}{\eta_* \alpha_P} \right|$  the second term becomes irrelevant, so that  $\ln |\alpha_C| \approx \ln |\alpha_P| + \ln |\eta_*|$ . When the value of the parent coefficient  $\alpha_P$  is small, in  $\ln |\alpha_C| = \ln |\alpha_*| + \ln \left| 1 + \frac{\eta_* \alpha_P}{\alpha_*} \right|$  the second term rapidly becomes irrelevant, so that  $\ln |\alpha_C| \approx \ln |\alpha_*|$ . Not only the asymptotes, but also the central behaviour is the one given by the model, as the line of maximum-probability of the histogram fits a shape:

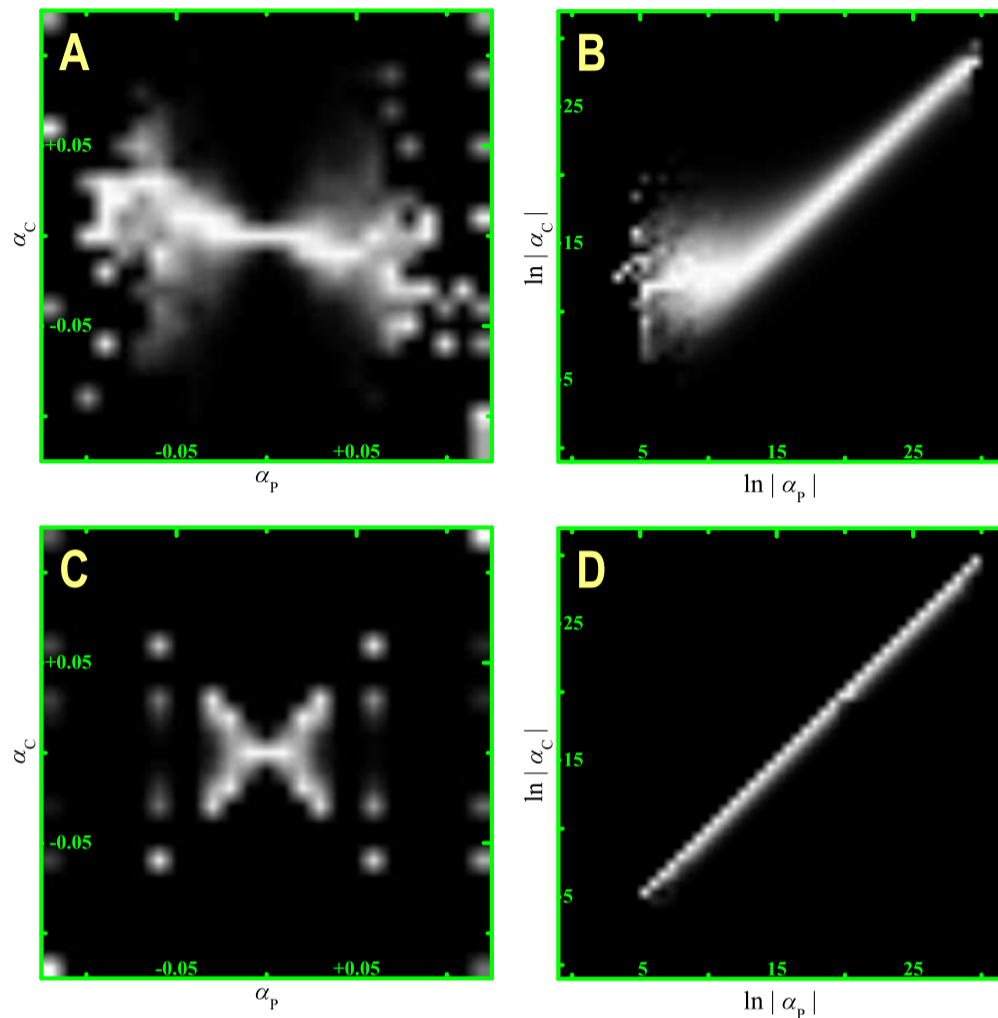
$$\ln |\alpha_C|_{\text{m.p.}} = \ln (|\alpha_*|_{\text{m.p.}} + |\eta_*|_{\text{m.p.}} \exp \ln |\alpha_P|) \quad (3.45)$$

where m.p. stands for *maximum probable*, i.e., these values are the probability maxima of their respective distributions. In addition, the amplitude of the fluctuations of  $\alpha_*$  is larger than that of  $\eta_*$ . That is why the left side of the histogram shows large dispersion that is reduced as  $\ln |\alpha_P|$  grows and tends to that of the optimal case in the right side.

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), Symlet (orders 4 to 8) and Battle-Lemarié (spline wavelets) (orders 1, 2, 3 and 6). Notice that Haar and Daubechies 1 coincide, while Symlet 1 to 3 also coincide with Daubechies 1 to 3 respectively, and for that reason we have not repeated them (see [99] for a description of these wavelet bases). 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 theoretical results presented before.

In Figure 3.3 we present the joint histograms of  $\ln |\alpha_C|$  vs.  $\ln |\alpha_P|$ , 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 behaviour seen in Figure 3.3, bottom left. In contrast, when an ensemble is analyzed with a suboptimal wavelet the histogram bends on the left to a horizontal line, as in Figure 3.3, bottom right. As the optimal and analyzing wavelets





**Figure 3.3:** Example that shows the influence of parent coefficients,  $\alpha_p$ , over child coefficients,  $\alpha_C$ , in the non-optimal and optimal cases. We plot the joint histograms of  $\alpha_C$  vs.  $\alpha_p$  and  $\ln |\alpha_C|$  vs.  $\ln |\alpha_p|$  for synthetic cascade data generated with the Coiflet-1 wavelet and analyzed, Eq. (3.14), 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  $25 \times 25$  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  $50 \times 50$  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 Section 4.2 for a detailed description of the process). The series has been obtained with Benzi *et al.*'s model [31]: (i) random, independent cascade variables are generated, (ii) from them we get the optimal coefficients following Eq. (3.16), and (iii) the series is synthesized from Eq. (3.11) using a wavelet that will be the optimal one by construction.

become more different, the amplitude of the term  $\alpha_*$  increases and hence the extension of the horizontal line in the joint histogram becomes longer.

In Table 3.3 we present the results of the mutual information  $I$  between  $\tilde{\eta}$  and  $\alpha_P$ , and the  $Q$  parameter as defined in Eq. (3.42) 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 higher. 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.

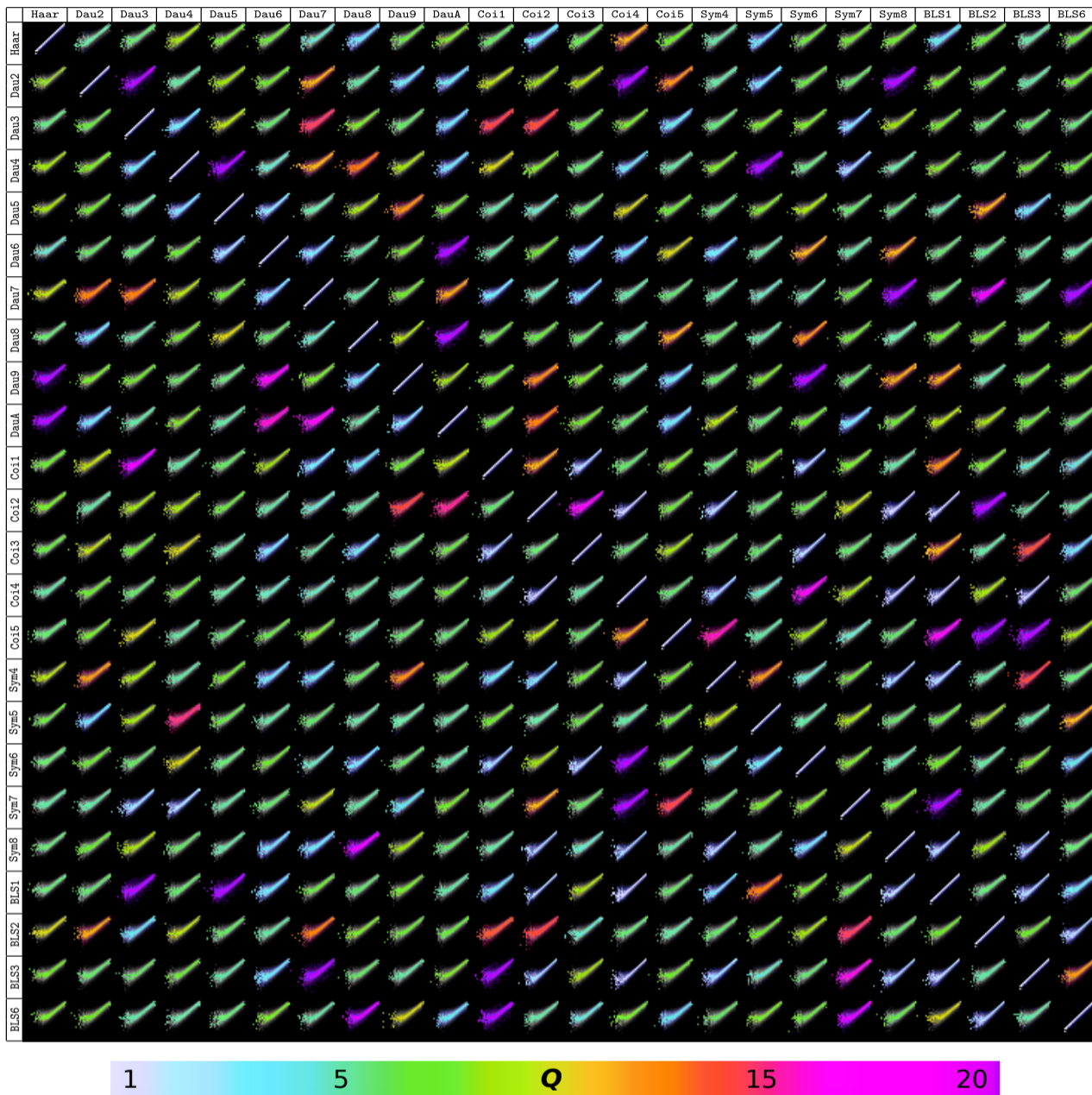
The  $Q$  parameter is obtained by means of the average of  $\tilde{\eta}$  and so, according to the Central Limit Theorem, it converges to its theoretical value with a standard deviation that depends on the number of samples  $N$  as  $N^{-\frac{1}{2}}$ ,  $\sigma_{\langle|\tilde{\eta}|\rangle} = \sigma_{|\tilde{\eta}|} N^{-\frac{1}{2}}$  (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  $0.62 N^{-\frac{1}{2}}$ .

The estimation of the mutual information  $I$  has an uncertainty of standard deviation also proportional to  $N^{-\frac{1}{2}}$ . The proportionality constant is  $\sqrt{\langle(\log_2 p_{x,y})^2\rangle + \langle(\log_2 p_x)^2\rangle + \langle(\log_2 p_y)^2\rangle}$  [8], which for our log-Poisson distribution is 5.66 bits. In addition, we have not taken into account other sources of uncertainty that do not depend

	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
Dau2	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
Dau3	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
Dau4	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
Dau5	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
Dau6	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
Dau7	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
Dau8	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
Dau9	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
DauA	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
Coi1	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
Coi2	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
Coi3	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
Coi4	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
Coi5	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
Sym4	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
Sym5	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
Sym6	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
Sym7	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
Sym8	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
BLS1	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
BLS2	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
BLS3	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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
BLS6	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 3.1:** 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). Each generated ensemble corresponds to 64 series of 4096 points 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 0.002 for  $Q$  and 0.02 bits for  $I$ .



**Figure 3.4:** Joint histograms of  $\ln \alpha_C$  vs.  $\ln \alpha_P$  for synthetically generated cascade data. This is the exhaustive version of Figure 3.3 here showing every possible generating wavelet – analysis wavelet combination. In each histogram, each column has been normalized so that vertical slices correspond to the probability distribution function of  $\ln \alpha_C$  conditioned to the value of  $\ln \alpha_P$  in the horizontal axis. This way, dark corresponds to zero probability and bright corresponds to maximum probability. Values range  $-20$  to  $1$  (nondimensional) in both axes, the same for all the histograms. Each histogram has  $30 \times 30$  bins. The histograms have been arranged as in Table 3.3, i.e., generation wavelet (row) vs. analysis wavelet (column), so that the main diagonal corresponds to the optimal wavelet cases. In addition, we have coloured them depending on the optimality parameter  $Q$ , Eq. (3.42) (colour brightness indicates the probability and chrominance channels indicate the value of  $Q$ , according to the palette shown below the histograms). Generated ensembles consist of 64 series of 4096 points following a log-Poisson cascade with  $D_\infty = 0$  and  $h_\infty = -\frac{1}{2}$ .

on  $N$ , namely sampling discretization and the influence of unsampled tails.

As stated in the caption of Table 3.3, the absolute uncertainty for  $I$  is 10 times that of  $Q$ , even when 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.



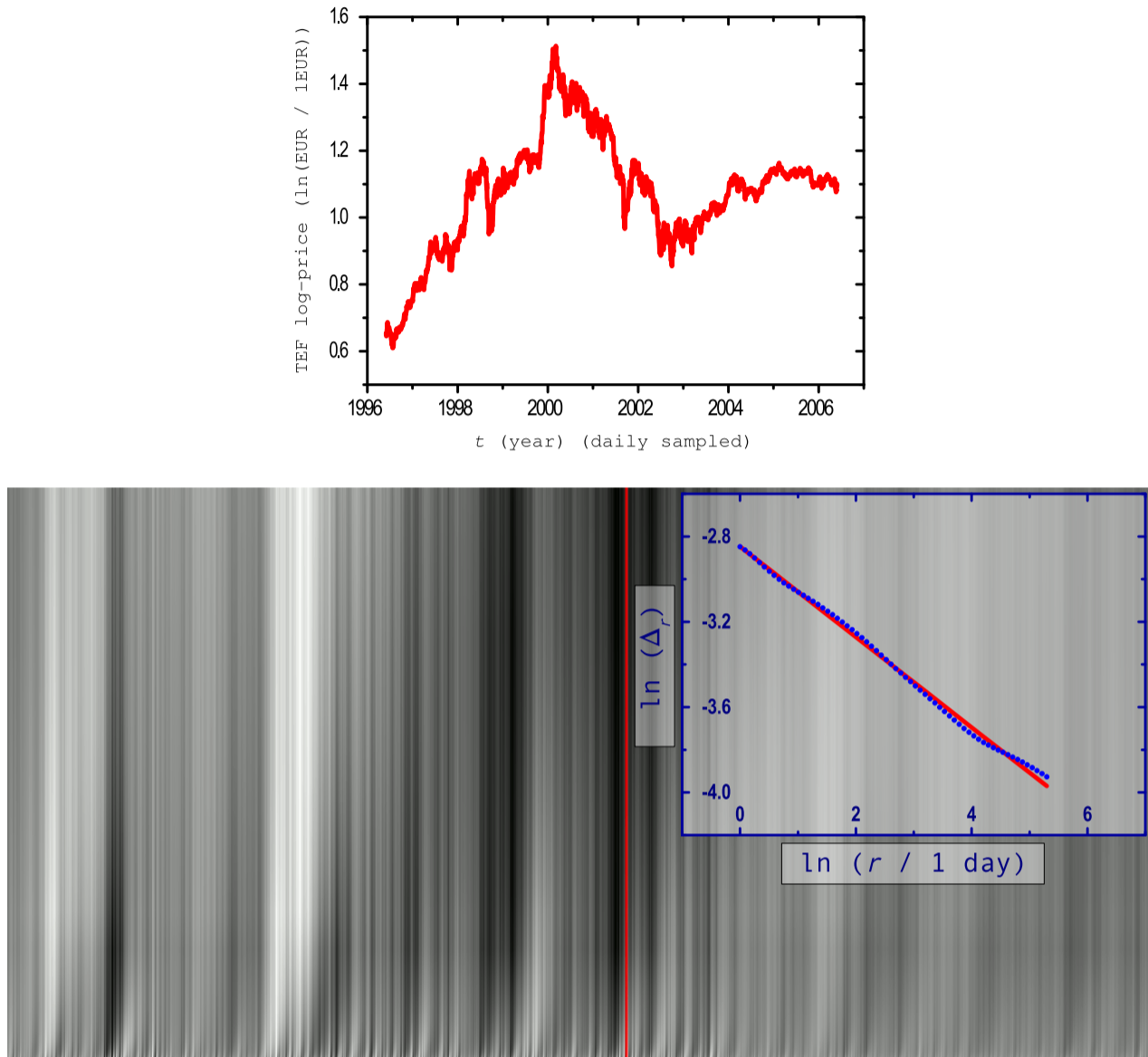
## 4 Cascades are multifractal

### 4.1 Singularity scaling

A sign of scale invariance is the fact that the distribution of some quantities subrogated at a variable scale evolves with the scale, but the distributions at different scales are related by the scale ratio only. In particular, the moments of the distribution are related by a factor that is a power law in terms of the scale ratio.

The main assumption behind the **microcanonical multifractal formalism (MMF)** is the existence of power-law scaling at each point of the system. These local power laws are governed by the local *singularity exponents*, which sometimes are simply called singularities, even when the exponent is positive and the behaviour is not singular but regular. In Figure 4.1 we show an example of this scaling.

*Singularity components* are the sets formed by all the points that have the same value of singularity exponent. As we can see, in this kind of systems singularity components are dense sets. In the next section we show that singularity components are fractal, so that the system is multifractal under the MMF. This means that the distribution of singularities evolves with the scale following another power law, where the exponent of this power law is the *fractal dimension* of the component [113].



**Figure 4.1:** Multifractal time series corresponding to the daily quotation of Telefónica in the Spanish market (top) and scalogram of its increments (bottom). The processed variable is the logarithm of the price in Euros, so that their increments are relative and non-dimensional (these are called *returns*). Increments have been normalized by the scale ( $r$ -day returns) so that they represent the finite-scale derivative. The scalogram is a level plot that shows the value of the increment (black corresponds to the minimum; white corresponds to the maximum) depending on the time (horizontal) and the scale of increment (vertical). Horizontal axis goes from 1997-11-24 to 2006-06-23, a total of 2152 (trading) days. Vertical axis goes from the scale of 1 day to that of 200 days. For proper contrast, each scale has been normalized so that its minimum is black and its maximum is white. Inset: for the randomly picked day #1164 (red line of the scalogram) it shows that the increment at scale  $r$  follows a power law:  $\Delta_r(t) = A(t) r^{h(t)}$ .  $h$  is often called Hölder exponent or Hurst exponent and it takes a value of -0.21 in this point (as we normalize by the scale, this corresponds to a divergence in the derivative; the logarithm of the price is  $s(t) \sim r^{h(t)+1}$ , i.e., it has a positive exponent and so a regular behaviour). Regression coefficient is 0.998, meaning that the estimation is of very high quality. Similar results are obtained in all the points. To stabilize the numerical calculation of the derivative, we have filtered it by projecting through a Gaussian wavelet (see Section 2.2 for details).



## 4.2 Connection of the microcanonical cascade with the singularity formalism

A variable is said to be multifractal in the MMF sense [6] if an intensive functional  $\mathbf{T}_r$  acting on this variable (see Section 3.1.1) can be characterized by local scaling relations of the type:

$$\mathbf{T}_r(\vec{x}) = \alpha(\vec{x}) r^{h(\vec{x})} + o\left(r^{h(\vec{x})}\right) \quad (4.1)$$

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  $\vec{x}$  [5, 6]. A signal verifying Eq. (4.1) is said “multifractal” (in the MMF 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 (Hausdorff) singularity spectrum of the signal [113].

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 [23] proved that the knowledge of  $D(h)$  granted the knowledge of the distribution of the cascade variables  $\eta$  through the knowledge of the multiscaling exponents  $\tau_p$ , as expressed by Eq. (3.2). It follows that  $\tau_p$  is related to the singularity spectrum of the multifractal through a Legendre transform:

$$\tau_p = \inf_h \{p h + d - D(h)\} \quad (4.2)$$

which is known as the Parisi-Frisch formula and is the cornerstone of the canonical multifractal formalism [33, 37]. An interesting corollary of Eq. (4.2) 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  $\tau_p$ , namely:

$$D_L(h) = \inf_p \{p h + d - \tau_p\} \quad (4.3)$$

The function  $D_L(h)$  is the so-called Legendre singularity spectrum [23, 114, 115], 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.

Having to pass through Eq. (4.3) to obtain the singularity spectrum is not generally an issue when we have an analytical model for  $\tau_p$ . However, the picture is completely different when we calculate  $\tau_p$  from empirical data: this requires estimation of high-order positive and negative moments of  $\mathbf{T}_r$ , what is not available unless we have huge amounts of data to process. When the second derivative of  $\tau_p$  has small variation with  $p$ , this fact limits the range of solvable singularity exponents and so it prevents the obtaining of the tails of  $D(h)$ , meaning that often we can only access its central part [37, 99]. This limitation can be important, especially if we take into account that the left-most point corresponds to the most informative fractal component, the so-called *most singular manifold*, whose precise estimation is of crucial relevance since it governs the dynamics of the system and can reconstruct the whole signal [92].

A more direct approach to obtain the  $D(h)$  that eliminates the necessity of imposing convex spectra is that of the MMF. When the cascade variables are accessible, the MMF method to obtain the  $D(h)$  consists in calculating the limit as  $\kappa \rightarrow 0$  of the distribution of cascade singularity exponents. The cascade singularity exponents are defined

as follows:

$$h_\kappa = \log_\kappa \eta_\kappa = \frac{\ln \eta_\kappa}{\ln \kappa} \quad (4.4)$$

where  $\eta_\kappa$  is the multiplicative cascade variable that relates  $\mathbf{T}_r$  with  $\mathbf{T}_L$ ,  $\kappa = r/L$ , as in Eq. (3.5). The cascade singularity exponents represent the singularity exponents in the same sense of Eq. (4.1) when they are obtained at the resolution level [6], i.e., when the scale ratio  $\kappa$  is the one that compares the largest (whole-domain wide) scale  $L$  with the smallest (resolution-level) scale  $r$ , meaning that  $r \ll L$  or equivalently  $\kappa \rightarrow 0$ . As the singularity components  $F_{h_\kappa}$  are of fractal character, the distribution of singularity exponents at a given observation scale behaves as [113]:

$$\rho(h_\kappa) \sim \kappa^{d-D(h_\kappa)} \quad (4.5)$$

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

$$\lim_{\kappa \rightarrow 0} \frac{\ln \rho(h_\kappa)}{\ln \kappa} = d - D(h) \quad (4.6)$$

where:

$$h \equiv h_0 = \lim_{\kappa \rightarrow 0} h_\kappa \quad (4.7)$$

This obtaining of the  $D(h)$ , based on the scaling of the singularity distribution, is inspired by other alternatives to derive singularity spectra that avoid passing through Eq. (4.2) by calculating the scaling of measures [75, 116, 117].

*Lemma: The singularity spectrum derived according to Eq. (4.6) coincides with the Legendre spectrum, Eq. (4.3), when the singularity spectrum is convex.*

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

$$h_\kappa = \frac{\ln \eta_\kappa}{\ln \kappa} \quad (4.8)$$

As the cascade variable  $\eta_\kappa$  is derived from a multifractal signal, the limit in Eq. (4.6) exists and it is  $d - D(h)$  (the Hausdorff spectrum of the signal) [6]. Therefore, the distribution of  $h_\kappa$  has a leading order  $\kappa^{d-D(h_\kappa)}$  as follows:

$$\rho(h_\kappa) = A_\kappa \kappa^{d-D(h_\kappa)} + o(\kappa^{d-D(h_\kappa)}) \quad (4.9)$$

for small values of  $\kappa$ . Recalling here Eq. (3.4) we have:

$$\tau_p = \lim_{\kappa \rightarrow 0} \frac{\ln \langle \eta_\kappa^p \rangle}{\ln \kappa} \quad (4.10)$$

We then expand it to find that:

$$\begin{aligned} \tau_p &= \lim_{\kappa \rightarrow 0} \frac{1}{\ln \kappa} \ln \left( \int dh_\kappa \kappa^{h_\kappa p} \rho(h_\kappa) \right) \\ &= \lim_{\kappa \rightarrow 0} \frac{1}{\ln \kappa} \ln \left( \int dh_\kappa \kappa^{h_\kappa p} A_\kappa \kappa^{d-D(h_\kappa)} \right) \\ &= \lim_{\kappa \rightarrow 0} \inf_{h_\kappa} \{h_\kappa p + d - D(h_\kappa)\} \\ &= \inf_h \{hp + d - D(h)\} \end{aligned} \quad (4.11)$$

where we used the saddle-point approximation. Notice that Eq. (4.11) is analogous to Eq. (4.2). Recalling that the inverse of a Legendre transform on convex functions is another Legendre transform, if we obtain now the Legendre spectrum, Eq. (4.3), 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:

$$\rho(\ln \eta_\kappa) = \frac{1}{\sqrt{2\pi\sigma_\kappa^2}} e^{-\frac{1}{2}\left(\frac{\ln \eta_\kappa - \mu_\kappa}{\sigma_\kappa}\right)^2} \quad (4.12)$$

Hence, the  $\tau_p$  as defined in Eq. (3.2) are given by:

$$\tau_p = \frac{\mu_\kappa}{\ln \kappa} p + \frac{\sigma_\kappa^2}{2 \ln \kappa} p^2 \quad (4.13)$$

Let  $h_m = \mu_\kappa / \ln \kappa$  and  $\sigma_h^2 = -2\sigma_\kappa^2 / \ln \kappa$  (remember that  $\kappa < 1$ ), so Eq. (4.3)) leads to the singularity spectrum  $D(h)$ :

$$D(h) = d - \left( \frac{h - h_m}{\sigma_h} \right)^2 \quad (4.14)$$

Let us show now that Eq. (4.6) leads to the same expression. Notice that Eq. (4.4) 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. (4.12) to obtain:

$$\frac{\ln \rho(h_\kappa)}{\ln \kappa} = \left( \frac{h - h_m}{\sigma_h} \right)^2 + \frac{\ln \sqrt{\frac{-\ln \kappa}{\pi \sigma_h^2}}}{\ln \kappa} \quad (4.15)$$

and the second term vanishes as  $\kappa \rightarrow 0$  leading to Eq. (4.14). It follows that Eq. (4.6) 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  $\lambda$ . Then the distribution of  $\ln \eta_\kappa$  is:

$$\rho(\ln \eta_\kappa) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \delta(\ln \eta_\kappa - h_\infty \ln \kappa - n \ln \beta) \quad (4.16)$$

which is discrete, i.e., it only takes nonzero values for some values of  $\ln \eta_\kappa$ . The parameter  $h_\infty$  is the singularity exponent of the Most Singular Component (MSC) [5, 118], while the parameter  $\lambda$  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  $\tau_p$  are given by:

$$\tau_p = ph_\infty + (d - D_\infty)(1 - \beta^p) \quad (4.17)$$

and through Eq. (4.3) the singularity spectrum is:

$$D(h) = D_\infty + (d - D_\infty)\omega(h)(1 - \ln \omega(h)) \quad (4.18)$$

with

$$\omega(h) = -\frac{1}{\ln \beta} \frac{h - h_\infty}{d - D_\infty} \quad (4.19)$$

Let us now apply Eq. (4.6). From Eq. (4.4) and Eq. (4.16), the  $h_\kappa$  deviates from the most singular exponent  $h_\infty$  in an integer number  $n$  of contributions  $\log_\kappa \beta$ , namely

$$h_\kappa = h_\infty + n \underbrace{\frac{\ln \beta}{\ln \kappa}}_{\Delta h_\kappa} \quad (4.20)$$

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

$$\begin{aligned} \omega(h_\kappa) &= \frac{n}{\lambda} \\ &= \frac{1}{d - D_\infty} \frac{n}{(-\ln \kappa)} \\ &= -\frac{1}{\ln \beta} \frac{h_\kappa - h_\infty}{d - D_\infty} \end{aligned} \quad (4.21)$$

Notice that  $\omega(h_\kappa)$  is positive and proportional to  $\Delta h_\kappa$ . We now recall Eq. (4.16) to obtain:

$$\frac{\ln \rho(h_\kappa)}{\ln \kappa} = \frac{-\lambda + n \ln \lambda - \ln n!}{\ln \kappa} \quad (4.22)$$

Hence, according to Eq. (4.6), the singularity spectrum is:

$$D(h) = d - (d - D_\infty) + \lim_{\kappa \rightarrow 0} \frac{n \ln \lambda - \ln n!}{-\ln \kappa} \quad (4.23)$$

Where  $h = h_{\kappa \rightarrow 0}$  as in Eq. (4.7). For any  $h_\kappa$  different from  $h_\infty$ , i.e.,  $\Delta h_\kappa \neq 0$ , when  $\kappa$  goes to 0,  $n$  grows accordingly, because  $n$  is proportional to  $(-\ln \kappa)$ . So the limit  $\kappa \rightarrow 0$  is equivalent to  $n \rightarrow \infty$ :

$$D(h) = D_\infty + \lim_{n \rightarrow \infty} \frac{n \ln \lambda - n \ln n + n - \ln(\sqrt{2\pi n})}{-\ln \kappa} \quad (4.24)$$

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

$$D(h) = D_\infty + (d - D_\infty) \lim_{n \rightarrow \infty} (\ln \lambda - \ln n + 1) \omega(h_\kappa) \quad (4.25)$$

which, as  $\omega(h_\kappa) = n/\lambda$ , leads to Eq. (4.18).

### 4.3 Regularization of diverging measures

Another useful extension of the MMF singularity analysis comes into scene when it is faced to diverging measures. Such behaviour should not be expected when the analyzed signal is an observable coming from a real physical system. However, it is not unusual to find it in popular mathematical models as widespread as the Brownian motion [119] among others.

When performing singularity analysis from a scalogram (see Figure 4.1) the operator used is a key point. Increments, even if projected through a wavelet, are sometimes not enough to filter the singularity exponent from polynomial long-range correlations. In these cases, the scale-tunable variable analyzed is a measure of the signal [120, 7]. A good measure for this purpose is:

$$\mu_r(t) \equiv \int_{B_r} dt' |\nabla s|(t') \quad (4.26)$$

A standard finite-difference discretisation at resolution scale  $\delta$  of Eq. (4.26), is the following:

$$\mu_r^{(\delta)}(t) = \sum_{i=1}^N \delta \left| \nabla^{(\delta)} s(t + i\delta) \right| \quad (4.27)$$

where  $N = r/\delta$  is the effective number of points that contribute to the measure. The gradient,  $\nabla s(t)$ , is approximated at resolution  $\delta$  as:

$$\nabla^{(\delta)} s(t) = \frac{\Delta_\delta s(t)}{\delta} \quad (4.28)$$

where  $\Delta_\delta s(t) = s(t + \delta) - s(t)$ .

For well-behaved gradients, all the dependency on  $\delta$  is removed when limits are taken, i.e.,

$$\nabla^{(\delta)} s(t) \xrightarrow{\delta \rightarrow 0} \nabla s(t) \quad (4.29)$$

and  $\mu_r^{(\delta)}$  will converge to the actual measure  $\mu_r$  as  $\delta \rightarrow 0$ . However, this convergence requires that the gradient  $\nabla s$  is well-defined, at least in a distributional sense. That is,  $\nabla s$  can diverge to  $\infty$  at some points, but it has to follow a well-defined distribution. In other words, the



curve  $|\nabla s|(t)$  can diverge in a zero-measure set of points, but it has to be integrable.

Consider two measures defined on the same interval, approximated with two different scale quanta  $\delta, \delta'$ . If  $\nabla s$  is well-defined, both converge to the same quantity, namely:

$$\mu_r^{(\delta)}(t) = \mu_r^{(\delta')}(t) \quad \delta, \delta' \rightarrow 0 \quad (4.30)$$

and both approximate  $\mu_r(t)$ .

However, when the gradient  $\nabla s$  is not well-defined, its discrete estimation  $\nabla^{(\delta)}s$  does depend on  $\delta$ . This is typically the case of fractal functions with infinite curve length, i.e., infinite  $\mu_r$ . If  $\nabla s$  is not defined in a distributional sense, neither is  $\mu_r$ , and the numerical, finite-size estimates of the measure at a given point  $t$ ,  $\mu_r^{(\delta)}(t)$  and  $\mu_r^{(\delta')}(t)$ , cannot converge to a fixed quantity:

$$\mu_r^{(\delta)}(t) \neq \mu_r^{(\delta')}(t) \quad \delta, \delta' \rightarrow 0 \quad (4.31)$$

Therefore,  $\mu_r^{(\delta \rightarrow 0)}(t)$ , will not scale as  $r^{1+h(t)}$  as expected. On the contrary, it usually grows proportionally to the number of points  $N$ , leading to an incorrect estimation of singularity exponents:  $h_{\text{est}}(t) = 0$ .

To solve this problem, one option would be to scale the infinitesimal  $\delta$  according to  $r$ , i.e.,  $\delta = r/N$ , thus leading to the right scaling in  $r$ ,  $\mu_r^{(\delta)}(t) \sim r^{1+h(t)}$ . However, the signals to be analyzed are sampled at a fixed resolution scale  $\delta$  and we cannot change it.

We can proceed in a different way, providing a redefinition of the measure, well-behaved and giving access to the true scaling properties of the signal. In order to give a physical meaning to the measure,

we must require that  $\mu_r^{(\delta)} = \mu_r^{(\delta')}$  as  $\delta, \delta' \rightarrow 0$ . So, we propose to regularize the measure in the following way:

$$\mu_r \equiv \lim_{\delta \rightarrow 0} \frac{\mu_r^{(\delta)}}{N^\alpha} \quad N = \frac{r}{\delta} \quad (4.32)$$

where  $\alpha$  is an exponent to regularize the behaviour of  $\mu_r$ . The value of  $\alpha$  is not known *a priori*, but can be easily estimated from two different discretizations,  $\delta$  and  $\delta'$ , which are assumed to be very small:

$$\frac{\mu_r^{(\delta)}}{N^\alpha} = \frac{\mu_r^{(\delta')}}{N'^\alpha} \quad (4.33)$$

with  $r = N\delta = N'\delta'$ . The expected value of  $\mu_r^{(\delta)}$  is:

$$\langle \mu_r^{(\delta)} \rangle = N^{1+h_{\text{est}}} \langle \mu_\delta^{(\delta)} \rangle \quad (4.34)$$

where  $h_{\text{est}}$  is the incorrectly estimated singularity exponent. From Eq. (4.27), it follows that  $\langle \mu_\delta^{(\delta)} \rangle = \langle |\Delta_\delta s| \rangle$  and taking expected values on Eq. (4.33) we obtain that:

$$\delta^{\alpha-h_{\text{est}}-1} \langle |\Delta_\delta s| \rangle = \delta'^{\alpha-h_{\text{est}}-1} \langle |\Delta_{\delta'} s| \rangle \quad (4.35)$$

so we can easily estimate  $\alpha$  from:

$$\alpha = 1 + h_{\text{est}} - \frac{\log(\langle |\Delta_{\delta'} s| \rangle / \langle |\Delta_\delta s| \rangle)}{\log(\delta'/\delta)} \quad (4.36)$$

Notice that we can always apply the generalized measure definition, Eq. (4.32), that is, we do not need to know *a priori* whether the signal  $s$  has an undefined gradient or not. If the standard measure is well-behaved  $\mu_r^{(\delta)}$  and  $\mu_r^{(\delta')}$  directly approximate  $\mu_r$  and, from Eq. (4.33), it follows that  $\alpha = 0$  and the correct singularity exponent  $h_{\text{est}} = h$  is directly estimated (notice that  $h$  is the singularity exponent referred to the gradient, not to the signal).



## 5 Application to stock market series

**Stock market series** are a kind of self-similar signals that can be represented by microcanonical cascades in both the logarithm of the price and the volatility. In the following we verify this and obtain their respective closest-to-optimal wavelets. As a promising application in forecasting, we derive the distribution of the value of next point of the series conditioned to the knowledge of past points and the cascade structure, i.e., the stochastic kernel of the cascade process. These results have been published in [10].

Stock markets, as well as many other trading markets, are formed by a great diversity of interacting agents, each with their own characteristics, such as reacting times, budgetary constraints and so on. As a consequence, and due to the large amount of agents taking part in a typical market, many econometric indicators behave in a complex, scale-invariant fashion, a feature that has been taken into account in many different models [69, 63, 72]. However, scale invariance can be exploited not only in the design of models, but also in analysis tools capable of extracting new information from time series of dynamical systems. One of the most promising theories for the description of scale-invariant data concerns multifractal systems [71, 95], and more particularly multifractal systems in the microcanonical approach [95, 96, 6]. With the aid of *microcanonical cascades* one can maximize the amount of information that some scales convey about the others. Besides, it is possible to provide an analytical model de-

scribing the evolution of the series, and produce high quality forecasts for both the returns and volatility evolution of stock market data.

In Section 5.2.1 the best wavelet from a bank of standard wavelet bases is obtained. Then, the direct cascade model furnished by the optimal wavelet is used to derive analytically the stochastic kernel, namely the distribution of future values conditioned by the known past values, and the entire Section 5.3 is devoted to this goal.

## 5.1 Description of the data and notation

We have processed a group of data belonging to the Spanish stock market (IBEX). The group is formed by daily series of 35 different assets (those with the largest liquidity in the Spanish market) during approximately ten years (from June, 1996 to June, 2006, although some series are slightly shorter) containing a total of 76663 points. We do not try to correct systematic deviations by any mean. In that sense, we always identify the ending of a session as the instant just preceding the opening of the following, no matter the actual time interval between them (sometimes several non-working days). An example of this series has been shown in Figure 4.1.

We are interested in relative variations of the price, i.e., the ratio of the absolute value to the absolute variation. For that reason, we will work on series formed by logarithms of prices. In this way the absolute variation between two consecutive instants (approximately, the derivative with respect to time) approximates the relative variation for the original stock series.

These series have been shown to have multifractal properties in the sense of MMF [95, 96, 6], so they are appropriate for the present study. Throughout the paper, time series will be represented by a dis-

crete, time-ordered collection of values  $x_1, x_2, \dots$ , where  $x_n = x(t_n)$  and the sampling times  $t_n$  are equally spaced,  $t_i = t_0 + n\Delta t$ . For description and forecast, we will take a finite number  $N$  of known past values, forming a vector  $\vec{x} = (x_1, \dots, x_N)$ . The unknown value at the following time instant (which would correspond to  $x_{N+1}$ ) will be denoted with a different letter,  $y$ , to emphasize that this is the forecast.

## 5.2 Scale invariant properties of the stock market cascade

As we have already commented, cascade variables  $\eta$  are not scale invariant, and neither is their distribution. However, they can be related to scale invariant quantities through the introduction of the Microcanonical Multifractal Formalism (MMF) [6].

According to Eq. (3.7), any wavelet coefficient  $\alpha_{j,k}$  is distributed as the product of  $j$  independent dyadic cascade variables, in the way:

$$\alpha_{j,k} \doteq \prod_{j'=0}^{j-1} \eta_{j-j', \left[\frac{k}{2^{j'}}\right]} \alpha_{0,0} \quad (5.1)$$

If the variables  $\eta_{j,k}$  are infinitely divisible, they verify:

$$\eta_{j,k} = \left(\frac{1}{2}\right)^{\tilde{h}_{j,k}} \quad (5.2)$$

where  $\tilde{h}_{j,k}$  is a scale-invariant quantity, the transition singularity exponent between the scales  $j$  and  $j-1$  at the point  $k$  [8, 6]. We can thus define a punctual estimate of the singularity exponent associ-

ated to the whole cascade at the point  $2^{-j}k$  by taking logarithms in Eq. (5.1) and normalizing by the scale factor  $j \log 2$ , namely:

$$\hat{h}_{j,k} = -\frac{\log \alpha_{j,k}}{j \log 2} \doteq h_{j,k} - \frac{\log \alpha_{0,0}}{j \log 2} \quad (5.3)$$

where  $h_{j,k}$  is a dimensionless, scale invariant field known as singularity exponent, which is the average of the transition singularity exponents,

$$h_{j,k} = \frac{1}{j} \sum_{j'=0}^{j-1} \tilde{h}_{j-j', \left[ \frac{k}{2^{j'}} \right]} \quad (5.4)$$

as we have presented in Section 4.2.

The distribution of singularity exponents as obtained at a scale  $j$  is not scale invariant. However, it can be related to the singularity spectrum of the underlying multifractal hierarchy [22, 37]. Let  $r = 2^{-j}$  be the scale ratio associated to  $h_{j,k}$ , according to Eq. (5.4). The distribution of values of  $h_{j,k}$ ,  $\rho(h_{j,k})$ , verifies:

$$\rho(h_{j,k}) = A_0 r^{d-D(h_{j,k})} \quad (5.5)$$

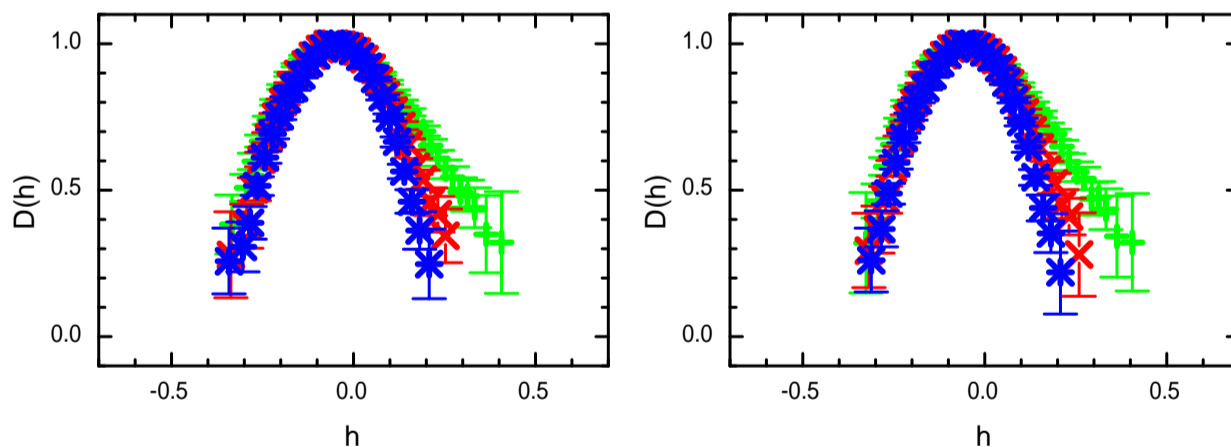
where  $d$  is the dimension of the embedding space ( $d = 1$  for time series) and  $D(h)$  is the function relating the value of the singularity exponent,  $h$ , with the fractal dimension of the associated singularity component. The maximum dimension corresponds to the fractal component associated to the maximum of the distribution  $\rho(h_{j,k})$ . This maximum dimension is known, because it coincides with the dimension of the support of the multifractal, i.e., it coincides with  $d$  when the support of the multifractal is the whole space. Then the singularity spectrum can be directly retrieved from Eq. (5.5) by a log-log



transformation [113, 118, 6]:

$$D(h_{j,k}) = d + \frac{\rho(h_{j,k})/\rho_0}{j \log 2} \quad (5.6)$$

where  $\rho_0$  is the maximum value of the function  $\rho(h_{j,k})$ . The singularity spectrum is a global scale invariant quantity, so for any value  $h_0$  and any  $j, k, j', k'$  we should observe  $D(h_{j,k} = h_0) = D(h_{j',k'} = h_0)$ . In fact, the equality of these factors at different scales is one of the conditions for MMF to be valid [6]. In Fig. 5.1 we show that the experimental singularity spectra obtained from the series of returns and from the series of volatilities (derived from our IBEX 35 dataset) are coincident within the experimental uncertainty at three different scales (notice that the right tail is always worse determined; see discussion in [118]). Remarkably enough, return and volatility singularity spectra are very similar. This should be expected because volatility is a measure of the amplitude of returns. As singularity exponents describe function regularity [6], those of returns approximately coincide with those of volatilities.



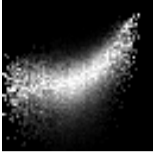

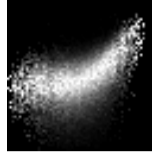
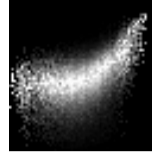
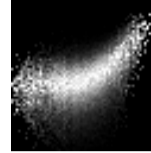
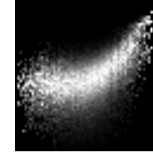
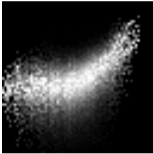
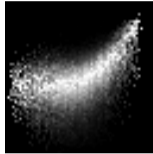
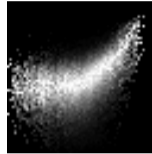
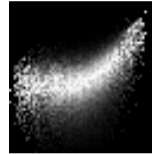
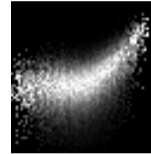
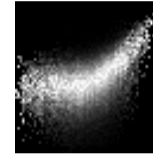
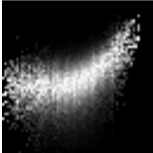
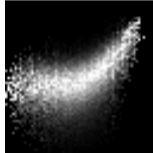
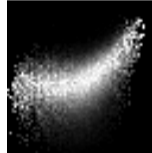
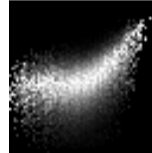
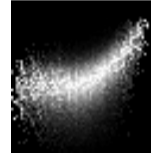
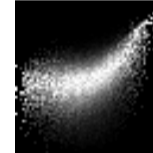
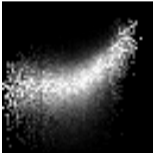
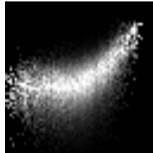
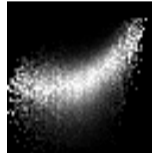
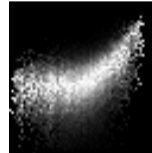
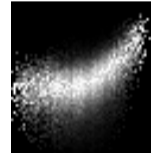
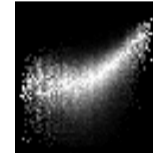
**Figure 5.1:** Singularity spectra derived from returns (left) and volatilities (right). The absolute values of returns and the squares of volatilities have been averaged at three different time windows, i.e., they have been projected with a box wavelet. The scales (sizes of the windows) are as follows: +: 2 days; ×: 4 days; \*: 8 days. For simplicity, we use a box wavelet, as we have seen that the marginal distributions of singularity exponents,  $\rho_{h_{j,k}}$ , are the same with almost any wavelet (the optimal wavelet basis is relevant only for calculation of the parent/child joint distribution).

Let us finally remark that the knowledge of the singularity spectrum  $D(h)$  provides information of the distribution of the cascade variables at any scale. By the application of Eq. (5.3) and Eq. (5.5) we can know the distribution of  $\alpha_{j,k}$  at any scale  $j$ .

### 5.2.1 Optimal wavelet for Spanish stock market series

We have constructed a bank of 24 orthonormal dyadic wavelet bases. The bases included in this bank are some of the most frequently used in the wavelet literature: Haar, Daubechies of orders 2 to 10 (signatures Dau2–9 and DauA), Symlets of orders 4 to 9 (signatures Sym4–9), Coiflets of orders 1 to 5 (signatures Coi1–5), and spline or Battle-Lemarié of orders 1,2,3 and 6 (signatures BLS1–3 and BLS6). We have searched which wavelet in this bank is closest to optimality for our dataset according to the  $Q$  criterion. Results are summarized in Table 5.2.1.

It is observed that the best wavelet (in terms of optimality) in the bank is the Battle-Lemarié wavelet of order 3. A similar experience conducted on volatility series throws that the optimal wavelet is a different one, Symlet of order 7 (in that case, Sym7 attains a  $Q = 2.22$  while for other wavelets of the bank the average value of  $Q$  is 2.95 and the farthest from optimality is Haar with  $Q = 5.71$ ). Should one expect to have a relation between the optimal wavelet for log-prices and that of volatility? In fact the answer is yes. Volatility is the amplitude of return variations, i.e., it represents some kind of modulus of the derivative of log-prices, and the optimal wavelet of the derivative series is the derivative of the optimal wavelet of the series, as explained in [91]. As shown in Fig. 5.2, Symlet of order 7 is very similar to the derivative of Battle-Lemarié of order 3. This confirms

BLS3  1.87	Sym6  1.95	Coi3  1.95	Sym4  1.99	Coi5  2.00	BLS1  2.01
Coi4  2.06	Coi1  2.09	Dau4  2.10	Dau3  2.11	Dau7  2.18	Dau6  2.20
Sym8  2.20	Dau8  2.24	Coi2  2.27	Sym5  2.35	Dau10  2.44	Dau2  2.45
BLS2  2.46	Dau9  2.46	Sym7  2.56	Dau5  2.57	BLS6  2.60	Haar  2.77

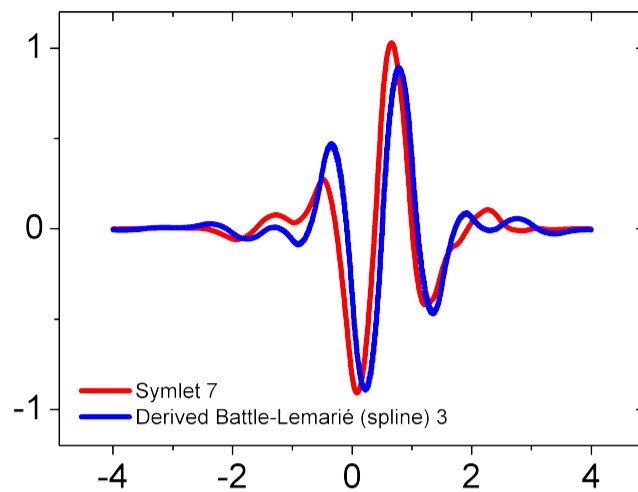
**Table 5.1:** Histograms of child wavelet coefficient  $\alpha_C$  conditioned by the value of its parent coefficient  $\alpha_P$  (in logarithmic scale) and values of  $Q$  for different wavelet bases. They are sorted from closest to farthest from optimality, according to the value of  $Q$ . All histograms range from -6 to 3.5 in both axes. The elbow shape is a direct consequence of loss of optimality, so its position depends on the degree of optimality (for Battle-Lemarié 3, the best wavelet, it is located the leftmost, at  $-3$ , about  $\frac{1}{3}$  of the horizontal axis, while for Haar, the worst wavelet, it is the rightmost, at  $-1$ , about  $\frac{1}{2}$  of the horizontal axis).

the validity of the cascade description in terms of optimal wavelets for both series.

### 5.3 Conditioned distribution of quotation values according the cascade model

#### 5.3.1 Settings

We will try now to determine the distribution of the series value  $y$  conditioned to the knowledge of the vector  $\vec{x}$  of  $N$  previous events,



**Figure 5.2:** Comparison of the optimal wavelet obtained for volatility series (Symlet of order 7) and the time derivative of optimal wavelet obtained for the logarithm of the quotation price (Battle-Lemarié of order 3).

which we will denote by  $\rho_y(y|\vec{x})$ . Due to the high dimensionality of this function, it cannot be estimated in general and it rather needs to be modelled. We will make use of the cascade properties that we have introduced in the previous sections to model this conditioned PDF. First, we assume that our wavelet basis is optimal, so:

$$\alpha_{j,k} = \eta_{j,k} \alpha_{j-1, \lfloor \frac{k}{2} \rfloor} \quad (5.7)$$

where the variables  $\eta_{j,k}$  are all identically distributed according to a known PDF  $\rho_\eta$  and each variable  $\eta_{j,k}$  is independent of  $\alpha_{j-1, \lfloor \frac{k}{2} \rfloor}$ . As  $\alpha_{j-1, \lfloor \frac{k}{2} \rfloor}$  can be further decomposed as  $\eta_{j-1, \lfloor \frac{k}{4} \rfloor} \alpha_{j-2, \lfloor \frac{k}{4} \rfloor}$ , we conclude that  $\eta_{j,k}$  is independent of its grand-parent  $\eta_{j-2, \lfloor \frac{k}{4} \rfloor}$  and by induction, it is independent of all its ancestors. Notice however that this does not imply that  $\eta_{j,k}$  is independent of  $\eta_{j,k'}$ ; the horizontal correlations must be studied and implemented.

Due to the linearity in the definition of the wavelet coefficients,  $\alpha_{j,k}$  can be expressed as a linear function of  $y$ , with coefficients depending on  $\vec{x}$ , namely:

$$\alpha_{j,k} = a_{j,k} y + b_{j,k} \quad (5.8)$$

Hence, the cascade variables are expressed as:

$$\eta_{j,k} = \frac{a_{j,k} y + b_{j,k}}{a_{j-1, \lfloor \frac{k}{2} \rfloor} y + b_{j-1, \lfloor \frac{k}{2} \rfloor}} \quad (5.9)$$

As there is a deterministic relation between each cascade variable and the variable  $y$ , the conditioned distribution of the cascade variables is expressed as a combination of delta functions, namely:

$$\rho(\{\eta_{j,k}\} | y, \vec{x}) = \prod_{j,k} \delta\left(\eta_{j,k} - \frac{a_{j,k} y + b_{j,k}}{a_{j-1, \lfloor \frac{k}{2} \rfloor} y + b_{j-1, \lfloor \frac{k}{2} \rfloor}}\right) \quad (5.10)$$

If we integrate this distribution with  $\rho_y(y|\vec{x})$  we will obtain the distribution of cascade variables conditioned by  $\vec{x}$  only, namely:

$$\rho(\{\eta_{j,k}\} | \vec{x}) = \int dy \rho_y(y|\vec{x}) \rho(\{\eta_{j,k}\} | y, \vec{x}) \quad (5.11)$$

We need to propose a model for  $\rho(\{\eta_{j,k}\} | \vec{x})$  so we can solve for  $\rho_y(y|\vec{x})$ .

### 5.3.2 The model

We propose the following model for  $\rho(\{\eta_{j,k}\} | \vec{x})$ :

$$\rho(\{\eta_{j,k}\}|\vec{x}) = \kappa \left( \prod_{j,k} \rho_{\eta}(\eta_{j,k}) \right) \chi_{\vec{x}}(\{\eta_{j,k}\}) \quad (5.12)$$

where  $\chi_{\vec{x}}(\{\eta_{j,k}\})$  is a set function (it takes only the values 0 and 1) that restricts the values of the variables  $\eta_{j,k}$  to a subset of really accessible values. That is, we assume that the probabilities are independent but not every possible value of  $\eta_{j,k}$  will be visited once  $\vec{x}$  is fixed, which explains the normalization prefactor  $\kappa$ . Nevertheless, we assume that the pyramid is large enough not to modify the shape of the marginal distributions.

#### The solution

Recalling Eq. (5.11) and using Eq. (5.10) we have:

$$\rho(\{\eta_{j,k}\}|\vec{x}) = \int dy \rho_y(y|\vec{x}) \prod_{j,k} \delta\left(\eta_{j,k} - \frac{a_{j,k}y + b_{j,k}}{a_{j-1, \lfloor \frac{k}{2} \rfloor}y + b_{j-1, \lfloor \frac{k}{2} \rfloor}}\right) \quad (5.13)$$

We will next show that the following function:

$$\rho_y(y|\vec{x}) = \kappa \prod_{j,k} \rho_{\eta}\left(\frac{a_{j,k}y + b_{j,k}}{a_{j-1, \lfloor \frac{k}{2} \rfloor}y + b_{j-1, \lfloor \frac{k}{2} \rfloor}}\right) \quad (5.14)$$

verifies the proposed model, Eq. (5.12). The values of  $\eta_{j,k}$  actually visited are those of the shape  $\left\{ \eta_{j,k} = \frac{a_{j,k}y + b_{j,k}}{a_{j-1, \lfloor \frac{k}{2} \rfloor}y + b_{j-1, \lfloor \frac{k}{2} \rfloor}} \right\}_{j,k}$  where  $y$

can vary but must be the same for all  $\eta_{j,k}$  in the same realization of the pyramid. Substituting Eq. (5.14) in Eq. (5.13) we have:

$$\rho(\{\eta_{j,k}\}|\vec{x}) = \kappa \int dy \prod_{j,k} \rho_{\eta}(\eta_{j,k}) \delta\left(\eta_{j,k} - \frac{a_{j,k} y + b_{j,k}}{a_{j-1, \lfloor \frac{k}{2} \rfloor} y + b_{j-1, \lfloor \frac{k}{2} \rfloor}}\right) \quad (5.15)$$

If we assume that the pyramid is large enough so that the distribution of values  $\eta_{j,k}$  across the pyramid is not dependent on the particular value of  $y$  considered, we can take the product of PDFs outside the integral and hence:

$$\rho(\{\eta_{j,k}\}|\vec{x}) = \kappa \prod_{j,k} \rho_{\eta}(\eta_{j,k}) \int dy \prod_{j,k} \delta\left(\eta_{j,k} - \frac{a_{j,k} y + b_{j,k}}{a_{j-1, \lfloor \frac{k}{2} \rfloor} y + b_{j-1, \lfloor \frac{k}{2} \rfloor}}\right) \quad (5.16)$$

Notice that the remaining integral is a set function with support on the values of  $\eta_{j,k}$  which are accessible only. Hence,

$$\chi_{\vec{x}}(\{\eta_{j,k}\}) = \int dy \prod_{j,k} \delta\left(\eta_{j,k} - \frac{a_{j,k} y + b_{j,k}}{a_{j-1, \lfloor \frac{k}{2} \rfloor} y + b_{j-1, \lfloor \frac{k}{2} \rfloor}}\right) \quad (5.17)$$

and so Eq. (5.12) follows. We conclude that Eq. (5.14) is the expression of the distribution of  $y$  conditioned by the past values  $\vec{x}$  according to the cascade model.

### Asymptotic limits and stability of the solution

The stochastic model given by Eq. (5.14) is very appealing, but its application to real data requires some care. First of all, let us realize that the model is based on the assumption that the wavelet used is optimal. This is a bit delicate for the case in which the denominators  $a_{j-1, [\frac{k}{2}]}y + b_{j-1, [\frac{k}{2}]}$  vanish, as for these, as shown by the conditioned histograms in Table 5.2.1, the cascade model is no longer valid. In fact, for those values of  $y$  making the denominator to vanish the numerator should also vanish, what implies  $a_{j, k}/a_{j-1, [\frac{k}{2}]} = b_{j, k}/b_{j-1, [\frac{k}{2}]}$ . Hence, such a constraint should be implemented in an appropriate numerical scheme in order to obtain stable results.

Another interesting limit is when  $|y| \rightarrow \infty$ . In that case, the solution collapses to a fixed value,

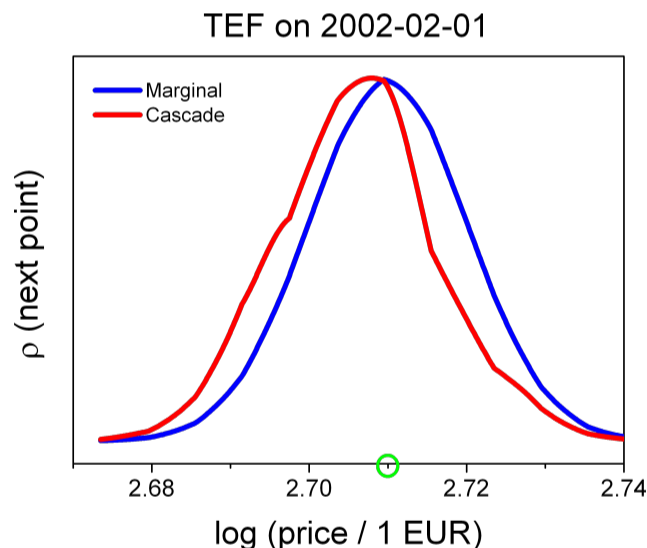
$$\rho_y(y|\vec{x}) \xrightarrow{|y| \rightarrow \infty} \kappa \prod_{j, k} \rho_\eta \left( \frac{a_{j, k}}{a_{j-1, [\frac{k}{2}]}} \right) \quad (5.18)$$

In order to define an integrable distribution,  $\rho_y(y|\vec{x}) \xrightarrow{|y| \rightarrow \infty} 0$ , so this has two consequences. First, there exists a finite maximum value  $\eta_\infty$  for the variable  $\eta$ , so  $\rho_\eta(\eta > \eta_\infty) = 0$ . Second, for at least one  $j, k$ , we must have  $a_{j, k}/a_{j-1, [\frac{k}{2}]} > \eta_\infty$ . The first condition is in fact trivially verified, as discussed in [95, 6]: a physical signal has always this finite maximum. The second property implies that a well-realized cascade must have one of the ratios  $a_{j, k}/a_{j-1, [\frac{k}{2}]}$  large enough. This property can be used as a control check on the validity of the cascade model to a given case. Let us finally remark that if the two conditions are fulfilled hence the range of valid values of  $|y|$  is bounded, so there is a minimum and a maximum possible value of  $y$ .



### 5.3.3 Conditioned distribution and maximum likelihood forecasting

Since  $\rho_y(y|\vec{x})$  takes into account the presence of the cascade, it should be a better predictor than just considering the marginal distribution of returns  $\rho_r(r)$  centred on the last point:  $\rho_{\hat{y}}(\hat{y}) = \rho_r(\hat{y} - x_N)$ .<sup>1</sup> In Fig. 5.3 we show the distributions  $\rho_y(y|\vec{x})$  and  $\rho_{\hat{y}}(\hat{y})$  for a randomly chosen point of the TEF series. We can see that our modelled distribution,  $\rho_y(y|\vec{x})$ , is visibly skewed, meaning that the cascade at this point is making negative returns more likely than positive ones. We also see that this distribution is narrower than the return-derived distribution, i.e., for this point the cascade structure is implying a reduction of volatility. Similar results are seen in other points of all the series.



**Figure 5.3:** Distribution  $\rho_y(y|\vec{x})$  for the TEF series on February 1st 2002 (light/red), compared to the marginal distribution of returns added to the last point  $\rho_{\hat{y}}(\hat{y})$  (dark/blue). The actual logarithm of price that day was 2.71 (circle).

The presented model is not rigid in the sense that it does not forecast an exact future value, but a distribution of possible future values.

<sup>1</sup>With the marginal distribution of returns  $\rho_r(r)$  we can construct a simple predictor of  $y$  from the last price  $x_N$  as follows:  $\hat{y} = \vec{x}_N + r$ , with a random return  $r$  chosen according to  $\rho_r(r)$ .  $x_N$  is the last element of vector  $\vec{x}$  and hence the element just preceding  $y$ . Obviously, the distribution of  $\hat{y}$  is just the distribution  $\rho_r$  shifted in its argument by  $-x_N$ :  $\rho_{\hat{y}}(\hat{y}) = \rho_r(\hat{y} - x_N)$ .

This allows us to forecast not only the most likely future value but also the instant volatility and higher-moment information conditioned by a given past. If we concentrate on the most probable value only, this is a maximum likelihood estimation of the future value. We have calculated these most likely future values sequentially and compared them to the actual observed values, all along the series and for all the series. Since volatility changes over time and also between series, we have calculated the difference between actual and predicted values, divided by the estimated instantaneous one-day volatility. This quotient is around 0.85, similar for all the series (between 0.84 and 0.87). This means that the knowledge furnished by our model allows reducing expected risk in about 15% from an estimation of volatility at the same point.

#### **5.4 A cascade inference for time series: discussion**

In this chapter we have addressed some important problems in the analysis of econometric time series. First, we have shown that stock market time series can be described in terms of cascade processes. We have shown that both log-prices and volatility are multifractal observables following a multiplicative process across scales. This result reveals the existence of a hierarchical structure whose analysis can be relevant not only to improve the statistical characterization of variables but also to forecast them. To complete the analysis, we have used the ratio  $Q$  (Section 3.2), a new criterion to measure the optimality degree of a wavelet in an accurate, robust and little data-demanding way, especially appropriate for real data consisting of limited datasets. With this optimality degree estimator we have found the closest to optimal wavelet among a bank of standard wavelets for

the logarithm of quotation price (Battle-Lemarié 3) and the volatility series (Symlet 7).

We have also faced explicitly the problem of forecasting. In this context, we have provided a theoretical model able to compute the probability distribution of an unknown point of the series conditioned to the knowledge of previous events making use of the cascade properties and knowing the optimal wavelet basis. As far as we know, this is the first theoretical approach in these terms and opens a new door to the problem. We have shown that in practice the distribution evolves and changes width and skewness, i.e., at some points the cascade favours positive returns while in others it favours negative returns, or similarly it implies a volatility increase at some points and decrease in others. Additionally, a simple maximum likelihood estimator shows a discrepancy smaller than the volatility at the same point.



## 6 Application to oceanographic data

In ocean turbulence, as it happens in many scale invariant systems, structures persist across scales. As we have seen in Section 3.1, this is notably seen in wavelet coefficients: they present similar shape through different scales, even though these are orthogonal components. This phenomenon is modelled as a multiplicative cascade, Chapter 3, an effective mechanism that transfers energy, information or an analogous quantity from the largest scales to the smallest scales.

We generically call parent coefficient  $\alpha_P$  the wavelet coefficient at a given scale and child coefficients  $\alpha_C$  the wavelet coefficients at the subsequent scale and corresponding position. Oceanographic data suggest a multiplicative relation between parent and children, with a saturation regime where additive noise becomes dominant. This is called linear model and was first proposed in [83] through chlorophyll maps, an almost passive scalar advected by the flow, see [82]. A characteristic feature of the linear model is the conditioned histogram of  $\log |\alpha_C|$  vs.  $\log |\alpha_P|$ , Figure 3.3, which has been reported in other completely unrelated cascade processes such as that of natural images [58] and stock market series [10], Chapter 5.

## 6.1 Sea Surface Temperature (SST)

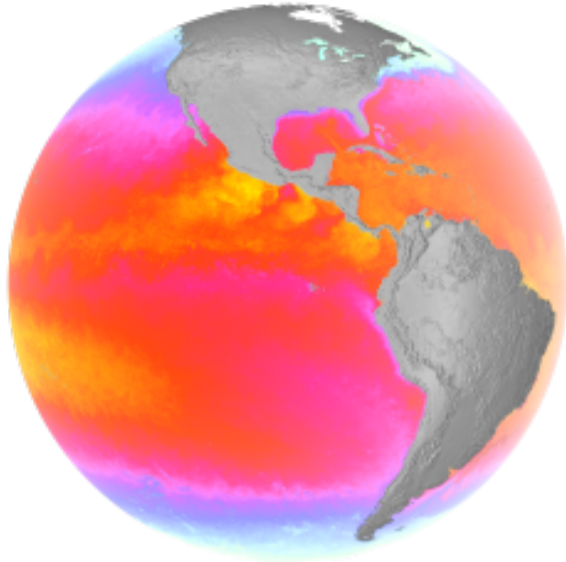
Sea surface temperature is a key variable in ocean circulation and global climate. Atmospheric heating is mostly done from Earth's surface (land and sea) and this is the main cause of wind that drags ocean surface. In addition, SST at the scales of study here can be considered as an almost passive scalar, being advected by the flow, and so tracing the velocity field cascade [121, 51, 122].

In this chapter, we show an analysis of SST daily maps from OSTIA project.<sup>1</sup> Ocean Surface Temperature and Ice Analysis (OSTIA) is a project leaded by the British Met Office and it provides global high resolution SST maps. These data is produced by combining global data of many different satellites (infrared and microwave radiometers) together with *in situ* observations. Combination is done with optimal interpolation model [123]. Data have a nominal resolution of  $1/20^\circ$ , although we subsample them to  $1/10^\circ$  as they show significant smoothing artefacts at  $1/20^\circ$ . The analyzed global daily maps range from 2006-04-01 to 2008-06-18.

## 6.2 Wavelets in 2D signals

Two-dimensional functions, i.e., functions that are defined in a  $\mathbb{C}^2$  (or  $\mathbb{R}^2$ ) domain, can also have a wavelet representation. The main difference from the one-dimensional case (Section 2.2) is that now each dyadic scale step reduces the degrees of freedom by  $\frac{1}{4}$  (instead of  $\frac{1}{2}$ ). As a consequence, for any scale, every detail space has three times the amount of components of the approximation space at the same scale. In the most general case, there is no simple way to represent the bases of the detail spaces (at most we can derive the approximation bases

<sup>1</sup>[http://ghrsst-pp.metoffice.com/pages/latest\\_analysis/ostia.html](http://ghrsst-pp.metoffice.com/pages/latest_analysis/ostia.html)



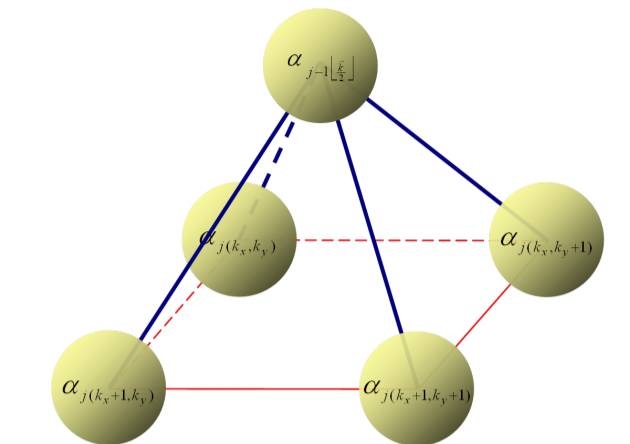
**Figure 6.1:** Image shows Sea Surface Temperature (SST) satellite map projected on the sphere. SST image was produced from MODIS (Moderate Resolution Imaging Spectroradiometer) aboard NASA's Earth Observing System (EOS) Terra satellite on January 18th 2001.

from a father wavelet  $\phi$ ). A more restricted case is the *separable* case, in which there are three different mother wavelets  $\psi^{\{1,2,3\}}$  that spawn three orthogonal detail spaces (thus separable). We have a particular case of this when a 1D wavelet is ported to 2D: any 1D wavelet basis has its natural extension to 2D (and  $ND$  in general).<sup>2</sup> In this extension, the one-dimensional mother wavelet  $\psi$  is represented in three different orientations:  $\psi^{\text{hor}}$ ,  $\psi^{\text{ver}}$ ,  $\psi^{\text{diag}}$  able to represent any 2D signal. An example of 2D wavelet representation has been shown in Figure 3.1.

### 6.3 Optimal wavelet for SST

As seen in previous chapters, the microcanonical cascade filters redundancy and allows inter-scale inference. Not all wavelets are equivalent,

<sup>2</sup>The converse is not true: separable 2D wavelets are more general and have cases where there is no 1D equivalent.



**Figure 6.2:** Schematic representation of the microcanonical cascade in the 2D case. For each parent coefficient  $\alpha_P$  there are four child coefficients  $\alpha_C$ .

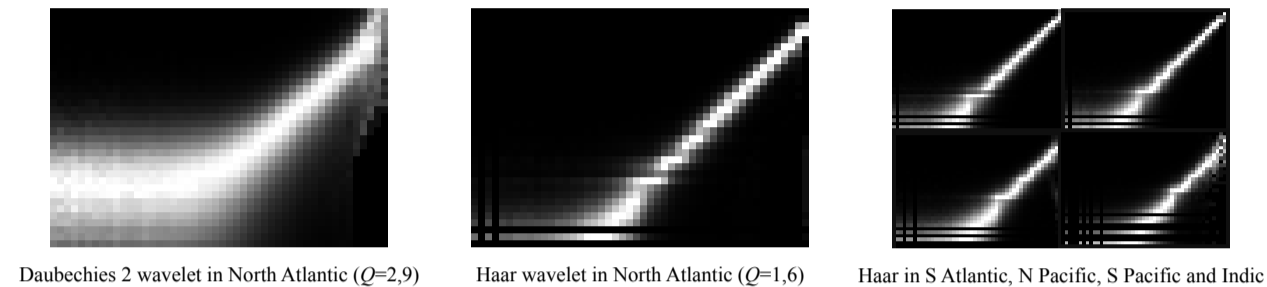
but we need the optimal wavelet of the system to avoid noise-like saturation.

Considering the linear model Eq. (3.1) the most desirable case would be that with no saturation (no  $\alpha_*$ ), as then the child coefficients could be easily inferred from their parent with minimal dispersion (that of the randomness of the multiplicative variable  $\eta$ , which in fact is small). In previous chapters, we have analytically shown that if the generating process is purely multiplicative, but wavelet coefficients are obtained with a different wavelet than that of the generation, the result is the linear model. That generating wavelet is called optimal wavelet. As already seen, parameter  $Q$  is a stable characterizer of optimality. It is  $Q \geq 1$  always, and  $Q = 1$  for the optimal wavelet only.

We have calculated the values of  $Q$  in OSTIA SST maps for a set of 24 standard wavelets – the same as those listed in Page 77 and used in Section 3.3 and Section 5.2.1. Among these wavelets, Haar wavelet is found to be the closest to optimal one (see Figure 6.3). This is seen both visually and in terms of  $Q$ . The same result is observed for the different oceanic regions. As Haar is a very compact wavelet, it represents the temperature gradient better than other wavelets. Tem-



perature gradients are highly relevant in both ocean and atmosphere dynamics due to the prevalence of thermal wind as the dominant geostrophic component. [124, 125, 84]

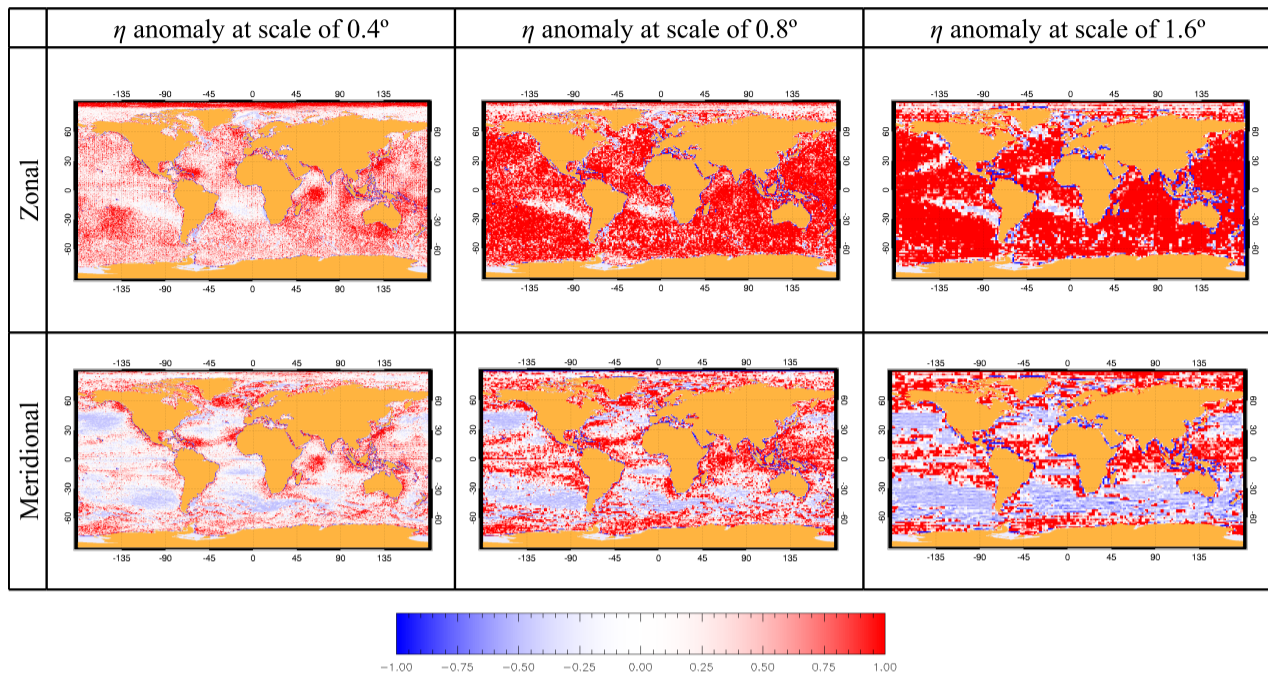


**Figure 6.3:** Joint histograms of child wavelet coefficients conditioned to a value for the parent wavelet coefficient. For Haar wavelet, the histogram is almost linear and saturation regime affects only the small leftmost area. It is the closest to optimal wavelet from the standard wavelets checked (these are the same as in Chapter 5). The same results are obtained in all the oceanic regions.

## 6.4 Local anomalies

Wavelet coefficients decompose the 2D signal into three components: zonal, meridional and a cross contribution that corresponds to the diagonal compensation [11]. As structures in the zonal direction dominate the other directions (correlation length is maximum along the zonal direction), the cross contribution component mostly follows the zonal component (that is why it is not shown). Anomalies of the cascade variable  $\eta$  at different scales are shown. Anomalies are calculated as the relative difference between the observed value and the theoretical average value for an ideal cascade process, which is  $\frac{1}{2}$ . See Figure 6.4.

We observe that at roughly the mesoscale, heat transfer between scales as measured by  $\eta$  is rather homogeneously distributed for the zonal components, with slightly positive anomalies all over the ocean and slight negative anomalies along tropical and equatorial wind-



**Figure 6.4:** Global maps of the local anomalies of cascade values of ocean turbulence, derived from SST monthly averages. Depending on the coefficients compared, we have zonal and meridional components.

driven currents. On the contrary, meridional components have strong positive anomalies distributed on the limits of tropical and subtropical gyres, and very weak negative anomalies over the rest. We conclude that heat transfer between scales is rather inhomogeneously distributed on the ocean, with a preference for the mesoscale and some greater scales, and centred around significant boundary currents.

In temperature cascades, the multiplicative variable can be regarded as heat transfer between scales. In conclusion, analysis of Sea Surface Temperature maps shows that local anomalies of the cascade give information about energy exchange across scales in oceanic gyres, upwelling zones, tradewind driven currents and other currents, something that could be used in their characterization.

## 7 Conclusions

*“I’m astounded by people who want to know the universe when it’s hard enough to find your way around Chinatown”*

— Woody Allen

Complex systems are abundant in our natural environment. In linear systems, the equations of their dynamics can be very difficult to solve, but if they cannot be described with a single characteristic scale, at least they can be described by a set of few characteristic scales that are totally decoupled from each other. However, this takes on a completely different flavour in non-linear systems, where scales are coupled and appropriate multiscale analysis is in order. This is the case of complex systems and, more particularly, scale invariant systems. In these, the approach to their solution is different, and it usually involves a multiscale basis. In this context, wavelets are one of the most used representation paradigms.

The research context of complex systems and, particularly, scale invariant systems and multifractals has been in constant evolution over the last few years. Theoretical advances, either statistical (stochastic processes and probability distributions) or geometrical (function analysis and measure theory), along with fancy signal-processing algorithms suited to scale invariant data (and additionally handling aliasing, discretization and other artefacts of experimental data), have

originated new tools for multifractal characterization of systems. While ten years ago the only methods available were statistical (ranging from rough numerical calculation of structure functions to more elaborate methods like WTMM [37]), by the start of this thesis project, development of geometrical methods had begun (most notably, the microcanonical multifractal formalism (**MMF**) [5]). Geometrical methods have a clear advantage over statistical methods: they characterize each point of the system and thus they permit new applications such as reconstruction and prediction of signals [92, 95, 6], i.e., not only statistical characterization. Additionally, geometrical methods provide statistical characterization with much less need of data than statistical methods [118].

In the present thesis, we have given solutions to several problems related to signal analysis of experimental data, under the framework of MMF. In this way, we have worked on the generalization and improvement of MMF, as well as its applications to the inference and forecasting of systems that follow a cascade process. In particular, we have described applications to two very different systems: stock-market series and ocean turbulence. The representation of the signal as a microcanonical cascade plays a crucial role in these applications. This representation can be achieved with one particular wavelet called optimal wavelet. The most relevant theoretical achievements are the regularization of diverging multifractal measures [7], the establishment of the bridge between multiplicative variables in microcanonical cascade processes and local singularity exponents [8], and the design of accurate and robust measure of wavelet optimality for a given dataset [8].

Regarding the developed applications, on stock-market time series, we have inferred the distribution of future returns conditioned by the

cascade and we have shown that a prediction based on this inference improves that of an ARIMA model.<sup>1</sup> From the distribution of future returns, future volatility and value-at-risk can be reliably forecasted [10, 12]. On ocean data we have characterized dynamical aspects from optimal wavelet cascade analysis. In particular, we have observed that anomalies in the cascade of sea surface temperature show particular points of heat transfer between structures at different scales in the zones of wind-driven currents, also in the gyres [10].

To achieve this, we have introduced a new formalism, that of microcanonical cascades, that marries the cascade formalisms with the microcanonical multifractal formalism. The theoretical and computational results can be summarised as this cascade–singularity connection (Section 4.2), extension of multifractal measures to functions of non-total support or with diverging measures (Section 4.3) and a robust criterion to obtain optimal wavelets (Section 3.2). Application to stock-market series consisted in forecasting of future points distribution, while application to ocean turbulence consisted in characterization of ocean surface dynamics from SST maps and characterisation of inter-scale energy exchanges. We have seen that this analysis identifies tradewind-driven currents and frontogenesis points.

Both understanding – combined with appropriate modelling – of dynamics and design of inference/forecasting algorithms have crucial importance for the anticipation of changes in natural phenomena. In this context, the chain formed by the three steps followed during the thesis, namely multifractal characterization first, then obtaining of the optimal wavelet and finally design of inference algorithms, summarizes the direction we have followed to tackle the study of econometric time series and ocean maps.

---

<sup>1</sup>ARIMA: autoregressive integrated moving average. More precisely, in that approach it was a neural-network-weighted, seasonal ARIMA, which is an enhanced version of the simplest ARIMA.

## 7.1 Future work

In research studies, answering questions often raises new ones. All the milestones achieved in this study point to a line for forthcoming research. This line encompasses both theoretical and empirical approaches. In the most theoretical area there is the need to determine whether all multifractals (*sensu* MMF) are describable as microcanonical cascades or, more precisely, whether they always have an optimal wavelet. While some artificial multifractal models cannot have a complete cascade representation, it seems that natural systems, which are compliant with Parisi-Frisch's statistical-geometrical duality [23], do have this property. Another theoretical proof that is lacking is one that would explain the formation of the same singularity spectrum attractor in so many different signals (an observation reported in [9]), which makes them reconstructible from a small, most singular fractal component.

In the future lines of work there is significantly more than new theorems to prove. There are also many promising methodological developments that we have already improved and that we expect to further improve in future research. The most important development is continuous wavelet optimization. We have defined a discriminating, robust measure of optimality (the ratio  $Q$  defined in Eq. (3.42)) and we know the degrees of freedom that define a wavelet (in the form of QMF or biorthogonal QMF). Implementation of these degrees of freedom in a continuous optimization algorithm for  $Q$  leads to non-trivial, nonlinear constraints that hinder the optimization. We have already attempted this optimization with advanced algorithms and we are seeking solutions to these issues.

Finally, the most promising developments that we foresee are new applications. The signal processing techniques presented are aimed towards advanced characterization of signal properties. While they also provide nice descriptive and modelling frameworks, the application of these techniques to actual problems in empirical data is what makes them so powerful. In this sense, we have shown that the microcanonical cascade formalism allows forecasting time series. If we introduce the effects of horizontal correlations then the forecasting horizon could be increased. Furthermore, by introducing the presence of multifractal sources discussed in [96], long-term dynamical changes could also be included.

Applications to geophysical turbulence are even more promising. We have seen that a microcanonical analysis of the temperature cascade gives information about heat transfers at the ocean surface, but this characterization goes further, since temperature is a scalar mostly advected by the flow in such a way that its multiscale analysis gives direct information of the velocity cascade. To demonstrate this, we are now studying chlorophyll maps to find a coincidence with temperature similar to that found in [82]. We are also analyzing the oceanic flow from the 2+1 dimensional perspective (time evolution of temperature maps) to improve its dynamic characterization. Indeed, multifractal analysis of sea surface temperature has been used to provide information about the ocean velocity field [84] and analysis of the chlorophyll concentration cascade has been used to reconstruct its acquisition gaps [83]. These characterizations are likely to be improved by the use of optimal wavelets.

There are dozens of other applications to study. The reconstructibility of multifractal structure and the minimal redundancy representation that the optimal wavelet achieves lead naturally to the field

of signal compression. The design of recent compression algorithms for multifractal signals such as speech and natural images is already exploiting their multiscaling properties. A microcanonical cascade representation is expected to improve these algorithms.



# Appendix



## Appendix A

### QMF construction, constraints and biorthogonal generalization

The value of  $\int \phi^*(t) dt$  is fixed due to orthonormality conditions, since we require that  $\phi \in \mathbb{R}$ . In the following, we will see that orthonormality, Eq. (2.12), imposes  $|\int \phi(t) dt| = 1$ . Let us prove it in the following lines: first of all, let us define the Fourier transforms:

$$\hat{\phi}(\nu) = \int_{-\infty}^{\infty} \phi(t) e^{-i2\pi\nu t} dt \quad (\text{A.1})$$

and<sup>1</sup>

$$\hat{h}_\nu = \sum_l h_l e^{-i2\pi\nu l} \quad (\text{A.2})$$

Also note that the Fourier transform has the well-known properties:

$$\mathcal{F}[s(at)] = \frac{1}{|a|} \hat{s}\left(\frac{\nu}{a}\right) \quad (\text{A.3})$$

$$\mathcal{F}[s(t - k)] = \hat{s}(\nu) e^{-i2\pi\nu k} \quad (\text{A.4})$$

---

<sup>1</sup> Sometimes, the discrete Fourier transform is defined also discrete, with  $\nu = \frac{m}{N}$  where  $m$  is the discrete frequency and  $N$  is the size of the series. However, in this case  $N$  can be arbitrarily large, ideally  $\infty$ , something that converts  $\nu$  to a continuous variable ranging  $[0, 1)$ .

Therefore, Eq. (2.20),  $\phi(t) = \sum_l h_l \sqrt{2} \phi(2t - l)$ , becomes in Fourier space:

$$\hat{\phi}(\nu) = \sum_l h_l \sqrt{2} \mathcal{F}[\phi(2t - l)] \quad (\text{A.5})$$

i.e.,

$$\begin{aligned} \hat{\phi}(\nu) &= \sum_l h_l \sqrt{2} \frac{1}{2} \hat{\phi}\left(\frac{\nu}{2}\right) e^{-i\pi\nu l} \\ &= \frac{1}{\sqrt{2}} \hat{\phi}\left(\frac{\nu}{2}\right) \hat{h}_{\frac{\nu}{2}} \end{aligned} \quad (\text{A.6})$$

An iteration of this formula allows building the scaling function from the  $h$  filter:  $\hat{\phi}(\nu) = \hat{\phi}(0) \prod_{j=1}^{\infty} \left(2^{-\frac{1}{2}} \hat{h}_{\frac{\nu}{2^j}}\right)$  and an equivalent expression with  $\hat{g}_{\frac{\nu}{2}}$  as the first factor builds the wavelet.

The next step consists in calculating the autocorrelation of the scaling function. Directly from its definition,  $a(\tau) = \int \phi^*(t) \phi(t + \tau) dt$ , it follows that its Fourier transform is:

$$\begin{aligned} \hat{a}(\nu) &= \hat{\phi}^*(\nu) \hat{\phi}(\nu) \\ &= \left| \hat{\phi}(\nu) \right|^2 \end{aligned} \quad (\text{A.7})$$

The autocorrelation  $a(\tau)$  coincides with  $\langle \phi_{0,0} | \phi_{0,\tau} \rangle$  when  $\tau \in \mathbb{Z}$ . That is why we define a discrete autocorrelation  $b(\tau)$  that equals  $a(\tau)$  when  $\tau$  is integer and zeros otherwise:

$$b(\tau) = \begin{cases} a(\tau) & \tau \in \mathbb{Z} \\ 0 & \tau \notin \mathbb{Z} \end{cases} \quad (\text{A.8})$$

or, in a more formal way<sup>2</sup>,

$$b(\tau) = \sum_n a(\tau) \delta(\tau - n) \quad (\text{A.9})$$

and now we use a well-known property of the Fourier transform of a Dirac's Comb:

$$\mathcal{F} \left[ \sum_n \delta(t - nT) \right] = \frac{1}{T} \sum_n \delta\left(\nu - \frac{n}{T}\right) \quad (\text{A.10})$$

to obtain the Fourier transform of  $b(\tau)$ , i.e.,

$$\begin{aligned} \hat{b}(\nu) &= \hat{a}(\nu) * \sum_n \delta(\nu - n) \\ &= \sum_n \int_{-\infty}^{\infty} \hat{a}(\nu') \delta(\nu - n - \nu') d\nu' \\ &= \sum_n \hat{a}(\nu - n) \end{aligned} \quad (\text{A.11})$$

But, on the other hand,  $a(\tau \in \mathbb{Z}) = \langle \phi_{0,0} | \phi_{0,\tau} \rangle = \delta_{\text{Kron.}}(\tau)$ , which implies that  $b(\tau) = \delta(\tau)$ , and so  $\hat{b}(\nu) = 1$ . Therefore, combining Eq. (A.7) and Eq. (A.11),

$$1 = \sum_{n=-\infty}^{\infty} \left| \hat{\phi}(\nu - n) \right|^2 \quad \forall \nu \quad (\text{A.12})$$

This relation is very general. In fact, it also applies to complex scaling functions and has a direct equivalent in the biorthogonal case (where  $a(\tau)$  is then defined as a cross-correlation  $a(\tau) = (\overline{\phi} \star \phi)(\tau) = \int \overline{\phi}^*(t) \phi(t+\tau) dt$  and all the preceding equations hold) with  $\hat{\phi}^*(\nu) \hat{\phi}(\nu)$  instead of  $\left| \hat{\phi}(\nu) \right|^2$ . In addition to the scaling function, the same ex-

<sup>2</sup> Strictly speaking, both definitions do not coincide. All the deltas appeared until this point are properly Kronecker tensors more than Dirac's Deltas. However, the properties of the continuous Fourier transform to be used in the following lines require Dirac's Deltas.

pression is valid for the wavelet and, in the derivation to follow, for the  $g$  filter.

Eq. (A.12) is valid for any  $\nu$ , in particular for  $\nu = 0$ . Combining Eq. (A.6) with Eq. (A.12),

$$\begin{aligned} 1 &= \sum_n \left| \hat{\phi}(n) \right|^2 \\ &= \sum_n \frac{1}{2} \left| \hat{\phi}\left(\frac{n}{2}\right) \right|^2 \left| \hat{h}_{\frac{n}{2}} \right|^2 \end{aligned} \quad (\text{A.13})$$

Notice that  $\hat{h}_\nu$ , as defined in Eq. (A.2), is the continuous Fourier transform of a discrete series, something meaning that it is  $2\pi$  periodic, i.e.,  $\hat{h}_{\nu+1} = \hat{h}_\nu$ . Therefore, in the expression above only two values of  $\hat{h}_{\frac{n}{2}}$  are possible, depending on the parity of  $n$ :  $\hat{h}_0$  for even  $n$  or  $\hat{h}_{\frac{1}{2}}$  for odd  $n$ . We split the sum in even ( $n = 2n'$ ) and odd ( $n = 2n' + 1$ ) terms. As  $n, n'$  ranges all  $\mathbb{Z}$  from  $-\infty$  to  $\infty$  in both sums.

$$1 = \frac{1}{2} \left| \hat{h}_0 \right|^2 \underbrace{\sum_{\substack{\text{even } n \\ n' = \frac{n}{2}}} \left| \hat{\phi}(n') \right|^2}_1 + \frac{1}{2} \left| \hat{h}_{\frac{1}{2}} \right|^2 \underbrace{\sum_{\substack{\text{odd } n \\ n' = \frac{n-1}{2}}} \left| \hat{\phi}\left(n' + \frac{1}{2}\right) \right|^2}_1 \quad (\text{A.14})$$

where we use Eq. (A.12) to simplify the sums to 1. Now, if we invert Eq. (A.6) at  $\nu = 0$  also requiring  $\hat{\phi}(0) \neq 0$ , we obtain that:

$$\hat{\phi}(0) = \frac{1}{\sqrt{2}} \hat{\phi}(0) \hat{h}_0 \quad \Rightarrow \quad \hat{h}_0 = \sqrt{2} \quad (\text{A.15})$$

i.e., the value of  $\hat{h}_0$  is fixed, and so they are the other integer and half-integer values: due to Eq. (A.14),  $\hat{h}_{\frac{1}{2}} = 0$ . Therefore, we can go

back to Eq. (A.6) knowing now that:

$$\hat{\phi}(1) = \frac{1}{\sqrt{2}} \hat{\phi}\left(\frac{1}{2}\right) \underbrace{\hat{h}_{\frac{1}{2}}}_0 = 0 \quad (\text{A.16})$$

$$\hat{\phi}(2) = \frac{1}{\sqrt{2}} \underbrace{\hat{\phi}(1)}_0 \hat{h}_1 = 0 \quad (\text{A.17})$$

$$\hat{\phi}(3) = \frac{1}{\sqrt{2}} \hat{\phi}\left(\frac{3}{2}\right) \underbrace{\hat{h}_{\frac{3}{2}}}_0 = 0 \quad (\text{A.18})$$

$$\hat{\phi}(4) = \frac{1}{\sqrt{2}} \underbrace{\hat{\phi}(2)}_0 \hat{h}_2 = 0 \quad (\text{A.19})$$

...

Only  $\hat{\phi}(0)$  remains. Therefore, Eq. (A.12) leads to:

$$1 = \left| \hat{\phi}(0) \right| \quad (\text{A.20})$$

Since the wavelet coefficients have a physical meaning in a multifractal signal, it seems reasonable to require that  $\int \phi(t) dt \in \mathbb{R}$ . If not, we could still have signals with only real wavelet coefficients, but that would not be the case for every signal. Possibly this assumption is even more restrictive and requires that the wavelet and the scaling function are entirely real-valued. However, this is what we have already assumed in the beginning of the section.

In the biorthogonal case,  $1 = \hat{\phi}^*(0) \hat{\phi}(0)$ . A totally equivalent derivation can be done for the wavelet. The only difference is that, due to  $\int \psi(t) dt = 0^3$ ,  $\hat{g}_{\text{even}} = 0$  and  $|\hat{g}_{\text{odd}}| = \sqrt{2}$ , and further derivations lead to a general relation between  $h$  and  $g$  for cases that are not linear in phase.

---

<sup>3</sup> In fact, this is a requirement for the wavelet to be admissible (so that the wavelet series converges), but can also be obtained from the fact that, as  $\hat{\phi}(0) \neq 0$ ,  $(\phi \star \psi)(\tau \in \mathbb{Z}) = 0 \Rightarrow \sum_n \hat{\phi}^*(\nu + n) \hat{\psi}(\nu + n) = 0 \Rightarrow \hat{\psi}(0) = 0$ .

## A.1 Biorthogonal generalization

A relaxation of the phase linearity or the orthonormality implies a double optimization with non-trivial constraints. Let us regard a generalization to a biorthogonal case. First, we recall Eq. (A.14) but in its biorthogonal shape and for any  $\nu$ :

$$1 = \frac{1}{2} \hat{h}_\nu^* \hat{h}_\nu + \frac{1}{2} \hat{h}_{\nu+\frac{1}{2}}^* \hat{h}_{\nu+\frac{1}{2}} \quad (\text{A.21})$$

Its obtaining comes from Eq. (A.6) and its dual  $\hat{\phi}(\nu) = \frac{1}{\sqrt{2}} \hat{\phi}(\frac{\nu}{2}) \hat{h}_{\frac{\nu}{2}}^*$ , following the same expansion shown before, and the simplification  $\nu = 0$  in Eq. (A.13) is not done, being instead:

$$1 \stackrel{\text{Eq. (A.12)}}{=} \sum_n \hat{\phi}(\nu' - n) \hat{\phi}(\nu' - n) \quad \forall \nu' \quad (\text{A.22})$$

$$= \sum_{\frac{\nu'}{2}=\nu} \sum_n \frac{1}{2} \hat{\phi}(\nu - \frac{n}{2}) \hat{\phi}(\nu - \frac{n}{2}) \hat{h}_{\nu-\frac{n}{2}}^* \hat{h}_{\nu-\frac{n}{2}} \quad (\text{A.23})$$

For  $\nu = 0$ , Eq. (A.6) and its dual lead to  $\hat{h}_0^* = \hat{h}_0 = \sqrt{2}$  and, due to Eq. (A.21),  $\hat{h}_{\frac{1}{2}}^* = \hat{h}_{\frac{1}{2}} = 0$ . There is another constraint for  $\nu = \frac{1}{4}$ , provided that  $h$  and  $\bar{h}$  are real, i.e.,  $\hat{h}_{\nu+\frac{1}{2}} = \hat{h}_{\frac{1}{2}-\nu}^*$  and so Eq. (A.21) becomes:

$$\begin{aligned} 1 &= \Re \left[ \hat{h}_{\frac{1}{4}}^* \hat{h}_{\frac{1}{4}} \right] \\ &= \Re \left[ \hat{h}_{\frac{1}{4}} \right] \Re \left[ \hat{h}_{\frac{1}{4}} \right] + \Im \left[ \hat{h}_{\frac{1}{4}} \right] \Im \left[ \hat{h}_{\frac{1}{4}} \right] \end{aligned} \quad (\text{A.24})$$

or  $1 = \hat{h}_{\frac{1}{4}} \hat{h}_{\frac{1}{4}}$  if  $h$  is symmetrical (and so  $\hat{h}$  is real). In analogy,

$$\begin{aligned} 1 &= \Re \left[ \hat{h}_{\frac{3}{4}} \hat{h}_{\frac{3}{4}}^* \right] \\ &= \Re \left[ \hat{h}_{\frac{3}{4}} \right] \Re \left[ \hat{h}_{\frac{3}{4}} \right] + \Im \left[ \hat{h}_{\frac{3}{4}} \right] \Im \left[ \hat{h}_{\frac{3}{4}} \right] \end{aligned} \quad (\text{A.25})$$



or  $1 = \hat{h}_{\frac{3}{4}} \hat{h}_{\frac{3}{4}}$  if  $h$  is symmetrical.

### Phase linearity

The phase linearity condition is then obtained from the fact that:

$$1 = \frac{1}{2} \hat{g}_{\nu}^* \hat{g}_{\nu} + \frac{1}{2} \hat{g}_{\nu+\frac{1}{2}}^* \hat{g}_{\nu+\frac{1}{2}} \quad (\text{A.26})$$

which gives  $\hat{g}_0^* = \hat{g}_0 = 0$  and  $\hat{g}_{\frac{1}{2}}^* \hat{g}_{\frac{1}{2}} = 2$ , and the dual orthogonalities:

$$0 = \sum_n \hat{\psi}^*(\nu - n) \hat{\phi}(\nu - n) \quad (\text{A.27})$$

$$0 = \sum_n \hat{\phi}^*(\nu - n) \hat{\psi}(\nu - n) \quad (\text{A.28})$$

i.e.,

$$0 = \frac{1}{2} \hat{g}_{\nu}^* \hat{h}_{\nu} + \frac{1}{2} \hat{g}_{\nu+\frac{1}{2}}^* \hat{h}_{\nu+\frac{1}{2}} \quad (\text{A.29})$$

$$0 = \frac{1}{2} \hat{h}_{\nu}^* \hat{g}_{\nu} + \frac{1}{2} \hat{h}_{\nu+\frac{1}{2}}^* \hat{g}_{\nu+\frac{1}{2}} \quad (\text{A.30})$$

or multiplying Eq. (A.29) by Eq. (A.30):

$$\hat{g}_{\nu}^* \hat{g}_{\nu} \hat{h}_{\nu}^* \hat{h}_{\nu} = \hat{g}_{\nu+\frac{1}{2}}^* \hat{g}_{\nu+\frac{1}{2}} \hat{h}_{\nu+\frac{1}{2}}^* \hat{h}_{\nu+\frac{1}{2}} \quad (\text{A.31})$$

now combining this with Eq. (A.21) and Eq. (A.26) results in

$$1 = \frac{1}{2} \hat{g}_{\nu}^* \hat{g}_{\nu} + \frac{1}{2} \hat{h}_{\nu}^* \hat{h}_{\nu} \quad (\text{A.32})$$

and here we can impose the biorthogonal phase linearity<sup>4</sup>:

$$\bar{g}_l = (-1)^l h_{1-l}^* \quad (\text{A.33})$$

$$g_l = (-1)^l \bar{h}_{1-l}^* \quad (\text{A.34})$$

## A.2 A possible wavelet dual

As seen, a QMF filter does not univocally define its dual orthogonal filter<sup>5</sup>. However, given a filter, there is a usual strategy to build one of their possible dual filters [126]. To do this, notice that Eq. (A.12) requires orthonormality, but in a biorthogonal case we will have:

$$q(\nu) = \sum_n \left| \hat{\phi}(\nu - n) \right|^2 \quad (\text{A.35})$$

which is the same as:

$$1 = \sum_n \frac{\left| \hat{\phi}(\nu - n) \right|^2}{q(\nu - n)} \quad (\text{A.36})$$

<sup>4</sup>In the general case, Eq. (A.29) and Eq. (A.30) respectively imply that we can express:

$$\hat{g}_{\nu+\frac{1}{2}}^* = -\frac{\hat{g}_\nu^*}{\underbrace{\hat{h}_{\nu+\frac{1}{2}}}_{\bar{\lambda}(\nu)}} \hat{h}_\nu \quad \hat{g}_{\nu+\frac{1}{2}} = -\frac{\hat{g}_\nu}{\underbrace{\hat{h}_{\nu+\frac{1}{2}}^*}_{\lambda(\nu)}} \hat{h}_\nu^*$$

where  $\lambda$  and  $\bar{\lambda}$  are  $2\pi$  periodic,  $\pi$  antiperiodic, i.e.,  $\lambda(\nu + \frac{1}{2}) = -\lambda(\nu)$  (due to Eq. (A.29) and Eq. (A.30)) and of unitary modulus (due to Eq. (A.32), except for  $\hat{\phi}(0) = 0 \Rightarrow \hat{h}_0 = 0 \Rightarrow \lambda = 0$ ):  $\lambda(\nu) = e^{-i2\pi\Phi(\nu)}$ . Then, the restriction to  $\Phi$  (nothing to do with our scaling function  $\phi$ ) is:  $\Phi(\nu + \frac{1}{2}) = \Phi(\nu) + \frac{1}{2}$ . Phase linearity is obtained imposing  $\bar{\Phi}(\nu) = \Phi(\nu) = \nu$ .

<sup>5</sup> However, due to Eq. (2.25),  $\sum_l \bar{h}_l h_{l+2n} = \delta(n)$ . Then let  $H(l) = h_l \sum_p \delta(l - p)$  and let  $a(n) = \int \bar{H}^*(l + n) H(l) dl \sum_p \delta(n - 2p) = \delta(0)$ , in Fourier space we have that:

$$1 = \hat{a}(\nu) = \underbrace{\hat{H}^*(\nu) \hat{H}(\nu)}_{c(\nu)} * \left( \frac{1}{2} \sum_p \delta\left(\nu - \frac{p}{2}\right) \right)$$

i.e.,

$$2 = \sum_p \int c(\nu - \nu') \delta\left(\nu' - \frac{p}{2}\right) d\nu' = \sum_p c\left(\nu - \frac{p}{2}\right) = \sum_p \hat{H}^*\left(\nu - \frac{p}{2}\right) \hat{H}\left(\nu - \frac{p}{2}\right)$$

where we used that, by definition in Eq. (A.35),  $q(\nu)$  is  $2\pi$  periodic. Then a scaling function that satisfies Eq. (A.22) is:

$$\hat{\phi}(\nu) = \frac{\hat{\phi}(\nu)}{q^*(\nu)} \quad (\text{A.37})$$

and since  $q^* = q$ ,

$$\begin{aligned} \hat{h}_\nu &= \hat{h}_\nu \frac{q(\nu)}{q(2\nu)} \\ &= \hat{h}_\nu \frac{\sum_n \prod_j \frac{1}{2} \left| \hat{h}_{\frac{\nu+n}{2^j}} \right|^2}{\sum_n \prod_j \frac{1}{2} \left| \hat{h}_{\frac{2\nu+n}{2^j}} \right|^2} \end{aligned} \quad (\text{A.38})$$

**Numerical implementation** Numerical difficulties can easily arise in a numerical calculation of Eq. (A.38). In fact, the dummy index  $n$  ranges from  $-\infty$  to  $\infty$ , while  $j$  goes from 1 to  $\infty$ . Despite so many terms contributing, high  $|n|$  terms vanish quickly due to the influence of factors near  $\hat{h}_{\frac{1}{2}} = 0$ . A rough analysis of

$$\sum_n \prod_j \frac{1}{2} \left| \hat{h}_{\frac{\nu-n}{2^j}} \right|^2$$

shows that the  $n = 0$  term is mainly influenced by the first ( $j = 1$ ) factor, the others quickly approaching 1 (as  $\nu' =: \frac{\nu-n}{2^j} \rightarrow 0$ ). The other terms are also highly influenced by their first factors (roughly, until  $|\nu'| < \frac{1}{2}$ ). In particular, terms where  $|\nu - n|$  is very high have many factors far from this regime and so many chances that some  $\nu'$  (modulo 1) is close to  $\frac{1}{2}$ .

An interesting strategy (to be tested) may consist in imposing a maximum value for  $|n|$ . This limit and the discretisation limit of  $\hat{h}_\nu$

(limit of the  $h_k$  series,  $K$ ) naturally impose a limit to  $j$ :

$$J_{\max} = \lceil \log_2 (K(n_{\max} + 1) - 1) \rceil + 1 \quad (\text{A.39})$$

with the square brackets meaning ‘integer part of’. When  $|n|$  is high, the actual value of  $\nu$  has little relevance: we can numerically compute a known  $q(\nu)$  and subtract the excess from its theoretical value to all the other terms. If  $\hat{h}_{\frac{1}{2}} = 0$  is required, no numerical result other than  $q(0) = 1$  is possible, and the same happens for the  $\nu = \frac{1}{2}$  case. However, the  $\nu = \frac{1}{4}$  and  $\nu = \frac{3}{4}$  cases may be useful: for the most general, non-symmetric case, Eq. (A.24) and Eq. (A.25) lead to:

$$\begin{aligned} 1 &= \Re \left[ \hat{h}_{\frac{1}{4}}^* \hat{h}_{\frac{1}{4}} \right] \\ &= \Re \left[ \frac{q(\frac{1}{4})}{q(\frac{1}{2})} \hat{h}_{\frac{1}{4}}^* \hat{h}_{\frac{1}{4}} \right] \\ &= \frac{q(\frac{1}{4})}{q(\frac{1}{2})} \left| \hat{h}_{\frac{1}{4}} \right|^2 \end{aligned} \quad (\text{A.40})$$

and analogously,

$$1 = \frac{q(\frac{3}{4})}{q(\frac{1}{2})} \left| \hat{h}_{\frac{3}{4}} \right|^2 \quad (\text{A.41})$$

where we have simplified the denominator, as  $q(\nu)$  is  $2\pi$  periodic. All the terms that contribute to  $q(\nu)$  are real and positive, also  $\hat{h}_{\frac{3}{4}} = \hat{h}_{\frac{1}{4}}^*$ .

## Bibliography

- [1] G. Nicolis and I. Prigogine. *Self-organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. John Wiley and Sons, 1977.
- [2] H.E. Stanley. *Introduction to phase transitions and critical phenomena*. Oxford Science publications, Oxford, UK, 1987.
- [3] B. B. Mandelbrot. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 62(02):331–358, 1974. doi: 10.1017/S0022112074000711.
- [4] C. Meneveau and K.R. Sreenivasan. Simple multifractal cascade model for fully developed turbulence. *Physical Review Letters*, 59:1424–1427, 1987. 10.1103/PhysRevLett.59.1424.
- [5] A. Turiel and N. Parga. The multi-fractal structure of contrast changes in natural images: from sharp edges to textures. *Neural Computation*, 12:763–793, 2000.
- [6] A. Turiel, H. Yahia, and C. Pérez-Vicente. Microcanonical multifractal formalism: a geometrical approach to multifractal systems. Part I: Singularity analysis. *Journal of Physics A*, 41:015501, 2008. doi: 10.1088/1751-8113/41/1/015501.
- [7] O. Pont, A. Turiel, and C. J. Pérez-Vicente. Application of the microcanonical multifractal formalism to monofractal systems.

- Physical Review E*, 74:061110–061123, 2006. doi: 10.1103/PhysRevE.74.061110.
- [8] O. Pont, A. Turiel, and C. Perez-Vicente. On optimal wavelet bases for the realization of microcanonical cascade processes. under revision in *European Physical Journal*, arXiv:0805.4810, 2009.
- [9] O. Pont, A. Turiel, and C. Perez-Vicente. Empirical evidences of a common multifractal signature in economic, biological and physical systems. *Physica A*, 388(10):2025–2035, February 2009. doi: 10.1016/j.physa.2009.01.041.
- [10] O. Pont, A. Turiel, and C. J. Pérez-Vicente. Description, modeling and forecasting of data with optimal wavelets. *Journal of Economic Interaction and Coordination*, 2009. ISSN 1860-711X (Print) 1860-7128 (Online). doi: 10.1007/s11403-009-0046-x.
- [11] O. Pont, V. Nieves, and A. Turiel. Description of ocean dynamics with optimal wavelet cascade processes. To be submitted to *Journal of Geophysical Research*, 2009.
- [12] O. Pont, A. Turiel, and C. Pérez-Vicente. Volatility forecasting methods with optimal wavelets. in preparation, 2009.
- [13] J. Ballabrera-Poy, B. Mourre, C. Gabarró, O. Pont, and A. Turiel. Microcanonical multifractal analysis of numerical simulations of the ocean. *Journal of Physical Oceanography*. In preparation.
- [14] A. N. Kolmogorov. The local structure of turbulence in an incompressible viscous fluid for very large reynolds numbers. *Dokl. Akad. Nauk. SSSR*, 30:301–305, 1941.
- [15] A. N. Kolmogorov. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible

- fluid at high reynolds number. *Journal of Fluid Mechanics*, 13(01):82–85, 1962. doi: 10.1017/S0022112062000518.
- [16] A. M. Oboukhov. Some specific features of atmospheric turbulence. *Journal of Fluid Mechanics*, 13(1):77–81, 1962. doi:10.1017/S0022112062000506.
- [17] A. M. Yaglom. On the local structure of a temperature field in a turbulent flow. *Dokl. Akad. Nauk SSSR*, 69:743, 1949.
- [18] A. M. Yaglom. Effect of fluctuations in energy dissipation rate on the form of turbulence characteristics in the inertial subrange. *Dokl. Akad. Nauk SSSR*, 166:49–52, 1966.
- [19] R.H. Kraichnan. Small-scale structure of a scalar field convected by turbulence. *Physics of Fluids*, 11(5):945–963, May 1968.
- [20] E. A. Novikov and R. W. Stewart. The intermittency of turbulence and the spectrum of energy dissipation. *Izv. Akad. Nauk SSSR, Ser. Geofiz.*, 3:408–413, 1964.
- [21] E. A. Novikov. Infinitely divisible distributions in turbulence. *Physical Review E*, 50:R3303, 1994.
- [22] U. Frisch. *Turbulence: The legacy of A.N. Kolmogorov*. Cambridge Univ. Press, Cambridge MA, 1995.
- [23] G. Parisi and U. Frisch. On the singularity structure of fully developed turbulence. In M. Ghil, R. Benzi, and G. Parisi, editors, *Turbulence and Predictability in Geophysical Fluid Dynamics. Proc. Intl. School of Physics E. Fermi*, pages 84–87, Amsterdam, 1985. North Holland.
- [24] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani. On the multifractal nature of fully developed turbulence and chaotic systems. *Journal of Physics A*, 17:3521–3531, 1984.

- [25] R. Benzi, S. Patarnello, and P. Santangelo. Self-similar coherent structures in two-dimensional decaying turbulence. *Journal of Physics A: Mathematical and General*, 21:1221–1237, 1988.
- [26] R. Benzi, S. Ciliberto, C. Baudet, G. RuizChavarria, and C. Tripicciono. Extended self similarity in the dissipation range of fully developed turbulence. *Europhysics Letters*, 24:275–279, 1993.
- [27] R. Benzi, S. Ciliberto, C. Tripicciono, C. Baudet, F. Massaioli, and S. Succi. Extended self-similarity in turbulent flows. *Physical Review E*, 48:R29–R32, 1993.
- [28] A. Babiano, G. Boffetta, A. Provenzale, and A. Vulpiani. Chaotic advection in point vortex models and two-dimensional turbulence. *Physics of Fluids*, 6(7):2465–2474, 1994.
- [29] R. Benzi, S. Ciliberto, and C. Baudet G. RuizChavarria. On the scaling of three dimensional homogeneous and isotropic turbulence. *Physica D*, 80:385–398, 1995.
- [30] A. Arneodo, C. Baudet, F. Belin, R. Benzi, B. Castaing, B. Chabaud, R. Chavarria, S. Ciliberto, R. Camussi, F. Chilla, B. Dubrulle, Y. Gagne, B. Hébral, J. Herweijer, M. Marchand, J. Maurer, J. F. Muzy, A. Naert, A. Noullez, J. Peinke, F. Roux, P. Tabeling, W. van de Water, and H. Willaime. Structure functions in turbulence, in various flow configurations, at reynolds number between 30 and 5000, using extended self-similarity. *Europhysics Letters*, 34(6):411–416, 1996.
- [31] R. Benzi, L. Biferale, A. Crisanti, G. Paladin, M. Vergassola, and A. Vulpiani. A random process for the construction of multiaffine fields. *Physica D*, 65:352–358, 1993.
- [32] B. Castaing, Y. Gagne, and E.J. Hopfinger. Velocity probability density functions of high reynolds number turbulence. *Physica*



- D*, 46:177–200, 1990.
- [33] J. F. Muzy, E. Bacry, and A. Arneodo. Wavelets and multifractal formalism for singular signals: Application to turbulence data. *Physical Review Letters*, 67:3515–3518, 1991.
- [34] E. Bacry, J. F. Muzy, and A. Arneodo. Singularity spectrum of fractal signals from wavelet analysis: exact results. *J. of Stat. Phys.*, 70:635–673, 1993.
- [35] B. Dubrulle. Intermittency in fully developed turbulence: log-poisson statistics and generalized scale covariance. *Physical Review Letters*, 73:959–962, 1994.
- [36] Z. S. She and E. Leveque. Universal scaling laws in fully developed turbulence. *Physical Review Letters*, 72:336–339, 1994.
- [37] A. Arneodo, F. Argoul, E. Bacry, J. Elezgaray, and J. F. Muzy. *Ondelettes, multifractales et turbulence*. Diderot Editeur, Paris, France, 1995.
- [38] B. Castaing. The temperature of turbulent flows. *Journal de Physique II*, 6:105–114, 1996.
- [39] A. Arneodo. Wavelet analysis of fractals: from the mathematical concepts to experimental reality. In G. Erlebacher, M. Yousuff Hussaini, and L.M. Jameson, editors, *Wavelets. Theory and applications*, page 349. Oxford University Press. ICASE/LaRC Series in Computational Science and Engineering, Oxford, 1996.
- [40] J. Arrault, A. Arneodo, A. Davis, and A. Marshak. Wavelet-based multifractal analysis of rough surfaces: application to cloud models and satellite data. *Phys. Rev. Lett.*, 79:75–79, 1997.
- [41] A. Arneodo, E. Bacry, S. Jaffard, and J.F. Muzy. Singularity spectrum of multifractal functions involving oscillating singular-

- ities. *Journal of Fourier Analysis and Application*, 4(2):159–174, 1998.
- [42] S. G. Roux, A. Arneodo, and N. Decoster. A wavelet-based method for multifractal image analysis. iii. applications to high-resolution satellite images of cloud structure. *Eur. Phys. J. B*, 15:765–786, 2000.
- [43] L. Chevillard, B. Castaing, E. Leveque, and A. Arneodo. Unified multifractal description of velocity increments statistics in turbulence: Intermittency and skewness. *Physica D*, 218:77–82, 2006.
- [44] P. Chainais. Infinitely divisible cascades to model the statistics of natural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:2105–2119, 2007.
- [45] D. Schertzer and S. Lovejoy. Physically based rain and cloud modeling by anisotropic, multiplicative turbulent cascades. *Journal of Geophysical Research*, 92:9692–9714, 1987.
- [46] Y. Tessier, S. Lovejoy, and D. Schertzer. Universal multifractals in rain and clouds: theory and observations. *Journal of Applied Meteorology*, 32:223–250, 1993.
- [47] T. Falco, F. Francis, S. Lovejoy, D. Schertzer, B. Kerman, and M. Drinkwater. Universal multifractal scaling of synthetic aperture radar images of sea-ice. *IEEE Transactions on Geosciences and Remote Sensing*, 34:906–914, 1996.
- [48] D. Schertzer and S. Lovejoy. Universal multifractals do exist!: comments on ” a statistical analysis of mesoscale rainfall as a random cascade”. *Journal of Applied Meteorology*, 36:1296–1303, 1997.
- [49] L. Seuront, F. Schmitt, Y. Lagadeux, D. Schertzer, and S. Lovejoy. Universal multifractal analysis as a tool to characterize mul-

- tiscale intermittent patterns: examples of phytoplankton distribution in turbulent coastal water. *Journal of Plankton Research*, 21:877–922, 1999.
- [50] S. Lovejoy, D. Schertzer, Y. Tessier, and H. Gaonac’h. Multifractals and resolution-independent remote sensing algorithms: the example of ocean colour. *International Journal of Remote Sensing*, 22(7):1191–1234, 2001.
- [51] A. Davis, A. Marshak, and W. Wiscombe. Wavelet based multifractal analysis of non-stationary and/or intermittent geophysical signals. In E. Foufoula-Georgiou and P. Kumar, editors, *Wavelets in Geophysics*, pages 249–298. Academic Press, New York, 1994.
- [52] L. A. N. Amaral, A. L. Goldberger, P. Ch. Ivanov, and H. E. Stanley. Scale-independent measures and pathologic cardiac dynamics. *Physical Review Letters*, 81(11):2388–2391, September 1998.
- [53] Y. Ashkenazy, J. A. Hausdorff, P. C. Ivanov, and H. E. Stanley. A stochastic model of human gait dynamics. *Physica A*, 316:662–670, 2002.
- [54] P. Ivanov, L. Amaral, A. Goldberger, S. Havlin, M. Rosenblum, Z. Struzik, and H. Stanley. Multifractality in human heartbeat dynamics. *Nature*, 399:461–465, 1999.
- [55] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk. A multifractal wavelet model with application to network traffic. *IEEE Transactions on Information Theory*, 45(4):992–1018, April 1999.
- [56] V. J. Ribeiro, R. H. Riedi, M. S. Crouse, and R. G. Baraniuk. Multiscale queuing analysis of long-range-dependent network traffic. In *INFOCOM (2)*, pages 1026–1035, 2000.

- [57] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch. The multiscale nature of network traffic: Discovery, analysis, and modelling. *IEEE Signal Processing Magazine*, 19(3):28–46, 2002.
- [58] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions in Image Processing*, 8:1688–1701, 1999.
- [59] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky. Random cascades on wavelet trees and their use in modeling and analyzing natural images. *Applied Computational and Harmonic Analysis*, 11:89–123, 2001.
- [60] A. Srivastava, A.B. Lee, E.P. Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *J. of Mathematical Imaging and Vision*, 18:17–33, 2003.
- [61] B. B. Mandelbrot. The variation of certain speculative prices. *Journal of Bussiness*, XXXVI:392–417, 1963.
- [62] B.B. Mandelbrot, A. Fisher, and L. Calvet. A multifractal model of asset returns. *Cowles Foundation Discussion Paper No. 1164*, 1997.
- [63] L. Calvet and A. Fisher. Forecasting multifractal volatility. *Journal of Econometrics*, 105:27–58, 2001.
- [64] L.E. Calvet and A.J. Fisher. *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press Advanced Finance Series. Elsevier Science & Technology Books, 2008. ISBN 0121500136.
- [65] R. N. Mantegna and H. E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376:46–49, 1995.

- [66] R. N. Mantegna and H. E. Stanley. Turbulence and financial markets. *Nature*, 383:587–588, 1996.
- [67] P. Gopikrishnan, V. Plerou, Y. Liu, L. A. N. Amaral, X. Gabaix, and H. E. Stanley. Scaling and correlation in financial time series. *Physica A*, 287:362–373, 2000.
- [68] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H.E. Stanley. Scaling and memory in volatility return intervals in financial markets. *Proceedings of the National Academy of Sciences*, 102:9424–9428, 2005.
- [69] A. Arneodo, J.-F. Muzy, and D. Sornette. “direct” causal cascade in the stock market. *European Physical Journal B*, 2:277–282, 1998.
- [70] J. F. Muzy, J. Delour, and E. Bacry. Modelling fluctuations of financial time series: from cascade process to stochastic volatility model. *Euro. Phys. Journal B*, 17:537–548, 2000.
- [71] J.-F. Muzy, D. Sornette, J. Delour, and A. Arneodo. Multifractal returns and hierarchical portfolio theory. *Quantitative Finance*, 1:131–148, 2001.
- [72] J. Perelló, J. Masoliver, and J.P. Bouchaud. Multiple time scales in volatility and leverage correlations: a stochastic volatility model. *Journal of Mathematical finance*, 11:27–50, 2004.
- [73] Josep Perelló, Jaume Masoliver, Andrzej Kasprzak, and Ryszard Kutner. Model for interevent times with long tails and multifractality in human communications: An application to financial trading. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 78(3):036108, 2008. doi: 10.1103/PhysRevE.78.036108.
- [74] A.B. Chhabra, C. Meneveau, R.V. Jensen, and K.R. Sreenivasan. Direct determination of the  $f(\alpha)$  singularity spectrum

- and its application to fully developed turbulence. *Physical Review A*, 40(9):5284–5294, November 1989.
- [75] A. Chhabra and R.V. Jensen. Direct determination of the  $f(\alpha)$  singularity spectrum. *Phys. Rev. Lett.*, 62:1327–1330, 1989.
- [76] S. Jaffard. Pointwise smoothness, two-microlocalization and wavelet coefficients. *Publicacions Matemàtiques*, 35:155–168, 1997.
- [77] S. Jaffard. Some mathematical results about the multifractal formalism for functions. In C. K. Chui, L. Montefusco, and L. Puccio, editors, *Wavelets: theory, algorithms, and applications (Taormina, 1993)*, volume 5 of *Wavelet Anal. Appl.*, pages 325–361. Academic Press, San Diego, CA, 1994.
- [78] J.F. Muzy, E. Bacry, and A. Arneodo. The multifractal formalism revisited with wavelets. *Int. J. of Bifurcation and Chaos*, 4:245, 1994.
- [79] S. Jaffard. Multifractal formalism for functions. I. Results valid for all functions. *SIAM Journal of Mathematical Analysis*, 28(4):944–970, 1997.
- [80] Z. R. Struzik. Determining local singularity strengths and their spectra with the wavelet transform. *Fractals*, 8(2):163–179, June 2000.
- [81] J. Isern-Fontanet, A. Turiel, E. Garcia-Ladona, and J. Font. Microcanonical multifractal formalism: application to the estimation of ocean surface velocities. *Journal of Geophysical Research*, 112:C05024, 2007. doi: 10.1029/2006JC003878.
- [82] V. Nieves, C. Llebot, A. Turiel, J. Solé, E. García-Ladona, M. Estrada, and D. Blasco. Common turbulent signature in sea

- surface temperature and chlorophyll maps. *Geophysical Research Letters*, 34:L23602, 2007. doi: 10.1029/2007GL030823.
- [83] C. Pottier, A. Turiel, and V. Garçon. Inferring missing data in satellite chlorophyll maps using turbulent cascading. *Remote Sensing of Environment*, 112:4242–4260, 2008. doi: 10.1016/j.rse.2008.07.010.
- [84] A. Turiel, J. Solé, V. Nieves, J. Ballabrera-Poy, and E. García-Ladona. Tracking oceanic currents by singularity analysis of micro-wave sea surface temperature images. *Remote Sensing of Environment*, 112:2246–2260, 2008.
- [85] J. Grazzini, A. Turiel, and H. Yahia. Entropy estimation and multiscale processing in meteorological satellite images. In *Proc. of ICPR 2002*, volume 3, pages 764–768, Los Alamitos, CA, 2002. IEEE Computer Society.
- [86] A. Turiel, J. Grazzini, and H. Yahia. Multiscale techniques for the detection of precipitation using thermal IR satellite images. *IEEE Geoscience and Remote Sensing Letters*, 2(4):447–450, October 2005. doi:10.1109/LGRS.2005.852712.
- [87] J. Grazzini, A. Turiel, H. Yahia, and I. Herlin. A multifractal approach for extracting relevant textural areas in satellite meteorological images. *Environmental Modelling and Systems*, 22:323–334, 2007.
- [88] H. Yahia, A. Turiel, N. Chrysoulakis, J. Grazzini, P. Prastacos, and I. Herlin. Application of the multifractal microcanonical formalism to the detection of fire plumes in noaa-avhrr data. *Int. J. Remote Sens.*, 29(14):4189–4205, 2008. ISSN 0143-1161. doi: <http://dx.doi.org/10.1080/01431160701840174>.
- [89] A. Turiel, G. Mato, N. Parga, and J. P. Nadal. Self-similarity properties of natural images. In *Proc. of NIPS'97*, volume 10,

- pages 836–842. MIT Press, 1997.
- [90] A. Turiel, G. Mato, N. Parga, and J. P. Nadal. The self-similarity properties of natural images resemble those of turbulent flows. *Physical Review Letters*, 80:1098–1101, 1998.
- [91] A. Turiel and N. Parga. Multifractal wavelet filter of natural images. *Physical Review Letters*, 85:3325–3328, 2000.
- [92] A. Turiel and A. del Pozo. Reconstructing images from their most singular fractal manifold. *IEEE Trans. on Im. Proc.*, 11:345–350, 2002.
- [93] N. Parga, J. M. Delgado, and A. Turiel. Predicting receptive fields in v1 from natural scene properties. In *Society for Neuroscience Abstracts*, volume 29, pages 8–12, 2003.
- [94] J.M. Delgado, A. Turiel, and N. Parga. Receptive fields of simple cells from a taxonomic study of natural images and suppression of scale redundancy. *Neurocomputing*, 69(10-12):1224–1227, June 2006.
- [95] A. Turiel and C. Pérez-Vicente. Multifractal geometry in stock market time series. *Physica A*, 322:629–649, May 2003.
- [96] A. Turiel and C. Pérez-Vicente. Role of multifractal sources in the analysis of stock market time series. *Physica A*, 355:475–496, September 2005.
- [97] A. Turiel and C. Perez-Vicente. Dynamical decomposition of multifractal time series as fractal evolution and long-term cycles: applications to foreign currency exchange market. In *Complexus Mundi: Emergent patterns in Nature*, pages 73–82, 2006.
- [98] I. Daubechies. *Ten lectures on wavelets*. CBMS-NSF Series in App. Math. Capital City Press, Montpelier, Vermont, 1992.



- [99] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd Edition, 1999.
- [100] S. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.
- [101] S. Mallat and S. Zhong. Wavelet transform maxima and multi-scale edges. In Mary Beth Ruskai et al, editor, *Wavelets and their applications*. Jones and Bartlett, Boston, 1991.
- [102] J. Paret and P. Tabeling. Intermittency in the two-dimensional inverse cascade of energy: Experimental observations. *Physics of Fluids*, 10(12):3126–3136, 1998.
- [103] J. Paret, M.-C. Jullien, and P. Tabeling. Vorticity statistics in the two-dimensional enstrophy cascade. *Physical Review Letters*, 83(17):3418–3421, October 1999.
- [104] M.-C. Jullien, P. Castiglione, and P. Tabeling. Experimental observation of Batchelor dispersion of passive tracers. *Physical Review Letters*, 85(17):3636–3639, October 2000.
- [105] J. Huang and D. Mumford. Statistics of natural images and models. In *Proc. CVPR*, pages 541–547, 1999.
- [106] O. Schwartz and E.P. Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825, 2001.
- [107] A. Turiel, J.-P. Nadal, and N. Parga. Orientational minimal redundancy wavelets: from edge detection to perception. *Vision Research*, 43(9):1061–1079, 2003.
- [108] T. Carleman. Sur le problème des moments. *Comptes rendus Acad. Sci. Paris*, 174:1680, 1922.
- [109] A. N. Kolmogorov. Dissipation of energy in a locally isotropic turbulence. *Dokl. Akad. Nauk. SSSR*, 32:16–18, 1941.

- [110] Alexandre Arenas and Alexandre J. Chorin. On the existence and scaling of structure functions in turbulence according to the data. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4352–4355, March 2006. doi: <http://dx.doi.org/10.1073/pnas.0600482103>.
- [111] S. Mallat and W. L. Huang. Singularity detection and processing with wavelets. *IEEE Trans. in Inf. Th.*, 38:617–643, 1992.
- [112] W. Rudin. *Real and Complex Analysis*. Mc Graw Hill, New York, USA, 1987.
- [113] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley and sons, Chichester, 1990.
- [114] Mogens H. Jensen, Leo P. Kadanoff, Albert Libchaber, Itamar Procaccia, and Joel Stavans. Global universality at the onset of chaos: Results of a forced rayleigh-bénard experiment. *Phys. Rev. Lett.*, 55(25):2798–2801, Dec 1985. doi: [10.1103/PhysRevLett.55.2798](https://doi.org/10.1103/PhysRevLett.55.2798).
- [115] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B.I. Shraiman. Fractal measures and their singularities: the characterization of strange sets. *Physical Review A*, 33:1141–1151, 1986. doi: [10.1103/PhysRevA.33.1141](https://doi.org/10.1103/PhysRevA.33.1141).
- [116] Charles Meneveau and K. R. Sreenivasan. Measurement of  $f(\alpha)$  from scaling of histograms, and applications to dynamical systems and fully developed turbulence. *Physics Letters A*, 137(3):103 – 112, 1989. ISSN 0375-9601. doi: DOI: [10.1016/0375-9601\(89\)90093-5](https://doi.org/10.1016/0375-9601(89)90093-5).
- [117] J. M. Gutiérrez and M. A. Rodríguez. A new exact method for obtaining the multifractal spectrum of multiscaled multinomial measures and ifs invariant measures. *Chaos, Solitons*

- É* *Fractals*, 11(5):675 – 683, 2000. ISSN 0960-0779. doi: DOI:10.1016/S0960-0779(98)00154-4.
- [118] A. Turiel, C. Pérez-Vicente, and J. Grazzini. Numerical methods for the estimation of multifractal singularity spectra on sampled data: a comparative study. *Journal of Computational Physics*, 216(1):362–390, July 2006.
- [119] A. Einstein. Investigations on the theory of the brownian movement. *Annalen der Physik*, 17:549–560, 1905.
- [120] A. Turiel and C. Pérez-Vicente. Multifractal measures: definition, description, synthesis and analysis. a detailed study. In J.-P. Nadal, A. Turiel, and H. Yahia, editors, *Proceedings of the "Journées d'étude sur les méthodes pour les signaux complexes en traitement d'image"*, pages 41–57, Rocquencourt, 2004. INRIA.
- [121] W.J. Emery, A.C. Thomas, M.J. Collins, W.R. Crawford, and D.L. Mackas. An objective method for computing advective surface velocities from sequential infrared satellite images. *Journal of Geophysical Research*, 91:12865–12878, 1986.
- [122] R. B. Scott and F. Wang. Direct evidence of an oceanic inverse kinetic energy cascade from satellite altimetry. *J. Phys. Oceanogr.*, 35(9):1650–1666, September 2005.
- [123] R.W. Reynolds and T.M. Smith. Improved global sea surface temperature analyses using optimal interpolation. *Journal of Climate*, 7:929–948, 1994.
- [124] J. Pedlosky. *Geophysical fluid dynamics*. Springer-Verlag, 1987.
- [125] J. Pedlosky. *Ocean Circulation Theory*. Springer, 1998.

- [126] F. Truchetet. *Ondelettes pour le signal numérique*. Hermes Sciences Publicat., 1998.



UNIVERSITAT DE BARCELONA

