Multiscale Statistical Analysis of Self-Similar Processes with Applications in Geophysics and Health Informatics

A Dissertation
Presented to
The Academic Faculty

by

Bin Shi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Industrial and Systems Engineering
Georgia Institute of Technology
March 2005

Multiscale Statistical Analysis of Self-Similar Processes with Applications in Geophysics and Health Informatics

Approved by:

Dr. Brani Vidakovic
School of Industrial and Systems Engineering
Georgia Institute of Technology, Advisor

Dr. Francois Sainfort
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Gabriel Katul
School of Environment and Earth Sciences
Duke University

Dr. Xiaoming Huo
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Julie Jacko
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 5 April 2005

To my wife

To my father and mother

To Ms. Peng, my middle school teacher

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In this dissertation, we address the statistical analysis under the multiscale framework for the self-similar process. Motivated by the problems arising from geophysics and health informatics, we develop a set of statistical measures as discriminative summaries of the self-similar process. These measures include Multiscale Schur Monotone (MSM) measures, Geometric Attributes of Multifractal Spectrum (GAMFS), Quasi-Hurst exponents, Mallat Model and Tsallis Maxent Model. These measures are used as methods to quantify the difference (or similarities) or as input (feature) vectors in the classification model. As the cornstone of GAMFS, we study the estimation of multifractal spectrum and adopt a Weighted Least Squares (WLS) schemes in the wavelet domain to minimize the heteroskedastic effects , which is inherent because the sample variances of the wavelet coefficients depend on the scale. We also propose a Combined K-Nearest-Neighbor classifier (Comb-K-NN) to address the inhomogeneity of the class attributes, which is indicated by the large variations between subsets of input vectors. The Comb-K-NN classifier stabilizes the variations in the sense of reducing the misclassification rates. Bayesian justifications of Comb-K-NN classifier are provided.

GAMFS, Quasi-Hurst exponents, Mallat Model and Tsallis Maxent Model are used in the study of assessing the effects of atmospheric stability on the turbulence measurements in the inertial subrange. We also formulate the criteria for success in evaluating how atmospheric stability alters the MFS of a single flow variable time series as a statistical classification model. We use the multifractal discriminate model as the solution of this problem. Also, high frequency pupil-diameter dynamic measurements, which are well documented as measures of mental workload, are summarized using both GAMFS and MSM. These summaries are further used as the feature vector in the Comb-K-NN classifier. The serious

inhomogeneity among subjects in the same user group makes classification difficult. These difficulties are overcome by using Comb-K-NN classifier.

CHAPTER I

INTRODUCTION

Classes of random processes which intrinsically have invariant statistical properties at different scales are often referred as self-similar process. Self-similar process has been very common in science, finance, manufacturing and other areas since the last two decades. Analyzing self-similar process was often requested for purpose of scientific research, medical study, business management, quality improvements etc. However, the non-trivial structures of the self-similar process make it hard to directly use the classical statistical model. Self-similarity is always associated with scaling invariance, mono- and multi-fratality, long-rang dependent, heavy tailness etc. Therefore, it is necessary and urgent to develop some sophisticate models to address these structures associated with the self-similar process. This thesis focus on contributing some statistical models under the multi-scale framework to analyze the self-similar process with several case studies arising from geophysics and health informatics.

1.1 Self-similarity, Scaling, Fractality and Long Range Dependence

The notion of self-similarity is pioneered by Mandelbrot (1968), in which the sequence of stock returns was studied. Mandelbrot argued that Brownian motion is not an adequate statistical description of the true stochastic process generating securities returns. The notion of self-similar process is defined as follows.

Definition 1.1.1 *A random process $X(t), t > 0$ is called self-similar if for any $a > 0$, there exists $b > 0$ such that*

$$X(at) \stackrel{d}{=} bX(t). \tag{1}$$

We assume that all processes discussed are real valued and defined on the same parameter space. Two processes $X(t)$ and $Y(t)$, *equal in all finite dimensional distributions*, will be denoted as $X(t) \stackrel{d}{=} Y(t)$. This means that for any selection of "times" $0 \leq t_1 < t_2 < \ldots t_k < \infty$ random vectors $(X(\omega, t_1), \ldots, X(\omega, t_k))$ and $(Y(\omega, t_1), \ldots, Y(\omega, t_k))$ have the same distribution. Informally, processes equal-in-distribution are statistically indistinguishable.

Random process $X(t)$ is called stochastically continuous at t_0 if $\lim_{h \to 0} P(|X(t_0 + h) - X(t_0)| > \epsilon) = 0$, for any fixed $\epsilon > 0$.

Also, we consider processes not to be trivial. [1]

Lamperti (1962) proved the result,

Theorem 1.1.1 *(Lamperti, 62.) If random process $X(t), t \geq 0$ is nontrivial, stochastically continuous at 0, and self-similar, then there exists unique $H \geq 0$ such that $b = a^H$. If $X(0) = 0, a.s.$ then $H > 0$.*

Standard definition of self-similar processes is as follows,

Definition 1.1.2 *Process $X(t)$, $t \geq 0$ is self-similar, with self-similarity index H (H-ss) if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.*

Uniqueness of H is not obvious from this definition, although, H is unique by the Lamperti's theorem. Also, from Definition 1.1.2 it follows $X(0) = 0$.

This definition implies that sample paths (t, X) and $(at, (1/a)^H X)$ are statistically equivalent to each other, which tells that $X(t)$ is statistically similar to the its dilated version. Fig. 1 illustrates the self-similarity using fractional Brownian motion ($H = 1/3$)(discussions about fractional Brownian motion is postpone to Chapter 2), which is the classical model for self-similar process.

---

[1] Process $X(t)$ is trivial if the distribution of random variable $X(\omega, t)$, t fixed is a point mass measure. For example, $X(t) = $ const or $X(t) = \sin(t)$ would be examples of trivial processes.

(a) Orginal sample   (b) Dilated version of subsample

Figure 1: Self-similar process $X(t)$ and the dilated version of its subsample. 2048 points were sampled between 0 and 1 for $X(t)$ as shown in (a). 512 points were sampled between 0.25 and 0.5 as subsample and the subsample is dilated as shown in (b).

The notion of long-range dependent (LRD) is defined through spectral density. However, the discussion of spectral density in time series usually requires the stationarity and unfortunately the self-similarity usually implies nonstationarity. On the other side, it is possible define a counterpart of a spectral density of a nonstationary process if, for example, linear filtering will produce a stationary process. In such cases, we define a pseudo (quasi) spectral density as a function of spectral density of filtered stationary process and transfer function of a filter. An example is spectrum for the fractional Brownian motion.

Definition 1.1.3 *A stationary process $Y(t)$ is called long-range dependent (LRD) process if its autocorrelation function or spectral density behave as*

$$\gamma_Y(h) \sim C_\gamma |h|^{\alpha-1},\ h \to \infty, \alpha \in (0,1), \tag{2}$$

or

$$f_Y(\omega) \sim C_f |\omega|^{-\alpha},\ \omega \to 0, \alpha \in (0,1), \tag{3}$$

where C_γ and C_f are two related constants.

These two relations are equivalent, subject to mild asymptotic monotonicity assumptions on γ.

Next, we will make link between self-similarity and LRD.

Let $X(t), t \in R$ be H-ss process. If its increments are stationary, i.e, if the distribution of $X(t+h) - X(t)$ is independent of t, it will be called H-sssi process.

The following theorem gives the form of autocorrelation function of any H-sssi process with finite second moment.

Theorem 1.1.2 *Let $X(t), t \in \mathbb{R}$ be an H-sssi process for which $E|X(1)|^2 < \infty$. Then,*

$$\gamma(t,s) = EX(t)X(s) = \frac{E|X(1)|^2}{2} \left[|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right]. \tag{4}$$

On the other hand, Let $X(t)$ be an H-sssi process with $0 < H < 1$ and $E|X(1)|^2 < \infty$. Define stationary sequence of random variables $Y(n)$ as

$$Y(n) = X(n+1) - X(n).$$

It is well known that $Y(n)$ is asymptotically second-order self-similar, i.e., the second order moments of Y_n and aggregated time series $Y^{(m)}$ coincide. The series $Y^{(m)}(k)$ is defined as series of averages of non-overlapping blocks of size m from the sequence $Y(n)$,

$$Y^{(m)}(k) = \frac{Y(km - m + 1) + \cdots + Y(km)}{m}, \tag{5}$$

It is easy to see that if $1/2 < H < 1$, the asymptotic behavior of $Var Y^{(m)}$ is influenced by asymptotic behavior of $\gamma_Y(n)$. Indeed,

$$Var Y^{(m)} \sim 1/m \, \gamma_Y(0) + \sum_{k=1}^{m-1} k^{2H-2}(m-k) \sim m^{2H-2}.$$

Informally, $Y^{(m)}(k)$ and $Y(n)$ look similar at all scales.

Fractality is borrowed from the fractal geometry to describe the self-similar process. The self-similarity leads to the facts that the whole process $X(t)$ can not be distinguished

from it is subpart $c^H X(t/c), \ \ c > 1$ in statistical sense. This coincides the original defini-tions of fractality about the geometric objects.

More generally, multifractal measures are proposed to model temporal heterogeneity of the self-similar process.

Definition 1.1.4 *A random process $X(t), t > 0$ is called Multifractal Process if for any $a > 0$, there exists a random function $M(a)$ such that*

$$X(at) \stackrel{d}{=} M(a)X(t). \tag{6}$$

Here the scaling (self-similar) factor $M(a)$ is a random variable, whose distribution does not depend on the particular instant t. Exact Self-similar processes is a degenerate example of multifractal, with $M(a) = aH$ and sometimes is referred as mono-fractal pro-cess. The generalized self-similarity index is defined $H(a) = \log_a M(a)$. Therefore we can rewrite the above relation 6 as: $X(at) \stackrel{d}{=} aH(a)X(t)$: In contrast to self-similar processes, the index $H(a)$ is a random function of a.

An alternative definition, which allows more convinces, about the multifractal process can be described as:

Definition 1.1.5 *A random process $X(t), t > 0$ is called Multifractal Process if it has stationary increments and satisfies*

$$E(|X(t)|^q) = c(q)^{\tau(q)+1}, \ \ for \ all \ \ t \in T; q \in Q \tag{7}$$

where T and Q are intervals on the real line, $\tau(q)$ and $c(q)$ are functions with domain Q. Moreover, we assume that T and Q have positive lengths, and that $0 \in T$, $[0 \ 1] \subseteq Q$.

Function $\tau(q)$ is the *scaling function* for multifractal process. All $\tau(q)$ has the intercept -1, which is implied by $E(|X(t)|^q) = 0$ at $q = 0$. As a special case, monofractal has the linear scaling function $\tau(q) = Hq - 1$. It is also shown that the $\tau(q)$ is always a concave function for all multifractal function.

5

Another immediate consequence of self-similarity is that we can not find a characteristic scale of time or space (frequency) which is more important to describe the process than others. This fact is usually referred as scaling. In scaling process, all scales incorporate identically information about the dynamics of the stochastic process. Therefore, it is not the major task to identify just a few scales in modeling the self-similarity, instead, the similarities across between scales are of great interests.

1.2 Motivation Problems

In this section, problems arising from geophysics, health informatics and manufacturing will be given to illustrate the importance and challenges of analyzing the self-similar process.

1.2.1 Geophysics Problems

Theoretical arguments based on the Navier-Stokes equations have suggested that turbulence is representative of complex processes rich in variability across broad range of time scales. This scaling behavior has been verified by many numerical and field experiments. Statistically modeling of the scaling behavior is of great interests in geophysics community. In turbulence research, there are many unresolved scientific questions, some of which depend on very much the understanding of the self-similarity of the turbulence measurements.

A set of turbulence measurements have been obtained by the team from Duke University. Time series measurements of u,v w, and T were collected over a grass-covered forest clearing at Duke Forest near Durham, North Carolina. The measurements were collected on June, 12-16, 1995 at $5.2\,m$ above the grass surface using a GILL triaxial sonic anemometer. In Geophysics convention, the velocities u,v and w are called turbulence velocities while temperature velocities T is referred as turbulence scaler. The components of turbulence velocities u, v and w are usually orthogonal each other after necessary rotation operation on the raw data. Component u is also aligned with the direction of the wind.

The time series plots of the four components of a typical turbulence measurement in the stable regime are shown in Fig. 2. Apparently, these plots suggest some interesting characteristics of turbulence measurements, i.e. very noisy, intermittent, non-stationary, highly peaky etc. On the other hand, if we study the turbulence in the frequency domain, as shown in Fig. 3, a consistent phenomena, logarithm linearity, is observed over these four components. This logarithm linearity in the frequency domain is known as the scaling behavior or power law. This is related to famous Kolmogorov (1941) view of universal scaling, which interpreters the physical process of the energy cascade over a range of scales.

Figure 2: The time series plots of the four components of a typical turbulence measurement under stable condition.

The departure from the so-called Kolmogorov (1941) view of universal scaling and subsequent refinements (Kolmogorov, 1962) is now supported by numerous observations and theoretical arguments regarding the anomalous scaling in measured structure functions,

Figure 3: Spectral density in the log-log scale for the data shown in Fig. 2

particularly for passive scalars (Pumir and Shraiman, 1995, Sreenivasan and Antonia, 1997, Celani et al, 2000, Warhaft, 2000, Antonov and Honkonen, 2001), and static pressure (Albertson et al., 1998). The anomalous scaling is commonly attributed to short-circuiting of the energy cascade process due to the existence of organized large-scale features such as ramp-like structures, which are influenced by boundary conditions, and themselves directly influence small scale turbulence (Warhaft, 2000, Celani and Vergassola, 2001).

There is a need in geophysics to quantify whether boundary conditions influence the statistical properties of the inertial subrange for the atmospheric surface layer (*ASL*). This influence is also called satiability effects. In this thesis, this quantification is studied in several ways. By appropriate modeling these turbulence measurements, we expect to quantify the stability effects with both accuracy and meaningful physical interpretation.

1.2.2 Health Informatics Problems

In the human visual system, pupil functions as a gain control device, which response to the external stimuli, such as luminance change, color and pattern change, onset of motion, attention, social signaling etc., in a very subtle way. In Human-Computer-Interaction (HCI), pupillary response (in terms of the dynamic pupil size) is becoming an important mental workload measure, which is useful in designing the computer or equipments for some special needs. HCI community is interested in the mental workload of low vision user. Low vision user usually have vision impairment, which should causes different pupillary response from normal user during some computer interaction tasks, such as iconic target selection , drag-and-drop, reading, searching, object positioning etc. The interests in the low vision users stem from the large magnitude of eye disorder population in USA. Almost One-third (80 million) of Americans are recognized with ocular abnormality that may affect visual function. A deep investigation of how the vision ability affects the pupillary response is clearly important and valuable. This helps better understand the unique interaction needs and behaviors of individuals with varying abilities.

To study the pupillary response in low vision user group, an experiment has been done on different individuals within four categories of vision ability (See Chapter 4 for details). The pupillary response was collected when they were performing some simple computer tasks. Fig. 4 illustrates the pupillary response (pupil size) time series for individuals from four different vision ability categories. It is intuitive that there are statistical differences in the pupillary response among these categories. However, it turns out to be hard to quantify these difference by only using the simple descriptive statistics, such as mean and standard deviations. This is caused by the long range dependence effects of the pupil signal. A simple example could be used to illustrated this. We simulated a fractional Gaussian Noise (fGn) $\{X_i, i = 1, 2, \cdots, 2048\}$ with $H = 0.8$, which is a popular model for stationary long range dependent time series in many literatures. Here $X_i \sim N(0, 1)$. The time series and its sample autocorrelation function are presented in Fig. 5. We run 1000 replicates of such a fGn process and estimate the sample mean and sample standard deviations from each of them. Those sample statistics are summarized in Fig. 6. It is apparent that the sample standard deviation is seriously biased. The sample mean looks unbiased however the estimates are quite unstable since it could rand from -0.5 to 0.5 very likely. This example tells that the long range dependence easily destroys the accuracy of the sample mean and other sample moments. This fact makes it hard to tell the differences among various groups using such simple statistics as sample mean and sample standard deviation because the sample autocorrelations of the pupil size dynamic measurements from wide range of subjects universally suggest the long range dependent structure.

An important characteristic of the pupil size measurements, is the apparent scaling behavior, which is similar to the turbulence data. We illustrate this in Fig. 7.

1.3 Thesis Contributions

The overall objective of this dissertation is to model the self-similarity process with statistical accuracy and to apply the theory to solve the scientific, engineering and medical

Figure 4: Pupillary response (pupil size) time series for individuals from four different vision ability categories. The details of this experiment and categories will be delayed in Chapter 4.

Figure 5: A simulated fractional Gaussian Noise with $H = 0.8$ (left panel) and the corresponding sample autocorrelations (right panel).

Figure 6: Sample mean and sample standard deviation for the simulated fractional Gaussian Noise with $H = 0.8$ shown in Fig. 5

Figure 7: Spectral densities for the data shown in Fig. 4 in the log-log plot

problems. In particular, we study and develop techniques that are able to derive meaningful summaries, which include global and local scaling measures, multifractal broadness, slopes and Hurst, multiscale Schur monotone measures etc., from self-similar processes. As a by-product, we investigate the combined k-nearest-neighbor classification. The main contribution of the this dissertation can be summarized as follows:

- We propose to apply global and local scaling measures (e.g. quasi-Hurst exponent, distributional properties of the wavelet coefficients, and Tsallis's thermostatic entropy measures) to quantify the differences (similarities) between self-similar processes. Quasi-Hurst generalizes the the idea of defining self-similarity parameters and extend this to the non-stationary self-similar process. Probabilistic models of the wavelet coefficients are studied in the context of Mallat's model and Tsallis Maxent model. A theoretical link between these two models is derived. These measures are used to assess whether atmospheric stability impacts both local and global inertial subrange scaling in the the atmospheric surface layer (*ASL*).

- We propose a Weighted Least Squares (WLS) scheme in wavelet domain to estimate the multifractal spectrum (MFS). By applying this, we minimize the heteroskedastic effects which is inherent because the sample variances of the wavelet coefficients depend on the scale. We also derived the discriminative measures including the left slope (rise), Hurst exponent (maxima), and broadness. We attempted to use these summaries to quantify how atemospheric stability affects turbulent velocity and temperature fluctuations in the atmospheric surface layer. These summary measures are applied to the MFS of velocity and temperature time series collected in the atmospheric surface layer for a wide range of atmospheric stability conditions. In a similar spirit, we develop a multifractal discrimination model (MDM) to discriminate the HCI user group using the measures of the left slope (rise), Hurst exponent (maxima), and broadness in MFS estimated from the high frequency pupil size data.

- We propose a combined k-Nearest-Neighbor (Comb-k-NN) classification model to address the inhomogeneous property of discriminative vectors. This method is proposed to minimize and stabilize the misclassification rate in training and test sets with the goal of improving classification accuracy. It is also justified by a Bayesian paradigm. We use Comb-k-NN as ingredients in the classification problem for the HCI users based on high frequency pupil size data. Reasonable accuracy has been achieved.

- We propose a multiscale Schur monotone (MSM) measure as summaries of self-similar processes. MSM measure is derived to characterize the disbalance properties of the data distribution at different frequency scales. In contrast to the global disbalance measures, the MSM measure carries information not only about the disbalance characteristics of the data, but also about its correlation structure. Thus, the MSM summary is more likely to be more sensitive to the differences in dynamics between different self-similar processes than any other single measure, such as correlation and global Schur Monotone measures. MSM measures also allow the adaptive summary of the process through wavelet basis selection and the choice of scale resolutions.

1.4 Thesis Organization

This dissertation is organized as follows:

In Chapter 2, we provide mathematical preliminaries including wavelet transform, self-similarity, multifractal theory, long range dependence. Definition and properties of self-similar process are provided.

In Chapter 3, we survey different mechanisms that explain what causes self-similarity in processes.

In Chapter 4, we introduce the multiscale Schur monotone (MSM) measure as summaries of self-similar processes. Meanwhile, we provide the details of the combined k-Nearest-Neighbor (Comb-k-NN) classification model. As an application, we use MSM

and Comb-k-NN to classify the HCI users based on the dynamic pupil size measurements.

In Chapter 5, we define the broadness, slopes (rises) and Hurst exponent (maxima) for multifractal spectrum (MFS). Also, the estimation of the multifractal spectrum using Weighted Least Squares (WLS) in the wavelet domain is introduced to minimize the heteroskedastic effects and achieve robust estimations. Numerical issues about computing the broadness, rises and maxima are discussed. These summary measures are applied to the MFS of turbulent (velocity and temperature) time series collected in the atmospheric surface layer for a wide range of atmospheric stability conditions.

In Chapter 6, we adopt the mulitfractal discrimination model in the classification problem, in which classifying the pupil-size measurements for HCI users is presented as an application example.

In Chapter 7, we assess the effects of atmospheric stability on the turbulence in the inertial subrange using both global and local scaling measures. The quasi-Hurst exponent, Mallat's Model and Tsallis MaxEnt model are studied in details.

Last, Chapter 8 contains concluding remarks and future research directions.

1.5   References

1. Albertson, J.D., Katul, G.G., Parlange, M.B., and Eichenger, W.E. (1998). Spectral scaling of static pressure fluctuations in the atmospheric surface layer: the interaction between large and small scales, *Physics of Fluids*, **10**, 1725-1732.

2. Antonov, N.V. and Honkonen, J. (2001). Anomalous scaling in two models of passive scalar advection: Effects of anisotropy and compressibility, *Phys. Rev. E*, **63**, 1-7.

3. Celani, A., Lanotte, A., Mazzino, A. and Vergassola M. (2000). Universality and saturation of intermittency in passive scalar turbulence, *Phys. Rev. Lett*, **84**, 2358-2388.

4. Celani, A., and Vergassola M. (2001). Statistical geometry in scalar turbulence, *Phys. Rev. Lett*, **86**, 424-427.

5. Jacko, J. A., Barreto, A. B., Marmet, G. J., et al. (2000). Low vision: The role of visual acuity in the efficiency of cursor movement. *Paper presented at the Fourth International ACM Conference on Assistive Technologies (ASSETS 2000)*.

6. Jacko, J. A., Barreto, A. B., Scott, et al. (2002). Macular degeneration and visual icon use: deriving guidelines for improved access. *Universal Access in the Information Society*, 1(3), 197-206.

7. Jacko, J. A., Dixon, M. A. et al.(1999). Visual profiles: A critical component of universal access. *Paper presented at the Human Factors in Computing Systems*, Pittsburg, PA.

8. Jacko, J. A., Moloney, et al. (in press). Multimodal feedback as a solution to ocular disease-based user performance decrements in the absence of functional visual loss. *International Journal of Human-Computer Interaction*.

9. Jacko, J. A., Rosa, R. H. et al(2000). Visual impairment: The use of visual profiles in evaluations of icon use in computer-based tasks. *International Journal of Human-Computer Interaction*, 12(1), 151-164.

10. Jacko, J. A., Scott, I. U., Sainfort, F., Barnard, L., Edwards, P. J., Emery, V. K., Kongnakorn, T., Moloney, K. P., & Zorich, B. S. (2003a). Older adults and visual impairment: What do exposure times and accuracy tell us about performance gains associated with multimodal feedback? *CHI Letters*, 5(1), 33-40.

11. Jacko, J. A., Scott, I. U., Sainfort, F., Moloney, K. P., Kongnako, T., Zorich, B. S., & Emery, V. K. (2003b). Effects of multimodal feedback on the performance of older adults with normal and impaired vision. *Lecture Notes in Computer Science (LNCS)*, 2615, 3-22.

12. Kolmogorov, A. N. (1941), Local structure of turbulence in an incompressible fluid for very large Reynolds numbers, *Dokl. Akad. Nauk. SSSR*, *30*, 299–303, 1941.

13. Kolmogorov, A. N. (1962), 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*, 82–85, 1962.

14. Lamperti, J. W. (1962). Semi-satble processes, *Trans. Amer.Math.Soc.* **104**, 62-78.

15. Mallat, S. (1989). A Theory for Multiresolution Signal Decomposition: the Wavelet Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**:674-693.

16. Madndelbrot, B. et al. (1968), Fractional Brownian Motion, Fractional Noise and Applications, *SIAM review*, 10, 422-437.

17. Pumir, A., and Shraiman, B. I. (1995). Persistent small scale anisotropy in homogeneous shear flows, *Physical Review Letters*, **75**, 3114-3117.

18. Ramosa, F.M., Rosaa , R.R., Netoa, C.R. *et al* (2001). Non-extensive statistics and three-dimensional fully developed turbulence *Physica A*, **295**, 250.

19. Schraiman, B. I., and Siggia, E. D. (2000). Scalar Turbulence, *Nature*, *405*, 439-446.

20. She, Z. and Leveque, E. (1994). Universal scaling laws in fully developed turbulence, *Phys. Rev. Lett.,* **72**, 336.

21. Someria, J. (2001). Unweaving the whirls, *Nature*, **413**.

22. Sreenivasan, K. R., and Antonia, R. A. (1997). The phenomenology of small scale turbulence, *Annual Reviews of Fluid Mechanics*, **29**, 435-472.

23. Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.

24. Warhaft, Z. (2000). Passive scalars in turbulent flows, *Annual Reviews of Fluid Mechanics*, **32**, 203-240.

CHAPTER II

PRELIMINARIES: FBM AND WAVELET ANALYSIS FOR SELF-SIMILAR PROCESS

2.1 Fractional Brownian Motion (fBm) and Fractional Gaussian Noise (fGn)

Fractional Brownian motion (fBm) is generalization of Brownian motion (Wienner Process). Brownian motion $B(t)$ is standardly defined as random process satisfying:

1. $B(0) = 0$;

2. for any choice n and $0 \leq t_1 < t_2 < \cdots < t_n$, the increments $B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$ are independent and stationary;

3. $B(t)$ is Gaussian random variable with zero mean and variance t, and

4. $B(t)$ is a continuous function of t, a.s.

It is easy to check that Brownian motion is an 1/2-sssi process, since $B(t) = a^{-1/2}B(at)$ satisfies properties (1)-(4) as well.

Brownian motion is a Gausian process and Gaussian processes are fully determined by their second order properties. Therefore, Brownian motion is the unique Gaussian process having covariance function $\gamma(t, s) = \min\{t, s\}$.

The theorem 1.1.2 gave the covariance structure for an H-sssi process. If such a process is gaussian, it is unique and it is called fractional Brownian motion (fBm).

Definition 2.1.1 *A zero mean gaussian process $B_H(t)$ is called fractional Brownian motion with Hurst exponent H, if $EX(t)X(s) = \frac{E|X(1)|^2}{2}\left[|t|^{2H} + |s|^{2H} - |t - s|^{2H}\right]$, where $E|X(1)|^2 = \frac{\Gamma(2-2H)\cos(\pi H)}{\pi H(1-2H)}$.*

The process $B_H(t)$ is unique, in the sense that class of all fractional Brownian motions with exponent H coincides with the class of all Gaussian H-ss processes. However, a Gaussian process is H-ss with independent increments, if and only if it $H = 1/2$, i.e., if it is a Brownian motion.

The difference process, $Y(n) = B_H(n+1) - B_H(n)$ is called fractional Gaussian noise (fGn). As for more general H-sssi processes, covariance function of fGn is

$$\gamma(h) = \frac{E|X(1)|^2}{2} \left[(h+1)^{2H} - h^{2H} + (h-1)^{2H} \right]. \tag{8}$$

An alternative definition of fractional Brownian motion can be given via stochastic integration. Mandelbrot and Van Ness (1968), Taqqu (1986) defined fBm as the process for which:

(i) $B_H(0) = 0$, and

(ii) $B_H(t) = 1/C_H \cdot \left[\int_{-\infty}^{0} [(t-s)^{H-1/2} - (-s)^{H-1/2}]B(ds) + \int_{0}^{t} (t-s)^{H-1/2} B(ds) \right]$,

where $B(dt)$ is the Wiener measure, and $1/C_H = \Gamma(H + 1/2)/(\Gamma(2H + 1)\sin(\pi H))^{1/2}$.

This representation can be discretized, in the sense that discrete counterpart of $B(dt)$ is normal noise, and as such it provides a way to simulate fBm.

Sample paths of fractional Brownian motion are behaving similarly to those of standard Brownian motion. They are continuous almost surely for all $H \in (0, 1)$ and nowhere differentiable. The fractal (Hausdorff) dimension of sample paths is $D = 2 - H$. That means that for small H (say, $H < 0.5$) the sample paths are quite irregular and *space-filling*. It is interesting that sample paths of fractional Brownian motions are also continuous in H, a result of Peltier and Lévy-Véhel (2000).

There is a Central Limit Theorem (CLT) for fBm. Let $\Delta(i) = B_H(\frac{i+1}{n}) - B_H(\frac{i}{n})$, for $i = 1, \ldots, n-1$. Let $\Delta_{1:n} \leq \ldots \Delta_{n:n}$ be the corresponding order statistics. Define the polynomial $VB_{H,n}(t) = \sum_{i=0}^{[nt]-1} \Delta_{i:n} + (nt - [nt])\Delta_{[nt]:n}$. Phillpe and Thilly (2000) demonstrated that

$$\frac{VB_{H,n}(t)}{n^{1-H}\sqrt{C}} \to L(t), \tag{9}$$

Figure 8: Simulated paths of fractional Brownian motion, (a) $H = 1/4$, (b) $H = 1/2$, and (c) $H = 3/4$.

where $L(t) = -\frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}\Phi^{-1}(t)$, and Φ is the standard Gaussian cdf. This result can be utilized to estimate H.

2.1.1 Simulation of fGn/fBm

There is several methods for simulating fGn/fBm. In fact if fGn is simulated, then a discretized realization of fBm can be obtained as partial sums of fGn, and also, if we have realization if fBm, a fGn can be obtained by differencing. Here we describe the method Based on Circulants since it has overwhelming advantages comparing to other methods and we universally use this method to generate the sample paths of fBm/fGn in this disseration. This battery of this method is based on embedding the covariance matrix Γ of any stationary process in a larger cieculant matrix, C. It is initially proposed by Davies and Harte (1987). See also Wood and Chan (1994). If realization of length N is to be generated, then the size of circulant M should be a power of 2 exceeding $2(N-1)$, i.e., $M = 2^p \geq 2(N-1)$.

Define matrix C as

$$
C = \left\|
\begin{array}{ccccc}
c_0 & c_1 & c_2 & \dots & c_{M-1} \\
c_{M-1} & c_0 & c_1 & \dots & c_{M-2} \\
c_{M-2} & c_{M-1} & c_0 & \dots & c_{M-3} \\
. & . & . & & . \\
. & . & . & & . \\
. & . & . & & . \\
c_1 & c_2 & c_3 & \dots & c_0
\end{array}
\right\|
\tag{10}
$$

where $c_i = \begin{cases} \gamma(i/N) & 0 \leq i < M/2 \\ \gamma((M-i)/N) & M/2 \leq i < M - 1 \end{cases}$

By standard result in matrix theory, the matrix C can be diagonalized by the unitary Fourier transformation matrix, $C = F'\Delta F$, where

$$
F = 1/\sqrt{M} \left\|
\begin{array}{ccccc}
1 & 1 & 1 & \dots & 1 \\
1 & \omega & \omega^2 & \dots & \omega^{M-1} \\
1 & \omega^2 & \omega^4 & \dots & \omega^{2(M-1)} \\
. & . & & . & \\
. & . & & . & \\
. & . & & . & \\
1 & \omega^{M-1} & \omega^{2(M-1)} & \dots & \omega^{(M-1)(M-1)}
\end{array}
\right\|
\tag{11}
$$

where $\omega = e^{-2\pi i/M}$ and Δ is a diagonal matrix whose diagonal is a discrete Fourier transformation of the first column in C (the diagonal is also the set of eigenvalues of C). Now, C can be represented as $F'\Delta^{1/2}F \cdot F'\Delta^{1/2}F$, since F is unitary, and one can generate a Gaussian vector that has $\mathcal{MVN}(0,C)$ distribution via $F'\Delta^{1/2}FZ$, for $Z \sim \mathcal{MVN}_M(0,I)$. Thus, matrix multiplication is replaced by FFT and the calculational complexity of this algorithm is $O(N \log N)$.

The algorithm can be described as follows:

1. Fourier Transformation of $(c_0, \ldots, c_{M-1})$ is the diagonal of $\Delta : \delta_k = \sum_{j=0}^{M-1} c_j \omega^{ij}$. Eigenvalues need to be non-negative. If, at this step, there exist a negative value of δ_k, one should double M, and start again.

2. Simulate FZ in two parts: (i) For $1 \leq j < M/2$, generate standard normal random variables U_j and V_j and define conjugate pairs $W_j = 1/\sqrt{2}(U_j + iV_j)$ and $W_{M-j} = 1/\sqrt{2}(U_j - iV_j)$ and (ii) Fill-in the "gaps" with two standard normal random variables U and V: $W_0 = U$ and $W_{M/2} = V$.

3. Find $Y(k/N) = 1/\sqrt{M} \sum_{j=0}^{M-1} \sqrt{\delta_j} W_j \omega^{jk}$, using FFT again.

2.2 Wavelets and Self-Similar Processes

Wavelets are capable tools in detecting self-similarity in the signals. In this section we discuss some properties of self-similar signals in the wavelet domain. Some important pioneering work in this area was done by Flandrin and his collaborators [Flandrin 1989a,b, 1992a,b; Flandrin and Gonçlavès, 1993; Abry, Gonçlavès and Flandrin 1993]. The body of recent literature is quite large.

2.2.1 Wavelets and Stationary Processes

Wavelets and stationary processes are well researched area. Highlights of the research include several topics: preservation of stationarity, whitening property of wavelets and assessing correlations in the wavelet domain, wavelet shrinkage in the presence of stationary noise, to list a few.

Let $X(t)$ be a second order process with autocorrelation function $\gamma(t, s)$. The discrete wavelet transformation of $X(t)$ is a discrete random field

$$\{d_{jk}, \ j, k \in \mathbb{Z}\} = \left\{ \int_{\mathbb{R}} X(t) \psi_{jk}(t) \, dt, \ j, k \in \mathbb{Z} \right\}, \tag{12}$$

which is well defined if the path integrals in (12) are defined and

$$\int_{\mathbb{R}} \sqrt{\gamma(t,t)} \, |\psi_{jk}(t)| \, dt < \infty. \tag{13}$$

Thus, when (13) is satisfied

$$Ed_{jk}d_{j'k'} = \iint_{\mathbb{R}^2} \gamma(t,s)\psi_{jk}(t)\psi_{j'k'}(s) \, dt \, ds. \tag{14}$$

If the process $X(t)$ is stationary, then (14) becomes

$$Ed_{jk}d_{j'k'} = \iint_{\mathbb{R}^2} \gamma(h)\psi_{jk}(s)\psi_{j'k'}(s+h) \, ds \, dh, \tag{15}$$

or in the Fourier domain,

$$\begin{aligned} Ed_{jk}d_{j'k'} &= \frac{1}{2\pi} \int_{\mathbb{R}} f(\omega)\Psi\left(\frac{\omega}{2^j}\right) \overline{\Psi\left(\frac{\omega}{2^{j'}}\right)} \\ &\quad \cdot e^{-i\omega k2^{-j}} e^{i\omega k'2^{-j'}} 2^{-j/2} 2^{-j'/2} \, d\omega, \end{aligned} \tag{16}$$

where Ψ and f are the Fourier transformations of ψ and γ, respectively.

Relations (15) and (24) are critical in assessing the whitening property of wavelet transformations for in case of stationary processes as well as the second order properties of wavelet coefficients of self-similar processes.

2.2.1.1 Approximation and Production of Stationarity

For a stationary process $X(t)$, $t \in \mathbb{R}$ and its wavelet approximation $X_m(t) = \sum_k c_{mk}\phi_{mk}(t)$, holds

$$E|X_m(t) - X(t)|^2 \to 0 \text{ when } m \to \infty, \tag{17}$$

i.e., degraded process $X_m(t)$ goes in the mean-square to the original.

Indeed, for r-regular wavelet ϕ and a wavelet-based reproducing kernel of V_m, $\mathbb{K}_m(t,s) = 2^m \sum_k \phi(2^m t - k)\phi(2^m s - k)$,

$$
\begin{aligned}
E|X(t) - X_m(t)|^2 &= EX(t)^2 - 2EX(t)X_m(t) + EX_m(t)^2 \\
&= \gamma(0) - 2\int \gamma(t-s)\mathbb{K}_m(t,s)\,ds \\
&\quad + \iint \gamma(u-s)\mathbb{K}_m(s,t)\mathbb{K}_m(t,u)\,du\,ds
\end{aligned}
$$

and

$$
\int \gamma(u-s)\mathbb{K}_m(t,s)\,ds \to \gamma(u-t),
$$

uniformly on bounded sets, implying (17). In general case $X_m(t)$ is not projection of $X(t)$ on V_m since the sample paths may not be $\mathbb{L}_2$ integrable.

It is well known that wavelet transformations of stationary processes and sequences yield level-wise stationary sequences of coefficients.

More general, wavelet transformation of a process with stationary increments yields a stationary sequences of wavelet coefficients in any fixed level j. The following lemma is straightforward.

Lemma 2.2.1 *Let $X(t)$, $t \in \mathbb{R}$ be a process with stationary (in strong sense) increments, i.e., for all finite-dimensional vectors $(X(t+h_1) - X(t), \ldots, X(t+h_k) - X(t)) \overset{d}{=} (X(h_1) - X(0), \ldots, X(h_k) - X(0))$, for all $t \in \mathbb{R}$. Then for a fixed j, $d_{j,k}$ is a stationary sequence.*

Proof: Select arbitrary $m \in \mathbb{Z}$ and fix j. Then

$$
\begin{aligned}
d_{j,k} &= \int_{\mathbb{R}} X(t)\psi_{jk}(t)dt \\
&= \int_{\mathbb{R}} X(t - 2^{-j}m)\psi_{jk}(t - 2^{-j}m)dt \\
&= \int_{\mathbb{R}} X(t - 2^{-j}m) \cdot 2^{j/2}\psi(2^j(t - 2^{-j}m) - k)dt \quad\quad\quad (18) \\
&= \int_{\mathbb{R}} \left(X(t - 2^{-j}m) - X(2^{-j}m)\right)\psi_{j,k+m}(t)dt \quad [\text{ because } \int \psi_{jk} = 0] \\
&\overset{d}{=} \int_{\mathbb{R}} \left(X(t) - X(0)\right)\psi_{j,k+m}(t)dt \quad [\text{ because increments are stationary}] \\
&= d_{j,k+m}.
\end{aligned}
$$

Standard arguments probabilistic arguments involving characteristic functions of finite linear combinations of d's [see Abry, Flandrin, Taqqu, and Veitch, 2000] lead to conclusion that finite-dimensional distributions are free of m, implying that the sequence $\{d_{j,k}, k \in \mathbb{Z}\}$ is stationary.

2.2.2 KL Expansions and Whitening of Stationary Processes

The represenation

$$X(t) = \sum_{j,k} d_{jk} \psi_{jk}(t),$$

with

$$d_{jk} = \int_{\mathbb{R}} X(t) \psi_{jk}(t) \, dt. \tag{19}$$

would remind the Karhunen-Loève (KL) representation if the d_{jk} are uncorrelated. It turns out that wavelet transformations are indeed decorrelating and because of that, wavelets are sometimes called *approximate KL expansions.*

The KL expansions convert continuous-time random processes to sequences of uncorrelated random variables. More formally, if $X(t), t \in \mathbb{R}$ is a real, mean-square continuous, zero-mean process, then there exists an orthonormal system of nonrandom functions $\phi_1(t), \phi_2(t), \ldots$, and a set of uncorrelated random variables $Z_1, Z_2, \ldots$, such that

$$X(t) = \sum_i Z_i \phi_i(t),$$

where $Z_i = \int_{\mathbb{R}} X(t) \phi_i(t) \, dt$. The functions $\phi_i(t)$ are eigenfunctions of an associated integral operator

$$(\Gamma f)(t) = \int_{\mathbb{R}} \gamma(t, s) f(s) \, ds, \tag{20}$$

where $\gamma(t, s) = EX(t)X(s)$ is the autocovariance function. The K-L transformation is "energy packing." The first k coefficients, corresponding to the k largest eigenvalues of

26

the operator (20) minimize the MSE among all orthogonal (unitary in the complex case) transformations.

Even though the K-L is an elegant theory, it has limited applicability. As a rule, it is difficult to obtain exact solutions of the eigenvalue problem. The transformation is data-dependent and approximate solutions are computationally intensive. For a discussion see Wornell (1992).

For a given (wide-sense) stationary process, one can construct a wavelet-like biorthogonal system so that the associated coefficients are uncorrelated, a feature similar to K-L expansions.

The informal statement "Wavelets whiten data" can be well formalized if the data on input is a stationary process.

2.2.3 Wavelet Analysis of Self-Similar Processes

Let $\{X(t), t \in \mathbb{R}\}$ be a H-ss process. Then for a fixed level j,

$$d_{jk} \stackrel{d}{=} 2^{-j(H+1/2)} d_{0,k}. \tag{21}$$

Indeed,

$$
\begin{aligned}
d_{jk} &= \int X(t) 2^{j/2} \psi(2^j t - k) dt \\
&= \int X(2^{-j}u) 2^{j/2} \psi(u - k) d(2^{-j}u) \\
&\stackrel{d}{=} 2^{-jH-j/2} \int X(u) \psi(u - k) du \\
&= 2^{-j(H+1/2)} d_{0,k}.
\end{aligned}
$$

Note that $1/2$ in the exponent is because of $\mathbb{L}_2$ normalizing of wavelets, for $\mathbb{L}_1$ normalized wavelets $\psi_{j,k} = 2^j \psi(2^j - k)$ the scaling is $(2^{-j})^H$.

Let $X(t)$ be a H-sssi process. Then $d_{0,k}, k \in \mathbb{Z}$ is stationary sequence and for any $k, \mathbb{E}d_{0k} = 0$ and $\mathbb{E}d_{0k}^2 = \mathbb{E}d_{00}^2$. Then,

$$\mathbb{E}d_{jk}^2 = C \, 2^{-j(2H+1)}, \tag{22}$$

27

with $C = \mathbb{E}d_{00}^2$. The re-expression of equation (22) as

$$\log_2 \mathbb{E}d_{jk}^2 = -(2H + 1) \cdot j + C',\tag{23}$$

is a basis for wavelet based estimation of H, as we will see later.

If the process is LRD, i.e., its spectra is behaving as $|\omega|^{-\alpha}$ at 0, then from (24)

$$\begin{aligned}
\mathbb{E}d_{jk}^2 &= \frac{2^{-j}}{2\pi} \int f(\omega)|\Psi(2^{-j}\omega)|^2 d\omega \\
&\sim c \cdot 2^{-\alpha j} \int \frac{|\Psi(\omega)|^2}{|\omega|^\alpha} d\omega,
\end{aligned}$$

This relation is a basis for estimating α (or $H = \frac{1+\alpha}{2}$) by taking logarithms of both sides.

If in (24), $j = j'$, we can explore covariance structure of wavelet coefficients belonging to a single level and separated by distance of $|k - k'|$.

$$\mathbb{E}d_{jk}d_{j'k'} = \frac{1}{2\pi} \int_{\mathbb{R}} f(\omega) \left| \Psi\left(\frac{\omega}{2^j}\right) \right|^2 \cdot e^{-i\omega(k-k')2^{-j}} d\omega,\tag{24}$$

The correlation $\mathbb{E}d_{jk}d_{j'k'}$ is thus a function of the difference $(k - k')$, and the asymptotic behavior of integral in (24) when $|k - k'| \to \infty$ is influenced by behavior of $f(\omega)\left|\Psi\left(\frac{\omega}{2^j}\right)\right|^2$ as $\omega \to 0$.

As we saw before, if the wavelet has exactly N vanishing moments, then its Fourier transformation is differentiable N times at the origin, and for $0 \leq i \leq N - 1$, $\Psi^{(i)}(0) = 0$, and $\Psi^{(N)}(0) \neq 0$. By Taylor series argument, the behavior of $|\Psi(\omega)|$ matches that of $|\omega|^N \cdot |\Psi^{(N)}(0)|$ at origin. On the other end, the spectra of LRD process is singular at zero and behaves as $|\omega|^{-\alpha}$. Thus, when the decomposing wavelet has N vanishing moments, singularity at 0 of spectra is compensated; the behavior at $\omega \sim 0$ of $f(\omega)\left|\Psi\left(\frac{\omega}{2^j}\right)\right|^2$ is $|\omega|^{2N-\alpha}$.

Two comments. (i) If $N > \alpha/2$ then sequence of wavelet coefficients is not LRD, even-though the original signal is, and (ii) looked as Fourier pair, the autocorrelations $\mathbb{E}d_{jk}d_{j'k'}$ behave as $|k - k'|^{\alpha-N-1}$ when $|k - k'| \to \infty$.

This means, the LRD processes are better decorrelated with wavelets possessing more vanishing moments.

28

2.2.4 LogScale Diagrams or Scalograms

The totality of squared wavelet coefficients represents the energy content of the zero-mean signal and expected levelwise energies form wavelet counterpart of Fourier spectra. The following definitions are as in Abry, Flandrin, Taqqu, and Veitch (1998, 2000).

Definition 2.2.1 *Theoretical Wavelet Spectra of a process $X(t)$ with stationary increments is the sequence*

$$e(j) = \mathbb{E}(d_{j,\bullet}), \tag{25}$$

where because of stationarity, $d_{j,\bullet}$ stands for an arbitrary coefficient from the level j. The plot of $(j, e(j))$ is referred as Theoretical Logscale Diagram.

Let $E(j)$ be an estimator of $e(j)$. Then, the plot of (j, E(j)) is called simply Logscale Diagram or Scalogram.

Because of linearity of wavelet transformation, the wavelet spectra gives complete second-order description of the random process.

The logscale diagram was originally proposed and utilized as a tool for detecting and estimating periodicities, since, as in the case of Fourier counterparts, energy is large at dominant scales. In our context, logscale diagram is used to estimate the scaling exponent of a signal.

Definition 2.2.2 *Let, for some $q \in \mathbb{R}$*

$$e_q(j) = \mathbb{E}|d_{j,\bullet}|^q, \tag{26}$$

The plot of $(j, e(j))$ is referred as q-th Order Theoretical Logscale Diagram.

Let $E_q(j)$ be an estimator of $e_q(j)$. Then, the plot of $(j, E_q(j))$ is called q-th Order Logscale Diagram or Scalogram.

2.2.5 Detection and Assessment of Scaling

By observing the data at time domain it is impossible to detect scaling without resorting to such tools such as, for example, Structure Functions, Spectrograms, Logscale Diagrams, Structure Functions, q-th order Logscale Diagrams, etc.

If, for example, in Fourier Log-Spectrograms or Logscale Diagrams it is possible to align a straight line with particular slope of $-\alpha$, over duration of several decades (octaves, "binary-decades") then the scaling in the data is present.

The task is to estimate the range of scaling, in frequences for Fourier tools or in scales for wavelet-based tools, as well as the slope and its variation.

The key tool is of course the linear regression analysis for which the goodness of fit is assessed by standard regression measure, such as R^2. Plethora of tests are available here including nonparametric goodness of fit tests (χ^2, Kolmogorov-Smirnov), bootstrap, empirical MSE, etc.

It is possible also to automatically select the range of scales $[j_1, j_2]$ for which the goodness of fit measure is minimized. The selection of this range is important since in many situations estimation of the slope is non-robust to change in j_1 and/or j_2. This robustness is influenced by several factors, including quality of data, a slope close to zero, presence of a periodicity, or injection of energy at a particular scale.

Especially critical is the selection of low scale j_1. The high variability of spectra at low scales is influenced by several factors – some of which have nothing to do with the nature of data. For example, in the Logscale diagrams points at low scales are obtained by averaging substantially less empirical values of energy (squared wavelet coefficients). The difference in the number of averaged values is huge, for example if the scale $j_2 = 10$ averages 1024 energies, the scale $j_1 = 3$ averages only 8 values.

By the assessment of scaling we consider two tasks: identification of the phenomenon (model selection) and evaluation of the scaling slope.

As in any model selection task, it is important to have some prior information about the

data – because that may be decisive in the model selection and subsequent inference.

If the observed scaling α is in $(0, 1)$ [slope of $-\alpha$], and we believe (or confirm) that data are stationary – then the estimated slope α corresponds to the LRD parameter α and the link with Hurst exponent is $H = \frac{1-\alpha}{2}$.

If , on the other hand, the estimated slopes are greater than one, and the data may have been obtained as cumulative process and/or are not stationary (stock market prices, turbulence measurements, seizure data) then they correspond to an stationary increments or even stationary k-th increments, $k > 1$. Another case is, as in some synthetic dynamic systems or ON-OFF processes, that the data may be fractal.

In this cases $H = \frac{\alpha-1}{2}$ and in case of fractal, H coincides with the regularity (in Hölder sense) of the non-differentiable sample paths. For instance, the turbulence signals give slope of $-5/3$, defining $H = 1/3$.

2.2.6   Estimation of the Scaling Law

We discuss methods for estimating the scaling exponent α or equivalently corresponding Hurst exponent, H. We limit our discussions in the wavelet-based methods.

The wavelet method has some connections to the variance plot. Recalling the variance expression when studying aggregated series X^j over dyadic blocks of size 2^j , $j = 0, 1, 2, ...,$

$$Var[X^j] \sim 2^{j(2H-2)}, j = 0, 1, 2, ... \tag{27}$$

This estimator could be very biased (not very reliable). In the wavelet framework we study differences of aggregated series. In the most simplest case we compute the difference between to points in non-overlapping blocks of size 2 (the Haar wavelet).

Let Y^{j+1} be the series made by difference Y^j (Y^0 is the data series at highest time resolution).

$$Y_k^{j+1} = (Y_{2k}^j - Y_{2k-1}^j)\frac{1}{\sqrt{2}}, \quad k = 1, 2, ..., \frac{N}{2^j} \text{and} j = 0, 1, 2... \tag{28}$$

Then the variance of Y^j decay according to a similar power-law as above.

$$Var[Y^j] \sim 2^{j(2H-1)}, j = 0, 1, 2, ... \tag{29}$$

.

It turns out that the variance is equal to the second order moment, i.e. the expectation of Y is zero. In the frequency domain, $\mathbb{E}[(Y^j)2]$ corresponds to the signal energy in a frequency band depending on j.

2.3 References

1. Abry, P. and Flandrin, P. (1994). On the initialization of the discrete wavelet transform algorithm. *IEEE Signal Processing Letters,* **1**, 32–34.

2. Abry, P. and Flandrin, P. and Taqqu, M. S. and Veitch, D. (1998). Wavelets for the analysis, estimation and synthesis of scaling data, In *Self-similar Network Traffic and Performance Analysis,* K. park and W. Willinger, eds., Wiley Interscience.

3. Abry, P. and Gonçalvès, P. and Flandrin, P. (1993). Wavelet-based spectral analysis of $1/f$ processes Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, pp. 237-240.

4. Abry, P. and Gonçalvès, P., and P. Flandrin, P. (1995). Wavelets, spectrum analysis and 1/f processes Wavelets and Statistics, Lecture Notes in Statistics, Vol. 103, pp. 15-29, Springer-Verlag, NY.

5. Abry, P. and Sellan, F. (1996). The wavelet-based synthesis for fractional Brownian motion - Proposed by F. Sellan and Y. Meyer: Remarks and fast implementation. *Applied and Computational Harmonic Analysis,* **3(4)** 377-383.

6. Abry, P., Veitch, D., and Flandrin, P. (1998). Long Range Dependence: Revisiting Aggregation with Wavelets. *Journal of Time Series Analysis,* **19**, 253-266.

7. Abry, P. and Veitch, D. (1998). Wavelet analysis of long-range-dependent traffic *IEEE Transactions on Information Theory,* **44(1)**, pp. 2-15.

8. Bacry, E., Arneodo, A., Muzy, J-F., and Graves, P-V. (1994). Wavelet Analysis of DNA Sequences, *SPIE* , vol. 2569 489-498.

9. Barnard, G. A. (1956). Discussion on Hurst, *Proc. Inst. Civ. Eng.,* **5**, 552–553.

10. Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics, *Journal of Econometrics,* **73**, 5-59.

11. Benedetto, J. J. and , D., (1995). Wavelet analysis of spectrogram seizure chirps, Proc. *SPIE,* 2569, 512-521.

12. Beran, J. (1994). *Statistics for Long-Memory Processes,* Chapman and Hall, London.

13. Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society,***57,** 659-672.

14. Bhattacharya, R.N., Gupta, V. K., and Waymire, E. (1983), The Hurst effect under trends, *Journal of Applied Probability,* **20**, 649-662.

15. Cox, D. R. (1984). Long-range dependence: A review, in: Statistics: An Appraisal,eds.H.A.Davidand H.T. David (Iowa State University Press, 1984) pp. 55-74.

16. Davies, R.B. and Harte, D.S. (1987). Test for Hurst effect. *Biometrika* **74**, 95-101.

17. Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). *Biological sequence analysis, probabilistic models for proteins and nucleic acids*, Cambridge U. Press.

18. Feller, W. (1951). The asymptotic distribution of the range of sums of independent random variables. *Ann. Math. Statsit.* **22**, 427-432.

19. Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Trans. on Information Theory,***38(2)**.

20. Geweke, J. F., and Porter-Hudak, S. (1983). The estimation and application of long memory time series models, *Journal of Time Series Analysis*, **4(4)**, 221-238.

21. Granger, C. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics,* **14,** 227-238.

22. Granger, C.W.J. and Joyeux, R.,(1980). An introduction to long-memory time series, *Journal of Time Series Analysis,* **1**, 15-30.

23. Gonçalvès, P and Abry, P. (1997). Multiple-window wavelet transform and local scaling exponent estimation, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, **4**, 3433-3436.

24. Haslett, J. and Raftery, A. E. (1989). Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource (with Discussion). *Applied Statistics*, **38**, 1-50.

25. Hosking, J. R. M. (1981). Fractional differencing, *Biometrika,* **68,** 165-176.

26. Hausdorff J.M. *et al.* (1996). Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *J. Appl. Physiol.* **80**, 1448-1457.

27. Hurst, H.E., (1951), Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineering,* **116**, 770-799.

28. Leland, W. E., Taqqu, M. S., Willinger, W., and Wilson, D. V. (1994). On the self-similar nature of Ethernet traffic (Extended version), *IEEE/ACM Transactions on Networking,* **2** 1-15.

29. Mandelbrot, B. B., (1975). Limit theorems of the self-normalized range for weakly and strongly dependent processes.*Z. Whar. verw. Geb.,* **31,** 271-285.

30. Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian Motions, *Fractional Noises and Applications, SIAM Review,* **10** , 422-437.

31. Mandelbrot, B., and Wallis, J., (1968). Noah, Joseph and operational hydrology, *Water Resources Research,* **4,** 909-918.

32. Marshall, B., McEver, R. and Zhu, C. (2001). Kinetic rates and their force dependence on the P-Selectin/PSGL-1 interaction measured by atomic force microscopy. Proceedings of ASME 2001, Bioengineering Conference, BED - Vol. 50.

33. Peltier, R. F. and Lévy Véhel, J. (1994). A New Method for Estimating the Parameter of Fractional Brownian Motion, *Rapport de Recherche n. 2396, INRIA, Le Chesnay Cedex.*

34. Peltier, R. F. and Lévy Véhel, J. (1995). MultiFractional Brownian Motion: Definition and Preliminary Results, *Rapport de Recherche n. 2645, INRIA, Le Chesnay Cedex.*

35. C-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simon, and H.E. Stanley (1992). Long-range Correlations in Nucleotide Sequences, *Nature,* **356,** 168-170.

36. Philippe, A. and Thilly, E. (2000). Identification of self-similar Gaussian process by using convex rearrangement, *Pub IRMA. LILLE,* **52**.

37. Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics,* **53**, 165-188.

38. Taqqu, M. S. (1986). A Bibliographical Guide to Self-Similar Processes and Long-Range Dependence, in Eberlein, E. and Taqqu, M. S., eds., (1986) *Dependence in Probability and Statistics*, Birkhauser, Boston, 137-162.

39. Taqqu, M. S., Teverovsky, V. and Willinger, W. (1995). Estimators for Long-Range Dependence: An Empirical Study, *Fractals,* **3**, 785-798.

40. Wood, A.T.A. and Chan, G. (1994). Simulation of Stationary Gaussian Processes in $[0,1]^d$. *Journal of Computational and Graphical Statistics,* **3,** 409-432.

CHAPTER III

WHAT CAUSES SELF-SIMILARITY?

The sources of self-similarity in the observations are discussed in a general sense. We survey six types of distinct mechanisms, which may be regarded as the sources of self-similarity. These mechanisms are not subject-specific and they can be used in different application contexts, though it might be originally proposed in a certain discipline.

3.1 Introduction

Statistical tools have been extensively used in the analysis of self-similar data arising in many application fields. Many efforts in the past have been dedicated in estimation of parameters (*e.g.* Hurst exponent) which describes the structure of the process. However, there are few explorations about the sources of the self-similarity in the data. The study of what causes self-similarity in the data is important because of three reasons. (1) It provides convincing justifications of modeling the data as self-similar process more than the pure statistical analysis; (2) The better knowledge about the source of self-similarity, the more accurate we can fit the data into an appropriate model and according the more physical explanation we can give about the estimated parameters; (3) The understanding of the source makes it easy to analyze the data and may provide more meaningful insights to the people in the application area.

In this chapter, we are going to survey the possible sources of self-similarities. The survey is set to in a general sense and should be shared among the people with different application interests. We provide six distinct mechanisms to address the possible sources of self-similarities.

3.2 Aggregation

3.2.1 AR Processes

Granger (1980) shows that the aggregation of many stationary processes would potentially exhibit self-similarity. Suppose that there are N stationary Autoregressive processes of order 1 (AR(1)), denoted by

$$X_t^{(m)} = \phi^{(m)} X_{t-1}^{(m)} + \varepsilon_t^{(m)}, \quad \text{for } t = 0, 1, 2, \cdots \quad \text{and} \quad m = 1, 2, \cdots, N. \quad (30)$$

If parameters $\phi^{(m)}, m = 1, 2, \cdots, N$ are i.i.d from Beta(p, q) distribution, which has the pdf

$$f(\phi) = \frac{2}{B(p,q)} \phi^{2p-1} (1 - \phi^2)^{q-1}, \quad \text{for } 0 \leq \phi \leq 1, \quad (31)$$

the aggregate process $X_t = \sum_{m=1}^{N} X_t^{(m)}$ exhibits the self-similarity with Hurst exponent $H = 1.5 - q/2$ as $N \to +\infty$.

To illustrate this aggregate process, we provide a simulation examples in Fig. 9. We generate 100 sample paths of the AR(1)process.

Following the idea of Granger, Oppenheim and Viano (2001) extended the analysis to the case of AR(p), with $p > 1$, processes and proved that the aggregation process will converge to the self-similar process with seasonal components under certain assumptions on the parameters.

Recently, Zaffaroni (2004) also studied the case for AR processes driven by both disturbances for each individual process and common disturbances for all processes. The result shows that the distribution assumption is not necessary to be Beta, instead it can be anything with a semi-parametric form, i.e.

$$f(\phi; b) \sim c_b (1 - \phi)^b, \quad \text{for } 0 \leq \phi \leq 1, \quad (32)$$

where c_b is a normalizing constant.

3.2.2 Poisson Shot Noise

The shot noise model, originally introduced by Rice (1944) and Parzen (1962), is extensively used in the areas of computer engineering (Lewis, 1964), geophysics (Vere-Jones,

Figure 9: Aggregation Example1

1970), acoustics (Kuno and Ikegaya, 1973), risk theory (Klüppelberg and Mikosch, 1995) and finance (Samorodnitsky 1995). This model is characterized by the superposing of "shot effects" which initiate at random times and persist through random durations, possibly infinite.

A Poison shot noise process, $S(t, t \geq 0)$, is defined as

$$S(t) = \sum_{i=1}^{N(t)} X_i(t - T_i) + \sum_{i=-1}^{-\infty} [X_i(t - T_i) - X_i(-T_i)], \ \ t \geq 0 \tag{33}$$

where $X_i(t)$, $i \in \{\cdots, -3, -2, -1, 1, 2, 3, \cdots, N(t)\}$ are i.i.d stochastic processes with $X_i(t) = 0$ for $t < 0$, $N(t)$ is the two-sided Poisson process with rate $\alpha > 0$, and T_i's are points of Poisson process $N(t)$.

Klüppelberg and Kühn (2004) show that Poisson shot noise model weakly converge to limit of fractional Brownian Motion, which has been known as a famous example of self-similar processes.

3.2.3 Renewal Process

As known to most of us, the renewal process N_t is a counting process defined as

$$N_t = \sum_{n=1}^{\infty} I(S_n > t) \ \ \text{and} \ \ S_n = \sum_{i=1}^{n} T_i, \tag{34}$$

where T_i's are the i.i.d non-negative random variables (*i.e.* renewal interval).

Taqqu and Levy (1986) show that the superposition of a large number of identical renewal processes will produce self-similarity in the aggregate process, though the inter-renewal-time (T_i) distribution has to be heavy-tailed, i.e., it has infinite variance.

3.3 Random Walks

3.3.1 Correlated Random Walks

A correlated random walk (CRW) model is a discrete processes such that the law of each move is ruled by the value of the previous move. We note that the decay of correlation for such processes is exponential, but that a mixture of these processes may lead to walks

exhibiting self-similarity. Taqqu (1975) show that the superposition of a large number of CRW processes yields a discrete Gaussian process with a long-range dependent memory structure, so that its scaling limit is the fractional Brownian motion.

3.3.2 Random Walks in Random Sceneries

A random walk in random sceneries (RWRS) is a stochastic process $\{K_n\}_{n \geq 1}$ with definition

$$K_n := \sum_{k=1}^{n} \delta([S_k]), \tag{35}$$

where S_k is a standard random walk, $[S_k]$ is an integer such that $S_k \leq [S_k] \leq S_k + 1$, and $\delta(x)$'s are i.i.d random variables with zero mean and unit variance. Kesten and Spitzer (1979) and Bolthausen (1989) proved that $\{K_n\}$ can be rescaled so as to converge in law to a nondegenerate process. A consequence of these results is that the proper normalization is $\frac{K_n}{n^{3/4}}$ if K_n is defined in the real line space. This is equivalent to say that it converge to self-similar process with Hurst exponent $H = 3/4$. Kesten and Spitzer (1979) also show that K_n may converge to the self-similar process with Hurst exponent $H > 3/4$ if the unit-variance is dropped and random steps $(S_k - S_{k-1})$ and δ's are i.i.d random variables from α-stable distribution with indices (not necessary to be the same for $(S_k - S_{k-1})$ and δ) $0 < \alpha \leq 1$. Piau (2002) showed that, even if good integrability properties hold, the convergence to the self-similar process with Hurst exponent $H > 3/4$ can still be obtained assuming that the transition probability in the random walk S_k has a suitable nonparametric form. This result is extended in Piau (2004) to achieve the convergence to the self-similar process with Hurst exponent $1/2 \leq H < 3/4$ through an further modification on the nonparametric form of the transition probability in the random walk S_k. Though there is a natural extension of Kesten and Spitzer (1979) results to obtain convergence to the self-similar process with Hurst exponent $H < 1/2$ (anti-persistent process), no published work reports the convergence to the anti-persistent process without the α-stable assumption. Employing such a model, Cox (1984) explained the self-similarity in a stochastic process of the coefficients of mass per

unit length of worsted yarn (two hundred sections of the yarn are randomly chosen with lengths ranging from 1cm to 100m).

3.4   Nonlinear Dynamics

A classical example of self-similarity caused by nonlinear dynamics is turbulence.

Kolmogorov (1941) developed his theory, often referred to as **K41** theory, for *locally isotropic* turbulence. Let $\mathbf{x} = (x_1, x_2, x_3)$ be the position vector and $\mathbf{u} = (u_1(\mathbf{x}), u_2(\mathbf{x}), u_3(\mathbf{x}))$ be the velocity components. Locally isotropic turbulence describes the flow in which the probability distribution of the relative velocity differences

$$\Delta\mathbf{u}(\mathbf{r}) = \mathbf{u}(\mathbf{x} + \mathbf{r}) - u(\mathbf{x}), \tag{36}$$

is independent of time, and invariant under translations, reflections, and rotations. The fundamental objects in **K41** theory are *structure functions.* Structure functions are closely related to correlations of two-point velocity differences,

$$\langle \Delta\mathbf{u}(\mathbf{r})^2 \rangle = 2\sigma_u^2(1 - \rho_u(r)). \tag{37}$$

A (longitudinal) structure function of order p is defined as

$$D_p(r) = \langle ||\Delta\mathbf{u}(\mathbf{r})||^p \rangle \tag{38}$$

where the angular brackets denote time averaging.

A functional description for the moments of velocity differences and thus for the structure functions can be derived using *dimensional analysis* and leads to

$$D_p(r) = C_p[\langle \epsilon \rangle \, r]^{\frac{p}{3}}, \tag{39}$$

where C_p is a universal constant. For the third-order structure function, it can be inferred directly from the Navier-Stokes equations that $C_3 = -\frac{4}{5}$. From (39) it follows that structure functions possess scaling behavior. Let the symbol $\propto$ denote "proportional to." Then,

$$D_p(r) \propto r^{\zeta_p}. \tag{40}$$

40

The exponent ζ_p is called the *scaling exponent.* The **K41** theory gives the simple model $\zeta_p = \frac{p}{3}$.

Similarly, as for the structure functions, a description of the energy of the turbulent fluctuations per unit of mass of fluid in scales r can be derived from the hypotheses and by dimensional analysis,

$$E_r \propto (r)^{\frac{2}{3}}. \tag{41}$$

Via the Fourier transform of E_r, which results in the spectral density $\phi(k)$, the celebrated "$-\frac{5}{3}$ law" for the power spectrum is obtained,

$$\begin{aligned} E_k &= 2R^{-1}k^2\phi(k) \\ &\propto k^{-\frac{5}{3}}. \end{aligned} \tag{42}$$

3.5 *Critical Phenomena*

3.5.1 Ising Model

A notable example is the Ising model, which has been used in describing the thermodynamic systems. Ising tried to explain certain empirically observed facts about ferromagnetic materials using a model of proposed by Lenz (1920). It was referred to in Heisenbergs (1928) paper which used the exchange mechanism to describe ferromagnetism. The Ising Model considers an idealized system of interacting particles, arranged onto a regular planar grid. Each particle can have one of two magnetic spin orientations, generally labelled up (+1) and down (-1). Each particle interacts only with its nearest neighbors; the contribution of each particle to the total energy of the system depends upon the orientation of its spin compared to its neighbors. Adjacent particles that have the same spin (-1;1) or (1; 1) are in a lower energy state that those with antithetic spins (1; -1) or (-1; 1). Given the spin orientations of all particles in the system, one may compute the total energy. If the variable x_i denotes the spin of particle i, then the total energy of the system (Hamiltonian) is

$$M_\Lambda = -J \sum_{i \sim j \in \Lambda} x_i x_j, \tag{43}$$

41

where $i \sim j$ denotes the a pair of neighbor indices defined in a large d-dimensional lattice $\Lambda \subset Z^d$ and the constant J represents the strength of the interaction between x_i and x_j. Statistical mechanics states that the probability of any particular configuration c in the configuration space $\mathcal{C}$ is proportional to

$$\exp\left\{-\frac{M_\Lambda(c)}{kT}\right\}, \tag{44}$$

where the parameter T is temperature in $^\circ K$ and the value k is the Boltzmanns constant, a fundamental constant in physics. From (44) it follows that low energy configurations are more probable, but the influence of energy is more pronounced at low temperatures. At low temperatures, the energy of a configuration is very important in determining its likelihood, and so the most likely states are those with lowest energy. At high temperatures, energy is less important, and so the states with high entropy are not unlikely, in fact there are far more states that are disordered than ordered.

Under normal condition, the normalized variable

$$M_\Lambda^* = \frac{M_\Lambda - E[M_\Lambda]}{|\Lambda|^{\frac{1}{2}}}, \tag{45}$$

converges to a non-degenerate random variable, where $|\Lambda|$ is equal to the number of elements in Λ. This implies that S_i is at most weakly dependent. However, there is always a critical temperature such that the normalized variable

$$M_\Lambda^* = \frac{M_\Lambda - E[M_\Lambda]}{|\Lambda|^\alpha}, \quad \text{for } 1/2 < \alpha < 1, \tag{46}$$

converges to a non-degenerate random variable, which is the result of S_i being a self-similar process with Hurst exponent α. This model with self-similarity is actually used in many natural process. For example, this is actually a model for the transition from a liquid phase to a gaseous phase.

3.6 Randomly Killed Processes

Reed and Hughes (2002) show that if stochastic processes with exponential growth in expectation are killed or observed randomly, the killed or observed states exhibiting self-similarity. Suppose the expectation of stochastic process X_t has the form

$$E(X_t|X_0) = X_0 \mu^{\psi(t)} \tag{47}$$

where μ is a real number and $\psi(t)$ and real-valued function of t. Assuming the probability of killing the states is

$$P(\text{killed at time} \geq t) = e^{-\nu t}. \tag{48}$$

where the ν is called hazard rate. In the case of X_t being a Geometric Brownian Motion (GBM), the killed process has self-similarity with Hurst index approximately

$$H = \frac{1}{1-s}, \tag{49}$$

where s is a root of the equation $\frac{1}{2}\sigma^2 s^2 + (\mu - \frac{1}{2}\sigma^2)s - \nu = 0$ and μ and σ are the drift and volatility parameters of the GBM.

Reed and Hughes (2004) proposed a simple birth-and-death process model to explain the self-similarity in the size of gene and protein family. Though this is derived in a genomic setting, it could be easily extended to other scenarios. In this model, new genes in a family arise from mutations of existing genes independently and at random at a fixed probability rate, individual genes in a family can be eliminated independently and at random at a fixed probability rate, and new families arise from the random splitting of existing families at a fixed probability rate. They showed that the states of a homogeneous birth-and-death process after exponentially distributed times forms a self-similar process with Hurst index as a function of the birth rate (λ), death (μ), splitting rates (ρ), which is $\frac{1}{1+\rho/(\lambda-\mu)}$.

3.7 Differential Equation

3.7.1 Supra Diffusion

It is well known that Brownian Motion (BM) is governed by a diffusive equation and can be explained as random walk as well. In BM the number of jumps is assume to be proportional to the time elapsed. If we make the number of steps grows at t^2, the states of the random walk would be converge to self-similar processes. Actually, this random walk can be stated in

$$\frac{\partial P(x,t)}{\partial t} = D(t)\frac{\partial^2 P(x,t)}{\partial x^2} \tag{50}$$

where $D(t)$ is a time dependent diffusion constant and $P(x,t)$ is the probability that the random walk has the displacement x at time t. The analytical solution of (50) can be obtained through Fourier transform and it is

$$P(x,t) = \frac{1}{\sqrt{4\pi d(t)}} \exp\left[\frac{x^2}{4d(t)}\right], \tag{51}$$

where $d(t) = \int D(t)dt$. This result implies that

$$Var(x(t)) \propto d(t) \tag{52}$$

. Therefore, the resulting processes $x(t)$ have varying self-similarities depending on the form of $D(t)$.

3.7.2 Stochastic Differential Equation

Chen *et al* (1997) modeled the body movement $x(t)$ of synchronizing one's limb to a periodic environmental signal as Stochastic Differential Equation

$$\frac{d^2x(t)}{dt^2} + \left(\frac{dx(t)}{d(t)}\right)^3 - \frac{dx(t)}{d(t)} + \frac{dx(t)}{d(t)}x(t)^2 + x(t) = \beta[\sin(\omega t) - x(t-\tau)]^2 + \sqrt{Q}\xi_t, \tag{53}$$

where β is the strength of the coupling between input periodic stimulus $\sin(\omega t)$ and history movement $x(t-\tau)$, ξ_t represents Gaussian white noise of unit variance, and Q is noise amplitude. Numerical simulation shows that with appropriate parameters β, ω, and Q the simulate time series exhibits apparent self-similarity, which is consistent with the finding obtained from the observed data.

3.7.3 Differential Equation with a Singular Term

Vicsek and Vicsek (1995) consider a differential equation with a singular term,

$$\frac{\partial h(x,t)}{\partial t} = \Delta^2 h(x,t) - B \ln(|\Delta h| + A), \tag{54}$$

where the positive parameter A is used to control the largest possible value of the singular term and B is used to monitor the relative strength of a singular term. By numerical integration, Vicsek and Vicsek (1995) show that the discretized version of this simple deterministic partial differential equations with singular terms (model (54)) exhibit rich spatiotemporal behaviour representing a mixture of stochastic and deterministic regimes. The fractal roughness of the growing surface $h(x,t)$ for $B = 0.002$ is indicated by a power-law relationship and the self-similarity exponent is found to be $H \approx 0.7$

In a directed polymer problem, Zhang (1992) adopted a model that reads

$$\frac{\partial h(x,t)}{\partial t} = \Delta^2 h(x,t) - \ln(\Delta h|), \tag{55}$$

and the corresponding self-similarity exponent is found to be $H \approx 0.2$.

3.8   Conclusions

In this chapter, we surveyed the mechanism of producing self-similarities in the observed data. These mechanisms are representative physics or mathematics which have been commonly used in exploring the problems in science, engineering and other areas. The links between these mechanisms and the observed self-similarities suggest that the self-similar process is not a enforced statistical model solely based on data. Also this survey provides researchers with different possible directions to explain the self-similarities in the data. For example, the self-similarities in the pupil data could be explained from different points of view.

3.9 References

1. Arrosmith, D. and Place, C. (1990). *An lntrodwcrion to Dynamical System*, Cambridge: Cambridge University Press.

2. Bak, P., Tang, C., and Wiesenfeld, K. (1987). Self-organized criticality: an explanation of 1/f noise, *Phys. Rev. Lett.*, **59**, pp. 381.

3. Bolthausen, E. (1989). A central limit theorem for two-dimensional random walks in random sceneries, *Ann. Probab.*, **17(1)**, pp. 108115.

4. Chen, Y., Ding, M., and Kelso, J. A. S. (1997). Long Memory Processes (1/f type) in Human Coordination, *Phys. Rev. Lett.*, **79**, pp. 4501-4504.

5. Cox, D.R. (1984). Long-range dependence: A review, *An appraisal. proceedings 50th Anniversary Conference. H.A. David, H.T David (eds.). The IOWA university Press* , pp. 55-74.

6. Granger, C.,(1980). Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics*, **14**, pp. 227238.

7. Kesten, H. and Spitzer, F. (1979). A limit theorem related to a new class of self-similar processes, *Z. Wahrsch. Verw. Gebiete*, **50(1)**, pp. 525.

8. Klüppelberg, C. and Kühn, C. (1999). Fractional Brownian motion as a weak limit of Poisson shot noise processes-with applications to finance, *Stochastic Processes and their Applications*, **113(2)**, pp.333-351

9. Klüppelberg, C. and Mikosch, T. (1995) Explosive poisson shot noise processes with applications to risk reserves, *Bernoulli,* **1(2)**, pp. 125-147.

10. Kuno, A. and Ikegaya, K. (1973). A statistical investigation of acoustic power radiated by a flow of random point sources, *J. Acoustic Soc. Japan,* **29**, pp. 662-671.

11. Lewis, P. (1964). Branching Poisson Process Model for the Analysis of Computer Failure Patterns, *Journal of the Royal Statistical Society, Series B*, **26**, pp. 398-456.

12. Oppenheim, G. and Viano, M.-C., (2001). Aggregation of random parameters Ornstein-Uhlenbeck or AR processes: some convergence results, Preprint.

13. Parzen, E. (1962).*Stochastic Processes,* Holden-Day, San Francisco, 1982.

14. Piau, D. (2002). Scaling exponents of random walks in random sceneries, *Stochastic Process. Appl.*, **100**, pp. 325.

15. Piau, D. (2004). Further scaling exponents of random walks in random sceneries, *Stochastic Process. Appl.*, **112**, pp. 145–155.

16. Reed, W.J. and Hughes, B.D.(2002). From gene families and genera to incomes and internet file sizes: why power-laws are so common in nature, *Phys. Rev. E*, **65**, pp. 67-103.

17. Reed, W.J. and Hughes, B.D.(2004). A model explaining the size distribution of gene and protein families, *Matematical Biosciences*, **189(1)**, pp. 97-102.

18. Rice, R. (1944). Mathematical anlysis of random noise, *Bell Syst. Tech. J.,*, **23**, pp. 282-332.

19. Samorodnitsky, G. (1996). A class of shot noise models for financial applications, *Proceeding of Athens International Conference on Applied Probability and Time Series, Volume 1: Applied Probability,*, pp. 332-353, Springer.

20. Taqqu, M. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process, *Z. Wahr. und Verw. Gebiete*,**31**, 287-302.

21. Taqqu, M. (1997). Proof of a fundamental result in self-similar traffic modeling, *Computer Communication Review,* **27**, pp. 5-23.

22. Taqqu, M. and Levy, J. (1986). Using renewal processes to generate long-range dependence and high variability, *Dependence in Probability and Statistics: A Survey of Recent Results, volume 11 of Progress in Probability and Statistics*, pp. 73-89. Birkhäuser, Boston, 1986.

23. Vere-Jones, D. (1970). Stochastic models for earthquake occurrence, *Journal of Royal Statistical Society, Series B*, **32**, pp. 1-62.

24. Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.

25. Vicsek, M. and Vicsek, T. (1995). Intermittency, stochastic growth and phase transition in a simple deterministic partial differential equation with a singular term, *J. Phys. A: Math. Gen.*, **28**, pp. 311-316.

26. Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies, *Journal of Econometrics*, **120(1)**, pp.75-102.

27. Zhang, Y. C. (1992). No Title, *J. Physiqlie*, **I (2)**, pp. 2175.

CHAPTER IV

MULTISCALE SCHUR MONOTONE DESCRIPTORS FOR HIGH FREQUENCY PUPIL DATA

This chapter addresses the problem of classifying users with different visual abilities based on their pupillary response while performing computer-based tasks. Multiscale Schur Monotone (MSM) summaries of high frequency pupil diameter measurements are utilized as feature (or input) vectors in this classification. Various MSM measures, such as Shannon, Picard, and Emlen entropies, the Gini coefficient and the Fishlow measure, are investigated to assess their discriminatory characteristics. A combination of classifiers, motivated by a Bayesian paradigm, is proposed to minimize and stabilize the misclassification rate in training and test sets with the goal of improving classification accuracy. In addition, the issue of wavelet basis selection for optimal classification performance is discussed. The members of the Pollen wavelet library are included as competitors. The proposed methodology is validated with extensive simulation and applied to high-frequency pupil diameter measurements collected from 36 individuals with varying ocular abilities and pathologies. The expected misclassification rate of our procedure can be as low as 21% by appropriately choosing the Schur Monotone summary and using a properly selected wavelet basis.

4.1 Introduction

The HCI (human-computer interaction) community is interested in understanding the unique interaction needs and behaviors of individuals with visual impairments who retain visual capabilities, albeit at a below 'normal' level (Biglan *et al.*, 1988). Therefore, there is a need for methods and procedures that can provide meaningful classification of individuals with varying visual abilities. In the human visual system, the pupil functions as a gain control

device, which responds to external stimuli, such as luminance changes, color and pattern changes, onset of motion, attention and social signaling, in a very subtle way. It is widely accepted (Backs and Walrath, 1992; Hess and Polt, 1960; 1964) that pupillary response (in terms of the dynamic pupil size) is an important measure of mental workload, which may be useful for classifying users with different abilities.

However, the pupil has an extremely complex control mechanism, which is moderated by several variables (Sahraie and Barbur, 1997), as well as various neural control pathways (Barbur, 2003). As such, it is very difficult to tease out the underlying differences in mental workload from point differences in pupil diameter. The inherent complexity of pupillary behavior requires that robust and valid measures be developed to extract the meaningful components from dynamic pupil behavior. While smoothing large aberrations in data values, and using global or local means may be suitable in helping to highlight even slight changes in pupil diameter for short, simple tasks, this averaging typically does not work for longer, more complex tasks that will inherently include more natural fluctuations in pupillary response and a larger number of confounding, non-cognitive effects. This being said, it is necessary to develop analytical techniques that can isolate these small changes in pupillary behavior. A more sensitive tool for the analysis of pupil measurement data may provide a solution to this problem and provide a unique characterization of interaction for individuals who are aging and/or have visual impairments.

This study examines the dynamic pupillary behavior of four groups of individuals, in which known performance differences were exhibited, during a computer-based task. Additionally, this study aims to examine if these behavioral differences can be sufficiently modeled for purposes of user classification, proposing the application of low dimensional summaries of high frequency data. Specifically, a summary measure called the Multiscale Schur Monotone (MSM) measure is derived to characterize the disbalance properties of the data distribution at different frequency scales. The MSM measure carries information not only about the disbalance characteristics of the data, but also about its correlation structure.

Thus, the MSM summary is likely to be more sensitive to the differences in visual functioning between users than any other single measure, such as correlation and global Schur Monotone measures. The combination of classifiers is proposed to address the inhomogeneous discriminatory information in the pupil diameter measurements.

The remainder of this chapter presents these MSM measures and their application in the classification of individuals with varying visual functioning. Section 4.2 derives a meaningful summary for high-frequency measurements for the purpose of classification, with wavelet transform and Schur Monotone measures briefly reviewed. Additionally, the concept of Schur Monotone summaries in the multiscale domain (MSM) is introduced and its application is illustrated via examples. Section 4.3 describes the classifier combining procedure and provides a Bayesian justification. Section 4.4 discusses the high frequency pupil diameter measurements used in this study. Section 4.5 illustrates the use of the MSM summaries of the high frequency pupillary behavior to classify the users. The K-nearest-neighbor (K-NN) classifier, equipped with combining techniques, is used for classification. Finally, Section 4.6 discusses the factors affecting the classification performance and the practical implications of the findings for the research in HCI.

4.2   Schur Monotone Summaries

In this section, we briefly review wavelet transforms. Next, the concepts of Schur Monotone ordering and Schur Monotone (SM) measures are presented. Then, we introduce the Multiscale Schur Monotone (MSM) measure as a natural way to combine multiscale representations and Schur ordering and give two illustrative examples to demonstrate this new measure.

4.2.1   Wavelet Transforms

Discrete wavelet transformations (DWT) have become indispensable tools in the analysis of data with complex stochastic structure. The DWT is an appropriate tool to model

non-stationary, non-Gaussian and long memory measurements. DWTs are simply linear transformations. Let y be a data vector of dimension (size) n. To avoid algorithmic complications we assume that n is an integer power of 2. The vector d representing a wavelet transform of vector y can be written as

$$d = Wy,$$

where W is an orthogonal matrix of size $n \times n$.

In practice, the choice of the orthogonal matrix W is related to the selection of wavelet basis, and ultimately, to the selection of a wavelet filter needed to implement the transform. The details on theories and statistical applications of wavelets can be found in (Vidakovic, 1999). Due to the properties of the wavelet functions, W usually admits a factorization in terms of a series of sparse matrices. A fast algorithm based on filtering to (equivalently) factorize the matrix W and calculate the wavelet-transformed vector d was proposed by (Mallat, 1989). This algorithm is easily implemented and is a part of many standard wavelet packages, such as the WAVELAB module for MATLAB by Stanford University (http://www-stat.stanford.edu/~wavelab/).

4.2.2 Schur Monotone (SM) Ordering

Schur Monotone orderings are based on the SM measure of a vector. This is used to order the vectors in terms of the their "disbalancing" characteristics. The definition of Schur ordering utilizes the inverted order statistic of two normalized vectors with non-negative components. For a pair of n-dimensional vectors x and y with non-negative components, the Schur ordering is defined as

$$x \prec y \quad \text{if} \quad \begin{cases} \sum_{i=1}^{k} x_{[i]} < \sum_{i=1}^{k} y_{[i]}, & k = 1, ..., n-1 \\ \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \end{cases}, \tag{56}$$

with $x_{[i]}$ and $y_{[i]}$ being the ith largest components of x and y respectively. When $x \prec y$, then it is said that x is Schur majorized by y.

4.2.3 Schur Monotone Measures

The Schur Monotone measure is a scalar value assigned to a vector that is sensitive to the Schur Monotone order. There are many available Schur Monotone measures, which have been previously used in economics and biology. In fact, any function ϕ such that

(1) $x \prec y \Longleftrightarrow \phi(x) \leq \phi(y)$**, and** $\phi(ax) = \phi(x)$ **for all** $a > 0$**,**

OR

(2) $x \prec y \Longleftrightarrow \phi(x) \geq \phi(y)$**, and** $\phi(ax) = \phi(x)$ **for all** $a > 0$

can be used to measure the disbalance of a vector x. If the first condition is satisfied, $\phi(\cdot)$ is called a Schur convex measure. If the second condition is true, $\phi(\cdot)$ is called a Schur concave measure. In both cases, $\phi(\cdot)$ is a Schur Monotone measure. In this chapter, we are interested in a SM measure defined as

$$\phi_2(x) = -\sum_{i}^{n} \log \frac{x_i}{S},$$

where $S = \sum_i x_i$. This SM measure is usually called the Picard entropy (Picard, 1979), which is slightly different from Shannon's entropy (Shannon, 1948). Other SM measures utilized in this study include Gini's coefficient (Gini, 1912), Fishlow's measure (Fishlow, 1973) and Emlen's modified entropy measure (Emlen, 1973). There is a comprehensive theoretical description and comparison of different SM measures in Marshall and Olkin, 1979. The choice of Picard entropy is substantiated by the relatively good performance in discriminations as shown in Section 4.5.

4.2.4 Multiscale Schur Monotone Measures

As previously noted, Schur Monotone (SM) measures have been popular in economics and biology. SM summaries usually measure disbalance (or non-uniformity) in an observed vector. Therefore, this measure is expected to provide good discriminative information if the analyzed vectors have different uniformity characteristics. Unfortunately, in some practical examples, the global disbalance (in the time domain) among the data vectors are

too weak to result in powerful discriminatory information. However, if we transform data to the time-scale (wavelet) domain and compare the disbalance at corresponding frequency scales, the discriminatory power may increase. This increase in sensitivity is apparently due to the unmasking of the balance caused by the interplay of different scale structures and the trends in the data. Through DWTs, the data vector is transformed to several wavelet coefficient vectors at different frequency scales (also called resolution levels). Therefore, we are able to define the SM measure at each level, with each measure summarizing the disbalance information of the data vector within distinguishable scales. This natural concept is named the Multiscale Schur Monotone (MSM) measure. The computation of MSM is illustrated in Figure 10.

Figure 10: Computation Diagram of Multiscale Schur Monotone Measures.

We provide an example to illustrate the case when the Multiscale Schur Monotone measure is beneficial compared to global, time domain disbalance measures. The exemplary datasets are the two functions:

$$f(t) = \texttt{Doppler} + fGn(H = 0.2);$$

$$g(t) = \text{a fixed permutation of } f(t),$$

where `Doppler` is a standard nonstationary testing function commonly used in nonparametric regression (Donoho and Johnson, 1994) and $fGn(H = 0.2)$ is a fractional Gaussian noise with Hurst exponent $H = 0.2$ (Mandelbrot *et al.*, 1968). Time series plots of the typical data simulated from $f(\cdot)$ and $g(\cdot)$ are presented in Figure 11. Clearly, these two functions do not differ from each other as far the Schur Monotone measures in the time domain is concerned, since this measure is invariant with respect to permutation. However, if we map the data into the wavelet domain and compute the MSM measures, the different scale levels become evident in their disbalancing measures. The apparent differences are demonstrated by Figure 12, which is obtained by the analysis of simulated data. We simulated 200 sample paths from $f(\cdot)$ and $g(\cdot)$ respectively, each of length $N = 2048$. Next, the MSM measures were computed for each sample path. The disbalance measure employed here within each scale in the MSM is the Picard entropy as defined in 4.2.3, though other disbalance measures will show similar results. To examine the differences of the MSM measures between $f(\cdot)$ and $g(\cdot)$, we provide box plots of the MSM at each scale, which are included in Figure 12. The ability to distinguish $f(\cdot)$ and $g(\cdot)$ using MSM measures can be explained by the disbalancing property of DWTs. The total inequality exhibited in the time domain is allocated to different frequency scales depending on the correlation structure. Statistically speaking, $f(\cdot)$ has higher values of the MSM in the first three scales than $g(\cdot)$ with large probability. This pattern is more pronounced in the finer scales because of the smoothing effect of wavelet filtering.

4.3 K-*Nearest-Neighbor Classifiers and Their Combinations*

The Nearest Neighbor (NN) method is one of the simplest ideas for modeling the regression (or classification) function between the response and predictor variables. It can be expressed as

$$\hat{y}(x_j) = \frac{1}{K} \sum_{x_i \in N_K(x_j)} y_i, \tag{57}$$

Figure 11: Typical time series plot for the data simulated from functions $f(t)$ and $g(t)$, $t = 1, 2, ..., 2048$

Figure 12: Boxplots of MSM measures for 200 replicates of simulated sample paths generated from $f(\cdot)$ and $g(\cdot)$, each with $t = 1, 2, ..., 2048$. The three panels correspond to the finest three scales. The righthandside panel represents the finest scale, while the lefthandside panel corresponds to the coarsest level.

55

where $\hat{y}(x_j)$ is the fitted value of the response at x_j and $N_K(x_j)$ is the set of first K nearest points to x_j in the predictor variable space. In our classification problem, the response variable y (user group) is categorical and takes only discrete values (e.g., control, group #1, etc.). The closeness concept used here is based on Euclidean distance. The Nearest Neighbor method makes minimal assumptions on the underlying data and is very flexible with respect to finding an arbitrary boundary. The crucial part of K-NN modeling is the tuning of the parameter K. It is well known (Hastie *et al.*, 2001) that in classification problems with $K = 1$, the misclassification rate is zero for training data set. However, the classification boundary resulting from the Nearest Neighbor method depends very much on the data adequacy of the training set. As a result, the boundary is often very wiggly and unstable in the test set. In other words, K-NN classifiers often have a large amount of variance in the prediction for the independent test data set.

An individual classifier usually performs best for certain types of data. However, due to the complexity of certain types of datasets and/or those with a small number of sample paths, the true properties of a population are not adequately described by a single classifier. In other words, the inhomogeneities of the dataset make it difficult to find a single K-NN classifier that optimally fits the data into. Although optimal results are achieved from individual classifiers, each classifier describes the dataset by emphasizing certain local aspects of features. The misclassification of aspects of different classifiers does not necessarily overlap. The non-overlapped misclassified measurements suggest that those individual classifiers provide complementary information for the prediction. Therefore, a scheme using a combination of the classification rules may result in better predictive performance. A diagrammatic representation of the classifier combining procedure is presented in Figure 13. In this chapter, the classifiers to be combined are K-NN with different tuning parameters K. We employed $R = 8$ and $C_1, C_2, ..., C_R$ are K-NN classifiers with $K = 3, 4, \ldots, 10$ in our simulation studies afterward.

The combined classifier was originally proposed as an ad hoc procedure, which was

Figure 13: Diagram of Classifier Combining

then justified by Bayesian decision theory (Kittler *et al.*, 1998). Consider a classification problem where four classes ($y = 0, 1, 2, 3$) are to be distinguished. Suppose that there are R possible classifiers available, which we note as $C_1, C_2, ..., C_R$. Each input x_j is assumed to have prior probability $P[y(x_j) = k]$, with $k = 1, 2, ..., 4$. According to Bayesian theory, the predicated class $\widehat{y(x_j)}$ of measurement j with feature vector x_j, $j = 1, 2, ..., N$ is

$$\widehat{y(x_j)} = \arg \max_{k \in \{0,1,2,3\}} P[y(x_j) = k | C_1, C_2, ..., C_R]. \tag{58}$$

Using Bayes theorem, the posteriori probability in (58) is

$$
\begin{aligned}
P[y(x_j) = k | C_1, C_2, ..., C_R] &= \frac{P[C_1, C_2, ..., C_R | y(x_j) = k] P[y(x_j) = k]}{P[C_1, C_2, ..., C_R]} \\
&= \frac{P[C_1, C_2, ..., C_R | y(x_j) = k] P[y(x_j) = k]}{\sum_{m=0}^{3} P[C_1, C_2, ..., C_R | y(x_j) = m] P[y(x_j) = m]}.
\end{aligned}
\tag{59}
$$

Several combination rules can be derived from (59), based on different assumptions on the model probability distribution $P[C_1, C_2, ..., C_R | y(x_j) = k]$ and the prior probability $P[y(x_j) = k]$. These combining rules are summarized in Table 1. The final decision from combined classifiers is

$$\widehat{y(x_j)} = \arg \max_{k \in \{0,1,2,3\}} G(k), \tag{60}$$

where the decision criteria function $G(.)$ is chosen from one of six rules in the second column of Table 1 .

Table 1: Summary of combining rules for multiple classifers

Rule	Decision Criteria $G(k)$	Assumptions
Product	$\dfrac{\prod_{i=1}^{R} P[y(x_j)=k|C_i]}{(P[y(x_j)=k])^{R-1}}$	$P[C_1, C_2, ..., C_R|y(x_j)=k] = \prod_{i=1}^{R} P[C_i|y(x_j)=k]$
Mean	$\dfrac{\sum_{i=1}^{R} P[y(x_j)=k|C_i]}{R}$	$P[y(x_j)=k|C_i] = P[y(x_j)=k](1+\delta_{ki}), \quad \delta_{ki} << 1$
Median	$\text{median}\{P[y(x_j)=k|C_i], \\ i=1,2,...,R\}$	$P[y(x_j)=k|C_i] = P[y(x_j)=k](1+\delta_{ki}), \quad \delta_{ki} << 1$
Max	$\max\{P[y(x_j)=k|C_i], \\ i=1,2,...,R\}$	$P[y(x_j)=k|C_i] = P[y(x_j)=k](1+\delta_{ki}), \quad \delta_{ki} << 1$ $\dfrac{\sum_{i=1}^{R} P[y(x_j)=k|C_i]}{R} \approx \max_{i=1}^{R} P[y(x_j)=k|C_i]$ $P[y(x_j)=0] = P[y(x_j)=1] = ... = P[y(x_j)=3]$
Min	$\min\{P[y(x_j)=k|C_i], \\ i=1,2,...,R\}$	$P[C_1, C_2, ..., C_R|y(x_j)=k] = \prod_{i=1}^{R} P[C_i|y(x_j)=k]$ $\prod_{i=1}^{R} P[y(x_j)=k|C_i] \approx \min_{i=1}^{R} P[y(x_j)=k|C_i]$ $P[y(x_j)=0] = P[y(x_j)=1] = ... = P[y(x_j)=3]$
Majority Voting	$\sum_{i=1}^{R} \Delta(y(x_j)=k|C_i)$	$P[y(x_j)=k|C_i] = P[y(x_j)=k](1+\delta_{ki}), \quad \delta_{ki} << 1$ $\Delta(y(x_j)=k|C_i) = \mathrm{I}(k = \arg\max_{k\in\{0,1,2,3\}} P(y(x_j)=k|C_i))$ $P[y(x_j)=0] = P[y(x_j)=1] = ... = P[y(x_j)=3]$

4.4 High Frequency Pupil Dataset

In this section, we briefly describe the datasets used in this study and how the data was preprocessed.

4.4.1 Dataset description

The equipment used to collect the pupillary response data during this study was the Applied Science Laboratories (ASL) Model 501 head-mounted optics system. Pupil size was recorded, at a rate of 60 Hz, for each participant over 105 trials of a computer-based task using a graphical user interface (GUI). A camera records the pupil image, which has been brightened by a near-infrared beam that illuminates the interior of the eye. Pupil size is assessed as the number of pixels attributed to the pupil's image, which has been determined by real-time edge detection processing of the image. Actual pupil diameter measurements (in millimeters) are then calculated by multiplying each pixel count by a scaling factor, which is based on the physical distance of the camera from the participant's eye.

The dataset is comprised of pupillary response data streams for 36 individuals, as described in Table 12. In this table, N refers to the number of individuals comprising this user group. Visual acuity refers to the range of Snellen visual acuity scores of the better eye for participants of each group. AMD refers to the presence (Yes) or absence (No) of the disease AMD (age-related macular degeneration) in individuals within each group. The number of data sets refers to the number of 2048-length data sets that were obtained from the data streams for each group. For this study, data was collected from four groups of individuals, classified by visual acuity and the presence or absence of AMD. Visual acuity, an individual's ability to resolve fine visual detail, was assessed via the protocol outlined in the Early Treatment of Diabetic Retinopathy Study (ETDRS) (University of Maryland School of Medicine, 2002). The experimental protocol from this study is fully described in studies by Jacko and colleagues (Jacko *et al.*, 2003).

Table 2: Group characterization summary

Group	N	Visual Acuity	AMD?	Number of Data Sets
Control	19	20/20 - 20/40	No	111
#1	6	20/20 - 20/50	Yes	59
#2	5	20/60 - 20/100	Yes	57
#3	6	> 20/100	Yes	124

4.4.2 Preprocessing

Studies of pupillary responses are faced with the problem of how to remove blink artifacts. A blink generally lasts about 70-100 msec. (producing an artifact spanning 4-6 observations under 60 Hz sampling) during which time the camera registers a loss and a pupil diameter of zero is recorded. Thus, it is generally straightforward to detect and eliminate these contiguous zero observation artifacts from the record. However, on either side of a blink, one may also observe highly unusual recordings because the pupil may be measured inaccurately when the eyelid partially obscures the pupil. The result may be an impossibly small value for the pupil's size.

To ensure that the analysis is conducted on pupil constriction or dilation and not on misleading discontinuities caused by blinks or partial blinks, one either remove the blink observations from the data entirely or replace them with interpolated values. Blinks (i.e., zero recordings) have been found to account for approximately 3-4% of all observations, with partial blinks accounting for another 1% of the total number of observations. The blink-removal procedure removes all observations having zero values (i.e., the blink) as well as any extreme values that occur within six additional observations on either side of the zero value (i.e., partial blinks). Figure 24 presents a preprocessed result of the typical measurements from a healthy subject (control group).

Because of the difficulty in collecting these measurements, especially from individuals with AMD, the original datasets were cut into equal length pieces to exploit their usage. Another reason for the segmentation is that the original data streams were not equally

Figure 14: Typical measurements with different resolutions from a heathy subject (control group)

Figure 15: Sample autocorrelation of Figure 24 (top) measurements

long among participants. The segmentation is conducted after the 'Six Law' filtering, as mentioned above. After segmentation and necessary outlier detection, the overall dataset contains the sum of 351 segments of measurements. Each segment has a length of 2048 readings. The distribution of the number of the segments among the four groups (Control, #1, #2 and #3) is reported in Table 12.

4.5 User Classification using Multiscale Schur Monotone Measures

In this section, we attempt to classify the user groups based on their high frequency pupil diameter measurements. Due to the high dimensionality of these measurements, it is necessary to derive low dimensional summaries from these measurements to characterize these users. The disbalance feature of the pupil-diameter measurements could be a good summary measure to describe the eye behaviors during computer-interaction tasks. However, by simply looking at the statistics of the SM measure in the time domain as shown in Table 3, no significant differences were found to distinguish these four groups. To increase the sensitivity of the disbalance measures, we employed MSM measures. This choice is also motivated by the fact that there are apparent long range correlations within these measurements as displayed in Figure 25. The MSM measure characterizes the measurement by considering both the disbalance and correlation structure simultaneously, which is not possible in the time domain. Summary statistics of MSM measures are provided in Table 4. As we can see from this table, the differences among these groups are reflected by the MSM measures, especially at the fine scales. For example, at the finest scale, level 1, the mean MSM measures of group #1 and #2 are much smaller than that of the control group. Group #3 tends to have similar mean disbalance at level 1 as the control group while they apparently have different medians. These results are interesting, as the MSM measures may provide evidences of the erratic effect that ocular diseases (in the case of groups #1

and #2), particularly central field deficiencies, have on pupillary response behavior, as previously discussed. While the distinction between groups #1 and #2 and the control group is expected, given the presence of ocular diseases in the experimental groups, the similarity of the control group and group #3 is unexpected. As seen in Table 12, the control group and group #3 are the most diverse with respect to both the presence of AMD and the level of visual acuity. However, the MSM measures at level 2 indicate considerably more distance between these two groups.

The addressed differences in the MSM measures imply the discriminatory information. To fully integrate this information, we propose to use the combined K-NN classifiers to develop a statistical classification procedure to automatically distinguish the MSM measures of the different user groups.

Table 3: Summary statistics of Schur Monotone Measures (10^4) in the time domain

	Control	#1	#2	#3
Min	1.563	1.5624	1.5625	1.5632
Mean	1.5619	1.5617	1.5618	1.5618
Median	1.5617	1.5617	1.5617	1.5618
Max	1.5616	1.5616	1.5616	1.5616
Std. Dev.	0.0003	0.0002	0.0002	0.0002

A 5-fold cross-validation scheme is used to guarantee the robustness of our procedure. The datasets described in Section 4.4 are randomly divided into two parts: 80% of the measurements are used in a training set, which is used to estimate the classification model; and the remaining 20% of the measurements are regarded as a test set, used to validate the classification model. The default scheme of combining K-NN classifiers (see Section 4.3) is employed to classify the MSM measures computed from pupil diameter measurements. The cross-validation is repeated twenty times in order to estimate the mean and standard deviation of the misclassification rate. The classification result is summarized in Table 14. The MSM measures at the three finest scales are considered here as input vectors in the classification model and the wavelet basis used in DWT is the Daubechies wavelet

Table 4: Group characterization summary in terms of the Schur Monotone measures in the wavelet domain. Level 1-4 represent the first, second, third, and fourth finest scales respectively.

	Statistics	Level4	Level3	Level2	Level1
Control	Min	583.02	1358.64	5467.64	19404.85
Group	Mean	470.32	1100.71	3739.06	15526.69
	Median	459.57	1093.07	3629.92	15703.12
	Max	390.71	944.39	2861.20	12335.15
	Std Dev.	54.45	106.53	578.94	2154.64
Group	Min	564.35	1559.91	5169.51	19261.05
#1	Mean	447.78	1073.71	3458.55	13626.51
	Median	444.43	1006.24	3282.31	14814.35
	Max	389.91	875.12	1954.61	6770.97
	Std Dev.	37.34	166.40	1001.68	3610.46
Group	Min	516.56	1297.45	4777.68	18714.94
#2	Mean	429.85	1030.24	3423.02	14160.38
	Median	431.18	1024.13	3538.62	15387.57
	Max	380.67	885.53	2187.12	7871.65
	Std Dev.	30.24	79.17	728.43	3287.20
Group	Min	733.43	2366.90	9005.10	25460.81
#3	Mean	452.76	1304.61	4738.77	15989.98
	Median	432.37	1198.93	4800.60	18413.07
	Max	375.40	872.74	1933.66	4266.89
	Std. Dev.	58.81	385.70	2215.99	7435.45

with two vanishing moments. The product combining rule, which assumes independence among the classifiers to be combined, does not work very well in our study because we combine a family of K-NN classifiers with different tuning parameters, which are quite likely dependent on each other. Therefore, the results for the product combining rule are not reported here and thereafter. The rest of the combining rules mentioned in Section 4.3 work comparably in terms of the misclassification rate for the test set, although the MAX rule seems to slightly outperform the others.

Table 5: Error rates after combining Nearest Neighbor classifiers using MSM measures at three finest levels

	rule	mean	median	**max**	min	majority voting
Test	avg.	0.2566	0.2592	**0.2408**	0.2592	0.25
Error	std.	0.0345	0.0344	0.036	0.0382	0.0462
Training	avg.	0.2474	0.2551	**0.245**	0.2257	0.2489
error	std.	0.0098	0.0089	0.0104	0.0166	0.0093

4.6   Discussions

In this section, we discuss the possibility of improving the classification performance of the default model used in Section 4.5 by choosing the appropriate coarsest level and wavelet basis in the DWT.

4.6.1   Choice of the Coarsest Level

The number of scales included in the MSM measure is a parameter to be decided in our classification model. This is equivalent to choosing the coarsest level in DWT, which affects the size of the input vector and is further related to the fitting quality of the classifier. Too many scales in the coarsest levels may result in overfitting, while not enough DWT levels results in oversmoothing. Table 15 illustrates the fact that if we use MSM measures from the two finest scales, the performance will be decremented at least 3%, which implies an underfitted case. On the other hand, if we include MSM measures at the first four scales,

the performance will still be decremented at least 2% - a case of overfitting. Therefore, MSM measures at the first three finest scales are the optimal choice for our pupil diameter classification.

Table 6: Error rates after combining Nearest Neighbor classifiers using MSM measures at two finest levels (underfitting)

	rule	**mean**	median	max	min	majority voting
test	avg.	**0.2763**	0.2796	0.2875	0.2855	0.2987
error	std.	0.0347	0.0316	0.035	0.0334	0.0328
training	avg.	**0.232**	0.236	0.2465	0.232	0.2523
error	std.	0.0105	0.0088	0.0112	0.0097	0.0081

Table 7: Error rates after combining the Nearest Neighbor classifiers using MSM measures at four finest levels (overfitting)

	rule	**mean**	median	max	min	majority voting
test	avg.	**0.2605**	0.2664	0.2678	0.2711	0.2704
error	std.	0.0463	0.0471	0.043	0.0442	0.0504
training	avg.	**0.2342**	0.2384	0.2391	0.223	0.237
error	std.	0.0116	0.0123	0.0134	0.0127	0.0116

4.6.2 Wavelet Basis Selection

The wavelet basis has substantial influence on the transformed coefficients of pupil diameter measurements and is, therefore, an important factor in determining the classifier quality. We formulate an optimization study to search for the best wavelet basis, which results in the most accurate classification. The search will be limited to the Pollen wavelets library. Pollen wavelets are a family of wavelet basis with a continuous mapping from $[0, 2\pi]^{N-1}$ to a set of "wavelet solutions" in terms of the quadratic mirror filters of $h = \{h_0, h_1, \cdots, h_{2N-1}\}$, where N is the number of vanishing moments. Pollen representation of all wavelet solutions of length 4 ($N = 2$) and length 6 ($N = 3$) is given in Tables 8 and 9. The Daubechies wavelet family is included in the Pollen library as a special case.

There are many measures of classifier performance. Some popular measures include scatter-matrix and Bayesian risk, among others. Though these separability measures are

Table 8: Pollen paramterization for $N = 2$ (four-tap filters).$[s = 2\sqrt{2}]$

n	h_n for $N = 2$
0	$(1 + \cos(\theta) - \sin(\theta))/s$
1	$(1 + \cos(\theta) + \sin(\theta))/s$
2	$(1 - \cos(\theta) + \sin(\theta))/s$
3	$(1 - \cos(\theta) - \sin(\theta))/s$

Table 9: Pollen paramterization for $N = 3$ (six-tap filters).$[s = 2\sqrt{2}]$

n	h_n for $N = 3$
0	$[1 + \cos(\theta_1) + \cos(\theta_2) + \sin(\theta_1) - \sin(\theta_2) - \cos(\theta_1 - \theta_2) - \sin(\theta_1 - \theta_2)]/2s$
1	$[1 - \cos(\theta_1) + \cos(\theta_2) + \sin(\theta_1) - \sin(\theta_2) - \cos(\theta_1 - \theta_2) + \sin(\theta_1 - \theta_2)]/2s$
2	$[1 + \cos(\theta_1 - \theta_2) + \sin(\theta_1 - \theta_2)]/s$
3	$[1 + \cos(\theta_1 - \theta_2) - \sin(\theta_1 - \theta_2)]/s$
4	$[1 - \cos(\theta_1) + \cos(\theta_2) - \sin(\theta_1) + \sin(\theta_2) - \cos(\theta_1 - \theta_2) - \sin(\theta_1 - \theta_2)]/2s$
5	$[1 + \cos(\theta_1) - \cos(\theta_2) - \sin(\theta_1) + \sin(\theta_2) - \cos(\theta_1 - \theta_2) + \sin(\theta_1 - \theta_2)]/2s$

optimal (or almost optimal) under certain assumptions, computational issues like matrix inversion and prior statistical knowledge about the data often make this impractical. For a detailed discussion of these measures, the readers are directed to Fukunaga, 1990. A more practical and easily implemented measure of the separability is the misclassification rate associated with the choice of wavelet filter $H = (h_0, h_1, ..., h_n)$.

As a result, our search procedure focuses on minimizing the misclassification rate in the test set with respect to the wavelet filters. The first search is done in the Pollen library with $N = 2$. The results are presented in Figure 16. As shown in this figure, the performance varies up to about 9% with different values of θ and the best performance is achieved around $\theta = 100°$. The scaling and wavelet functions corresponding to this optimal Pollen wavelet basis are plotted in Figure 17. To compare the performance of the different pollen wavelet basis with a different number of vanishing moments, the search is conducted for Pollen wavelets with $N = 3$. The results are shown in Figure 18. For $N = 3$, there is more variability in performance with these different parameters than those in the case of $N = 2$, resulting in worse overall performance compared with $N = 2$. This can be attributed

to the locality and smoothness of the wavelet bases. The Pollen wavelet with $N = 3$ is smoother and, hence, tends to smooth the data more than $N = 2$. It may be the case that some of the discriminatory information has been smoothed, which causes the classification performance to become worse.

Figure 16: Misclassification rates using a Pollen wavelet basis with different parameter $\theta \in [0, 2\pi]$. The classifiers here are the combined k-NN with MsSC measure input vectors. The error rates shown in the figure are the average values for 20 randomly selections of test set from the whole dataset. The minimum error rate here is 21.32%, which is achieved at $\theta = 100°$.

Figure 17: GT wavelet basis (Four-tap Pollen wavelet basis with $\theta = 100^o$).

Figure 18: Misclassification rates using Pollen wavelet basis with different parameter $\theta \in [0, 2\pi]^2$. The classifiers here are the combined k-NN with MsSC measure input vectors. The error rates shown in the figure are the average values for 20 randomly selections of test set from the whole dataset. The error rates here are obviously quite larger than 4-tap Pollen wavelet filter. The 6 curves correspond to different values of θ_1.

4.7 Conclusions and Future Work

The classification procedure for the user group is developed utilizing the Multiscale Schur Monotone measures in the wavelet domain and ad hoc classifier combination schemes. We investigated the performance of different Multiscale Schur Monotone measures in this particular user classification problem. The Picard entropy measure works best among the considered candidates. We also considered the stabilization of misclassification rates by using combinations of single basic classifiers. This heuristic procedure implies some approximation of Bayesian model averaging. Our user classification example validates this procedure through the relatively low misclassification rate, which resulted in the randomly selected test set. Additionally, we studied the problem of searching for the optimal wavelet basis among certain candidate wavelet families. Those families includes Daubechies and Pollen. In the Pollen wavelet family (limited to four tap filters), we found that the basis with parameter $\phi = 100°$ achieves the best classification performance for the pupil-diameter measurements. Overall, the expected misclassification could be at least around 21% by choosing the appropriate wavelet basis and Multiscale Schur Monotone measure.

This exciting result is of much importance in the HCI community.

The utility of these analytical tools for applied research in HCI has tremendous potential, as user classification is of primary importance in this field of research. The use of novel statistical methods, as shown in this chapter, shows promise for the ability to use complex physiological data from users to better understand their unique needs and behaviors. While further data collection is needed to help increase the amount of data being analyzed, the initial results suggest that the presence of ocular disease and/or acuity loss does result in dynamic, complex differences in pupil behavior. In essence, MSM measures can be used to 'tease-out' differences in the pupillary behavior of individuals with and without ocular disease, possessing a range of visual acuity. The results not only show the fairly reliable classification or distinction of individuals with and without ocular disease (AMD), as the separation of the control group and group #1 illustrates, but the results also illustrate finer distinctions amongst groups with a similar ocular disease, but with varying visual functioning (e.g., visual acuity), as the separation of groups #1, #2, and #3 illustrates. This ability to separate these groups, based on dynamic pupillary behavior, illustrates the usefulness of these analytical procedures for user classification.

One of the overreaching goals of this study was to examine the use of high-frequency pupillary behavior as a means of quantitatively assessing differences between users during performance of a computer-based task. The results of this study show great potential toward this goal, as MSM measures were employed to distinguish the user groups. This distinction between user groups was used to generate a promising predictive model of user classification. The future implications of this study include the application of these, and similar, analytical tools for other high frequency physiological data, such as eye movement, heart rate and brain signals.

4.8 References

1. Backs, R.W. & Walrath, L.C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. Applied Ergonomics, 23, 243-254.

2. Barbur, J. L. (2003). Learning from the pupil - studies of basic measurement mechanisms and clinical applications. L. M. Chalupa & J.S. Werner (Eds.), Cambridge, MA: MIT Press, in press.

3. Biglan, A. W., Van Hasselt, V. B., & Simon, J. (1988). Visual impairment. In V. B. Van Hasselt (Ed.), *Handbook of Developmental and Physical Disabilities* pp471-502 New York, NY: Pergamon Press.

4. Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425-455.

5. Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. NY:Academics.

6. Hastie, T., Tibshirani, R. and Friedman, J. (2001). *Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer-Verlag, New York.

7. Hess, E. H. & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, **132**, 349-350.

8. Hess, E. H. & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science*, **143**, 1190-1192.

9. Jacko, J. A., Barreto, A. B., Marmet, G. J., et al. (2000). Low vision: The role of visual acuity in the efficiency of cursor movement. *Paper presented at the Fourth International ACM Conference on Assistive Technologies (ASSETS 2000)*.

10. Jacko, J. A., Barreto, A. B., Scott, et al. (2002). Macular degeneration and visual icon use: deriving guidelines for improved access. *Universal Access in the Information Society*, 1(3), 197-206.

11. Jacko, J. A., Dixon, M. A. et al.(1999). Visual profiles: A critical component of universal access. *Paper presented at the Human Factors in Computing Systems*, Pittsburg, PA.

12. Jacko, J. A., Rosa, R. H. et al. (2000). Visual impairment: The use of visual profiles in evaluations of icon use in computer-based tasks. *International Journal of Human-Computer Interaction*, **12(1)**, 151-164.

13. Jacko, J. A., Scott, I. U., Sainfort, F., Barnard, L., Edwards, P. J., Emery, V. K., Kongnakorn, T., Moloney, K. P., & Zorich, B. S. (2003). Older adults and visual impairment: What do exposure times and accuracy tell us about performance gains associated with multimodal feedback? *CHI Letters*, 5(1), 33-40.

14. Kittler, J. et al (1998), On Combining Classifiers, *IEEE Transactions on pattern analysis and machine learning intelligence* , 20(3):228-239.

15. Loewenfeld, I. E. (1999). The pupil: Anatomy, physiology, and clinical applications (2nd ed.). Oxford, UK: Butterworth-Heinemann.

16. Mallat, S. (1989). A Theory for Multiresolution Signal Decomposition: the Wavelet Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**:674-693.

17. Madndelbrot, B. et al. (1968), Fractional Brownian Motion, Fractional Noise and Applications, *SIAM review*, 10, 422-437, 1968.

18. Marshall, A. & Olkin, I. (1979). *Inequalities: theory of Majorization and its applications,* Academic Press.

19. Sahraie, A., & Barbur, J. L. (1997). Pupil response triggered by the onset of coherent motion. *Graefes Archives of Clincal Experimental Ophthalmology*, **235**, 494-500.

20. University of Maryland School of Medicine, Department of Epidemiology and Preventative Medicine. (1980). Early treatment diabetic retinopathy study, Manual of Operations, Chapter 12 (pp. 1-15). Baltimore, MD: ETDRS Coordinating Center.

21. Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, **41**, 167-174.

22. Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.

CHAPTER V

MULITFRACTAL DISCRIMINATION MODEL FOR TURBULENCE

Numerous studies have already demonstrated the multifractal properties of turbulence within the inertial subrange. How atmospheric stability (i.e. a boundary condition) alters the multifractal spectrum (MFS) of turbulent velocity and temperature fluctuations in the atmospheric surface layer (ASL) remains to be investigated. A challenge to estimating the MFS from time series via traditional regression approaches is the heteroskedastic problem because the variance of the error term are shown to depend on scale. Using a combination of Discrete Wavelet Transforms (DWT) and a Weighted Least Squares (WLS) scheme, heteroskedastic effects are minimized and a robust estimator of the scaling parameters needed to compute the MFS is derived. Next, to quantify the effects of atmospheric stability on the MFS function, we examine how stability impacts three canonical geometric attributes of the MFS - left slope (rise), mode, and broadness. For stationary conditions, the MFS mode can be related to the scaling or Hurst exponent. These geometric attributes of MFS (or GAMFS) are computed for the three velocity components and temperature time series collected in the ASL for a wide range of atmospheric stability conditions. The evaluation of how atmospheric stability alters the MFS of a single flow variable is conducted via ANOVA tests and Box-plots. We found that atmospheric stability impacts most the MFS mode for all four variables. In particular, we showed that the MFS mode appears smaller for stable than for unstable conditions. We also showed that the left slope is the least impacted GAMFS by atmospheric stability. Finally, we demonstrated that the GAMFS for temperature under near neutral and stable stability conditions significantly depart from their velocity counterparts. The latter finding indicates clear dissimilarity in the MFS statistical properties of temperature and velocity within the ASL inertial subrange. This dissimilarity

between temperature and velocity within the inertial subrange is consistent with numerous laboratory studies that utilized higher order structure function approaches.

5.1 Introduction

One of the main challenges to describing land-atmosphere momentum and sensible heat turbulent exchange is attributed to the nonlinearities in the Navier-Stokes equations, which lead to rich spectral variability at numerous turbulent scales and possible non-local interactions among them. Recent attempts to quantify the multiscale properties of such variability made use of wavelet transforms as an analyzing tool (e.g. [*Meneveau,* 1991; *Addison,* 2002]) and the multifractal formalism as a theoretical tool (e.g. [*Sreenivasan,* 1991; *Meneveau,* 1991; *Frisch,* 1995; *Antonia and Sreenivasan,* 1997; *Yakhot and Sreenivasan,* 2004]). In particular, the multifractal framework assumes that local singularities exist within turbulence [*Frisch and Parisi*, 1985] but are nested in fractal sets (see [*Riedi,* 1999; 2002] for reviews). This assumption deviates from the classical and widely used Kolmogorov $K41$ scaling in which turbulence is assumed to posses a constant scaling behavior characterized by a single Hurst exponent of $1/3$ ([*Mandelbrot,* 1968; *Frisch,* 1995]). The anomalous scaling (i.e. multifractality) of turbulence is commonly attributed to short-circuiting of the energy cascade due to the existence of organized large-scale features such as ramp-like structures, which are influenced by boundary conditions, and themselves directly influence small scale turbulence ([*Warhaft,* 2000; *Celani and Vergasola,* 2001]). Hence, a logical question to explore is whether measures of multifractality, such as the multifractal spectrum (MFS), are affected by boundary conditions such as surface heating and friction velocity or a combination thereof such as atmospheric stability. The MFS describes the "richness" of the local irregularity or disorder in turbulence in terms of local strengths of singularities. Furthermore, $K41$, and models that employ $K41$ (e.g. subgrid closure in Large Eddy Simulations; high frequency spectral corrections to eddy correlation measurements, synthetic turbulence generators, etc...) implicitly assume that the MFS is independent of atmospheric stability for velocity within the inertial subrange. Finally, while the inertial

subrange similarity (or dissimilarity) between temperature and velocity has already been explored in numerous studies via the structure function approach (e.g. [*Warhaft*, 2000; *Katul et al.*, 1997]), few studies considered this similarity (or dissimilarity) for the MFS within the context of different atmospheric stability regimes, the subject of this study.

In this chapter, we seek to address the following problem: how does atmospheric stability alter the multifractal spectrum of atmospheric surface layer (ASL) turbulence. Towards this end, we use longitudinal (u), lateral (v), and vertical (w) velocity along with virtual potential temperature (T) time series measurements collected in the ASL for a wide range of atmospheric stability conditions of a grass-covered surface.

There are two obstacles to addressing this problem: The first deals with the estimation of the MFS from limited time series, and the second deals with assessing the statistical significance of whether atmospheric stability alters the MFS. To address the first issue, we explore the estimation of MFS via discrete wavelet transforms (DWT) because of (i) the availability of fast algorithms (vis-a-vis box counting), (ii) automatic separation of trends and fluctuations, and (iii) the widely demonstrated effectiveness in various applications ([*Arneodo et al.,* 1998; *Vehel et al.,* 1997]). To address the second issue, we propose a general approach that does not a priori assume a cascade model but directly analyzes the effects of atmospheric stability on the geometric attributes of the MFS.

However, before we describe these approaches, a brief summary of the experiment and resulting data is presented next.

5.2   Data

Time series measurements of u, v, and vertical w velocity components, and air temperature T were collected at the Duke Forest near Durham, North Carolina. The measurements were conducted June, 12-16, 1995 at 5.2 m above a grass surface. The measured time series were sampled at $56Hz$ and were subsequently divided into 19.5 minute intervals

to produce $N = 65,536$ time measurement per flow variable per run. We use an ensemble of 95 runs, 6 collected in a stable regime, 23 in a near-neutral regime, and the rest in unstable atmospheric stability conditions. Further details about the experimental setup, atmospheric stability conditions, inertial subrange identification, and instrumentation details can be found elsewhere [*Katul et al.*, 1997; 2000]. To overcome the limited sample size of stable and neutral runs, each time series was further segmented into 4096 data points for the purposes of MFS estimation, described next.

5.3   Wavelet-based Multifractal Spectrum

Unlike power spectra, which describe the distribution of signal energy, the MFS describes the distribution of local singularities expressed in terms of the so-called Holder exponents. Geometrically, a local singularity at time t_o can be visualized as a relation between neighborhood fluctuations of a function $g(t)$ and two bounding curves as shown in Fig. 1. To estimate such a singularity, construct two curves given by $g(t_o) \pm c|t - t_o|^{\alpha}$, where c is a constant. The maximum value of α that locally bounds $g(t)$ in the neighborhood of t_o between these two curves is the local singularity (see bottom-left of Fig. 1). Note that when α is large, the curvatures are narrow thus limiting local fluctuations (see bottom-right of Fig. 1). On the hand, when α is small, the two curves have small curvature, thus allowing $g(t)$ to take on large local excursions (i.e. irregular). As t_o slides across the time series, the distribution of the resulting α is described by MFS. Formal definitions of local singularities are discussed elsewhere [*Seuret and Vehel,* 2001; *Pesquet-Popescu and Vehel,* 2002]. While this geometric interpretation is intuitive, it is computationally not feasible.

Several authors have explored the link between the size of the wavelet coefficient and Holder exponent, and hence, it is natural to consider wavelets as a basis for computing MFS ([*Meneveau,* 1991; *Muzy et al.,* 1993]). The wavelet-based calculation of MFS relies on

Figure 19: The geometric interpretation of local scaling exponents for a time series. The distribution of α derived from the time series at various times is the Multifractal spectrum.

the concepts of *partition function* , but for fast estimation algorithms, relies on the concept of *Legendre transform*. In brief, the partition function $T(q)$ is defined as

$$T(q) = \lim_{j \to -\infty} \log_2 E|d_{j,k}|^q, \tag{61}$$

where $d_{j,k}$ is the $\mathcal{L}_1$-normalized wavelet coefficient at level (or scale index) j and time (or position index) k, and q is the order of moment. Note that such normalization of the wavelet coefficient departs from the standard $\mathcal{L}_2$ normalization often employed in fast wavelet algorithms. We emphasize that q is real and can be positive or negative though negative moments do not have clear physical interpretations. While this definition of $T(q)$ is of theoretical significance, it also cannot be readily employed to time series with finite durations.

To circumvent this limitation, it has been proposed that the local singularity strength can be measured in terms of the wavelet coefficients via [*Gonçalvès et al.,* 1998]

$$\alpha(t) = \lim_{k2^j \to t} \frac{1}{j} \log_2 |d_{j,k}| \tag{62}$$

77

With α determined, the MFS, denoted by $f(\alpha)$, measures the distribution of $\alpha(t)$ within a time series and can be obtained using the standard box-counting technique,

$$f(\alpha) = \lim_{\epsilon \to 0} \#\{\alpha(t) : \alpha - \epsilon < \alpha(t) < \alpha + \epsilon, -\infty < t < \infty\}. \tag{63}$$

For an efficient algorithm that estimates the MFS by weighted least squares and the Legendre transform in the wavelet domain, see next section.

5.4 Estimation of MFS

Although it is feasible to estimate the MFS using (72) and (63), the method is not practical due to the difficulty in approximating the finite limit and the computational expense. A useful tool to improve the estimation efficiency is the Legendre transform. The Legendre transform of the partition function is

$$f_L(\alpha) = \inf_q \{q\alpha - T(q)\}. \tag{64}$$

It can be shown that $f_L(\alpha)$ converges to the true MFS using the theory of large deviations ([*Ellis,* 1984]).

From a practical point of view, we still require a robust estimator of the partition function. If we rearrange (61), it becomes,

$$E|d_{j,k}|^q \sim 2^{jT(q)}, \quad \text{as} \ \ j \to -\infty \tag{65}$$

It was shown that the qth moment of the wavelet coefficients of a power law process ([*Arneodo et al.*, 1998]) satisfies the following equation:

$$E|d_{j,k}|^q = C_q 2^{jqH} \tag{66}$$

where H is the so-called self-similarity exponent and C_q is a constant depending only on q. Comparing (65) and (66), one can relate the partition function estimation with the self-similarity exponent estimation problem. It is a standard practice to use linear regression to

identify the self-similarity exponent H since the values $E|d_{j,k}|^q$ can be easily obtained by a moment-matching method thereby facilitating the estimation of $T(q)$. Formally,

$$\log_2 \widehat{S_j(q)} \sim jT(q) + \varepsilon_j, \tag{67}$$

where $\widehat{S_j(q)} = \frac{1}{N2^{-j}} \sum_{k=1}^{N2^{-j}} |d_{j,k}|^q$ is the empirical q^{th} moment of the wavelet coefficients (N is the length of the entire time series) and the error term ε_j is introduced from the moment matching method when replacing the true moments with the empirical ones. Simple ordinary least square (OLS) is the most convenient choice for estimating the partition function. However, the bias can be large in some extreme cases since the variance of the empirical q^{th} moments is not constant with respect to scale j. The variance of $\log_2 \widehat{S_j(q)}$ is

$$Var(\log_2 \widehat{S_j(q)}) = \frac{A(q)}{N2^{-j}} + \frac{B(q)}{N^2 4^{-j}} + \cdots \tag{68}$$

where $A(q)$ and $B(q)$ are constants depending only on the underlying distribution function of the finest wavelet coefficient. Therefore, it is clear that the regression problem in (67) is a heteroskedastic problem in which the variances of the error terms are not constant across the scales. Even though the OLS solution of $T(q)$ is still unbiased and consistent asymptotically, it is no longer robust due to the heteroskedasticity. To resolve this problem, [*Audit et al.*, 2002] proposed a weighted least squares (WLS) method to obtain unbiased estimates of $T(q)$. The WLS estimator down-weights the squared residuals for scales with large variances, in proportion to those variances. If one finds $w_j = Var^{-1}(\log_2 \widehat{S_j(q)})$, a WLS estimator of $T(q)$ is given by

$$\widehat{T(q)} = \frac{\sum_{j=1}^{J} w_j \sum_{j=1}^{J} jw_j \log_2 \widehat{S_j(q)} - \sum_{j=1}^{J} jw_j \sum_{j=1}^{J} w_j \log_2 \widehat{S_j(q)}}{\sum_{j=1}^{J} w_j \sum_{j=1}^{J} j^2 w_j - (\sum_{j=1}^{J} jw_j)^2} \tag{69}$$

In practice, the exact analytical formula $Var(\log_2 \widehat{S_j(q)})$ is too complicated to be used directly. However, if the N is reasonably large (as is the case here), the first term in (68) will be dominant. Therefore, the approximate weights are $p_j = N2^{-j}$.

This WLS estimator results in a variance given by,

$$Var(\widehat{T(q)}) = \frac{A(q)C(J)}{N} + \frac{B(q)D(J)}{N^2} \tag{70}$$

79

where the constants $C(J)$ and $D(J)$ can be evaluated from the formula provided in [*Audit et al.*, 2002].

Once the $T(q)$ is estimated, the next step is to perform the Legendre transform. Since $\frac{\partial(\alpha q - T(q))}{\partial q} = \alpha - T'(q)$ and $T''(q) < 0$ (Gonçalvès *et al.*, 1998), the maximum value of $\alpha q - T(q)$ is achieved at $q = T'^{(-1)}(\alpha)$. So, performing the Legendre transform is divided into two steps: First, the numerical derivative of $T(q)$ is obtained using finite differencing; then, the value of Legendre spectrum at $\alpha = \widehat{T'(q)}$ is evaluated. We point out that the Legendre transform is not able to estimate the MFS value at arbitrary singularity strength α. The set of the MFS values are determined by a pre-specified vector of q values. The more q values used, the finer the MFS will appear, i.e., the resolution of the spectrum is determined by the "(order) sampling frequency" of the moments. Note that the resolution of the q vector does impact the calculation of broadness measure through c_f.

5.5 *Geometric Attributes of the Multifractal Spectrum*

To assess how atmospheric stability impacts the MFS, we first introduce three geometric features of the MFS and then track how variations across stability classes result in variations in the three features. These three features originate from the fact that the MFS of a fractional Brownian Motion or an fBm process (i.e. mono-fractal) is known to consist of three canonical descriptors: the vertical line (slope=$+\infty$), the maximum point, and the right slope (slope=-1) ([*Gonçalvès et al.,* 1998]). Mathematically, this fact is stated in the following theorem.

Theorem 5.5.1 *For the standard fractional Brownian Motion (fBm) with Hurst exponent H, its multifractal spectrum $f(\alpha)$, which describes the distribution of local singularities (i.e. α's), is*

$$f(\alpha) = \begin{cases} -\infty, & \alpha < H \\ 0, & \alpha = H \\ H - \alpha, & \alpha = H \end{cases} \tag{71}$$

80

The maximum point corresponds to the mode and the vertical line and the right slope are thought to be inherent features that distinguish fBm from a multifractal process. To illustrate, consider the two processes in Fig. 2. One of them is a synthetically generated fBm while the other one is taken from the turbulence measurements (widely believed to be a multifractal process). Several observations can be drawn from this figure:

1) for the fBm process, it is rare to obtain a perfect spectrum (i.e. vertical line, maximum point at the Hurst exponent, and a unity right slope) because of imperfections in the MFS estimation processes.

2) the left slope of the fBm MFS is much closer to the vertical line when compared to the MFS of turbulence.

3) the width spread of the fBm MFS is much smaller than that of the turbulence MFS indicating lack of richness in singularity indices for the fBm process (as expected).

Figure 20: Multifractal Spectra for mono- (dashed line) and multi-fractal (solid line) processes. The dotted line indicates the theoretical slope of the spectrum for an fBm process (mono-fractal) with a Hurst exponent of $1/3$

In short, this figure demonstrates that even for a multifractal process, the key attributes of the MFS can be quantitatively described with a left slope, mode, and width spread.

While the left and right slopes can be obtained from linear regression, the width spread cannot be computed automatically. This difficulty is attributed to (i) automatically locating the start and end points of the width spread, and (ii) the discreteness of the MFS. The former is difficult conceptually while the latter is difficult computationally. There are many ways to define the width spread. In this chapter, we define the broadness of the MFS as follows:

Suppose that α_1 and α_2 are two roots that satisfy the equation $f(\alpha) + c_f = 0$, with c_f being a constant and $\alpha_1 < \alpha_2$. The broadness at c_f of the MFS is defined as $B = \alpha_2 - \alpha_1$. This definition is also presented in Fig. 3. The deviation from the mono-fractal can be computed according to this Broadness measure since it posts a universal standard on the width spread (i.e. defined at level c_f). It is worth to point out the constant c_f can be adjusted empirically to insure that this measure is well defined for all analyzed signals. For this experiment, we found that $c_f = -0.2$ is optimum for broadness calculations. The choice of c_f is correlated with the choice of the q discretization and the inherent data characteristics. The discreteness may also produce difficulties in the computation of broadness because of the difficulty in obtaining exact roots of the equation $f(\alpha) + 0.2 = 0$ from the set of discrete values of α. To bypass this problem, we compute the minimum value of $|f(\alpha) + 0.2|$ with respect to α instead of directly solving the equation.

5.6 The Effects of Atmospheric Stability Conditions on the MFS

In this section, how atmospheric stability influences these three geometric attributes of the MFS (or GAMFS) for all the four flow variables is discussed. Towards this end, we computed the broadness, the mode, and left slope for each measurement run and each flow variable. Table 1 summarizes these GAMFS characteristics for u, v, w, and T.

Figure 21: Canonical features of the multifractal spectrum defining the geometric attributes (GAMFS), including the left slope, the Broadness measure, and the mode.

Table 10: Ensemble summary statistics of the three geometric components of the MFS for the three velocity components and temperature and for each stability class.

Variable	Stability		Left slope	Mode	Broadness
	Stable	Mean	0.341	0.375	0.631
		Std.	0.069	0.061	0.205
u	Neutral	Mean	0.348	0.409	0.612
		Std.	0.067	0.067	0.183
	Unstable	Mean	0.340	0.407	0.631
		Std.	0.053	0.062	0.180
	Stable	Mean	0.357	0.359	0.564
		Std.	0.113	0.095	0.202
v	Neutral	Mean	0.337	0.396	0.594
		Std.	0.046	0.053	0.151
	Unstable	Mean	0.343	0.399	0.632
		Std.	0.054	0.068	0.187
	Stable	Mean	0.343	0.369	0.612
		Std.	0.044	0.074	0.167
w	Neutral	Mean	0.354	0.378	0.589
		Std.	0.045	0.062	0.143
	Unstable	Mean	0.349	0.392	0.653
		Std.	0.057	0.073	0.183
	Stable	Mean	0.340	0.331	0.589
		Std.	0.073	0.111	0.184
T	Neutral	Mean	0.321	0.280	0.613
		Std.	0.066	0.066	0.241
	Unstable	Mean	0.371	0.413	0.623
		Std.	0.081	0.139	0.203

83

To investigate whether the left slope, the mode, and the broadness vary with stability class, ANOVA tests were conducted and the corresponding p-values are reported in Table 2. This analysis demonstrated that atmospheric stability significantly impacts:

    1) the mode of the MFS for all four variables.

    2) the broadness for v and w.

    3) the left slopes of T.

Table 11: p-values from an ANOVA test to determine whether atmospheric stability statistically impacts GAMFS for all four flow variables. Statistically significant effects are indicated by a p-value < 0.05

Variable	Left slope	Mode	Broadness
u	0.576	0.003	0.922
v	0.216	0.001	0.026
w	0.597	0.046	0.02
T	0	0	0.502

The fact that the MFS of all four variables is impacted by atmospheric stability suggests that the mode (or the scaling exponent) may be the most sensitive indicator among the three GAMFS. Recall that for $K41$ theory, the mode must be $1/3$. Here, we found that the mode of all three velocity variables and all stability classes exceed $1/3$. Only for unstable atmospheric stability conditions does the MFS mode for T exceed $1/3$. For neutral conditions, it appears that the mode of the MFS drops significantly below $1/3$ (i.e. high intermittency).

Beside the ANOVA tests, we also provide Box-plots to examine how stability alters the mode and broadness between two different stability classes (Fig. 4 and Fig. 5 respectively). In Fig. 4, we found that the modes corresponding to the stable regime are generally smaller

than those corresponding to the unstable regime and this appears general for all flow variables. For velocity, the main difference between the two stability classes is that surface heating contributes to the production of turbulent kinetic energy (TKE) for unstable conditions and the destruction of TKE for stable conditions. The smaller exponent for stable conditions indicates a higher irregular or intermittent phenomena - at least when compared to their unstable counterparts. Possible causes may be attributed to the fact that pockets of weak to no turbulence within the time series are interrupted by mechanical production of TKE thereby producing higher irregularity and more frequent small local scaling exponents.

Finally, the fact that the MFS mode for temperature significantly departs from its velocity counterpart for near-neutral and stable conditions suggests clear dissimilarities between inertial scale thermal and velocity fluctuations. These results are not surprising as several studies already suggested statistical dissimilarity between T and u fluctuations within the inertial subrange (e.g. see review by [*Warhaft,* 2000]). What is new here is that the quantification of these dissimilarities is conducted within the context of MFS. For example, the MFS mode for T is consistently smaller than the MFS mode of u, v, and w for near neutral and stable conditions, suggesting more "local irregularities" in the T time series (see Fig. 1). This MFS analysis further confirms the global scaling analysis conducted via higher order structure functions elsewhere [*Warhaft,* 2000], which argued that T is more intermittent than u. However, what is surprising is the fact that the largest difference between velocity and temperature MFS modes are for near-neutral conditions, when temperature acts as a passive scalar.

Figure 22: Box-plots of the MFS mode versus stability class for all four flow variables. In these plots, '+' indicates outlier, the length of the bars reflects the variability, and the lines in the middle of these bars correspond to the median values. The dashed line represents $K41$ theory.

Figure 23: Same as before but for the MFS Broadness.

5.7 Conclusions

Theoretical arguments based on the Navier-Stokes equations and statistical analysis of numerous experiments have shown that ASL turbulence, the dominant transport agent in momentum and heat transfer between the land surface and the atmosphere, is multifractal. How boundary conditions at the land surface, such as surface heating or shear stresses, impact the multifractal properties of ASL turbulence within the inertial subrange remains to be investigated. This study is among the first to address the problem of quantifying the effects of atmospheric stability on the multifractal spectrum (MFS) of turbulent flow variables u, v, w and T within the ASL inertial subrange. We used a wavelet-based estimator of the MFS in conjunction with a weighted least squares to improve the robustness of the MFS estimation. To efficiently compare the MFS for different variables and stability conditions, we defined a set of geometric attributes (GA) that describe the main features of the MFS (referred to as GAMFS). The GAMFS of u, v, w and T within different stability are then estimated and empirically compared using descriptive statistics. ¿From the ANOVA tests, we found that stability conditions impact all four variables in terms of the mode of MFS while broadness and left slope are less sensitive to the stability conditions. We also showed that the MFS mode for temperature significantly depart from their velocity counterpart for stable and even more so for near-neutral conditions suggesting clear dissimilarity in thermal and velocity inertial subrange statistical properties. This conclusion lends indirect support to simulation results that showed extra irregularities (or anomalous scaling) in small scale scalar concentration statistics originate from the "passive scalar mixing process" itself and not from the complexities in the turbulent velocity [*Shraiman and Siggia,* 2000].

While we showed that the certain components in GAMFS parameters in the inertial subrange vary with atmospheric stability class for velocity and temperature, several thorny issues remain to be explored in the future. For example, can the stability dependence of the GAMFS components guide new subgrid models for Large Eddy Simulations of the atmospheric boundary layers; assist in analytical high frequency spectral corrections to

turbulent flux measurements for different stability regimes; or aid in the development of low-dimensional dynamical models that describe the interaction between high frequency turbulence statistics and atmospheric stability?

¿From a methodological point of view, we proposed an approach that can detect shifts in the multi-scale properties of stochastic processes (e.g. MFS) with potential applications that go beyond ASL turbulence. For example, the GAMFS can shed some light on several on-going debates about the similarity between ASL turbulence and the price dynamics in the foreign exchange market (e.g. [*Mantegna and Stanley,* 1996]), or about the connection between precipitation formation and their MFS properties (e.g. [*Carsteanu et al.*, 1999]).

5.8 References

1. Addison, P. (2002). *The illustrated wavelet transform handbook: Introductory theory and applications in the science, engineering, medicine, and finance,* 353 pp, Institute of Physics Publishing, Philadelphia, USA.

2. Arneodo, A, E. Bacry, S. Jaffard, and J.F. Muzy (1998). Singularity spectrum of multifractal functions involving oscillating singularities, *Journal of fourier analysis and applications*. **4**, 159-174.

3. Audit, B, E. Bacry, J.F. Muzy, and A. Arneodo (2002). Wavelets based estimators of scaling behavior, *IEEE, Trans. in Information Theory*, **48**, 2938-2954.

4. Carsteanu, A., V. Venugopal, and E. Foufoula-Georgiou (1999). Event-specific multiplicative cascade models and an application to rainfall, *Journal of Geophysical Research-Atmospheres*, **104**(D24), 31611-31622.

5. Celani, A, and M. Vergassola (2001). Statistical geometry in scalar turbulence, *Phys. Rev. Lett*, **86**, 424-427.

6. Ellis, R. (1984). Large deviations for a general class of random vectors, *Ann. Prob.*, **12,** 1-12.

7. Frisch, U and G. Parisi (1985). Fully developed turbulence and intermittency. *Proceeding of Intenational School in Physics*.

8. Frisch, U. (1995). Turbulence, 296 pp, *Cambridge University Press*.

9. Gonçalvès, P., R.H. Riedi and R.G. Baraniuk (1998). Simple Statistical Analysis of Wavelet-based Multifractal Spectrum Estimation. *Proceedings of the 32nd Conference on 'Signals, Systems and Computers'*, Asilomar, Nov 1998

10. Katul, G, I. Hsieh and J. Sigmon (1997). Energy-inertial scale interaction for temperature and velocity in the unstable surface layer, *Boundary Layer Meteorology*, **82**, 49-80.

11. Katul, G, B. Vidakovic and J. Albertson (2000). Estimating global and local scaling exponents in turbulent flows using wavelet transformations, *Physics of Fluids*, **13**, 241-250.

12. Madndelbrot, B. et al. (1968). Fractional Brownian Motion, Fractional Noise and Applications, *SIAM review*, **10**, 422-437.

13. Mantegna, R.N. and H.E. Stanley (1996). Turbulence and financial markets, *Nature*, **383**, 587-589.

14. Meneveau, C. (1991). Analysis of turbulence in the orthonormal wavelet representation, *J. Fluid Mech.,* **232**, 469–520

15. Muzy, J.F., E. Bacry , and A. Arneodo (1993). Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method, *Phys. Rev. E*, **47**, 875.

16. Pesquet-Popescu, B. and J. Vehel (2002). Stochastic fractal models for image processing, *IEEE Signal Process. Mag.*, **September**, 48-62.

17. Riedi, R.(1999)., Multifractal processes, *Stoch. Proc. Appl.*, preprint.

18. Riedi, R. (2002). *Multifractal Processes, Long range dependence: Theory and Applications*. Eds. Doukhan, Oppenheim and Taqqu, (In-Press).

19. Seuret, S. and J. Véhel (2001). The local Hölder functions of a continuous function, *Project Fractales*, INRIA.

20. Shraiman, B. and E. Siggia (2001). Scalar Turbulence, *Nature (review)*, **405**, 639-46.

21. Sreenivasen, K. R. (1991). Fractals and multifractals in fluid turbulence, *Ann. Rev. Fluid Mech.*, **23**, 539-600.

22. Sreenivasan, K. R. and R. A. Antonia (1997). The phenomenology of small-scale turbulence, *Annu. Rev. Fluid. Mech.,* **29**, 435-472.

23. Véhel, J., E. Lutton, and C. Tricotn (1997). *Fractals in engineering from theory to industrial applications*, Springer Verlag.

24. Warhaft, Z. (2000). Passive scalars in turbulent flows, *Annual Reviews of Fluid Mechanics*, **32**, 203-240.

25. Yakhot, V., and Sreenivasan K. (2004). Towards a dynamical theory of multifractals in turbulence, *Physica A*, **343**, 147-155.

CHAPTER VI

MULITFRACTAL DISCRIMINATION MODEL FOR HIGH FREQUENCY PUPIL DATA

Multifractality present in high frequency pupil diameter measurements, usually connected with irregular scaling behavior and self-similarity, is modeled with statistical accuracy and discriminatory power. The Multifractal Discrimination Model (MDM) is proposed to determine the ocular pathologies based on the pupil diameter measurements during the interaction with graphical user interfaces. The MDM consists of two parts: (1) multifractal spectrum and (2) combined k-nearest-neighbor classifier. Analysis based on descriptive statistics and kernel density estimation is provided to obtain the statistical description of the inherited multifractality. The multifractal spectrum is used to discriminate pupil behavior measurements from four groups differing in ocular pathology. The spectrum broadness and maximum (Hurst exponent), two measures characterizing the multifractal spectrum of observations, are proposed as the distinguishing features among the groups. The combined nearest neighbor classifier is a model free and robust classifier, which is thought to be a appropriate classifier for the accurate prediction of ocular pathologies of the pupil diameter measurements.

6.1 Introduction

The discipline of human computer interaction (HCI) strives to evaluate and improve user performance and interaction with information technologies for many different users in many different contexts. Mental workload has long been recognized as an important component of human performance during interaction with complex systems (Gopher & Donchin, 1986),, such as computers. Notably, extreme levels of workload (high and low) have been shown predictive of performance decrements for different users under different conditions. To this end, this study examines the workload of visually-impaired users when performing

a computer-based task.

Previous investigations have examined the interactions of users with the visual impairments caused by age-related macular degeneration (AMD) (Jacko et al, 2001, 2003a, 2003b). AMD is one of the leading causes of visual impairment and blindness for individuals 55 years of age and older (The Schepens Eye Research Institute, 2002). Since the majority of information offered by computers is presented visually on a screen, these users are at a clear disadvantage. Research efforts directed towards the characterization of computer interaction for users with visual impairments can provide designers with the knowledge to better anticipate user needs in the development of information technologies.

AMD affects central, high-resolution vision, which has a large impact on the individual's ability to perform focus-intensive tasks, such as using a computer (The Center for the Study of Macular, 2002). Researchers have found that users with AMD tend to perform worse than normally-sighted users, as measured by performance metrics such as task times and errors, on simple computer-based tasks (Jacko et al, 2001, 2003a, 2003b). However, little work has been done to examine how these performance decrements are affected by increases in mental workload due to sensory impairments. Mental workload has long been recognized as an important component of human performance during interaction with complex systems (Gopher & Donchin, 1986).. Notably, extreme levels of workload (high and low) have been shown predictive of performance decrements. Measures of workload can be performance-based, survey-based, or physiologically assessed.

Pupil diameter is a well-documented, physiological measure of mental workload (see Loewenfeld (1999) and Andreassi (2000)). While research has shown pupillary activity to be related to changes in mental workload and task difficulty in a number of domains (Backs, 1992, Kahneman, 1973, Beatty, 1982, Marshall et al, 2002), the complex control mechanism of the pupil has made it difficult to extract the small, meaningful signals, related to changes in mental workload from the larger, overall noisy signal of pupillary activity (Barbur, 2003.). This being said, it is necessary to develop analytical techniques that can isolate

these small changes in pupillary behavior. A more comprehensive analysis of the pupil measurement may provide a solution to this problem and provide a unique characterization of interaction for individuals with AMD.

The development of analytical tools for high frequency data lends strong support to the analysis of the pupil signal. The high frequency pupil measurements share many important features with other extensively studied measurements, such as the turbulence (Shi et al, 2003), internet traffic (Abry & Veitch, 1998) and high frequency financial time series (Mandelbrot et al, 1997). This type of measurements are considered as selfsimilar signals, which are always connected with fractals. Fractal signals are usually divided into two classes – the mono-fractal signal and the multifractal signal. Although a multifractal signal model has been applied in many other fields, no previous work has been done with pupil measurements. This chapter addresses the modeling of pupil measurements, which are intractable using traditional statistical models and we propose a Multifractal Discrimination Model (MDM) to predict the ocular pathologies of the pupil measurements. We describe a multifractal spectral model to fit the pupil measurements and then extract the signal features from this model in order to discriminate the measurements coming from the different visual acuity groups. The challenge of this problem is due to the complexity of the pupil measurement (high-dimensional, irregular etc), the non-Gaussian distribution of the multifractal spectral characteristics and the difficulty of building a stable classifiers for multi-class data. The choice of taking the multifractal spectral characteristics as the classifier input is convinced by the descriptive statistics. To overcome the difficulty of building a multi-class classifier for non-Gaussian data, the combination of the predictors from individual model-free classifiers is adopted for more robust and accurate results.

The chapter is organized as follows. The dataset is described in Section 6.2. Section 6.3 includes the description of multifractal spectrum model and the features based on the multifractal spectrum. Discriminate analysis of the pupil measurements using the multifractal model is presented in Section 6.4. Section 6.5 provides conclusions.

6.2 Pupil-diameter Measurement

In this section, we briefly describe the datasets and how the data is preprocessed to fit the further analysis.

6.2.1 Datasets description

The equipment used to collect pupillary response data during this study was the Applied Science Laboratories (ASL) Model 501 head-mounted optics systems. Pupil size was recorded, at a rate of 60 Hz, for each participant over 105 trials of a computer-based task using a graphical user interface (GUI). A camera records the pupil image, which has been illuminated by a near-infrared beam that illuminates the interior of the eye. Pupil size is assessed as the number of pixels attributed to the pupil's image, which has been determined by real-time edge detection processing of the image. Actual pupil diameter measurements (in millimeters) are then calculated by multiplying each pixel value by a scaling factor that is based on the physical distance of the camera from the participant's eye.

The dataset is comprised of pupillary response data streams for 36 individuals, as described in Table 12. In this table, N refers to the number of individuals comprising this user group. Visual acuity refers to the range of Snellen visual acuity scores (assessed by ETDRS) of the better eye for participants of each group. AMD? refers to the presence (Yes) or absence (No) of this ocular disease in individuals within each group. Number of data sets refers to the number of 2048-length data sets that were obtained from the data streams for each group. For this study, data was collected from four groups of individuals, classified by visual acuity and the presence or absence of age-related macular degeneration (AMD). Visual acuity, an individual's ability to resolve fine visual detail, was assessed via the protocol outlined in the Early Treatment of Diabetic Retinopathy Study (ETDRS) (University of Maryland School of Medicine, 2002). The experimental protocol from this study is fully described in studies by Jacko and colleagues (Jacko, 2003a).

94

Table 12: Group characterization summary.

Group	N	Visual Acuity	AMD?	Number of Data Sets
Control	19	20/20 - 20/40	No	111
#1	6	20/20 - 20/50	Yes	59
#2	5	20/60 - 20/100	Yes	57
#3	6	20/100	Yes	124

6.2.2 Preprocessing

Studies of pupillary response are faced with the problem of how to remove blink artifacts. A blink generally lasts about 70-100 msec. (producing an artifact spanning 4-6 observations under 60 Hz sampling) during which time the camera registers loss and a pupil diameter of zero is recorded. Thus, it is generally relatively straightforward to detect and eliminate these contiguous zero observation artifacts from the record. However, on either side of a blink, one may also observe highly unusual recordings because the pupil may be measured inaccurately as the eye lid partially obscures the pupil. The result may be an impossibly small value for the pupil's size.

To insure that the analysis is conducted on pupil constriction or dilation and not on misleading discontinuities caused by blinks or partial blinks, one must either remove the blink observations from the data entirely or replace them with linearly interpolated values. Blinks (i.e., zero recordings) have been found to account for approximately 3-4% of all observations. Partial blinks account for another 1% of the total number of observations. The blink-removal procedure removes all observations having zero values (i.e., the blink) as well as any extreme values that occur within six additional observations on either side of the zero value (i.e., partial blinks). Figure 24 presents a preprocessed result of the typical measurements from a heathy subject (Control Group). Because of difficulty of collecting the measurements, especially from individuals with AMD, the original datasets were cut into equal length pieces to exploit their usage. Another reason of the segmentation is that the original measurements are not equally long. The segmentation is conducted after the 'Six Law' filtering, which was mentioned above. The dataset contains the sum of 351

Figure 24: Typical measurements with different resolutions from a heathy subject (Control Group)

Figure 25: Sample autocorrelation of the measurements in Figure 24

segments of measurements after segmentation and necessary outlier detection and each have the length of 2048. The distribution of the number of the segments among the four groups (Control, #1, #2 and #3) is reported in Table 12.

6.3   *Multifractality Features*

In this section, we discuss the concept of multifractality and the definition of the multifractal spectrum and analyze the features of the multifractal spectrum from the perspective of discrimination.

6.3.1   Scaling and multifractal spectrum

Many measurements encountered in nature, industry, science, etc. are characterized by complex scaling behavior, namely multifractality. Multifractals are signals that, instead of a single irregularity index H (usually the worst overall index of irregularity) typical of monofractality, possess a continuous range of Hurst exponents. Prime examples of multifractals are turbulence measurements where the deviation from the constant scaling, characterized by a Hurst exponent of 1/3 and called the Kolmogorov K41 law, is explained by multifractality of such measurements (Mandelbrot, 1968).

Wavelet-based energy spectrum is a commonly used tool to check the scaling behavior of the signal. This spectrum describes the second order statistics (i.e. variance) of the signal at different scales (frequency points). The linearity (or curvature) of this spectrum reflects the fractality of the signal and this connection could be utilized to the estimation of the Hurst exponent of the signal. The exact definition of the wavelet-based energy spectrum and its estimation could be found in Vidakovic (1999). Figure 26 shows the wavelet-based energy spectrum of a typical pupil measurement. This spectrum suggests that the fractal behavior exists in pupil measurement and the multifractal model is possible to recover the inherent features.

Figure 26: Wavelet-based energy spectrum for the measurements in Figure 24 The slope in the intermediate scales is found to be -1.9484 corresponding to the Hurst exponent of 0.4742. The hockey-stick effect in the finest two scales is caused by quantization and possible smoothing of high frequencies of the measuring instrument.

The measure of multifractality is given by multifractal spectra that describe the "richness" of the signal in terms of various Holder regularity indices. The term spectrum connotes the spectral decomposition of the signal into components characterized by their irregularity. Thus, multifractal analysis is not focused on the irregularity/self-similarity of the data set as measured by a single parameter, but rather on a measure of inhomogeneity of such a parameter. In recent years, the multifractal formalism is implemented with wavelet tools (Arneodo et al., 1998, Riedi, 2002) and hence could be efficiently used in practice. The wavelet-based multifractal spectrum is based on the local singularity strength measure:

$$\alpha(t) = \lim_{k2^j \to t} \frac{1}{j} \log_2 |d_{j,k}| \tag{72}$$

where $d_{j,k}$ is the wavelet coefficient at scale j and location k. It has been shown (Jaffard, 1995) that the wavelet coefficients can carry the scaling behavior of the process if the wavelet is more regular than the process and the local singularity strength measure (72) converges to the local Holder index the process at time t. As the name tells, $\alpha(t)$ indicates the oscillation of process at time t. Small values of $\alpha(t)$ reflect the more irregular behavior

at time t. It can be imagined that any inhomogeneous process has a collection of local singularity strength measures and their distribution $f(\alpha)$ formulates the multifractal spectrum. The detailed estimation procedure of the wavelet-based multifractal spectrum is directed to (Gonçalvès, 1998).

6.3.2 Features based on multifractal spectrum

Theoretically, the multifractal spectrum of fBm (a representative of mono-fractal) consists of three geometric parts: the vertical line, the maximum point and the right slope. The maximum point corresponds to the Hurst exponent and the vertical line is thought to be an inherent feature, which distinguishes fBm from the multifractal process. However, it is rare to obtain such a perfect spectrum in practice. Even for the simulated fBm, due to error of estimation, its spectrum may deviate from the theoretical form, as shown in Figure 27. Even with the lack of precise estimation of the spectrum, the deviation from the vertical line could be still utilized in the discrimination between the mono- and multi-fractal processes. In Figure 27, two type processes are presented in the multifractal spectra. One is the fBm and the other is the turbulence measurement, which is widely believed to be a multifractal process. Comparing with the turbulence measurement, the fBm is much closer to the vertical line and this closeness may be quantified by the left slope of the spectra. Another important difference between these two spectra is the width spread of the spectra. It is obvious that the width spread of the fBm is much smaller than that of the turbulence measurement.

Despite the existence of the estimation error, the spectrum can be approximately described by two slopes and one point without loss of the discriminant information. Alternatively, we can also approximate the spectrum by the left slope, the maximum point and the width spread. A typical multifractal spectrum is described as shown in Figure 28.

The left and right slopes can be obtained easily using the linear regression technique.

Figure 27: Multifractal Spectra for mono- (dash line) and multi-fractal (solid line) processes (The dotted line indicates the theoretical slope of the spectrum for monofractal process)

However, it is not as straightforward to define the width spread automatically. The difficulties are related to two aspects - one being how to locate the start and end points of the width spread, while the other is what to do with the discreteness of the spectrum. It is easy to see that the former is difficult conceptually, while the latter is computationally difficult. There are many ways to define the width spread. In this chapter, we give one definition of width spread and we name the width spread the *broadness* of the spectrum.

Definition 6.3.1 *Suppose that α_1 and α_2 are two roots which satisfy the equation $f(\alpha) + 0.2 = 0$ and $\alpha_1 < \alpha_2$, the the broadness of multifractal spectrum is defined as $B = \alpha_2 - \alpha_1$, where $f(\alpha)$ is the spectrum function in terms of Holder regularity indices α's.*

This definition is also graphically presented in Figure 28. The deviation from the monofractal could be fairly compared according to this Broadness measure since it posts a universal standard on the width spread. It is worth to point out the threshold value 0.2 used in this definition could be adjusted empirically in the practice analysis to insure that this measure is well defined for all analyzed signals.

As mentioned earlier, the discreteness may produce difficulties in the computation. The problem is that it may be hard to find the exact roots of the equation $f(\alpha) + 0.2 = 0$

100

among the discrete values of α's. To get around this, we try to find the minimum value of $|f(\alpha) + 0.2|$ with respect of α instead of solving the equation directly.

Applying our idea about extracting the spectral features from the pupil measurements, we obtain the broadness, Hurst exponent and left slope for each measurement. Table 13 summarizes the spectral characteristics of the pupil datsets that we are using in our study. We will use this result in Section 6.4.

Figure 28: Approximate description of the spectral characteristics

Table 13: Summary statistics of the multifractal spectral characteristics for our pupil datasets

Group		Left slope	Hurst	Broadness
Control	Mean	0.5053	0.4177	0.8591
	Median	0.4725	0.4153	0.7668
	Std.	0.1658	0.1517	0.4956
#1	Mean	0.3787	0.3561	0.8404
	Median	0.3701	0.3214	0.7266
	Std.	0.0738	0.1511	0.6796
#2	Mean	0.4049	0.4233	0.6989
	Median	0.3908	0.4104	0.6804
	Std.	0.1105	0.0985	0.1655
#3	Mean	0.484	0.3965	1.348
	Median	0.4608	0.3926	0.8562
	Std.	0.139	0.1723	1.1761

6.4 Pupillary Responses Analysis

As mentioned previously, we attempt to find the inherent features which can separate the measurements with different ocular pathologies from each other. The empirical evidence (e.g. wavelet-based energy spectrum) has shown that Pupil-diameter measurements possess self-similarity and fractality. Hence, it is natural to apply the multifractal spectra to discriminate these measurements.

We have discussed the features of multifractal spectra in section 6.3. The most important feature of the spectra is the maximum point, which corresponds to the Hurst exponent. The Hurst exponent is a measure of "roughness" of the self-similar process. The Hurst exponent coincides with the Holder regularity index, and signals with H close to 0 look quite irregular and intermittent while for H close to 1, the signals look smooth. Such an important property of the Hurst exponent enables us to explain the dynamics of pupillary behavior. Informally speaking, large values of Hurst exponent correspond to less dynamic changes in pupil size ("frozen eye") while low values of the exponent indicate bursty and frequent changes. Therefore, the Hurst exponent could discriminate the measurements. The boxplots of the Hurst exponents for the four groups are shown in Figure 6.4. According to this figure, the group #1 have exponents much lower than the control group, which reflects that the individuals from this group have more irregular pupillary responses than those from the control group.

As can be seen in Figure 6.4, the Hurst exponent could not completely discriminate the groups. This motivates us to introduce other discriminatory quantities. Another measure we just defined is the broadness, which is able to distinguish the deviation from monofractality. Broadness describes richness in the distribution of the Hurst exponent. Pupil-diameter measurements with narrow multifractal spectra are close to monofractals (i.e., the scaling is quite uniform over all scales). The boxplot of the broadness are given in Figure 6.4. It is very hard to tell the difference among the four groups. However, the last group #3

Figure 29: Boxplot for the Hurst exponent of multifractality.

Figure 30: Boxplot for the Broadness measure of multifractality.

significantly differs in terms of the broadness from other experimental groups (#1, #2). Group #3 has relatively high large broadness measure, which indicates that the pupillary responses of the individuals from this group deviates from monofractal much more than groups #1 and #2. Physiologically speaking, more change patterns of the pupil dynamics exist in group #3.

Neither the Hurst exponent nor the Broadness measure are able to achieve the complete discrimination separately. Thus, we need to increase the analysis into the 2D plane, analyzing the data with both measures simultaneously. Figure 31 presents the centroid points for the four groups. These four points look nearly evenly distributed on the plane. From this figure, we can see that the Hurst exponent from the control group is relatively large although it is not the largest. Only group #2 has a larger Hurst exponent than the control group. Comparing these two groups, we can tell that the Pupil-diameter measurement from the control group is further from monofractal than group #2 since the broadness measure of group #2 is the smallest. Therefore, we can claim that the pupillary responses of individuals from the control group is very smooth although the fractal properties are relatively inhomogeneous, which implies the causes of the regularity are quite rich. Group #1 is located on the very left-bottom side of the plane and hence it represents measurements with much more irregular dynamics and homogeneous fractal properties, which indicates the cause of the irregularity is relatively simple. Group #3 located near the top left side signifies that the measurements are quite irregular and have inhomogeneous fractal properties, which indicates the cause of the irregularity is not single.

To further address how the Hurst exponent and Broadness could be the discriminating measures, we estimate bivariate kernel densities of them for each group. The contours of these densities are given in the Figure 32. This figure includes all the information shown in Figure 31. The variability information of the two measures within each group, however, provides more discriminating features. It is easy to see that measures in group #2 are very compact while those in group #3 are dispersed.

Figure 31: Centroid points from bivariate measures:Hurst exponent and Broadness

Another important task in the analysis of these measurements is classifier building. Among the many candidates, the k-nearest-neighbor classifier is chosen because it is model free. The original datasets are divided into two parts, one of them is assigned to training set and the other is used to test the trained classifier. The training set includes a 90% randomly selected sample of each group from the whole datasets and the rest is taken to be the test set. To choose the nearest neighbor parameter k, the classifier is built as a learning process. The learning curve, which includes the test error and training error corresponding to different parameters k, is given in Figure 6.4. Although, relatively low training error could be achieved by choosing small k, the test error is too big for a practically useful classifier. To overcome these drawbacks, we adapt the model by combining techniques. Model combining is a technique of combining the predictions from different classifiers. The results have shown to be promising. For the details of this combining technique, the reader is directed to Xu (1992. The advantage of using model combining is due to its ability of overcoming the instability of the single classifier. In our study, the single k-nearest-neighbor classifier is not very accurate and robust according to Figure 6.4. By applying the model combining technique to these k-nearest-neighbor classifiers ($3 \leq k \leq 10$), the test errors get much smaller as we can see from Table 14. Although the combining rules do not make much

Figure 32: Joint density of both the Hurst exponent and Broadness

Figure 33: Learning Curve of K-Nearest Neighbor Classifier

difference, the result from mean-combining rule is shown to be optimal among the alternatives. Up to now, all analysis is done according to two features: the Hurst exponent and Broadness. To demonstrate how an additional measure may affect the classifier quality, we add the left slopes into the feature vectors and the result of classification design is reported in Table 15. It is apparent that both the test and training errors decrease a lot as the new feature is added (e.g. the test errors drops down about 6%).

Table 14: Error rate after combining the nearest neighbor classifiers

	rule	mean	median	max	min	majority voting
Training	mean	0.42	0.43	0.44	0.42	0.46
Errors	std. dev.	0.01	0.01	0.02	0.02	0.01
Test	mean	0.51	0.53	0.52	0.52	0.55
Errors	std. dev.	0.09	0.07	0.08	0.09	0.07

Table 15: Error rate after combining the nearest neighbor classifiers(adding slope feature)

	rule	mean	median	max	min	majority voting
Training	mean	0.407	0.414	0.417	0.401	0.432
Errors	std. dev.	0.013	0.012	0.015	0.015	0.013
Test	mean	0.446	0.450	0.439	0.459	0.475
Errors	std. dev.	0.051	0.045	0.048	0.057	0.050

6.5 Conclusions

The overreaching goal of this detailed analysis was to determine if individuals with different ocular pathologies exhibit quantifiable differences in their interaction with graphical user interfaces. These distinctions between classes of users can enable developers to design improved interfaces for more efficient and effective human-computer interactions. Pupillary behavior is an informative, yet complex, means of quantifiably assessing differences in the interaction behaviors of users.

Measurements of pupil diameter during task performance is one way to study the effects of mental workload on users. However, the inherent complexity of pupillary behavior requires that robust and valid measures should be developed to extract the meaningful components of the data stream in order to characterize those changes in pupillary behavior that distinguish changes in mental workload. In this way, the relative mental workload of users with different visual capabilities can be examined. These distinctions between user needs can be used to modify visual interfaces and interaction paradigms in order to best adapt information technologies for users with visual impairments.

In this chapter, we study how to incorporate characteristics of the multifractal spectrum into the modeling and discrimination of the pupil-diameter high frequency measurement. The multifractal process was validated to be appropriate in the analysis of the pupil-diameter measurements. By decomposing the spectrum into describable parts, the feature extraction is discussed to do further discrimination. The concept of the Broadness of a multifractal spectrum was defined. The analysis based on the Hurst exponent and Broadness measures gave distinguishable characteristics of the pupillary responses from the individuals with different visual acuity ranges. The model-free classification method, k-nearest-neighbor classifier, is applied with the model combining technique to build a robust and accurate classifier.

6.6 References

1. Abry & Veitch(1998).Wavlet analysis of long rang depedent traffic,*IEEE Trsanctions on Information Theory*, **44**, 2-15.

2. Andreassi, J. L. (2000). Psychophysiology: Human behavior & physiological response (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

3. Arneodo, A., Bacry, E., Jaffard, S., and Muzy J. F. (1998a). Singularity spectrum of multifractal functions involving oscillating singularities. *Journal of fourier analysis and applications,* **4(2)**, 159– 174.

4. Backs, R.W. & Walrath, L.C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, **23**, 243-254.

5. Barbur, J. L. (2003). Learning from the pupil - studies of basic measurement mechanisms and clinical applications. L. M. Chalupa & J.S. Werner (Eds.), Cambridge, MA: MIT Press.

6. Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin,* **91**, 377-381.

7. Beran, J. (1994). Statistics for Long Memory Processes, New York: Champan & Hall.

8. Coeurjolly, J.F.(2000). Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study *JSS*, **07**.

9. Gonalvs, P., Riedi, R. H. and Baraniuk, R. G. (1998). Simple Statistical Analysis of Wavelet-based Multifractal Spectrum Estimation, *Proceedings of the 32nd Conference on 'Signals, Systems and Computers'*, Asilomar.

10. Gopher, D., & Donchin, E. (1986). Workload- An Examination of the Concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (pp. 41-1 - 41-49). New York, NY: John Wiley and Sons.

11. Hastie, T. and Tibshirani, R.(1996) Discriminant Analysis by Gasssian mixtures, *J. Royal. Statist. Soc. B*, **58**, 155-176.

12. Hurst, H. E.(1951). Long-Term Storage Capacity of Reservoirs., *Proc. American Society of Civil Eng.,* **76(11)**.

13. Jacko, J. A., Scott, I. U. , Barreto, A. B., Bautsch, H. S., Chu, J. Y. M., & Fain, W. B.(2001). Iconic visual search strategies: A comparison of computer users with AMD versus computer users with normal vision. *Proceedings of the 9th International Conference on Human-Computer Interaction*, New Orleans , LA, August 5-10, 423-427.

14. Jacko, J. A., Barreto, A. B., Scott, I. U., Chu, J. Y. M., Vitense, H. S., Conway, F. T., & Fain, W. B. (2002). Macular degeneration and visual icon use: deriving guidelines for improved access. *Universal Access in the Information Society*, **1(3)**, 197-206.

15. Jacko, J. A., Scott, I. U., Sainfort, F., Barnard, L., Edwards, P. J., Emery, V. K., Kongnakorn, T., Moloney, K. P., & Zorich, B. S. (2003a). Older adults and visual impairment: What do exposure times and accuracy tell us about performance gains associated with multimodal feedback? *CHI Letters*, **5(1)**, 33-40.

16. Jacko, J. A., Scott, I. U., Sainfort, F., Moloney, K. P., Kongnako, T., Zorich, B. S., & Emery, V. K. (2003b). Effects of multimodal feedback on the performance of older adults with normal and impaired vision. *Lecture Notes in Computer Science (LNCS)*, **2615**, 3-22.

17. Jaffard, S. (1995). Local behavior of Riemanns function, *Contemporary Mathematics*, **189**, 287307.

18. Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

19. Loewenfeld, I. E. (1999). The pupil: Anatomy, physiology, and clinical applications (2nd ed.). Oxford, UK: Butterworth-Heinemann.

20. Madndelbrot, B. *et al* (1968). Fractional Brownian Motion, Fractional Noise and Applications, *SIAM review*, **10**, 422-437.

21. Mandelbrot, B.B., Calvet, L., Fisher, A.(1997). A Multifractal Model of Asset Returns. Working Paper. Yale University. Cowles Foundation Discussion Paper #1164.

22. Marshall, S. P., Pleydell-Pearce, C. W., & Dickson, B. T. (2002). Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts, In Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS '03).

23. Riedi, R. (2002). Multifractal Processes, Long range dependence: Theory and Applications, Eds. Doukhan, Oppenheim and Taqqu, (In-Press).

24. Shi, B., Vidakovic, B., Katul, G. and Albertson, J.(2003). "Assessing the Effects of Atmospheric Stability on Inertial Subrange Turbulence using Multiscale Approaches", Working paper, ISyE, Georgia Tech.

25. The Center for the Study of Macular Degeneration, University of California, Santa Barbara. (2002, January 25). Biology of AMD. Retrieved October 15, 2002, from http://www.csmd.ucsb.edu/faq/faq.html

26. The Schepens Eye Research Institute, Harvard Medical School, Harvard University. Macular degeneration: Your questions answered. Retrieved November 12, 2002, from http://www.eri.harvard.edu/htmlfiles/md.html

27. University of Maryland School of Medicine, Department of Epidemiology and Preventative Medicine. (1980). Early treatment diabetic retinopathy study, Manual of Operations, Chapter 12 (pp. 1-15). Baltimore, MD: ETDRS Coordinating Center.

28. Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.

29. Xu, L., Krzyzak, A., and Suen, C.Y. (1992). Methods of combining multiple classifiers and their application to handwriting recognition, *IEEE Trans. SMC*, **22**, 418-435.

CHAPTER VII

ASSESSING THE EFFECTS OF ATMOSPHERIC STABILITY ON TURBULENCE USING LOCAL AND GLOBAL MULTISCALE APPROACHES

The conceptual framework for modelling the inertial subrange is strongly influenced by the Richardson cascade, now the subject of various reinterpretations. One apparent departure from the Richardson cascade is attributed to boundary conditions influencing large-scale motion, which in turn, can directly interact with smaller scales thereby destroying the universal statistical scaling attributes of the inertial subrange. Investigating whether boundary conditions and inertial subrange eddies interact continues to be an active research problem in contemporary turbulence research. Using longitudinal (u), lateral (v), and vertical (w) velocities co-located with temperature (T) time series measurements collected in the atmospheric surface layer (*ASL*), we evaluate whether the inertial subrange is influenced by different stability regimes. The different stability regimes are proxies for different boundary conditions, as upper boundary condition force the mechanical shear and lower boundary condition force surface heating and buoyancy. The novelty of the present work lies in its combined use of global and local scaling properties (e.g. quasi-Hurst exponent, distributional properties of the wavelet coefficients, and Tsallis's thermostatic entropy measures) to assess whether atmospheric stability impacts both local and global inertial subrange scaling for velocity and temperature.

7.1 Introduction

The structure of turbulence in the inertial subrange has received much research attention over the past 50 years (Someria, 2001).. This strong interest is attributed to the possible emergence of universal or quasi-universal theories of turbulence (Frisch, 1995, Sreenivasan and Antonia, 1997, Frisch et al., 1998, Schraiman and Siggia, 2000), which is a

research area of interest in many fields (including finance). The inertial subrange encompasses eddies much larger than the viscous dissipation scales yet much smaller than the integral length scale (L_I) of the flow. The basic premise for the emergence of universal scaling is that large-scale anisotropic forcing characteristics (i.e. boundary effect conditions) are lost during the Richardson cascade process, thereby achieving local isotropy and universality at sufficiently smaller inertial scales (Biferale etc., 2001). However, several experiments and simulations over the past 2 decades suggest persistent effects of large scale anisotropies at these so-called inertial scales, even for very high Reynolds numbers and after many cascading steps (Warhaft, 2000). The departure from the so-called Kolmogorov (Kolmogorov, 1941) view of universal scaling and subsequent refinements (Kolmogorov, 1962) is now supported by numerous observations and theoretical arguments regarding the anomalous scaling in measured structure functions, particularly for passive scalars (Pumir and Shraiman 1995, Sreenivasan and Antonia, 1997, Celani et al, 2000,Warhaft, 2000, Antonov and Honkonen, 2001), and static pressure (Albertson et al., 1998).. The anomalous scaling is commonly attributed to short-circuiting of the energy cascade process due to the existence of organized large-scale features such as ramp-like structures, which are influenced by boundary conditions, and themselves directly influence small scale turbulence (Warhaft, 2000, Celani and Vergassola, 2001). Several theoretical arguments have been put forth to explain the apparent departure between experiments and $K41$. For example, Qian (1997, 1999) demonstrated that 1) K41 scaling can only be attained at infinite Reynolds number ($R_\lambda = \lambda \times \sigma_u / \nu$; where λ is the Taylor microscale, σ_u is the longitudinal velocity standard deviation, and ν is the kinematic viscosity), 2) the effects of finite R_λ on the structure function statistics decay slowly with increasing R_λ (e.g. Qian's work demonstrated that the decay is on the order of $R_\lambda^{-\mu}$ for third order structure function, and $\mu \leq 6/5$), and 3) the energy injection mechanism may be important. Recently, Gagne et al. (2004) indirectly confirmed all three findings experimentally using data collected for different flow types (i.e. different injection mechanism) and different R_λ.

113

Hence, it is clear that in addition to R_λ, surface heating (or cooling), and thereby atmospheric stability class, can be responsible for differences in kinetic energy injection (or removal in the case of stable flows) mechanism, thereby, introducing significant departure from $K41$.

To quantify whether boundary conditions (or energy injection mechanism) influence the statistical properties of the fine scale structure of the atmospheric surface layer (*ASL*), we utilize two methods: (i) a global scaling self-similarity index, and (ii) a scale-wise evolution of non-parametric estimates of probability densities in the wavelet domain. Each of these measures will be applied to turbulence time series measurements collected for three atmospheric stability regimes: unstable, near-neutral, and stable stratification. An ANOVA-type technique will then be applied to asses the significance of atmospheric stability factors on these two multi-scale measures. The main novelty of this work is in utilizing both global and local multi-scale measures to assess the effects of atmospheric stability on the statistical structure of the inertial subrange of several flow variables. However, before we describe these multi-scale methods, a brief description of the experiment and the data set will be provided.

7.2 Data

Time series measurements of longitudinal (u), lateral (v), and vertical (w) velocities and air temperature (T) were collected over a grass surface at the Blackwood Division of the Duke Forest near Durham, North Carolina. The measurements were collected on June, 12-16, 1995 at $5.2\ m$ using a GILL triaxial sonic anemometer. The measurements were sampled at $f_s = 56Hz$ and were subsequently divided into 19.5 minute intervals to produce $N = 65,536$ time measurement per flow variable per run. Our choice of 19.5 minute intervals for a run is a compromise between the need for stationary conditions at long time scales and maximizing the statistical sample size within a given run. The experiment resulted in an ensemble of 95 runs over a wide range of atmospheric stability conditions

ranging from near convective to stable atmospheric flows (Katul et al., 1997). Using the atmospheric surface layer stability parameter, $\xi (= z/L)$, these 95 runs were then classified into one of the three stability classes: stable, near neutral, and unstable. Here, L is the Obukhov length and z is the height from the ground surface. With this classification, 6 runs were collected in stable stability conditions, 23 runs were collected under near-neutral atmospheric stability conditions, and the rest were collected under unstable atmospheric stability conditions. We then selected 6 runs from the near neutral class and 6 runs from the unstable stability class such that the ensemble mean wind speed $(\overline{U})$ of these 6 runs were comparable to those 6 runs collected under stable atmospheric stability conditions. The scale-wise analysis only utilizes these 18 runs while the global analysis makes use of all 95 runs. The wind speed ensemble mean and standard deviation of the 6 runs for unstable, stable, and neutral conditions are provided in Table 16. Further details about the experimental setup, atmospheric stability conditions, the various measures used in the inertial subrange identification, and instrumentation details can be found elsewhere (Katul et al., 1997, Katul et al., 2001b) though we note that R_λ exceeded 10^3 for all runs (Katul et al., 1997).

Table 16: Mean and standard deviation of the 6 runs for unstable, stable, and neutral wind speed

	Unstable	Neutral	Stable
Mean($\overline{U}$)	1.8096	1.9782	1.814
Std.($\overline{U}$)	0.0789	0.2502	0.4678

7.3   Methods of Analysis

In this section, the methodologies for quantifying the effects of ξ on the local and global scaling properties of the inertial subrange are described. These methods include: (i) zero-crossing analysis to estimate the so-called quasi-Hurst exponent, and (ii) scale-wise parameter fitting of exponential power distributions, $\mathcal{EPD}$, in the wavelet domain. We also show

that the latter approach has theoretical links to the well-known Tsallis entropy measure. Linking these empirical exponential power distribution parameter to the Tsallis entropy measure is of interest here given the recent theoretical results that demonstrate how a non-linear Fokker-Plank equation can reproduce essential features about the inertial cascade via Tsallis entropy.

These analytic tools will be utilized to assess whether ξ affects the inertial subrange. We note that these measures are sensitive to different assumptions; hence, if ξ significantly affects the inertial subrange, it is likely to be resolved by these two methods.

7.3.1 Global Index: Quasi-Hurst Exponent

The Hurst exponent (after British hydrologist H. E. Hurst (1951)) is a measure of "roughness" of self-affine time series. There are many methods to estimate the Hurst exponent of a stationary process. The zero-crossing method (Coeurjolly, 2000) is based on counting the number of zero crossings Z_N, producing an estimate of the Hurst exponent given by:

$$\hat{H} = \frac{1}{2}\{1 + \log_2(1 \pm |\cos(\pi S_N)|)\}, \tag{73}$$

where $S_N = Z_N/(N-1)$ is an average number of zero-crossings for the differenced time series of length N, and sign + (alternativelly –) in $\pm$ is taken if the true exponent H is above (below) 1/2. Usually, it is not difficult to decide whether the true exponent H is above or below $1/2$ by observing the time series, unless the true H is close to 1/2. It was also demonstrated (Coeurjolly, 2000) that the $\hat{H}$ is asymptotically Gaussian for the fractional Brownian motion (fBm) process when the true Hurst exponent does not exceed $3/4$.

The estimation of $\hat{H}$ via (73) is valid only for time series with stationary increments. For time series lacking stationary increments, as may be the case in turbulence time series measurements, we call $\hat{H}$ the *quasi-Hurst* exponent. One of the attributes distinguishing turbulence signals from fBm is the distinction between the quasi-Hurst and Hurst exponents. Theoretically, the quasi-Hurst and Hurst exponents coincide for fBm (since fBm has stationary increments). We empirically confirmed this convergence using 1000 fBm paths

constructed with a true hurst exponent of 1/3. The resulting average $\hat{H}$ is 0.3331 with standard deviation of 0.06. The p-value of a t-test for the equality of $\hat{H}$ to $1/3$ is 0.9. Such a large p-value suggests that the $\hat{H}$ is an adequate estimator of the Hurst exponent for an fBm process. On the other hand, the quasi-Hurst exponent for turbulence time series measurement is quite variable and significantly exceeds $1/3$. This discrepancy may be utilized to diagnose how atmospheric stability alters the global scaling parameter. That is, by analyzing deviations of $\hat{H}$ from 1/3 (or the Kolmogorov scaling) for turbulence measured under different stability conditions, a logical basis for tracking how ξ impacts global scaling of inertial subrange turbulence can be developed.

7.3.2 Local Index: Evolutionary Models of Scale-Wise Empirical Densities of Wavelet Coefficients

Another method for assessing the effects of atmospheric stability on the inertial subrange is the sensitivity of the probability density function (pdf) of the wavelet coefficients of a given flow variable to variations in ξ. It is a case-verified fact that the scale-wise distribution of wavelet coefficients appear similar for a variety of signals and images. Typically, their empirical distributions are symmetric with a sharp peak at zero. Guided by this opulent evidence, Mallat (1989) proposed modelling a "typical" wavelet coefficient X by an exponential power family of distributions, $\mathcal{EPD}(\alpha, \beta)$, having the following pdf:

$$f(x) = Ke^{-(|x|/\alpha)^{\beta}}, \tag{74}$$

where α is the scale parameter, β is the shape parameter, and K is a normalizing constant given by $K = \beta/(2\alpha\Gamma(1/\beta))$. In the context of wavelet modelling, this approach is often referred to as Mallat's model and reduces to Gaussian for $\beta = 2$, to double exponential for $\beta = 1$ and trivially to uniform for $\beta = 0$. Hence, by investigating how ξ affects α and β at various scales, we can assess whether atmospheric stability impacts the scale-wise wavelet coefficients. Such coefficients, belonging to single scale, can be thought of as independent due to the de-correlation property of discrete wavelet transformations (DWT). To estimate

these pdf parameters in $\mathcal{EPD}(\alpha, \beta)$, a moment-matching method is adopted. The method is based on matching the theoretical moments, given by

$$E(|X|) = 2K\frac{\alpha^2\Gamma(\frac{2}{\beta})}{\beta} \quad \text{and} \quad E(X^2) = 2K\frac{\alpha^3\Gamma(\frac{3}{\beta})}{\beta}, \tag{75}$$

with their empirical counterparts: $\frac{1}{N}\sum|x_i|$ and $\frac{1}{N}\sum x_i^2$. Computation of these estimators is not difficult and involves solving one non-linear equation, and the empirical counterparts for two quantities in (75) can be readily obtained. While the Mallat distribution model appears arbitrary, we explore in the appendix its potential relationship with the Tsallis Maxent distribution, now receiving significant attention in turbulence research (e.g. Beck,2002). Beck (2002) demonstrated that the scaling properties of high Reynolds number turbulence, including the probability densities and scaling exponents can be well described using non-extensive measures.

From the appendix, we show that Mallat's model parameters can be related to the non-extensive Tsallis entropy parameters, which themselves are shown (e.g. Beck, 2002) to be connected to parameters from a nonlinear Fokker-Plank equation with random frictional force. In particular, we show in the appendix that when β is assumed from a particular prior, the marginal likelihood obtained from the $\mathcal{EPD}$ model in a Bayesian fashion is the Maximum Tsallis Entropy solution. Because Mallat's model parameters can be easily linked to geometric attributes of the wavelet coefficient probability distributions, much of our discussion will focus on the effect of atmospheric stability on Mallat model parameters (scale by scale). However, these model parameters do have an unambiguous physical interpretation within the context of Maximum Tsallis Entropy distribution.

7.3.3 Fitting the fBm to Mallat's Model

We generated 100 independent runs of fBm using the Wood-Chan algorithm and computed their wavelet coefficients. We normalized the coefficients at each scale to have a unit variance and fitted the Mallat model to these normalized coefficients. At each scale, we

obtained 100 shape parameters. Finally, we performed the nonparametric ANOVA analysis (Kruskal-Wallis test) to assess whether the scale index affects the shape parameters. We found a high p-value (=0.7180) that indicates the shape parameter is invariant to scale. This finding is crucial for turbulence data as increased intermittency with scale can be readily measured as scale-wise shift in the Mallat shape parameters. To illustrate, we show the fitted $\mathcal{EPD}$ of the scale-wise wavelet coefficients for the fBm in Fig. 1. Unlike turbulence, these pdf's are nearly Gaussian at all levels (scales), as theoretically expected.

Figure 34: The comparison between Tsallis Maxent distribution (dash), the empirical distribution (dash dot), and Mallat's model (solid) at the three finest scales of the discrete wavelet transformation. The left panel is the coarsest scale, while the finest scale is at the right panel. The fitted parameters for the three scales are $q = 1.3263, 1.3575, 1.3688$, $\beta = 0.8133, 0.6467, 0.5767$ respectively from the left to the right.

7.4 $\mathcal{EPD}$ Distribution as Tsallis MaxEnt Solution

The traditional entropy theory is based on Shannon's definition; more general measures, such as Renyi and Tsallis (sometimes called Tsallis-Havrda-Charvat (THC), or Havrda-Charvat entropy, since the definition of this entropy measure was first given in (Havrda and Charvat, 1967)) have also been proposed. Practical applications and theoretical implications of Tsallis-type entropy are active research areas in the physical sciences, especially turbulence (Beck, 2001, Beck, 2002, Arimitsu, T. and Arimitsu,2000, Arimitsu, T. and Arimitsu,2002a, Arimitsu, T. and Arimitsu,2002b, Ramosa et al., 2001). For example, *Tsallis thermostatistic* is based on Tsallis entropy which is a generalization of the Shannon (Boltzmann-Gibbs) entropy. The Tsallis entropy (Tsallis 1988, Tsallis et al., 1998) is given by

$$S_q = \frac{1}{q-1}\left(1 - \int_{-\infty}^{\infty} p(x)^q dx\right),\tag{76}$$

where $p(x)$ is the probability density of the microstate x of the system (in our case, the microstate refers to the individual velocity of the turbulence) and q is the non-extensive parameter (also regarded as a measure of the information incompleteness). The ordinary Shannon entropy is obtained as a special case when $q \rightarrow 1$. Tsallis' measure of entropy is more flexible than Shannon's due to its adaptive parametrization. We find that the maxent (maximum-entropy) probability distribution of a random variable X representing the turbulent measurements is constrained as follows:

$$\int_{-\infty}^{\infty} p(x)dx = 1, \text{ and}\tag{77}$$

$$\int_{-\infty}^{\infty} p(x)\epsilon_x dx = E,\tag{78}$$

121

where E is the energy content of x and should be a known constant, and ϵ_x's denote the energies at various microstates. The exact definition of ϵ_x in reality depends on the condition of the flow; roughly it can be taken as $\epsilon_x = x^2/2$ (x is assumed to be the velocity at this microstate). The general maximizing solution or Maxent distribution has the following form (Prato et al., 1999):

$$p(x) = \frac{1}{K_q} \left(1 + (q-1)\beta_1\epsilon_x\right)^{\frac{1}{1-q}}, \tag{79}$$

where

$$K_q = \int_{-\infty}^{\infty} \left(1 + (q-1)\beta_1\epsilon_x\right)^{\frac{1}{1-q}} dx \tag{80}$$

is the so-called partition function and $\beta_1 = 2/(5-3q)$ is a suitable inverse *temperature-like* variable. Interestingly, the value of q can be related to the properties of the Maxent distribution as follows.

If $1 < q < 3$, we could evaluate (80) with $\epsilon_x = x^2/2$ as

$$\begin{aligned}
K_q &= \int_{-\infty}^{\infty} \left(1 + \frac{1}{2}(q-1)\beta_1 x^2\right)^{\frac{1}{1-q}} dx \\
&= \left[\frac{\pi}{\beta_1(q-1)}\right]^{0.5} \frac{\Gamma(1/(q-1) - 1/2)}{\Gamma(1/(q-1))},
\end{aligned} \tag{81}$$

where $\Gamma(\cdot)$ represents the standard Gamma function. Furthermore, the second moment is calculated when $1 < q < 5/3$ as

$$EX^2 = \frac{2}{\beta_1(5-3q)}. \tag{82}$$

However, this second moment tends to infinity if $q \geq 5/3$ and the so-called heavy-tailed distribution, which is universally recognized as a basic characteristic of turbulence, is recovered in this case.

If $q > 3$, the integral defining K_q would diverge and hence the probability density function does not exist.

If $q \to 1$, the distribution in (79) converges to normal.

If $q < 1$, the distribution in (79) would resemble a cut-off distribution. Below, we briefly explore further links between Mallat's model and the Tsallis Maxent model.

Comparison of canonical forms of Mallat's and Tsallis models indicate close relationship between the $\mathcal{EPD}$ family and the Maxent distribution. Both of them include the uniform, normal, and Laplace distributions as special cases, as well as an infinite number of distributions with arbitrary variances and kurtosis. We demonstrate this link empirically using turbulent velocity time series. First, the turbulence time series is decomposed into three successive finest scales using discrete wavelet transformation. These scales are within the inertial subrange as discussed in (Katul et al., 2001a). We utilize the Daubechies 4-tap filter to ensure a balance of localization in time and frequency domains (i.e. a compromise between the Haar and Fourier bases). Next, we estimate the non-extensive parameter q for each scale using the relationship between q and the moments,

$$EX^m = \frac{1}{2^m} \prod_{j=0}^{m-1} \frac{5 + 2j}{4 + j - (3 + j)q},$$ (83)

where X is a Maxent distributed random variable. To give an estimator of q, we utilize the kurtosis κ, which is usually defined as:

$$\kappa = \frac{EX^4}{(EX^2)^2}.$$ (84)

We substitute the expressions for EX^4 and EX^2 evaluated by (83) and solve (84) with respect to q. The solution is:

$$q = \frac{7\kappa - 15}{5\kappa - 9}.$$ (85)

Once the parameter q is evaluated, the complete form of the normalized Tsallis Maxent distribution of unit variance, which is assumed to be the theoretical distribution of wavelet coefficients, is specified. The empirical pdf's of the scale-wise wavelet coefficients are compared with the corresponding Tsallis Maxent distribution in Fig. 2. In addition, we also fit Mallat's model for the scales-wise coefficients. For simplicity, the variances of these scales-wise coefficients have been normalized to unity.

The results shown in Fig. 2 suggest an almost perfect match among these three types of pdf's, especially at the tails. The large departures from the measured pdf around the center points is attributed to the "zoom effect" of logarithmic representation. The turbulence data we analyzed in Fig. 2 is for a longitudinal velocity run collected in stable atmospheric conditions. The pdf's of wavelet coefficients for the other flow variables v, w and T and for different stability regimes behave similarly as those in Fig. 2. We conclude that the marginal distribution of the turbulence wavelet coefficients at a fixed scale well match the maximum Tsallis entropy distribution. At the same time, the $\mathcal{EPD}$-model also fits well the empirical pdf. This empirical closeness demonstrates the inherent link between the $\mathcal{EPD}$ and the Tsallis Maxent distribution, explored next.

It is demonstrated that the Maxent solution is the marginal likelihood obtained from the $\mathcal{EPD}(\alpha, \beta)$ model when the prior on β is the Inverse Gamma. In other words, the Maxent solution is the scale mixture of $\mathcal{EPD}$ distributions with the inverse gamma as mixing distribution.

Consider a random variable X distributed as the Exponential Power Family with conditional pdf given by

$$f(x|\alpha, \beta) = K \exp\left(-(|x|/\alpha)^\beta\right).$$

(86)

Under the Bayesian paradigm, the scale parameter α is considered random and given a prior distribution. The marginal likelihood distribution of X given β is then obtained by integrating out α. Assume that $\lambda = \alpha^{-1}$ has prior distribution $\text{Gamma}\left(\frac{n}{2}, \frac{n}{2\lambda_0}\right)$ with density

$$g(\lambda) = \frac{1}{\Gamma(\frac{n}{2})} \left(\frac{n}{2\lambda_0}\right)^{\frac{n}{2}} \lambda^{\frac{2}{n}-1} \exp\left\{-\frac{n\lambda}{2\lambda_0}\right\},$$

(87)

with $\lambda_0 = E(\lambda)$.

Combining (86) and (87) we have

$$f(x|\beta) = \int_{-\infty}^{\infty} f\left(x|1/\lambda, \beta\right) g(\lambda) d\lambda.$$

(88)

There exists a closed form solution of this integral given by

$$f(x|\beta) = \frac{1}{K_q}(1 + (q-1)\lambda_1|x|^\beta)^{\frac{1}{1-q}},$$

(89)

where

$$q = 1 + \beta/(\beta n/2 + 1) \tag{90}$$

$$\lambda_1 = \frac{\beta}{1 + \beta - q}\lambda_0, \tag{91}$$

and

$$K_q = \int_{-\infty}^{\infty} (1 + (q-1)\lambda_1|x|^\beta)^{\frac{1}{1-q}} dx. \tag{92}$$

Hence, when the shape parameter is assumed from a particular prior, the marginal likelihood obtained from the model $\mathcal{EPD}$ in a Bayesian fashion is the Maximum Tsallis Entropy solution.

7.5   *Atmospheric Stability Effects on the Inertial Subrange*

In this section, we discuss the effects of atmospheric stability on the quasi-Hurst exponent (global scaling) and the parameters of the $\mathcal{EPD}$ distribution (local scaling). As mentioned before, the quasi-Hurst exponent ($\hat{H}$) is a global fractal index of a time series. Systematic variability of global scaling property with respect to stability factors can be used as indicator of interactions between boundary conditions and inertial subrange scaling.

For each stability class, a descriptive summary of the quasi-Hurst exponent of the four turbulent flow variables is reported in Table 17. These exponents are all calculated by the zero-crossing method. Interestingly, we found that for u and v, the scaling exponents exceed those computed from $K41$ (i.e. 0.333) and do not vary appreciably with stability class. In fact, the scaling exponents for u and v lead to fractal dimensions (in a mono-fractal model (Frisch, 1995)) on the order of 2.86. When compared to other intermittency models, these exponents ($0.42-0.44$) are somewhat higher than predicted by the She-Leveque (She, 1994) or K62 (Kolmogorov, 1962) models (0.39) hinting that external effects that are not entirely connected with classical internal intermittency buildup are at play.

Table 17: Mean and standard deviation of the quasi-Hurst exponent $\hat{H}$'s for the four flow variables. The numbers in the brackets are the statistics for the selected six runs for each stability regime described in the experimental setup and are reported here for reference.

Turbulence	Temperature		u		v		w	
$\hat{H}$	mean	std.	mean	std.	mean	std.	mean	std.
unstable	0.419 (0.417)	0.053 (0.043)	0.444 (0.446)	0.049 (0.026)	0.403 (0.404)	0.051 (0.045)	0.392 (0.328)	0.086 (0.032)
neutral	0.457 (0.431)	0.047 (0.044)	0.454 (0.449)	0.028 (0.040)	0.395 (0.399)	0.034 (0.060)	0.2515 (0.285)	0.059 (0.095)
stable	0.388 (0.388)	0.052 (0.052)	0.419 (0.419)	0.050 (0.050)	0.437 (0.437)	0.038 (0.038)	0.356 (0.356)	0.045 (0.045)
All	0.427 (0.412)	0.055 (0.047)	0.446 (0.438)	0.045 (0.040)	0.403 (0.413)	0.047 (0.049)	0.356 (0.323)	0.098 (0.067)

The exponents for the w time series for unstable and stable conditions are bounded by K41 and the She-Leveque (or K62) predictions. Surprisingly, for neutral flows (and for w only), the exponent was lower than K41. The difference between the global exponents for u and w clearly suggests anisotropy not consistent with $K41$ assumptions.

Upon comparing the global exponents of T and u, we find that the two exponents significantly differ for unstable conditions at the 99% level (using a student t-test) but not for near-neutral flows. This analysis lends support to a recent study by Aivalis et al. (2002) who found that for strongly unstable conditions, two scaling ranges emerge in higher order structure functions. One towards the smaller scales, which is the classical inertial range, while the other for larger scales in which the effects of buoyant convective production may still persist. Aivalis et al. (2002) attributed the buoyant-convective scaling to the active role of temperature in the production of turbulent kinetic energy (see also Katul and Parlange (1994) for further evidence of the active role of temperature using multiple scalars). We note that for stable flows, the u and T global exponents do not differ at the 99% level. This may be attributed to the small number of runs ($= 6$) for statistical significance testing, or alternatively, that the weak removal of turbulent kinetic energy for stable flows is occurring at larger scales than the injection of kinetic energy at near-convective conditions.

We tested the significance of the ANOVA components for the above quasi-Hurst exponents to quantify how the atmospheric stability condition affects the global fractal geometry of each variable. The results are summarized by the p-values, which are used again to test the equality of the quasi-Hurst exponents under three different stability conditions.

It is clear from Table 17 that the stability does not impact the global scaling of the longitudinal and lateral velocities significantly and this is supported by their relatively large p-values (0.22 and 0.16 respectively). Meanwhile, the air temperature and vertical velocity are significantly influenced by atmospheric stability and the corresponding ANOVA test p-values are quite small (0.0027 and 0 respectively). Hence, we conclude that atmospheric stability has a statistically discernable effect on the global scaling properties of the inertial

subrange of T and w.

Table 18: Mean and standard deviation of β's for all four flow variables and stability classes. Recall that $\beta = 2$ for a Gaussian process.

Turbulence		u		v		w		Temperature	
β		mean	std.	mean	std.	mean	std.	mean	std.
$j = 16$	unstable	0.8195	0.0986	0.8912	0.0824	0.8493	0.0731	0.701	0.1715
	neutral	0.8246	0.074	0.8138	0.1479	0.8296	0.1357	0.6565	0.1127
	stable	0.9241	0.1124	0.8642	0.0883	0.8183	0.1082	0.6315	0.1573
	All	0.8561	0.1031	0.8564	0.1087	0.8324	0.103	0.663	0.1433
$j = 15$	unstable	0.7204	0.1788	0.8044	0.122	0.7709	0.0804	0.7072	0.157
	neutral	0.7723	0.1094	0.7196	0.1457	0.7367	0.0845	0.8054	0.234
	stable	0.9068	0.1831	0.8394	0.173	0.8062	0.161	0.747	0.2365
	All	0.7999	0.1712	0.7878	0.1487	0.7713	0.1117	0.7532	0.2038
$j = 14$	unstable	0.7624	0.2286	0.8978	0.1768	0.7918	0.1356	1.1785	0.3625
	neutral	0.8636	0.1802	0.8776	0.1404	0.7587	0.0688	1.3546	0.4593
	stable	0.9563	0.1556	0.8808	0.1949	0.7892	0.0832	1.0551	0.3587
	All	0.8608	0.1967	0.8854	0.162	0.7799	0.0953	1.1961	0.3931

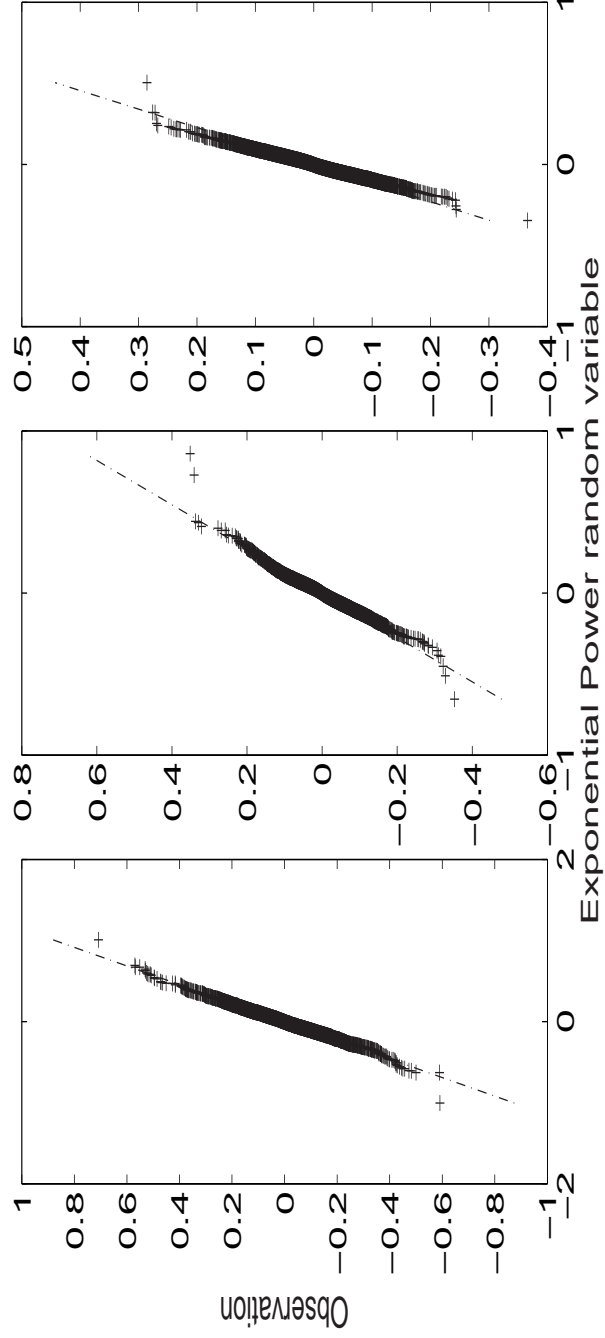We use Mallat's $\mathcal{EPD}$ model in the wavelet domain to further explore this finding at the finest scales. The first step is to determine the pdf parameters of the normalized wavelet coefficients in the $\mathcal{EPD}$ model at each scale. After estimating the parameters for each flow variable, we can assess the significance of atmospheric stability at a specific scale. In this analysis, we restrict the scale-wise calculations to the finest 3 (out of 16) detail levels to insure that these wavelet coefficients fall within the inertial subrange (identified as in Katul et al.(1997, 2001a)) and ensure minimum distortions due to Taylor's hypothesis.

Both pdf plots (Fig. 2) and Q-Q plots (Fig. 3) show that $\mathcal{EPD}$ model fits well the turbulence measurements. Thus, it is natural to regard the scale-wise parameters of this model as local scale indices of turbulence. We estimated these parameters for each of the runs within each stability class and tested their equality across different atmospheric stability classes. The results from this analysis should be taken with caution because the parameter comparisons across stability regimes for a specified scale index (j) assume that the eddy sizes represented by j are the same irrespective of stability (here - the larger the j the smaller the turbulent scale). While every effort was made to select the 6 runs across different stability classes with comparable mean wind speeds, the distortions due to Taylor's hypothesis may dependent on atmospheric stability itself. That is, a scale index of $j = 15$ may not precisely reflect identical eddy sizes for unstable, near neutral, and stable atmospheric stability because of differences in the turbulent intensities (despite equality in mean wind speed). What makes this analysis robust to such a limitation is that the use of orthonormal wavelet transforms means that the scale $j = 15$ is not precisely mapped onto one unique frequency or wavenumber (as is the case with Fourier spectra) but a distribution of frequencies (or wavenumbers) set by the wavelet basis. Hence, as long as the relative distortions attributed to the use of Taylors hypothesis across different stability classes (for the same mean wind) are much smaller than $\ln(2)$, the scale index j may represent the mean eddy sizes across different atmospheric stability regimes. Furthermore, we are restricting this local scale-wise analysis to the three finest scales known to be least affected

by distortions due to finite turbulent intensity.

Figure 35: The logarithms of the empirical density and Mallat's model at different scales for a typical fBm having a Hurst exponent of 1/3. Four panels show the pdf analysis at the four finest scales in the DWT. The top-left panel is the coarsest scale while the bottom-right panel is the finest scale. The β's at different scales are estimated from the data and are $1.7891, 1.9388, 1.9307$ and 1.9242 respectively. Note that these pdf's are approximately normal.

Figure 36: The Q-Q plots of the measured and modelled random variables at different scales in the wavelet domain for a typical flow variable u. The left panel shows the coarsest scale while the right panel shows the finest scale. The same turbulence measurement time series are used as in Fig. 34
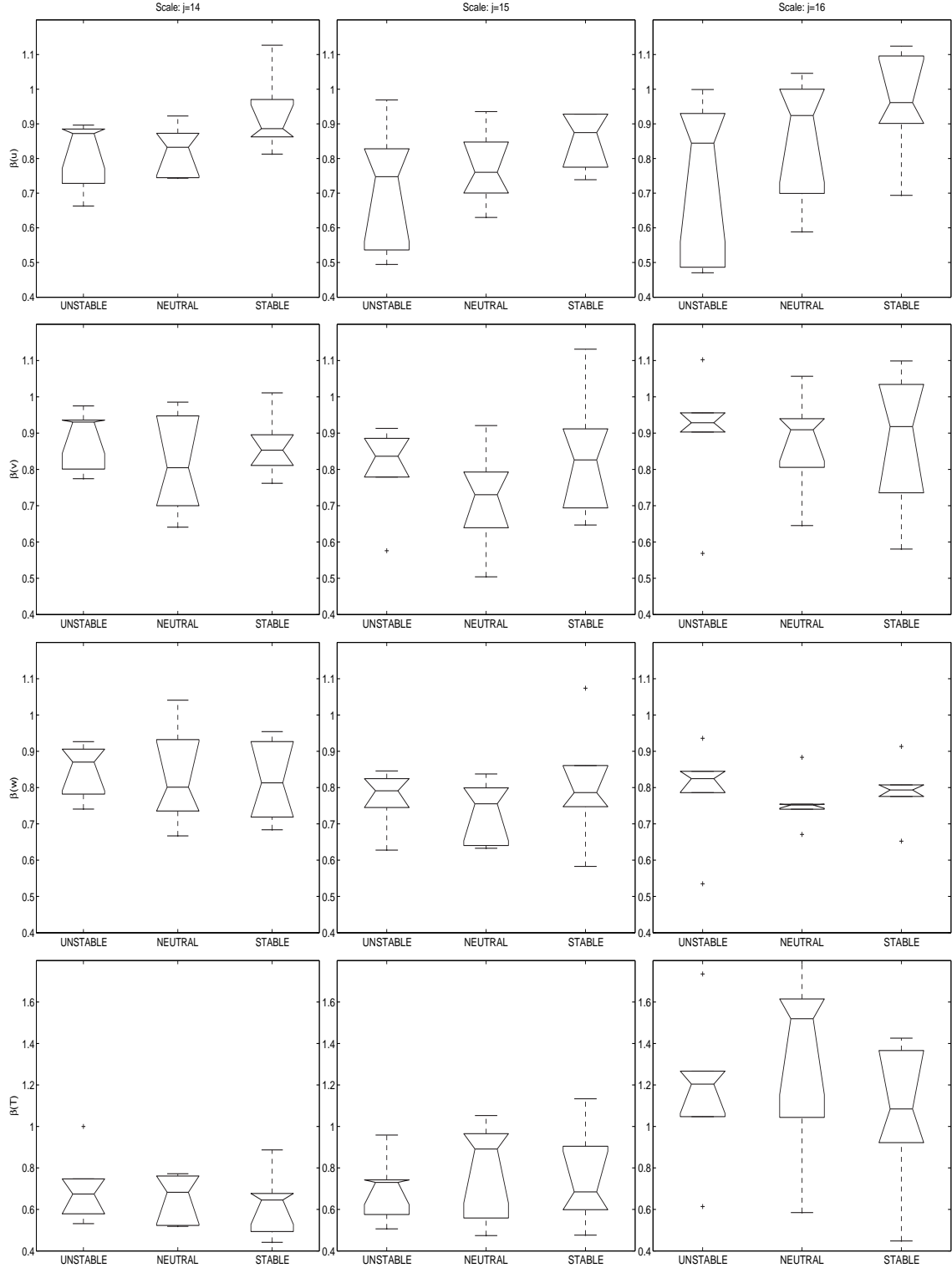
For each of the 6 runs within each stability class, we fitted the Mallat model for j=14,15, and 16 (out of 16) and all four flow variables. Table 18 gives the mean and standard errors of the shape parameters for each flow variable and stability class. Unlike the global analysis, we are not able to provide the ANOVA p-values since we only have six runs for each regime. This small number of runs makes the statistical inference difficult. As a remedy procedure, we provide the Box plots for these shape coefficients to compare the difference across the stability regimes in Fig. 4. Figure 4 demonstrates that β for all the flow variables, stability regimes, and inertial scales is far from Gaussian ($\beta = 2$), and is closer to a double exponential ($\beta = 1$). The lowest β (i.e. most heavy-tailed) is for temperature at the finest level (irrespective of stability regime). Furthermore, the fastest change in β with scale (i.e. a measure of increased intermittency with decreasing scale) is also for the temperature irrespective of stability class. Hence, the local analysis here clearly demonstrates scale-wise dissimilarity between temperature and u within the inertial subrange.

7.6 Discussion

To explore how atmospheric stability impacts the global scaling properties, we show Box-plots (Fig. 5) of the quasi-Hurst exponents for all the 95 turbulence measurements.

In the w and T cases the ranges at different stability regimes differ. Their ranges in stable and neutral regimes are totally separated and the median quasi-Hurst values in three regimes are different. Among the three regimes, the median quasi-Hurst values in the stable case tends to be closer to $H = 1/3$, which is the theoretical value based on K41. For u and v, it is also noted that the median quasi-Hurst values in the unstable and neutral are very close to each other (i.e., the neutral measurements tend to be similar on average as the unstable ones in terms of the global fractal characteristics), while the those in the stable regime are quite different.

Another important finding is that the ranges of quasi-Hurst exponents in neutral regime

134

Figure 37: Box plots of the shape coefficients β in $\mathcal{EPD}$ model across different stability regimes. The four rows correspond to u, v, w and T respectively from the top to the bottom. The three finest scales of wavelet coefficients are used here. The results in the left panels are from the coarsest level ($j = 14$) of wavelet decomposition while the right panels refer to the finest level ($j = 16$).

Figure 38: Box plots of the quasi-Hurst exponents $\hat{H}$ for four flow variables T, u, v and w, from the top-left to the bottom-right respectively.
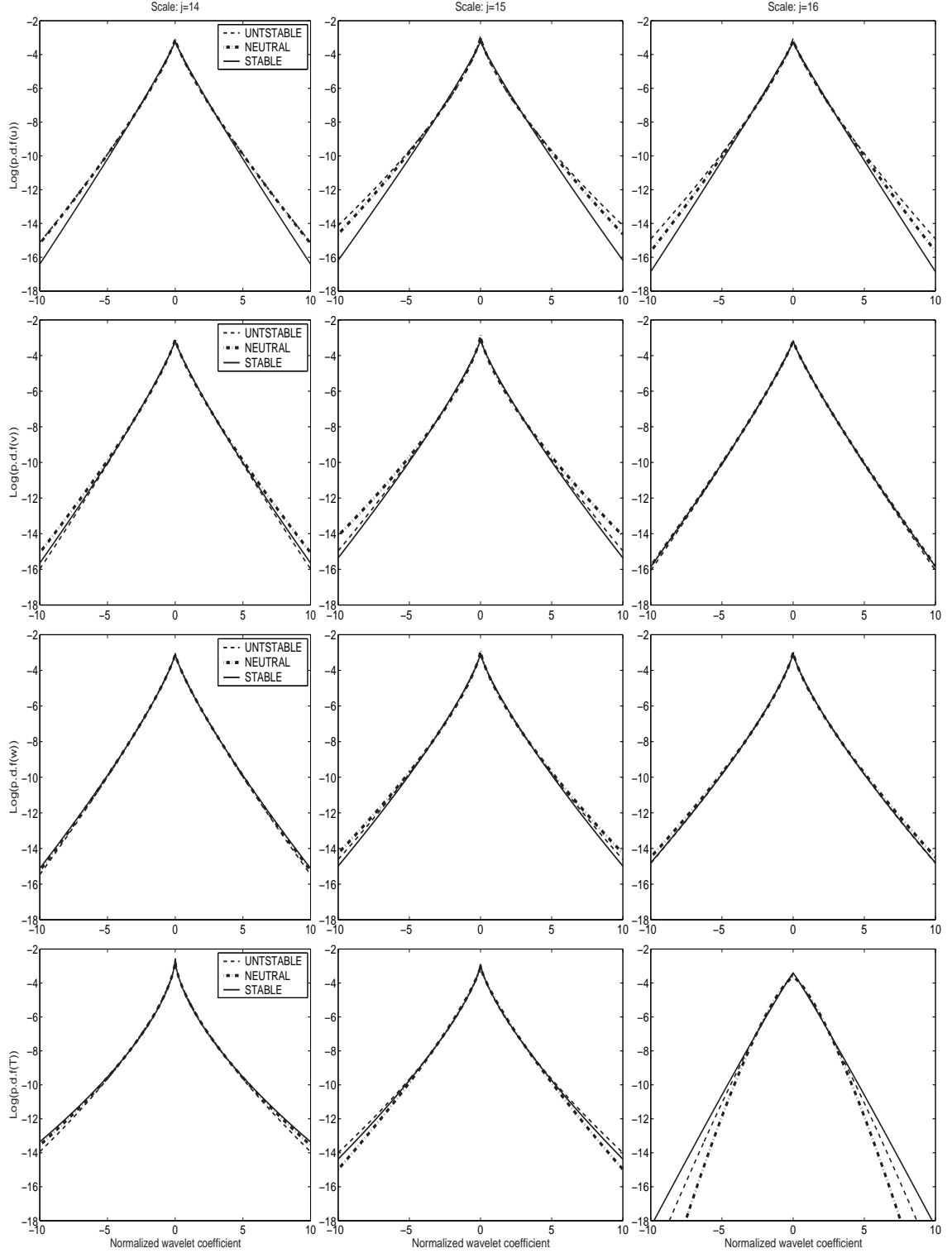
are much tighter for u than for the other flow variables. This implies that for neutral atmospheric stability conditions, the global fractal fluctuation of the horizontal velocity is much smaller than for the other variables. Also, the longitudinal velocity is characterized by greater quasi-Hurst exponents in the unstable and neutral regimes than in the stable one, which means that u is more intermittent in the stable regime. The fractal property of u in the stable regime tends to have more global fractal fluctuation than in unstable regime because the quasi-Hurst range is broader. However, this pattern seems opposite in the lateral velocity. This difference may be attributed to random shifts in wind direction. As an almost common phenomena for u, v and T, their quasi-Hurst values are generally larger than $H = 1/3$. However, the quasi-Hurst values of w is overall smaller than the other three flow variables, and, interestingly, the quasi-Hurst values in the neutral are much smaller.

The scale-wise comparison provides complementary evidence for the above arguments. The "average" pdf's (in Mallat's model) associated with atmospheric stability conditions of the four flow variables at the three finest scales are computed. To obtain such pdf's, we averaged the estimated parameters (including both the shape parameters and scale parameters) within a stability regime and use these averages as new parameters to specify the scale-wise pdf's for each flow variable (see Fig. 6). The plots in Fig. 6 suggest that u, v and T appear to be sensitive to atmospheric stability while w is not. The results here suggest that the local scaling parameters are more sensitive to the effects of atmospheric stability than the global scaling parameters.

7.7 Conclusions

This study assessed the effects of atmospheric stability on the inertial subrange structure of u, v, w, and T using both local and global measures in the wavelet domain. The global measure relied on a quasi-hurst exponent calculation while the local measure was based on the Tsallis thermostatistic entropy approach shown to be analogous to $\mathcal{EPD}(\alpha, \beta)$ family. The analysis here demonstrated the following:

Figure 39: The average logarithm pdf's associated with atmospheric stability conditions of the four flow variables (u, v, w, T) at first three finest scales. The four rows are corresponding to measurement u, v, w and T respectively from the top to the bottom. Three finest scales of wavelet coefficients are used here. The results in the left panels are from the coarsest level $(j = 14)$ of wavelet coefficients while the right panels refer to the finest level $(j = 16)$.

(i) Both $\mathcal{EPD}(\alpha, \beta)$ and the Tsallis thermostatistic entropy approach reproduces reasonably well the scale-wise velocity and temperature properties within the inertial subrange and for all stability classes, consistent with several recent studies.

(ii) The scale-wise analysis demonstrates that the distributional properties of the velocity and temperature within the inertial subrange are far from Gaussian and with tails even heavier than a double-exponential for all stability classes.

(iii) The local or scale-wise analysis clearly identified dissimilarities between temperature and velocity even within the same stability class.

(iv) The global measures were less sensitive to atmospheric stability than the scale-wise measures. In particular, the global measure identified the T and w components as the only flow variables whose inertial subrange has been statistically impacted by atmospheric stability. On the other hand, atmospheric stability clearly impacted the parameters of $\mathcal{EPD}(\alpha, \beta)$ for u, v and T. This analysis is in agreement with an earlier analysis by Katul et al. (2003) who demonstrated, via a Functional ANOVA (FANOVA), that atmospheric stability impacts the wavelet-based multifractal spectrum of u, v and T.

When conclusions (iii) and (iv) are taken together, it is clear that the inertial subrange of temperature is unambiguously impacted by atmospheric stability - both locally and globally. This analysis lends some support to the conclusions by Shraiman and Siggia (2000) who suggested that the statistical properties of 'scalar' turbulence are decoupled from those of the underlying velocity. That is, the added complexity in scalar turbulence (e.g., here added complexity refers to greater intermittency and overall sensitivity to boundary conditions) is not derived from the complexity in the velocity field, but also from the mixing process itself, and boundary conditions (or injection of energy). Finally, we note that the symbiotic use of global and local scaling analysis proposed here lends further confidence in the validity of the above conclusion for T.

7.8 References

1. Aivalis, K., Sreenivasan, K., Tsuji, Y., Klewicki, J., and Biltoft, C. (2002). Temperature structure functions for air flow over moderately heated ground, *Physics of Fluids*, **14**, 2439.

2. Albertson, J.D., Katul, G.G., Parlange, M.B., and Eichenger, W.E. (1998). Spectral scaling of static pressure fluctuations in the atmospheric surface layer: the interaction between large and small scales, *Physics of Fluids*, **10**, 1725-1732.

3. Antonov, N.V. and Honkonen, J. (2001). Anomalous scaling in two models of passive scalar advection: Effects of anisotropy and compressibility, *Phys. Rev. E*, **63**, 1-7.

4. Arimitsu, T. and Arimitsu, N. (2000). Analysis of fully developed turbulence in terms of Tsallis statistics, *Physical Review E*, **61**(3), 3237.

5. Arimitsu, T. and Arimitsu, N., Tsallis, C. (2002). statistics and Turbulence, *Chaos, Solitons and Frctals*, **13**, 149.

6. Arimitsu, T. and Arimitsu, N. (2002). PDF of velocity fluctuation in turbulence by a statistics based on generalized entropy, *Physica A*, **305**, 218.

7. Beck, C. (2001). On the small-scale statistics of Lagrangian turbulence, *Physica A*, **287**, 240.

8. Beck, C. (2002). Generalized statistical mechanics and fully developed turbulence, *Physica A*, **306** 189-198.

9. Biferale, L., and Vergassola, M. (2001). Isotropy vs anisotropy in small-scale turbulence, *Physics of Fluids*, **13**, 2139-2141.

10. Celani, A., Lanotte, A., Mazzino, A. and Vergassola M. (2000). Universality and saturation of intermittency in passive scalar turbulence, *Phys. Rev. Lett*, **84**, 2358-2388.

11. Celani, A., and Vergassola M. (2001). Statistical geometry in scalar turbulence, *Phys. Rev. Lett*, **86**, 424-427.

12. Coeurjolly,J.F. (2000), Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study *JSS* **07**.

13. Fan, J. and Lin, S.K. (1998). Test of significance when data are curves, *JASA*, **93**, 1007-1021.

14. Frisch, U. (1995). *Turbulence*, Cambridge University Press, 296 pp.

15. Frisch, U., Mazzino, A., and Vergassola, M. (1998). Intermittency in passive scalar advection, *Phys. Rev. Lett*, **80** 5532-5537.

16. Gagne, Y., Castaing, B., Baudet, C., and Malécot, Y. (2004). Reynolds dependence of third-order velocity structure functions, *Physics of Fluids*, **16**, 482.

17. Giostra, D. Cava, S. Schipa (2002), Structure Functions in a wall-turbulent shear flow, *Boundary Layer Meteorol.*, **103**, 337-359.

18. Gonçalvès, P., R. Reidi, and R. Baraniuk (1998). A simple statistical analysis of wavelet-based multifractal spectrum estimation., *Proceedings of Asilomar 32nd Conference on Signals, Systems and Computers, Monterey.*

19. Havrda, M.E. and Charvat, F. (1967). Quantification method of classification processes: Concept of structural α-entropy. *Kibernetica,* **3**, 30-35.

20. Hurst, H. E. (1951). Long-Term Storage Capacity of Reservoirs., *Proc. American Society of Civil Eng.,* **76(11)**.

21. Katul, G.G., and Vidakovic , B. (1996). The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering, *Boundary Layer Meteorology*, **77**, 153-172, 1996.

22. Katul, G.G., Hsieh, C.I., and Sigmon, J. (1997). Energy-inertial scale interaction for temperature and velocity in the unstable surface layer, *Boundary Layer Meteorology*, **82**, 49-80.

23. Katul, G. and Parlange, M. (1994). On the active role of temperature in surface layer turbulence, *Journal of the Atmospheric Science,* **51**, 2181.

24. Katul, G.G., and Vidakovic, B. (1998). Identification of low-dimensional energy-containing/flux transporting eddy motion in the atmospheric surface layer using wavelet thresholding methods, *Journal of the Atmospheric Sciences* , **55**, 377-389.

25. Katul, G.G., Vidakovic, B., and Albertson, J.D. (2000). Estimating global and local scaling exponents in turbulent flows using wavelet transformations, *Physics of Fluids*, **13**, 241-250.

26. Katul, G.G., Lai, C. T., Albertson, J.D., Vidakovic, B., Schafer, K., Hsieh, C.I. and Oren, C.I. (2001). Quantifying the complexity in mapping energy inputs and hydrologic state variables into land-surface fluxes,*Geophysical Research Letters*, **28**, 3305-3307.

27. Katul, G.G., Vidakovic, B., and Albertson, J.D. (2001). Estimating global and local scaling exponents in turbulent flows using wavelet transformations, *Physics of Fluids*, **13**,241.

28. Katul, G.G., Angelini, C., Canditiis, D.D., Amato, U., Vidakovic,B. and Albertson, J. (2003). Are the effects of large scale flow conditions really lost through the turbulent cascade?, *Geophysical Resrach Letters*, **29**.

29. Kolmogorov, A. N. (1941). Local structure of turbulence in an incompressible fluid for very large Reynolds numbers, *Dokl. Akad. Nauk. SSSR*, **30**, 299-303.

30. Kolmogorov, A. N. (1962). 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**, 82-85.

31. Mallat, S.G. (1989). A theoy for multiresolution signal decomposition: the wavelet reprentation. *IEEE Trans. on Patt. Anal. Mach. Intell.*, **11(7)**, 674-693.

32. Meneveau, C. (1991). Analysis of turbulence in the orthonormal wavelet representation, *Journal of Fluid Mechanics*, **232**, 469-520.

33. Meneveau, C, Sreenivasan, K. R., Kailasnath, P., and Fan, M. S. (1990). Joint multifractal measures: Theory and applications to turbulence, *Physical Review*, **41**, 894-913.

34. Meneveau, C. and Sreenivasan, K. R. (1991). The multifractal nature of turbulent energy dissipation, *Journal of Fluid Mechanics*, **224**, 429-484.

35. Nathan, R., Katul, G.G., Horn, H.S., Thomas, S. M., Oren, R., Avissar, R., Pacala, S. W., and Levin, S. A. (2002). Mechanisms of long-distance dispersal of seeds by wind, to appear, *Nature*.

36. Pumir, A., and Shraiman, B. I. (1995). Persistent small scale anisotropy in homogeneous shear flows, *Physical Review Letters*, **75**, 3114-3117.

37. Prasad, R. R., Meneveau, C. , and Sreenivasan, K. R. (1988). Multifractal nature of the dissipation field of passive scalars in fully turbulent flows *Physical Review Letters*, **61**, 74-77.

38. Prato,D. and Tsallis,C. (1999). Nonextensive foundation of Levy distributions, *Phys. Rev. E*, **60**, 2398.

39. Qian, J. (1997). Inertial range and the finite Reynolds number effect of turbulence, *Phys. Rev. E*, **55**, 337.

40. Qian, J. (1999). Slow decay of the finite Reynolds number effect of turbulence, *Phys. Rev. E*, **60**, 3409.

41. Ramosa, F.M., Rosaa , R.R., Netoa, C.R. *et al* (2001). Non-extensive statistics and three-dimensional fully developed turbulence *Physica A*, **295**, 250.

42. Schraiman, B. I., and Siggia, E. D. (2000). Scalar Turbulence, *Nature*, *405*, 439-446.

43. She, Z. and Leveque, E. (1994). Universal scaling laws in fully developed turbulence, *Phys. Rev. Lett.,* **72**, 336.

44. Someria, J. (2001). Unweaving the whirls, *Nature*, **413**.

45. Sreenivasan, K. R., and Antonia, R. A. (1997). The phenomenology of small scale turbulence, *Annual Reviews of Fluid Mechanics*, **29**, 435-472.

46. Tsallis, C. (1988). Possible generalizatopn of Boltzmann-Gibbs statistics, *J. Statist. Phys.* **52**, 479-487.

47. Tsallis, C., Mendes, R. S. and Plastino, A. R. (1998). NO TITLE, *Physica A*, **261**, 534.

48. Warhaft, Z. (2000). Passive scalars in turbulent flows, *Annual Reviews of Fluid Mechanics*, **32**, 203-240.

CHAPTER VIII

CONCLUSIONS

The dissertation is concluded as follows.

- Global and local scaling measures (e.g. quasi-Hurst exponent, distributional properties of the wavelet coefficients, and Tsallis's thermostatic entropy measures) are effective in quantifying the differences (similarities) between self-similar processes. Quasi-Hurst provides an global measures of self-similarities. Mallat model and Tsallis Maxent model examine the local behavior and it has been shown that these two models are related to each other in a Bayesian fashion. These measures are used to assess whether atmospheric stability impacts both local and global inertial subrange scaling in the the atmospheric surface layer and reveal some important facts about the turbulence.

- A Weighted Least Squares (WLS) scheme in wavelet domain to estimate the multifractal spectrum (MFS) is derived. This scheme is shown to minimize the heteroskedastic effects which is inherent because the sample variances of the wavelet coefficients depend on the scale. We also derived the discriminative measures including the left slope (rise), Hurst exponent (maxima), and broadness. These summary measures are applied to the MFS of velocity and temperature time series collected in the atmospheric surface layer for a wide range of atmospheric stability conditions. We successfully quantify the difference of turbulence series collected at different stability conditions. In a similar spirit, we develop a multifractal discrimination model (MDM) to discriminate the HCI user group using the measures of the left slope (rise), Hurst exponent (maxima), and broadness in MFS estimated from the high frequency

pupil size data. The discrimination is supported by some physiological interpretations.

- A combined k-Nearest-Neighbor (Comb-k-NN) classification model is proposed to address the inhomogeneous property of discriminative vectors. This method is proposed to minimize and stabilize the misclassification rate in training and test sets with the goal of improving classification accuracy. It is also justified by a Bayesian paradigm. We use Comb-k-NN as ingredients in the classification problem for the HCI users based on high frequency pupil size data. Reasonable accuracy has been achieved.

- We propose a multiscale Schur monotone (MSM) measure as summaries of self-similar processes. MSM measure is derived to characterize the disbalance properties of the data distribution at different frequency scales. In contrast to the global disbalance measures, the MSM measure carries information not only about the disbalance characteristics of the data, but also about its correlation structure. Thus, the MSM summary is more likely to be more sensitive to the differences in dynamics between different self-similar processes than any other single measure, such as correlation and global Schur Monotone measures. MSM measures also allow the adaptive summary of the process through wavelet basis selection and the choice of scale resolutions.

VITA

Bin Shi is a PhD candidate in the school of Industrial and Systems Engineering (ISyE). He was born in Jan, 1977. After finishing the high school in a small town of southern China, he went to Nanjing for his college. He graduated with highest honors from Southeast University (SEU) in 1999 and obtained his bachelor degree in Manufacturing. He was admitted to the graduate school of SEU with the entrance exam waived because of his proven academic excellence. In 2001, he decided to take the challenging journey and accepted the admission of the PhD program of ISyE at Georgia Tech. He was awarded a Master degree in operations research at Georgia Tech in 2003. His research interests include applied statistics, health informatics, financial modeling, manufacturing and geoscience.