Applying the Multi-Fractal formalism for extracting sub-LUN skew

D Gueorguiev Jan 7th , 2012

**Introduction**

The distribution of IOs along a LUN may have multi-fractal nature. The goal of this report is to explore the Wavelet Transform Modulus Maxima (abbrev. *WTMM*) method and the Histogram method for obtaining the singularity spectra of certain workloads. Also we will be looking to obtain an approximate closed form expression for the singularity spectra thereby estimating the Sub-LUN skew for various workload mixtures.

By obtaining the *multifractal singular spectrum* and *the multifractal singular exponents* of the generalized LUN workload we will be looking to approximate the sub-LUN skew with a parametrized closed form expression. Here the parameters are LUN size, write percent and IO size.

The steps in this process are as follows:

1. Write a code which implements the Histogram and WTMM methods to obtain singular spectrum and singular exponents of the fractal signal a.k.a *Devil staircase*.
2. Verify the computation process against a fractal distribution with known analytical expression for the singular spectrum and singular exponents
3. Create a module which parses btp/trace workload and transforms it into a Devil staircase signal which can be supplied as an input into the multifractal analysis code
4. Produce singular spectra for various LUN workloads with different LUN size, write percent, IO size, and interval duration
5. Create a closed form expression fitting the produced spectra for various values of the LUN size, write percent , IO size and trace duration.
6. Handle traces with significant number of sequential IOs. We hope that such workloads can be modeled as non-everywhere singular fractal signals i.e. fractal signals which have non-fractal component. Those signals can be analyzed relatively easily by the WTMM formalism (see [1]).

This report discusses step 1 and 2 of the process above which are completed.

**What is the multifractal singular spectrum of LUN IO workload?**

The singular exponent is a measure of the density of the IOs in particular area of the LUN or disk. *(give an example)*

The multifractal singular spectrum represents the distribution of the IO clusters with various densities .

**Histogram method**

*( provide brief discussion about the method and how it is used to extract skew information )***Wavelet transform modulus maxima method**

*( provide brief discussion about the method and how it is used to extract skew information )*

Figure 0.1: Simplified LUN IO layout. L – the LUN size; B – the block size

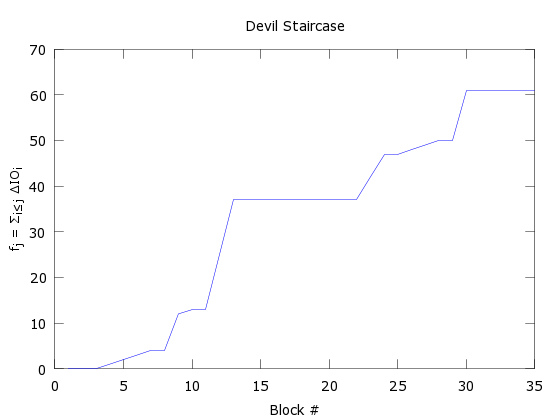


Figure 0.2: Devil staircase for the workload on Figure 1

**Generalized Cantor Set**

To illustrate the WTMM and the histogram methods and their application we will consider Generalized Cantor set as a signal source. Starting with a line of unit length subdivided into *b* sectors we discard sectors from it. For each of the present sectors we apply the same step recursively. Then on the -th iteration we will have a set of sectors each of length interspersed with holes between them. With each sector it is associated a non-zero measure obtained in the following way – at the first iteration the -th sector is assigned a value , . At the second iteration this sector is subdivided into sectors the measure of which is formed as , and so on. Thus on the -th iteration the measure of a sector takes the form , . The holes can be represented as sectors with zero measures. Then the fractal set can be uniquely characterized by its generating vector giving the measure distribution on the first iteration where can be zero in which case it indicates a hole at that location.

Figure 0.3: Generalized Cantor set with generating vector [p1, 0, p2, p3, 0, p4, 0, p5] . Here b = 8 and k = 3.

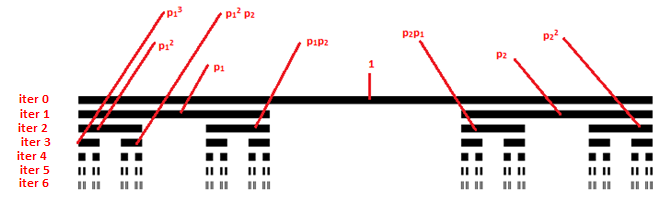


Figure 0.4: Cantor set with generating vector [p1 0 p2]. On the sixth iteration the fractal object is subdivided into number of sectors with non-zero measure.

**Examples**

Example 1 – generalized Cantor set with generating vector [p1 0 p2] where the weights are p1=0.6 and p2=0.4, with base b = 3. The fractal set is generated with 7 iterations, and there are total of N = 2187 sectors. The smallest feature is μmin = 1.6384 x 10-3 and the largest feature is μmax = 27.994 x 10-3 .

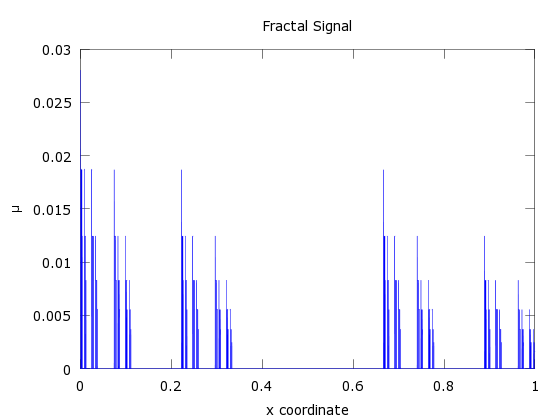


Figure 1.1: The fractal signal for Generalized Cantor set with generating vector [0.6 0 0.4]. To each sector of the fractal object it is prescribed a measure value which has the form where *n1*+*n2*=7

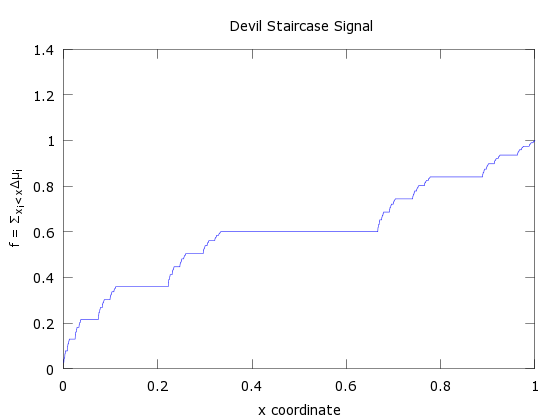


Figure 1.2 The Devil staircase signal for Generalized Cantor set with generating vector [0.6 0 0.4].

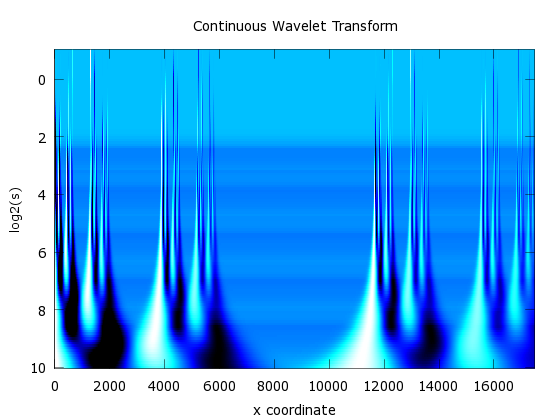
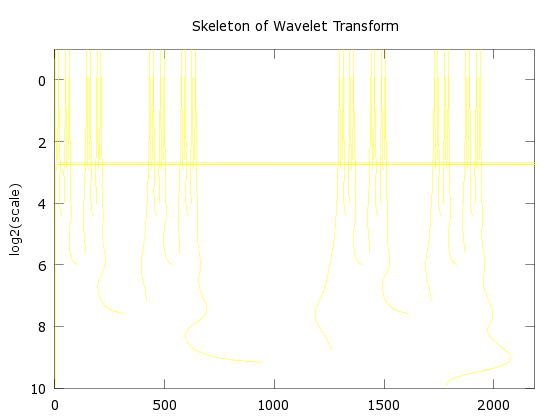
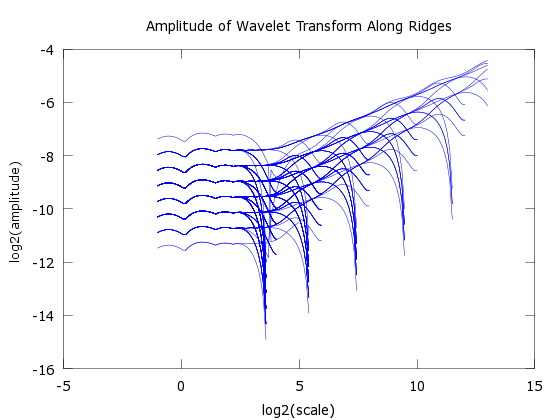
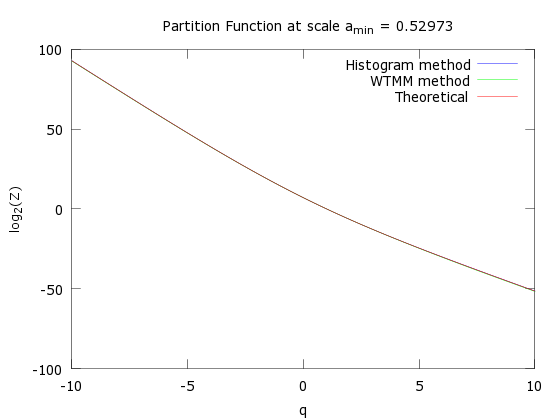
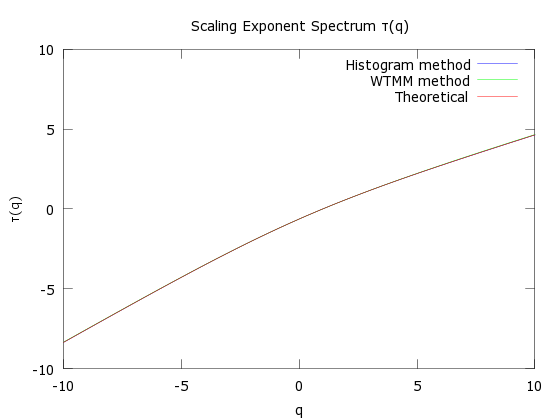


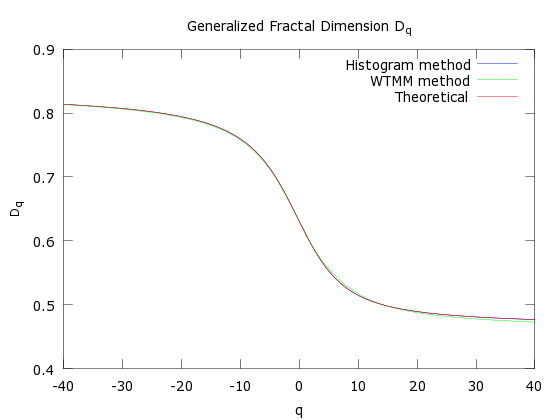
Figure 1.3 The continuous wavelet transform for Cantor set with generating vector [0.6 0 0.4].

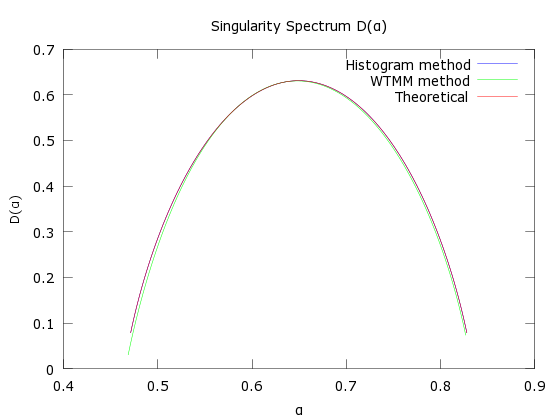


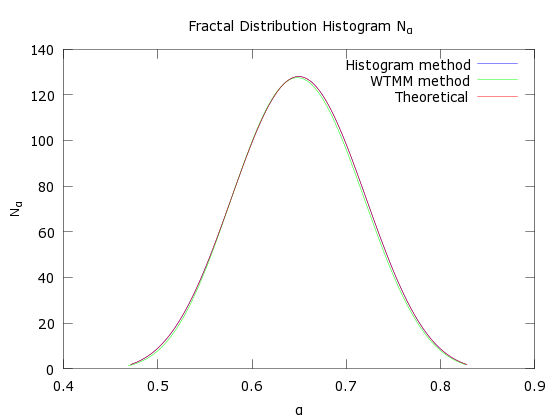


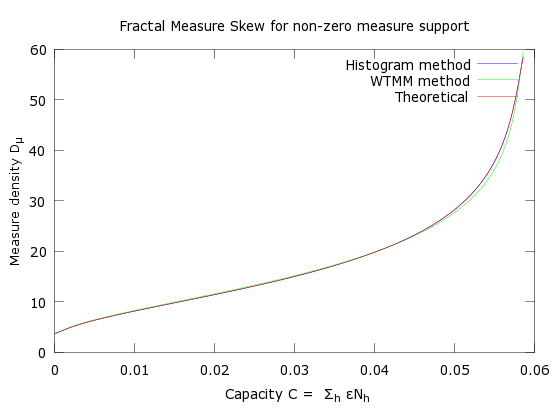




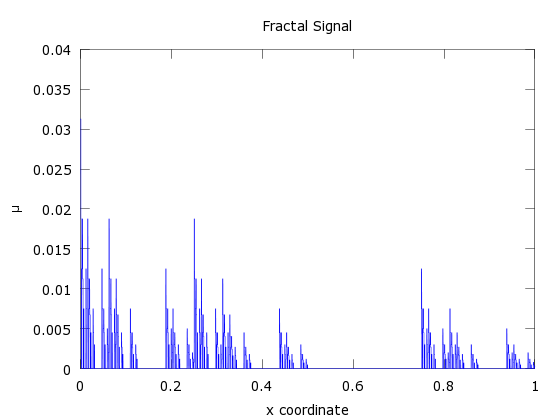


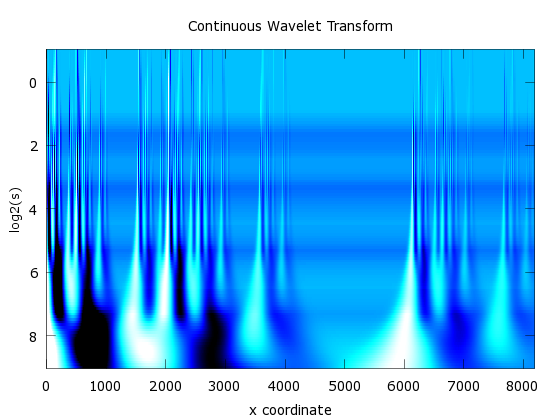
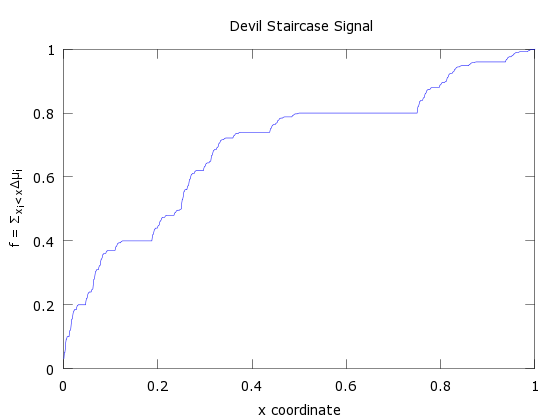


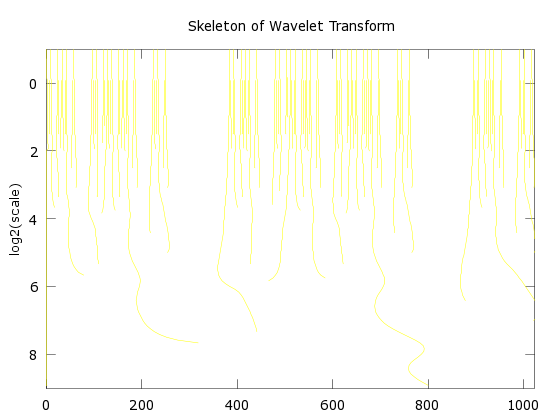


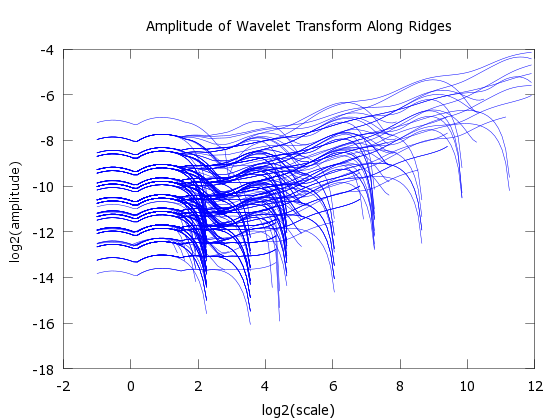


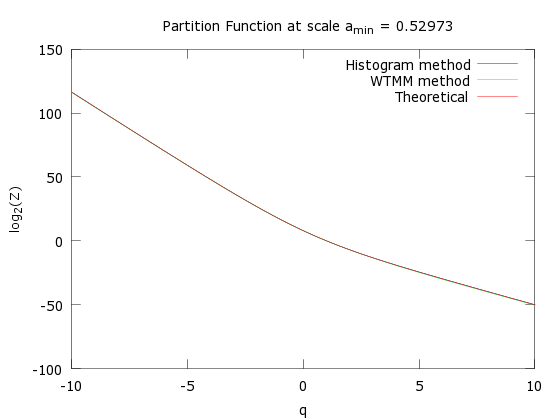
Example 2 – generalized Cantor set with generating vector [p1 p2 0 p3] where the weights are p1=0.5 and p2=0.3, p3=0.2 with base b = 4. The fractal set is generated with 5 iterations, and there are total of N = 1024 sectors. The smallest feature is μmin = 3.2 x 10-4 and the largest feature is μmax = 312.5 x 10-4 .

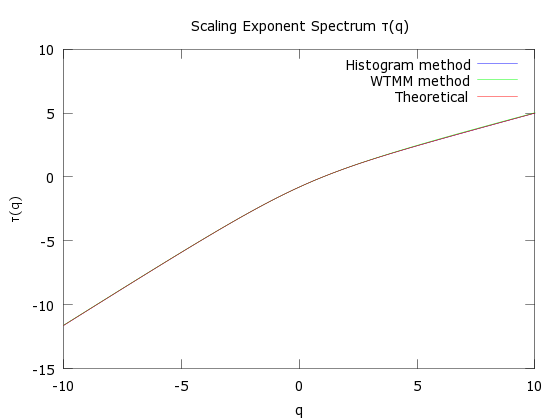


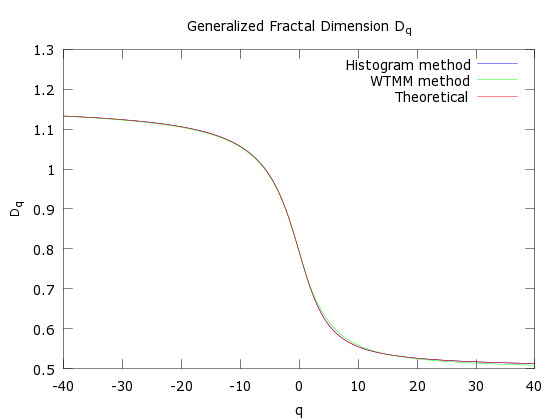


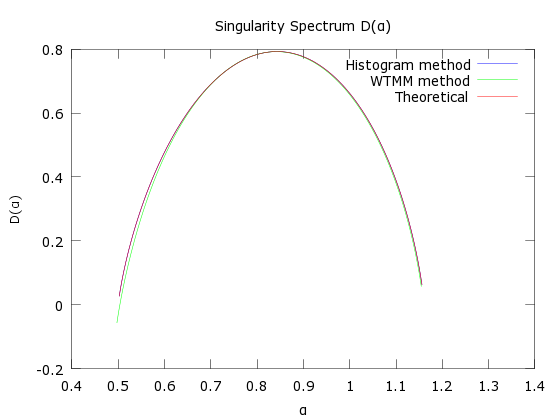


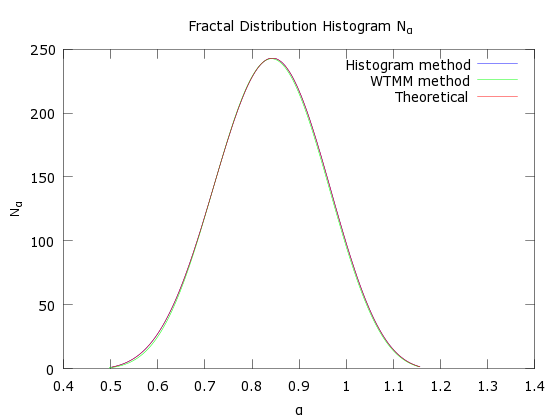


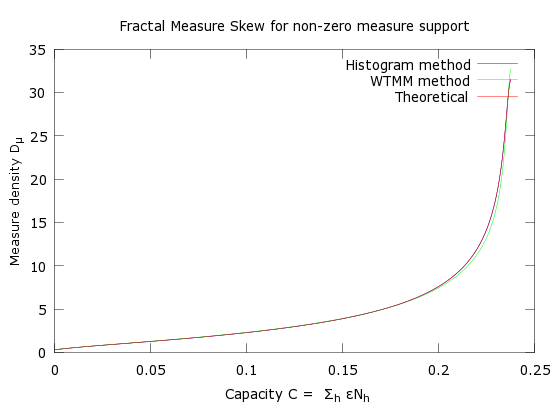




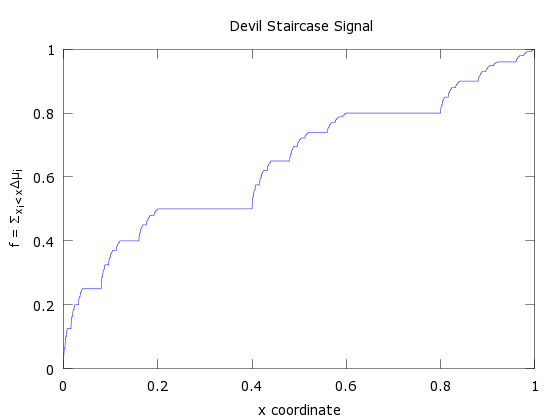


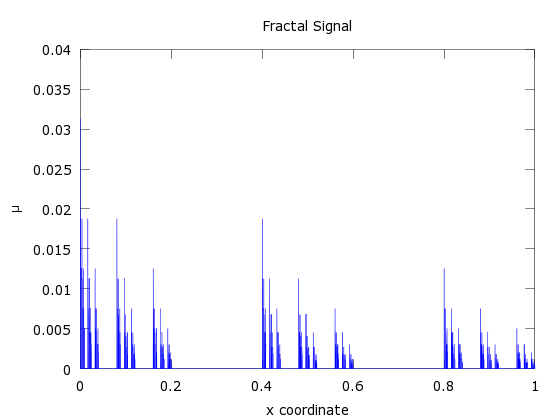


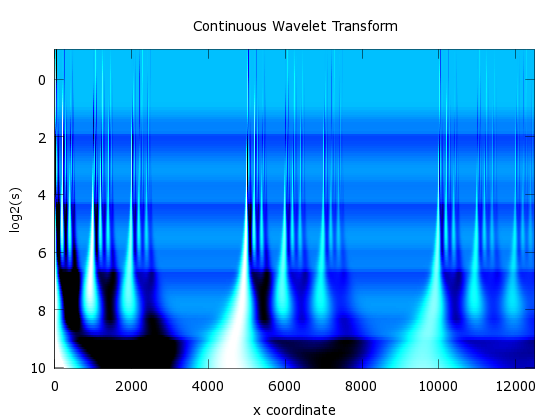


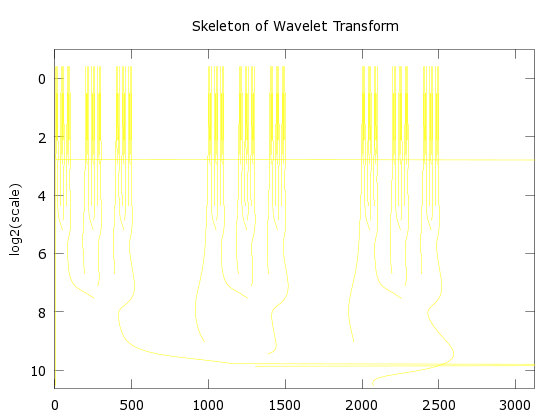


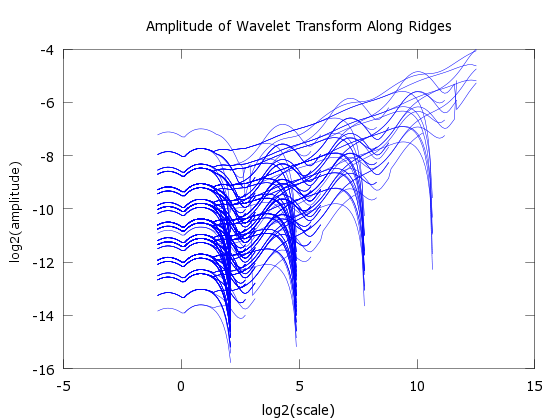
Example 3 – generalized Cantor set with generating vector [p1 0 p2 0 p3] where the weights are p1=0.5 and p2=0.3, p3=0.2 with base b = 5. The fractal set is generated with 5 iterations, and there are total of N = 3125 sectors. The smallest feature is μmin = 3.2 x 10-4 and the largest feature is μmax = 312.5 x 10-5 .

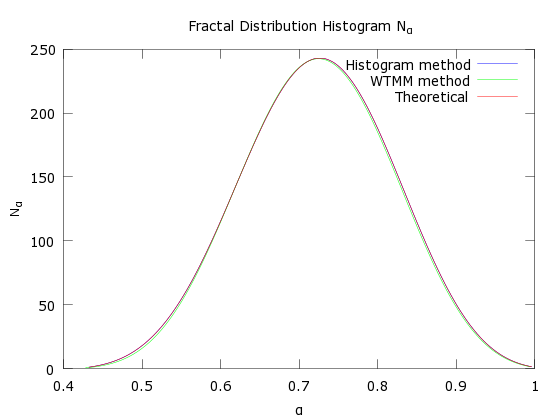
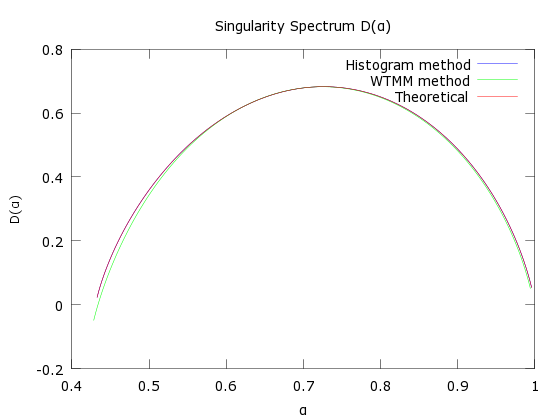
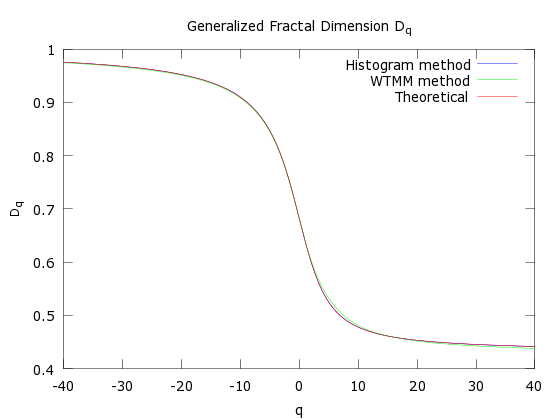
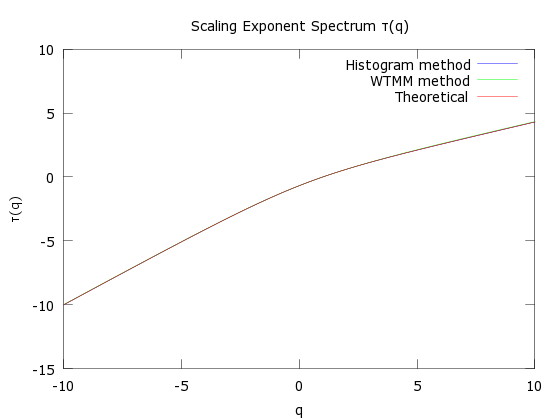
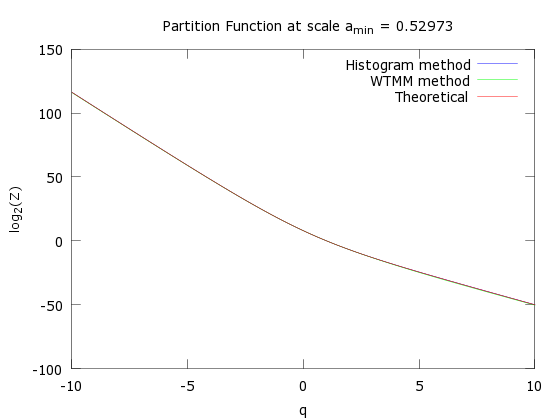


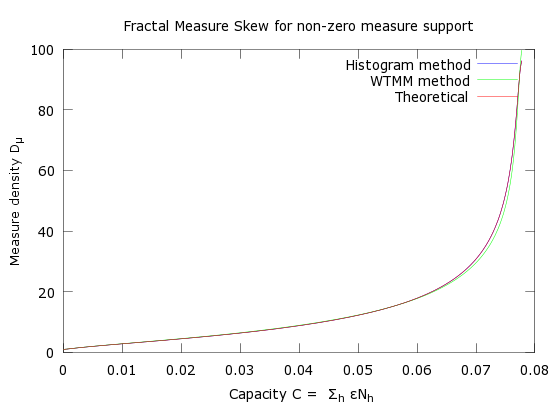




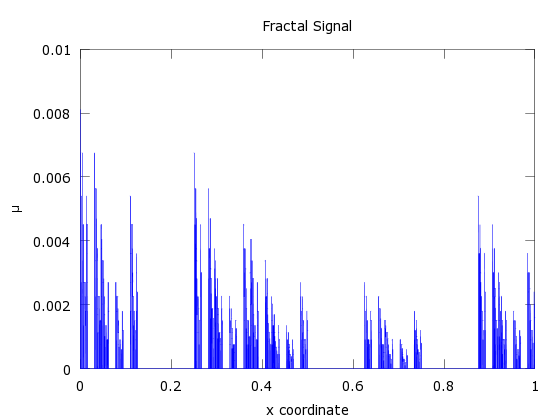


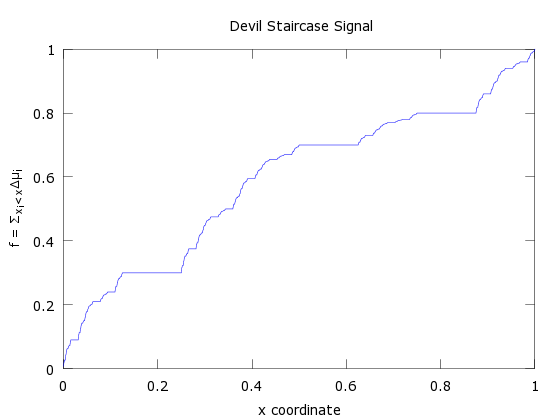


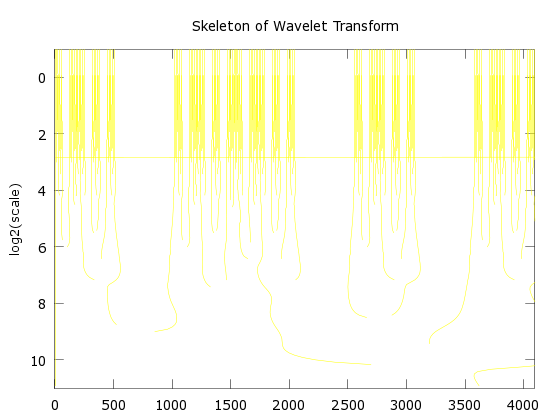


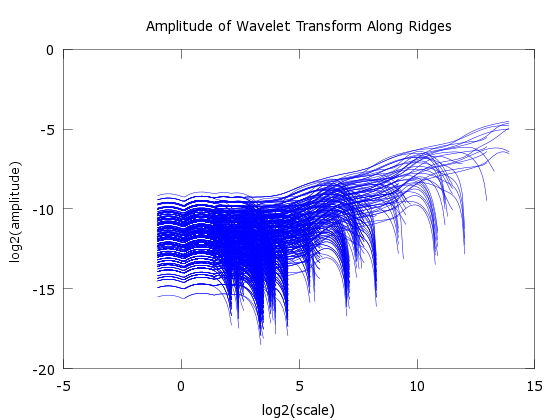


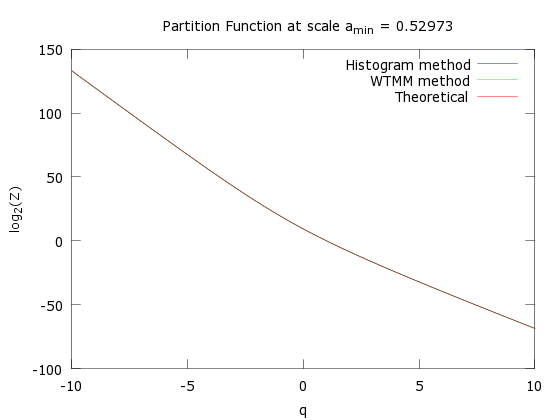
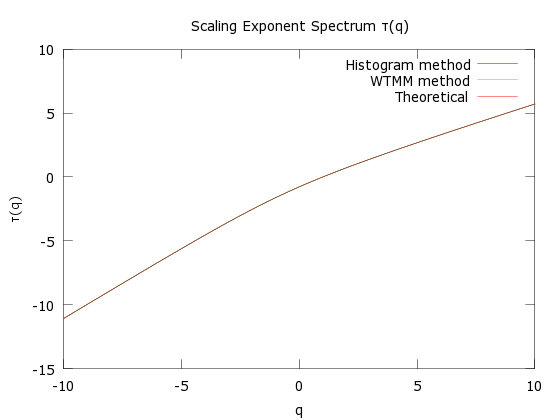
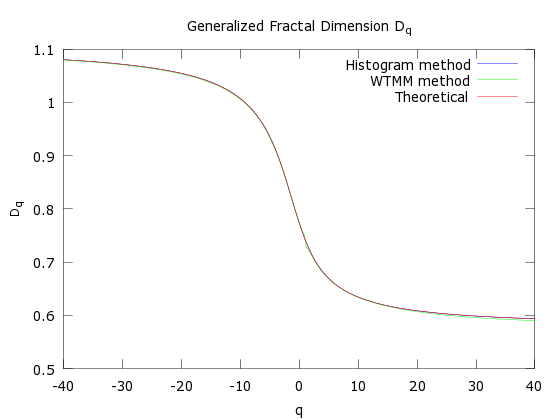
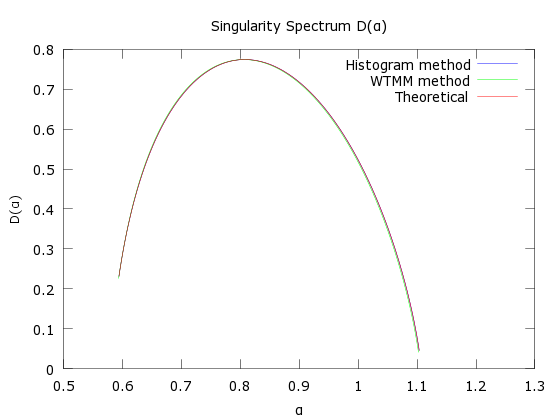
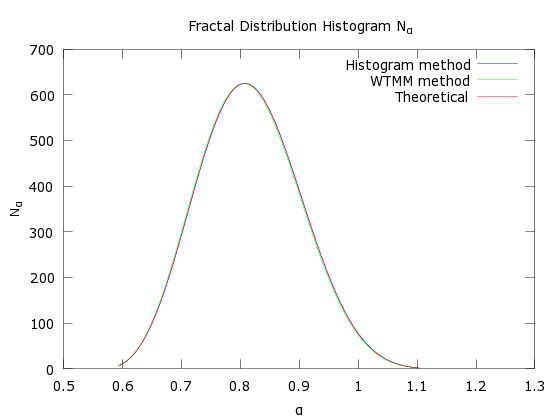
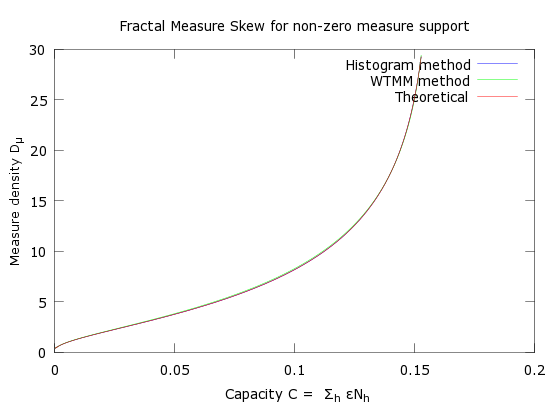
Example 4 – – generalized Cantor set with generating vector [p1 0 p2  p3 0 p4 0 p5] where the weights are p1=0.3 and p2=0.25, p3=0.15, p4=0.1, p3=0.2 with base b = 8. The fractal set is generated with 4 iterations, and there are total of N = 4096 sectors. The smallest feature is μmin = 3.2 x 10-4 and the largest feature is μmax = 312.5 x 10-5 .









**Summary**

advantages of the WTMM method over the Histogram method for obtaining the spectral distribution:

* it allows the proper treatment of sequential IOs / non-fractal components in the IO distribution by identifying the non-singular exponents in the wavelet transform of the fractal signal

disadvantages of the WTMM compared to the Histogram method:

Questions to be answered:

what is the period of the scale oscillations in the CWT plots? Check if removing/dampening of those oscillations will fix the issue with the ridge misidentification in example 3. If not then try setting smaller max scale lower than the kink scale.

Plot 𝛼i = (xi) and see if the values can clustered in m < n clusters

References

[1] The Multifractal formalism revisited with wavelets. J.F. Muzy, E. Bacry, A. Arneodo, *International Journal of Bifurcation and Chaos* Vol. 4, No. 2 (1994), pp 245-302

[2] Wavelet based multifractal formalism : applications to DNA sequences, satellite images of the cloud structure and stock market data in *The Science of Disasters*, 2002, eds A. Arneodo, B. Audit, N. Decoster, J.F. Muzy, C. Vaillant, *(Springer, Berlin)*, pp 26-102

[3] Direct determination of the f(𝛼) Singularity spectrum. , A. Chhabra, R. V. Jensen, *Physical Review Letters*, Vol. 62, Number 12, pp 1327-1330