CNR-IMATI Milano

December 14-16, 2004

WAVELETS AND SELF-SIMILARITY: THEORY AND APPLICATIONS

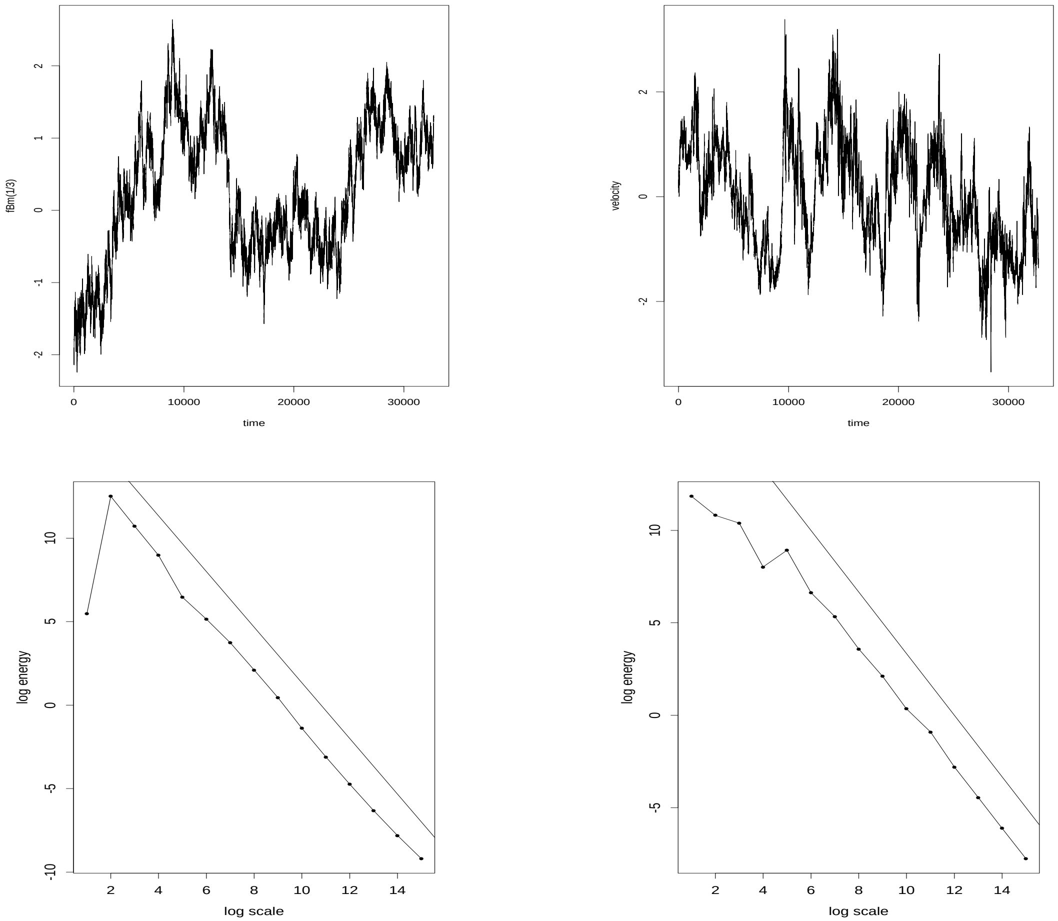
Lecture 6: BAMS in ACTION

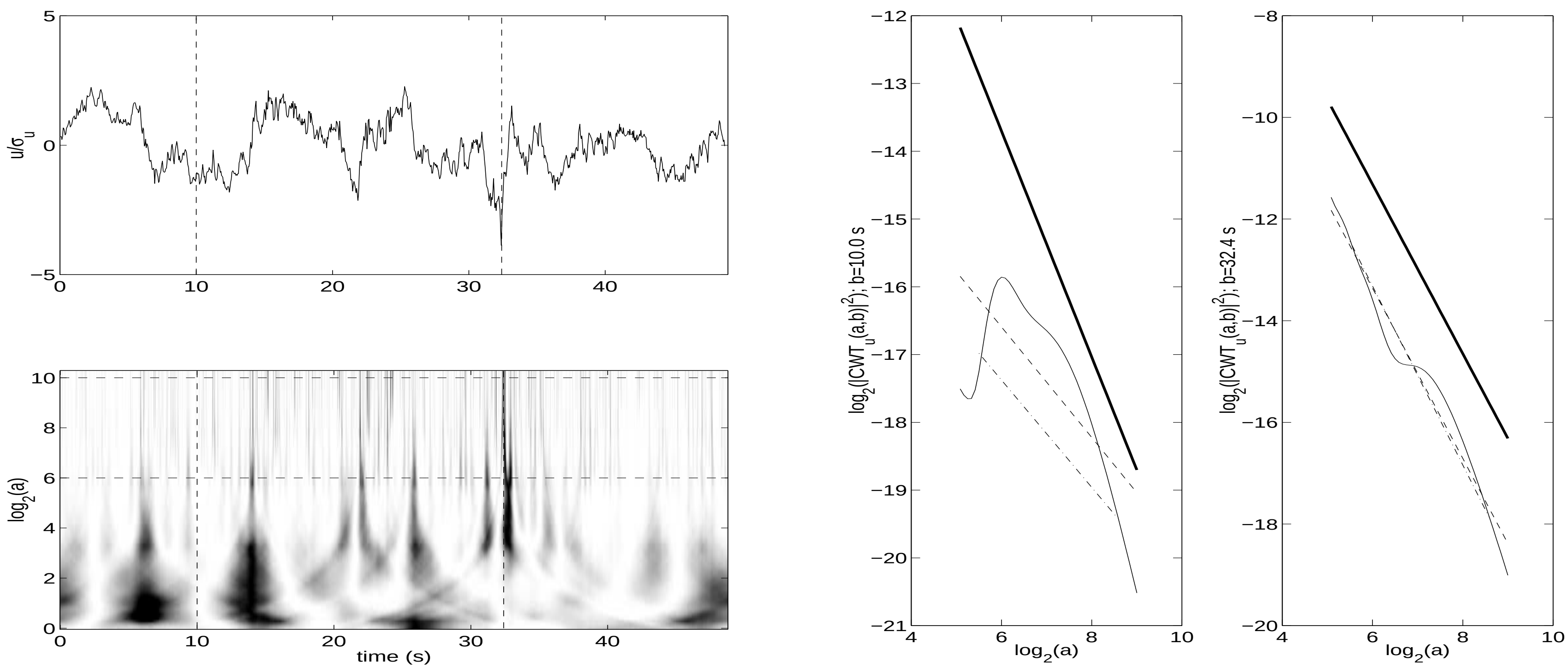
P L A N

1. **Global vs. Local**

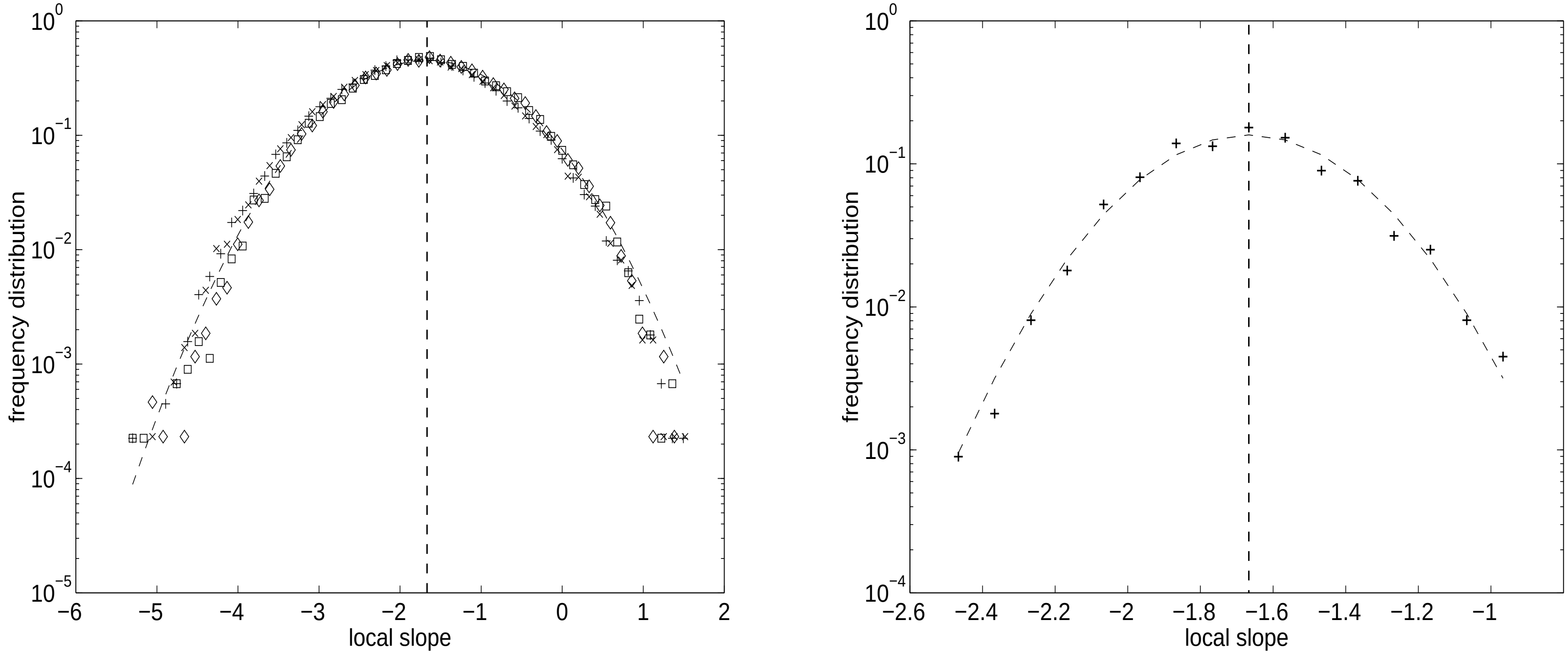
2. **BAMS in Action**

Global/Local scaling exponent in turbulence measurements





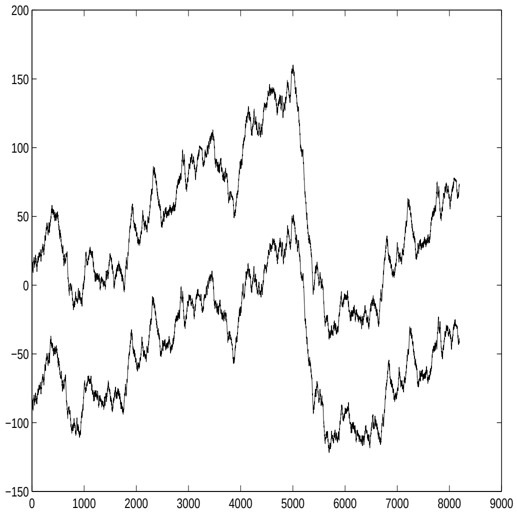
u variation with time for 50 seconds duration (top) and its $\mathcal{CWT}$ energy distribution in the time/frequency domain.

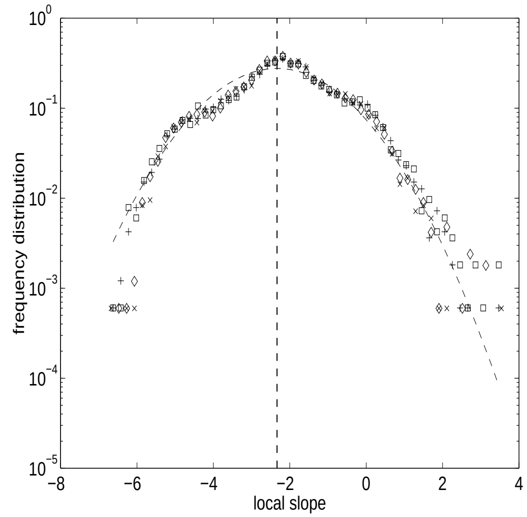


frequency distribution
local slope
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
−6 −5 −4 −3 −2 −1 0 1 2
frequency distribution
local slope
10^0
10^-1
10^-2
10^-3
10^-4
−2.6 −2.4 −2.2 −2 −1.8 −1.6 −1.4 −1.2 −1

Coding vs. Noncoding DNA Sequences

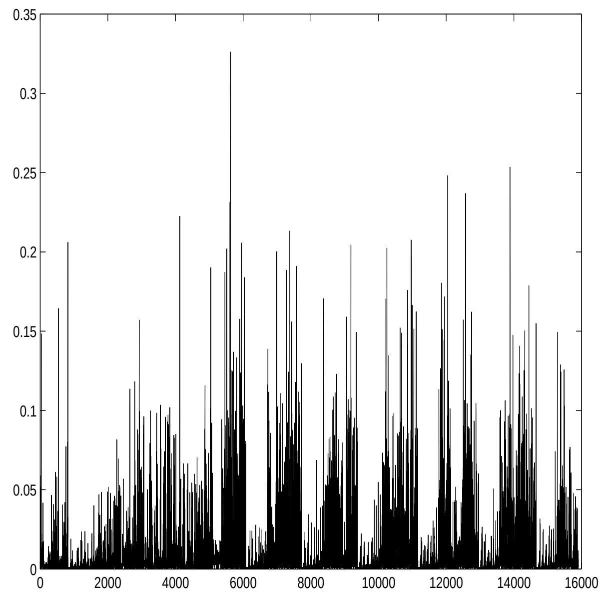
The DNA sequences can be recoded with respect to purines (A, G) v.s. pyrimidine (C, T) as the binary sequences ω in the alphabet $\{-1, 1\}$, where -1 corresponds to A, G and 1 coresponds to C, T.

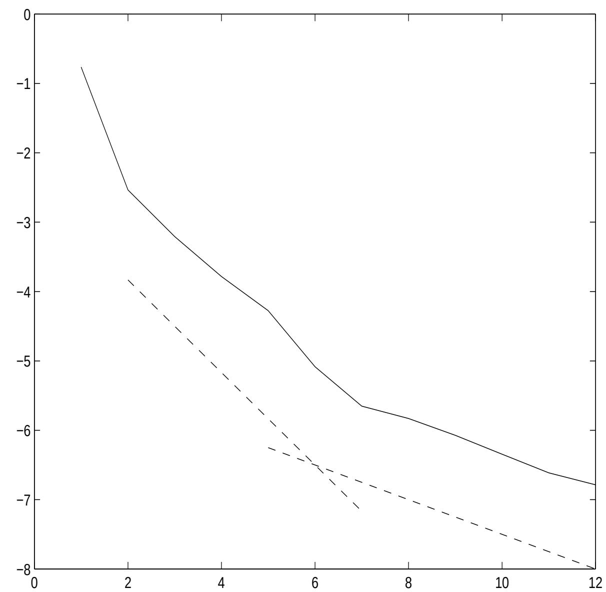




Internet Traffic Data

- Interarrival package times, Duke OIT
- Global power spectra: bifractal? Marginal? Model?





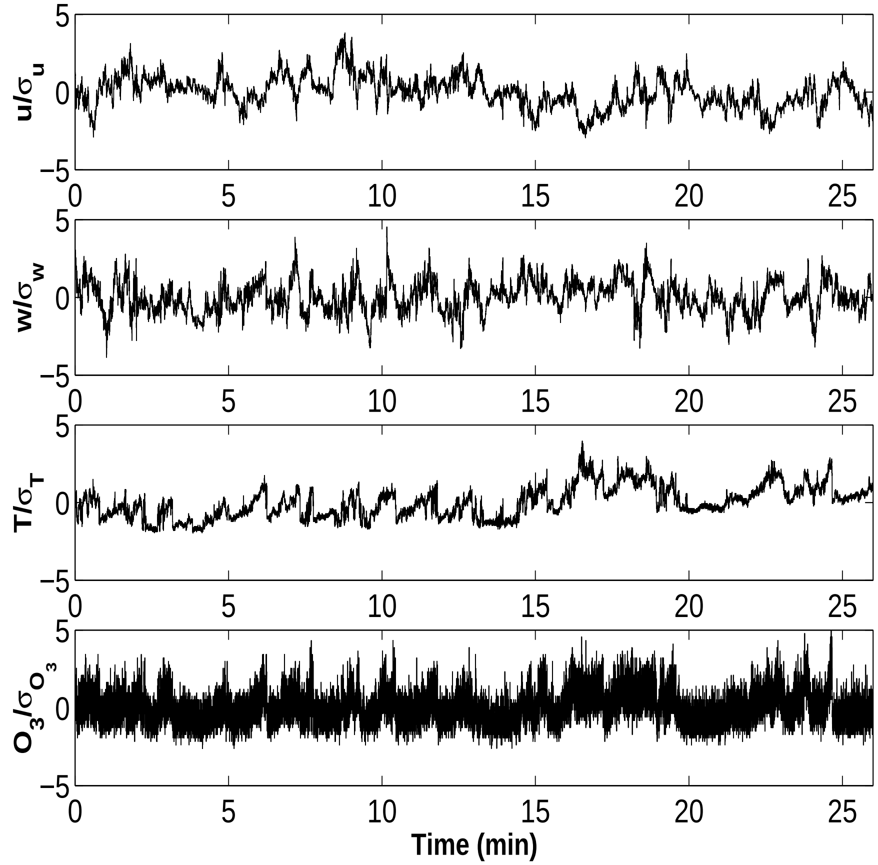
A CASE STUDY: BAMS AGAIN

■ Of interest: separation of the instrumentation noise from high frequency ozone concentration measurements sampled in the atmospheric boundary layer.

■ The proposed Bayesian model relies on BAMS (Bayesian Adaptive Multiresolution Shrinker)

■ Gas analyzers used in high frequency sampling of ozone concentration tend to convolve the signal with a noise assumed either white or autoregressive.

The separation between signal and noise is achieved by inverting wavelet coefficients, splitting in a fashion dictated by the statistical model which incorporates a **K41** type-power law.



Time series of simultaneously measured turbulent velocity components, air temperature, and ozone concentration. For comparison purposes, we normalized all the time series measurements to zero-mean and unit variance.

BAYESIAN MODEL

■ Suppose the observed data $\mathbf{y}$ (e. g., ozone concentrations) represent the sum of an unknown signal $\mathbf{s}$ and random noise ϵ. Coordinate-wise,

$$\mathbf{y}_i = \mathbf{s}_i + \epsilon'_i,\ i = 1, \ldots, n.$$

■ In the wavelet domain (after applying a nondecimated wavelet transformation W to the observed data), regression from the "time" domain becomes $d_{jk} = \theta_{jk} + \epsilon_{jk}$, where d_{jk}, θ_{jk}, and ϵ_{jk} are the j, k-th coordinates in the traditional nondecimated scale/shift wavelet-enumeration of vectors $W\mathbf{y}, W\mathbf{s}$ and $W\epsilon'$, respectively. Our assumption is that the coefficients d_{jk} can be considered independent, at least for high resolution levels.

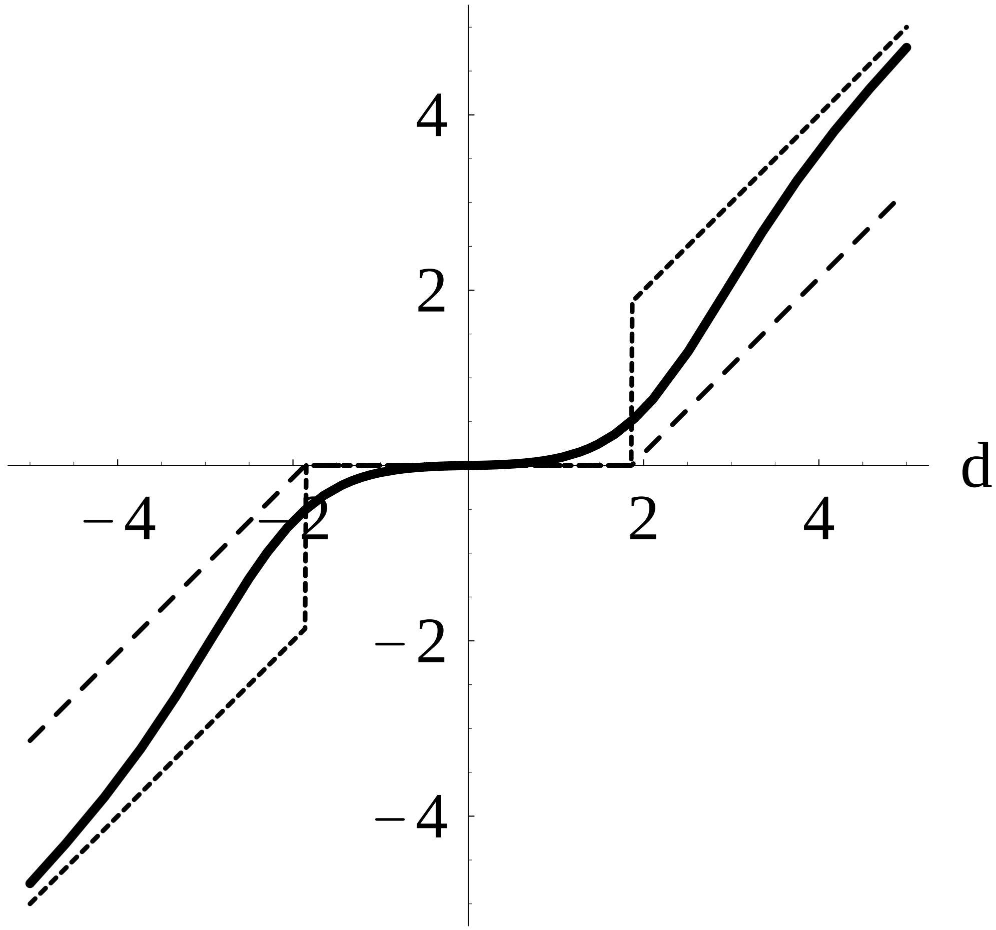
- We assume, as commonly done, that each coefficient d in the wavelet domain is affected by normal errors, and thus the conditional distribution of d given θ and σ^2, $[d|\theta,\sigma^2]$, is $\mathcal{N}(\theta,\sigma^2)$. In BAMS, the prior distributions on σ^2 and θ are chosen to be, respectively, an exponential one, $\mathcal{E}(\mu)$, and a mixture of a point mass at zero and a double exponential distribution, $\mathcal{DE}(0,\tau)$.

- The Bayes rule is:

$$\delta^*(d) = \frac{(1-\epsilon)\ m(d)\ \delta(d)}{(1-\epsilon)\ m(d) + \epsilon\ \mathcal{DE}\left(0, \frac{1}{\sqrt{2\mu}}\right)}. \tag{1}$$

where

$$m(d) = \frac{\tau e^{-|d|/\tau} - \frac{1}{\sqrt{2\mu}} e^{-\sqrt{2\mu}|d|}}{2\tau^2 - 1/\mu}.$$



Bayes rule $\delta(d)$ for $\epsilon = 0.9, \tau = 2$, and $\mu = 1/2$.

■ Vidakovic and Ruggeri (2001) proposed an empirical moment matching specification of parameters which worked well if the signal is smooth. Here, the parameter specification will reflect the fact that signals are self-similar with theoretically established Hurst exponent $H = 1/3$.

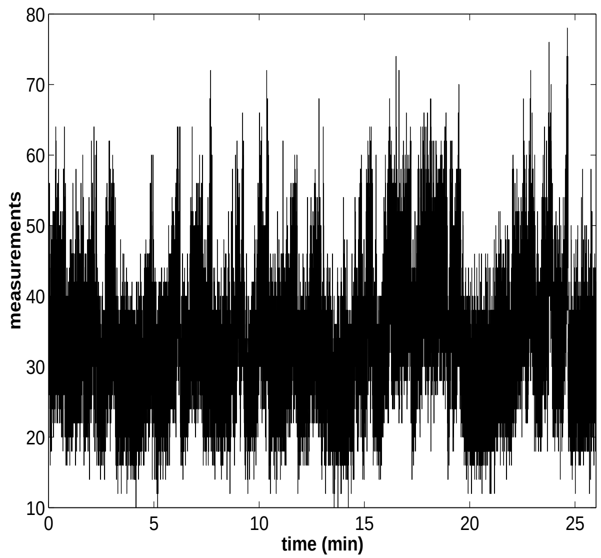
1. μ is the reciprocal of the mean for the prior on σ^2, or, equivalently, the square root of the precision for σ^2. We first estimate σ by a robust Tukey's $\texttt{pseudos} = (Q_3 - Q_1)/C$, where Q_1 and Q_3 are the first and the third quartile of the finest level of details in the decomposition and $1.3 \leq C \leq 1.5$. We propose $\frac{1}{\texttt{pseudos}}$ as a default value for μ; according to the Law of Large Numbers, this ratio should be close to the "true" μ.

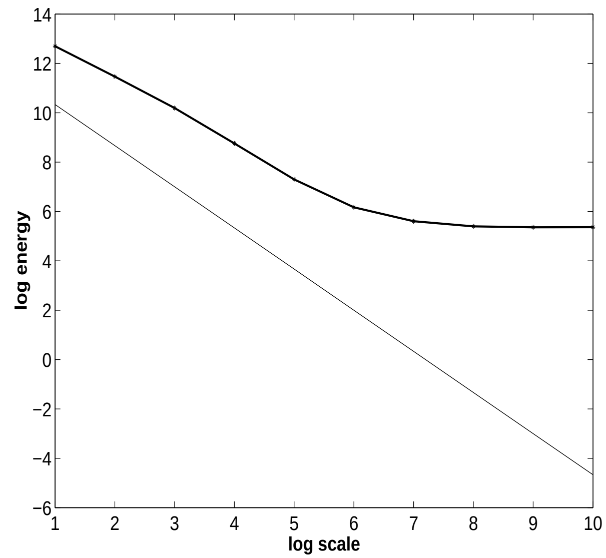
2. ϵ is the weight of the point mass at zero in the prior on θ and should depend on level j. If the signal is smooth ϵ should be close to 1 at the finest level of detail and near 0 at coarse levels. In our case the signal is not smooth and a decay in ϵ is unreasonable. Hence, we fixed $\epsilon = 0.5$ in all levels. With a constant ϵ, control of prior variance decay becomes easier though we emphasize that the value of $\epsilon = 0.5$ is not unique.

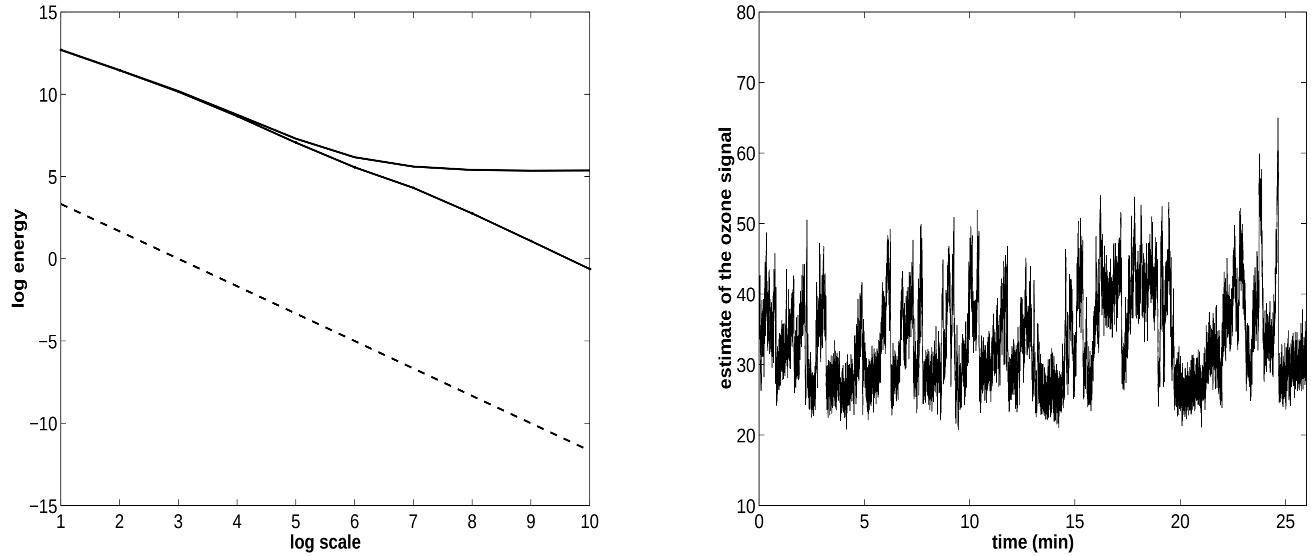
3. The parameter τ is the scale of the "spread part" in the prior. In the case of a double exponential prior, the variance of the signal part is $2\tau^2$. Because of assumed independence between the error and the signal parts, we have $\sigma_d^2 = 2(1-\epsilon)^2\tau^2 + 1/\mu$, where σ_d^2 is the variance of an observation d. According to **K41** exact power laws, the average energies of the signal in the wavelet domain decay in a log-linear fashion when increasing the resolution of the levels (or the level index). This provides a calibration method for eliciting prior variances on the signal coefficients: they decay proportionally to $2^{(-5/3)}$.

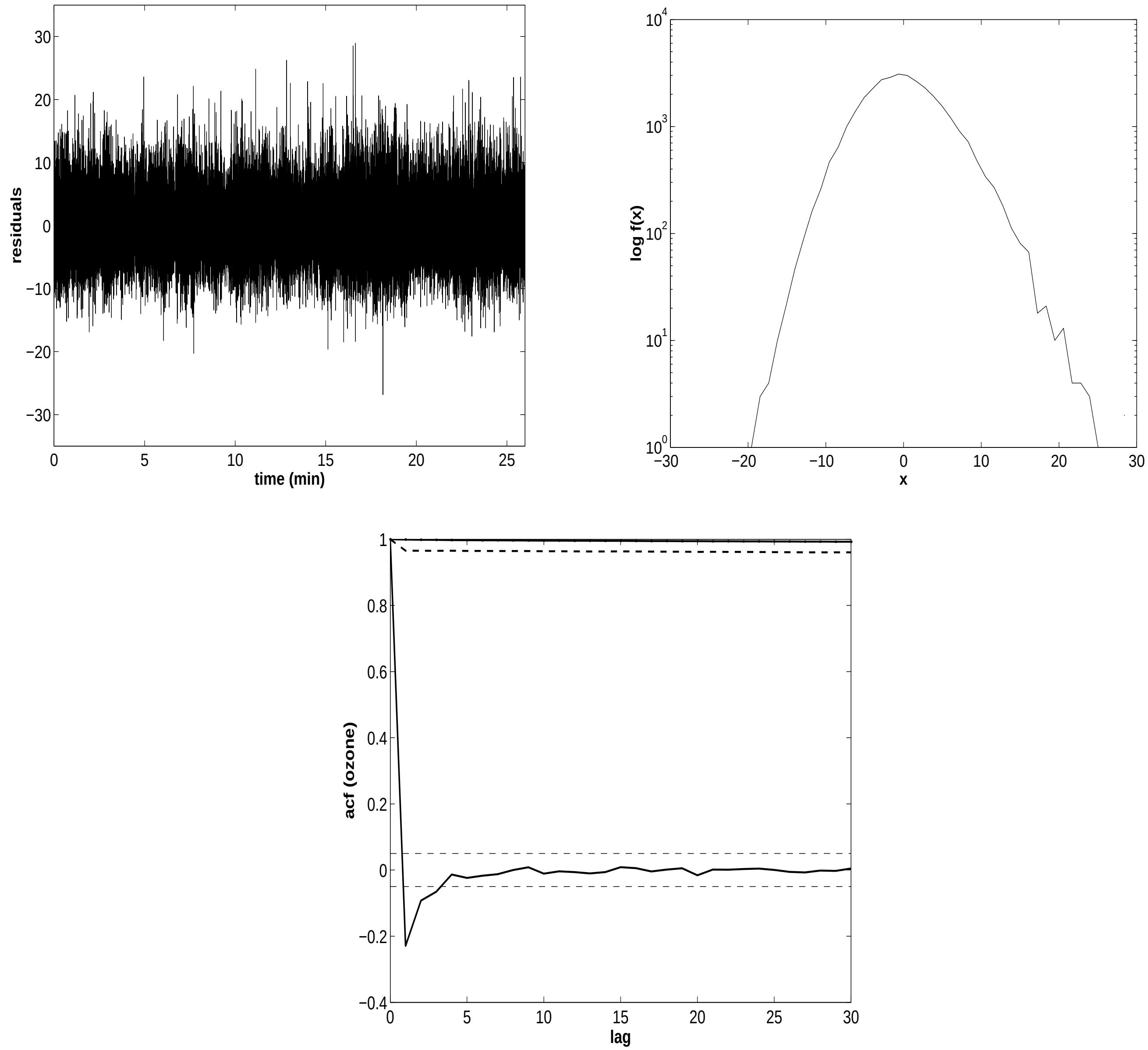
Unlike the thresholding rules that set small wavelet coefficients to 0, the rule (1) splits the coefficients as $d = \delta^*(d) + (d - \delta^*(d)) = \hat{\theta} + \hat{\epsilon}$. The $\hat{\theta}$-part corresponds to turbulence signal and exhibits energy spectra decaying with a slope of -5/3 ($H = 1/3$) due to pre-described scaling. On the other hand, the noise part coefficients, $\hat{\epsilon}$, exhibit "flat" spectra, as expected.

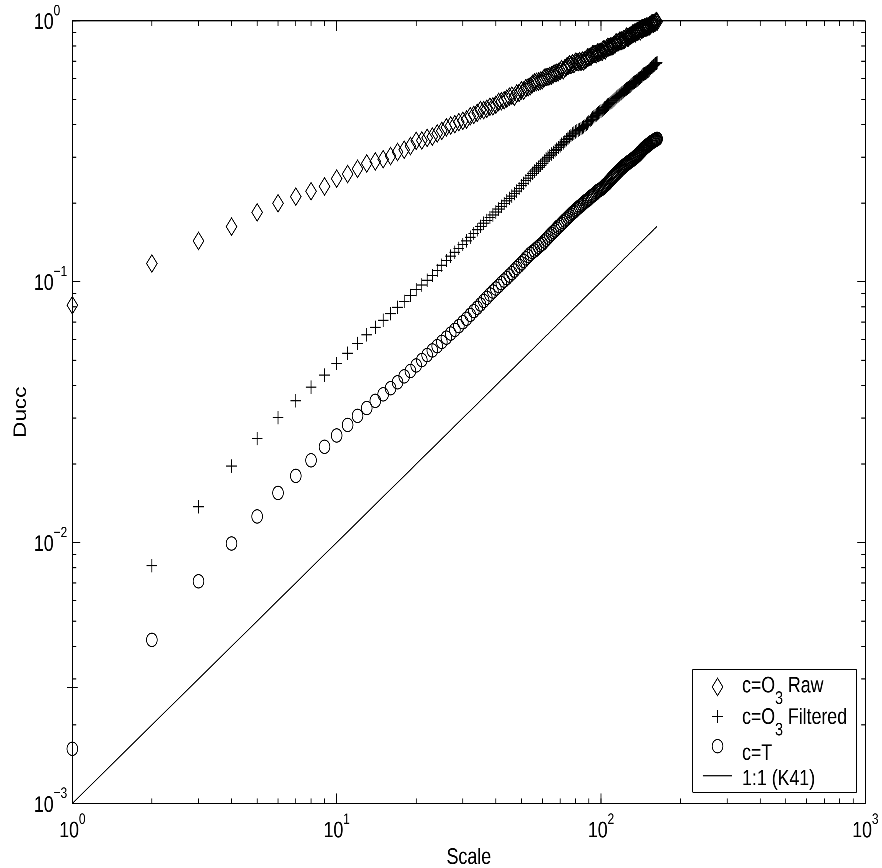
Filtering Ozone Data (A Cartoon)






log energy
log scale
estimate of the ozone signal
time (min)


residuals
time (min)
log f(x)
x
acf (ozone)
lag



Computed mixed structure functions ($D_{u'CC}$) for C representing filtered and raw O_3 time series with u'. For reference, the mixed structure function between T and u' is also shown along with τ^1 from **K41**.

■ Interactions between the filtered O_3 time series and turbulent velocity (u', w') are more consistent with theoretical predictions from turbulence theories (e.g. **K41**) than the unfiltered ozone time series.

■ Although marginally normal, the residuals e_k are not exactly "white." We found that in separated noise there are significant 2- to 3-lag autocorrelations. In the run we explored the empirical model for the noise is found to be

$$e_k - 0.2971 * e_{k-1} - 0.2056 * e_{k-2} - 0.1605 * e_{k-3} = Z_n,$$

where Z_n is a white noise time series.

■ This finding is in agreement with the physical properties of the chemiluminescent gas analyzer, in which some residual reactants from previous sampling times influence the present ozone concentration measurement.