Turbulence Observations in Heliospheric Space Plasmas

1. A basic turbulence toolkit

Luca Sorriso-Valvo

Nanotec/CNR, Rende, Italy

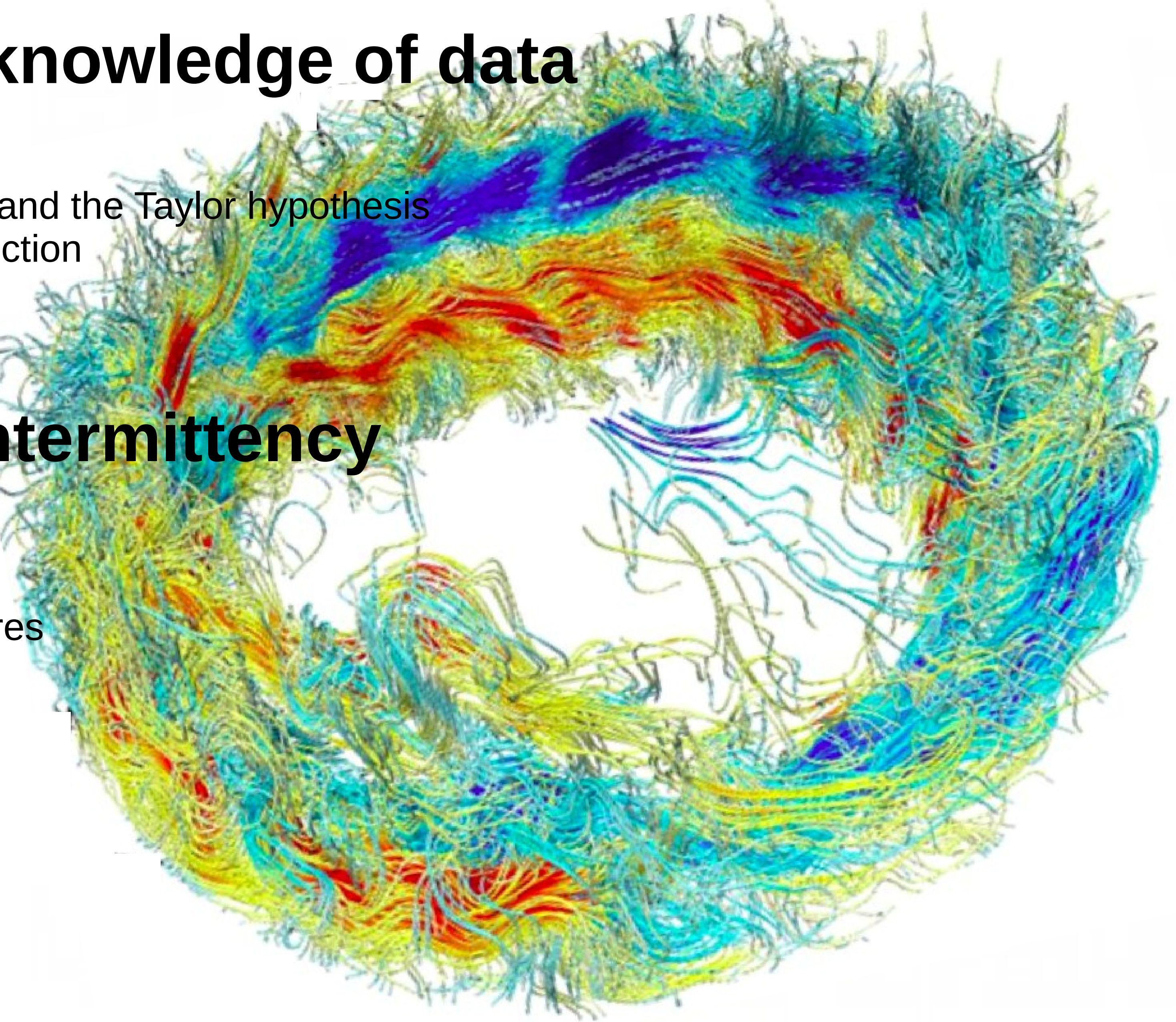
Summary

1. The basic knowledge of data

- Know your data
- Stationarity, Ergodicity and the Taylor hypothesis
- The autocorrelation function
- The energy spectrum
- The relevant scales

2. Tools for intermittency

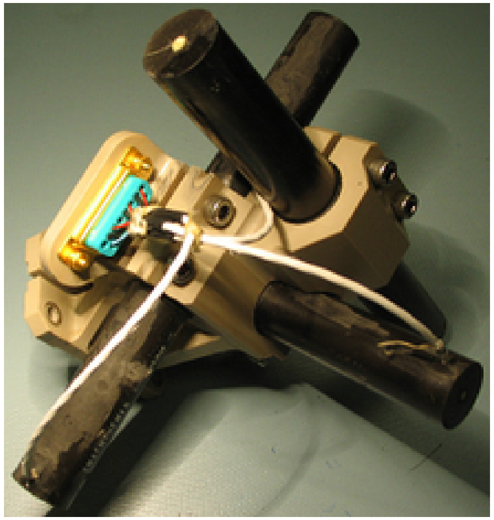
- Scaling of PDFs
- Structure functions
- Multifractal analysis
- Identification of structures



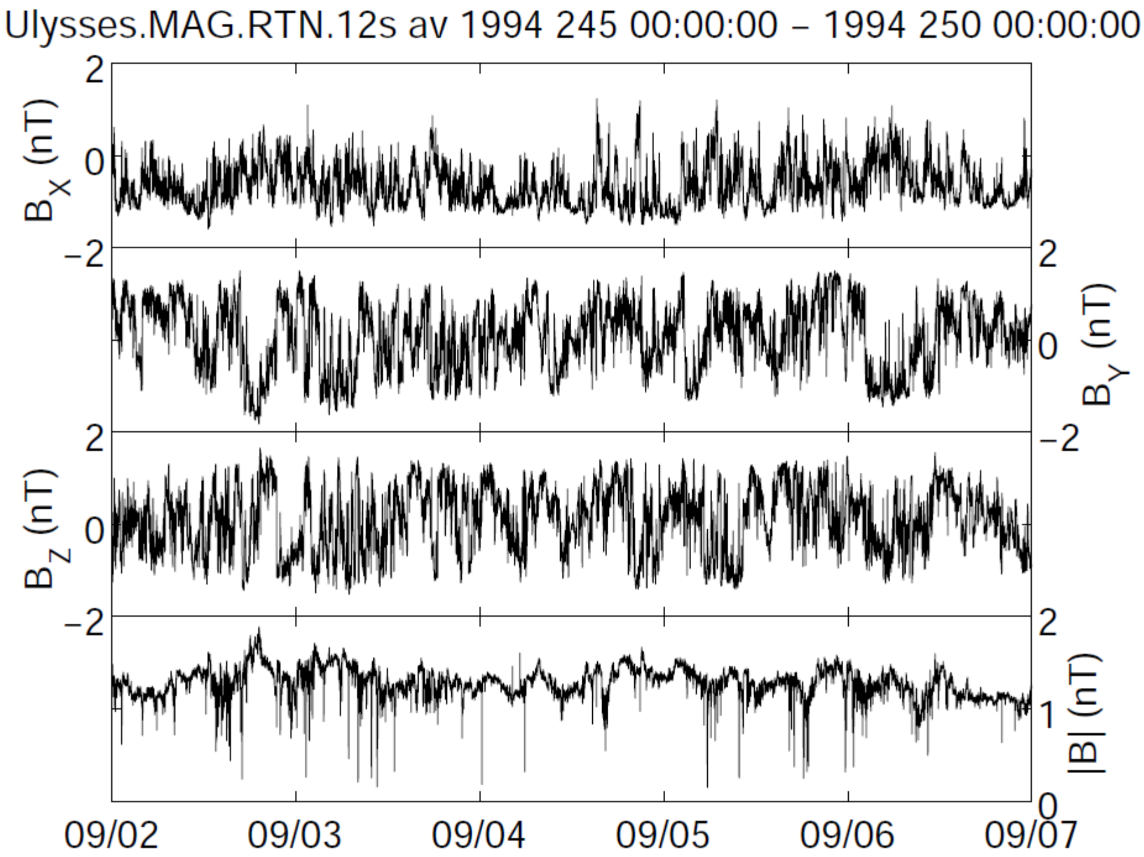
Turbulent measurements

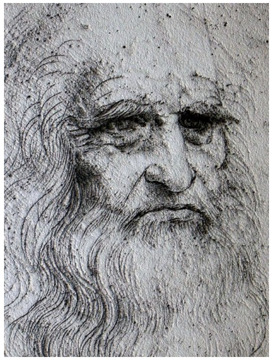
- Typical measurements of space plasma turbulence include velocity, density, temperature, magnetic field, composition, particle VDFs and other quantities.

- The usual data are time series $v(t_i)$ taken with sampling time Δt by some instrument, at a given point (Eulerian), or following a trajectory (Lagrangian).



Search coil magnetometer used on Cluster and THEMIS spacecrafts



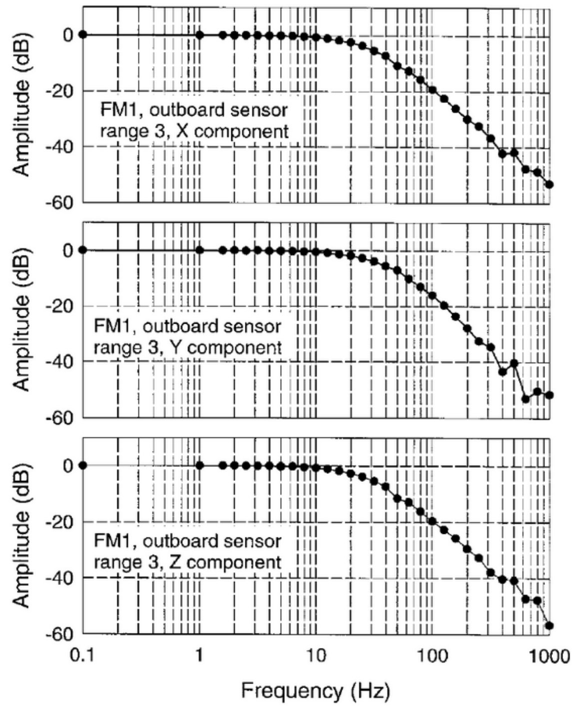


Know your data

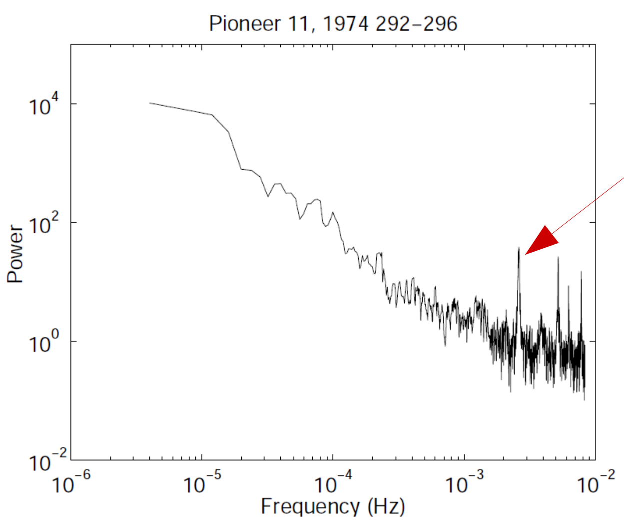
All measurements have approximations and caveats.

Know as much as possible about data before the analysis.

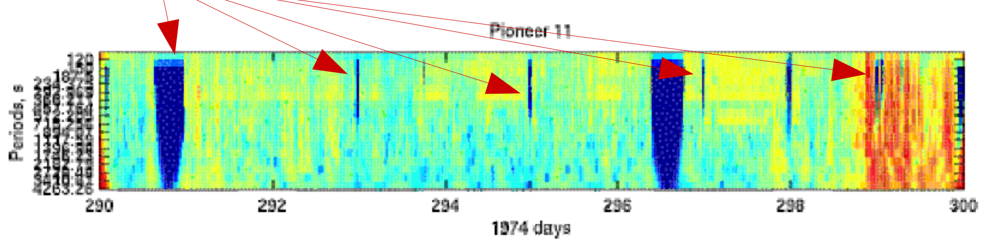
- Study instrumentation and data characteristics and the available literature (manuals, publications...). Talk to PIs and instrument teams.
- Have clear the geometry of the system (coordinate system, units, etc.).
- Be aware of the possible artifacts: **instrumental** (resolution, noise, digitization, etc.); **human** (calibration, reduction, manipulation - e.g. VDF moments); **spacecraft or telemetry** (spurious frequencies, data rate, etc.).



Analogue response of the three sensors of the one of the vector fluxgate magnetometers on Cluster spacecraft.



Pioneer 11 magnetic data: the peaks are not waves but spurious frequencies from spacecraft (wavelet analysis shows the periodic appearance of power)



Know your data/2

All measurements have approximations and caveats.

Know as much as possible about data before the analysis.

Deal with bad data: PLOT THE DATA!

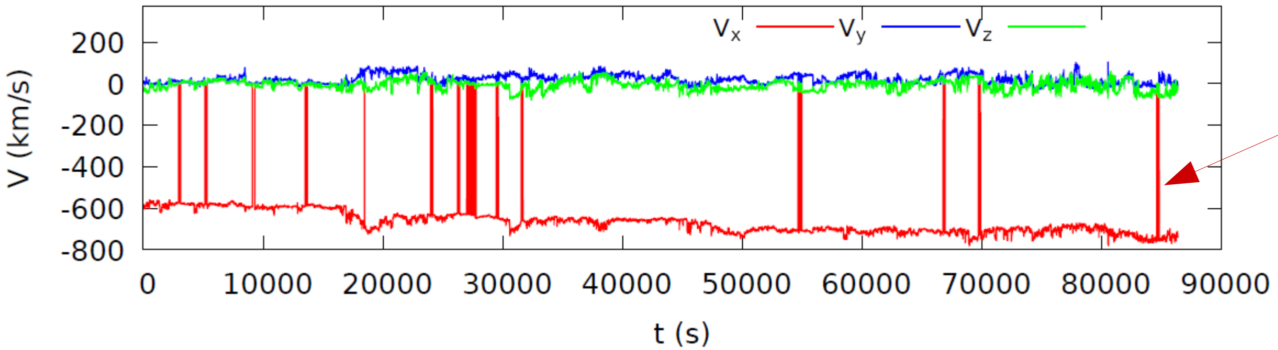
• Check for data gaps and bad points. Fill/replace (linear, spline fit; synthetic data…) or skip? It depends on the analysis…

• Check for synchronization problems: regularise time if needed (in particular for combined data analysis – e.g. velocity and magnetic field).

• De-trend? Trends might go into fake low frequency power or scale-dependent fluctuations.

• Normalize when needed (e.g.: solar wind density ~ R^{-2}).

• Remove the mean if needed.



WIND spacecraft velocity components: missing data points replaced by 0 show up in the radial component (not visible, but present in the other components)

Analysis Methods for Multi-Spacecraft Data, ed. G. Paschmann and P. W. Daly, ISSI, Bern, 1998

Know your data/3

All measurements have approximations and caveats.

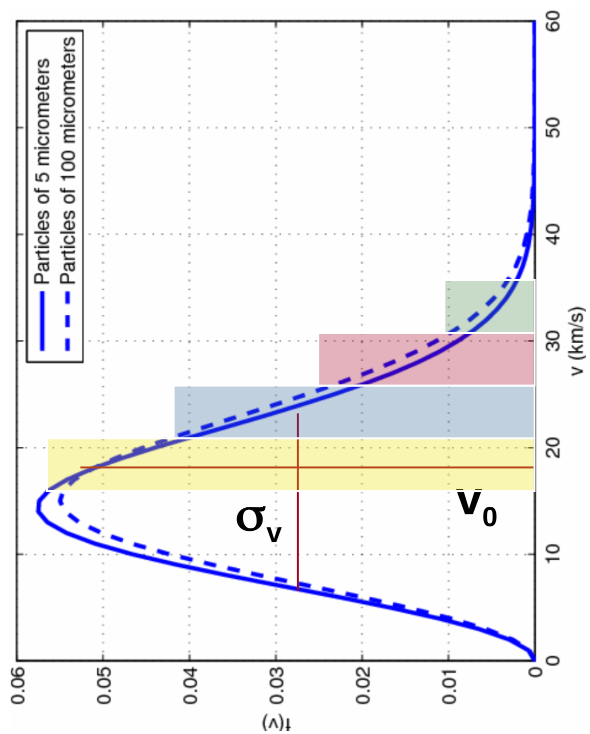
Know as much as possible about data before the analysis.

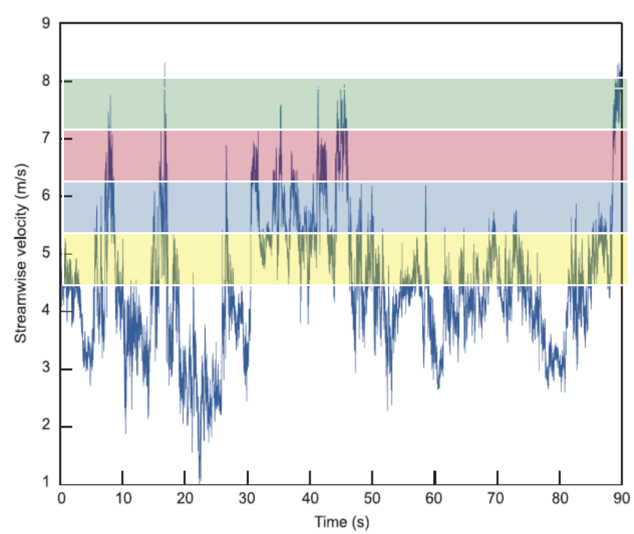
Compute basic quantities: data distribution function, average, standard deviation...

$$p(x) = \lim_{\Delta x \to 0}\left[\frac{\text{Prob}\,[x < x(t) \leqslant x + \Delta x]}{\Delta x}\right]$$

$$\int_{-\infty}^{+\infty} p(x)\,dx = 1$$

$$p(x) \geqslant 0$$





$$\mathrm{v}_0 = \langle \mathrm{v}(t) \rangle$$

$$\sigma_{\mathrm{v}} = \left\langle \left(\mathrm{v}(t) - \mathrm{v}_0\right)^2 \right\rangle^{1/2}$$

$$p(x) = \frac{N_x}{NW}$$

The Taylor hypothesis

Theoretical results on fluid turbulence are mainly in the space (wavevector) domain, rather than in time (frequency) domain. When dealing with turbulent time series, the **Taylor** “frozen-in” hypothesis must be valid to switch from time to space.

If turbulent fluctuations are small with respect to the mean flow, it is possible to neglect the (slow) time variations and assume that the field is “frozen” while it spans the probe.

$$\sigma_v / v_0 << 1$$

$$r \rightarrow v_0 t$$

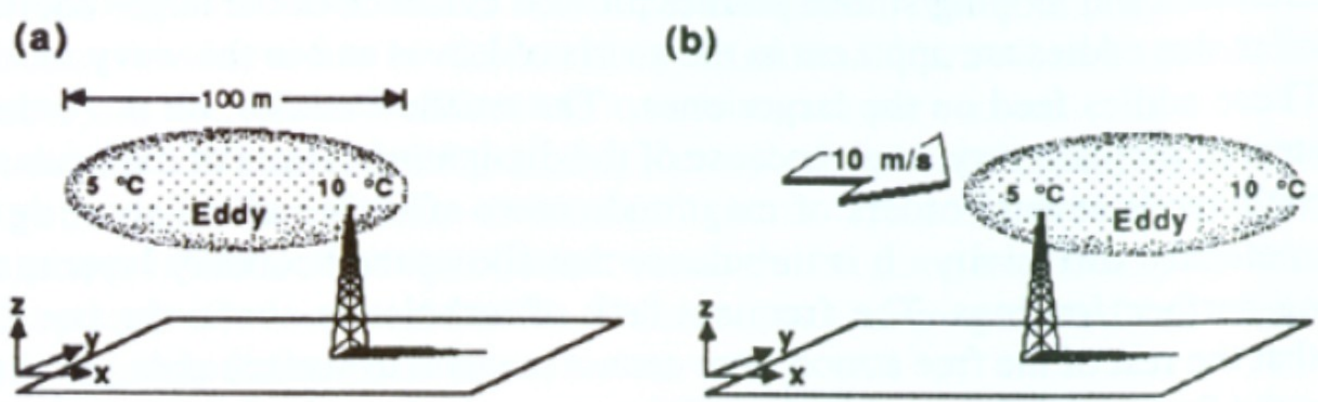


Fig. 1.4 Illustration of Taylor's hypothesis. (a) An eddy that is 100 m in diameter has a 5 ° C temperature difference across it. (b) The same eddy 10 seconds later is blown downwind at a wind speed of 10 m/s.

from Stull 1988

The Taylor hypothesis/2

In space plasmas there are complications (waves). A more formal version:

Frequency in the spacecraft frame $\omega_{s/c}$ is the combination of the plasma frame frequency ω and the plasma frame wave-vector ***k*** advected by the bulk flow $\boldsymbol{V}_{SW}$:

$$\omega_{s/c} = \omega + \boldsymbol{k} \cdot \boldsymbol{V}_{sw}$$

In SW turbulence, the typical wave mode is the Alfvén mode, for which $|\omega| \sim k\, V_A$

$V_{sw} \sim 400\text{–}500\ \mathrm{km\,s^{-1}}$

$V_A \sim 50\ \mathrm{km\,s^{-1}}$

→ $V_A \ll V_{sw}$ → $|\omega| \ll |\boldsymbol{k} \cdot \boldsymbol{V}_{sw}|$ → $\omega_{s/c} \sim \boldsymbol{k} \cdot \boldsymbol{V}_{sw}$

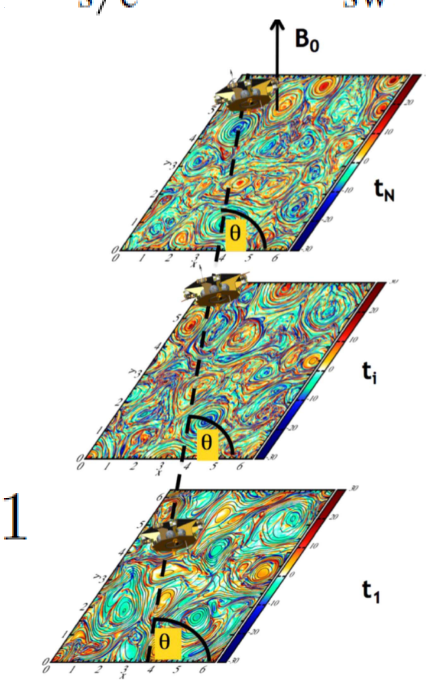
However, in other environments like the Earth's magnetosheath $v_{ms} \sim V_A$

$$\omega \sim \mathbf{k} \cdot \mathbf{v}_A = k\, V_A \cos\theta$$

$$\mathbf{k} \cdot \mathbf{v}_{ms} = k v_{ms} \cos\phi$$

→ Taylor is valid if $\cos\theta \ll \cos\phi$

In typical Alfvénic turbulence $\theta \sim \pi/2$ → Taylor is valid if $\cos\phi \sim 1$



Stawartz et al., JGR 121, 21 (2016); Perri et al., APJS 231, 4 (2017)

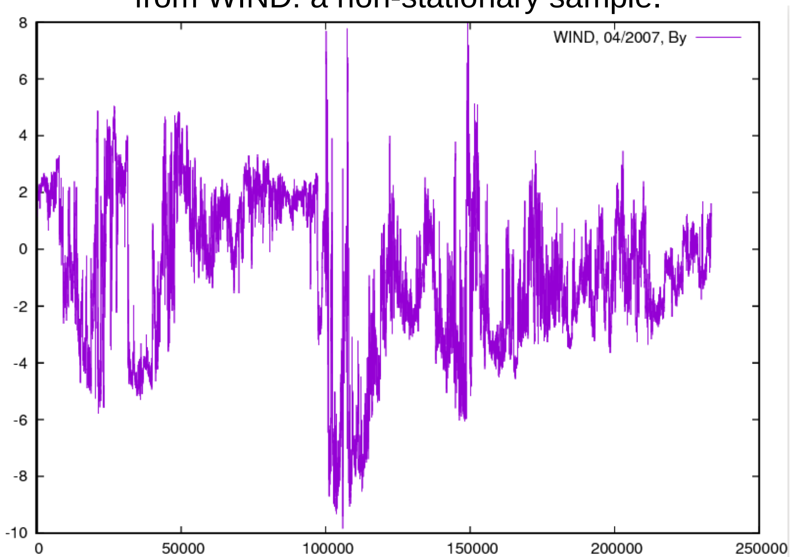
Statistical meaningfulness

Most of the statistical tools for turbulent data assume **ergodicity** and **stationarity**.
Although often verified in space plasmas, this is not always the case and should be checked.

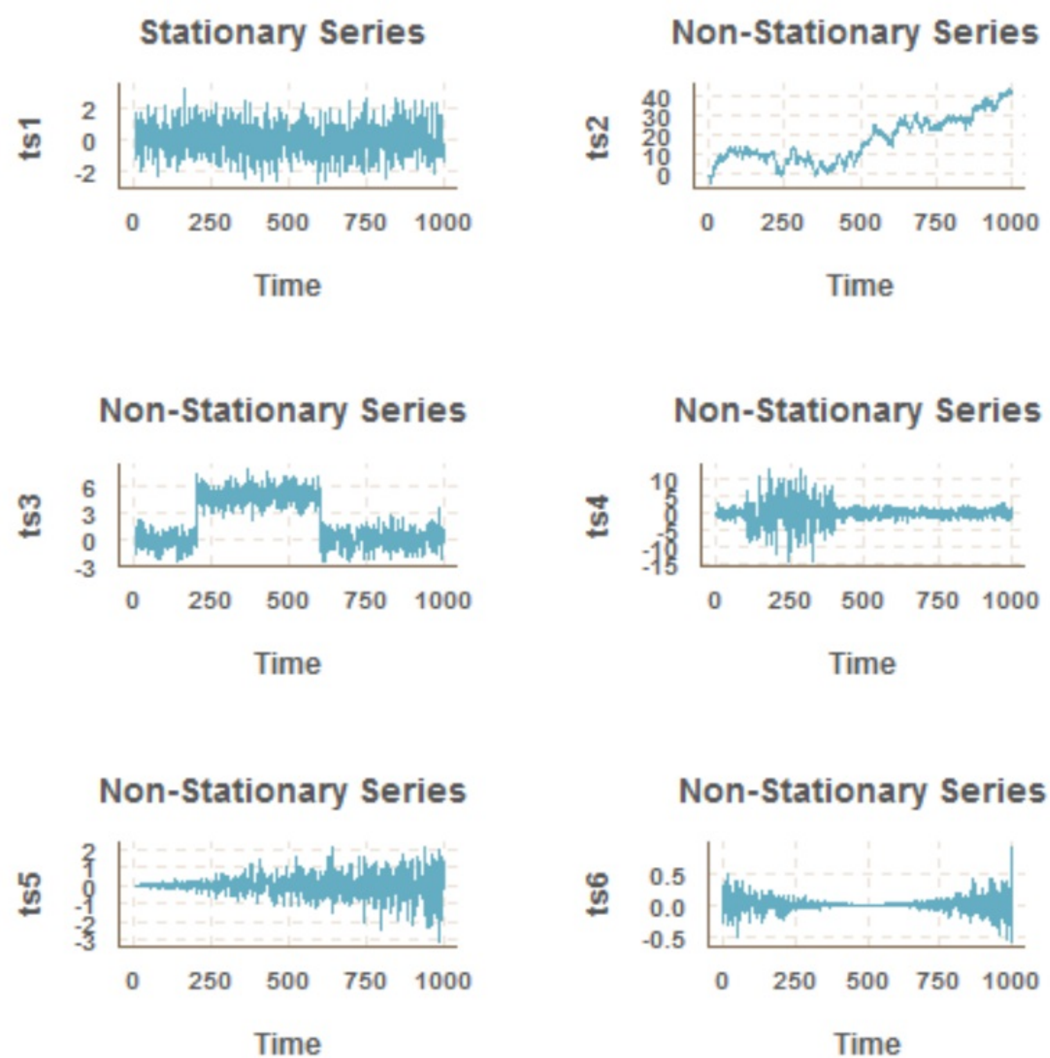
Stationarity

All statistical properties of a time series must be independent of time.
It corresponds to homogeneity of turbulence, via the Taylor hypothesis.
Weak stationarity tests are often used: the mean is constant and the autocorrelation function only depends on time lag.
Example: mixing fast and slow solar wind streams may lead to non-stationarity.

Example: magnetic field component *By* from WIND: a non-stationary sample.



Examples of non-stationary time series [https://stats.stackexchange.com]



Statistical meaningfulness/2

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 87, NO. A12, PAGES 10,347–10,354, DECEMBER 1, 1982

Stationarity of Magnetohydrodynamic Fluctuations in the Solar Wind

WILLIAM H. MATTHAEUS AND MELVYN L. GOLDSTEIN

"...variances, correlation functions, and power spectra can be meaningfully evaluated from *appropriately selected* finite data intervals."

Or: not all solar wind samples are stationary. (i) They should be long enough or short enough; (ii) Check for shocks and cross-sector boundaries; etc. (read the paper).

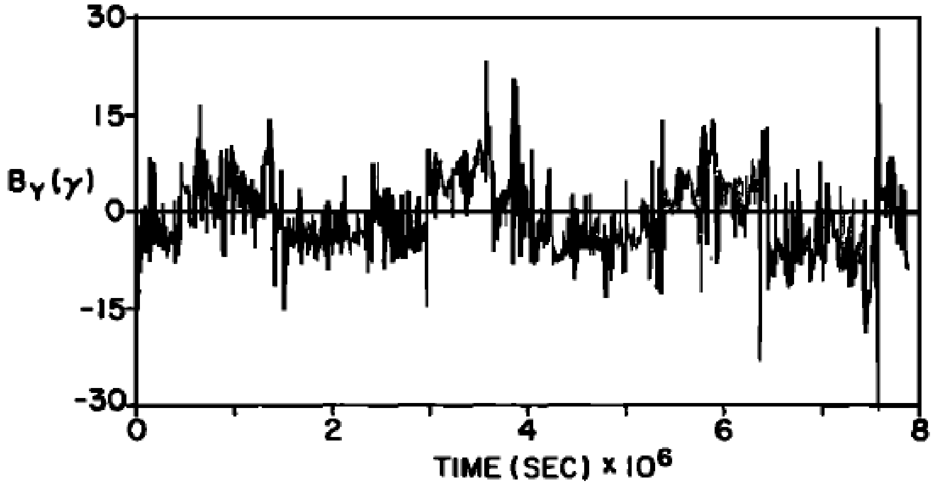


Fig. 7. The tangential (y) component of the magnetic field (in gammas) from the ISEE 3 magnetometer. The interval begins on January 7, 1979, and spans 92 days. A regular sector structure is evident.

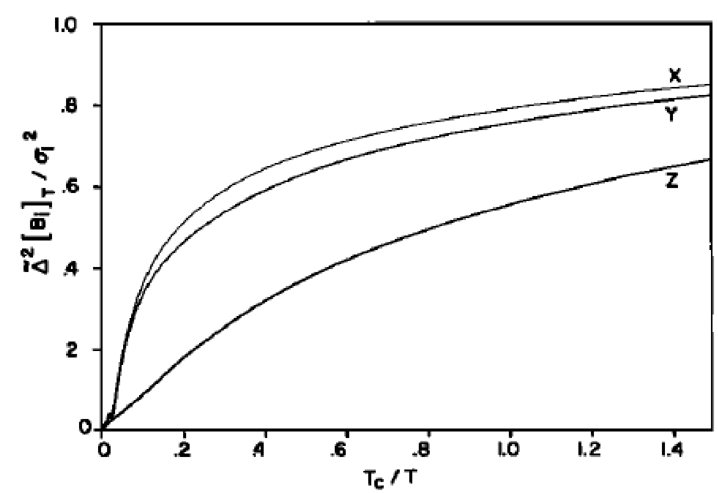


Fig. 3. The variance of the means ($\bar{\Delta}^2[B_i]_T$) for the x, y, and z field components of the 621 day interval are calculated for intervals of duration T and plotted as functions of T_c/T, where T_c is the total correlation time. The variance of the means are normalized by σ_i^2. Convergence of the estimates is evident.

Statistical meaningfulness/3

THE ASTROPHYSICAL JOURNAL, 714:937–943, 2010 May 1

STATIONARITY IN SOLAR WIND FLOWS

S. PERRI[1] AND A. BALOGH[1,2]

"...the stationarity assumption in the inertial range of turbulence on timescales of 10 minutes to 1 day is reasonably satisfied in fast and uniform solar wind flows, but in mixed, interacting fast, and slow solar wind streams the assumption is frequently only marginally valid."

Or: not all solar wind samples are stationary. (i) Hardly below 10 min samples (see also ergodicity issues); (ii) avoid very long samples; (iii) slow wind has to be checked more carefully; (iv) better not mix fast and slow wind together; etc. (read the paper).

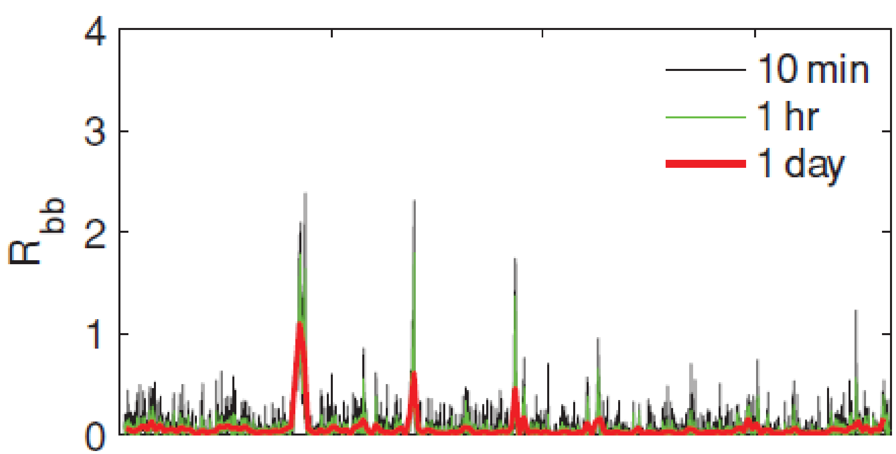
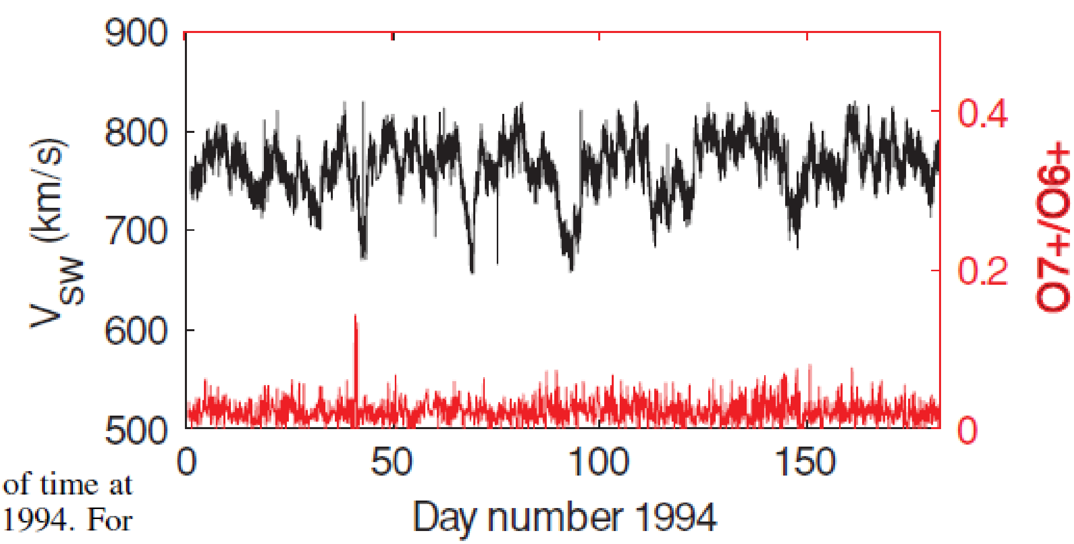


Figure 6. Autocovariance function of the $\delta B/B_0$ ratio as a function of time at three different time lags τ during a 182 day period in fast wind in 1994. For comparison, both the solar wind speed (black line) and the O^{7+}/O^{6+} ratio (red line) are shown in the bottom panel.



Statistical meaningfulness/4

Most of the statistical tools for turbulent data assume **ergodicity** and **stationarity**. Although often verified in space plasmas, this is not always the case and should be checked.

Ergodicity

The assumption that the properties of a finite sample (time series) converge to the properties of the process as the sample size increases (it’s true for all stationary data).
In turbulent time series, ergodicity is usually satisfied if the sample length includes *several eddy-turnover times*.
Example: short time series such as Cluster samples (~30 min in the solar wind) are not ergodic.

Note:
* Ergodicity requires long time series;
* Stationarity requires to avoid long time series;
→ in solar wind, there is need to find the right balance between these requirements.

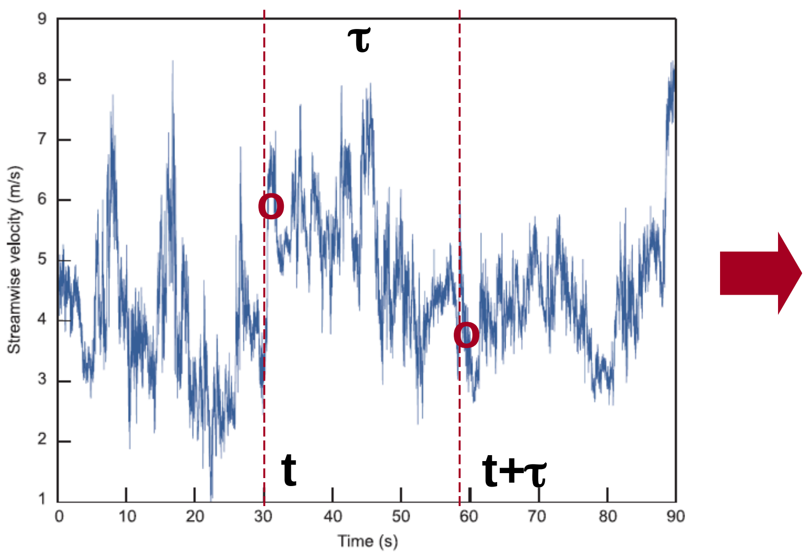
NASA/JPL/SwRI/MSSS/Gerald Eichstädt/Alexis Tranchandon/Solaris

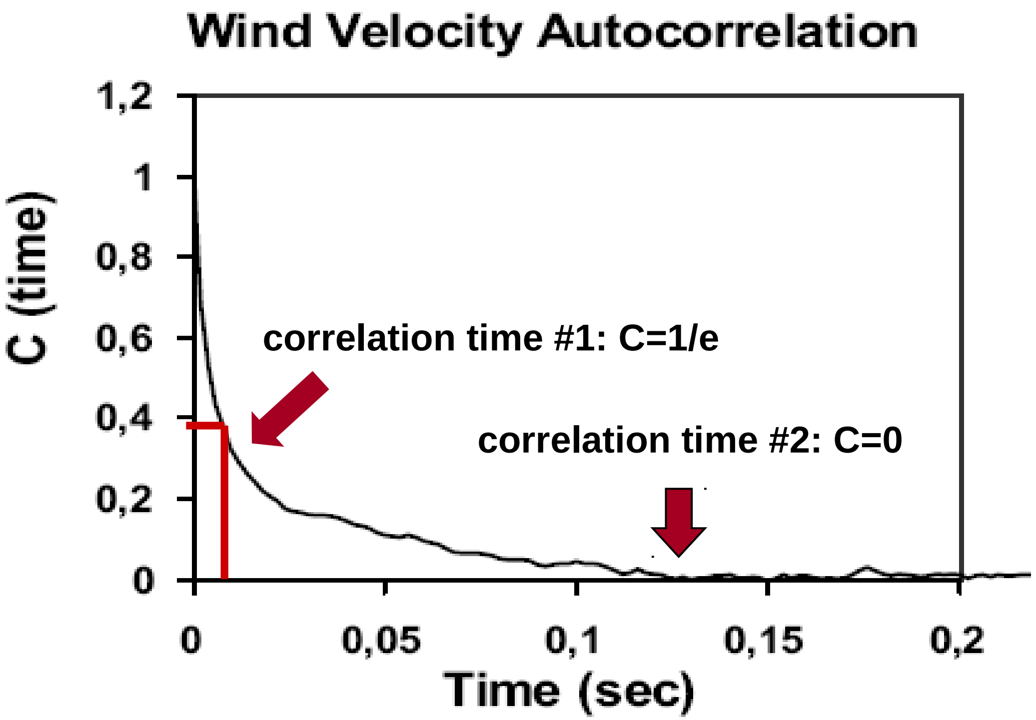
If ergodicity, stationarity or the Taylor hypothesis are not holding, then the analysis could be biased: quantitative results cannot be interpreted in terms of theoretical models of turbulence.

Autocorrelation

Most information about the statistical properties of a turbulent signal is included in the **autocorrelation function**

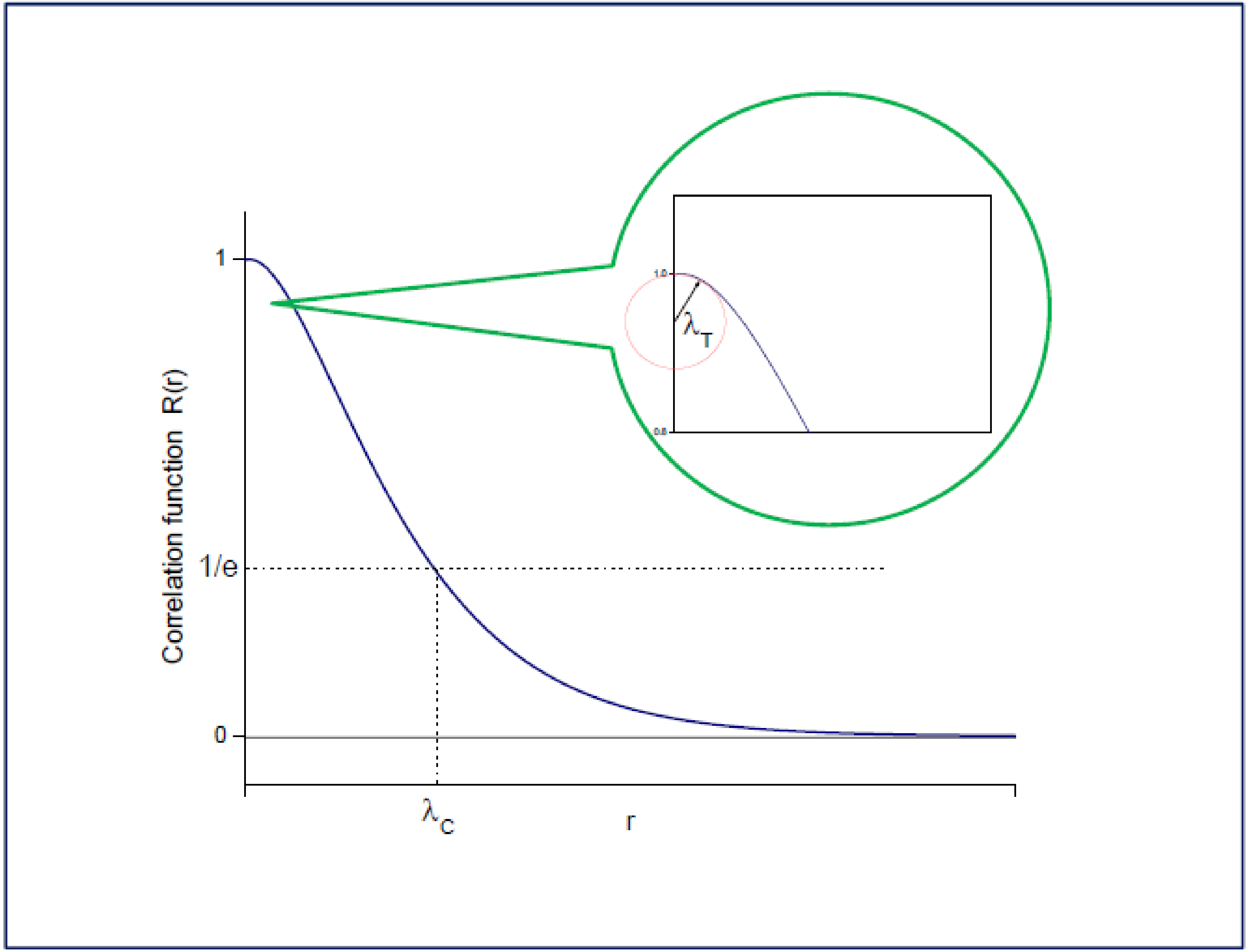
$$C_v(\tau) = \frac{\langle [v(t+\tau) - v_0][v(t) - v_0] \rangle}{\sigma_v^2}$$





The correlation time (different possible operational definitions) gives information about the "memory range" of the system.

Autocorrelation/2



The Taylor microscale can also be estimated using the autocorrelation function as the curvature radius near the origin (fig. from Bruno & Carbone Liv. Rev. Sol. Phys., 2013). Not easy in space plasma data.

Power Spectral Density

Similar information is carried by the power spectral density (PSD), in the frequency domain.

$$E_\nu(f) = |\hat{\nu}(f)|^2 = \int_{-\infty}^{\infty} C_\nu(\tau) e^{-2\pi i f \tau} d\tau$$

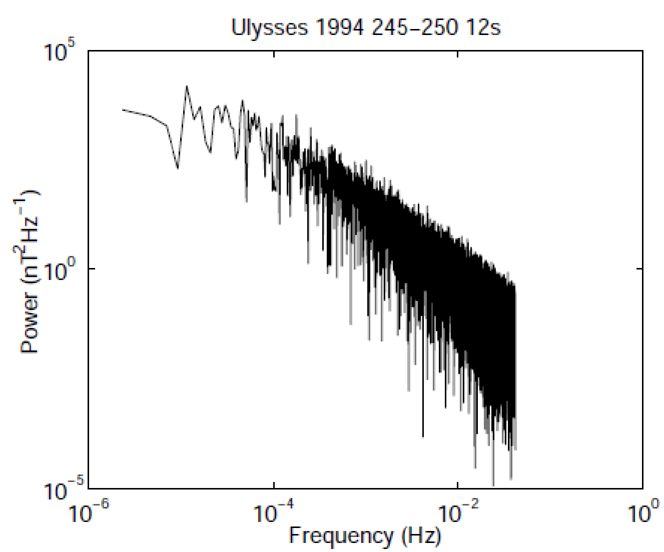
Various techniques can be used to estimate it: Fourier Transform, FFT, Fourier transform of the autocorrelation function (via the Wiener-Kinchine theorem, good for data with gaps), “multitaper” techniques, etc.

Parseval’s theorem:
The total variance of the time series is the sum of the squares of the Fourier coefficients.
Therefore, with the Fourier transform we can identify the variance in the signal at different frequencies

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2.$$

The highest measured frequency is the *Nyquist frequency*, $f_N = f_s/2$.
To access higher frequencies, must use higher time resolution data. No way around.

The lowest measured frequency is $f_1 = 1/T$
To access lower frequencies, must use longer data set. No way around.



Power Spectral Density/2

A few caveats:

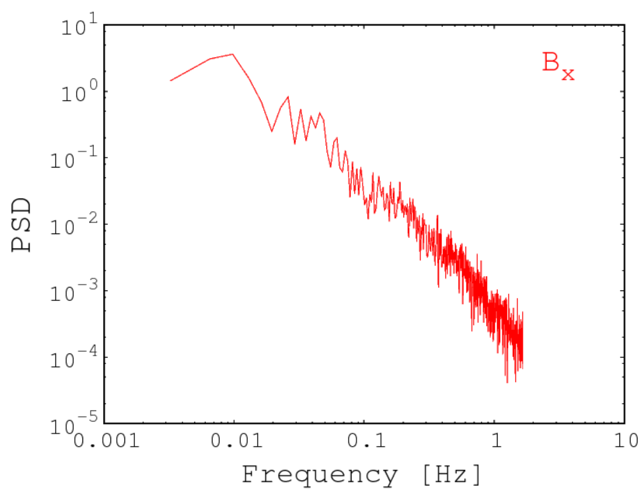
* For most estimators, signal must be stationary and “periodic”.

* Most of the FFT routines require power of 2 data points: use windows not to waste data.

* To increase PSD quality and reduce spurious fluctuations (noise), make abundant use of **windowing** (various types: Hamming, Hanning...). It also reduces the effects of discontinuities at the sample boundaries. It comes at the costs of losing low frequencies.

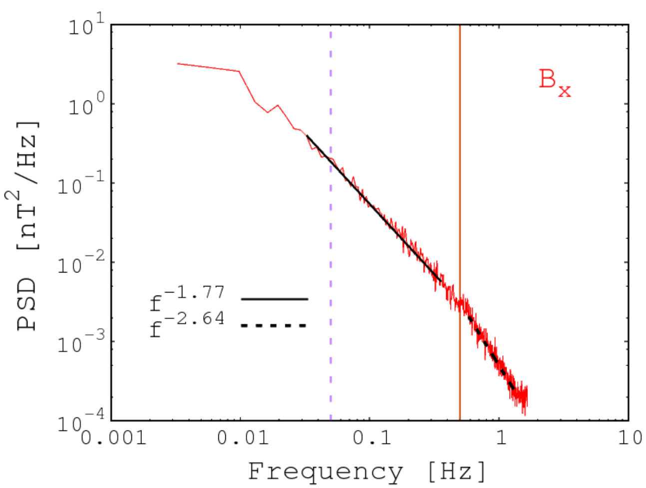
* Don’t forget to consider aliasing: power outside the accessible range will be distributed in the spectrum.

* Evaluation of error: for example use sub-samples to estimate convergence. Or use 95% confidence level (variance of PSD is χ^2 distributed).



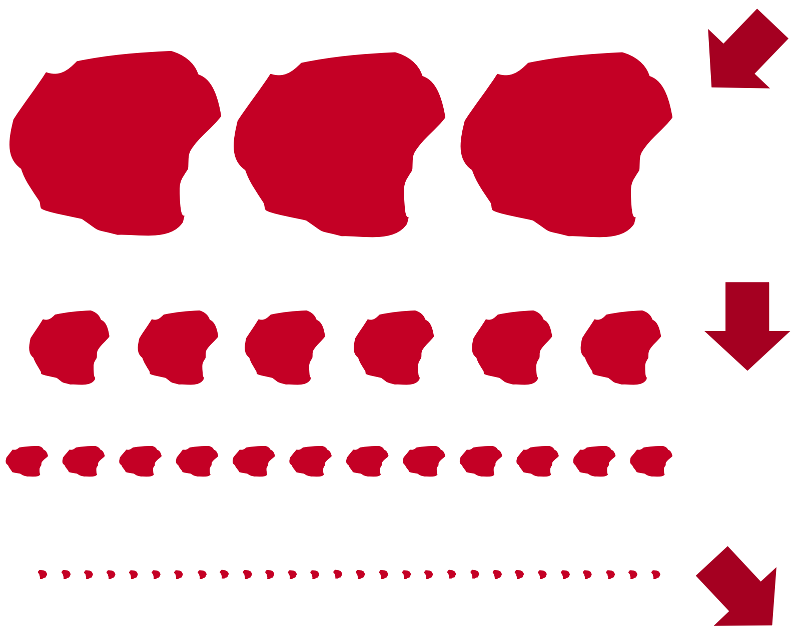
Example: PSD of the magnetic field component *Bx* from the same MMS data in the magnetosheath.

Left: no windowing; Right: 8 windows.



Power Spectral Density/3

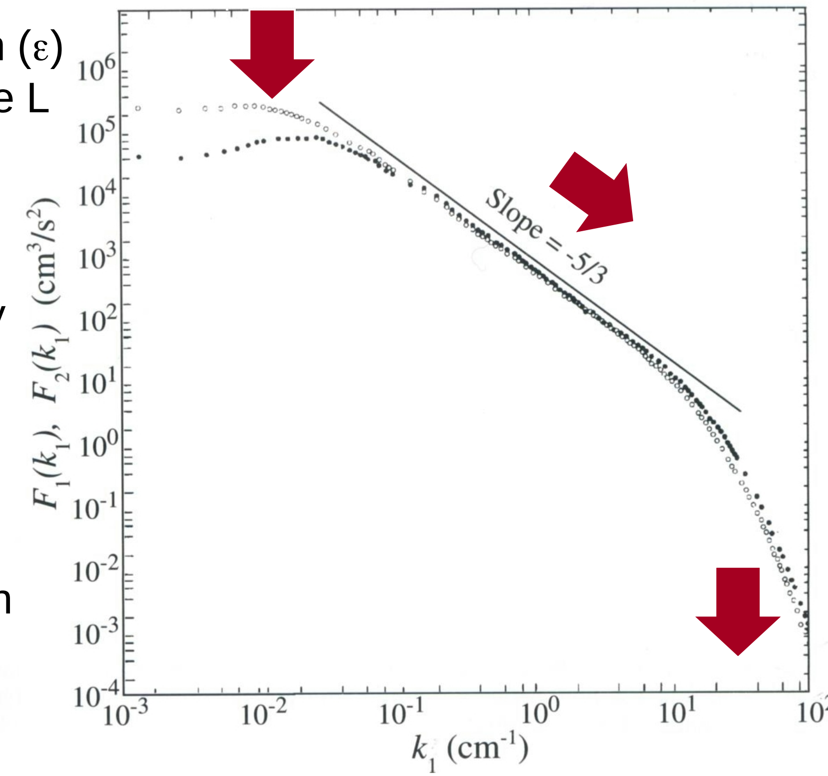
Kolmogorov 1941 phenomenology of turbulence predicts the power-law scaling of the PSD in the inertial range, representing the energy cascade across scales (the Richardson cascade):

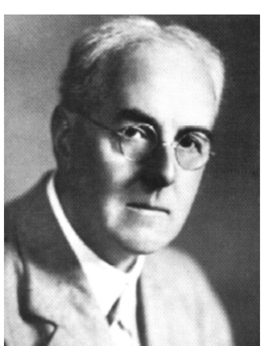


Energy injection (ε) at integral scale L

Non-linear energy transfer (ε) in the inertial range

Energy dissipation (ε) at small scale





$$E_v(k) \propto k^{-5/3}$$

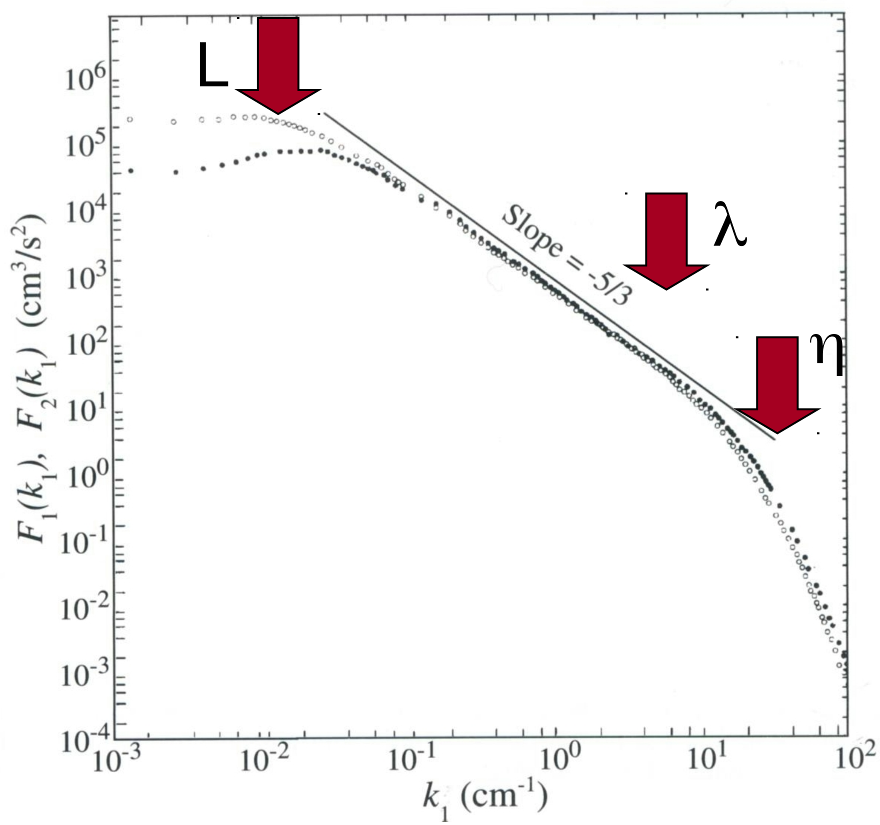
Taylor

$$E_v(f) \propto f^{-5/3}$$

The PSD scaling exponent may give insight on the phenomenology

Power Spectral Density/4

In Navier-Stokes turbulence, three fundamental scales can be introduced: the integral scale L (the eddy-turnover time, usually corresponding to the correlation scale); the Taylor microscale λ, representing the typical size of the intermittent structures; and the dissipation scale η. These can be identified in the spectrum:



Integral scale

$$L = \frac{\int_0^\infty dk\; E(k)/k}{\int_0^\infty dk\; E(k)}$$

Taylor microscale

$$\lambda = \frac{\int_0^\infty dk\; E(k)}{\int_0^\infty dk\; k^2 E(k)}$$

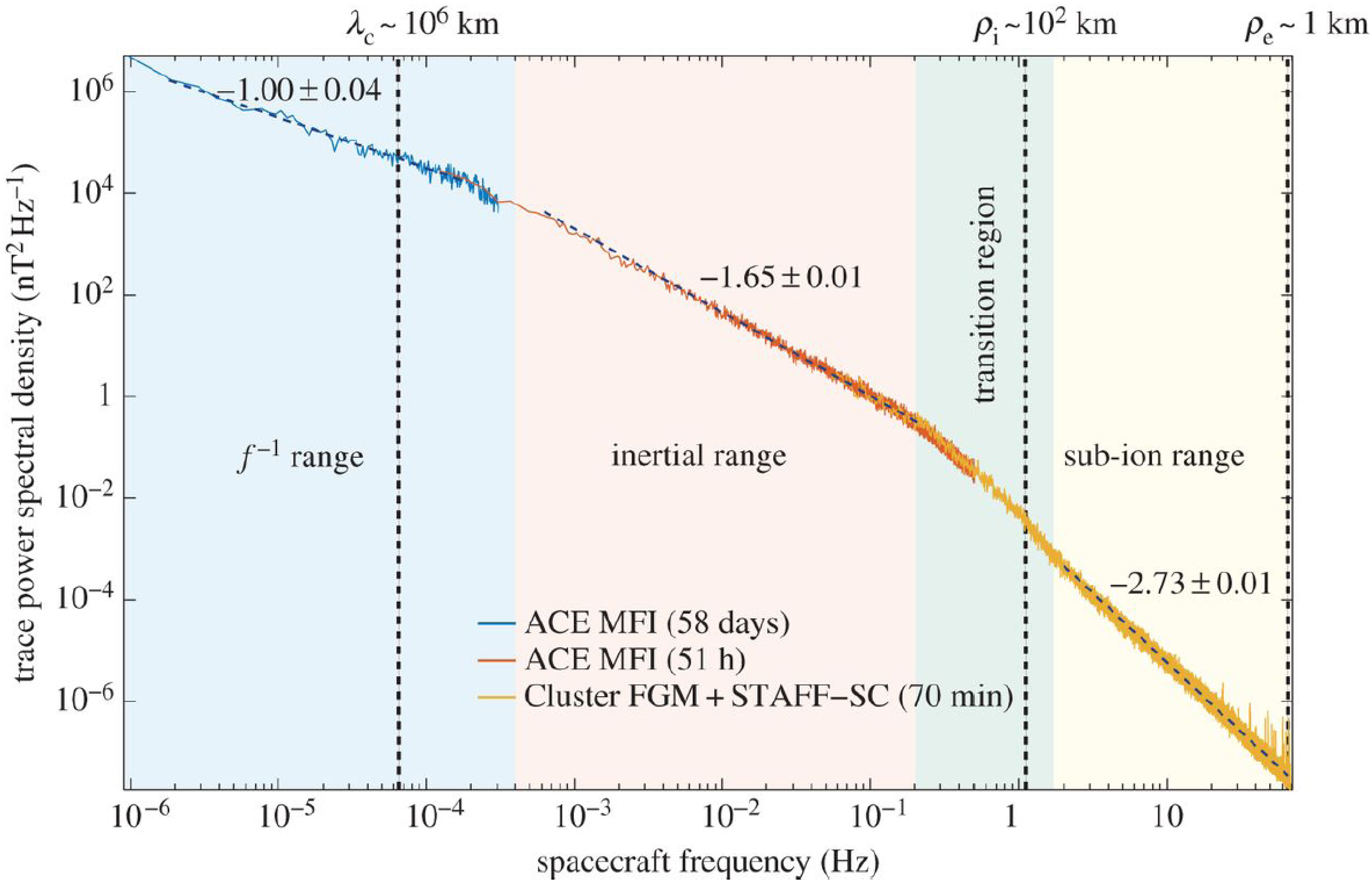
Dissipative scale

$$\eta = \left[\frac{\upsilon^3}{\varepsilon}\right]^{1/4}$$

In plasmas: more physical scales exist (e.g. the Alfvén time-scale, the typical ion and electron scales, etc.) and can often be identified through the PSD.

Power Spectral Density/5

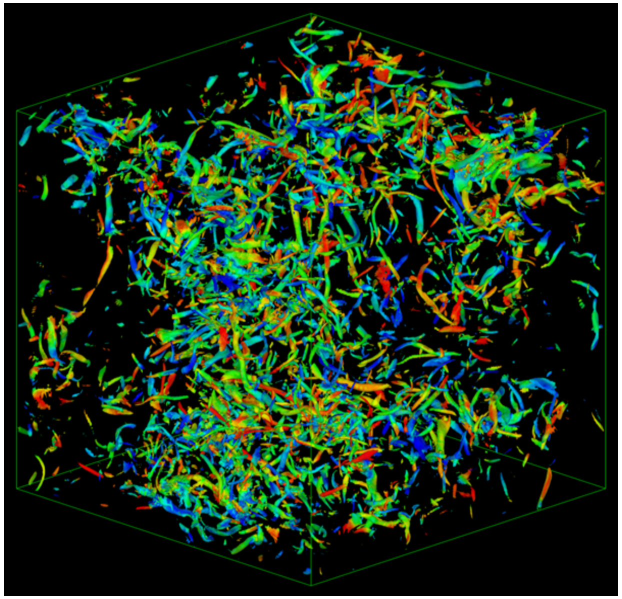
In solar wind turbulence spectra are more complex (more about this tomorrow)



Khurom H. Kiyani et al. Phil. Trans. R. Soc. A 2015;373:20140155

Intermittency: high order statistics

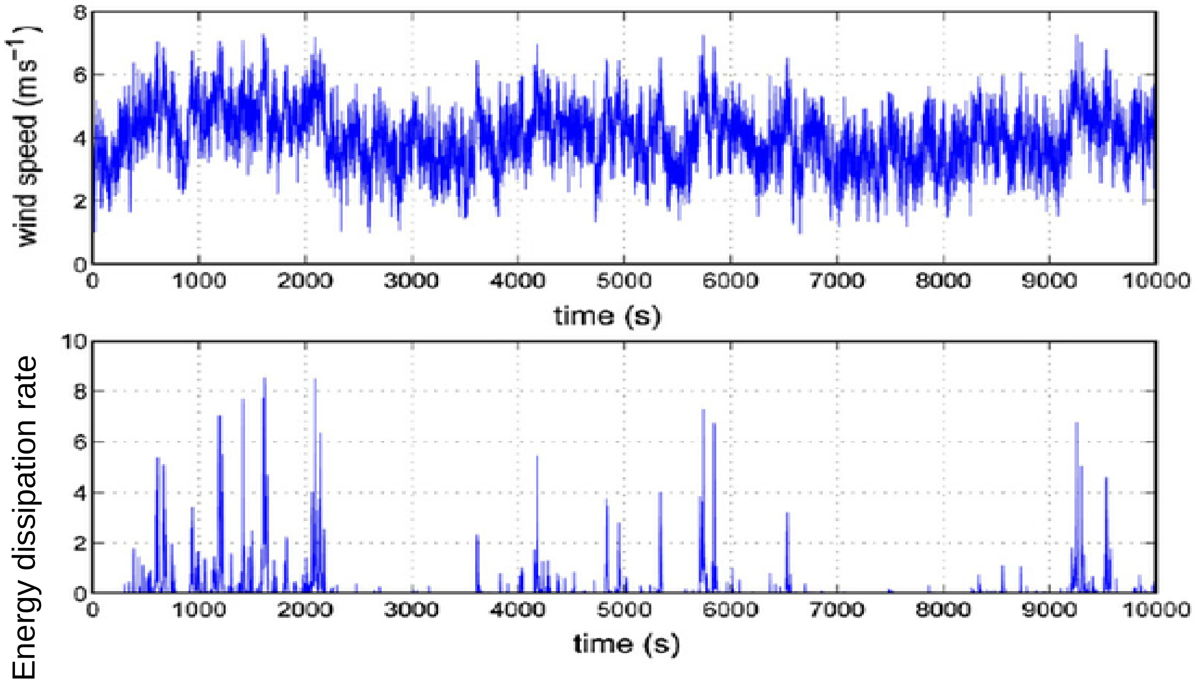
Turbulent dissipation in numerical simulations

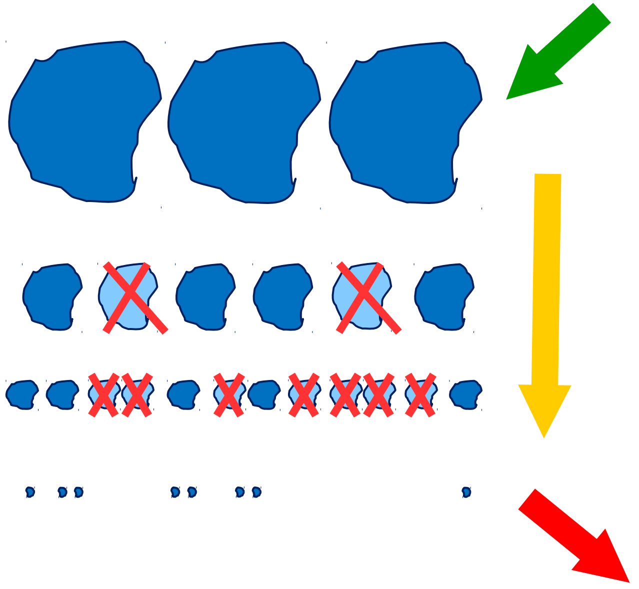


Universal feature of turbulence: **intermittency**

Energy dissipation is more efficient when it occurs in vortices distributed inhomogeneously in space.

Experimental observation of intermittent, bursty, multifractal dissipation fields.

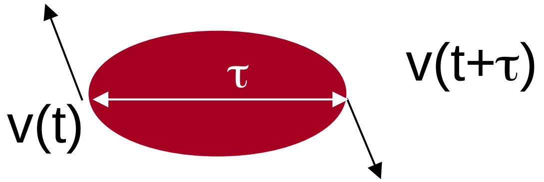


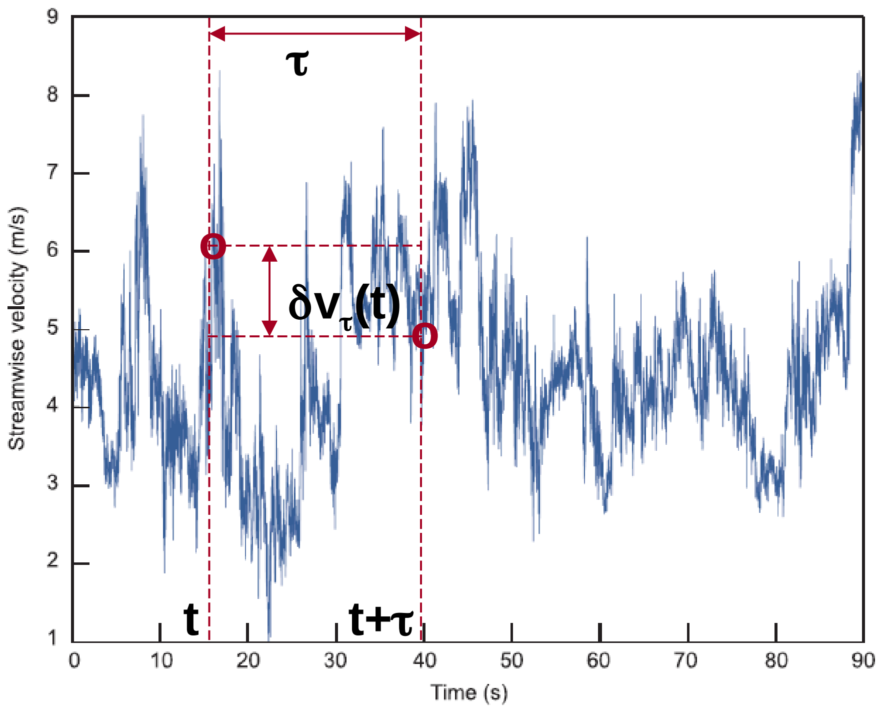


Intermittency: velocity increments

Turbulent fluctuations studied through the analysis of velocity differences at different time lags

$$\delta v_\tau = v(t+\tau) - v(t)$$

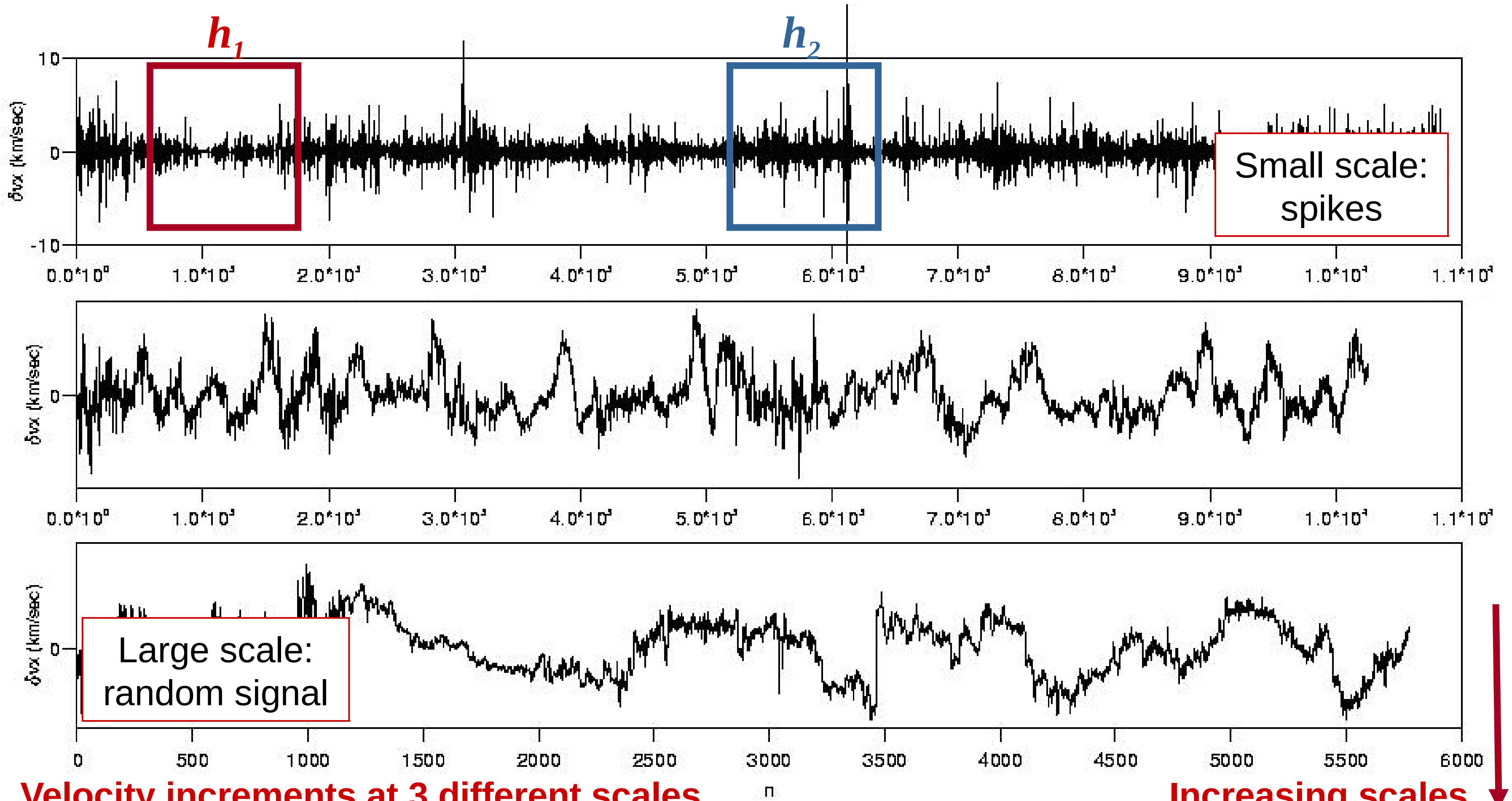




This variable describes the field fluctuations at a given scale, providing information about scale distribution of turbulent structures (e.g. vortices).

Correnspondence between the scale t and the spectral frequency f.

Intermittency: velocity increments/2



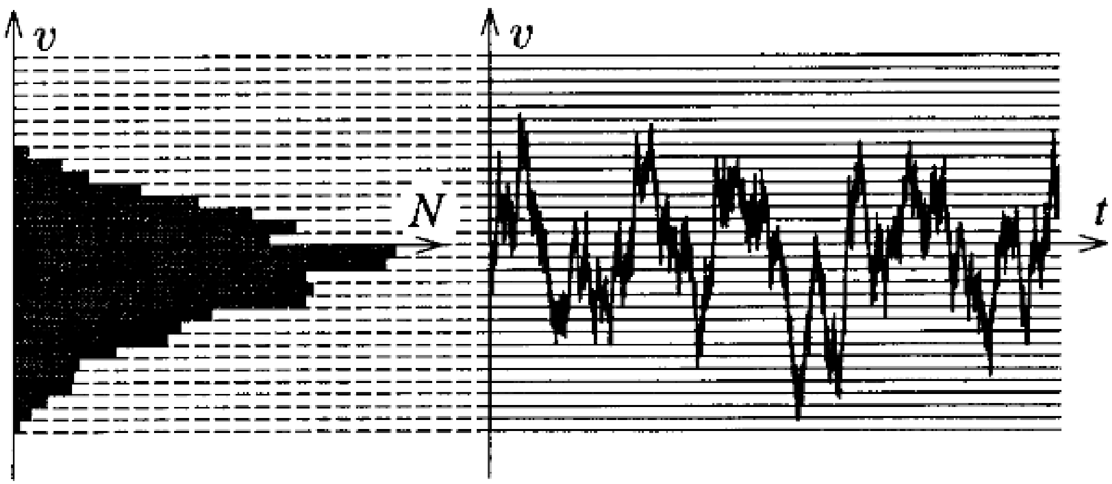
Velocity increments at 3 different scales

Increasing scales

Statistical properties are increasingly inhomogeneous as the scale decreases.
Second-order statistics (variance, spectrum) are not enough to describe the statistics: the whole PDF, or its high-order moments, are necessary.

Intermittency: PDFs

Probability Density Functions (PDFs) are usually computed as normalized histogram of data



$$P_i(\delta v_\tau) \equiv \frac{N_i(\delta v_\tau)}{N_{tot} A_i}$$

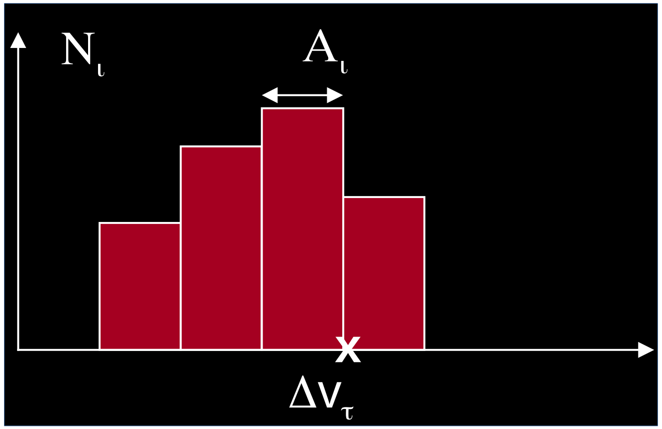
$N_i(\Delta v_\tau)$ is the number of fluctuations in the i-th bin $[\Delta v_\tau - A_i/2,\ \Delta v_\tau + A_i/2]$.

A_i is the size of the interval (bin).

N_{tot} is the total number of data used to compute the PDF.

* Fix number and size of bins, and variation range (in terms of sample standard deviations σ) to ensure significant statistics in each bin (at least ~10 points).
* Estimate error bars of the histogram as standard Poissonian deviation ($\sqrt{N}$), then propagate to the PDF.
* Plot Log(P), since it enhances tail details.
* To compare different scales and/or different databases, standardize the variables:

$$\delta v_\tau'(t) = \frac{\delta v_\tau(t) - \langle \delta v_\tau(t) \rangle}{\sigma_{\delta v_\tau}}$$



Intermittency: PDFs/2

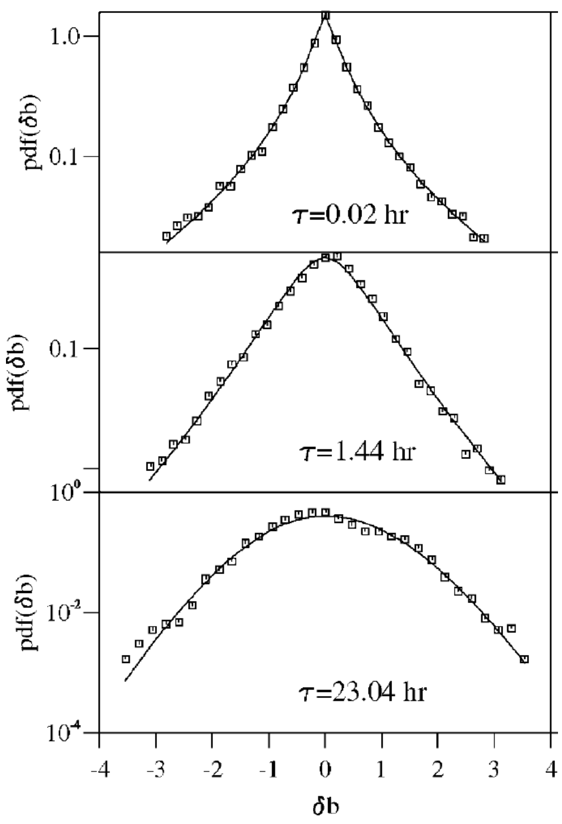
All the statistical information is contained in the PDFs

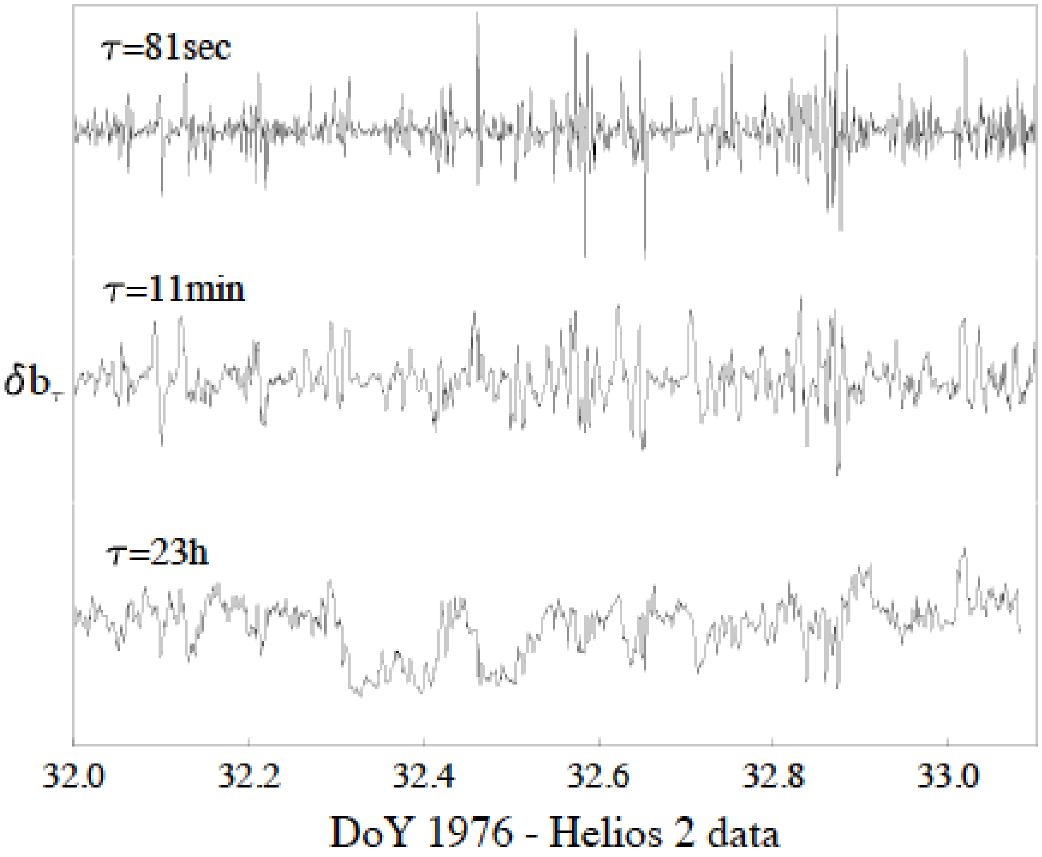
Example from plasmas: PDF of solar wind magnetic field increments at 3 different scales

Small scale: high tails

Large values of δb are more probable than for Gaussian fields: intermittency

Large scale: nearly Gaussian





Intermittency: moments

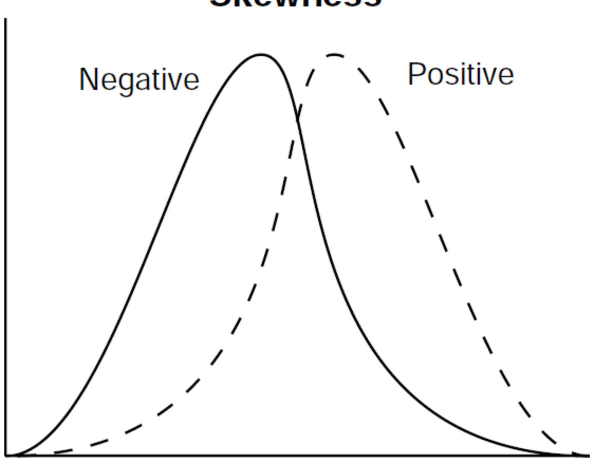
Alternative information can be extracted from PDFs through the estimation of the p-th order moments, called **structure functions**.

Moment	Name	Gaussian value
1	mean	0
2	variance	σ^2
3	skewness	0
4	flatness (kurtosis+3)	$3\sigma^2$
5	-	0
6	Order 3 hyperflatness	$15\sigma^2$

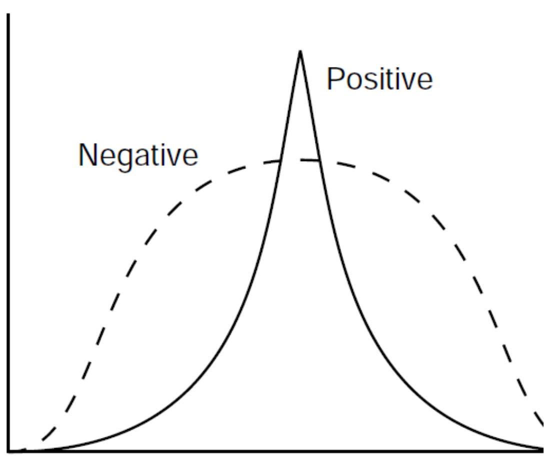
$$S_p(x) \equiv \int_{\infty}^{\infty} x^p P(x)dx \cong \langle x^p \rangle$$

Tips:
* Do not standardize variables when computing structure functions.
* Often necessary to use absolute value of fluctuations.
* Computation needs to be done very carefully (e.g.: sort data for summing; consider removing outliers, etc.).
* Check convergence (next slide).
* Try to estimate errors, e.g. by using sub-ranges.

Skewness

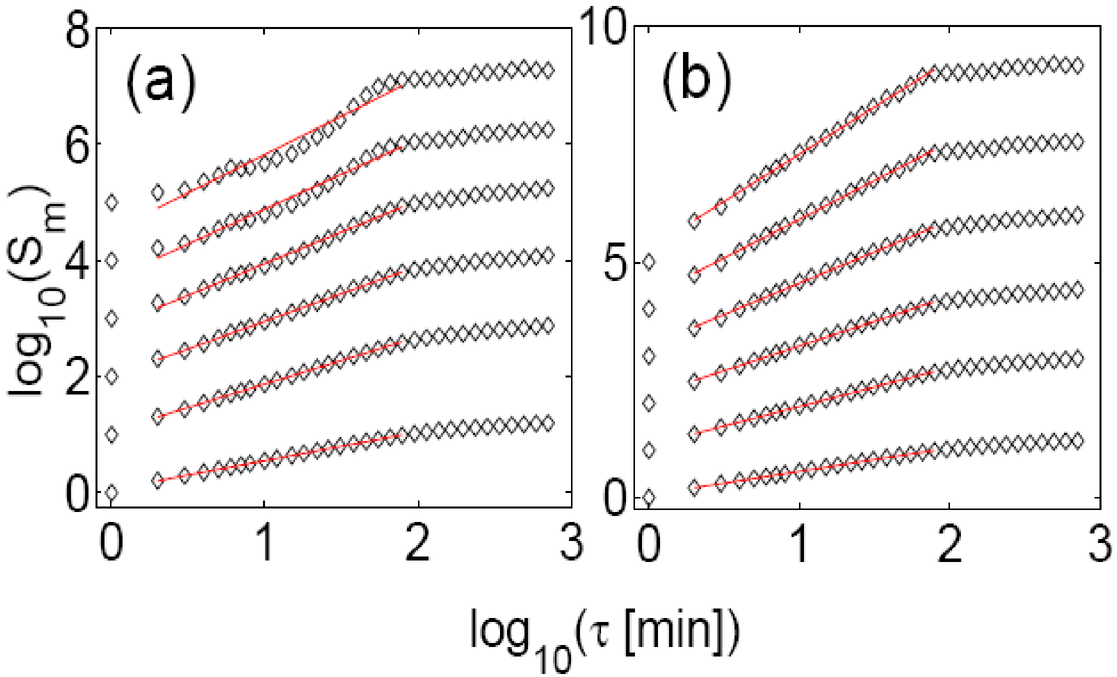


Kurtosis



Intermittency: moments/2

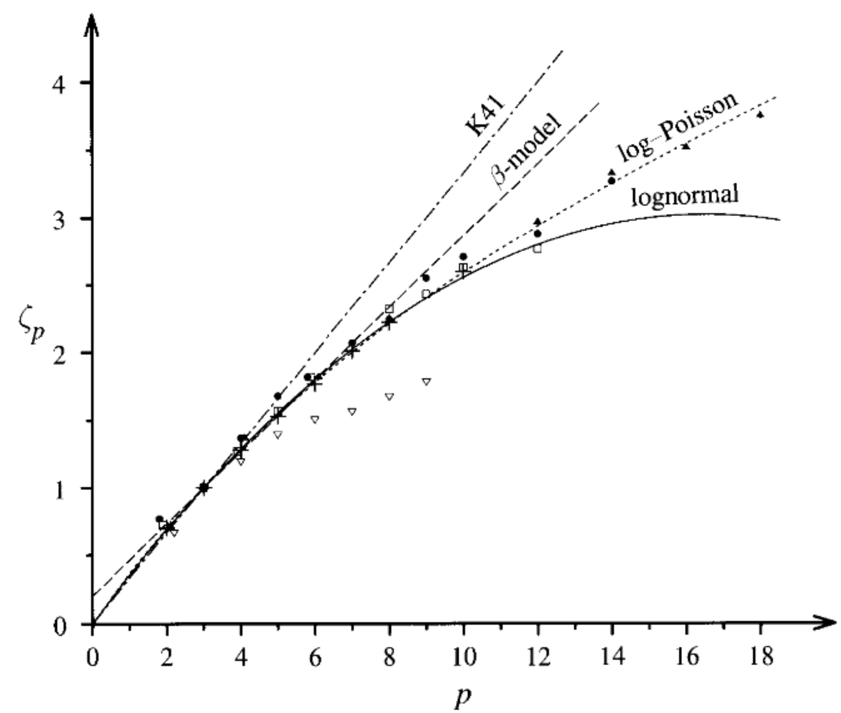
For turbulent flows, power-law scaling is observed in the inertial range: $S_p(\delta v_\tau) \sim \tau^{\xi(p)}$.



Intermittency effects can be quantified through the study of the structure function scaling exponents. Many models predict the anomalous scaling, and can be tested on data

For self-similar (Gaussian) PDFs $S_p(\delta v_\tau) \sim \tau^{p/3}$

Intermittency: anomalous scaling $S_p(\delta v_\tau) \sim \tau^{\xi(p)}$



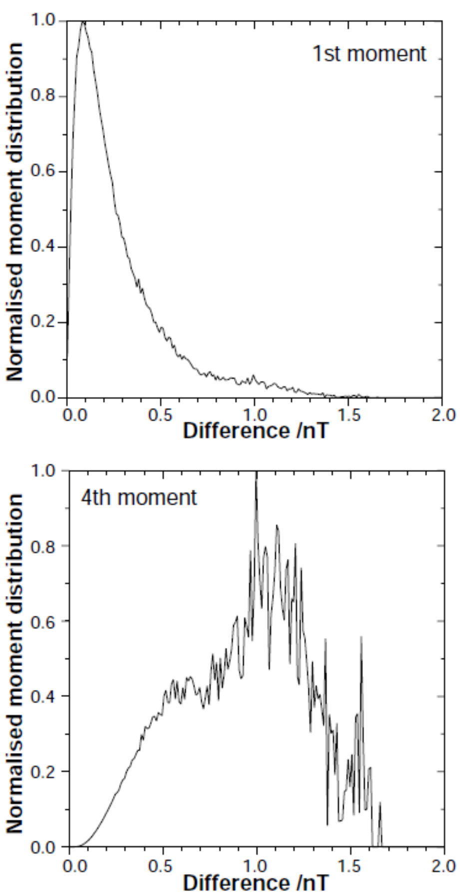
Intermittency: moments/3

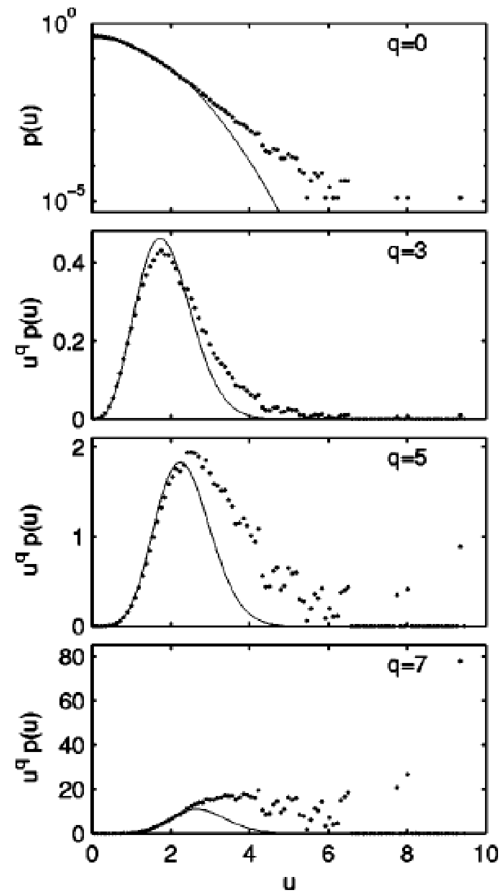
The problem of moments convergence

The estimate of structure functions is difficult for high orders p. Large datasets are required for statistical convergence. This needs to be checked (see also next slide).

Plot the function to be integrated: $x^p P(x)$ and observe its convergence at different p.

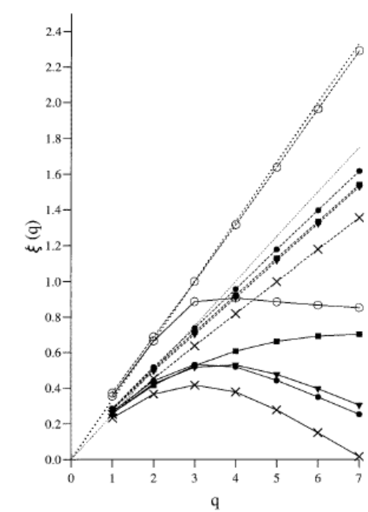
$$S_p(x) \equiv \int_{\infty}^{\infty} x^p P(x) dx \cong \langle x^p \rangle$$





Rule of thumb:
$p_{max} \sim \log_{10}(N_{tot})$

In solar wind turbulence, need to find a balance between stationarity, ergodicity, and statistical significance of moments



Intermittency: moments/4

The problem of moments convergence

PHYSICAL REVIEW E 70, 055302(R) (2004)

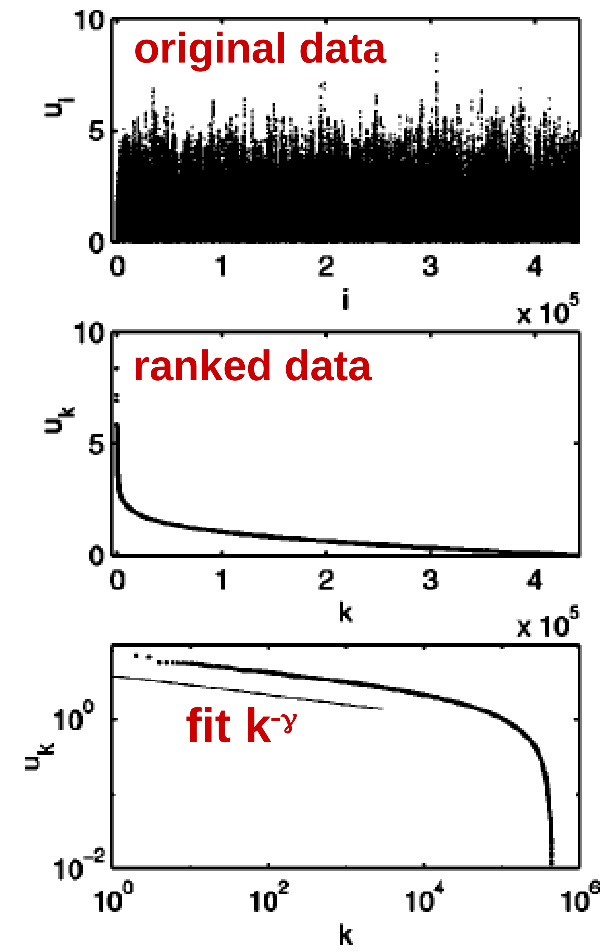
Can high-order moments be meaningfully estimated from experimental turbulence measurements?

T. Dudok de Wit*

Given the data time series, the plot of the ranked data shows power-law behavior (with exponent $-\gamma$) near larger values. Convergence of the structure function is ensured up to the order q_{max}, where:

$$u_k = \alpha \left(\frac{k}{N} \right)^{-\gamma},$$

$$q_{max} = \left\lfloor \frac{1}{\gamma} \right\rfloor - 1,$$



Intermittency: moments/4

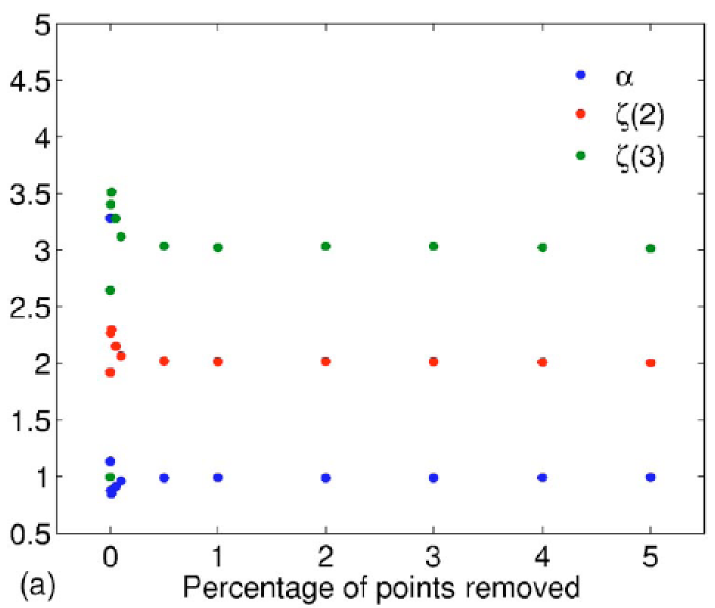
The problem of moments convergence

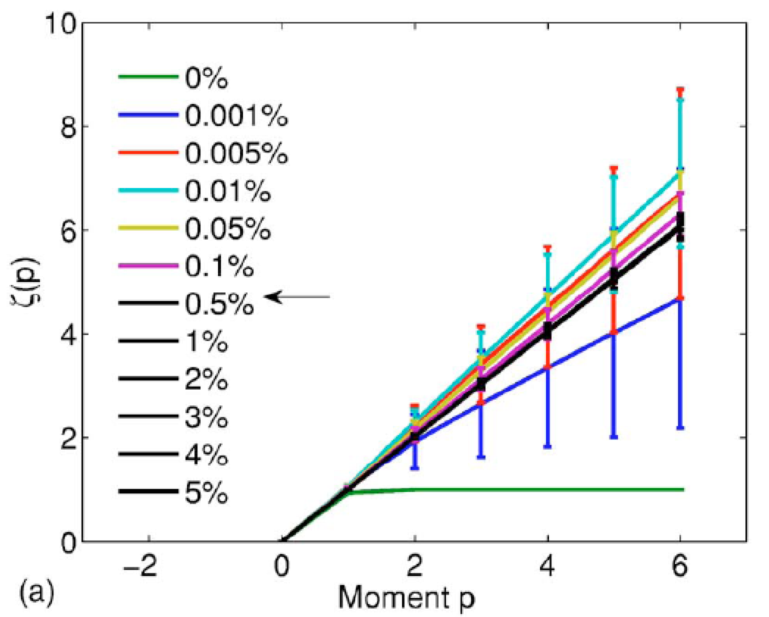
PHYSICAL REVIEW E **74**, 051122 (2006)

Extracting the scaling exponents of a self-affine, non-Gaussian process from a finite-length time series

K. Kiyani,* S. C. Chapman, and B. Hnat

It may be useful to remove the most extreme values, which control the high-order moments but could lack statistical significance. Convergence depends on the dataset, but is generally achieved remving 0.5-0.1% of data.





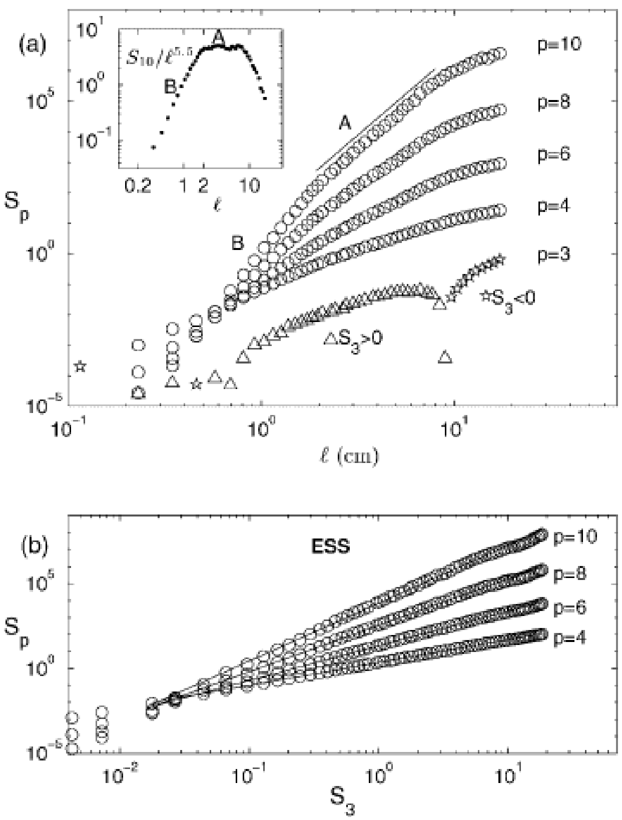
Intermittency: moments/5

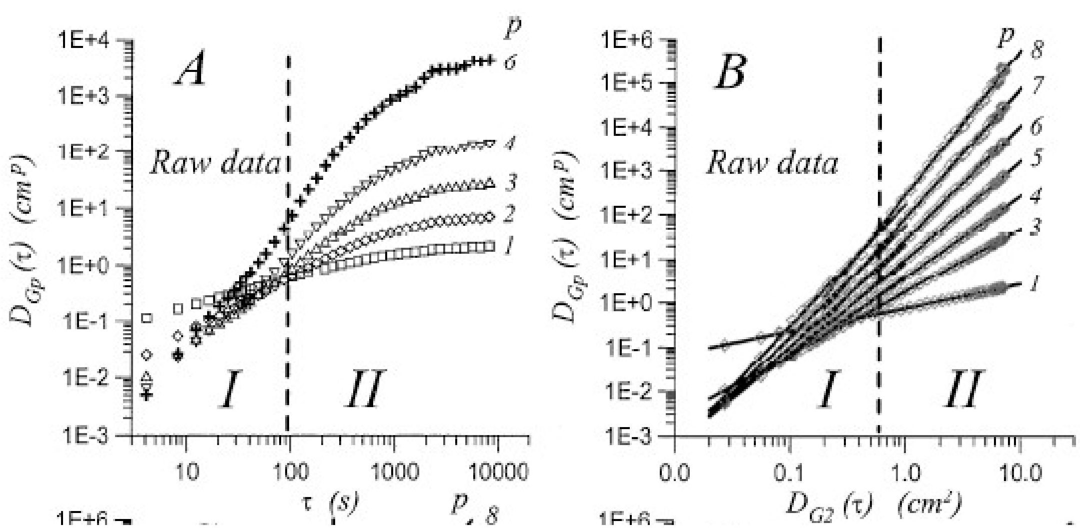
The problem of moments convergence

If, despite all cares, moments still do not have good power-law ranges, go magic and use **Extended Self-Similarity** (ESS), based on the linear Kolmogorov 4/5 exact law $S_3(\tau)= 4/5\varepsilon\tau$.

$$S_p(\tau) \propto \tau^{\xi(p)} \Rightarrow S_p(S_3) \propto S_3^{\xi(p)}$$

$$\zeta(p) = \xi(p)\zeta(3)$$





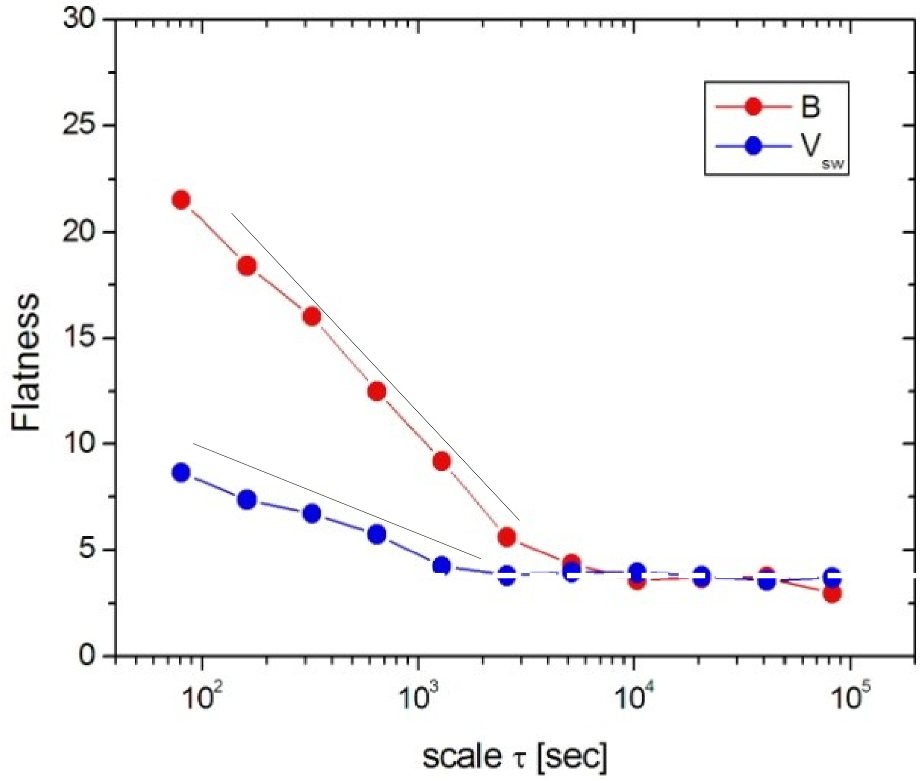
R. Benzi et al., Phys. Rev. E 48, 29 (1993).

Intermittency: kurtosis (or flatness)

A practical tool for quantitative analysis of intermittent turbulence: the scaling of the normalized fourth-order moment, the **flatness**: $F(\tau)=S_4(\tau)/S_2^2(\tau)$.

For Gaussian PDF $F(\tau)=3$, so that Kurtosis is often used $K(\tau)=F(\tau)-3$. Deviation from $F(\tau)=3$ is used as indication of intermittency.

"A random function is intermittent at small scales if the flatness grows without bound at smaller and smaller scales" (Frisch, 1995)



The flatness has usually power-law scaling in the inertial range, with exponent ~0.1. In the solar wind, exponent is larger and not universal (strongly dependent on shocks and other structures)

For good dataset, *n*th-order (n>2) hyperflatness can also be used:

$$H_n(\tau)=S_{2n}(\tau)/(S_n(\tau))^2$$

(Don't even think about it in solar wind data)

Intermittency: models

None of the above is useful unless there are models to describe the physics.

Lognormal (Frisch)

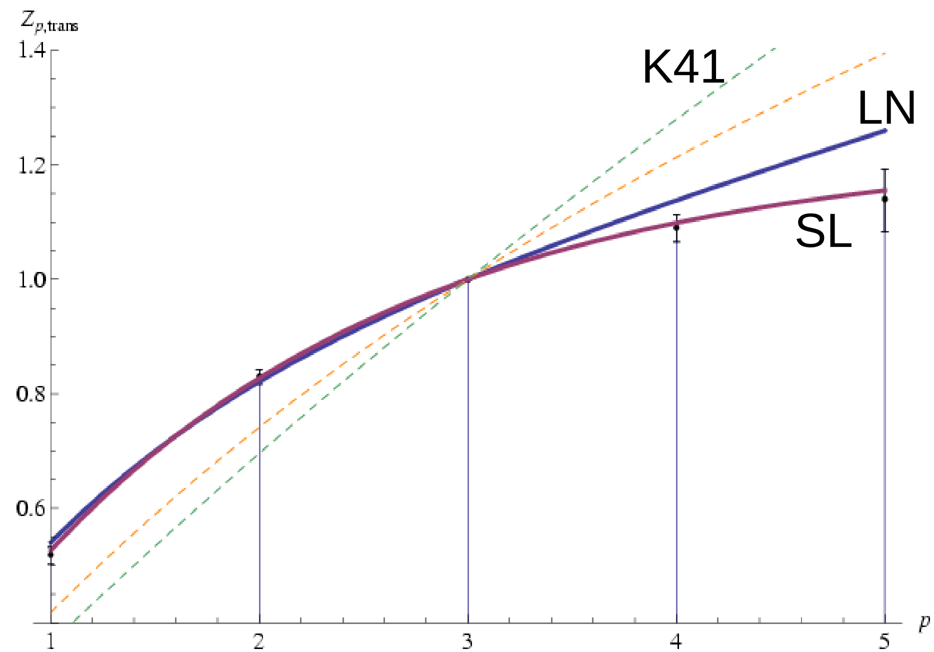
$$\zeta_p^{(LN)} = \frac{p}{3} + \frac{\mu}{18}\left(3p - p^2\right)$$

Random-β (Frisch et al.)

$$\zeta_p^{(\beta)} = \frac{p}{3} - \log_2 \langle \beta^{1-p/3} \rangle$$

She-Leveque

$$\zeta_p^{(SL)} = \frac{p}{9} + 2\left[1 - (2/3)^{p/3}\right]$$



Multifractal (Parisi & Frisch)

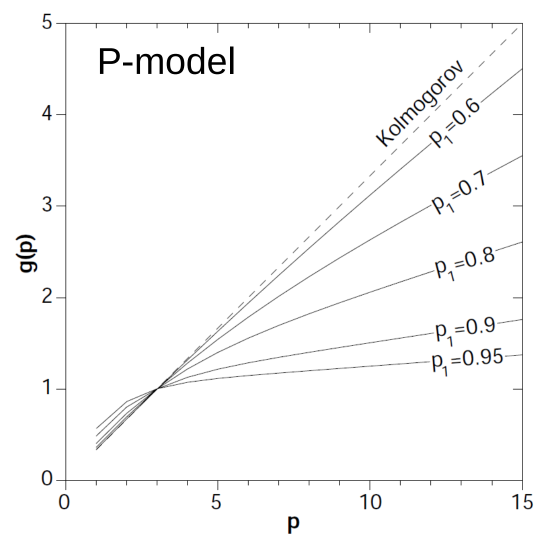
$$\zeta_p^{(MF)} = \inf_h \left[ph + 3 - D(h)\right]$$

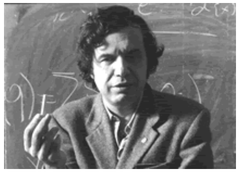
D(h) being the multifractal dimensions

P-model (Meneveau & Sreenivasan)

$$\zeta(m) = 1 - \log_2\left(p^{m/3} + (1-p)^{m/3}\right)$$

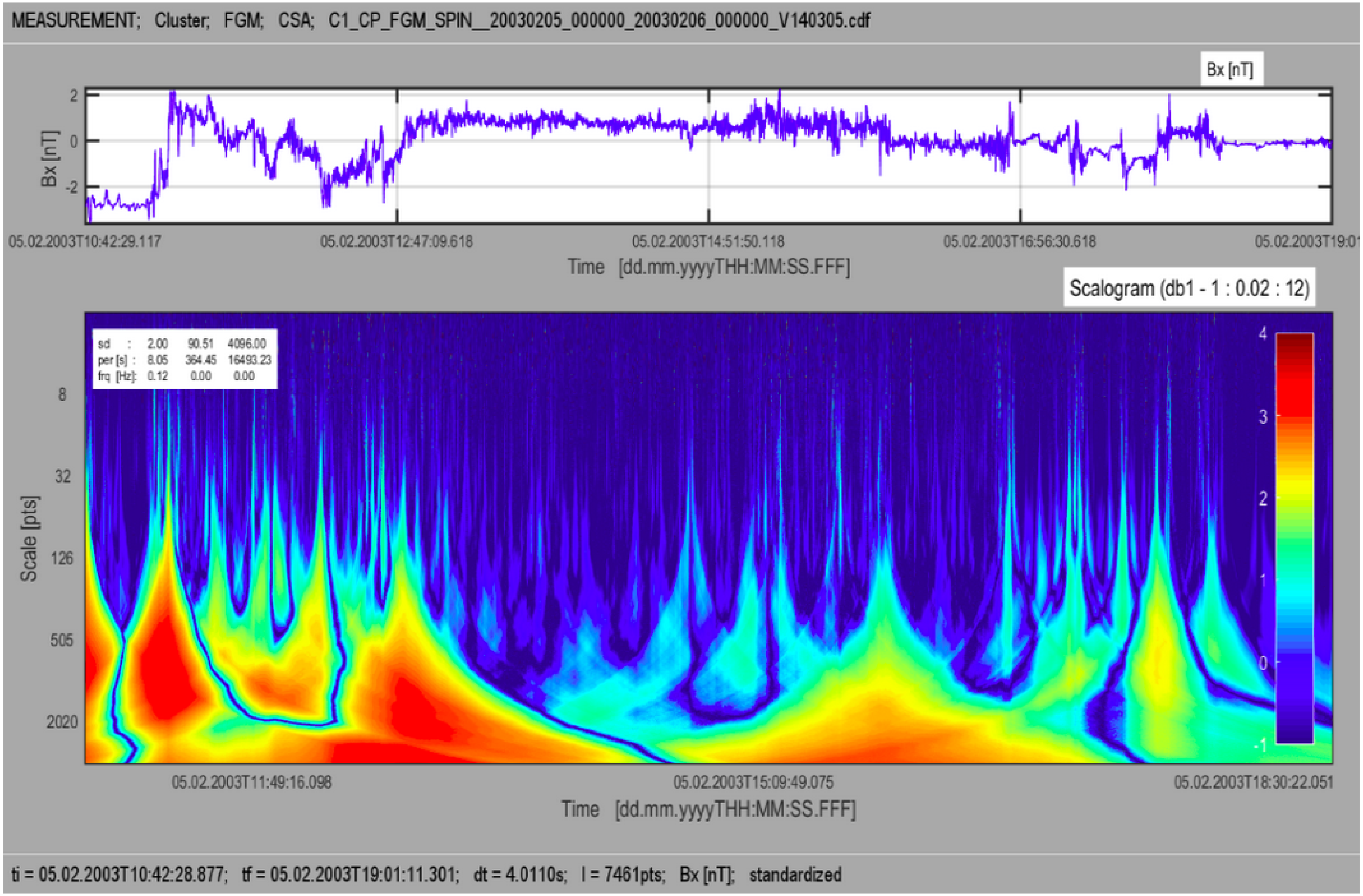
Parameter P gives quantitative measure of intermittency





Wavelets?

Almost all of the above is best done using wavelet transform, which add local information (trade-off with frequency accuracy though). Scalograms are useful.



Summary

- Know your data.
- Be careful with any data analysis tools.
- Possibly avoid black-box packages where you don’t control the tools.