Data analysis tools to investigate multi-scale turbulence based on high resolution plasma and electromagnetic field measurements

STORM (Solar system plasma Turbulence: Observations, inteRmittency and Multifractals) is an European FP7 project that investigates solar system plasma turbulence through a thorough survey of the existing data bases (e.g. Ulysses, Cluster, Venus Express). One of the goals of the project and one of its major deliverables is to build and test an integrated software library for nonlinear data analysis (INA) that cumulates analysis methods able to reveal the structure and topology of turbulent fluctuations. The library is versatile enough to ingest data from various missions and to apply a full package of analysis methods, ranging from lower order analysis, like the Power Spectral Density analysis, to higher order analyses like the Probability Distribution Functions and their moments (e.g. flatness), wavelet analysis and multifractals. The integrated library is equipped with a graphical user interface (see figure 1) that enables an interactive analysis and a full control of the analysis parameters. The library is extensively tested with ESA missions data (e.g. Cluster, Ulysses, Venus Express, see figure 2). A brief description of the analysis methods integrated at this moment in INA is given below:

1. **The Power Spectral Density (PSD)**: The power spectrum of the Fourier transform (the representation of the signal by sines and cosines) is traditionally used to estimate the distribution of power over frequency (temporal or spatial) ranges The library adopts the Welch algorithm that reduces the noise of the PSDs and is fully customizable by the user. The library also computes spectrograms that illustrate the variation of the spectrum in time/space.

2. **The Probability Distribution Functions (PDF)** The probability density function (PDFs) of a variable A at a scale $\tau=n\delta$ (n = integer, δ = resolution of measurements) is computed as the normalized histogram of the differences $\Delta A=A(t+\tau)-A(t)$. A normalized scale dependent measure related to the fourth order moment of the coarse-grained probability distribution function is called the Flatness. It indicates whether the data are more peaked or flatter relative to the Gaussian distribution and is a descriptor of intermittency.
3. **The Structure function** analysis: is based on the computation of ensemble averages of the moments of order *m* of the PDFs
$$S^m(\tau) = \langle |A(t+\tau) - A(t)|^m \rangle$$
where <> means ensemble average over the entire statistical set of differences. The structure function analysis reveals the topology and scale behavior of fluctuations of plasma and electromagnetic field variables and provides a measure of self-similarity and fractal behavior.
4. **The wavelet analysis** including the **Local Intermittency Measure** (LIM). The library implements a fully customizable representation of the time series in wavelet basis (all the classical mother wavelet functions are available). In addition to the wavelet representation the library also implements an estimator of the intermittency level as a normalized map of the wavelet coefficients
$$LIM(\tau, t) = \frac{|C(\tau, t)|^2}{\langle |C(\tau, t)|^2 \rangle}$$
where the average <…> is computed over the entire time interval for each scale (Farge).
5. **The Partition Function multifractal analysis** (PFMA) based on the concept of structure and partition function have been applied on space plasma data only recently (Pagel and Balogh, 2001; Burlaga, 2001; Voros et al., 2004, Yordanova et al., 2004; Macek et al., 2005). The multifractals are described by an infinite number of the generalized dimensions, Dq and by the multifractal spectrum $f(\alpha)$. The generalized dimensions Dq are calculated as a function of a continuous index q. that can be compared to a microscope exploring different regions of the singular measurements. In the case of turbulence cascade the generalized dimensions are related to the inhomogeneity of the energy transfer between different scales and provide information about dynamics of multiplicative process of cascading nonlinear structures. High positive values of q emphasize regions of intense energy transfer rate, while negative values of q accentuate low-transfer rate regions. The time series is decomposed in segments of size l and then each segment is associated to a nonlinear (eddy) structure. Thus to each eddy in the turbulence cascade one associates a probability measure. The slopes of the log-log plots of the generalized measure

versus scale l for different steps of the cascade curves correspond to the generalized multifractal dimensions, Dq. INA provides the generalized dimensions and the multifractal spectrum for the analyzed time series.

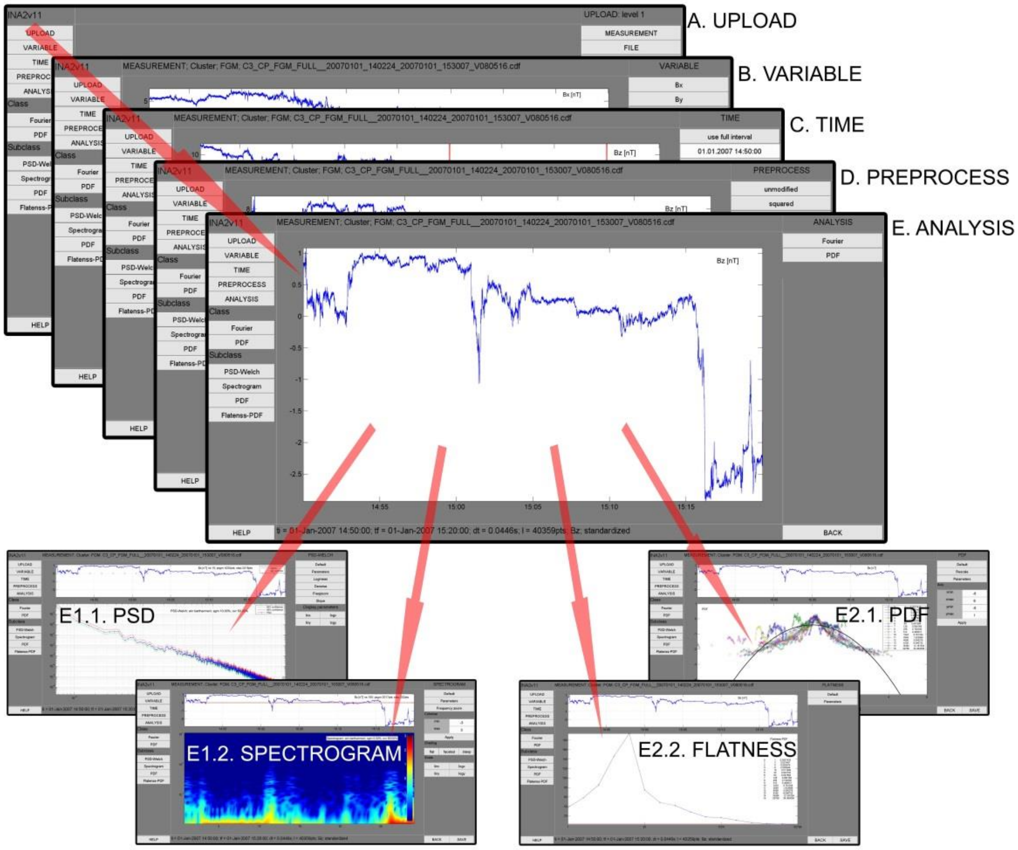

Figure 1: *Diagram representation of the structure of the Integrated Nonlinear Analysis tool (INA). The library is designed as a multi-layer structure allowing to upload various types of time series (in general produced by in-situ measurements), select the type of variable and time interval of interest and apply a full package of nonlinear analysis tools, like e.g. Power Spectral Density, Probability Distribution Functions, wavelet analysis, structure function analysis, multifractal analysis (from Kovacs, Munteanu et al., 2014).*

UPLOAD level 1	UPLOAD level 2	UPLOAD level 3
MEASUREMENT	Cluster Ulysses Venus Express ACE Geomagnetic Indices	FGM, CIS file upload VHM-FGM file upload MAG file upload MFI file upload AE, Dst file upload
FILE	CDF ASCII MAT	CDF import GUI ASCII import GUI MAT import GUI
SYNTHETIC	Sinusoids Other	Sin wave generator GUI Various synthetic signals

Figure 2: *Diagram illustrating the structure of the UPLOAD layer of INA adapted to ingest and analyze data from ESA missions like Cluster, Ulysses and Venus Express, as well as other types of data and synthetic signals designed for testing the structure of real time series (from Kovacs, Munteanu et al., 2014).*