

6. Event Detection Methods

The only reason for time is so that everything doesn't happen at once.

Albert Einstein

Everything happens to everybody sooner or later if there is time enough.

George Bernard Shaw

The phenomenon of self-organized criticality (SOC) can only be identified and validated by event statistics, which requires measurements of relevant observables, such as time scales and size scales of avalanches. The most defining predictions of the ideal SOC model are the scale-free powerlaw distributions of time and size scales, as well as the Poissonian randomness of waiting times. In this Chapter we focus on the methods of measuring time and size scale distributions, mostly based on events detected in astrophysical observations. Apart from particle in-situ measurements in the heliosphere, astrophysical observations generally provide light curves in some wavelength as input for event statistics. In a time series, an event can then be defined by a start time, a peak time, and an end time, which yields a convenient definition for the duration, peak flux, and integrated flux of an event, to be used for statistics of SOC avalanches in terms of durations, peak energies, and total energies. However, the devil is always in the detail. There is no sure-fire method of measuring a unique duration and a peak flux of an event from a light curve. There are numerous diabolical effects such as the separation of events from a fluctuating background, flux threshold biases, confusion from near-simultaneous events, the ambiguity of an event definition in multiple-peak light curves, instrumental irregularities and data gaps, to name just a few. The resulting distributions of measured values are often not robust, unless every event can be uniquely defined, but rather subject to the event definitions and detection algorithms. Thus, the methods used to detect events and to measure their parameters have a profound impact on the results whether we find powerlaw distributions or different functional forms, being our arbiters of SOC phenomena. The best way to investigate various measurement biases is always to conduct a numerical simulation of data and to test a detection algorithm with such well-defined data. The output of the resulting event statistics,

in the form of parameter distributions and correlations, can then be cross-compared with those of the input parameters and measurement biases can be quantified. In this Chapter we will start with the simulation of a time series that contains SOC events according to our standard model described in Section 3.1 and will scrutinize the performance and biases of various detection methods that have been previously used for SOC event statistics.

6.1 Test Data for Event Detection

In order to test various structure detection algorithms it is useful to simulate first a realistic model of a time series $f(t)$. Since we are mostly interested in SOC phenomena, we use our standard analytical model described in Section 3.1, which consists of avalanche events that have an exponential growth during the rise time and a linear decay after the peak. The canonical time profile of such a pulse is shown in Fig. 3.2. The pulses have random waiting times and random rise times.

We simulate a time series containing $n = 1,000$ pulses with an average waiting time of $\Delta t_0 = 5$ and a time resolution of $dt = 0.1$, so the time series contains $n\Delta t_0/dt = 50,000$ data points. The start times t_i are randomly distributed in the time interval $0 < t_i < t_{max} = n\Delta t_0 = 5,000$, produced by a numerical random generator. We verify that the distribution of waiting times Δt_i follows the prescribed random (exponential) distribution $N(\Delta t)$ (Eq. 5.1.1), which is shown in Fig. 6.1 (bottom row, left). The rise times have a mean of $\langle \tau \rangle = t_S = 1$ and are numerically generated with the function (Eq. 7.1.30),

$$\tau_i = -t_S \log(1 - \rho_i), \quad \rho_i = [0, 1], \quad (6.1.1)$$

where ρ_i is a uniformly distributed random number in the interval $[0, 1]$, which is drawn from a numerical random generator. We verify that the distribution of rise times τ_i follow the prescribed random (exponential) distribution $N(\tau)$ (Eq. 3.1.4), which is shown in Fig. 6.1 (third row, left).

Next we simulate the peak energies P_i according to Eq. (3.1.3),

$$P_i = W_0 \left[\exp\left(\frac{\tau_i}{\tau_G}\right) - 1 \right], \quad (6.1.2)$$

using the constants $W_0 = 1$ and $\tau_G = 1$. The resulting distribution of peak energies P_i has a powerlaw slope of $\alpha_P = 1.89$ (Fig. 6.1, third row right). The theoretical distribution of peak energies P_i is expected to be a powerlaw distribution with a slope of $\alpha_P = (1 + \tau_G/t_S) = 2.0$ (Eq. 3.1.28). This is compatible with the simulation, since we found that different random number representations typically change the powerlaw slope of the distributions in the order of about $\pm 10\%$.

Next we can simulate the decay times D_i of the pulses, which depend linearly on the peak energy P_i (Eq. 3.1.13),

$$D_i = t_D \left(\frac{P_i}{W_0} \right), \quad (6.1.3)$$

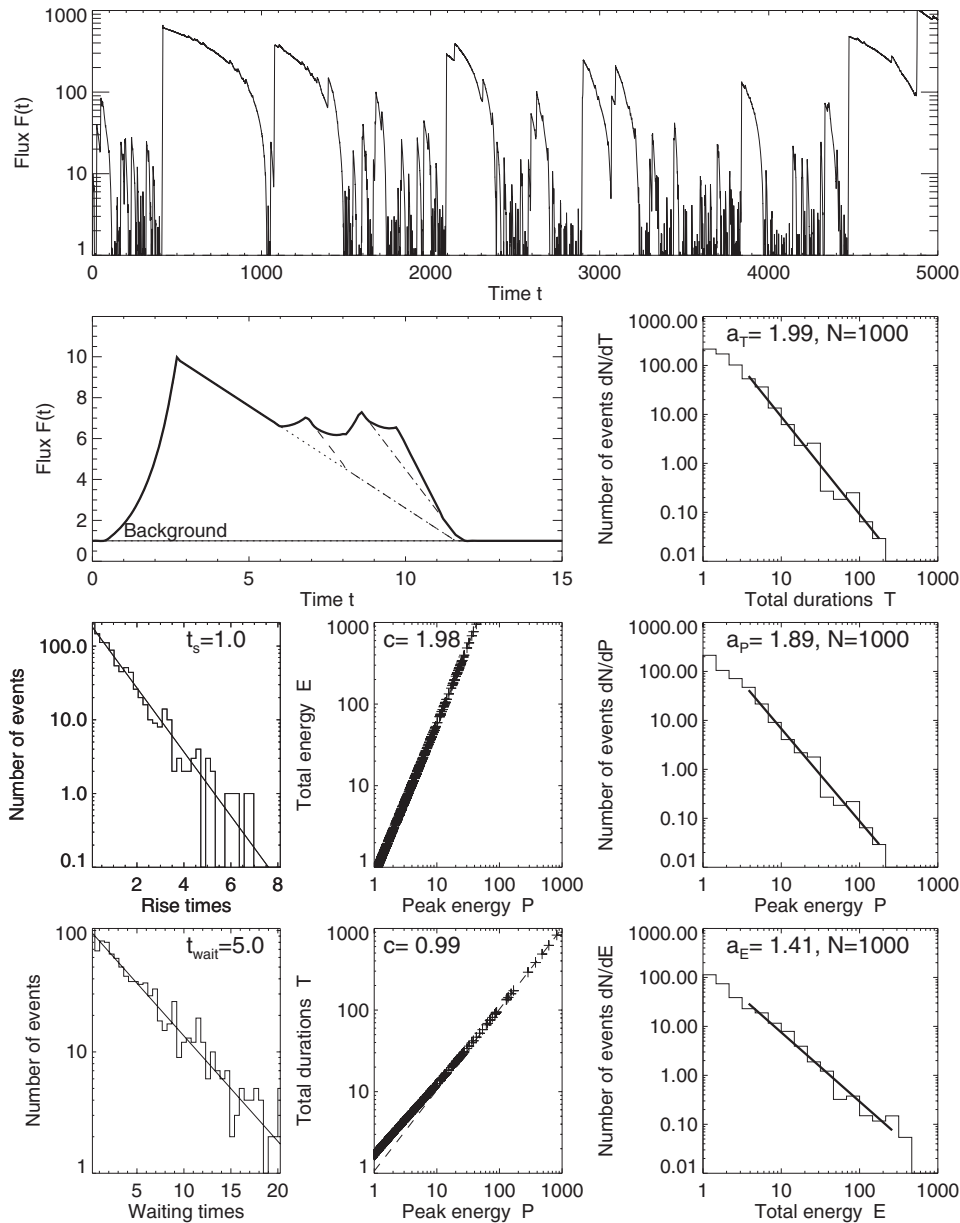


Fig. 6.1 Simulation of a time series (top panel) consisting of pulses with an exponential rise and a linear decay as defined in our analytical exponential-growth SOC model (Section 3.1), with the number of events $n = 1,000$, average waiting time $\Delta t_0 = 5$, time resolution $dt = 0.1$, growth time $\tau_G = 1$, average saturation time $t_S = 1$, and decay time $\tau_D = 1$. The different panels show the first pulses enlarged (second row left), the distributions or rise times and waiting times (left column), correlations between E , T , and P (middle column), and the frequency distributions $N(T)$, $N(P)$, and $N(E)$ with fitted powerlaws (right column).

using the constant $t_D = 1$. Adding the rise time τ_i and the decay time D_i together, we obtain the total pulse duration T_i (Eq. 3.1.14),

$$T_i = \tau_i + D_i . \quad (6.1.4)$$

The resulting distribution of total durations T_i has a powerlaw slope of $\alpha_T = 1.99$ (Fig. 6.1, second row right). This value is compatible with the theoretical expectation $\alpha_T = \alpha_P = (1 + \tau_G/t_S) = 2.0$ (Eq. 3.1.28).

Now we can calculate the total energy E_i (with the background level W_0 subtracted), according to Eq. (3.1.20),

$$E_i = P_i t_G - W_0 \tau_i + \frac{1}{2} P_i D_i . \quad (6.1.5)$$

The resulting distribution of total energies E_i has a powerlaw slope of $\alpha_E = 1.41$ (Fig. 6.1, bottom row right). Theoretically, we expect the relation $\alpha_E = (\alpha_P + 1)/2 = (2 + 1)/2 = 1.50$, which indeed is compatible. For our particular random representation shown in Fig. 6.1 we actually expect $\alpha_E = (\alpha_P + 1)/2 = (1.89 + 1)/2 = 1.44$, which is even closer to the measured value.

We can also plot the correlations between E and P and find a powerlaw fit of $E \propto P^{1.98}$ (Fig. 6.1, third row middle), which agrees with the theoretical limit of $E \propto P^2$ for large values (Eq. 3.1.28). Plotting the correlation between T and P we obtain $T \propto P^{0.99}$ (Fig. 6.1, bottom row middle), which agrees with the theoretical limit of $T \propto P^1$ for large values (Eq. 3.1.28).

Fig. 6.1 shows such a simulated flux time profile $f(t)$, with the full time series shown (Fig. 6.1, top), as well as an enlargement of the first pulses (Fig. 6.1, second row left). Thus we have a simulation of frequency distributions with well-defined values which can be cross-compared with those obtained from different pulse detection algorithms, as we will carry out in the following sections.

When we refer to the peak energy P and total energy E we follow the nomenclature of the SOC terminology, but actually mean the peak flux P and total (time-integrated) flux or fluence E when we detect events from the photon flux of an astrophysical time series. Strictly speaking, the energy rate is then defined in terms of the radiated energy $dE/dt = (n_{ph}/dt)hv$, where (n_{ph}/dt) is the number of photons per seconds. We will discuss other definitions of energies for SOC avalanches in Chapter 9 on physical SOC models.

6.2 Threshold-Based Event Detection

The probably most common and conceptually simplest method of defining events and measuring their duration in astrophysical time series is based on flux thresholds. The essential assumption is that background fluctuations have fluxes below the threshold, while events of interest exceed the flux threshold. This method of event detection is straightforward and pretty safe for large events and high thresholds, but systematically degrades when we detect weaker events, as desirable for SOC studies spanning over a large logarithmic range of time or size scales. In order to obtain some insight how methods of threshold-based event

detection affect the statistical distribution of SOC parameters, we apply this method to the test data shown in Fig. 6.1.

We detect an event simply by identifying the start and end times when the flux profile $f(t)$ crosses the threshold F_{th} at those times according to,

$$\begin{aligned} f(t) &< F_{th} \text{ for } t < t_{start} \\ f(t) &\geq F_{th} \text{ for } t_{start} < t < t_{end} \\ f(t) &< F_{th} \text{ for } t > t_{end} \end{aligned} \quad (6.2.1)$$

The total time duration T is then defined by the time difference, the peak energy P by the maximum flux during this time interval, and the total energy E by the integral of the flux above the threshold,

$$\begin{aligned} T_i &= t_{end} - t_{start} \\ P_i &= \max[f(t_{start}), \dots, f(t_{end})] - W_0 \\ E_i &= \sum_{t=t_{start}}^{t_{end}} [f(t_i) - W_0] dt \end{aligned} \quad (6.2.2)$$

The frequency distributions and parameter correlations obtained with this detection method is shown in Fig. 6.2, which can be compared with the input parameters of the simulated time series shown in Fig. 6.1. The biggest difference is that we detect only 172 pulses out of the 1,000 simulated pulses, where we lose progressively more pulses towards weaker fluxes, because they either are below the threshold, or occur near-simultaneously during larger pulses. This progressive insensitivity towards weaker events leads to a peak flux-dependent under-abundance of shorter pulse durations, and thus to an underestimate of the powerlaw slope, i.e., $\alpha_T = 1.48$ for a threshold of $F_{th} = 1.5W_0$ (Fig. 6.2, second row right), compared with $\alpha_T = 1.99$ of the input data. Similarly, also the powerlaw slope of the peak energy distribution is underestimated, i.e., $\alpha_P = 1.50$ (Fig. 6.2, third row right), compared with $\alpha = 1.89$ of the input data. Also the powerlaw slope of the total energy distribution is underestimated, i.e., $\alpha_P = 1.28$ (Fig. 6.2, bottom row right), compared with $\alpha = 1.41$ of the input data. The correlations between the parameters are less severely affected, because the threshold effects cancel out to some degree, so we find $E \propto P^{1.88}$ (instead of $\propto P^2$) and $T \propto P^{0.93}$ (instead of $\propto P^1$) (Fig. 6.2, left).

This example clearly demonstrates that a flux threshold leads to a significant underestimate of the powerlaw slopes, even for noise-free data (Fig. 6.2). The bias mostly occurs for pulse durations that are shorter than the average waiting time. Only for time series containing well-separated events this bias vanishes in the limit of infinitely small thresholds.

Light curves from astrophysical objects often have a low signal-to-noise ratio, in which case the rate of arriving photons at our detector exhibits fluctuations of comparable magnitude as the weakest SOC events we try to detect. The observed photon flux $F_{obs}(t)$ follows Poisson statistics,

$$f_{obs}(t_i) \approx f(t_i) + \rho(t_i) \sqrt{f(t_i)}, \quad (6.2.3)$$

where $\rho(t_i)$ is a random number with a mean of $\mu = \langle \rho(t_i) \rangle = 0$ and a standard deviation of $\sigma = 1$ in the Gaussian approximation (Section 4.1). The addition of this photon noise to the simulated time profile shows clearly a noisier background, while the signal-to-noise ratio is better at the peak of large pulses (Fig. 6.3, top). If we detect temporal structures with a threshold of $F_{th} = 4.0W_0$, we detect a total of $N = 1,015$ structures (Fig. 6.3), compared

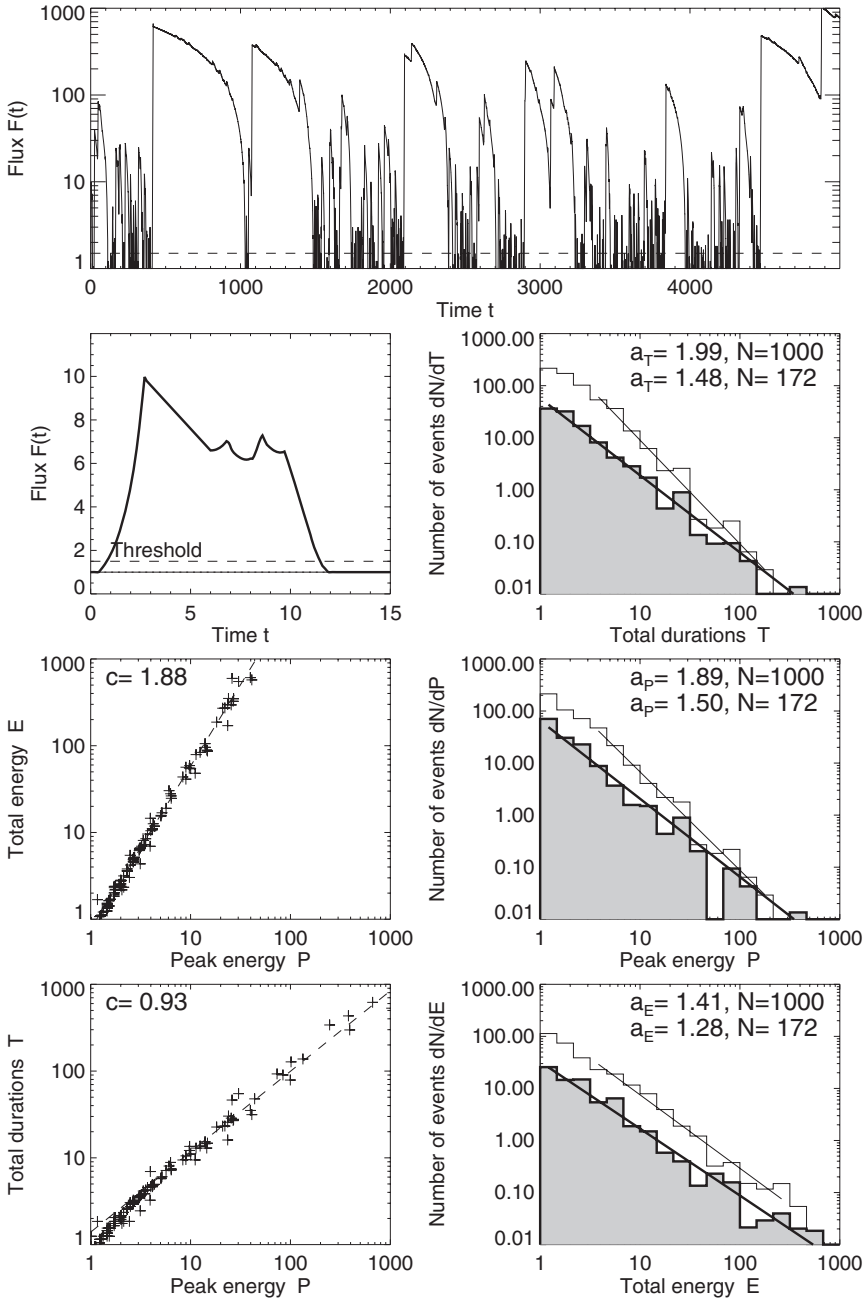


Fig. 6.2 The same time series as shown in Fig. 6.1 (top panel), with frequency distributions of event parameters (right) and correlations between event parameters (left), for events detected above a flux threshold of $F_{th} \geq 1.5W_0$. The histograms of detected structures are shown in gray, while the histograms of the input data (Fig. 6.1) are shown in white.

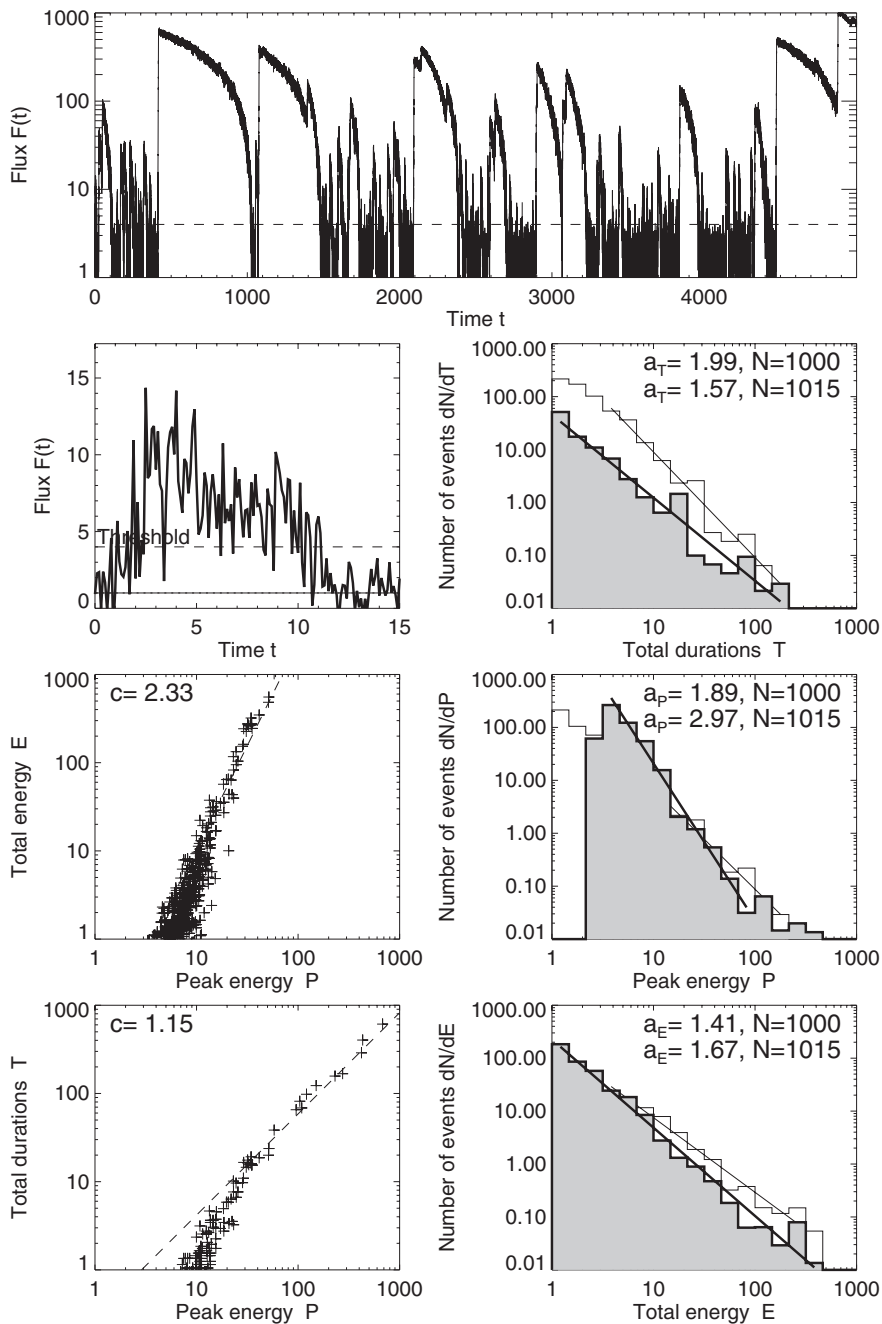


Fig. 6.3 The same time series as shown in Fig. 6.1 but with added Poissonian photon noise (top panel). The frequency distributions of event parameters (right) and correlations between the event parameters are shown (left) for events detected above a flux threshold of $F_{th} \geq 4.0W_0$.

with $N = 172$ in the noise-free data (Fig. 6.2), out of the simulated $N = 1,000$ pulses. The threshold of $F_{th} = 4.0W_0$ corresponds to a 3-sigma significance level (for a Gaussian distribution), which has a confidence level of 99.73%, so we expect about 0.27% fluctuations beyond the threshold level (i.e., positive and negative), or half of it (0.135%) above the positive threshold, which amounts to excess fluctuations in 68 time bins for our time series of $n = 50,000$ data points, which explains the observed excess of 15 events. If we compare the observed frequency distribution of time scales between the input parameters (Fig. 6.1) and the noisy data (Fig. 6.3), we can see that we detect an over abundance of events with small peak energies, leading to a much steeper slope for peak fluxes ($\alpha_P = 2.97$ instead of $\alpha_P = 1.89$), a slightly steeper slope for energies ($\alpha_E = 1.67$ instead of $\alpha_E = 1.41$), and a flatter slope for durations ($\alpha_T = 1.57$ instead of $\alpha_T = 1.99$). Consequently, event detection with noisy data lead to substantially modified frequency distributions, hence, we have to apply suitable procedures to suppress the photon noise. This is more of a problem for photon-starved astrophysical time series, but much less so for the generally photon-rich solar data.

An efficient and robust technique to get rid of photon noise is the smoothing of a time series with a boxcar function, which is defined as the replacement of the flux value $f_i^{sm} = f^{sm}(t_i)$ at each time point with the average within a ‘‘boxcar’’ centered around the time point t_i ,

$$f_i^{sm} = f^{sm}(t_i) = \sum_{j=i-n_{sm}/2}^{i+n_{sm}/2} f(t_j) . \quad (6.2.4)$$

We demonstrate this technique by applying a boxcar smoothing with a length of $n_{sm} = 2t_S/dt = 20$ datapoints to the noisy time series of Fig. 6.3, which is shown in Fig. 6.4. We notice that the data noise is mostly gone; an individual pulse shape is again recognizable with a well-defined rise and decay time (Fig. 6.4, second row left), and the resulting frequency distributions have similar powerlaw slope as the noise-free data (Fig. 6.2), i.e., for durations $\alpha_T = 1.42$ versus $\alpha_T = 1.48$, for peak energies $\alpha_P = 1.43$ versus $\alpha_T = 1.50$, for energies $\alpha_E = 1.23$ versus $\alpha_T = 1.28$. What is even more important, we detect a similar number ($N = 211$) of structures in the smoothed data for a threshold of $F_{th} = 1.5$ as in the noise-free data with the same threshold ($N = 172$), so we get rid of most false structures produced by the photon noise.

The consequences of time-overlapping pulses, photon noise, thresholding, and photon noise on the measurement of time scale distributions, as we exemplified here, apply specifically to pulses that have a powerlaw distribution in their amplitude and duration, as expected for the classical SOC model. Statistics of pulses with other amplitude and duration distributions may exhibit a different behavior. The analysis of astrophysical time series consisting of pulses with random (exponential) distributions in amplitude and durations has been studied in Scargle (1981), with a model called the *moving average model* therein, in contrast to the *autoregressive model* (Scargle 1981), where pulses are clustered in time and have a memory over their recent past. The effects of the event definition and threshold (see Fig. 5.4, middle column) on the measured time scale distribution was also quantitatively studied in a particular time series constructed from an MHD turbulence model (Buchlin et al. 2005), and the main drawbacks were found to be similar as demonstrated here: (i) the loss of small events that produces a cutoff for short time scales, (ii)

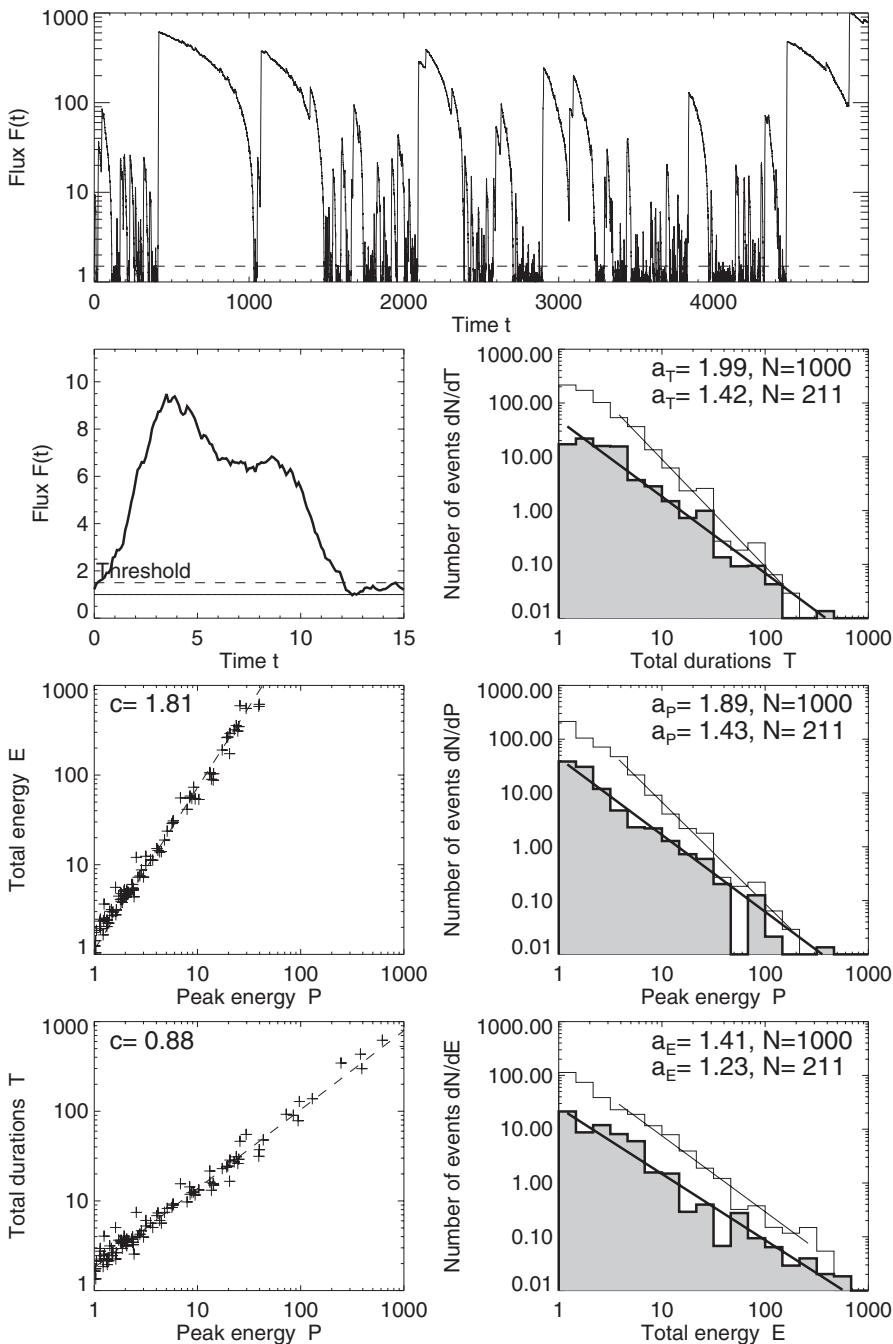


Fig. 6.4 The same simulation of a time series (top panel), correlations (left), and frequency distributions (right) as shown in Fig. 6.3, but a smoothing with a boxcar length of $n_{sm} = 2t_S/dt = 20$ datapoints has been applied.

the inability to separate closely-spaced events, and (iii) the adjustment of thresholds in nonstationary time series (with varying mean rates), which requires Bayesian statistics. Multi-level detection of pulses using the flux levels in the preceding and following time intervals has been applied in the automated detection of gamma ray bursts (Quilligan et al. 2002).

6.3 Highpass-Filtered Event Detection

In order to overcome the non-detection of weaker pulses that occur simultaneously during longer pulses, a fundamental difficulty with threshold-based detection methods (Section 6.2), it is sometimes useful to use a highpass filter, which filters out the slowly-varying time components, so that small pulses on top of a longer pulses can be detected. We demonstrate this method in Fig. 6.5, where we use the same smoothed time series as shown in Fig. 6.4, but subtract a moving-average time profile that is smoothed with a 20 times longer boxcar, $n_{sm2} = n_{sm1} \times 20$. The method of subtracting a smoothed time profile from the original data is also called *unsharp masking*, and is defined as,

$$f_i^{HP} = f^{HP}(t = t_i) = f_i - \sum_{j=i-n_{sm2}/2}^{i+n_{sm2}/2} f(t_j). \quad (6.3.1)$$

Since we already smoothed the original data with a smoothing boxcar n_{sm1} , which represents a lowpass filter, we have actually a *bypass filter*,

$$f_i^{BF} = f^{BF}(t = t_i) = \sum_{j=i-n_{sm1}/2}^{i+n_{sm1}/2} f(t_j) - \sum_{j=i-n_{sm2}/2}^{i+n_{sm2}/2} f(t_j), \quad (6.3.2)$$

where the lowpass filter constant has to be longer than the highpass filter constant, i.e., $n_{sm2} > n_{sm1}$. A bypass filter is essentially sensitive to time structures in the time range of approximately $n_{sm1} \times dt < T < n_{sm2} \times dt$. In our example shown in Fig. 6.5 we expect a lower cutoff of $T_1 = n_{sm1} * dt/2 = 1.1$ and an upper cutoff of $T_2 = n_{sm2} * dt/2 = 20$. As we can see from the number of detected time structures, the highpass filter method yields a more complete sample, i.e., $N = 939$ for a threshold of $F_{th} = 0.01 W_0$, compared with $N = 172$ in the noise-free data (Fig. 6.2). The resulting frequency distributions of the bypass-filtered structures are remarkably robust in retrieving the powerlaw slopes of fluxes ($\alpha_P = 1.74$ versus $\alpha_P = 1.89$) and total energies ($\alpha_E = 1.27$ versus $\alpha_P = 1.41$), and durations ($\alpha_T = 1.87$ versus $\alpha_P = 1.99$), but impose an upper cutoff at the filter time scale ($T \lesssim n_{sm2} dt = 40$) (dashed line in Fig. 6.5, second row right). Thus, except for the filter cutoff of time scales, this method yields robust powerlaw slopes and has a detection efficiency of $\approx 94\%$ for the example analyzed here. However, the parameter correlations are significantly distorted, i.e., $E \propto P^{1.44}$ (instead of P^2), and $T \propto P^{0.84}$ (instead of P^1) (Fig. 6.5, left). Apparently, the distortion of the parameter correlations cancel out to a large extent in the frequency distributions.

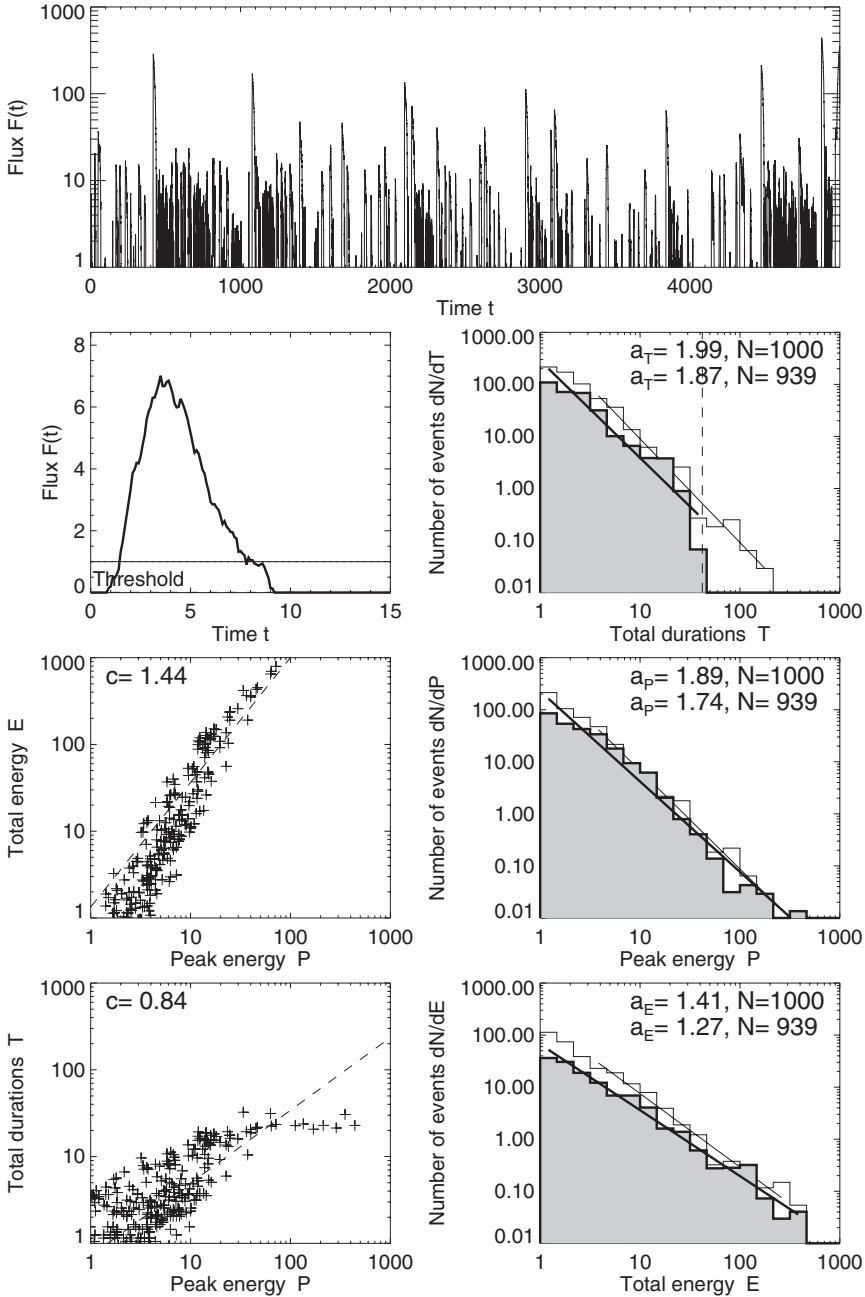


Fig. 6.5 The same simulation of a time series (top panel), correlations (left), and frequency distributions as shown in Fig. 6.3, except for application of a lowpass filter (smoothing) with a boxcar of $n_{sm1} = 2t_S/dt = 20$ datapoints and a highpass filter with a boxcar introduces a cutoff at $T \lesssim 40$ in the time duration histogram (marked with a dashed line in second row right).

Taking advantage of these properties, we could conceive an improved method by combining multiple bypass filters with adjacent but not overlapping time scale ranges, in order to obtain a complete sampling in each filter. Extending this method into the continuum limit we arrive at the so-called *multi-scale* methods, which include the wavelet method (Section 6.7) or principal component analysis (Section 6.8).

6.4 Peak-Based Event Detection

A method of pulse detection that is independent of flux thresholds is the detection of pulse peaks, which of course strongly depends on the pulse shape. In the case of noise-free data (Fig. 6.1), the detection efficiency could approach 100%, because every pulse has a single peak that can be separated temporally, except for near-simultaneous pulses within $\lesssim 2dt$, where dt is the time resolution of the data. In noisy data, however (Fig. 6.3), every pulse has multiple peaks, which even persist in smoothed (Fig. 6.4) and bypass-filtered data (Fig. 6.5). Effects of the event definition by peak times (Fig. 5.4, left) on the frequency distribution of pulse durations are studied in Buchlin et al. (2005) for a particular dataset of MHD turbulence. The problem is mostly to discriminate between peaks of significant pulses and noise peaks. As the enlarged pulse in Figs. 6.3 to 6.5 show, a single pulse can have a multitude of noise peaks. Generally, a decomposition of a multi-peak structure with n local peaks into a denoised structure is ambiguous, because there are $n(n-1)/2$ possible combinations to form subgroups, which can have a significant combined flux about the local background. Thus, a denoising method has first to be applied to the time series, such as smoothing (Fig. 6.4) or a Fourier lowpass filter (Fig. 6.6), before unique peaks can be attributed to individual pulse structures. A peak-based event detection is the less problematic the better the signal-to-noise ratio of the data is.

Once the peaks of significant structures have been identified with a unique peak time, we still have to find the start and end times in order to obtain the total duration. The start time can usually be found by tracking the next significant local minimum (valley) before the peak time. The end time can be estimated the same way by the next local minimum after the peak time, but since our pulses naturally have a longer decay time than rise time, there may be multiple peaks that occur during the decay time. In order to bypass those secondary structures, one would have to require the following minimum to be as low in flux as the flux minimum at the start time.

6.5 Fourier-Filtered Event Detection

A common method of denoising a time series is the application of a Fourier lowpass filter, which we demonstrate in Fig. 6.6. Strictly speaking, this is a special method of the more general category of threshold-based event detection methods (Section 6.2). We use a *Fast Fourier Transform (FFT)* to produce a power spectrum, apply a cutoff to frequencies, $\nu < \nu_{cutoff} = 1/t_{filter}$, or time scales of $T \gtrsim t_{filter}$, and apply the inverse Fourier transform to return a smoothed time profile, so it is a three-step process,

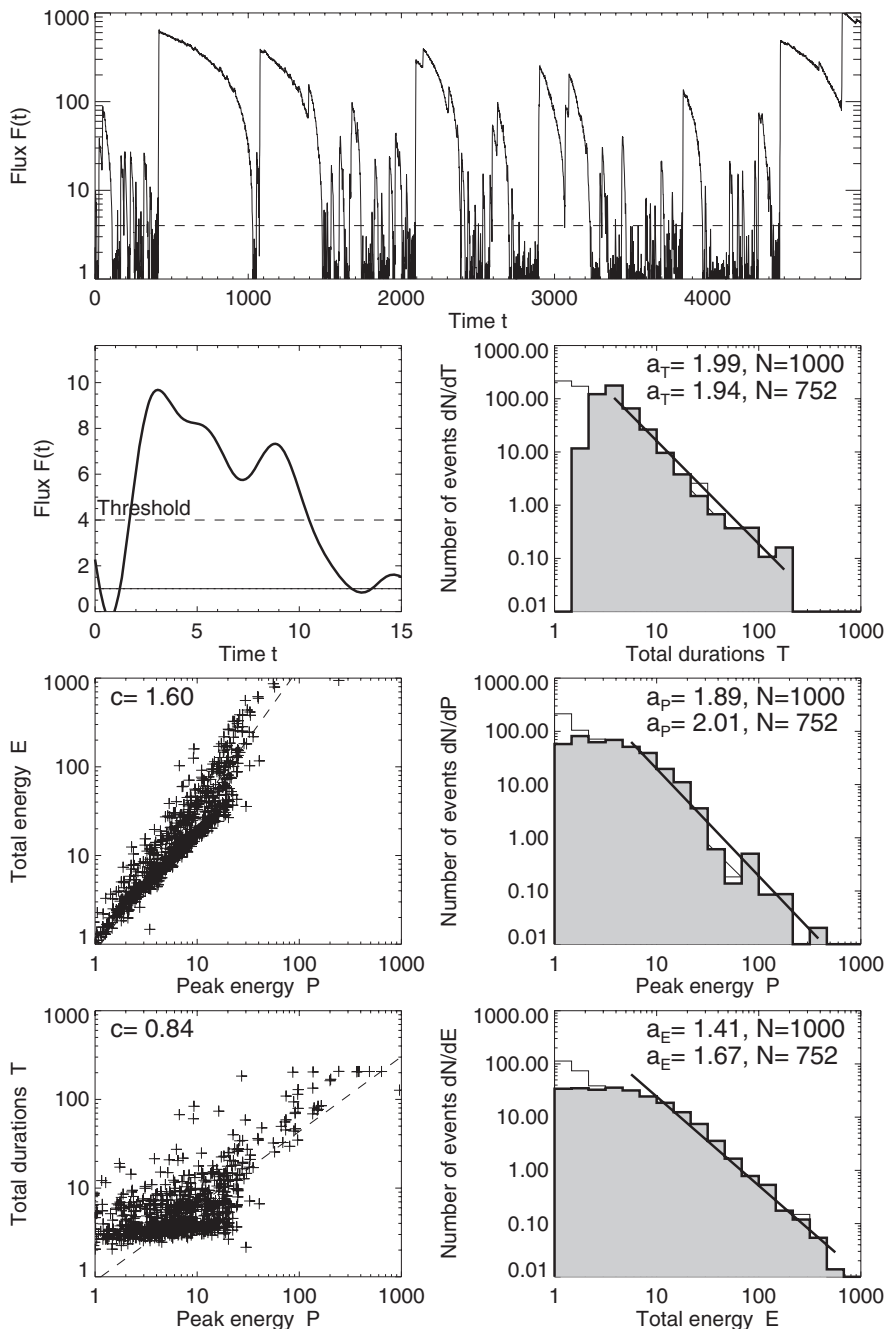


Fig. 6.6 The same simulation of a time series (top panel), correlations (left), and frequency distributions (right) as shown in Fig. 6.1, except for application of a Fourier lowpass filter with a filter passband for time scales $t \leq t_{filter} = 3$ and threshold $F_{th} = 4$.

$$\begin{aligned}
f(t) &\mapsto P(\nu) = FFT[f(t)], \\
P(\nu) &\mapsto P(\nu \geq \nu_{cutoff}), \\
P(\nu \geq \nu_{cutoff}) &\mapsto f_{lowpass}(t) = FFT^{-1}[P(\nu \geq \nu_{cutoff})].
\end{aligned} \tag{6.5.1}$$

Once we obtain the Fourier lowpass-filtered time profile, either a threshold-based (Section 6.2) or a peak-based (Section 6.4) method can be applied to produce statistics of peak fluxes and time durations.

In Fig. 6.6 we demonstrate a peak-based detection method applied to a denoised time profile using a Fourier filter with a cutoff of $t_{filter} = 3$. In addition we require a threshold of $F_{th} = 4$. The detection of a structure requires a local peak above this threshold. The start time is found at the next local minimum before the peak, while the end time is found at the next local minimum after the maximum that has a lower flux than at the start time. With this detection scheme we detect 752 (75%) structures in the example shown in Fig. 6.6. As the enlarged time profile in Fig. 6.6 shows, there are two local peaks during the first time segment of $t = [0, 15]$, which can be compared with the input of four structures that appear clustered (Fig. 6.1). Thus, the Fourier filter has some capability to discriminate near-simultaneous structures, but the number of discriminated structures depends on the filter cutoff. The resulting frequency distributions are relatively robust for this cutoff filter and flux threshold, i.e., for the duration $\alpha_T = 1.94$ versus $\alpha_T = 1.99$, for the peak energy $\alpha_P = 2.01$ versus $\alpha_P = 1.89$, and the total energy $\alpha_E = 1.67$ versus $\alpha_T = 1.41$ (Fig. 6.6). Thus the application of a Fourier lowpass filter is a quite robust technique in retrieving the correct slope of the frequency distributions, has a high detection efficiency ($\approx 75\%$), but has the disadvantage of missing time structures longer than the time scale cutoff ($T \gtrsim t_{filter}$). The results, of course, depend very much on the Fourier filter cutoff ν_{filter} and flux threshold F_{th} .

6.6 Time Scale Statistics from Power Spectra

A Fourier spectrum of a time series is primarily used to detect hidden periodicities, one famous example in astrophysics being the periodic signals from pulsars. Alternatively, a power spectrum can also serve to characterize the occurrence of time structures, such as the $1/f$ noise spectrum, which corresponds to a Poissonian (random) distribution of time scales (Section 4.6). In Section 4.8.2 we derived the specific function of power spectra $P(\nu)$ for randomly distributed pulses with a fixed or mean duration $\langle T \rangle$ according to the shot noise model, as well as for a powerlaw distribution $N(T)$ of pulse durations (Section 4.8.4). Therefore, we can invert the distribution of time scales $N(T)$ analytically from the power spectra $P(\nu)$ of a time series for those special cases.

We demonstrate the inversion of the time scale distribution $N(T)$ from the power spectrum $P(\nu)$ with the example simulated in Fig. 6.3. In Fig. 6.7 we show the Fourier spectrum of the noisy time series $f(t)$ simulated in Fig. 6.3, computed with a standard FFT. We fit a powerlaw spectrum and obtain a slope of $p = 1.03$ for the spectrum $P(\nu) \propto \nu^{-p}$. The time series was simulated with a growth time $\tau_G = 1$ and a mean saturation time $t_S = 1$, which yields a peak energy distribution with a powerlaw slope of $\alpha_P = (1 + \tau_G/t_S) = 2$. The distribution of total durations has a powerlaw slope of $\alpha_T = \alpha_P = 2$, while the distribution

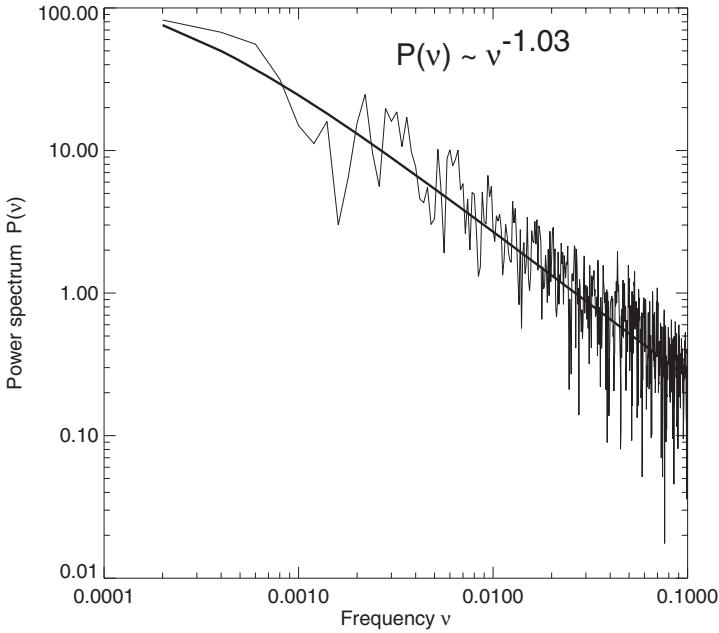


Fig. 6.7 Fourier power spectrum $F(v)$ of time series $f(t)$ shown in Fig. 6.3 (top panel). The spectrum is fitted with a powerlaw function, $P(v) \propto v^{-p}$ with a slope of $p \approx 1.03$.

of energies is $\alpha_E = (\alpha_p + 1)/2 = 1.5$ (Eq. 3.1.28). From the relation Eq. (4.8.23), i.e., $-\alpha_E(1 + \gamma) + \gamma = -\alpha_T$ we can constrain the correlation coefficient γ between energies and time durations, $E \propto T^{(1+\gamma)}$,

$$\gamma = \frac{\alpha_T - \alpha_E}{\alpha_E - 1}, \quad (6.6.1)$$

which yields $\gamma = (2 - 1.5)/(1.5 - 1) = 1$. The corresponding powerlaw index of the power spectrum, $P(v) \propto v^{-p}$, is then $p = (2 - \alpha_E)(1 + \gamma) = (2 - 1.5)(1 + 1) = 1.0$ according to Eq. (4.8.27), which agrees with our measurement of $p \approx 1.03$ obtained in Fig. 6.7. The inversion of the time scale distribution requires the knowledge of both the power spectrum with slope p and the correlation coefficient γ , as we can infer from Eqs. (4.8.22) and (4.8.27),

$$\alpha_T = \alpha_E(1 + \gamma) - \gamma = 2(1 + \gamma) - p - \gamma \quad (6.6.2)$$

so the correlation $E \propto T^{(1+\gamma)}$ between the total energies E and total durations T has to be measured too, at least for the larger pulses.

A power spectrum from a solar light curve observed with the GOES 6 satellite over a total of 32 months during the years 1991–1994 was measured by Ueno et al. (1997), who finds three different spectral components, which can be characterized with the following powerlaw slopes (Fig. 6.8): $p = 1.50 \pm 0.02$ in the frequency range of $v \geq 10^{-3.8}$ Hz ($\lesssim 1.8$ hrs), $p = 0.95 \pm 0.03$ in the frequency range of $10^{-4.7} < 10^{-3.8}$ Hz (1.8–14 hrs), and $p = 0.45 \pm 0.08$ in the frequency range of $v \leq 10^{-4.7}$ Hz ($\gtrsim 14$ hrs). If we translate

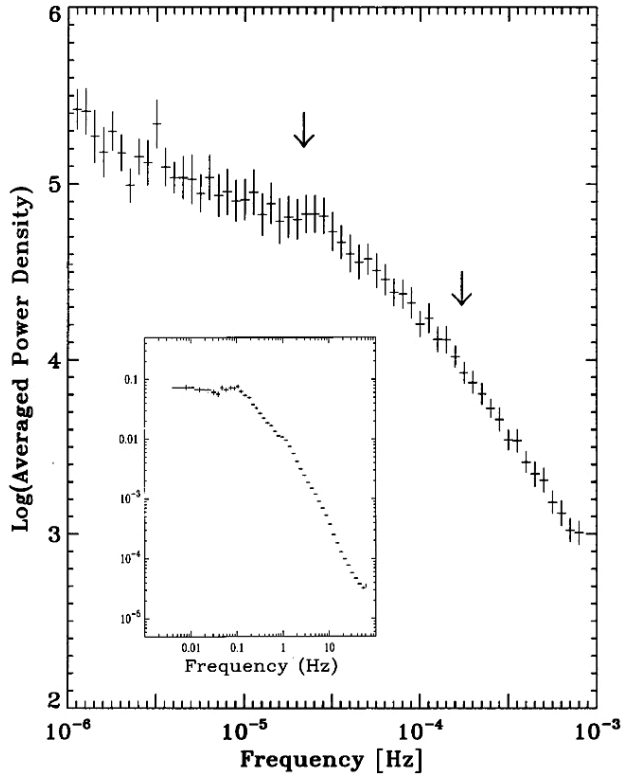


Fig. 6.8 Power spectrum from a solar soft X-ray time profile observed by GOES during 1991-1994, showing three segments with different powerlaw slopes of $p \approx 0.45, 0.95,$ and 1.50 (separated by arrows). The insert shows a power spectrum from Cygnus X-1 obtained by Negoro (1992) which has a similar spectrum (Ueno et al. 1997; reproduced by permission of the AAS).

these powerlaw slopes of the power spectrum p into the powerlaw slopes α_T of a time scale distribution, using $\gamma = 1$ as above, we would obtain values of $\alpha_T = 1.55, 2.05,$ and 2.55 in order of increasing time scales, which appears to be similar to those found from numerical SOC simulations of cellular automata by Lu and Hamilton (1991), where a mean slope of $p = 2.17$ was found, being somewhat flatter at smaller time scales and somewhat steeper at longer time scales (Eq. 2.6.15). Interestingly, this three-part power spectrum for the Sun resembles also a similar power spectrum observed from the black hole candidate Cygnus X-1 (Negoro et al. 1992; Ueno et al. 1997); see insert in Fig. 6.8.

Other applications of Fourier power spectra analysis to astrophysical time series address the restoration and enhancement of astronomical data (Brault and White 1971), unevenly spaced data (Scargle 1982, 1989), and uncertainties in the Fourier power spectrum due to noise (Hoyng 1976).

6.7 Wavelet-Based Time Scale Statistics

When analyzing non-periodic time structures, to be expected for randomly occurring SOC events, a Fourier decomposition of a time profile is not a natural tool, since the harmonic modes used in the expansions are themselves periodic, while the time profile is non-periodic. A better approach is to use other *spectral methods*, such as the *Windowed Fourier Transform*, a *wavelet-based method*, or a *multi-resolution method*. A windowed Fourier transform chops up a time series $f(t)$ into a sequence of windows and yields a Fourier spectrum $P(\nu, t_i)$ for every window t_i as a function of time. The wavelet transform can be considered as a generalization of the windowed Fourier transform, which also yields a gliding power spectrum as a function of time, but uses a better adapted functional decomposition of pulses in a time series, using a so-called *mother wavelet* function, rather than the harmonic sinusoidal functions used in the windowed Fourier transform. There is extensive literature on wavelet methods in general (e.g., Mallat 1989; Daubechies 1992; Meyer and Ryan 1993; Kaiser 1994, Chan 1995), which we do not review here.

Wavelet methods in the analysis of astrophysical time series have been introduced by Scargle (1993), and applied, e.g., to study geomagnetic time series (Kovacs et al. 2001; Vieira et al. 2003), solar helioseismology (Fröhlich et al. 1997), solar diameter variations (Vigoroux and Delache 1993), solar cycle variability (Watari 1995, 1996a; Polygiannakis et al. 2003), solar irradiance (Willson and Mordvinov 1999), solar chromospheric oscillations (Bocchialini and Baudin 1995), sunspot oscillations (Jess et al. 2007), solar hard X-ray flares (Aschwanden et al. 1998a; McAteer et al. 2007), solar radio bursts (Schwarz et al. 1998), stellar chromospheric oscillations (Frick et al. 1997), quasi-periodic oscillations in accretion disks (Scargle et al. 1993), or gamma-ray bursts (Young et al. 1995), just to name a few that deal with wavelet-based analysis of time series. In addition, wavelet-based analysis has also been applied in the spatial domain, especially in solar imaging data.

For statistics of SOC parameters, in particular for durations T of pulses, we are interested whether the output of standard wavelet algorithms, the time-dependent Fourier power spectral density $P(\nu, t_i)$, consisting of *scalegrams* $S(T)$ for each time interval $[t_i, t_{i+1}]$,

$$S(T) = \langle |P(\nu[T], t)|^2 \rangle, \quad \text{for } t_i < t < t_{i+1} \quad (6.7.1)$$

where $\nu[T] = 1/t$, can be transformed into a distribution $N(T)$ of time scales T . Such a transformation has been developed in Aschwanden et al. (1998a). The procedure is sketched in Fig. 6.9 and examples are simulated in Fig. 6.10. Essentially, a scalegram $S(T)$ can be considered as a convolution of a distribution function $N(T)$ of time scales T with a kernel function $p(T)$. The distribution $N(T)$ of time scales can then be obtained from the inversion of a scalegram $S(T)$ using the kernel function $p(T)$ that corresponds to a particular *mother wavelet* function. The schematic in Fig. 6.9 shows a rectangular distribution $N(T)$ of time scales and how the convolution of a (double-powerlaw) kernel function $p(T)$ produces the scalegram $S(T)$. The numerical simulations in Fig. 6.10 show artificial time profiles $f(t)$, the wavelet scalegrams $S(T)$, and the inverted time scale distributions $N(T)$ (histograms in Fig. 6.10, right), compared with the theoretical distributions $N(T)$ (delta-functions and Gaussians) that have been used as input in the generation of the

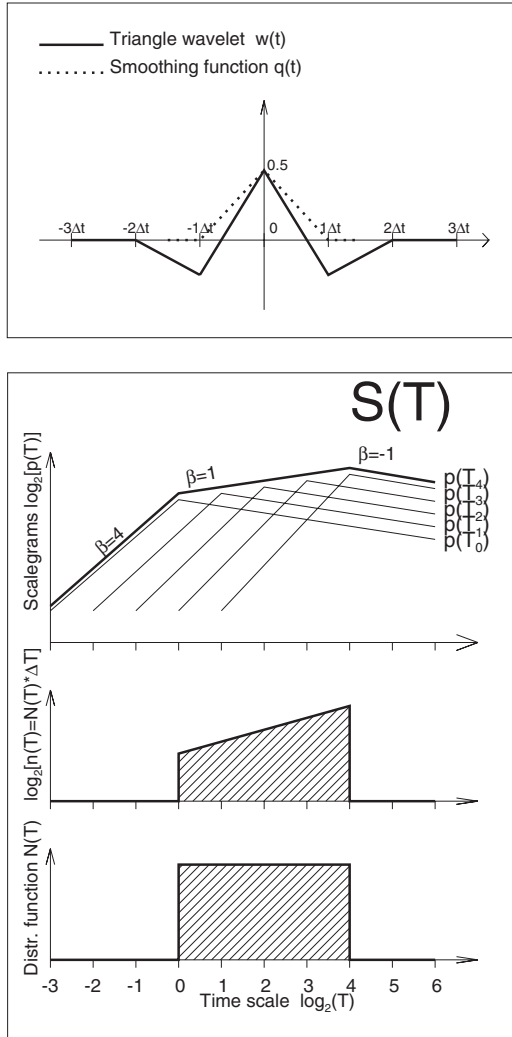


Fig. 6.9 *Top:* Triangle mother wavelet function $w(t)$ (thick line) and smoothing function $q(t)$ (dotted). *Bottom:* Schematic illustration of the convolution of a standard distribution function $N(T)$ (bottom) of time scales with kernel functions $p(T_i)$ that sum up to a scalegram $S(T)$ (Aschwanden et al. 1998a).

time profiles $f(t)$. These examples demonstrate that the wavelet-based inversion method can retrieve the original time scales.

A practical example of a wavelet scalegram of an observed time profile observed in a solar flare is shown in Fig. 6.11, along with the inverted time scale distributions $N(T)$ in four different time intervals. This method has been applied to the time profiles of 46 solar flare events and exponential distributions of time scales were found, in contrast to powerlaw-like distributions expected for SOC models. In the study of McAteer et al. (2007), the

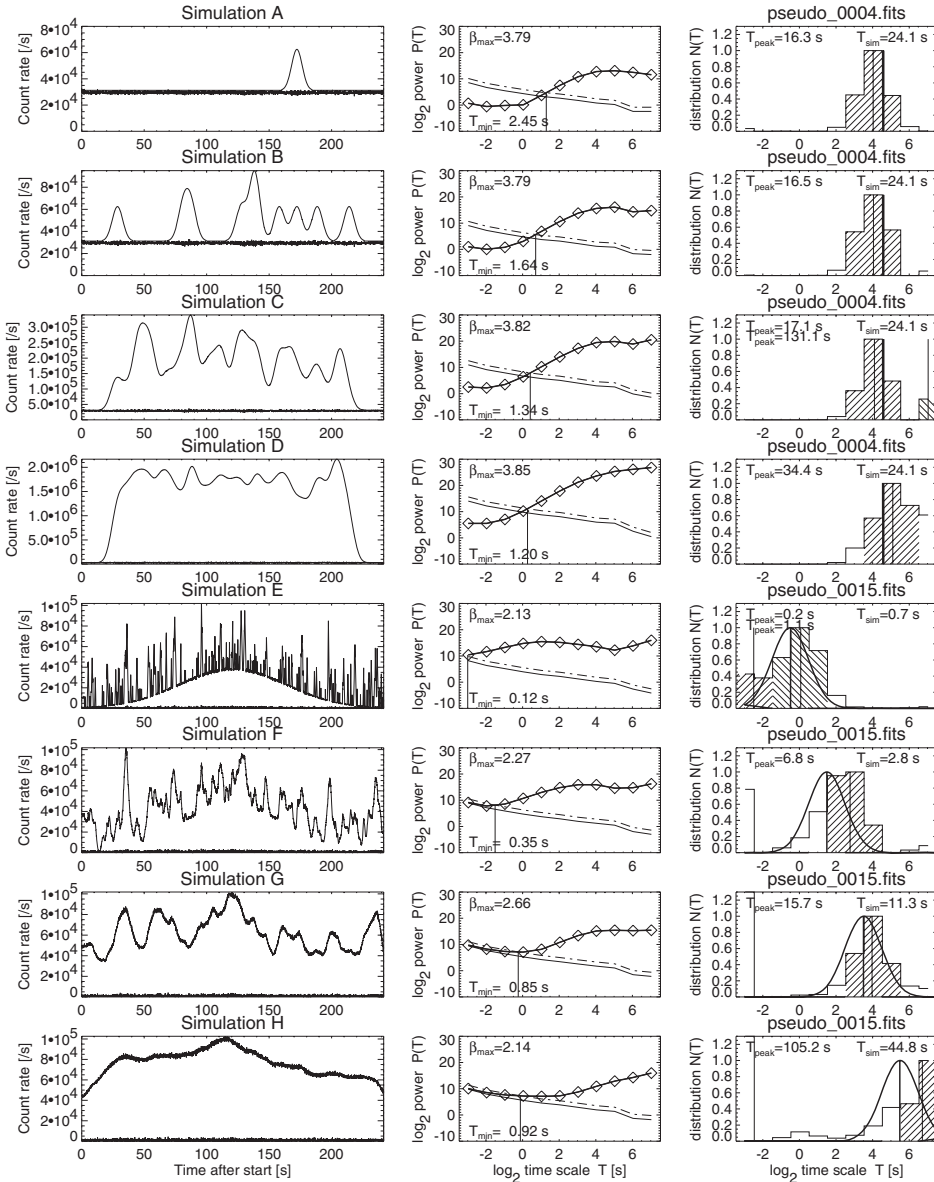


Fig. 6.10 Eight numerical simulations (A–H) of time profiles $f(t)$ (left panels), the computed wavelet scalegrams $S(T)$ (middle panels), and inverted time scale distribution functions $N(T)$ (right panels). The time profiles contain also the noise templates (left panels). The scalegrams (diamonds in middle panels) contain also the noise scalegrams (thin solid line) with the 3σ -limit (dashed line). The slope β_{max} is measured at the steepest part of the scalegrams. The inverted time scale distribution functions (histograms in right panels) are compared with the theoretical distribution functions (thick curve) used in the simulation of $f(t)$, with mean time scales T_{sim} , and are compared with the inverted peak times T_{peak} (weighted over hatched part of histogram) (Aschwanden et al. 1998a).

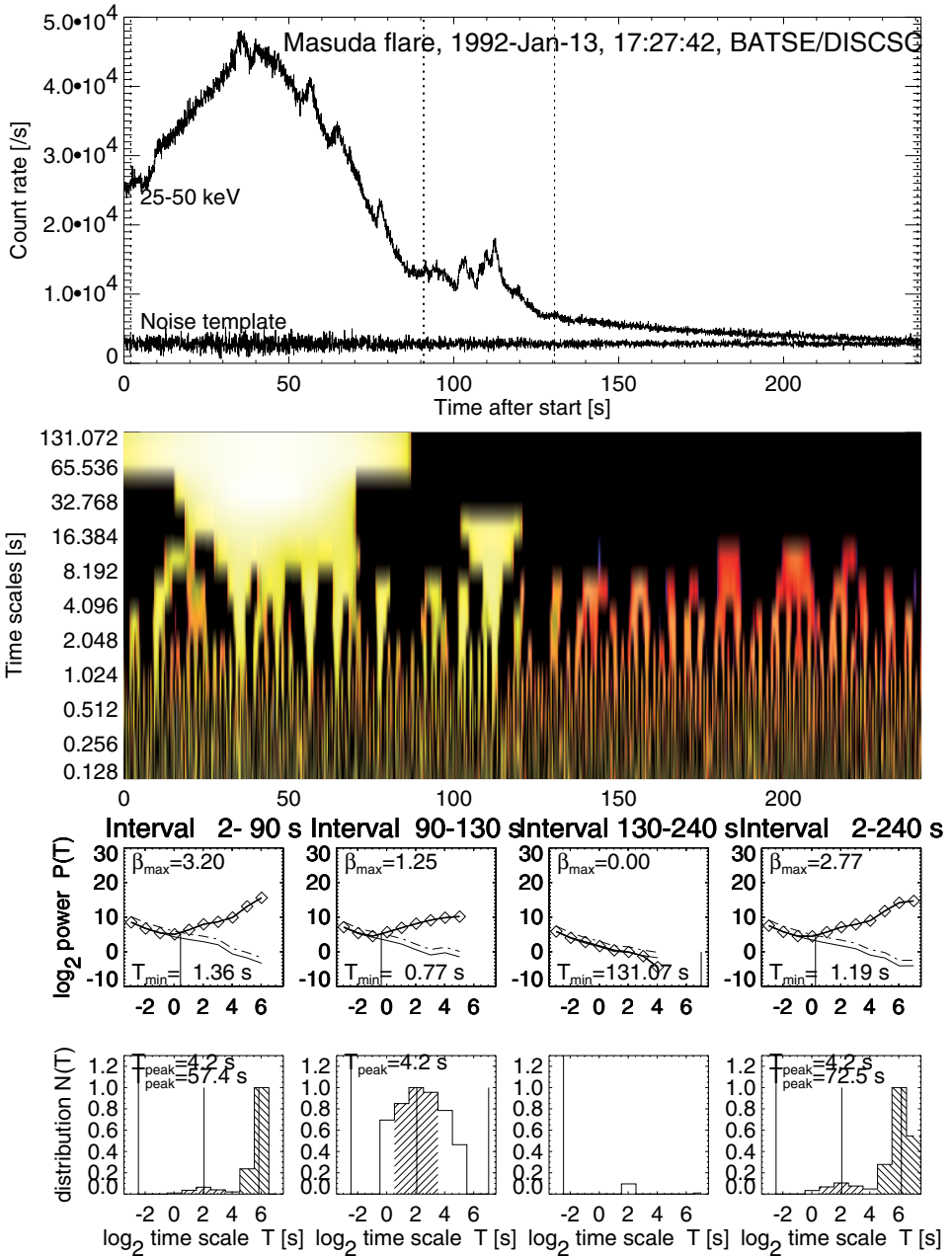


Fig. 6.11 Time profile (top), scalogram $P(T, t)$ (middle panel with grayscale), time-averaged scalegrams $S(T)$ (third row), and inverted time scale distribution functions $N(T)$ (fourth row) for the Masuda flare, 92-Jan-13, 17:27:42 UT, observed with *BATSE/CGRO* (Aschwanden et al. 1998a).

wavelet-based analysis of a solar flare revealed a Hölder exponent that indicates a high degree of memory between subsequent hard X-ray peaks, which is also in contrast to the supposed independent events in a SOC process. Wavelet-based statistics of time scales has not been exploited to the full extent yet, but appears to be a very promising method for obtaining statistics of temporal structures from convolved time series, which contain near-simultaneous events with differing time scales.

6.8 Principal Component Analysis

While a Fourier analysis decomposes a time series into harmonic sinusoidal components, and a wavelet analysis decomposes into stretched and shifted mother wavelet functions (e.g., a Mexican hat function), there is an even better adjusted decomposition method that attempts to find a minimum number of best-fit components, which is called *Principal Component Analysis (PCA)*, *Independent Component Analysis (ICA)*, *Proper Orthogonal Decomposition (POD)*, *Complex Empirical Orthogonal Function (CEOF) analysis*, *Hotelling transform*, or *Karhunen–Loève transform (KLT)*. The mathematical procedure transforms a number of possibly correlated variables into a smaller number of uncorrelated (independent, or orthogonal) variables, called principal components, and involves the calculation of eigenvalues of a data covariance matrix, or a singular value decomposition of a data matrix. The PCA method can be used for automated detection of spatial or temporal features.

In astrophysical time series (or image time series), the PCA method (or a PCA-like extension) has been applied, e.g., to solar EUV data to detect propagating waves (Terradas et al. 2004), to solar cycle synoptic data to characterize the “butterfly diagram” over ≈ 25 –50 years (Lawrence et al. 2005; Vecchio et al. 2005a), to solar magnetogram data to identify low-frequency oscillations in photospheric motion (Vecchio et al. 2005b), to interplanetary magnetic field polarity data (Cadavid et al. 2008), or to $\approx 17,000$ light curves of variable stars such as RR Lyrae’s, Cepheids, and Mira variables (Deb and Singh 2009).

An example of a PCA analysis is shown in Fig. 6.12 for a time series of solar EUV data (Terradas et al. 2004). A decomposition into oscillatory components with independent periods is attempted, called *empirical mode decomposition (EMD)*, which obtains from the analyzed signal six different components with different periods and amplitudes. The sum of the six decomposed components represent the original data within the data noise. The EMD method produces a decomposition into frequency band-limited components by using information from the signal itself instead of prescribing basis functions with fixed frequency, such as in Fourier or wavelet methods. The decomposition is not unique, but attempts to represent a time series with a minimum number of time scale ranges. Another example is shown in Fig. 6.13 for a time series from a variable star (Deb and Singh 2009). The full quasi-periodic time series was decomposed with both Fourier and PCA decomposition techniques. Fig. 6.13 shows the reconstruction of a fundamental mode (FU) Cepheid light curve using the first 1, 3, 7, and 10 principal components. It was found that 10 principal components contain nearly 90% of the variance in the data.

These examples demonstrate how an arbitrary time profile can be decomposed into a relatively small number of noise-free time profiles $f_T(t)$, each one having a characteristic time scale T within a prescribed bandwidth Δt . From these time profiles $f_T(t)$,

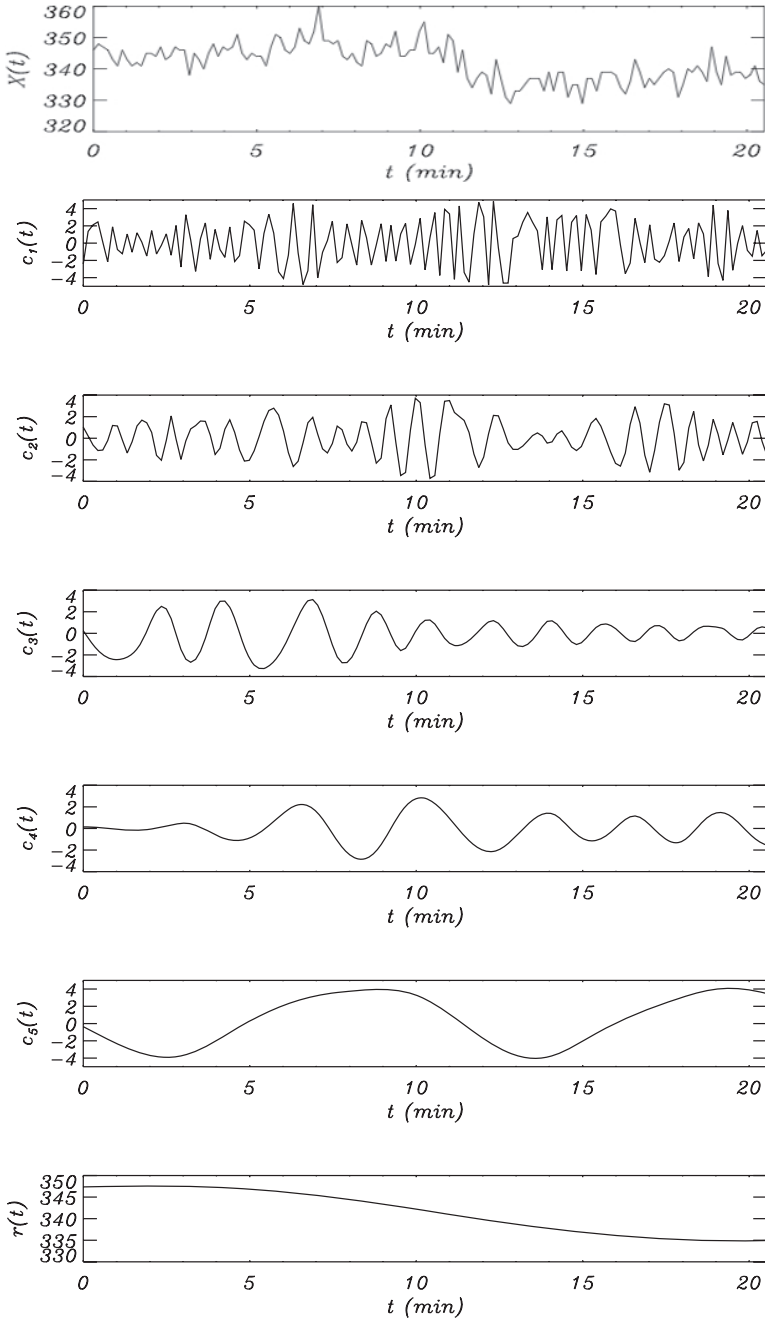


Fig. 6.12 A time series of EUV flux observed in a loop in the solar corona (top), decomposed into six principal components with increasing time scales, according to an empirical mode decomposition (EMD) method. Each component is not strictly periodic (as a Fourier mode), but has its own characteristic time scale within a small tolerance range, $T \pm \Delta T$ (Terradas et al. 2004; reproduced by permission of the AAS).

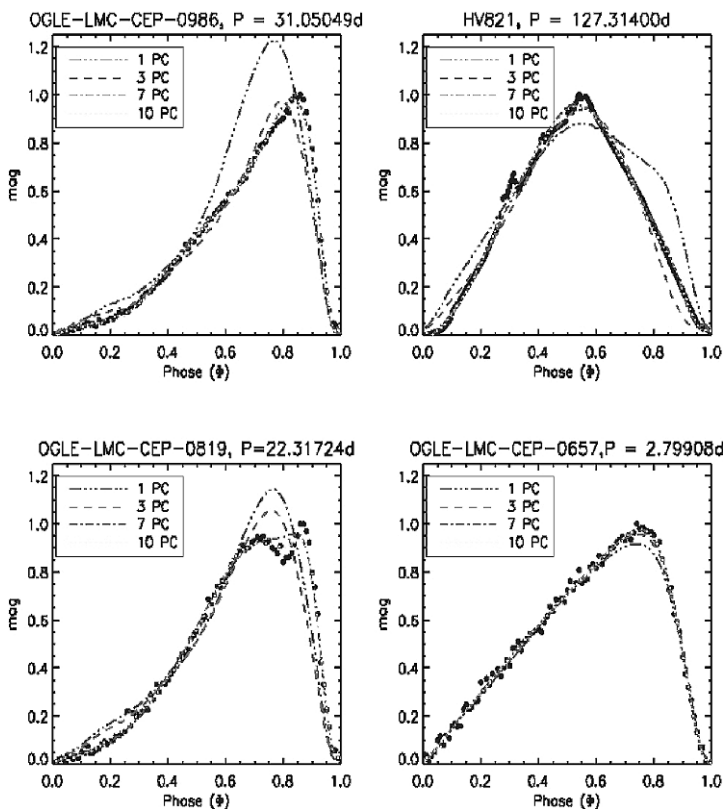


Fig. 6.13 Reconstruction of four FU Cepheid light curves using the first 1, 3, 7, and 10 principal components (Deb and Singh 2009).

the peak amplitudes P_T , and total time-integrated fluxes E_T of pulse structures with a time scale T can be sampled, using either a threshold-based or peak-based event detection method. This would allow us, after proper normalization, to obtain the frequency distributions of time scales $N(T)$, peak energies $N(P)$, and total energies $N(E)$. Therefore, the PCA method appears to be a useful and efficient method for SOC statistics, probably better adapting to unknown pulse shapes than Fourier-based and wavelet-based methods.

6.9 Image-Based Event Detection

In the previous sections we discussed SOC event statistics obtained from one-dimensional (1-D) time series data $f(t)$, as we usually obtain from astrophysical observations, but statistics of SOC events has also been inferred from time sequences of images $f(x, y; t)$, such as from magnetospheric or solar imaging data. Although the automated processing of three-dimensional (3-D) data $f(x, y; t)$ is more complex, the chief advantage for SOC statistics is the spatial separation of near-simultaneous events, which can conveniently be

discriminated in the space domain, while they coincide in the time domain. This is particularly important for SOC statistics because spatial correlations can introduce time clustering and deviation from Poissonian random statistics (e.g., aftershocks of earthquakes, or sympathetic solar flares), in contrast to spatially independent events that are expected to obey a true random behavior of waiting times. Event or feature detection in imaging data became a growing industry and for general introductions into digital image processing we refer to the textbooks of Gonzales and Woods (2008), Jain (1989), Castleman (1996), Jähne (2005), Woods (2006) and Mallat (2008), and more specifically for astrophysical data see Starck et al. (1998) and Starck and Murtagh (2002), or for image processing techniques and feature recognition in solar physics see Aschwanden (2009). Here we will discuss only a few examples that are most relevant for SOC statistics.

A threshold-based detection method of temporal features essentially involves a criterion $f(x, y; t) \geq F_{th}$ in 3-D space, but additionally requires the automated detection of spatially coherent and contiguous features. A common procedure is to perform an image segmentation in an image $f(x, y; t = t_i)$ that detects spatially coherent shapes above a prescribed threshold, with subsequent identification of co-spatial structures in the preceding or following images $f(x, y; t = t_{i-1}, t_{i+1})$ that have at least one pixel above the prescribed threshold in common. An example is given in Fig. 6.14, which shows a result of detected

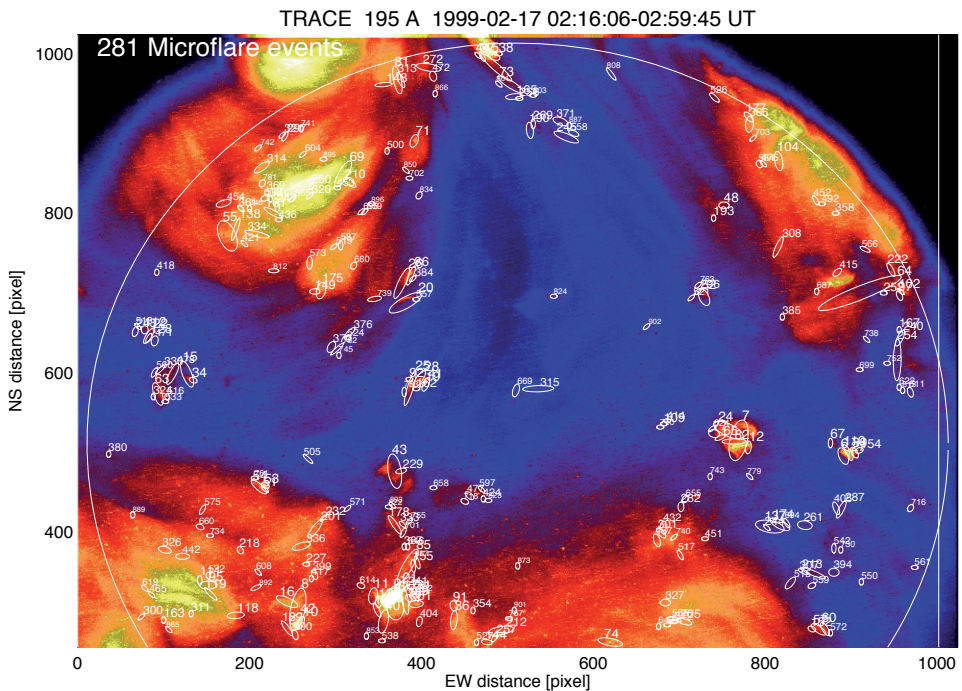


Fig. 6.14 This synthesized TRACE 195 Å image is summed from 22 images recorded during 1999-Feb-17 02:16:06–02:59:45 UT. The circle encompasses the analyzed field-of-view with a diameter of ≈ 8 arcmin. The numbered ellipses mark 281 flare-like events that fulfill the flare definition criterion, out of a total of 901 EUV brightening events. The geometric size and orientation of the ellipses is on scale, encompassing the simultaneously-varying pixels of a flare event (Aschwanden et al. 2000b).

solar nanoflares in a solar EUV data cube (Aschwanden et al. 2000a,b). Since SOC phenomena are dynamic events, the threshold criterion for an event detection is in this case not simply a flux threshold, but rather a variability threshold Δf_{th} , which can be defined in terms of a flux change that exceeds the level of random fluctuations,

$$f(x, y; t_{i+1}) - f(x, y; t_i) \geq \Delta f_{th} = 3\sigma_f, \quad (6.9.1)$$

where σ_f is the standard deviation of the photon Poisson noise in a time bin corresponding to the exposure time of the image. Examples of such variability maps are shown in Fig. 6.15, where it can be seen that those pixels with high fluxes (indicated with flux contours in Fig. 6.15) are not necessarily identical with those of significant flux variabilities from one to the next time frame. To ensure a proper tracking of a coherent event in time, only pixels that exhibit a co-spatial variability in the previous and/or subsequent time frame are considered as part of the same coherent event, or SOC avalanche (marked with diamonds in Fig. 6.15), while other pixels with significant variability occurring in one single time frame only are considered as event-unrelated (instrumental or unresolved) brightness fluctuations. The time evolution of such automatically traced features observed in two different wavelengths (Fig. 6.16) exhibits the typical fast rise and exponential decay of a solar flare event. The multi-wavelength coverage of these events moreover ensures the self-consistent physical evolution of an elementary solar flare process, which consists of a rapid impulsive heating phase with subsequent plasma cooling by thermal conduction and radiative cooling. It exhibits the typical exponential decay, which appears delayed in the wavelength with the cooler temperature. Proper definition of events are extremely important in image-based feature detection methods, because multi-dimensional data are more prone to erroneous event detections of unrelated other variabilities contained in the data than 1-D time series.

The numerical event detection code used for the examples shown in Fig. 6.14–6.16 was especially designed to detect solar microflares and nanoflares, which represent the faintest counterparts of solar flares, and thus are important to extend the dynamic range of frequency distributions of flare energies over nine orders of magnitude. Similar codes were also developed by Krucker and Benz (1998) and Parnell and Jupp (2000), which spurred controversial results on the powerlaw slopes in the nanoflare regime. A number of issues were considered that contribute to the initially discrepant results of powerlaw slopes, such as event definition, selection, and discrimination, sample completeness, observing cadence and exposure times, pattern recognition algorithms, threshold criteria, instrumental noise, wavelength coverage, fractal geometry, but also physical modeling issues of energy, temperature, electron density, line-of-sight integration, and fractal volume (e.g., Aschwanden and Parnell 2002; Benz and Krucker 2002). The issue of the correct powerlaw slope of the frequency distribution of nanoflare energies was further aggravated by the fact that the initially discrepant results scattered on both sides of the critical value (with a slope of $\alpha_E = 2$) that decides whether the energy of nanoflares is more important for coronal heating (if $\alpha_E > 2$). We will come back to this issue when we discuss physical energy models of SOC events in Section 9.3.

A similar task is the automated detection of solar *bright points*, which are small, bipolar magnetic fields in the photosphere and can be detected best in EUV. These events

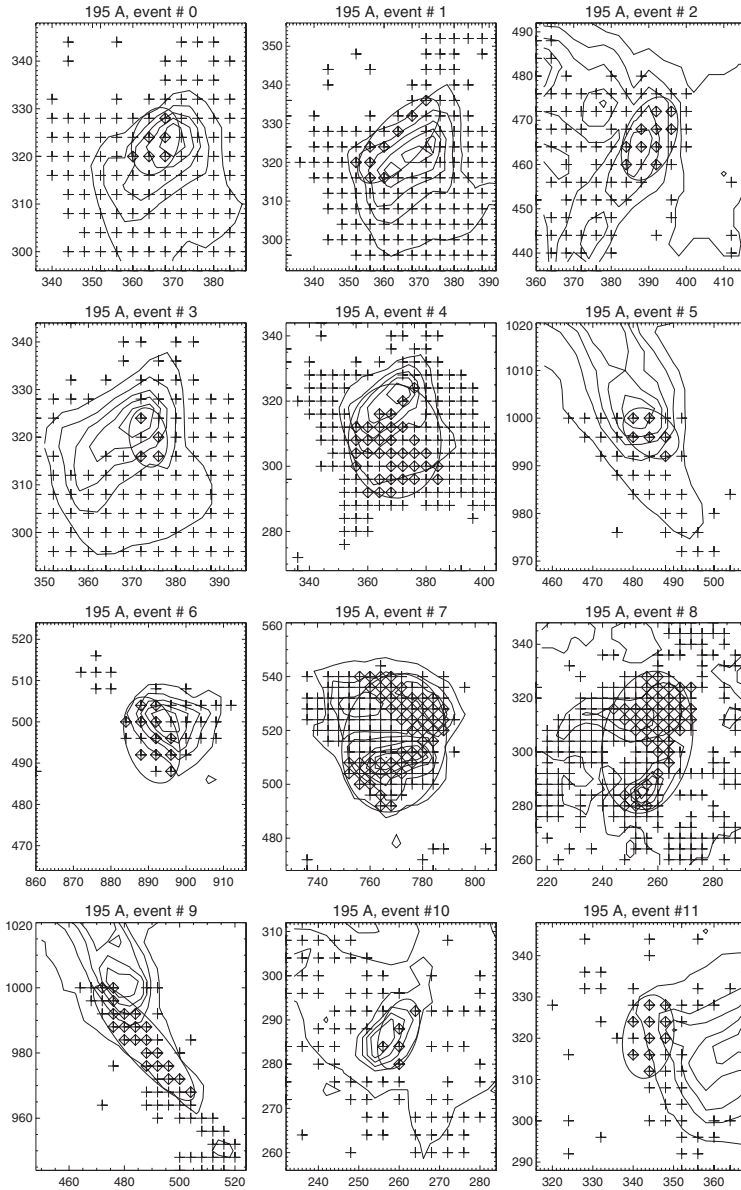


Fig. 6.15 The spatial clustering of the pattern recognition code is illustrated for the 12 largest events on 99-Feb-17, 02:15–03:00 UT. The contours outline local EUV intensity maps around the detected structures. The crosses mark the positions of macropixels with significant variability ($N_{\sigma} > 3$). The spatio-temporal pattern algorithm starts at the pixel with the largest variability, which is located at the center of each field of view, and clusters nearest neighbors if they fulfill the time coincidence criterion ($t_{peak} \pm \Delta t$). These macropixels that fulfill the time coincidence criterion define an event, marked with diamonds, and encircled with an ellipse. Each macropixel that is part of an event, is excluded in subsequent events. Note that events 0,1,3,11 belong to the same active region, where the four near-cospatial zones have peaks at different times and thus make up four different events (Aschwanden et al. 2000a).

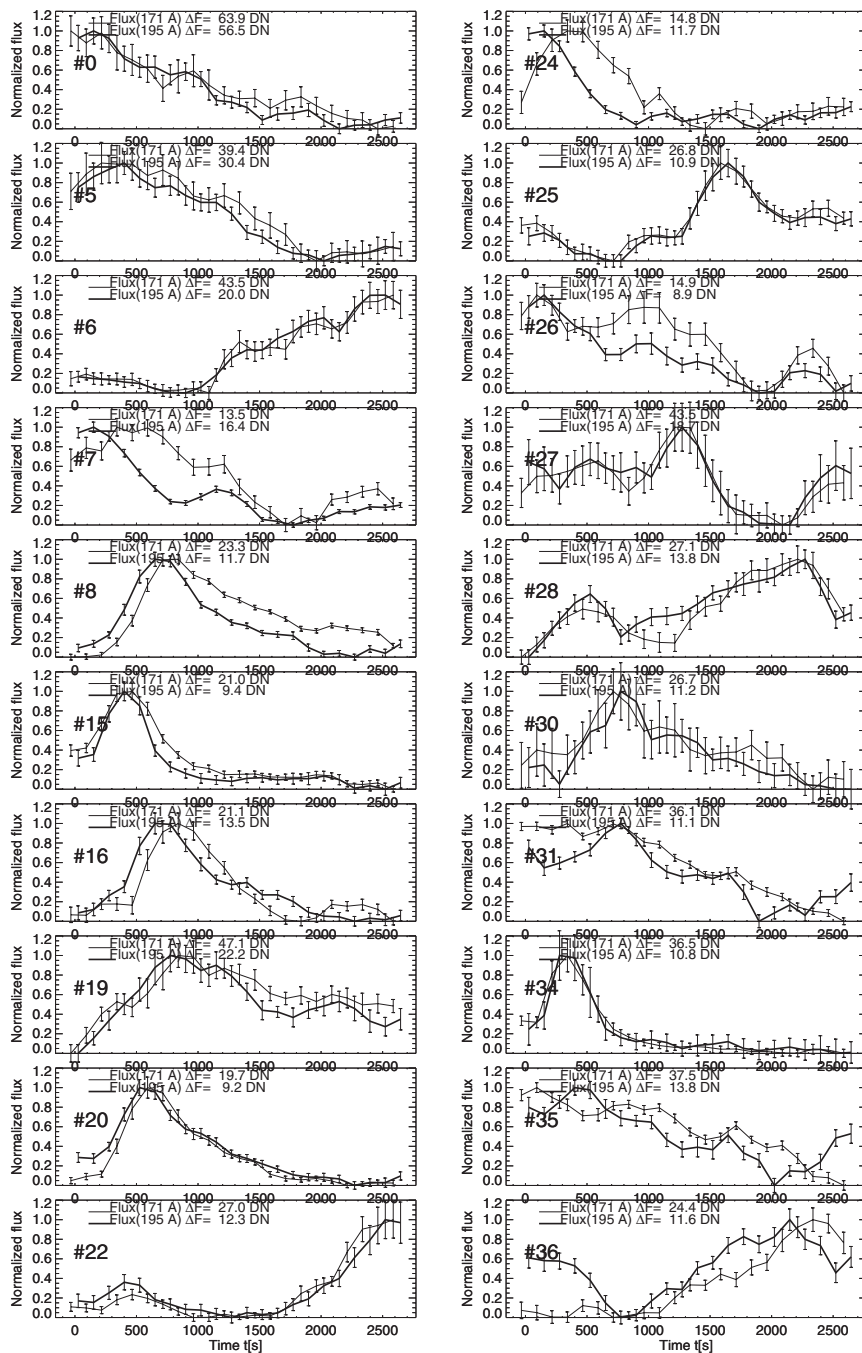


Fig. 6.16 Time profiles of the 171 Å (thin line) and 195 Å (thick line) flux of 20 EUV microflares. Both fluxes are normalized to unity, with the absolute fluxes indicated in each panel. The error bars include all instrumental and photon noise components. Note that the 171 Å flux is highly correlated with the 195 Å flux, but generally delayed, as expected for a plasma cooling process (Aschwanden et al. 2000b).

have been detected with a highpass-filter method with proper noise threshold estimates from a long-time data series over 9 years to the extent of an unheard number of 1.3×10^8 events (McIntosh and Gurman 2005). Such large statistics is extremely useful for SOC statistics, regarding the accurate functional form of the frequency distributions and their time-dependent changes.

Automated detection of *coronal mass ejection (CME)* events in coronagraph images represent a specially challenging task because of their highly transient nature and complex and inhomogeneous spatial morphology. A CME rapidly expands in 3-D space, the observed brightness becomes quickly diluted, and the morphology evolves from an initial fan pattern to a turbulent spherical shape, possibly containing multiple shock fronts with accelerating and decelerating speeds. Thus, typical SOC parameters like a peak flux P , total flux E , and duration T are difficult to define for such dramatically changing morphological structures. Even waiting times Δt of CMEs (Fig. 5.13) are problematic to measure, because multiple CMEs interfere with each other in an observed field-of-view. Frequency distributions of CMEs have only been sampled for their (angular) sizes, which were found to exhibit invariant powerlaw slopes during a solar cycle (Robbrecht et al. 2009). Thus, typical CME observables entail an angular width, an apparent latitude, and apparent velocities, which are not straightforward to translate into SOC parameters. These parameters would correspond to a geometric aspect ratio, location, and velocity of sandpile avalanches. Nevertheless, automated detection algorithms for CME events have been developed by using a threshold-segmentation technique of radial off-limb images (Olmedo et al. 2008), a wavelet-based multi-scale edge detection technique (Young and Gallagher 2008), or a Hough transform with a morphological opening operator (Robbrecht and Berghmans 2004). An example of a CME detection by Young and Gallagher (2008) is shown in Fig. 6.17. The optical brightness of CMEs is usually so weak that they can only be detected in running time-difference or in polarized brightness images. CME-related phenomena are so-called *EIT waves*, which according to one model propagate concentrically to the CME over the solar surface and can be traced by means of flux threshold-based detection of spherically propagating ring patterns (Podladchikova and Berghmans 2005).

Future automated detection algorithms of spatio-temporal patterns are expected to involve more artificial-intelligence algorithms or neural-network-learning techniques, which can adjust to the unknown or unquantified morphological shapes progressively with the increasing number of detected events.

6.10 Summary

Feature and event detection methods represent the input for SOC statistics and thus it is extremely important to simulate and understand their statistical biases on the resulting frequency distributions of SOC events. We simulated a time series that is particularly designed for typical SOC events, characterized by powerlaw distributions of amplitudes and durations, as well as by Poisson statistics of waiting times (Section 6.1). Armed with such test data, we simulate the detection biases for threshold-based event detection (Section 6.2), for both noise-free data and data affected by heavy photon noise. We test how smoothing of a time series, which is one option to suppress the data noise, affects the

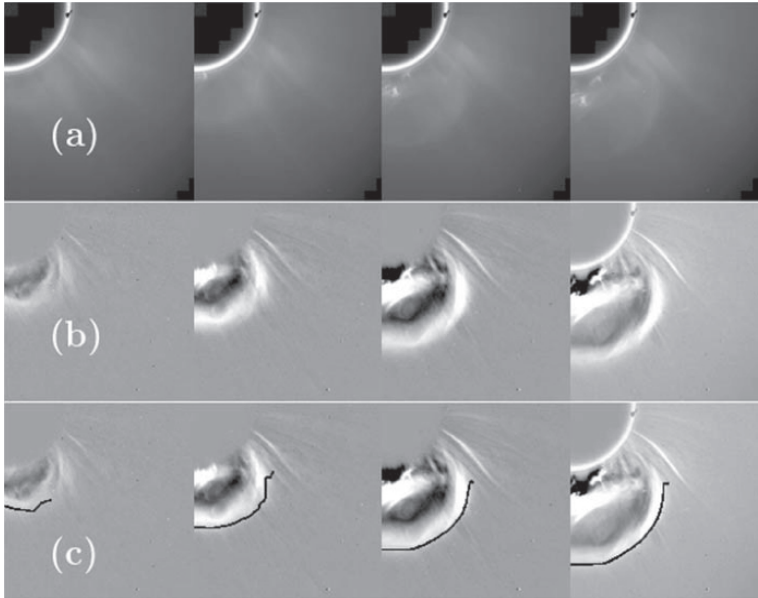


Fig. 6.17 Illustration of a CME edge detection in subsequent images. (a) The original LASCO C2 images, (b) running difference images of the LASCO C2 images, and (c) application of the multiscale edge detection algorithm to the sequence of the original images. The edges are the black lines displayed over the running difference images. The CME erupted on 18 April 2000, the times for the frames are (from left to right) 16:06 UT, 16:30 UT, 16:54 UT, and 17:06 UT (Young and Gallagher 2008).

resulting frequency distribution. To overcome the main disadvantage of threshold-based event detection, namely the loss of weak events in the presence of large events, we test a highpass-filtered and bypass-filtered detection method (Section 6.3). Other alternatives are peak-based detection methods (Section 6.4) and Fourier-filtered time series (Section 6.5). We demonstrate that Fourier power spectra $P(\nu)$ (Section 6.6) or wavelet-based methods (Section 6.7) can be used to retrieve the frequency distribution of time scales $N(T)$. A related method is the principal component analysis (Section 6.8), which has not been much used for SOC statistics yet. Finally, given the availability of imaging data in magnetospheric and solar physics, image-based spatio-temporal detection methods are appropriate for SOC statistics, which have the chief advantage of spatial discrimination of contemporaneous events. We illustrate such spatio-temporal feature recognition techniques for the detection of solar nanoflares and coronal mass ejections. The simulation and testing of any automated event detection technique cannot be taken seriously enough, because systematic errors and biases occur most dramatically for the weakest events, which populate the largest fraction of the logarithmic scale range covered in SOC statistics.

6.11 Problems

Problem 6.1: Simulate the time series described in Section 6.1 for longer mean waiting time (e.g., $\Delta t_0 = 10, 50, 100$) and find the scaling how the number of time-overlapping events reduces with increasing waiting time.

Problem 6.2: Using the test time series simulated in Problem 6.1, develop a simple peak-based event detection algorithm and test whether you can retrieve the frequency distribution of the input parameters for longer waiting times.

Problem 6.3: With a threshold-based event detection algorithm determine how the number of detected events scales with the threshold.

Problem 6.4: Calculate the powerlaw slopes of the frequency distribution of times for the three spectral segments shown in the power spectrum of Cygnus X-1 (Fig. 6.8), using Eq. (6.6.2) and assuming correlations of $E \propto T$ and $E \propto T^2$.

Problem 6.5: Discuss the pro's and con's of Fourier-based, Windowed Fourier transform, wavelet-based, and principal component analysis for periodic, quasi-periodic, and non-periodic time series.

Problem 6.6: Discuss and simulate two different strategies for an image-based event detection method: (1) Detect temporal structures in $n_x \times n_y$ time series $f(t)$ for every image pixel first and then identify co-spatial patterns; or (2) detect spatial structures in each image plane $f(x, y)$ first and then track co-spatial structures in time.