On Burst And Correlation Structure of Teletraffic Models (extended version)

Sándor Molnár, György Miklós
High Speed Networks Lab., Dept. of Telecommunications and Telematics
Technical University of Budapest, H–1111, Sztoczek u. 2., Budapest, Hungary
E-Mail: molnar@bme-tel.ttt.bme.hu, miklos@ttt-atm.ttt.bme.hu

Abstract

The clustering phenomenon of arrivals known as burstiness is an important and poorly understood characteristic of broadband traffic. Most of the B-ISDN services produce bursty traffic and the burstiness has a significant impact on the network performance. However, the nature of burstiness of the various traffic types can be very different.

Since, for the time being, there is no well established notion of burstiness there is need for finding an easily computable measure which can express the inherent nature of burstiness in such a way that it fits well to the practical network dimensioning tasks.

In this paper we review the most popular burstiness measures and give an analysis study by evaluating them with different kinds of traffic produced by a large set of traffic models. Our aim is to contribute to the understanding of the behaviour of the candidate burstiness descriptors which could take us a step closer to establishing a generally acceptable burstiness measure.

1 Introduction

Burstiness has been found a key characteristic of broadband traffic which plays a critical role in determining network performance [5, 18]. However, the nature of bursty traffic is still poorly understood, moreover, there is not even a single and widely-accepted notion of burstiness in the teletraffic literature.

Burstiness expresses the clustering phenomenon of arrivals, that is, when arrivals tend to form clusters with relatively short inter-arrival times within the cluster separated by relatively long intervals. It has a strong relationship to the correlation structure of the traffic. For example, strong positive correlations are a particularly major cause of burstiness. However, this relationship is rather complex and not well understood.

A simple class of burstiness measures takes only the first-order properties into account. These measures can be considered as different characteristics of the marginal distribution of the inter-arrival time. A set of candidates are the moments of that distribution. However, in practice the peak to mean ratio and the squared coefficient of variation are the most frequently used first-order measures in the teletraffic literature [5, 18].

Measures expressing second-order properties of the traffic are more complex. The indices of dispersion [7, 20] and the generalized peakedness [2, 3] are the most well known measures from this class. The indices of dispersion measures include the correlation properties of the traffic and can be very informative [7]. The generalized peakedness measure, which gives a complete second-order characterization of the traffic, takes into account the reaction of a system to a given traffic via the complementary holding time distribution of the system [13].

There are other approaches to capture the bursty nature of the traffic. Recently self-similarity has been identified as a dominant property of the traffic in several packet networks [11]. By the concept of self-similarity the Hurst parameter is also a candidate for burstiness measure

[11, 15, 16]. Other ways of finding a useful practical measure, e.g. based on capturing the queueing behaviour of the traffic is also a research topic [14].

Several traffic models have been proposed for capturing the correlation and bursty structure of broadband traffic [5, 18, 19, 21] but their complex burst-correlation analysis is not always performed. In this paper an analysis study is presented which collects the most popular tele-traffic models (renewal process, Markov Modulated Poisson Process, overflow process, batch renewal process, aggregated process, voice model, video AR+IPP model) and their detailed burst-correlation properties are analyzed. The study includes the most important burstiness measures (peak to mean ratio, squared coefficient of variation of inter-arrival times, higher moments of inter-arrival times, index of dispersion for intervals and counts).

The results give us a better understanding of the connections between the different burstiness and correlation characteristics. We present the corresponding results for the widely used traffic models revealing their burst modeling ability.

2 Traffic Sources

In this section a short description of the analyzed traffic sources is given. Each traffic source was implemented in such a way that it had a mean intensity of 1. Therefore the traffic sources in our simulation study differ only in the shape of traffic and not in the amount of traffic they produce.

2.1 Renewal Models

Renewal models [5] are favoured because of their simplicity: their inter-arrival times form a sequence of independent identically distributed random numbers. In this paper, a renewal model with *Erlang-2* inter-arrival distribution (`ERL2`) and one with *hyper-exponential* inter-arrival distribution (`IPP`) are considered.

2.2 Correlated processes

In contrast to the uncorrelated renewal processes, we used correlated linear, batch renewal and MMPP processes (some of the later processes are also correlated).

A *linear process* [1] is constructed from a renewal sequence $\{\varepsilon_i\}$: $X_i = \sum_{s=-\infty}^{\infty} w_s \varepsilon_{i-s}$ where $\{w_s\}$ is a sequence of coefficients. The usefulness of linear processes is that their correlation structure can be prescribed (although it cannot be arbitrary). In this paper we use a linear process with 12 positive correlation coefficients (`LINPOS`) and a linear process with negative lag 1 and positive lag 2 correlation coefficients (`LINNEG`). *Batch renewal* (`BREN`) processes [9, 10] are useful because they can represent a wide range of different correlation structures. Batch renewal processes are made up of batches of independent identically distributed batch sizes (i.e. number of simultaneous arrivals in a batch), independent identically distributed inter-batch intervals, and the batch sizes are also independent from the inter-batch intervals.

Markov Modulated Poisson Process (`MMPP`) models [4, 5, 21] have become very popular traffic models over the last years. An MMPP consists of a Poisson process whose rate is controlled by the state of a Markov process. In our study we use a 2-state MMPP.

2.3 Overflow and aggregation models

We considered an *overflow model* (`OVER`) to model the traffic which is overflown from a queue with finite buffer, Poisson arrivals and service. This traffic is known to be more bursty than the Poisson process since if one arrival sees the buffer full, the next arrival is more likely to see full buffer, too.

Aggregation models are important in ATM studies. In our analysis the superposition of 100 Interrupted Poisson Processes is used (`AGGR`).

2.4 Voice model

Our voice model (`VOICE`) is for a telecommunications line which is used by many customers (1000 in our example) making telephone calls.

One customer is represented with a hierarchical model: the customer is either in call state or in inter-call state. A call state is further divided into talk state and two types of silences: inter-word state and listen state. (See Fig. 1.) In our model, we used exponential distributions for the length of silences and talk periods (see [6] for parameters and justification of exponential distributions).

In this model, the traffic generated by one customer is bursty, however, the superposition traffic generated by many customers is smoothed.

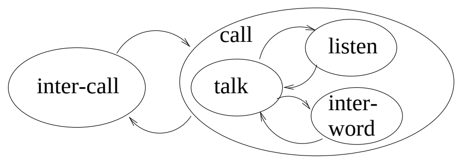

Figure 1: The hierarchical voice model for one customer

2.5 Video model

There are many types of video coders with different traffic characteristics. Our model (`VIDEO`) is for a coder which uses predictive coding. In this case, the frames where scene changes occur are much larger than the other frames.

We used an IPP+AR(1) model to capture this type of behaviour [17]. The IPP (Interrupted Poisson Process) model represents the infrequent and abrupt scene changes. The AR(1) (autoregressive model) captures the high correlation in the frame lengths when there are no scene changes. Our model was used to create the sequence of frame sizes. The arrivals in one frame are spread uniformly supposing constant rate over a frame.

3 Burst and Correlation Measures

In this section we give a short description of the investigated burst and correlation measures. Their application is discussed in the next section.

3.1 Measures based on the first order properties

One of the widely used measures is the *peak to mean ratio* (PMR) [5]. However, the definition and the applicability of a peak is not at all clear. In the case of unshaped traffic, the peak determined by the two closest arrivals may be very high and likely to correspond to two arrivals in consecutive slots in practice. In the case of shaped traffic, the peak gives more information on the shaper parameter settings than the traffic itself. Moreover, there is a variety of traffic structures corresponding to the same peak to mean ratio.

Also widely used is the *squared coefficient of variation* (SCV) of the inter-arrival times [5] which includes information from the first two moments and is defined as $C^2(X) = Var(X)/E^2(X)$ where X is the inter-arrival time.

Higher moments can also give important information about the traffic. For example, two traffic with the same first two moments but different third moment can produce very different queueing behaviour.

3.2 Measures based on the second order properties

Indices of dispersion measures are useful because they show the traffic variability over different scales and they capture the correlation structure. Two indices of dispersion measures are used: the *index of dispersion for intervals* (IDI) is related to the sequence of inter-arrivals; the *index of dispersion for counts* (IDC) is related to the sequence of counts of arrivals in consecutive time units [1, 7, 9].

The index of dispersion for indices is defined for a stationary inter-arrival sequence $\{X_i\}$ as follows:

$$J_k = \frac{Var(X_{i+1} + \ldots + X_{i+k})}{kE^2(X)}.$$

It is known that the IDI is constant for renewal processes, and it is easily seen that $J_1 = C^2(X)$.

In the definition, the variance of the sum of k consecutive inter-arrivals is taken. In the case of bursty processes, the short and long inter-arrivals are grouped together, and it causes the IDI to increase with increasing k. In fact, the increase or decrease in the IDI graph is directly related to the correlation of the inter-arrival sequence.

The index of dispersion for counts for a stationary process is defined as

$$I_t = \frac{V(t)}{E(t)} = \frac{V(t)}{tm}$$

where $V(t)$ and $E(t)$ are the variance and expected number of the arrivals in an interval of length t, and $E(t) = tm$, where m is the mean intensity of arrivals.

The IDC shows the variability of a process over different time-scales. It is defined so that it is constant 1 for the Poisson process; however, it is not in general constant for a renewal process. A well known connection between the IDI and IDC of a given stationary process is that the value at infinity of the two curves are equal.

The slope of the IDI and IDC curves are directly related to the correlation structure of the traffic [1, 7, 12]:

The slope of the IDI graph is

$$J_{k+1} - J_k = C^2(X) \frac{\sum_{l=1}^{k} l\rho_l}{\frac{k(k+1)}{2}}.$$

where ρ_l is the lag l correlation in the inter-arrival time sequence. In words, it is the squared coefficient of variation times the weighted average of the correlation coefficients from 1 to k, where the weights are just equal to the lag.

Similarly, the slope of the IDC graph is

$$I_{(k+1)\tau} - I_{k\tau} = \frac{V(\tau)}{E(\tau)} \frac{\sum_{l=1}^{k} l\rho_l(\tau)}{\frac{k(k+1)}{2}}.$$

where $\rho_l(\tau)$ is the lag l correlation in the sequence of counts in intervals of length τ, $E(\tau)$ and $V(\tau)$ are the mean and variance of counts in that length. In words, it is the variance to mean ratio for the number of arrivals in a period of τ times the weighted average of the correlation coefficients from 1 to k, where the weights are just equal to the lag.

4 Analysis Results

The various burst characterization methods listed in the previous section were applied to the traffic sources described in Section 2. For each traffic type the source parameters were set to get sample traffic sequences. (The detailed description of parameter sets of the models can be found in [12].) We used traffic sequences of 10000 inter-arrivals (all with mean intensity 1) in our simulation study. A sample intensity-time trace, the pdf of the inter-arrival times, correlation

	PMR	SCV	m_3
ERL2		0.500 ± 0.014	3.03 ± 0.12
IPP		5.03 ± 0.3	62 ± 14
LINPOS	1.234 ± 0.06	0.0031 ± 0.0004	1.013 ± 0.01
LINNEG	1.46 ± 0.02	0.0123 ± 0.0003	1.0367 ± 0.0027
BREN		4.7 ± 1.0	48 ± 18
MMPP		2.5 ± 0.3	23.0 ± 3.3
OVER		4.03 ± 0.17	40.6 ± 1.8
AGGR		1.07 ± 0.02	6.45 ± 0.9
VOICE	1.42 ± 0.09	0.0181 ± 0.004	1.058 ± 0.03
VIDEO	8.6 ± 2.2	1.42 ± 0.5	6.91 ± 1.9

Table 1: Peak to mean ratio (where applicable), squared coefficient of variation and third moment of inter-arrival times for the sample traffic sources. (The first moment was set to 1.)

sequence for the inter-arrival times and for the counts in successive intervals of unit length, the IDI and IDC graphs are shown at the end of the paper (see Fig. 1-10). The PMR, SCV and the third moment are listed in Table 1. Each computation was made 10 times and 90% confidence intervals are shown. 10 experiments may not be enough from a statistical point of view but it is very informative concerning what we can get in practice with quick computation.

The figures give us a quick summary of the properties of a wide range of teletraffic models.

4.1 Peak to mean ratio

Among the sources presented in this paper, the peak to mean intensity ratio as determined by the closest two arrivals can only be evaluated for the sources `LINPOS`, `LINNEG`, `VOICE`, `VIDEO`. In the case of the other sources, the peak is not limited from above since we use unslotted inter-arrival times. (In slotted time, the corresponding models would give a peak corresponding to two arrivals in consecutive slots.)

4.2 Squared coefficient of variation

The results show that the squared coefficient of variation can usually be computed accurately. Its use as a burstiness measure, however, is limited. This is shown, for example, by comparing the values for example of the `VIDEO` and `IPP` processes. It is much higher for the `IPP`, yet the `VIDEO` source is much more bursty because it includes sustained high intensities. (Compare Fig. 2 and Fig. 9.) Also, the low value of the squared coefficient of variation of the `VOICE` indicates a very smooth process, however this traffic is characterized by short and long-term fluctuations (see Fig. 8).

The squared coefficient of variation takes into account only the set of inter-arrival times; the order of the inter-arrivals are disregarded. We can think of burstiness as being caused by two factors [5]: the distribution and especially the tail of the inter-arrival times and the correlation between them. The squared coefficient of variation being a first-order measure takes into account only the inter-arrival distribution and the results indicate that it is not adequate.

4.3 Higher moments

The third moment tells us about the long inter-arrivals. For the sources where the inter-arrival times can be very high (e.g. `IPP`, `MMPP`, `OVER`), the third moment is much higher as compared to the sources where the inter-arrival time is bounded from above (e.g. `VOICE`, `VIDEO`).

4.4 Index of dispersion measures

From the figures it can be seen that the index of dispersion measures cannot be estimated as accurately as the squared coefficient of variation. But once we have the IDI and IDC of a traffic, they are very informative.

In the case of `MMPP`, `AGGR`, `VOICE`, `VIDEO` the IDI and IDC curves both increase. Together they imply that the low burstiness in short scales (shown by the relatively low squared coefficient of variation) increases over higher scales due to positive correlation. In the case of the `VIDEO` the quickly increasing curves and the high value at infinity imply that this is a very bursty source.

As seen from the `ERL2`, `IPP` and `OVER` IDI and IDC graphs, the constant IDI (meaning no correlation in the inter-arrival sequence) does not imply uncorrelated sequence of counts as seen from the fact that the IDC curve is decreasing in the case of `ERL2` and increasing in the case of `IPP` and `OVER`.

The example source of `BREN` (with our particular parameter settings) shows that the other case is also possible: in this case we have low correlation in the sequence of counts (IDC is horizontal) and high correlation in the sequence of inter-arrivals (IDI is increasing).

The IDC for the `LINPOS` and `LINNEG` sources are interesting because they exhibit waves. This can, however, be attributed to the very small support of the inter-arrival time distribution. This quasy-periodic nature is increased by the negative correlation for `LINNEG` which is why the IDC is more wavy in this case.

As for the accuracy of the curves, the IDI and IDC for the `voice` source shows that these curves can be highly inaccurate, even though the curve is only plotted to 10% of the simulated trace. The reason for this inaccuracy is that the `voice` source is characterized by very long fluctations. It folows that the accuracy of the index of dispersion curves is very dependend not only on the length of the data but on the data itself.

5 Conclusion

In this paper we have reported a burstiness and correlation analysis study by investigating the most important candidate measures of burstiness and applying them to the traffic generated by a large set of traffic models. The different burst-correlation modeling ability of the models are reported.

Our results indicate that the first-order measures like the squared coefficient of variations and the peak to mean ratio are not able to express the important bursty nature in several cases. These measures are frequently used in teletraffic and their limitations should be recognized.

From the second-order measures the indices of dispersion are investigated in this paper (for results relating to the generalized peakedness we refer to [12, 13]) and found to be very informative and useful in practice. A quick evaluation of IDI and IDC curves immediately reveals a lot of important correlation and burstiness information about the traffic. We believe that a good understanding of IDI and IDC properties gives us a very efficient tool to understand the nature of the traffic. However, for practical purposes we often need a simple scalar or a measure with few parameters. Choosing the important points of the IDI and IDC curves can be a candidate for such a measure and finding them is one of our future research topic.

References

[1] D. R. Cox, P. A. W. Lewis, "The Statistical Analysis of Series of Events", Methuen & Co Ltd, 1966.

[2] A. E. Eckberg, "Generalized Peakedness of Teletraffic Processes", *ITC-10*, 1984.

[3] A. E. Eckberg, "Approximations for Bursty (and Smoothed) Arrival Queueing Delays Based On Generalized Peakedness, *ITC-11*.

[4] W. Fischer, K. Meier-Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook", *Performance Evaluation 18* (1992) 149-171, North-Holland.

[5] V. S. Frost, B. Melamed, "Traffic Modeling For Telecommunications Networks", *IEEE Communications Magazine*, March, 1994.

[6] J. G. Gruber, 'A Comparison of Measured and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection', *IEEE Transactions on Commununications*, Vol. COM-30, No. 4, pp 728-738, 1982

[7] R. Gusella, "Characterizing the Variability of Arrival Processes with Indexes of Dispersion", *IEEE Journal on Selected Areas in Communications*, February 1991, Vol 9., No. 2.

[8] F. Kamoun, M. Mehmet Ali, "Statistical Analysis of the Traffic Generated by the Superposition of N Independent Interrupted Poisson Processes" *Third Canadian Workshop, Information Theory and Applications*, Rockland, Canada, May, 1993.

[9] D. Kouvatsos, R. Fretwell, "Batch Renewal Process: Exact Model of Traffic Correlation", *High Speed Networking for Multimedia Application*, Kluwer Academic Press, pp. 285-304, 1996.

[10] D. Kouvatsos, R. Fretwell, "Closed Form Performance Distributions of a Discrete Time $GI^G/D/1/N$ Queue with Correlated Traffic", *Proceedings of the 6th IFIP Conference on Performance of Computer Networks* (Istambul, 1995), Chapman and Hall Publishers, London.

[11] W. E. Leland, M. S. Taqqu, W. Willinger, D. W . Wilson "On the Self-Similar Nature of Ethernet Traffic (extended version)", *IEEE/ACM Transactions on Networking*, Vol.2, No.1, Feb. 1994.

[12] Gy. Miklós, "Burstiness and Correlation Measures in Teletraffic", *Master's Theses*, Technical University of Budapest, Faculty of Electrical Engineering and Informatics, 1997.

[13] S. Molnár, "Evaluation of Quality of Service and Network Performance in ATM Networks", *PhD dissertation*, Technical University of Budapest, Department of Telecommunications and Telematics, 1995.

[14] S. Molnár, I. Cselényi, N. Björkman, "ATM Traffic Characterization and Modeling Based on the Leaky Bucket Algorithm", *IEEE Singapore International Conference on Communication Systems*, Singapore, November 25-29, 1996.

[15] S. Molnár, A. Vidács, A. Hegedűs "Meeting a Challenge: Modeling Self-Similar LAN/MAN Traffic", *Proc. 8th IEEE Workshop on Local and Metropolitan Area Networks*, Berlin/Potsdam, Germany, Aug. 1996.

[16] S. Molnár, A. Vidács "On Modeling and Shaping Self-Similar ATM Traffic", *Proc. 15th International Teletraffic Congress*, Washington, D.C., USA, June, 1997.

[17] K. Onda, K. Nakagawa, "Approximation of Video Cell Traffic by AR(1) + IPP Model", *Electronics and Communications in Japan*, Part 1, Vol. 78, No. 8., 1995.

[18] R. O. Onvural, "Asynchronous Transfer Mode Networks, Performance Issues", 1994, Artech House.

[19] H. Saito, "Teletraffic Technologies in ATM Networks", 1994, Artech House.

[20] K. Srivam, W. Whitt, "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data", *IEEE Journal on Selected Areas in Communications*, Vol. 4, No. 6, September, 1986.

[21] G. D. Stamoulis, M. E. Anagnostou, A. D. Georgantas, "Traffic Source Models for ATM Networks: a Survey", *Computer Communications*, vol. 17, No. 6, June, 1994.

Figure 1: ERL2, Renewal model with Erlang-2 inter-arrival time distr.

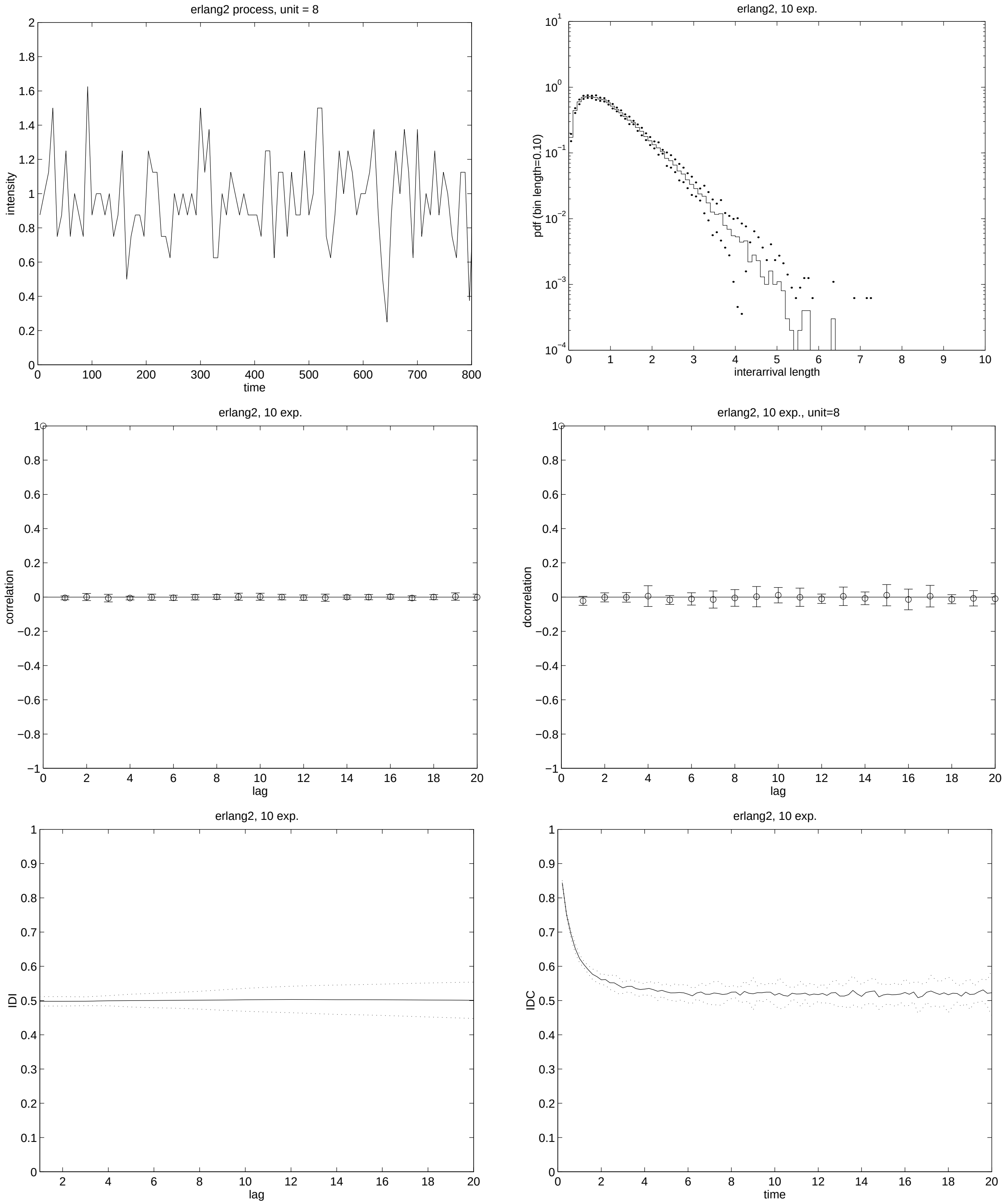

Figure 2: IPP, Renewal model with hyper-exponential inter-arrival time distr.

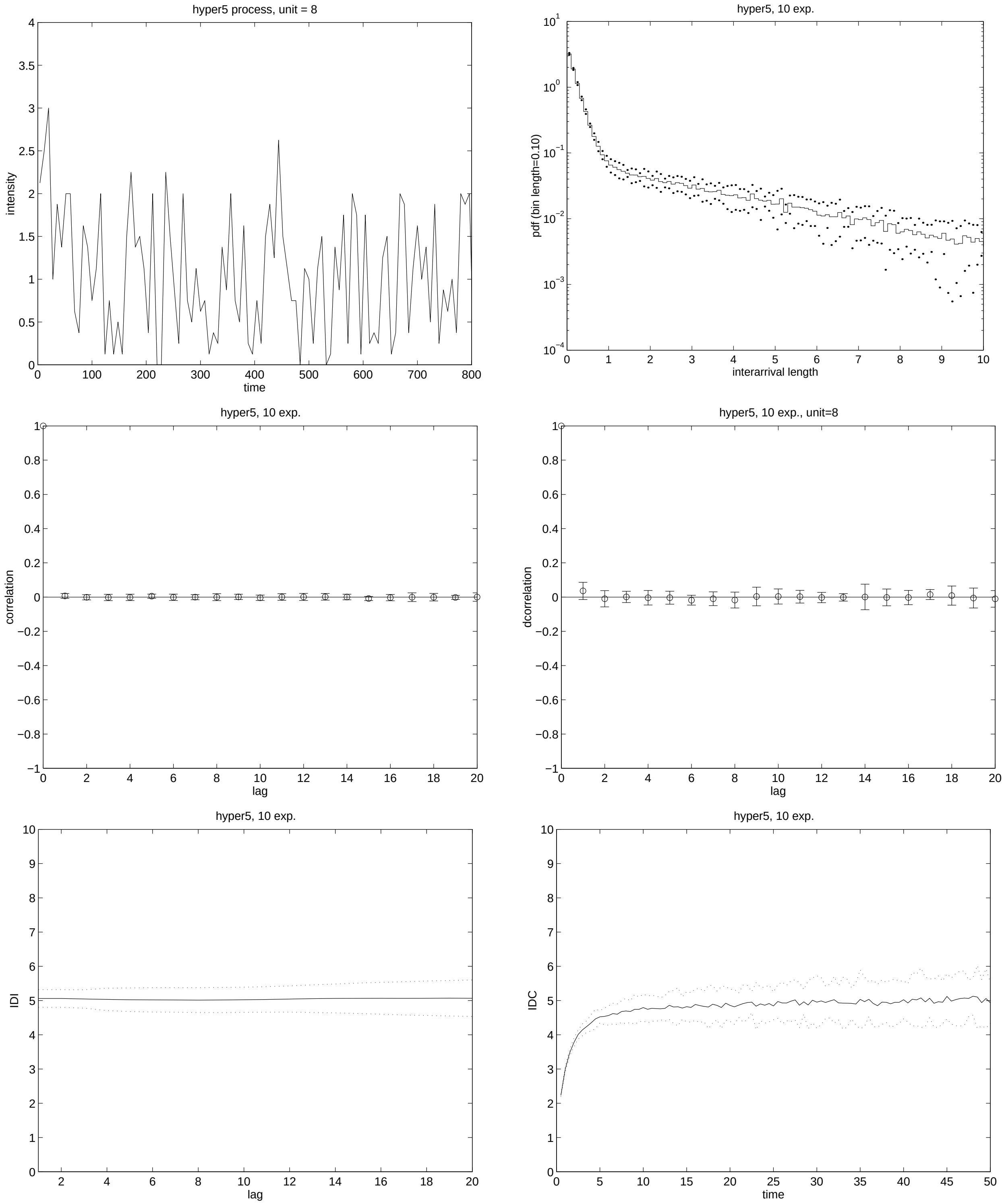

Figure 3: LINPOS, Linear process with 12 prescribed positive correlation coeff.

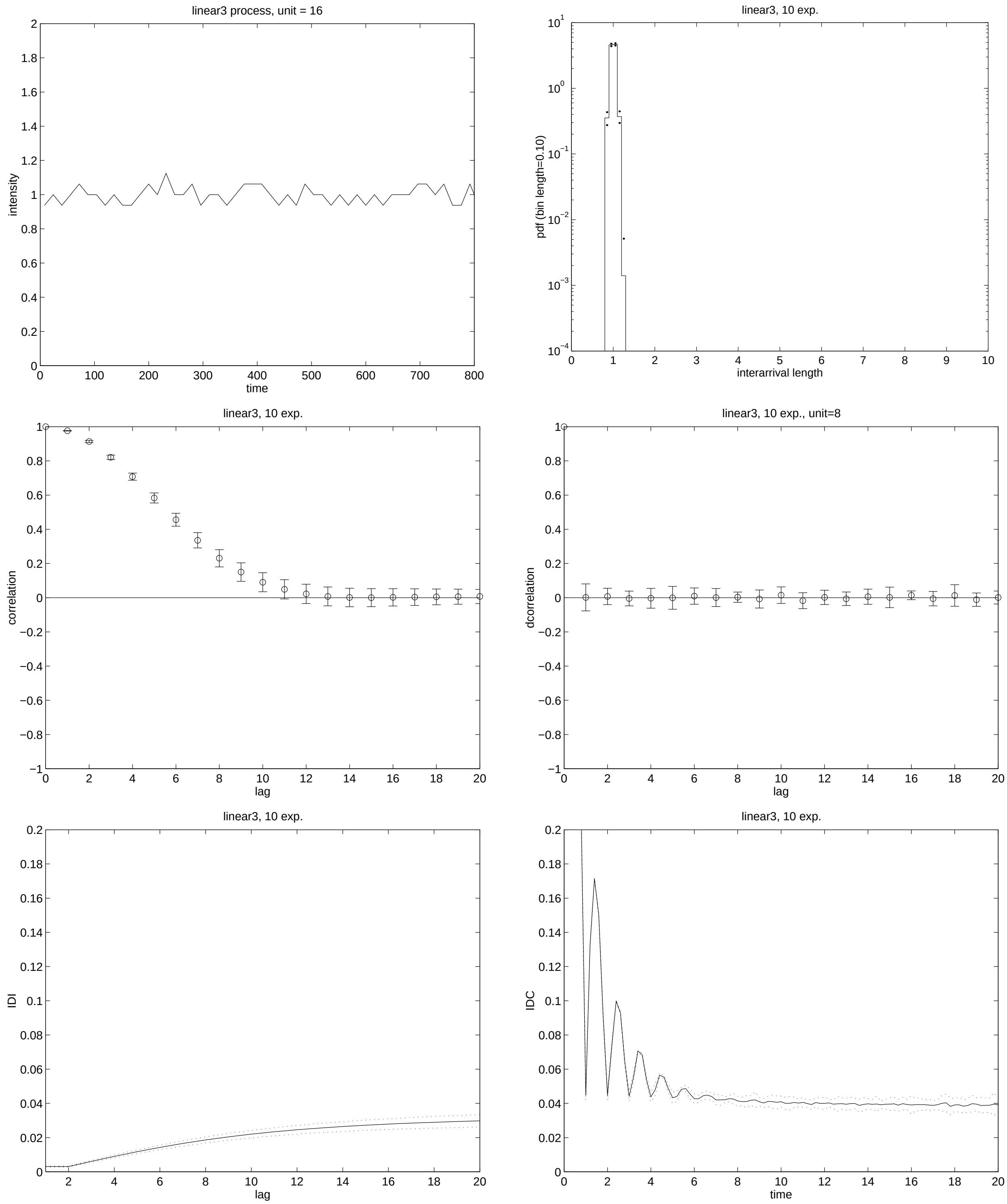

Figure 4: LINNEG, Linear process with negative lag 1 and positive lag 2 corr. coeff.

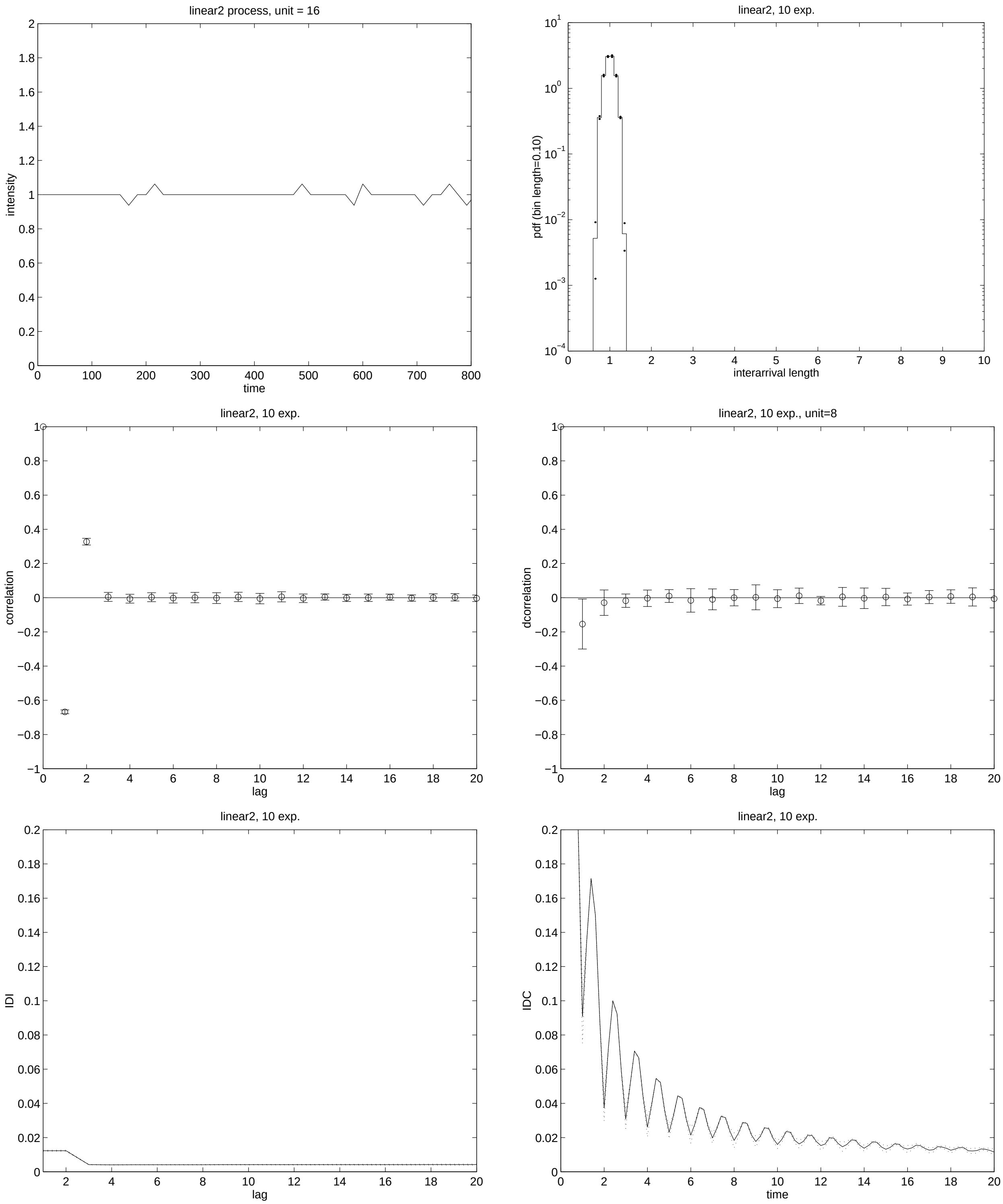

Figure 5: BREN, Batch renewal process

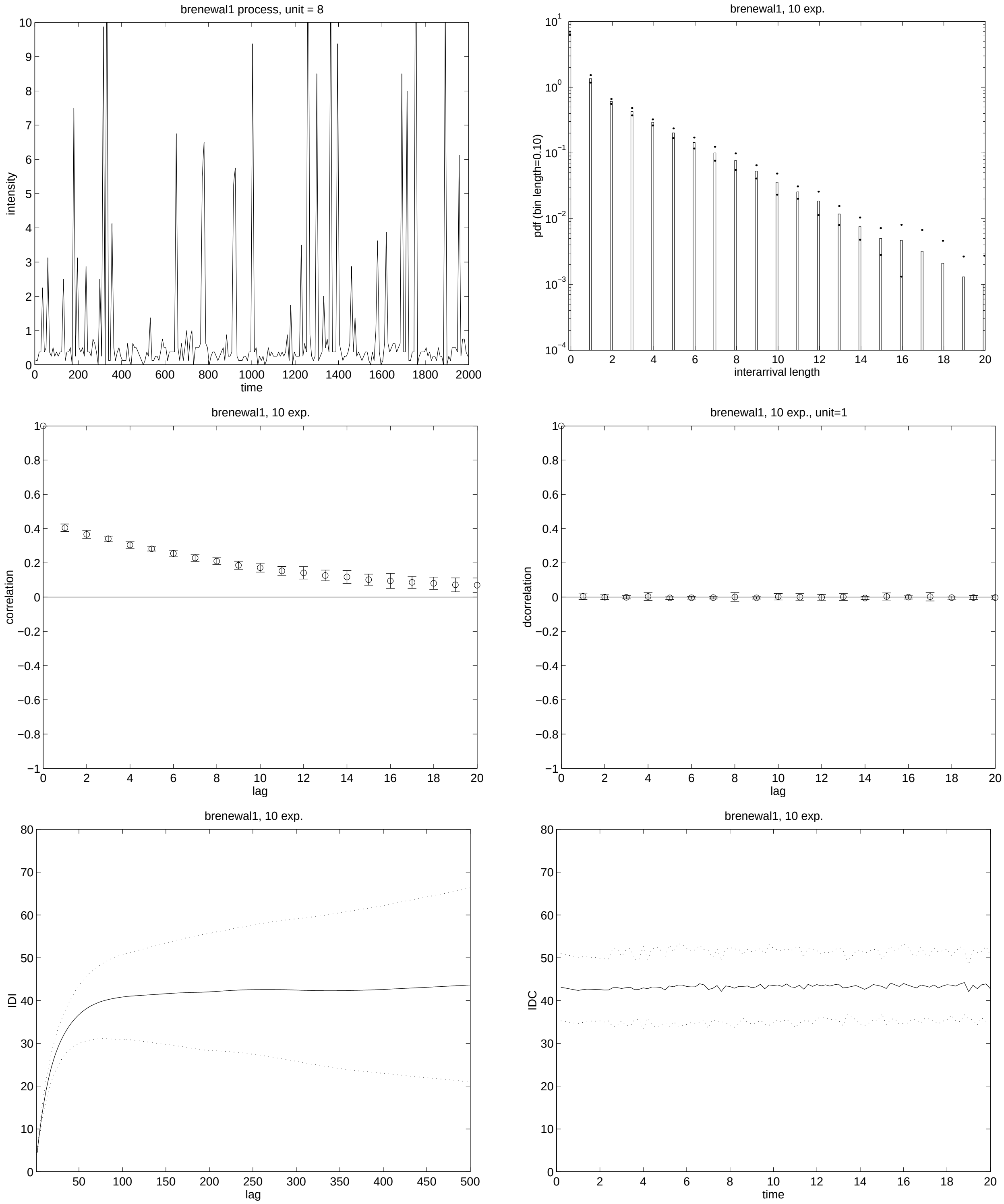

Figure 6: `MMPP`, Markov Modulated Poisson Process with two states

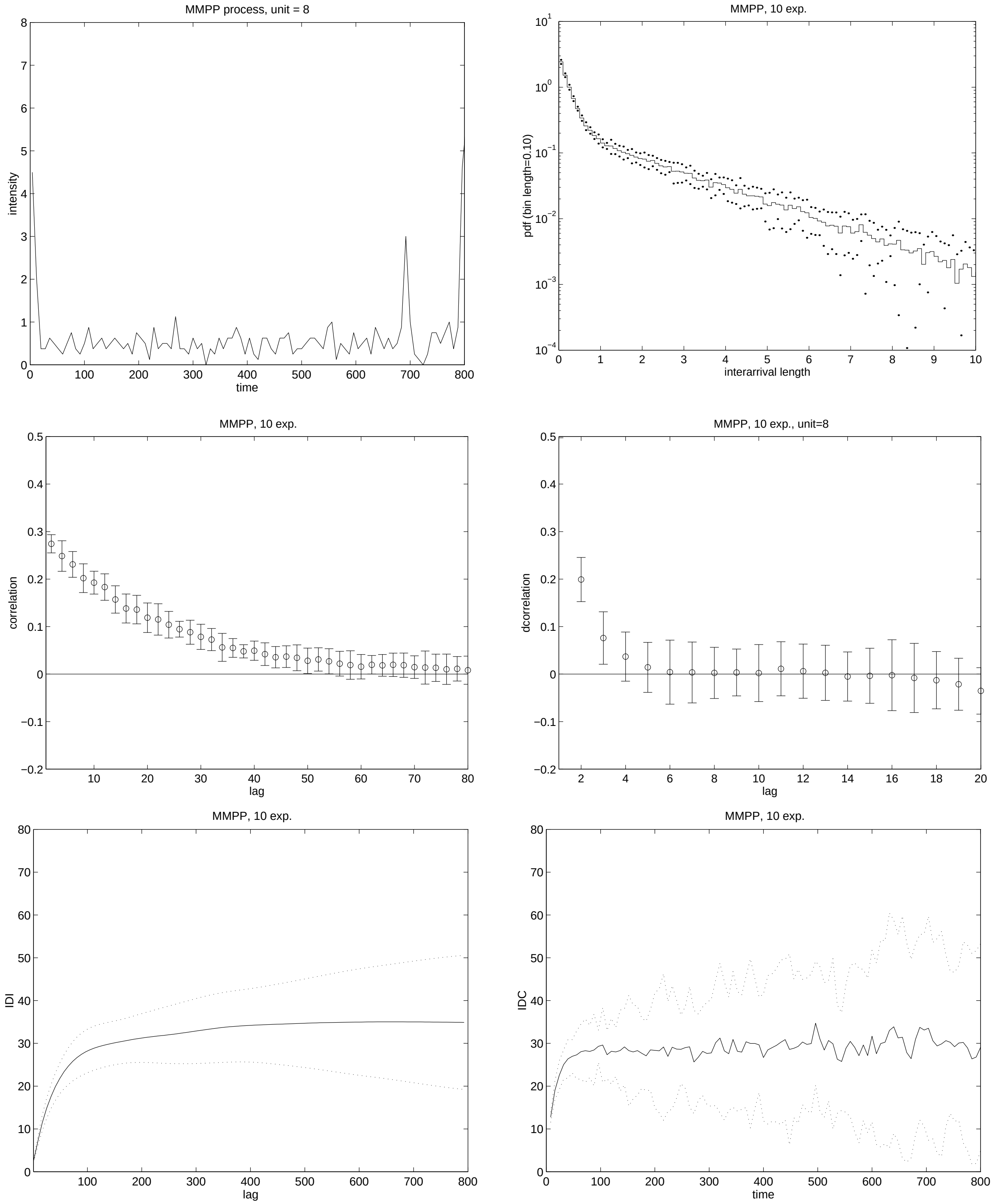

Figure 7: OVER, Overflow process

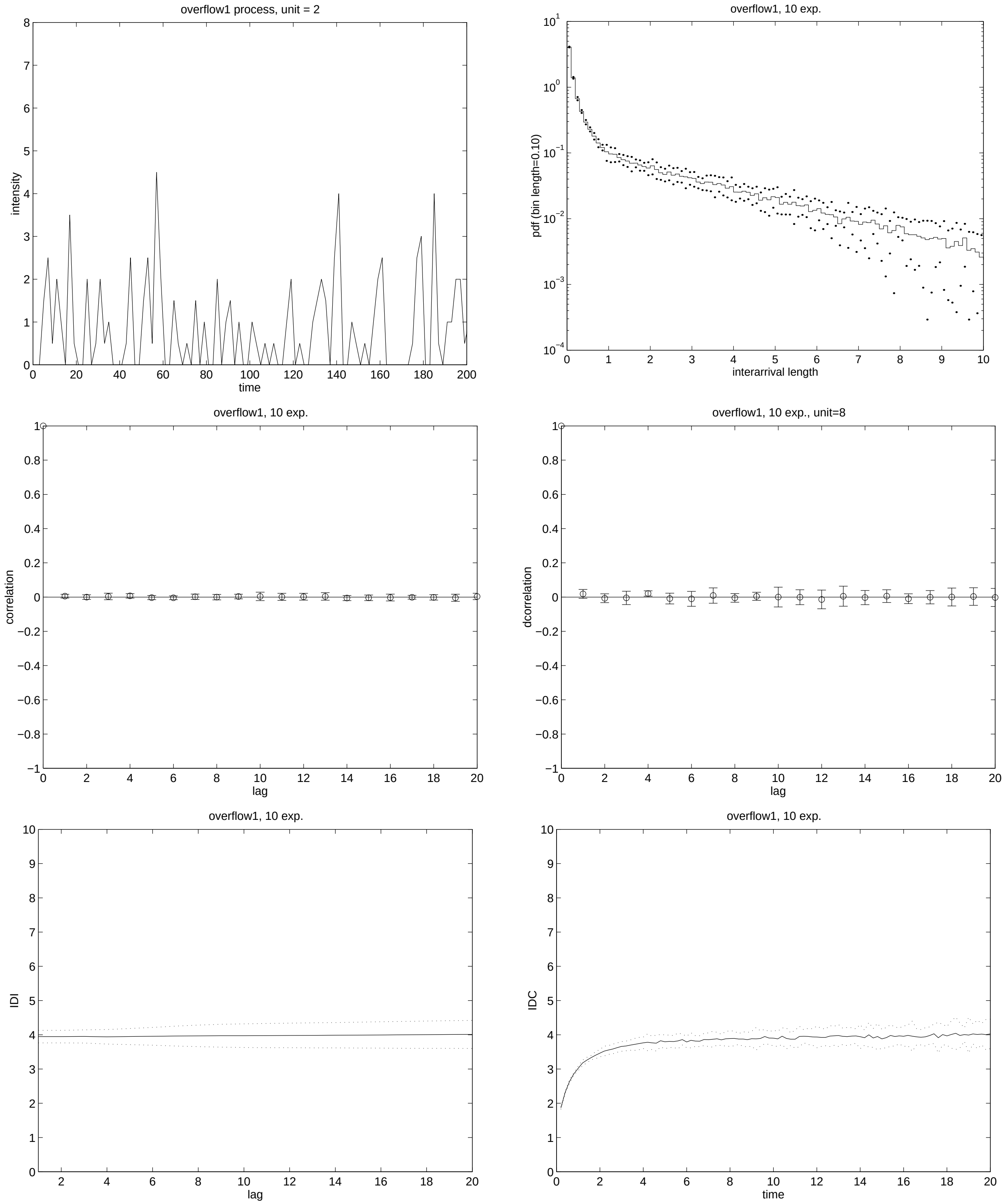

Figure 8: AGGR, Aggregated process, superposition of 100 IPPs

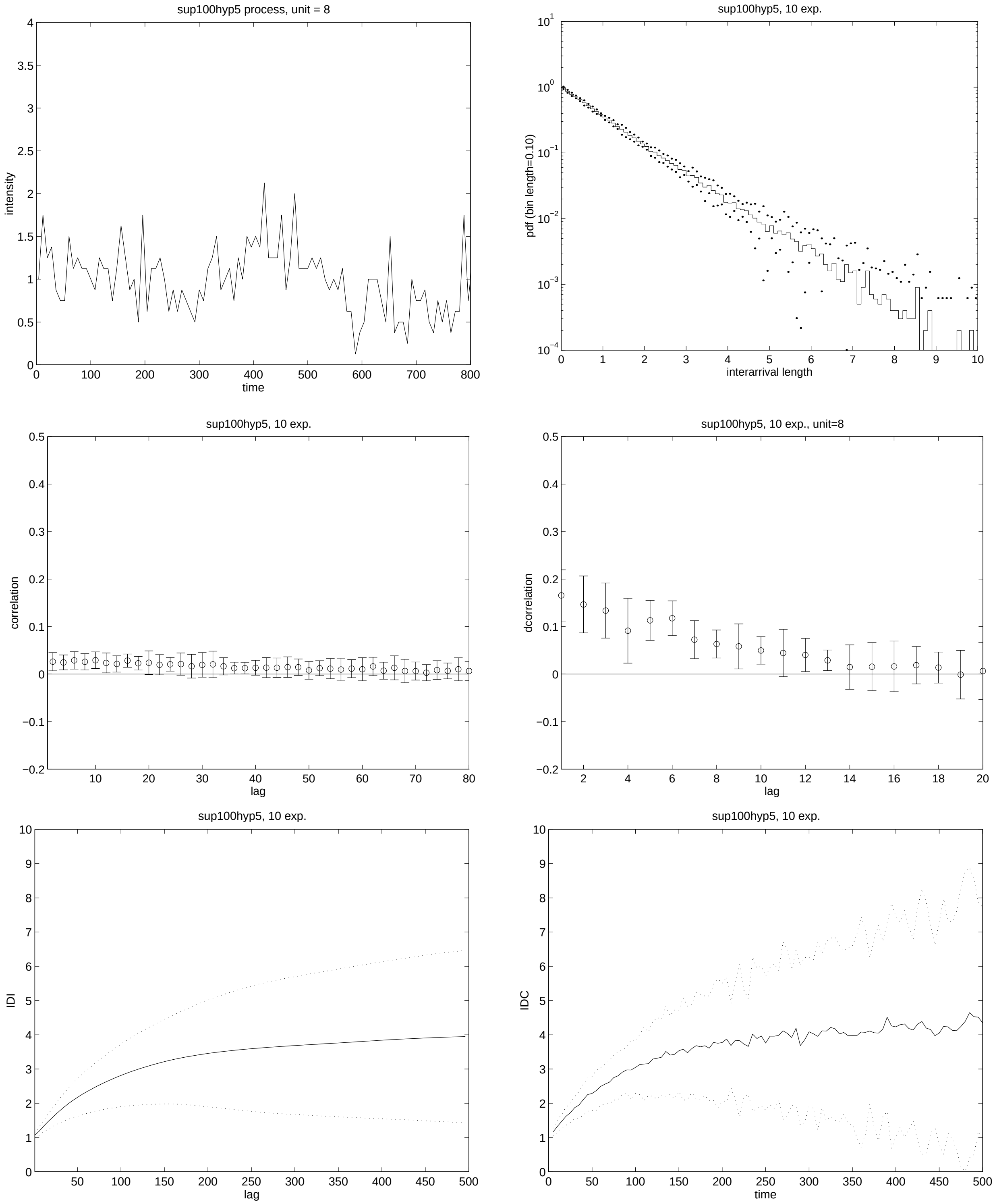

Figure 9: VOICE, Voice model with 1000 customers

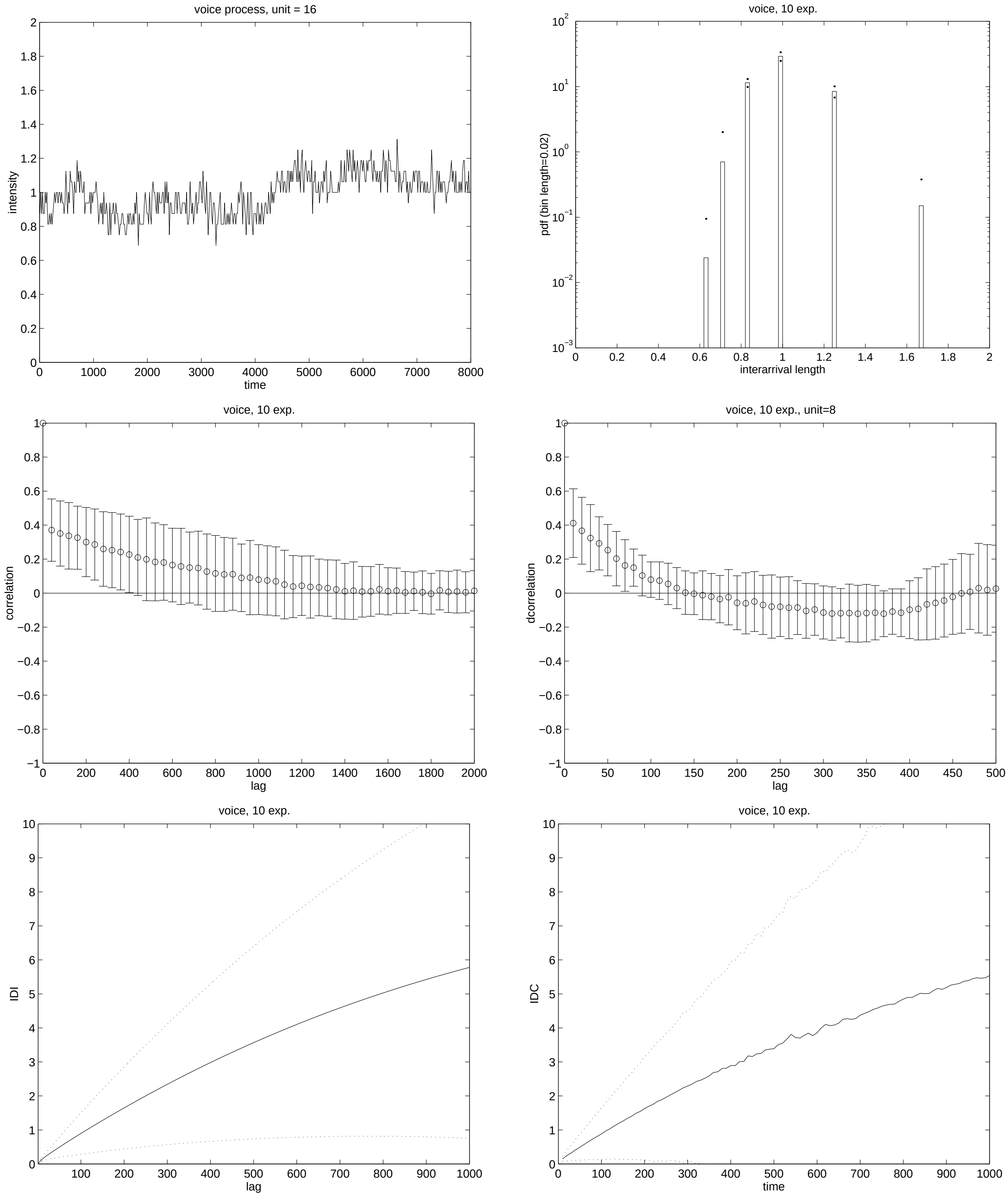

Figure 10: `VIDEO`, Video model