Extraction of a deterministic component from ROSAT X-ray data using a wavelet transform and the principal component analysis: Analysis of random pulses and simulated data

Ludwik Liszka, Steward Observatory, University of Arizona, Tucson, AZ, USA *

Abstract

The main purpose of this part of the study is to calibrate the method of separation of deterministic variations of an event series from its Poisson part. The method is used for search for deterministic structures in computer-generated random number sequences and also in the output of, what is believed, unpredictable physical processes, like noise-diode generated pulses or the radioactive beta-decay. It has been found that a beta-decay generated event series contains less deterministic structures than the other tested. Four typical deterministic components are then added to the beta-decay event series and the ampligram and time scale spectrum (analysis tools defined in Liszka & Holmström [1999]) is determined for each composite series. It is found that even a small deterministic component, adding only 10% or less events to the total series, reflects clearly in the ampligram and in the time scale spectrum. The second part of the paper deals with simulation of photon flux consisting of a random or deterministic sequence of Poisson-like, decaying, elementary building blocks. Again, a clear difference is found between time scale spectra of photon series consisting of randomly distributed building blocks and those consisting of quasi-regular repetition of building blocks. Finally, a technique is presented to remove the deterministic or Poisson-like part of the event series. This can be used to improve the beta-decay series, decreasing the amplitude of the deterministic component by a factor of ten, or to obtain wavelet spectra of pure deterministic variations of the photon flux.

1 Introduction

It has been shown that the short-term variations of photon counts recorded by the ROSAT X-ray satellite contain deterministic information, most likely corresponding to intensity variations of the source. Wavelet spectra combined with the principal component analysis seem to be an useful tool to study the short-term temporal variations in the photon counting rate (Liszka [1997], Liszka & Holmström [1999]), even in cases of very low counting rates. The present method is based on a wavelet transform combined with the principal component analysis. The time series of photon counts is transformed into the wavelet coefficient (spectral density) - time plane. An ampligram - a graph displaying phase relations in the time series at different spectral densities - is created. The ampligram may be used to determine the number of significant independent components in the data and to determine the time scale spectrum of the signal in the spectral density - time scale domain. An interesting property of the time scale spectrum is, that deterministic periodic or semi-periodic structures in the data are mapped on the graph as vertically elongated features, while purely stochastic structures are mapped as horizontally elongated features.

*Permanent address: Swedish Institute of Space Physics, Umeå Division, Sörfors 634, S-905 88 Umeå, Sweden

2 Analysis of random event series

The purpose of this part of the study is to calibrate the method and to understand the physical meaning of results. The first task was to check whether ideal random event series, which do not show any temporal structures, really exist.

The first case analyzed was the Table of 500 000 Random Numbers which was found on the Internet. The random numbers X_i from the beginning of the table were normalized to the maximum value and converted into a Poisson event series, where:

$$t_\mathrm{i} = t_{\mathrm{i}-1} - \frac{\ln X_\mathrm{ni}}{C} \qquad (1)$$

where t_i is time of event i, X_ni is a random number between 0 and 1 and C is the Poisson parameter. The event series is then converted into an event rate plot

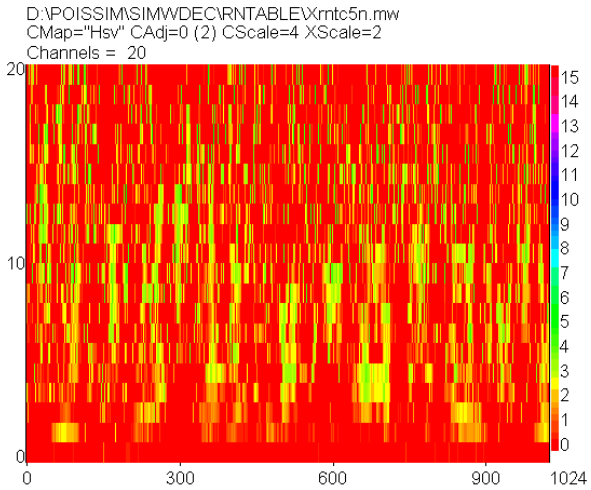


Figure 1: A low-20 ampligram of a 1024-point sample of an event rate curve constructed using a random number table. Semi-periodic structures correspond to the wavelet coefficient magnitude less than 12% of its maximum value. The vertical scale corresponds to the percentage of the maximum wavelet coefficient, the horizontal scale shows the bin number.

by counting events in consecutive time bins. A low-20 ampligram for a 512-point sample of such event rate plot is shown in Fig. 1

It may be seen that an event series constructed using a random number table is not a good Poisson series. Series constructed using simple random number generators usually attached to compilers give even worse results.

Fig. 2. shows an example of ampligram for a pulse series generated by a noise-diode based "random" pulse generator, earlier frequently used in particle physics. Pulses are counted in 1024 1 second bins. The ampligram of that pulse rate time series shows semi-regular components up to 20% of maximum wavelet coefficient magnitude.

The best results were obtained using a pulse series from the radioactive beta-decay, see Fig. 3. Pulses are counted in 1024 one second bins. The semi-regular structures are here, as a rule, less than 3% of maximum wavelet coefficient magnitude.

The pulse trains from the radioactive beta-decay are therefore used in the present study for calibration purposes. The time scale spectrum for a 1024-point ampligram of beta-decay data is shown in Fig. 4.

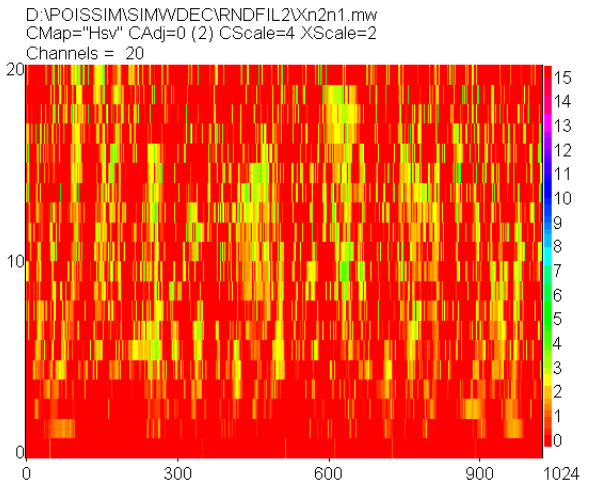


Figure 2: A low-20 ampligram of a 1024-point sample of a pulse series generated by a noise-diode based "random" pulse generator. The vertical scale corresponds to the percentage of the maximum wavelet coefficient, the horizontal scale shows the bin number (1 second bins).

3 Superposition of beta-decay data and simulated data

The next question is whether it is possible to separate deterministic components from pure stochastic variations of counting rate expressed in events/bin (the number of bins is used as a measure of time). For that purpose a synthetic pulse series, obtained by adding a beta-decay series to a weak deterministic component defined according to four templates presented in Fig. 5.

There are two cases of positive components and two cases of bipolar component, all with a period P bins and a maximum amplitude R counts/bin. When the sum of both series is < 0 for bipolar components, a zero value is put into that bin. The difference between cases 1 and 3, respective 2 and 3 is the width of the peaks. Cases 1 and 2 carry more energy than 3 and 4. On the other hand cases 1 and 3 entail an increase of number of events, while cases 2 and 4 entail mainly a redistribution of events in time. A sample of time scale spectra of ampligrams for the beta-decay data in Fig. 3 with superimposed all four templates of Fig. 5. for P=57 bins and R=3 counts/bin is shown in Fig. 6.

It may be seen that the peak at 57 bins is clearly visible for cases 1 and 2. The deterministic component can not be clearly seen for cases 3 and 4. On contrary, addition of narrow deterministic peaks to the beta-decay data increases number of horizontal struc-

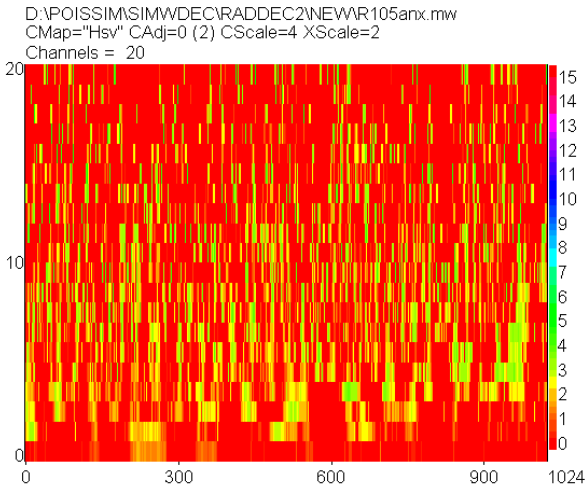


Figure 3: A low-20 ampligram of a 1024-point sample of a pulse series from the radioactive beta-decay. The vertical scale corresponds to the percentage of the maximum wavelet coefficient, the horizontal scale shows the bin number (1 second bins).

tures in cases 3 and 4. It must be remembered that a single-bin case will easily interact with the random part of the signal and thus generate horizontal structures. A consequence of the present observation is that the deterministic component must cover at least two consecutive time-bins in order to be detected as a such. Low-level single-bin events will only increase the random noise. It is interesting that in the multi-bin cases there is a clear difference between a bipolar case and a twice as large positive case. The case 2 in Fig. 6. with peak value R=3 would correspond to the case 1 with R=6. However, that case, shown in Fig. 7. a higher amplitude of the peak is obtained on the time scale spectrum and the stochastic structures are stronger suppressed than in the case 2.

The consequence of the present observation is that it is, in principle, possible to distinguish a case when a deterministic component is added to the background from a case when a deterministic process modulates the background. Already simple simulations indicate that is possible to separate deterministic components from pure stochastic variations of counting rate using a non-linear filtering technique based on a wavelet transform.

4 Analysis of simulated photon flux

In the case of X-rays from AGN there might be a physical source of deterministic variations of the photon flux. Pacholczyk and Stoeger ([1998]) propose "building blocks" in the X-ray photon flux from active galactic nuclei. In the present study each building block is assumed to be formed by a Poisson process, like that described by expression (1), but with the process intensity C decreasing with time according to:

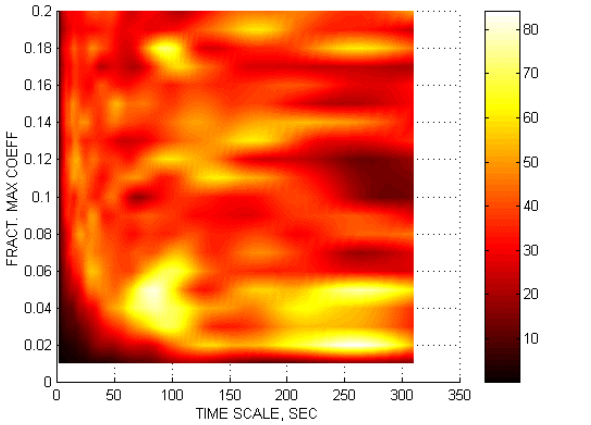


Figure 4: A time scale spectrum for a 1024-point ampligram of beta-decay data. The vertical scale in %/100.

$$C = C_{\mathrm{o}} t_{\mathrm{j}}^{-k} \qquad (2)$$

where C_{o} is the starting process intensity, t_{j} is the time of the j-th photon measured from the beginning of the building block and k is the exponent determining the rate of decay of the block. The simulations were performed assuming random values for the number of photons in each block, NP, the number of building blocks during the simulation period, NB, and a pair of values for C_{o} and k. An example of a single building block for NP=200 photons, C_{o}=5 and k=2 is shown in Fig. 8. Building blocks were superimposed randomly (Section 4.1) during the simulation period or following a semi-deterministic pattern (Section 4.2).

4.1 Randomly superimposed building blocks

A simulation made assuming NB=35, NP=214, C_{o}=12.6 and k=2.27 is presented in Fig. 9. Photon events are counted in 1 second bins during the simulation period. The plot of event rate is shown.

It may be seen that the time scale spectrum is dominated by horizontal structures. It is in agreement with results in the previous section.

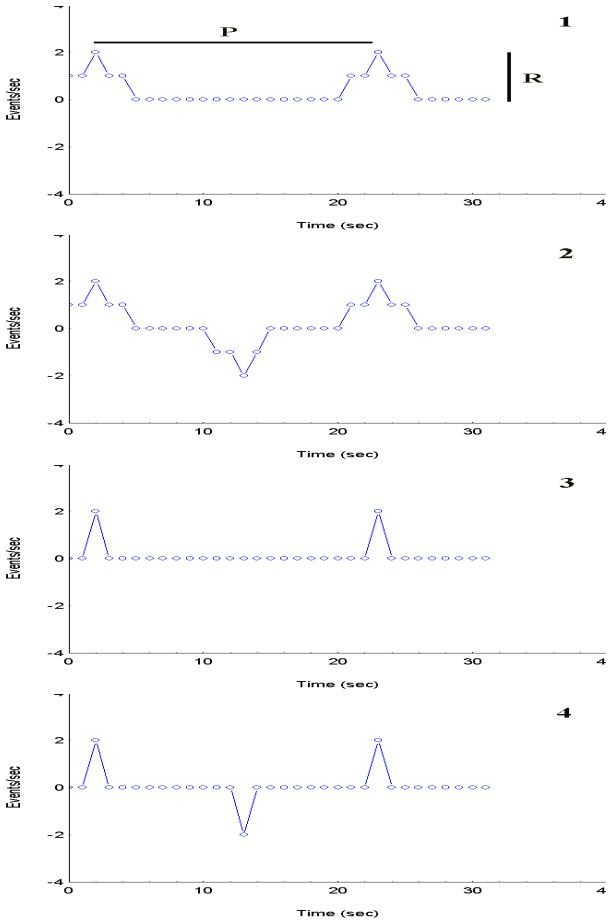


Figure 5: Four types of deterministic component of event rate added to the beta-decay stochastic series.

4.2 Semi-regular superposition of building blocks

When the building blocks are superimposed at semi-regular period the event rate plot still does not show any apparent regular properties. A result of simulation for NB=8, NP=238, C_o=5.0 and k=0.615 is shown in Fig. 11. Due to a low k the building blocks are slowly-decaying.

The period of occurrence is here rather stable, with the maximum deviation of 20% (40/200). The corresponding time scale spectrum is shown in Fig. 12. The spectrum is now dominated by a vertically elongated structure at 200 seconds. The harmonics of the 200 second component is also visible.

Another example with a larger number of building blocks (NB=17) of approximately same size (NP=238) but occurring with relatively large deviations from the average: 57 +/- 20 seconds. The maximum deviation is here 35%. The event rate is shown in Fig. 13.

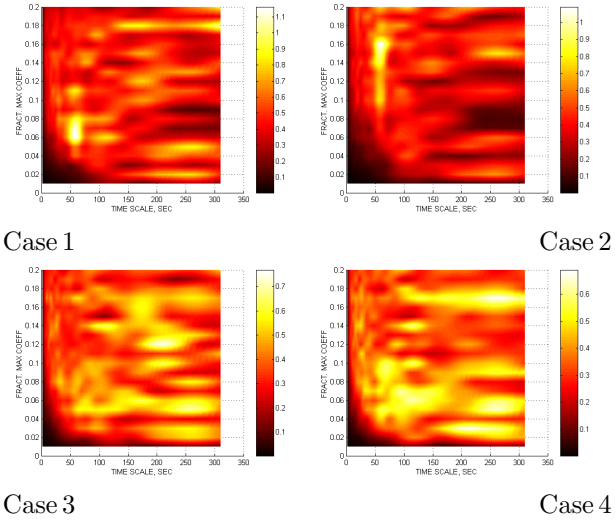


Figure 6: Time scale spectra of ampligrams for a superposition of the beta-decay data and the templates of Fig. 5. for P=57 bins and R=3 events/bin, multi-bin component (broad) in cases 1 and 2; single-bin component (narrow) in cases 3 and 4. The vertical scales in %/100.

The time scale spectrum (Fig. 14.) shows now both vertical and horizontal structures. In addition to a vertical structure at 57 seconds, there is a vertical structure at approximately 170 seconds, corresponding probably to a sub-harmonics of the repetition rate of building blocks.

The above simulations show that it is possible to detect a deterministic component, even in a process consisting of Poisson-like building blocks. Although the event rate graphs in Fig. 9. and Fig. 13. looks similar, the time scale spectrum of Fig. 14. differs significantly from that of Fig. 10.

4.3 A possibility to estimate number of building blocks

There is a simple method to estimate number of building blocks during the observation period from a wavelet scalogram. The scalogram differs from a frequency spectrum showing on vertical axis the time scale instead of frequency. The scalogram is assumed to be the most correct presentation of results of wavelet transform. In the present application a linear scale is used on the vertical axis instead of a logarithmic. The method is as follows: The ampligram obtained from the simulated data of Fig. 11. is used for the principal component analysis. It may be found that there should be 4 sig-

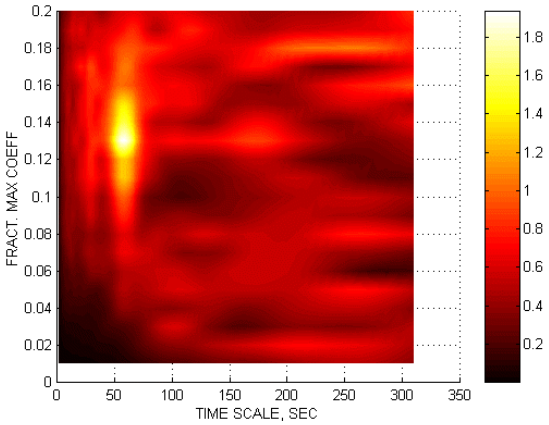


Figure 7: Time scale spectrum for case 1 (a positive component) and R=6. The vertical scale in %/100.

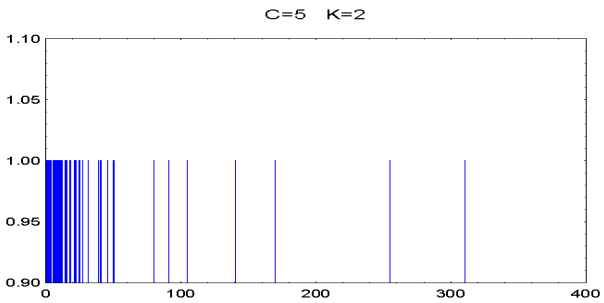


Figure 8: A single building block for NP=200 photons, C_o=5 and k=2. X-axis scale in seconds.

nificant principal components in the pattern of that ampligram (see Fig. 15).

It may be found that the determined principal components PC1 - 4 explain: 17.4, 8.6, 6.7 and 6.3% respectively, of the total variance in the ampligram matrix. However only component #1 and #3 may be efficiently filtered out and inversely transformed. Scalograms may be constructed for these inversely transformed components; see Fig. 16.

A rough estimate of the number of building blocks may be obtained counting intervals corresponding to the dominating scale along the intense trace on the left scalogram. It would correspond to 6 blocks at a repetition period of approximately 200 seconds. The physical meaning of independent components is that the algorithm searches for a set of blocks with approximately same repetition period. Blocks found at different spacings from the main train of blocks will be, in combination with adjacent blocks in the main set, interpreted as different components. The scalogram of Component #3 shows 2 additional blocks with a spacing slightly smaller than that in Component #1 and at least 1 with a spacing less than 100

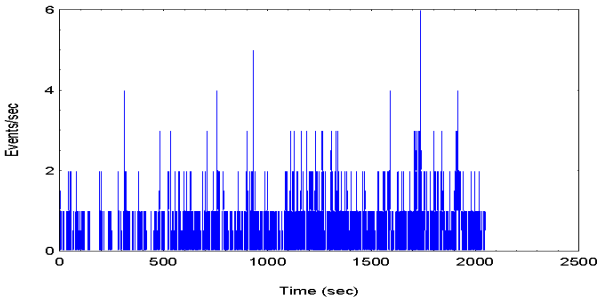


Figure 9: Event rate in the case of randomly occurring blocks. The time scale spectrum of the ampligram constructed for first 1024 seconds of the above simulation is shown in Fig. 10.

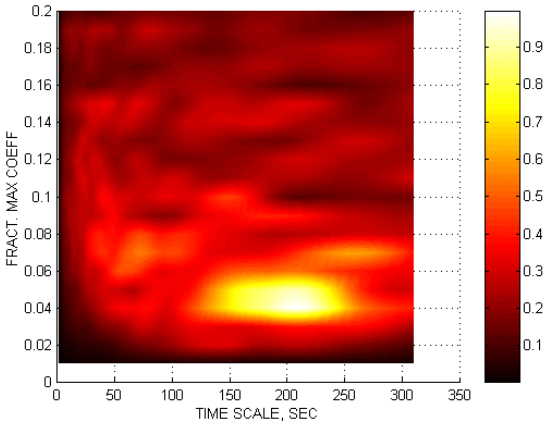


Figure 10: The time scale spectrum of the ampligram constructed for the data of Fig. 9. The vertical scale in %/100.

seconds. That estimate is not far from the simulation parameter of 8 blocks.

5 Regular structures in the beta-decay data

What is the cause of the semi-regular structure observed in the beta-decay data? One possibility is that the structure has an instrumental origin. The other possibility could be that it is a consequence of counting events in bins of finite width. Another question is if it is possible to rearrange events in the event series in such a way that the amplitude of the structure will be minimum. It has been found that the structure corresponding wavelet coefficients below 1% of the maximum value can not be removed. That structure is most likely due to the mentioned counting events in bins of a finite width. However, most of the structure in the beta-decay data, within the interval between 1

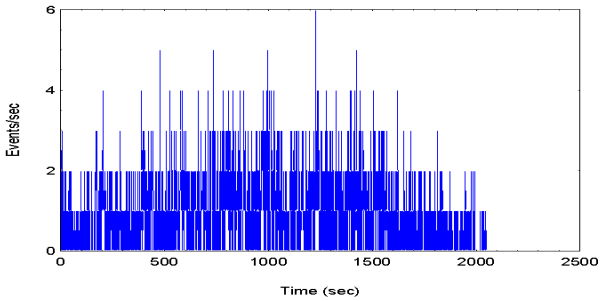


Figure 11: Event rate in the case of 8 building blocks superimposed with a period of 200 +/- 40 seconds.

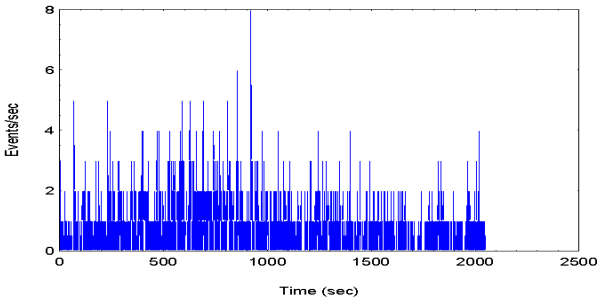


Figure 13: Event rate in the case of 17 building blocks superimposed with a period of 57 +/- 20 seconds.

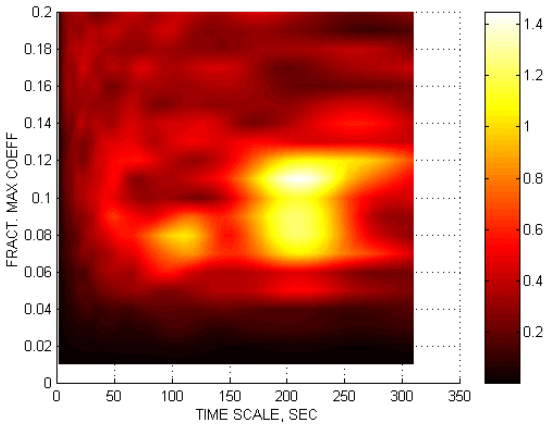


Figure 12: The time scale spectrum of the ampligram constructed for the data of Fig. 11. The vertical scale in %/100.

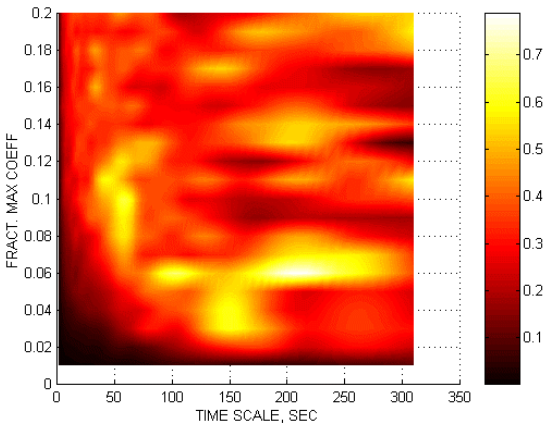


Figure 14: The time scale spectrum of the ampligram constructed for the data of Fig. 13. The vertical scale in %/100.

and 8% of the maximum wavelet coefficient may be suppressed by modifying the event rate data using the following procedure:

1. The original series of event rate is "high-pass filtered" at 8% level according to the procedure described in Part I, Section 3 of this study.

2. The residual component is corrected for offset, since the offset is removed before performing the wavelet transform.

3. The resulting event rates are then truncated to the nearest integer and a new, corrected, event rate series is obtained.

The choice of the filtering level may be tested comparing the kurtosis above and below the selected level. The corrected event rate series may be subject of the same analysis as the original data. The low-20 ampligram for the corrected data is shown in Fig. 17. It may be interesting to compare the sum over the range 1 - 8% of the ampligram for original beta-decay data and for the corrected data. The results for first 300 bins are shown in Fig. 18.

It may be seen that the amplitude of the structure within the 1-8% range is attenuated by the rearranging procedure. It may be interesting to know how much rearrangement is needed in order to obtain that improvement. Total number of events within the 1024-bins interval is 1077 . In the corrected series there is 1076 events. The events has thus been mainly moved between adjacent bins. For illustration, event counts for first 100 bins, both for original and corrected data, are shown in Fig. 19.

Subtracting both time series from each other it is possible to obtain the difference between the observations and a pure Poisson process (corrected data). The difference plot for the same part of the original beta-decay series as in Fig. 19. is shown in Fig. 20. The difference file may be conveniently used for further analysis.

An interesting result of the present exercise is that it seems be possible to remove most of the semi-regular

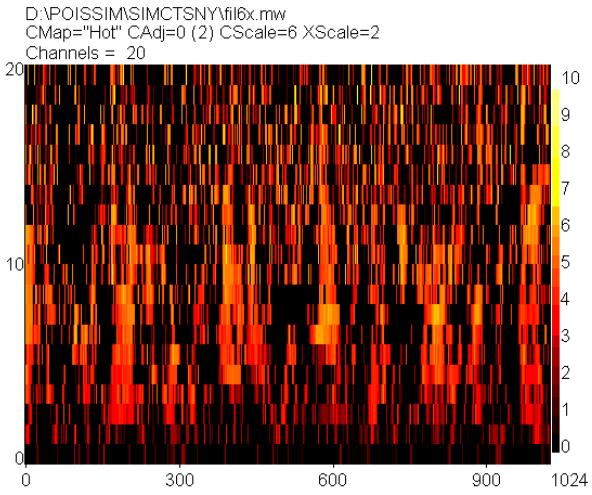
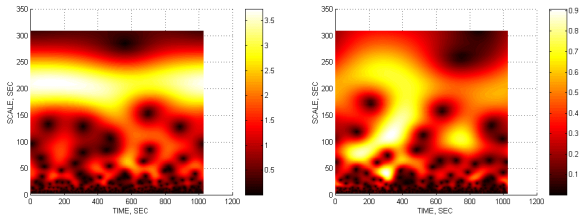


Figure 15: The ampligram of simulated data of Fig. 11.



Component #1 Component #3

Figure 16: Scalograms for components #1 and #3 separated from the ampligram of Fig. 15. using the principal component method. Observe the linear vertical scale.

structure in, what is usually assumed, random data. The present result may be of importance in cases when there is need of good quality random data, like for Monte-Carlo simulations or for encrypting purposes.

6 Separation of deterministic structures in simulated photon data

The above method to separate the deterministic components has been applied to simulated photon data in Section 4. In the case of random superposition of building blocks the difference between the original data and the pure Poisson part is very small. The wavelet spectrum of the difference time series is practically empty, see Fig. 21.

When the procedure is repeated for the simulated

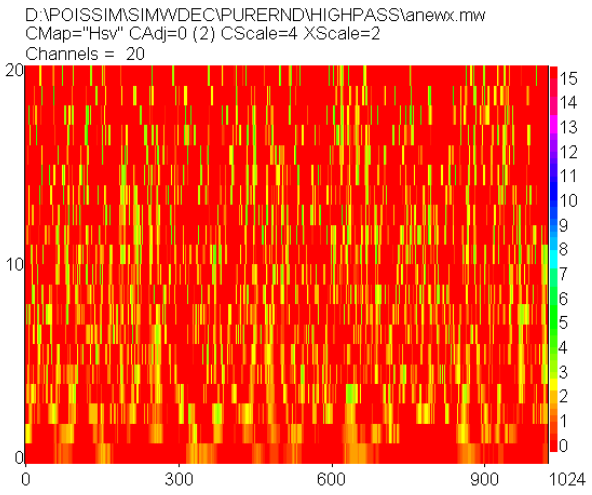


Figure 17: A low-20 ampligram of the same sample as shown in Fig. 3. but after rearranging the events in the series. The vertical scale corresponds to the percentage of the maximum wavelet coefficient, the horizontal scale shows the bin number.

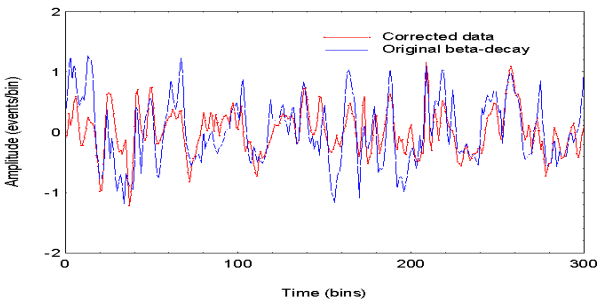


Figure 18: A comparison between the sums of ampligram between 1 and 8% for original data (blue line) and for the corrected data (red line)

data in Fig. 11 and Fig. 13 there is a significant amount of structures in the wavelet spectrum, see Fig. 22.

The results of above experiment demonstrate that a random superposition of Poisson-like building blocks there can not produce significant deterministic components. However, when building blocks are superimposed in a semi-regular manner, the resulting event rate has a significant deterministic component. That is another indirect proof of existence of deterministic components in the X-ray photon data. When the above procedure is applied to measured photon data, most frequently a significant deterministic component is obtained. The procedure seems to work on the data with an average event rate larger than 0.1 event per time bin. It means that it can be applied to a large number of X-ray sources observed by ROSAT even for as short time bins as 1 second. As an exam-

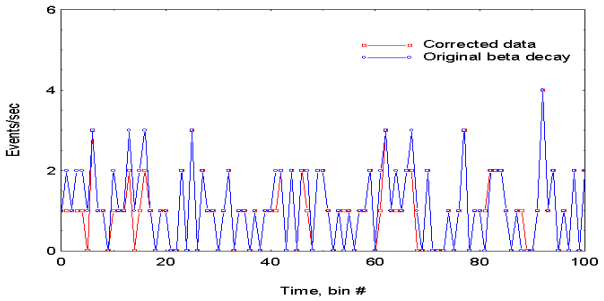


Figure 19: Number of counts during first 100 bins for the original beta-decay series (blue points) and for the corrected one (red points).

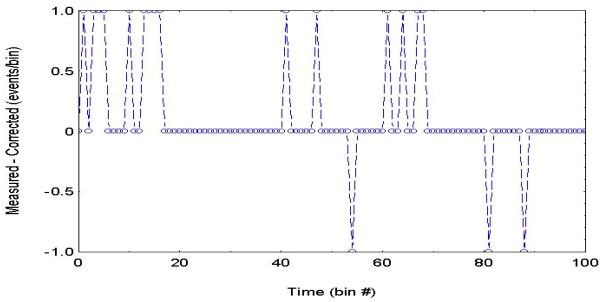


Figure 20: The difference between measured and corrected counts shown in Fig. 19.

ple a wavelet spectrum of the difference between the measured data and the pure Poisson component for NGC 5548, ROR 701246 is shown in Fig. 23.

If the observed photon flux would be a result of a random superposition of stochastic events, the above procedure would produce an empty spectrum, like that in Fig. 21.

7 Conclusions

The present study indicates that the method utilizing the ampligram on the time scale spectrum is capable to detect weak deterministic components in Poisson-like event data. Further on, for average event rates larger than 0.1 event per time bin it is possible to isolate the deterministic component and to calculate its wavelet spectrum. Another interesting consequence of the study is that weak deterministic components in naturally observed random event series, like beta-decay data may be removed. That may be a practical method to obtain very pure series of random events. Simulations of photon fluxes constituted of superimposed decaying Poisson-like elementary building blocks indicate that it is possible to distinguish between random superposition and semi-

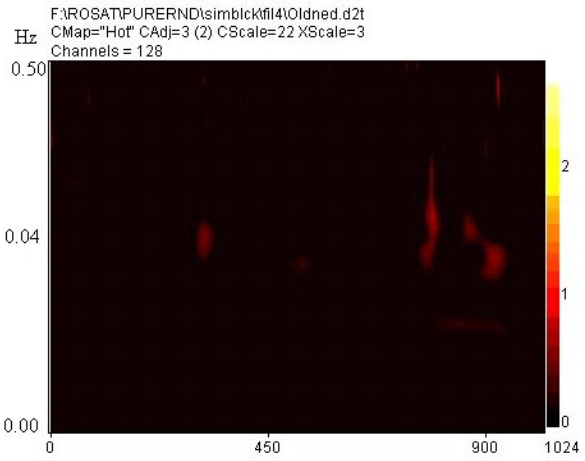


Figure 21: The wavelet spectrum of the difference between the original simulated data and the pure Poisson part of the data shown in Fig. 9. (random superposition of building blocks). The vertical scale is frequency in Hz.

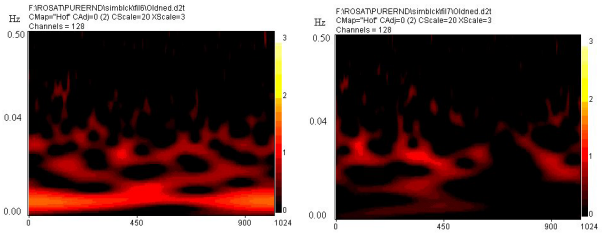


Figure 22: The wavelet spectrum of the difference between the original simulated data and the pure Poisson part of that data shown in Fig. 11 and Fig. 13. (semi-regular superposition of building blocks).

regular superposition of blocks.

8 Acknowledgements

The author is indebted to A.W. Wernik of Space Research Center, Warsaw, Poland, for support in the area of non-linear filtering using wavelets and to A. G. Pacholczyk and W. Stoeger of Steward Observatory, University of Arizona for continuing discussions in the course of this work and for support in matters concerning the AGN:s. The software used in this work has been developed by P-O. Nilsson, J. Karlsson and F. Rutqvist of the Umeå Division of the Swedish Institute of Space Physics. The random events data (noise diode generator and radioactive decay data) used to demonstrate the present analysis technique were kindly supplied by T. Lindblad of the Royal Institute of Technology, Stockholm.

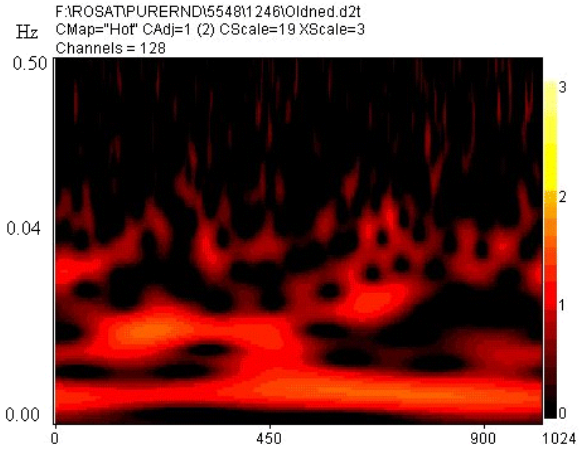


Figure 23: an example a wavelet spectrum of the difference between the measured data and the pure Poisson component for NGC 5548 (ROR 701246). The vertical scale shows the dilation number.

`http://www.irf.se/ume/pdf/236rep.pdf`

9 References

References

[1997] Liszka, L., 1997, A Study of Short-Term Temporal Variations of Photon Counts Recorded by the ROSAT X-ray Satellite, IRF Report No. 239 available at `http://www.irf.se/ume/pdf/239rep.pdf`.

[1999] Liszka L., Holmström M., 1999, A&AS 140, 125

[1998] Pacholczyk, A. G. and W. R. Stoeger, 1994, Active galactic nuclei. V. X-ray variability and the black hole cluster paradigm. Astrophysical Journal, 434, p. 435.