Nonlinear analysis techniques that were originally developed to measure the structure of galaxies in the universe are now improving doctors' ability to diagnose skin cancer, tumours and detect serious heart problems

New designs on complex patterns

Greg Morfill and Wolfram Bunk, Max Planck Institut für Extraterrestrische Physik, Germany

In most areas of everyday life we are faced with growing levels of information. When we have to make decisions we are often confronted by huge amounts of data that need to be understood. Usually we have to reduce the data to a few characteristics in order to understand those aspects that have the greatest influence. It is then easier for us to make a decision based on these "reduced measures".

At the same time, man-made systems are becoming more and more complex, particularly those in finance, economics and technology. And we are only just beginning to learn of the complexities and interdependencies in natural systems throughout physics, biology and ecology.

The message is clear. In reducing many different measurements obtained from such systems to a few "characteristics", we must not cut across the inherent interdependencies. If we do so, we will not be able to monitor some of the essential aspects of the system, let alone understand them.

This is easily demonstrated with the well known "double pendulum", in which the second pendulum is suspended from the first. In the so-called linear regime, where the amplitudes of oscillation are very small, the system can be described approximately by the superposition of two independently moving pendulums. However, as soon as the amplitudes are "finite" – which is practically always the case – this simple description fails. Thus, reducing a coupled system to two independent pendulums, we clearly leave out some of the essential physics – a fact that becomes very clear from the measurements of the motion of a double pendulum.

Therefore as the complexity increases, we need a more sophisticated approach to reduce the system to its essential components, and to then derive the corresponding characteristics from measurements. The situation becomes more difficult if we cannot directly measure the system, and instead have to infer its components and interactions indirectly. In such situations – which are the rule, rather than the exception – a new data-analysis strategy is needed. Such a strategy must go far beyond the classical techniques and be able to characterize complex nonlinear behaviour.

About ten years ago at the Max Planck Institute for Extraterrestial Physics in Garching, near Munich we began developing nonlinear analysis methods to quantify the irregular and highly structured distribution of galaxies in the universe. The analysis technique allows us to differentiate between different theoretical models and to compare these with measurements of large-scale galactic surveys. More recently, we have applied the same techniques to a variety of industrial and medical applications, and in particular to the early detection of skin cancer and risk assessment of heart attacks.

Reductionism and fractals

One of the most widely used classical reduction techniques is Fourier analysis. In such an analysis, a series of measurements that has been made at regular time intervals is broken down into its different frequency components. Often the dominant spectral features are then used as indicators or "measures" of the system. As with all reduction techniques, Fourier analysis clearly involves removing some of the information – a fact that may be fatal, as is shown in figure 1.

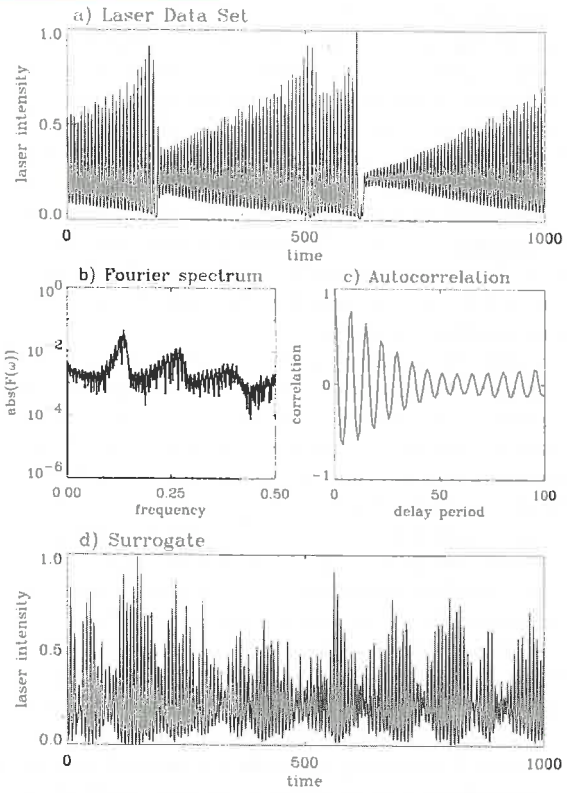


Fig. 1: Limitations of linear analysis
Example of the limitations of linear analysis techniques. The top panel (a) shows the time dependent variation of a laser intensity signal, which was analysed by Fourier (b) and autocorrelation techniques (c). The original data was then modified in such way that the linear properties remain, but the phase information was randomised. This "surrogate" data set is shown in the next panel (d). The Fourier spectrum and autocorrelation function that are derived from this signal are identical to those from the original data. This effect that two very different data sets may contain the same linear information highlights the limitations of linear analyses.

The problem is that Fourier analysis – together with power-spectral analysis, and wavelet and autocorrelation analyses – is a linear technique. As a result, it cannot describe the nonlinear aspects of the measurements obtained from complex systems. New techniques and methods have therefore been developed in recent years to overcome this shortfall, including neural networks and so-called nearest-neighbour, scaling-index and scaling-vector methods. In principle, they can provide "measures" for certain aspects of the nonlinearities contained in the measurements, thus allowing additional insights into the system under study.

The powerful nonlinear techniques we developed are based on the so-called scaling-index-method (SIM). Using this technique we can think of the "state" of a given system as a single point localized in an appropriate "state space". This state space is made up of the different "state variables". For example, each pixel in a static black and white image has three state variables – the x and y co-ordinates, and the greyscale – and hence a three-dimensional state space. Meanwhile a colour image has five state variables – x, y and the red, green and blue colour components. And a tomographic image of a solid body has four state variables – x, y, z and a greyscale.

A dynamic system may have many such variables that change with time. In this case, monitoring the trajectory of a system in state space can be an important tool, for example, in assessing how likely a person is to die from a sudden heart attack

In general the various states in a complex system do not occupy state space uniformly. The characteristics of the system usually mean that it will favour some regions in state space over others. In a colour image, for example, a local region with the same colour will lead to a sheet-like structure in state space, while a line will remain a line. Hence the complexity of a system may be measured by the structures found in the state space. The question then is how to quantify structure. The answer comes from "fractals" – geometric shapes that look similar at different scales. Many structures in nature can be quantified by their fractal geometry, including coastlines, mountains and clouds.

The scaling-index-method makes use of fractals to characterize the system. To calculate the local "fractal dimension" or the "scaling index", $a(x)$, we first draw a sphere of radius r around a given state at position x in state space. Next, we count the number of other states, $N(x,r)$, inside this sphere, and repeat the process for different values of r. The value of the scaling index, $a(x)$, is then extracted from a graph of $N(x,r)$ versus r using the scaling relation $N(x,r) \sim r^{a(x)}$. We assume that this scaling law holds in the range of r chosen, which is usually the case. (For most applications, it is not a problem if the scaling law does not hold exactly.)

We are able to determine the type of structure from the measured value of $a(x)$. For example, a point-like structure has $a \sim 0$; a line-like structure has $a \sim 1$; and a flat sheet has $a \sim 2$. Meanwhile curved lines or sheets have $1 \leq a \leq 2$ and $2 \leq a \leq 3$, respectively.

In addition, if the scaling law holds as r approaches zero, then we can regard the scaling index $a(x)$ as a local property of the system at position x. We can thus define a nonlinear property for each observed system state.

More importantly, however, is the fact that $a(x)$ is directly proportional to the gain in information we obtain by measuring this particular system state. For example, in a static system all the measured states occur at the same location in state space and $a(x) = 0$. A new measurement does not give us any new information about the system. On the other hand, in an evolving system each new observed state occupies a new location in state space and hence yields fresh information about the system.

Thus the scaling index is a very fundamental and independent property of each observed state because it retains information about the nonlinearity and complexity of the system. This makes it a prime candidate for the quantitative analysis of complex systems.

For specific applications there are a number of additional considerations, including the finite resolution, the scaling lengths and different sampling rates for different measurements.
(The Max Planck Society has patented the scaling-index-method and other associated analysis techniques.)

Diagnosing skin cancer

The occurrence of skin cancer has more than doubled in the last ten years. Minor surgery is almost always successful if the cancer is discovered early enough. However, the disease is invariably fatal if it is discovered too late. Reliable early detection is therefore a key issue. This is a formidable task as it involves detecting minute pigment changes on the skin and then differentiating between malignant and benign marks.

Currently dermatologists investigate pigment changes using magnified images of skin lesions. These images allow medical practitioners to identify sub-structures in the image with reasonable resolution. Experts can detect up to about 80% of early skin cancer signatures based on the structure and colour of a lesion. Less experienced practitioners, however, are not so successful.

Dermatologists have established four semi-quantitative risk indicators that can be used to analyse skin lesions based on the visual inspection of such images. These include the asymmetry of the skin lesion, the structure of its border, the colour variation across it and the differential structure inside it. These characteris-

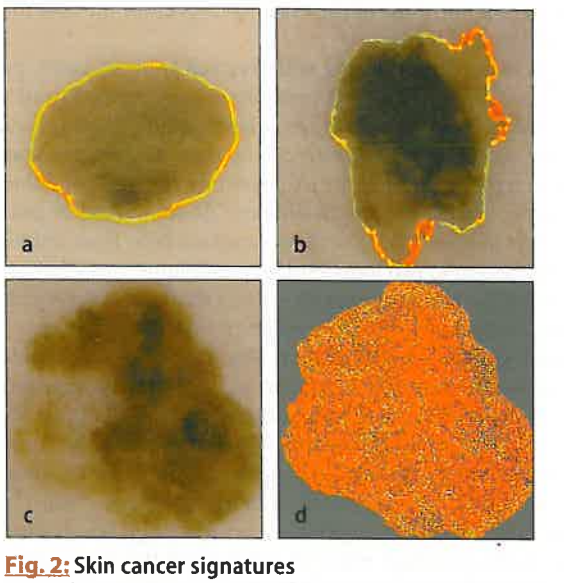


Fig. 2: Skin cancer signatures
Magnified images of benign (a) and malignant (b, c/d) skin lesions. The original size is ~ 11 mm x 11 mm. The border (a,b) is identified using the scaling index method (SIM) described in the text. It is colour-coded from red to blue, where "red" implies a diffuse (fuzzy) boundary (high a) and "blue" a sharp boundary (low a). It can be seen that the malignant skin lesion exhibits a large, connected fuzzy boundary region – a signature which is comparatively rare in benign lesions. Another example for the use of the structural "measure" to identify skin cancer is shown in the lower panels. The scaling index method enables a structural decomposition of the lesion. The colour-coded scaling indices (d) indicate the content of different structural elements of the skin lesion (c) and provide a powerful tool for diagnostic purposes.

tics are known as the "dermatoscopic ABCD rule". In co-operation with expert dermatologists, we have quantified the complex patterns in the images using the scaling-index-method. In these digital colour images, the state variables are the x and y position of each pixel, and the amount of red, green and blue it contains. The resulting structural measures can then be used to identify skin cancer at an early stage. For example, the border between pigmented and normal skin tends to be much more diffuse in harmful malignant skin lesions than in safe benign ones (see figure 2a and 2b) and characterising the structural details of the colour variations is a powerful tool for diagnostic purposes (see figure 2 c and d).

We have tested the scaling-index-method on over 740 images of malignant and benign skin lesions that were acquired by the dermatology centre at the University of Regensburg in Germany (figure 3). Each image was taken using a CCD camera combined with magnification optics and under the same lighting conditions. This assured that the sample was homogeneous from the point-of-view of data acquisition. Our method has improved the detection efficiency to over 90%, thus providing powerful "on site" diagnostic support to medical practitioners.

Such fast and accurate diagnostic support will be invaluable in screening tests or routine inspections. And the need for such support is likely to grow in the future as the trend towards "telemedicine" increases. Digital images of moles and freckles could be evaluated on site, while data from borderline candidates could be sent to an expert dermatology centre for further scrutiny and advice. This centre could double as a data archive that could be used for epidemiological studies and to develop new analysis techniques.

In this particular case, our scientific knowledge has been already transferred to industry and from there into everyday medical use. Since the middle of last year, Rodenstock Precisions has been marketing the digital imaging system, including the specialized software for the early detection of skin cancer.

Tumour diagnostics

In the previous example, we illustrated the power of general pattern-analysis techniques in identifying structures in colour images. More recently, we have extended the technique to tumour diagnostics. The aim is to provide doctors with better and faster methods to determine the volume and precise shape of tumours so that they can differentiate between live and dead tissue within a tumour, and can monitor how the tumour changes during treatment. Such information is invaluable, as it helps surgeons to plan operations according to a particular patient's needs. It can also be used, for example, to monitor the success of chemotherapy or radiation treatment, or to fine-tune the dose of drugs needed.

Currently experts base their decisions on 3-D data from X-ray or nuclear magnetic resonance tomography. However, it takes a trained expert over an hour to analyse these data by hand, which is far too long for routine use.

Our research aims to reduce this to a few minutes, and to provide experts with even more information about the tumour. First, we project the same 3-D tomography data – the x, y and z position of each pixel in the image, together with its greyscale – into a four-dimensional state space. We then perform the fractal analysis to identify the irregular border between the tumour and healthy tissue in 3-D, and to identify the substructure within the tumour.

We first tested our analysis method using computer tomography images of a cow's liver into which known samples of tissue had been implanted, and found that we could determine the volume of the implant to within 1%–3%. Since then, we have optimized the analysis strategy and techniques for all kinds of

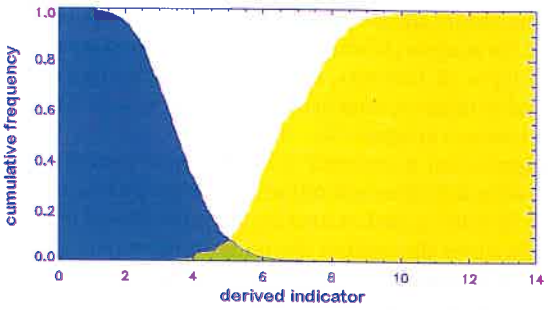


Fig. 3: Detection efficiency
Normalised distributions of benign (blue) and malignant (yellow) skin lesions plotted as a function of the quantitative "measure" (score) derived from the dermatoscopic ABCD-rule described in the text. The score is obtained using fractal dimensions (for the border B – as shown in figure 2a and 2b), entropy measures (for the colour variability – C) and local scaling properties (for an asymmetry feature – A, for the border – B, and for the differential structure – D, see figure 2c and 2d). The high level of discrimination between malignant and benign skin lesions is reproducible and has been tested on a sample of 749 skin lesions. The achieved overall accuracy is more than 90%.

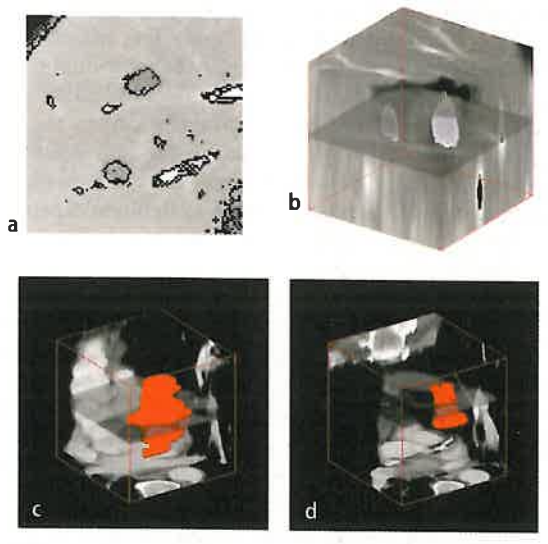


Fig. 4: Tumour diagnostics
The method is used to measure the volume of tumorous tissue. The upper left panel (a) shows one CT slice, with the identification of the boundary of two tumours (lines around the darker segments). The right panel (b) shows the 3D-reconstruction. The volume of the larger tumour was determined at 0.75 ml, the reconstruction "by hand" gave 0,77 ml – in good agreement with the computerised determination. Tumour volumetric analysis of the image data yields a very good method for a (noninvasive) quantitative monitoring of chemotherapeutic treatment. The 3D volume rendering of a segmented stomach cancer before (left) and after (right) chemotherapeutical treatment is shown (figure c and d). In this case it was found that the tumour volume has been reduced by 80 %.

tomography. This means that our software can, in principle, accompany future advances in technology and image resolution. However, the analysis technique has to be optimized separately for different types of tumours, for example in the liver or lungs, because of its inherent sensitivity to tissue structures. That this is possible is shown in figures 4a – d.

This application is currently still at an experimental stage. We need to perform further calibrations before we can be confident of routinely identifying and measuring the properties of tumours in various locations throughout the body, to determine growth or size reductions reliably and thus to contribute to a quantitative therapy. The first results, however, are very encouraging (figures 4c and d).

Dynamic pattern recognition

So far we have shown two different examples of image analysis for recognizing and quantifying complex, but stationary, patterns. The techniques can also be used to identify dynamical patterns in industrial systems or in the body's "biosignals", such as heart rate, brain waves, neural signals and variations in blood pressure.

These dynamical patterns need not be regular. For example, Morse code is based on a predetermined pattern of long and short signals that represent letters in the alphabet. A long chain of seemingly irregularly arranged "dots" and "dashes" can comprise a message, in the same way that a number of "symbols" can constitute a written text. In an analogous way, we believe that the dynamics of the biosignals provides information about the condition of a patient, ranging from normal states – such as when the patient is exercising or sleeping – to abnormal states, such as illness.

Our long-term aim is to decipher these "bio-messages" with sufficient accuracy so that we can identify any impending disorders before they become too serious. In the ideal situation, we would be able to "read" the information provided by the nerve traffic generated by the body's own sensors and the central nervous system. However, this is way beyond current technology in terms of both measurement and pattern recognition capabilities. We therefore have to restrict ourselves to what is possible and, preferably, use non-invasive or surface measurements. The available data fall basically into three categories – those that have a natural rhythm, those that are sampled continuously at a fixed rate, and those that occur in irregular bursts.

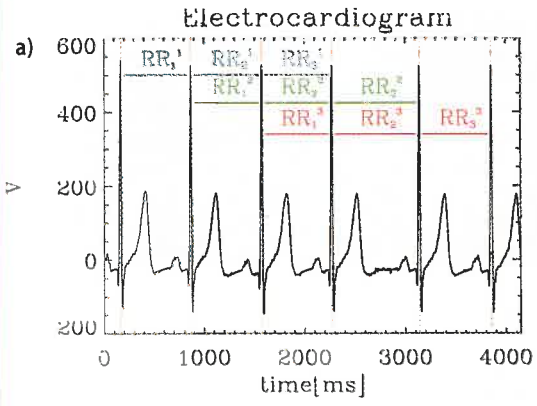
Risk assessment

Systems such as the heart have a natural rhythm or pacemaker. The heart pumps blood by contracting in response to electrical signals transferred between neighbouring cells in the cardiac muscle. This signal-transfer process, which involves a number of other electrophysiological processes, provides a complex signature of the "state" of the heart muscle. In this case, it makes sense to determine the variability in this rhythm very accurately, and to use the time between successive heartbeats as a measure for analysing the system. In principle, these timing data contain information about the pacemaker and signal transfer in the cardiac tissue.

Cardiologists can measure the electric signals using electrodes placed on the surface of the chest. They can identify various heart diseases from the resulting electrocardiogram (ECG) once the disease has developed, but not before. This is particularly true in the case of myocardial infarction, in which the blood supply to the heart is impeded by a blocked coronary artery. In addition, one of the major causes of death in the Western World is "sudden cardiac death", in which victims die suddenly from a seemingly stable situation. In both cases we assume that the heart has certain characteristics that can be detected sufficiently far in advance to allow preventive treatment.

If the processes involved are as complex as they appear, we have to use nonlinear analysis techniques to characterize the heart. In principle, one needs to typify the rich morphology of the ECG signals as accurately as possible, together with the levels of oxygen and carbon dioxide in a patient's blood and also the blood pressure. However, it is usually impractical to measure all of the quantities, and it may even be unnecessary for some purposes.

In practice it is sufficient to use the time between consecutive



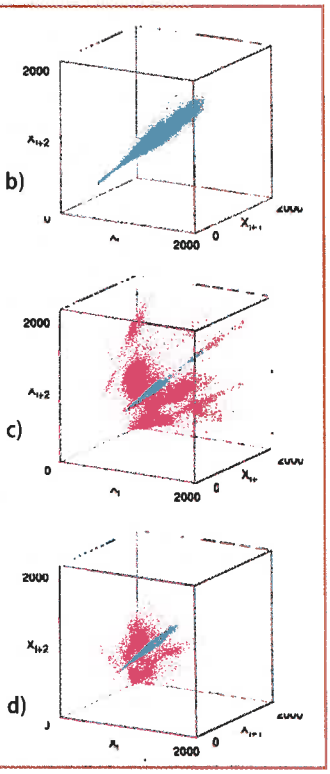


Fig. 5: Dynamic pattern recognition
a) The construction of an artificial 3-dimensional state-space representation of a single time series ECG measurement. One point in the state-space is given by three coordinates (xi = RR1i – the time between two heart beats, yi = RR2i – the time between the next two heart beats and zi = RR3i – the time between the next pair).
b) The state-space representation of a healthy heart containing 100000 measurements from an ECG taken over a period of 24 hours. The rapid heartbeats at the bottom-left were measured while the patient was jogging. The slow heartbeats at the top-right were measured as the patient slept. The homogeneity in structure suggests that we are dealing with a system that can be driven dynamically from one state to the next without major disruptions.
c) The state-space picture of the heartbeats of a patient who suffers from heart disease. The well-defined structure of the arrhythmias (red) appears to scale according to the healthy heart rate. This seems to be an inherently stable situation.
d) The state-space representation of the heartbeats of another patient with heart disease. The arrhythmias are not characterised by well-defined patterns and are much more irregular in structure than those of the previous patient (5c). A patient showing these signs is at risk from sudden cardiac death.

heartbeats. This is both the simplest measurement to make and the most accurate. Even this simple measurement contains a great deal of information. Its variability reflects the interplay between the central nervous system – which receives information from the various sensors in the body, processes them and then translates them into the optimum heart action – and the ability of the heart to comply. A rapid response indicates that the heart is healthy, while an impaired response signifies an unhealthy heart.

The most valuable data for identifying normal and abnormal behaviour patterns are from ECG measured over 24 hours. In this period, the heart experiences many different physical and psychological situations. In other words, the system has sampled a great many "states".

Our reduced analysis strategy involves investigating the pattern between heartbeats, and adding further information as and when required. We have found that by analysing the time interval between four consecutive heartbeats ("triplets"), we can accumulate a great deal of information that has powerful diagnostic and predictive value. We plot the ECG data in a 3-D state space, where each axis represents one of the consecutive beats. RR1, the time between the first two heartbeats, is plotted on the x axis; RR2, the time between the next two, is plotted on the y axis; and RR3, the time between the following pair, is shown on the z axis (see figure 5a).

If the heartbeat is perfectly regular over the time interval under consideration, the system will remain in an identical state and thus occupy exactly the same point in state-space. Over a period of 24 hours, however, the heart rate will change as a patient exercises or rests, for example, and the data will fill a volume in state space. Our analysis therefore concentrates on the occupation density and dynamics in state space (see figure 5b - d).

The characteristic state-space structure for a normal healthy heart is a club-shape structure that is centred along the diagonal. This structure shows that the rapid heartbeats join smoothly with the slow regular ones. This homogeneity in the structure suggests that we are dealing with a system that can be driven dynamically from one state to the next without major disruptions (figure 5b).

However, the state space structure looks very different for a patient suffering from a condition known as an arrhythmia, in which the electrical waves that drives the heart are impaired. The irregularity in the heartbeat is revealed clearly because these states occupy non-diagonal regions of state space. Surprisingly, we found that the arrhythmias appear to scale according to the healthy heart rate. In other words, the arrhythmias occur faster as the heart beats faster, thus creating the coherent off-diagonal structures.

Such structures signify that the electrical signals in the heart always travel along similar paths – for example, round a section of dead tissue in the heart wall – irrespective of the heart rate. This would appear to be an inherently stable situation, and indeed the patient was still well and in good shape 10 years after this particular ECG was taken.

The analysis can also be used to identify patients at serious risk from sudden cardiac death. In this case, the state space appears to have much less structure (figure 5d). And while the arrhythmias are identifiable, they do not scale with heart rate in the same way that they did in the previous patient. This suggests that the heartbeat irregularities do not correspond to constant, and hence stable, signal paths within the cardiac tissue. This patient died of a sudden cardiac arrest a few weeks after the ECG was taken.

Individual examples, such as those in figure 5, only illustrate the possibilities for new analysis techniques. In order to verify their usefulness in a clinical setting, we need to calibrate the technique using "training samples" – patients whose ECGs have been evaluated independently by experienced cardiologists. We then need to test the algorithms in blind studies where we do not know the clinical history of the patient.

Future outlook

The modern analysis methods that we have developed can identify and characterize complex spatial or temporal patterns in images or time-dependent signals. The techniques have a huge diagnostic potential in many applications, ranging from basic research through to engineering and medicine.

In the case of sudden cardiac death our preliminary studies have already shown encouraging results, improving the risk assessment by a factor of two. Such an improvement could potentially have a massive economic impact. It would ensure that only those patients at serious risk receive drug treatment or a defibrillator to regulate the electrical signals in the heart. Indeed, Karl Lauterbach, a health economist at the university of Cologne, has shown that even 20% improvement in risk assessment could save up to DM 3 billion a year in Germany alone.

We have also taken the first steps towards forming a series of industrial partnerships. BASF Pharma, for example, will use our methods to evaluate the effects that new drugs have on ECG measurements in pre-clinical drug trials. For a technique that was originally developed for basic research, nonlinear analysis has come a long way and has a bright future ahead.

Further reading

H Atmanspacher, H Scheingraber, and G Wiedenmann 1989 Determination of f(a) for a limited random point set Phys. Rev. A40 3954

A Horsch et al. 1997 Improving early recognition of malignant melanomas by digital image analysis in dermatology Medical Informatics Europe MIE 97, Sithonia, Greece, ed. C Pappas et al. IOS Press Amsterdam.

C Räth and G E Morfill 1997 Texture detection and texture discrimination with anisotropic scaling indices J. Opt. Soc. Am. A12 14.

G Schmidt and G E Morfill 1994 Complexity diagnostics in cardiology: methods PACE Vol. 17, 2336.

G Schmidt et al. 1996 Variability of ventricular premature complexes and mortality risk. PACE 19, 976.

Acknowledgement

The authors would like to thank R. Pompl, C. Räth, G. Wiedenmann, V. Demmel and R. Sachs; W. Stolz (University of Regensburg); and G. Schmidt and P. Gerhardt (Technical University Munich).

About the authors

Gregor Morfill is director of the Max Planck Institut für Extraterrestrische Physik, and since 1990 a professor at the Ludwig-Maximilians-University (LMU). His main research areas range from cosmic rays, astrophysics, plasma physics to complex system analysis. Although mainly a theoretician, he has recently taken up experimental research in "complex plasma physics" after the discovery of liquid and crystalline plasma states. Prof. Morfill has received many honours and has been elected Foreign Member of the Russian Academy of Sciences.

Wolfram Bunk studied physics and astrophysics at the LMU, where he obtained his doctoral degree in 1993, working on deconvolution techniques applied to astrophysical systems. In 1993 he joined the theory group of the MPE, and has been at the Centre for Interdisciplinary Plasma Science since 2000. His research is focussed on the development of new methods and measures for the description of complex systems.