



## Discerning non-autonomous dynamics



Philip T. Clemson, Aneta Stefanovska\*

Department of Physics, Lancaster University, Lancaster, LA1 4YB, United Kingdom

### ARTICLE INFO

#### Article history:

Accepted 22 April 2014

Available online 6 May 2014

editor: I. Procaccia

#### Keywords:

Non-autonomous systems

Complex systems

Nonlinear dynamics

Time series analysis

Inverse approach

### ABSTRACT

Structure and function go hand in hand. However, while a complex structure can be relatively safely broken down into the minutest parts, and technology is now delving into nanoscales, the function of complex systems requires a completely different approach. Here the complexity clearly arises from nonlinear interactions, which prevents us from obtaining a realistic description of a system by dissecting it into its structural component parts. At best, the result of such investigations does not substantially add to our understanding or at worst it can even be misleading. Not surprisingly, the dynamics of complex systems, facilitated by increasing computational efficiency, is now readily tackled in the case of measured time series. Moreover, time series can now be collected in practically every branch of science and in any structural scale—from protein dynamics in a living cell to data collected in astrophysics or even via social networks. In searching for deterministic patterns in such data we are limited by the fact that no complex system in the real world is autonomous. Hence, as an alternative to the stochastic approach that is predominantly applied to data from inherently non-autonomous complex systems, theory and methods specifically tailored to non-autonomous systems are needed. Indeed, in the last decade we have faced a huge advance in mathematical methods, including the introduction of pullback attractors, as well as time series methods that cope with the most important characteristic of non-autonomous systems—their time-dependent behaviour. Here we review current methods for the analysis of non-autonomous dynamics including those for extracting properties of interactions and the direction of couplings. We illustrate each method by applying it to three sets of systems typical for chaotic, stochastic and non-autonomous behaviour. For the chaotic class we select the Lorenz system, for the stochastic the noise-forced Duffing system and for the non-autonomous the Poincaré oscillator with quasi-periodic forcing. In this way we not only discuss and review each method, but also present properties which help to clearly distinguish the three classes of systems when analysed in an inverse approach—from measured, or numerically generated data. In particular, this review provides a framework to tackle inverse problems in these areas and clearly distinguish non-autonomous dynamics from chaos or stochasticity.

© 2014 Elsevier B.V. All rights reserved.

### Contents

1. Introduction.....	299
1.1. Motivation.....	299
1.2. Outline.....	300
2. Background.....	300
2.1. Dynamical systems and oscillators.....	300

\* Corresponding author.

E-mail address: [aneta@lancaster.ac.uk](mailto:aneta@lancaster.ac.uk) (A. Stefanovska).

2.2.	Complex systems .....	301
2.2.1.	Thermodynamics perspective.....	302
2.3.	Models of complex systems .....	302
2.3.1.	Lorenz system .....	302
2.3.2.	Duffing system .....	303
2.3.3.	Stochastic systems.....	304
2.3.4.	Non-autonomous systems .....	305
2.4.	Time series .....	307
2.4.1.	Inverse problems .....	308
2.4.2.	Nonstationarity versus non-autonomicity.....	308
2.5.	Simulated systems .....	308
2.5.1.	Parameters .....	309
3.	Frequency domain .....	310
3.1.	Fourier transform.....	310
3.2.	Analysis .....	311
3.3.	Summary .....	313
4.	Time domain.....	314
4.1.	Smoothing.....	314
4.1.1.	Moving average.....	314
4.1.2.	Moving polynomial.....	314
4.1.3.	Window functions .....	315
4.2.	Decomposition .....	316
4.2.1.	Empirical mode decomposition.....	316
4.2.2.	Wavelets.....	317
4.3.	Analysis .....	319
4.4.	Summary .....	320
5.	Time–frequency domain.....	320
5.1.	Instantaneous frequency.....	321
5.1.1.	Marked events .....	321
5.1.2.	Hilbert transform.....	321
5.2.	Short-time Fourier transform .....	322
5.3.	Wigner–Ville transform .....	323
5.4.	Continuous wavelet transform .....	324
5.5.	Analysis .....	326
5.6.	Summary .....	326
6.	Phase space.....	327
6.1.	Embedding .....	327
6.1.1.	Parameter estimation .....	327
6.1.2.	Limitations .....	328
6.2.	Lyapunov exponents.....	328
6.3.	Rényi dimensions.....	330
6.4.	Eigenvector decomposition.....	331
6.5.	Analysis .....	332
6.5.1.	Embedding .....	332
6.5.2.	Lyapunov exponents .....	333
6.5.3.	Correlation dimension.....	333
6.6.	Eigenvector decomposition.....	335
6.7.	Summary .....	335
7.	Complexity analysis.....	336
7.1.	Hurst exponent .....	336
7.2.	Detrended fluctuation analysis.....	336
7.3.	Multifractals.....	337
7.4.	Analysis .....	338
7.5.	Summary .....	338
8.	Interactions.....	339
8.1.	Background .....	339
8.1.1.	Couplings.....	339
8.1.2.	Phases and protophases .....	339
8.1.3.	Synchronization.....	339
8.1.4.	Phase coherence .....	340
8.2.	Phase dynamics methods.....	340
8.2.1.	Synchrogram.....	340
8.2.2.	Synchronization index.....	340
8.2.3.	Direction of coupling.....	341
8.3.	Bayesian inference approach .....	343
8.4.	Time–frequency approach .....	343

8.4.1.	Wavelet phase coherence .....	343
8.4.2.	Wavelet bispectrum .....	344
8.5.	Information theoretic approach.....	346
8.6.	Analysis .....	347
8.6.1.	Phase of complex systems.....	347
8.6.2.	Pseudo-complexity.....	348
8.6.3.	Coupled system examples.....	348
8.6.4.	Phase dynamics methods.....	349
8.6.5.	Wavelet phase coherence .....	351
8.6.6.	Wavelet bispectrum .....	352
8.6.7.	Information theoretic approach .....	354
8.7.	Summary .....	356
9.	Applications.....	357
9.1.	Acoustic turbulence .....	357
9.2.	Electrons on helium.....	357
9.3.	Cardiorespiratory system .....	358
10.	Conclusion .....	359
	Acknowledgements.....	360
	Appendix A. Matlab toolbox.....	360
	Appendix B. Numerical integration.....	360
	B.1. Heun scheme.....	360
	B.2. Lie scheme.....	360
	Appendix C. Derivations .....	361
	C.1. Formulation of the Fourier transform .....	361
	C.2. Least squares polynomial fitting.....	362
	Appendix D. Statistics .....	363
	D.1. Binning .....	363
	D.2. Surrogate data.....	363
	References.....	364

## 1. Introduction

### 1.1. Motivation

In nature there is an unfathomable number of systems that are explicitly dependent on time. Living systems are perhaps the most notable, simply because it is not possible to consider them in any state other than when they are undergoing time-dependent fluctuations. To do so would lead to interpretations of life that are not realistic, with the most extreme cases being when the interpretation corresponds to a ‘dead’ state. Such systems are in fact still capable of self-sustained oscillations but at the same time are under the influence of each other, leading to complex phenomena such as intermittent synchronization [1]. More subtle time-dependent effects are also observed such as the circadian rhythm [2], where natural processes in living systems are seen to fluctuate over an approximately 24 hour cycle. Even on the longest timescales life is still time-dependent, where the changes to the system are known as ageing [3].

Besides living systems, time-dependent dynamics is observed across a wide range of other fields. Climate science deals with time-varying fields, which must be treated as such to obtain the effective number of degrees of freedom [4]. General theories of complex systems and networks also consider individual elements as time-dependent systems, which is relevant to topics ranging from chemical reactions to power grids [5–7]. Since the discovery of an expanding universe, cosmology has also become the study of the largest time-dependent system: the universe itself.

Despite the ubiquity of time-dependent dynamical systems in nature, there has been relatively little work done on the analysis of time series from such systems. Mathematically they are known as *non-autonomous* systems, which are named as such because, unlike autonomous systems, in addition to the points in space over which they are observed they are also influenced by the points in time. Recently there has been much work on the direct ‘bottom-up’ approach to these systems, which includes the introduction of a subclass known as *chronotaxic* systems that are able to model the stable but time-varying frequencies of oscillations in living systems [8,9]. In contrast, the time series analysis of these systems, referred to as the inverse or ‘top-down’ approach, has not been studied in detail before. This is partly because non-autonomous systems can still be analysed in the same way as other types of systems in both the deterministic [10] and the stochastic [11] regime. However, it is now argued that this type of analysis is insufficient and that an entirely new analytical framework is required to provide a more useful picture of such systems. In the case of chronotaxic systems some methods have already been developed for the inverse approach and they have shown to be useful in analysing heart rate variability [12]. A general and dedicated procedure for analysing non-autonomous systems has still not been tackled though.

Due to the prevalence of non-autonomous systems, the approach to their analysis has far-reaching implications with the potential to influence research carried out across a huge and diverse range of fields. For example, this knowledge can be applied to recent research into brain dynamics that aims to map the structural and functional connections of the brain [13,14].

This map or *connectome* is a large step from previous descriptions of brain dynamics that have adopted a statistical approach and random field theory [15]. Instead, the description provided by the connectome is closer to the theory of dynamical systems, with the potential to revolutionize our understanding of the brain and to give insight into neurological disorders.

In addition to application to the brain, methods used to analyse non-autonomous systems have already been successfully applied to the cardiovascular system and resulted in a wealth of new information [16,17,8,18]. New research has also revealed whole host of cellular dynamics [19], including calcium [20,21] and mitochondrial oscillations [22,23]. The activity expected in this field in the coming years is likely to result in many more time series being recorded from systems at the cellular level, which in turn adds to the demand for methods to analyse them.

The application of methods based on time-dependent dynamics also extends beyond living systems. For example, they are invaluable to the investigation of rogue waves, which are transient phenomena that pose a great risk to both rigs and shipping [24,25]. Analysis of time series from non-autonomous systems is also required for many observed astronomical sources such as supernovae and gamma-ray bursts [26]. The review presented in this paper therefore has useful applications to a wide range of research.

## 1.2. Outline

This paper has been written to accomplish a number of aims. First and foremost it provides a detailed review of the various techniques that are available for analysing the time series data of dynamical systems. Matlab codes for most of these methods are provided online in the TACTS toolbox (see Appendix A). Secondly, the applicability of the techniques to several low-dimensional systems with complex dynamics is investigated. The often-overlooked class of non-autonomous systems is included here, for which it is argued that a different framework of time series analysis is needed. Lastly, the knowledge gained from the investigation into the time series analysis of complex systems is applied to real data from a number of scientific fields.

The first sections present an introduction to the ideas and terminology used in the discussion of nonlinear dynamical systems and an explanation of some of the fundamental concepts. After this, the systems to be considered in this paper are introduced, including a discussion of their integration and the choice of parameters.

Following the introduction, the time series analysis techniques are organized into sections corresponding to the *domains* in which they operate. At the end of each of these is an analysis section, where the implications of applying these methods to complex systems are shown. Hence, this provides an evaluation of the suitability of using these methods in inverse problems where the exact nature of the system is unknown.

## 2. Background

### 2.1. Dynamical systems and oscillators

A dynamical system is defined as the mechanism by which a state changes in time. The most basic form is given by

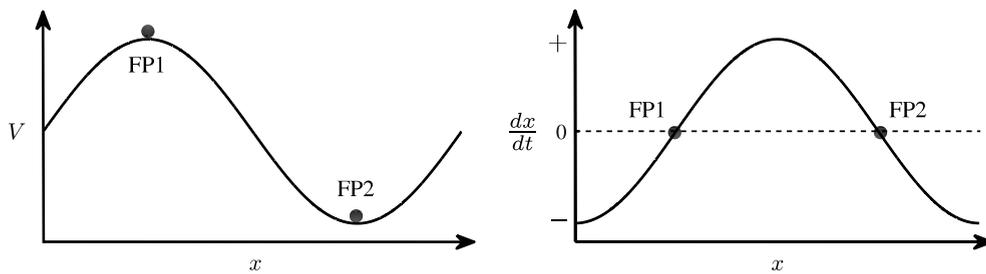
$$\frac{dx}{dt} = f(x), \quad (1)$$

where the infinitesimal change of the state  $x$  in time  $t$  has some dependence on the previous state [27]. There are of course countless trivial cases which satisfy this condition but the most interesting, and indeed the ones most extensively observed in nature, are systems described by nonlinear differential equations (see for instance [28] for biological examples or [5] for examples in chemistry). In this thesis the term “dynamical system” will almost exclusively refer to these types of systems.

An essential property of many dynamical systems in nature is that their states are bounded as  $t \rightarrow \infty$ . Without this property, sequential perturbations to a system would cause it to diverge uncontrollably from the original state. The region of bounded states is known as an *attractor*, since the system will be attracted to this region if its original state is within a neighbourhood. Once the system is within this region the attractor is said to be forward invariant, which means the system will remain on the attractor as it evolves forward in time.

The physics of attractors can be understood more clearly by considering a nonlinear system where the rate of change in some variable  $x$  over time is  $\frac{dx}{dt} = f(x)$ . When  $\frac{dx}{dt} > 0$  the system moves in the positive  $x$ -direction, while if  $\frac{dx}{dt} < 0$  it moves in the negative  $x$ -direction. Values of  $x$  where  $\frac{dx}{dt} = 0$  are defined as *fixed points* and, unless perturbed by an external force, the system does not move once on these points. If for nearby values of  $x$ ,  $\frac{dx}{dt}$  carries the system away from the point, then it is referred to as an *unstable* fixed point. However, if  $\frac{dx}{dt}$  brings the system back towards the point then it is referred to as a *stable* fixed point. This is analogous to a ball positioned either at the top of a hill or at the bottom of a valley and can be illustrated in terms of a potential  $V$ , as shown in Fig. 1.

One would expect the system to contain some inertia so that when it rolled down the potential from FP1 it would overshoot FP2 by some amount before returning to that point. The reason that this is not the case is because the system has no defined *acceleration*,  $\frac{d^2x}{dt^2}$ . Introducing this second order differential term to the original system is equivalent to assigning a new variable  $y = \frac{dx}{dt}$  and creating a 2-dimensional system  $\frac{dy}{dt} = f(x)$ . As a consequence, the system will exhibit more familiar Newtonian dynamics and it will indeed overshoot the fixed points where  $\frac{d^2x}{dt^2} \equiv \frac{dy}{dt} = 0$ . In the case of stable fixed



**Fig. 1.** The gradient of the potential for the system  $\frac{dx}{dt} = f(x)$  can be defined as  $\frac{dy}{dx} = -f(x)$ . Here two fixed points are present, with FP1 lying on a hill in the potential and FP2 lying in a valley. Any small perturbation to the system when at FP2 will bring it back to the original position, which makes it a stable fixed point. However, for FP1 a small perturbation will cause the system to “roll” down one of the slopes of the potential away from the original position, which makes it an unstable fixed point.

points such as FP2, this extra dimension enables the possibility of *oscillations* where the system continues to return to the fixed point in  $V$  only to overshoot each time. If no inertia is lost and the amplitude of the oscillations remains constant over time, this is known as a *closed orbit*.

The ‘position’  $x$  of the system and its ‘velocity’  $y$  describe the specific ‘state’ of the system, so that if there are no external perturbations, the subsequent state for the next time interval can be determined exactly using the corresponding equations for  $\frac{dx}{dt}$  and  $\frac{dy}{dt}$ . Plotting the subsequent states in the  $x - y$  plane, which is referred to as the ‘trajectory’ of the system, produces a loop representing a closed orbit. This is known as *phase space* representation and it will be returned to frequently in the thesis. Special types of closed orbits exist called *limit cycles*, which can be thought of as “fixed oscillations” in the same sense that was used for fixed points. If all nearby trajectories of the system in phase space move towards a closed orbit then it is a *stable limit cycle*. If all nearby trajectories move away from the closed orbit then it is an *unstable limit cycle*. There is also the possibility that the trajectories converge towards the orbit on one side but diverge on the other, in which case the limit cycle is referred to as being *half-stable* [29].

## 2.2. Complex systems

Following the introduction of the various ideas and terminology related to nonlinear dynamical systems, it is now time to turn to the actual focus of this paper.

There is no formal definition of complex systems that encompasses all related theories and areas of research. However, they can be loosely-defined as cases where a system is comprised of a large number of interacting parts. As opposed to the 1 or 2 dimensional systems discussed in the previous section, where the dynamics only occur in time, this definition suggests that the dynamics occur over space as well as time.

In the field of fluid dynamics the concept of turbulence, originally investigated in the work of Stokes and Reynolds [30,31] with later developments by Kolmogorov, Onsager and others [32–35], motivated many models and theories of complex systems. Turbulence is manifested as rapid variations in the pressure and velocity of the fluid flow, with apparently stochastic dynamics. However, the subsequent development of turbulence has led to it being extended to the electrically conducting fluids studied in astrophysics [36,37] and it has also been described using the more general Ginzburg–Landau equation [38,39]. Though it is still largely associated with fluids, turbulence is now used to describe dynamics not only in physics but in chemistry [5], financial markets [40], geophysics [41] and many other areas. In a similar way to complex systems, the original application to high-dimensional spatio-temporal systems has therefore become a small part of a much larger area.

The connection between the notions of complex systems/turbulence and the low-dimensional systems discussed in the previous section is not straightforward. However, an example used by authors such as Eckmann [42] and Haken [43] is able to provide an intuitive explanation. The example is of Taylor–Couette flow, where a liquid is positioned between two coaxial cylinders; one moving, the other fixed. When laser light is shone into the liquid, the reflected light is Doppler-shifted in frequency. This light provides a measure of the various oscillatory modes in the liquid. Initially, when the fluid is moving slowly, there are no oscillations so the frequency spectrum of the light is empty. However, as the speed of rotation is increased, oscillations in the fluid begin to take place. This is analogous to a dynamical system that is initially at rest but then moves onto a limit cycle. As the rotation is increased further, more oscillations appear until eventually the frequency spectrum is filled with oscillating modes and their harmonics. With the onset of turbulence, the discrete modes merge and the spectrum becomes a continuous distribution.

The above example illustrates how the complex dynamics generated by turbulence can be observed in the properties of a low-dimensional data series, despite the fact that it is in reality manifested in a very high-dimensional spatio-temporal system. Such a reduction in the dimension of turbulent systems is possible because the extra degrees of freedom are often not important to the dynamics [39]. Instead, using the theories developed by Cvitanović and others [39,44] they can be represented by a low-dimensional dynamical system. For the remainder of the paper, the term ‘complex systems’ will refer to these types of systems.

### 2.2.1. Thermodynamics perspective

The common empirical feature of complex systems is that when knowledge of the system is incomplete they appear to have unpredictable dynamics. An interesting perspective is that this coincides with the condition where it is more suitable to use statistical methods to understand the properties of the system, rather than to find approximate analytical solutions [45]. It is therefore arguable that any theory of complex systems should have a foundation in statistical physics.

Statistical physics was one of the great scientific accomplishments of the 19th and early 20th century. Through the work of Boltzmann and others [46], the ideas from probability theory were applied to allow many aspects of physics to be described by dimensionless quantities. The most important of these were the applications to thermodynamics, which have been used to explain a huge range of phenomena through the theories of mass and energy exchange between systems.

Despite having ‘dynamics’ in the name, the vast majority of the field of thermodynamics only considers static equilibrium states and quasi-static transitions between them. However, the states of dynamical systems as observed in phase space are analogous to the states of thermodynamic systems: the energy of the system can be thought of as a combination of its inertia (from the time derivatives in each dimension) and its position in a potential such as  $V$  in Fig. 1. This energy can be altered either by internal damping (i.e. friction—the transfer of useable energy into heat) or external forcing, which in turn changes the possible states of the system. For a detailed discussion of the connections between thermodynamics and dynamical systems theory see [47].

The most relevant thermodynamic quantity related to complex systems is *entropy*. This was introduced in the second law of thermodynamics, originally developed in the work by Sadi Carnot in 1824 [48]. This work highlighted the importance of the arrow of time in thermodynamic systems, where entropy increases whenever there is an irreversible change to a different state. In order to maintain the same entropy the system must stay in the same state or make only reversible changes to other states. For systems that are not in equilibrium the arrow of time becomes important and the increase in entropy is unavoidable.

In unperturbed dynamical systems, reversible processes are shown in phase space by closed orbits. Therefore, whenever the trajectory is not a closed orbit there is a continuous increase in entropy. In thermodynamics there are two ways which entropy can be defined—the macroscopic definition is given by

$$dS = \frac{dQ}{T} \quad (2)$$

for a reversible process, where  $dS$  is the change in entropy,  $dQ$  is the change in heat and  $T$  is the temperature of the system. Prigogine extended this definition by dividing internal and external contributions to the entropy [49]:

$$dS = d_e S + d_i S, \quad (3)$$

where  $d_e S$  is the entropy from interactions with external sources and is equivalent to (2), while  $d_i S$  is the entropy due to irreversible changes that occur internally in the system.

The other interpretation of entropy given in the microscopic definition,

$$S = k_B \ln \Omega, \quad (4)$$

where  $k_B$  is Boltzmann’s constant and  $\Omega$  is the number of microstates of the system (i.e. the sum of all possible states in which the system can exist), with bounds determined by the present amount of energy and number of particles. Since the macrostates of the system (e.g. temperature, volume, pressure) are not relevant to the present discussion the microstates will simply be referred to as ‘states’.

When the number of states is small, the probability that the system is in one of those states becomes large and the system appears more ordered. Therefore, in addition to being an indicator that the system is not in equilibrium, entropy can be used as a measure of the disorder. Highly disordered systems with many states have a large entropy while ordered systems with relatively few states have a small entropy.

Entropy was first applied to dynamical systems in the second half of the 20th century. In particular, a generalized version of entropy was provided by Alfréd Rényi in 1961 [50]:

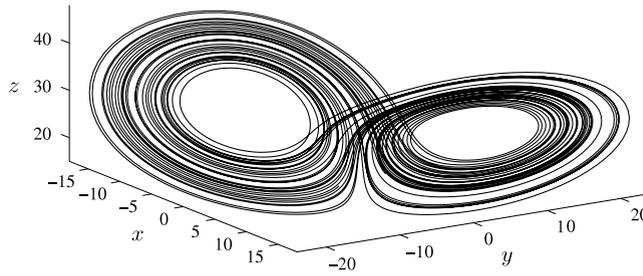
$$H_\alpha = \frac{1}{1-\alpha} \ln \sum_i p_i^\alpha, \quad (5)$$

where  $H_\alpha$  is the Rényi entropy of order  $\alpha$  ( $\alpha \neq 1, \alpha \geq 0$ ) and  $p_i$  are the probabilities for each state of the system. The order  $\alpha$  was introduced to the calculation to control the influence of less likely states on the entropy. As  $\alpha$  is increased, these states have less of an effect on the value of  $H_\alpha$ , while for  $\alpha = 0$  all states are treated equally. This formulation of the entropy is the one most commonly used in dynamical systems and will become important later on when characterizing the various complex systems studied in this paper.

## 2.3. Models of complex systems

### 2.3.1. Lorenz system

Since their discovery, chaotic systems have become one of the most studied class of systems which give rise to complexity. They are set apart from most of the other examples in that they are completely deterministic and influenced only by the



**Fig. 2.** A trajectory of the Lorenz system in phase space, revealing the strange attractor. The parameters used were  $\sigma = 10$ ,  $\beta = 8/3$  and  $\rho = 28$ . The system was integrated using the Heun scheme with  $h = 10^{-4}$  s with the initial conditions  $x_0 = 1.1$ ,  $y_0 = -1.9$  and  $z_0 = 13.1$ .

internal parameters of the system. This means that if the initial conditions are kept the same, a chaotic system will *always* follow the same trajectory. However, the dynamics is complex because the trajectory diverges exponentially if these initial conditions are altered very slightly, making the dynamics unpredictable even with the aid of computer simulation.

Chaotic systems can be defined both discretely (in the case of maps) and continuously where they are defined by a minimum of three differential equations. However, to carry out legitimate time series analysis it is important to have systems where the state can be precisely and accurately defined at any point in time. In the case of maps, the “dynamics” come about through an iterative process rather than an actual dependence on time and, hence, require additional interpolation schemes to approximate the states of the system between each iteration. For this reason, maps are therefore not considered.

The Lorenz system is one of the best-known continuous chaotic systems. It was originally formulated by Edward Lorenz in work relating to the problem of weather forecasting [51]. It is defined by the equations,

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(\rho - z) - y, \\ \frac{dz}{dt} &= xy - \beta z,\end{aligned}\tag{6}$$

where  $t$  is time;  $x$ ,  $y$  and  $z$  are variables that define the dynamics of the system in each of the three dimensions; and  $\sigma$ ,  $\rho$  and  $\beta$  are constant parameters. Despite its application in meteorology the Lorenz system is quite versatile, with another interpretation of its dynamics being that of a leaky waterwheel with water flowing steadily in from the top [29].

The best visual representation of the Lorenz system is given in the 3-dimensional phase space diagram, shown in Fig. 2. A structure known as a *strange attractor* appears after only a few cycles of the system. In the case of the Lorenz system, the strange attractor is a two-dimensional stable manifold and can be thought of as the next order after the 1-dimensional limit cycles discussed in the previous section. In this paper the dimension of this manifold will be referred to simply as the ‘dimension of the attractor’.

The Rössler system is another chaotic system that also features prominently in the literature. However, despite having similar properties to the Lorenz system there is one significant difference in that outside the bounds of the attractor the oscillations spiral off to infinity. This is because the Rössler system is not *globally stable*, meaning not all trajectories converge to the attractor as is the case in the Lorenz system. As a consequence, a problem arises when considering the forced version of the system as it becomes difficult to confine the trajectory within the stable region. The Rössler system is also specialized in that it exhibits a funnel-type attractor, which makes it difficult to define a full cycle of the system [52].

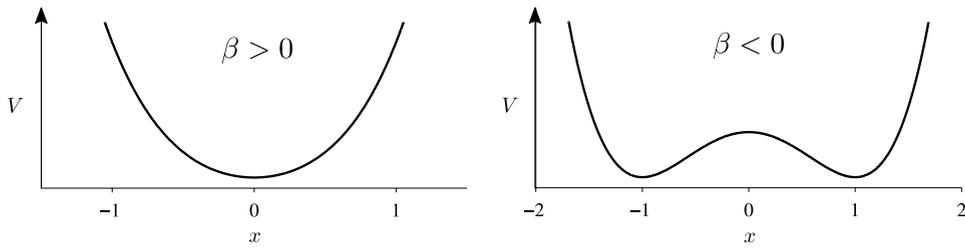
### 2.3.2. Duffing system

The unforced Duffing oscillator is a 2-dimensional system with a cubic nonlinearity. It is defined by the differential equation

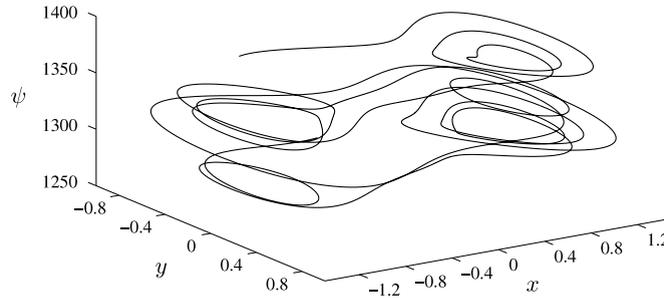
$$\frac{d^2x}{dt^2} + \delta \frac{dx}{dt} + \beta x + \alpha x^3 = 0,\tag{7}$$

where  $\alpha$ ,  $\beta$  and  $\delta$  are constants. The equation was originally investigated by Georg Duffing who was studying the real-world vibrations it could describe [53]. The equation has since been applied and adapted to produce many different types of dynamics, which are comprehensively described in [54].

The constant  $\delta$  defines the damping strength so that when  $\delta = 0$  and  $\alpha > 0$ , the oscillations stay on a limit cycle. The trajectory of the system either orbits a single fixed point located at the origin  $x = \frac{dx}{dt} = 0$  in phase space for  $\beta > 0$ , or orbits two additional fixed points at either side of the origin for  $\beta < 0$ . The potentials in each case are shown in Fig. 3.



**Fig. 3.** The shape of the potential for the unforced Duffing oscillator with  $\alpha = 1$  and different values of  $\beta$ .



**Fig. 4.** A trajectory of the periodically forced Duffing oscillator in phase space for  $\alpha = 1.0$ ,  $\beta = -1.0$ ,  $\gamma = 0.3$ ,  $\delta = 0.2$  and  $\omega = 1$ . The system was integrated with the Heun scheme using  $h = 10^{-4}$  s.

While this system may at first appear simple, its dynamics becomes far more complex by simply adding a periodic forcing term to Eq. (7),

$$\frac{d^2x}{dt^2} + \delta \frac{dx}{dt} + \beta x + \alpha x^3 = \gamma \cos(\omega t), \quad (8)$$

where  $\omega$  is the angular frequency of the periodic forcing term. This essentially adds the new dimension  $\phi$  and  $\psi$  to the system so that it can now be written as

$$\begin{aligned} \frac{dx}{dt} &= y, \\ \frac{dy}{dt} &= \gamma \phi - \delta y - \beta x - \alpha x^3, \\ \frac{d\phi}{dt} &= \psi, \\ \frac{d\psi}{dt} &= -\omega^2 \phi, \end{aligned} \quad (9)$$

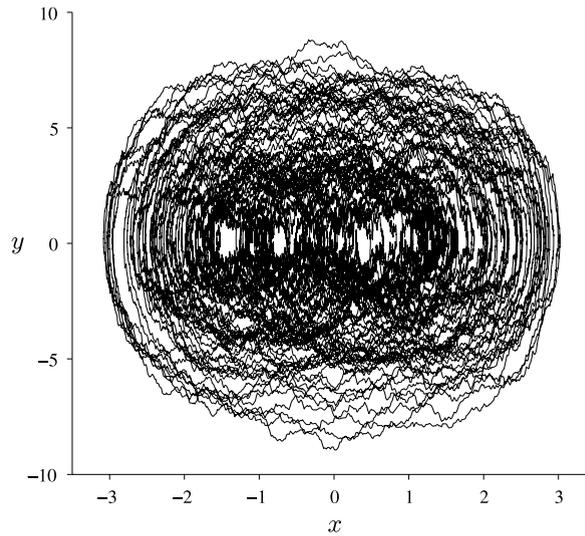
which as a 4-dimensional system provides a route to chaos. However, unlike the chaos of the Lorenz system, the dynamics of the periodically forced Duffing oscillator can appear even more complex since both the position of the fixed points and the shape of the attractor in the  $x$ - $y$  plane change periodically with time. The corresponding phase portrait is shown in Fig. 4.

The forced Duffing oscillator in fact models a periodically forced steel beam that oscillates between two magnets for  $\beta < 0$  [55,54,56]. In the case where  $\beta > 0$  it can be interpreted as a forced spring with the restoring force  $F = -\beta x - \alpha x^3$ .

For sufficiently small values of  $\alpha$ ,  $\gamma$  and  $\delta$ , a phenomenon known as nonlinear resonance also occurs [57,54]. In this case the system approaches the state where  $\frac{d^2x}{dt^2} + \beta x = 0$ , which is the equation for a harmonic oscillator where  $\beta \equiv \omega_0^2$ . The interaction between  $\omega$  and  $\omega_0$  therefore gives rise to a peak in the amplitude of the oscillations as  $\omega$  is varied. However, complexity can be observed in the system with or without this effect so it can be considered a special case. Therefore, for the purpose of this review only strong forcing will be considered with values of  $\alpha$ ,  $\gamma$  and  $\delta$  comparable to  $\beta$ .

### 2.3.3. Stochastic systems

Complexity can also be brought about simply by the influence of some “external” randomness. In a stochastic system the complex dynamics arise from the addition of a noise component to the equations describing the dynamics. While the characteristics of the noise, such as its spectral distribution, may be defined it is otherwise completely separate from the system itself and must be artificially generated. Stochastic systems are therefore strictly open systems, i.e. ones that interact with their environment.



**Fig. 5.** A trajectory of the noise forced Duffing oscillator in phase space for  $\alpha = 1.0$ ,  $\beta = 1.0$ ,  $\delta = 0.2$  and  $\xi = 0.2$ . The system was integrated using the Heun scheme with an appropriate adjustment to take into account the noise term [60].

The unforced Duffing oscillator (7) can be transformed into the stochastic system,

$$\frac{d^2x}{dt^2} + \delta \frac{dx}{dt} + \beta x + \alpha x^3 = \xi \eta(t), \quad (10)$$

where  $\eta(t)$  is a normalized Gaussian white noise source and  $\xi = \sqrt{2D}$  is a constant where  $D$  is the noise intensity. The addition of the stochastic term gives the appearance of complexity as well as other interesting behaviour due to the bistability of the system when  $\beta < 0$ . In this case even with low noise intensity ( $\xi \sim 0.1$ ), at events where  $x, \frac{dx}{dt} \rightarrow 0$  it is possible for the system to jump from one valley in the potential to the other, producing dynamics that resemble chaos. In the literature this phenomenon is known as stochastic resonance [58,59]. A phase portrait of the system for  $\xi = 0.2$  with  $\beta > 0$  is shown in Fig. 5.

While in chaotic systems the tried and tested method of analysis is to construct the attractor in phase space, for stochastic systems the noise term hides the underlying autonomous system, or at the very least increases the apparent complexity of the system. For this reason another approach is taken, which makes use of the Fokker–Planck equation [61]. Under this framework of analysis the dynamics of the system are divided into a deterministic *drift* and stochastic *diffusion*. Once these two effects are known it is then possible to determine the likelihood of a trajectory that the system may take.

The approach to stochastic systems has in the past proved controversial, since their integration can be approached in two different ways [62]. The Stratonovich interpretation seems to be useful for noise that is integrated in the continuous sense, while the Itô interpretation is more appropriate for when the noise is integrated discretely over finite timescales [11].

In the Itô interpretation the drift and diffusion terms of the Fokker–Planck equation may be extracted from time series data of the system [63,11,64]. Once the functions for these terms are known, the deterministic drift can be used to plot flow diagrams that reveal the existence of fixed points in phase space [63]. However, the inverse approach to stochastic systems extends far beyond this and a thorough review can be found in Friedrich et al. [11].

One of the main issues discussed in this paper is that, given only a single observable of the system such as  $x(t)$ , it is difficult to know which analytical approach should be taken. It is always important to recognize when a system is driven by stochastic dynamics and to apply this Fokker–Planck analysis instead. However, it is also easy to assume that a system is stochastic as soon as the methods based around deterministic systems fail. The analysis of time series from the next class of systems aims to highlight the problem of being quick to resort to stochastic methods.

#### 2.3.4. Non-autonomous systems

The main emphasis of this paper will be on a third class known as non-autonomous systems, which have been the subject of recent in-depth studies [65]. As in the stochastic case, they depend on some external input  $f(t)$  and are therefore strictly open systems. While the classic definition of “non-autonomous” includes the possibility of  $f(t)$  being a noise term, in this case we will take it to represent some deterministic function. However, unlike for the forced Duffing oscillator in (8), the forcing is not necessarily periodic. In fact periodic forcing is a specialized case of non-autonomity, where the underlying autonomous attractor of the system can still be extracted from a single trajectory by plotting  $x(t)$  and  $y(t)$  at the intervals  $t = t_0, t_0 + \frac{2\pi}{\omega}, t_0 + \frac{4\pi}{\omega}, \dots$ . To explore the more general issues of dealing with non-autonomous systems, the forcing should not be periodic but instead contain some inherent time-dependence. This modification means that the shape of the attractor becomes extremely difficult to track given only one trajectory of the system.

To best demonstrate the effects of non-autonomicity we should choose the underlying system to be simple and low-dimensional. The Poincaré system is a two-dimensional nonlinear oscillator that generates self-sustained oscillations and is defined in Cartesian coordinates by:

$$\begin{aligned}\frac{dx}{dt} &= -qx - \omega y, \\ \frac{dy}{dt} &= \omega x - qy, \\ q &= \alpha(\sqrt{x^2 + y^2} - a),\end{aligned}\tag{11}$$

where  $a$  and  $\alpha$  are positive constants and  $\omega$  is the angular frequency of the oscillations. The oscillations in this system always converge to the limit cycle given by  $r = \sqrt{x^2 + y^2} = a$  at a rate dependent on  $\alpha$  so long as the initial condition  $r_0 \neq 0$ . Non-autonomicity can be introduced to the system in the following way:

$$\begin{aligned}\frac{dx}{dt} &= -qx - \omega y, \\ \frac{dy}{dt} &= \omega x - qy + \gamma f_y(t), \\ f_y(t) &= \sin(2\pi t) + \sin(4t),\end{aligned}\tag{12}$$

where  $\gamma$  is a constant. The function  $f_y(t)$  is neither periodic nor stochastic so that the equations in (12) provide a clear illustration of a non-autonomous system. Nonperiodic dynamics is easily obtained in this system but these become complex for fairly high values of  $\gamma$  (meaning the amplitude of  $\gamma f_y(t)$  is several times greater than the unmodified Poincaré system) with  $\omega \approx \pi + 2$ , which is the average frequency of  $f_y(t)$ . The onset of complex dynamics is due to the disappearance of the limit cycle from the autonomous perspective of the system. This can be seen by transforming to polar coordinates,

$$\frac{dR}{dt} = -\alpha R(R - a) + \gamma \sin(\theta) f_y(t),\tag{13}$$

$$\frac{d\theta}{dt} = \omega - \gamma \frac{\cos(\theta)}{R} f_y(t),\tag{14}$$

where  $R \exp[i\theta] = x + iy$ . By defining the limit cycle as the point where  $\frac{dR}{dt} = 0$ , the position of the limit cycle is found to be

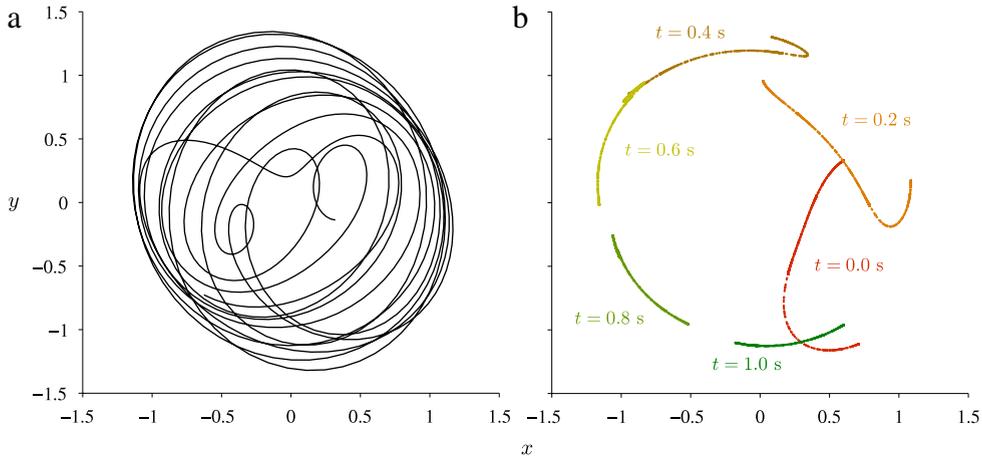
$$R = a \left( \frac{1}{2\alpha} \pm \frac{1}{\alpha} \sqrt{\frac{1}{4} + \frac{\gamma \sin(\theta) f_y(t)}{a^2 \alpha}} \right).\tag{15}$$

The existence of a square root in this equation means that for certain angles  $\theta$  and points in time  $t$  the limit cycle is not defined (i.e. it contains an imaginary part). While there is still a limit cycle at other points, its transient disappearance and reappearance leads to complex dynamics.

Fig. 6 shows the phase portrait of this system as well as the attractor viewed at a single point in time. It can be seen that when viewed only in space, the trajectory of the system is far from a limit cycle and qualitatively it is difficult to determine whether it is deterministic. However, when time is added to the representation it becomes clear that the attractor of the system is confined to a very small part of phase space at any one time. This indicates that the system is neither chaotic nor dominated by a stochastic component as this would lead trajectories to diverge from each other, filling phase space.

The phase space plots of the system illustrate one of the problems with analysing non-autonomous systems. The classical way of dealing with non-autonomous systems is to keep the representation of the system shown in Fig. 6(a) and simply add new dimensions to account for the time-dependent forcing, ‘unwrapping’ the trajectory so that paths no longer cross. This transforms it into autonomous form and was demonstrated before with the forced Duffing system in Eq. (9). In this case the system becomes 6-dimensional:

$$\begin{aligned}\frac{dx}{dt} &= -qx - \omega y, \\ \frac{dy}{dt} &= \omega x - qy + \gamma(u + v), \\ \frac{du}{dt} &= p, \\ \frac{dv}{dt} &= r,\end{aligned}\tag{16}$$



**Fig. 6.** Phase space representations of the forced Poincaré oscillator with  $a = 1$ ,  $\alpha = 5$ ,  $\omega = 5.3$  and  $\gamma = 2$ : (a) A single trajectory of the system over 10 s; (b) Snapshots of 1000 trajectories of the system, revealing the time-dependent attractor. The system was integrated using the Heun scheme with  $h = 10^{-4}$  s and the initial values of  $x$  and  $y$  were obtained from normalized Gaussian white noise.

$$\frac{dp}{dt} = -4\pi^2 u,$$

$$\frac{dr}{dt} = -16v.$$

However, we have to ask: what is actually gained by analysing the system this way? The variables  $u$ ,  $v$ ,  $p$  and  $r$  are completely independent of the rest of the system and in many real non-autonomous systems the dynamics that they model can be subject to change [66]. Including them as new dimensions artificially increases the apparent complexity of the system, making them indistinguishable from systems which actually do have a large number of interacting dimensions.

The solution is therefore to treat non-autonomous systems as a unique class and not as an extension of either the autonomous or stochastic realms. One of the goals of this paper is to motivate this new treatment and to consider which of the existing methods are either suitable or unsuitable for the analysis of these systems.

In a sense, the specific non-autonomous system presented here is the most deterministic out of all the systems considered. Complex dynamics only appear on account of the forcing term, but this term in fact introduces very little ‘external’ information. The fact that it is still composed of linear periodic functions means that given the frequencies of the two sine functions and the initial phases, only the value of  $t$  brings new information into the system. Indeed, the distribution of states in the  $x$ - $y$  plane occupies a relatively small area, which means all trajectories converge to a similar path no matter which initial conditions are chosen. Therefore, to verify their suitability for analysing non-autonomous systems, the methods investigated in the next sections should be able to detect the simplicity of this test system with respect to the other examples.

#### 2.4. Time series

In order to extract information about any system experimentally, repeat measurements must usually be made. For systems that are ergodic [67] there are two approaches. The first is to control the conditions of the system and to measure the response after a time lag. The second is to take an arbitrary set of initial conditions and to make  $N$  consecutive measurements over a time step  $\Delta t$  to obtain a *time series*, where the dynamics of the system (its time-dependent properties) are revealed. Both approaches are valid since with enough measurements they eventually give the same information. For example, chaos would immediately be detected using the first method since any small change in the initial conditions would result in large deviations of the measured values (providing the responses were taken after several cycles). Recording a single trajectory of the system in a time series would not be able to identify chaos so simply but it would be possible after mapping the time series to phase space and constructing the strange attractor [10] (a process which is discussed in detail in Section 6.1). On the other hand, a limit cycle oscillator would be clearly seen in the time series but would require responses from many trajectories if the first approach was used.

For complex systems, observing the dynamics in a time series is the easiest way to understand the system. Certainly for non-autonomous systems, measuring only responses would reveal completely different results depending on the time lag used, as indicated in Fig. 6(b). Time series are also better related to real world problems where the initial conditions of a system can rarely be controlled.

The two fundamental properties of any time series are the interval  $L = N\Delta t$  over which the system is sampled (the recording length) and the rate  $f_s = 1/\Delta t$  at which samples are taken (the sampling frequency). These properties place strict limits on the timescales over which the dynamics is observable. In the frequency domain, the Nyquist–Shannon sampling

theorem [68,69] states that the maximum observable frequency is equal to half the sampling frequency or *Nyquist frequency*,  $f_s/2$ . In addition, the lowest observable frequency is given by  $1/L$ . To obtain a frequency distribution that is continuous and extends between 0 and  $\infty$  then both  $L$  and  $f_s$  must be infinitely large, which cannot be met by the discrete and finite time series obtained from simulations and real world measurements.

#### 2.4.1. Inverse problems

Inverse problems relating to dynamical systems can be tackled in two ways: modelling the system in a set of equations or by making direct measurements from the system. For time series analysis the first case offers the ideal scenario, where the exact state of the system at any given time is known and all of the parameters can be set directly. In the latter case only a one-dimensional representation of the system trajectory is usually obtained, which means the phase and inner workings of the system are much more difficult to define. However, for the inverse problem both approaches are equally useful; the only difference is that when modelling it is the limitations of the model that add to the uncertainty, rather than the limitations of the data.

In this paper, time series from both modelled and measured systems will be studied. However, for the modelled systems most of the time series analysis will be applied to a single dimension of the system to mimic the problems of dealing with measured data. Since the properties of the system are known prior to the analysis, this provides a good test for the suitability of the methods to measured time series.

#### 2.4.2. Nonstationarity versus non-autonomicity

One of the defining features of complex time series is *nonstationarity*. This is defined statistically as when there is a systematic change in either the mean or variance of the time series (or both) [70]. However, more generally it can refer to when the properties of different sections of the same time series are shown to vary.

Confusion can arise when discussing both nonstationarity and non-autonomicity. The latter, as described in Section 2.3.4 refers simply to the effect of having an explicitly time-dependent term in the equations for a dynamical system. These are therefore two very different concepts, although of course they are not mutually exclusive: the time series of the non-autonomous system described here is also nonstationary. However, it is equally possible for non-autonomous limit cycle oscillators to produce stationary time series. It is therefore important not to interchange these two terms.

### 2.5. Simulated systems

Theoretical nonlinear systems must be integrated numerically in order to generate a trajectory and obtain a time series from it. During this process we face a similar problem to the finite sampling frequency in time series analysis. The accuracy of the integration is limited by the integration step  $h$ , with the accuracy increasing as  $h \rightarrow 0$  (neglecting computer precision). This is particularly problematic in chaotic systems which only require the slightest error to alter the trajectory of the system completely. Although there is no solution, integration schemes have been devised to deal with this limitation. In the analysis shown here, the test signals have been generated using the Heun scheme (see Appendix B.1).

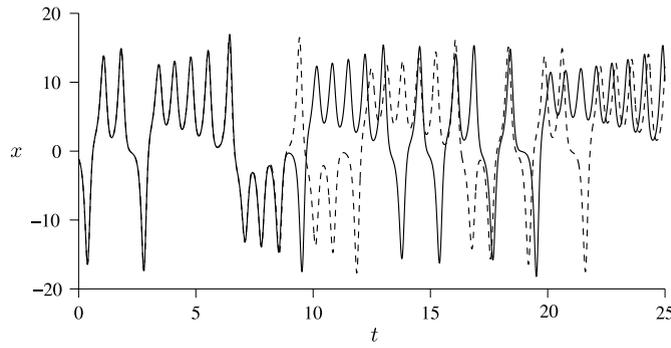
The actual choice of units for the time series generated from simulated systems is arbitrary. However, in this paper seconds, s, and hertz, Hz, have been adopted to help with comparisons between the simulated time series and those of the real world.

In simulations it is common to use adaptive integration schemes that alter the size of  $h$  to keep a constant local precision around each point. However, the calculation of this precision is an estimate itself so it can only help to reduce local error. With chaotic systems the global error (i.e. the difference between the approximated trajectory and the trajectory integrated using an infinitesimal  $h$  and infinite precision) is much more difficult to estimate but has recently been achieved for the Lorenz system up to 1000 s [71].

One way to deal with the problem of the global error is to use a completely different integration scheme, with completely different errors, and compare the two trajectories. The Lie integration method (see Appendix B.2) is a suitable alternative which works by applying a differential operator, rather than integrating the equations of the system directly. Although this formulation is not as intuitive as the Heun scheme, which can be thought of in the same way as the trapezium rule of integration, the contrast between the two methods makes each one an ideal test of the other.

If it is observed that the trajectories of the Heun and Lie integration schemes diverge then obviously at least one of the schemes must be inaccurate. By lowering  $h$  and integrating again, the least accurate trajectory should converge towards the more accurate one although care must be taken. Other than increasing the computation time, lowering  $h$  also increases the computational error since each point is only calculated to a finite number of digits ( $\sim 16$  in double precision) and causes a roundoff error. Therefore, in this case both integration schemes use a precision of 128 decimal digits.

The principle was tested for the Lorenz system, using the integration steps  $h = 10^{-5}$  s and  $h = 10^{-4}$  s for the Heun and Lie integration schemes respectively and the result is shown in Fig. 7. Despite the extremely high local precision of both schemes, it can be seen that the global error still accumulates rapidly and the trajectories deviate after only 12 cycles. However, many of the methods discussed in this review are tested using systems integrated with even less accuracy. To investigate the effects of these global integration errors on the subsequent time series analysis, the Lorenz system is again integrated using the same initial conditions as in Fig. 7 but with larger (less accurate) integration steps,  $h = 10^{-3}$  s,  $10^{-4}$  s.



**Fig. 7.** Integration of the Lorenz system for the initial conditions  $x_0 = -1.1, y_0 = -1.9, z_0 = 13.1$  using the Heun scheme with  $h = 10^{-5}$  s (solid line) and 8th-order Lie scheme with  $h = 10^{-4}$  s (dashed line) with 128 decimal digit precision.

The main effect that this inaccurate integration will have is a *time horizon* in the trajectory [29]. This means that at any point in the time series there will exist a window where, if the first points in the window were taken as the initial conditions, the integrated trajectory follows the actual trajectory. The size of the window will be dependent on  $h$  so that the system will only ‘remember’ a short part of the previous trajectory at any point for  $h = 10^{-3}$  s, but a much larger part for  $h = 10^{-5}$  s.

To further aid the investigation of the effect of different types of integration of the Lorenz system, a time series from an analogue Lorenz system was also generated. In this case the integration is ‘real’ as it can be, but inevitably there is also a small amount of noise in the system due to the imperfections of the electronic components. This was based on the circuit defined in [72] and used the methodology established in [73,74]. The simulation used type MC1458N operational amplifiers and type AD534LD analogue multipliers, where the output of each multiplier was automatically divided by a factor of 10 (which was taken into account in the design of the circuit). The analogue-to-digital conversion was performed at a sampling frequency  $f_s = 1000$  Hz.

It should be noted, however, that another theorem known as *shadowing* finds that integration can in fact be achieved for arbitrarily long time series [75]. Here, the idea is that while an integrated trajectory may not follow the actual trajectory for the exact initial conditions, there is another ‘actual’ trajectory of the system with slightly different initial conditions that shadows the integrated trajectory. In this case the effect of the time horizon seems much less noticeable.

### 2.5.1. Parameters

Based on the limitations of computer memory size and processing time, the standard properties of the time series used in this review are  $L = 1000$  s and  $f_s = 1000$  Hz (integrated using  $h = 0.0001$  s), meaning a time series with  $10^6$  data points. The exceptions to this are the time series of the Duffing system, for which the main oscillatory frequency is lower than in the other systems. To ensure for statistics that the number of points over a typical oscillatory cycle is the same in each case,  $f_s$  is reduced by a factor of 5 in the time series of the Duffing system. However, to ensure that there is still a similar total number of cycles, the length of the time series is also increased by a factor of 5.

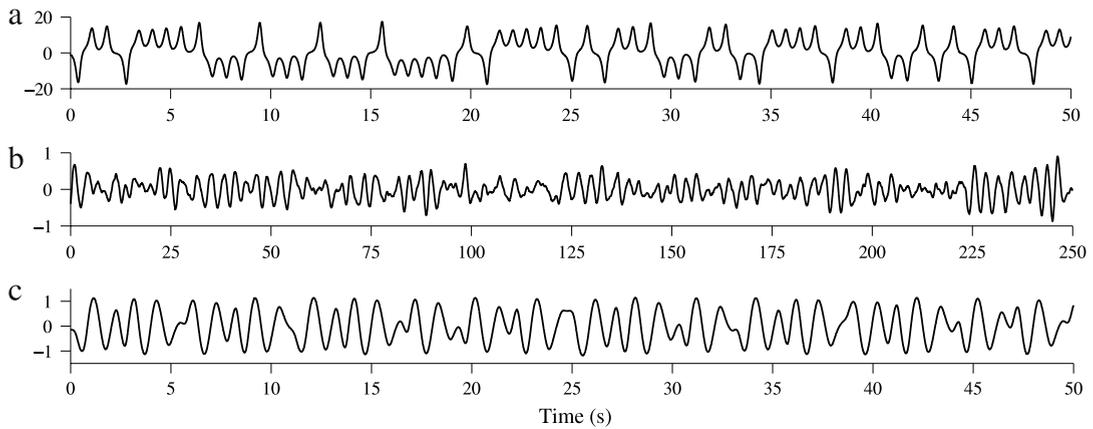
Time series from three complex systems are used as examples. The Lorenz system (6) with parameters  $\sigma = 10, \beta = 8/3$  and  $\rho = 28$  is used to demonstrate a standard chaotic system. The noise-forced Duffing oscillator (10) with the parameters  $\alpha = 1.0, \beta = 1.0, \delta = 0.2$  and  $\xi = 0.2$  is used as an example of a stochastic system. Lastly, the forced Poincaré oscillator (12) with parameters  $a = 1.0, \alpha = 5, \omega = 5.3$  and  $\gamma = 2$  is a demonstration of a non-autonomous system. Samples of the time series from these three main systems are shown in Fig. 8.

In addition to these three cases, time series of the same systems but with different parameters are also generated. The periodically forced Duffing oscillator (8) with parameters  $\alpha = 1.0, \beta = -1.0, \gamma = 0.3, \delta = 0.2$  and  $\omega = 1$  demonstrates chaos with periodically changing dynamics. The noise-forced Duffing oscillator (10) with the parameters  $\alpha = 1.0, \beta = -1.0, \delta = 0.2$  and  $\xi = 0.2$  is an example of a bistable stochastic system exhibiting stochastic resonance for  $\beta = -1.0$ . Additionally, the noise-forced Lorenz system,

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(\rho - z) - y + \xi\eta(t), \\ \frac{dz}{dt} &= xy - \beta z, \end{aligned} \quad (17)$$

and the periodically-forced and noise-forced Duffing oscillator,

$$\frac{d^2x}{dt^2} + \delta \frac{dx}{dt} + \beta x + \alpha x^3 = \gamma \cos(\omega t) + \xi\eta(t), \quad (18)$$



**Fig. 8.** The first 50 000 points for the time series of the three main complex systems: (a) Lorenz; (b) Stochastic Duffing; (c) Non-autonomous Poincaré. In all cases the initial conditions were chosen to avoid specialized dynamics, such as being either too close or far away from the fixed points. The systems were also integrated for 15 000 points prior to  $t = 0$  to avoid the transient features that occur if the initial conditions are not exactly on the attractor.

with the same parameters as before provide examples of chaotic *and* stochastic systems. The forced Poincaré oscillator can also be modified into the following stochastic system,

$$\begin{aligned} \frac{dx}{dt} &= -qx - \omega y, \\ \frac{dy}{dt} &= \omega x - qy + \gamma_y f_y(t) + \xi \eta(t), \\ f_y(t) &= \sin(2\pi t) + \sin(4t), \end{aligned} \quad (19)$$

with  $\xi = 0.2$ . The values for  $\xi$  chosen here are based on the strength of noise used in equations of stochastic nonlinear systems from previous work [76,77].

### 3. Frequency domain

A common way of visualizing the dynamics of a time series is by plotting its frequency spectrum. This view shows how oscillations and fluctuations in the time series are distributed over the different timescales/wavelengths. It has proved particularly valuable to the analysis of turbulence in fluids, where the energy cascade described by Kolmogorov is based on the distribution of kinetic energy among different wavelengths [33]. Later developments known as shell models have also taken advantage of the frequency domain to give a simpler description of turbulence than the one provided by the Navier–Stokes equations [78].

This section will introduce frequency domain representation of time series data via the Fourier transform. It will also provide details of the power spectrum, which is the form most often encountered.

#### 3.1. Fourier transform

The discrete Fourier transform (DFT) is one of the earliest and most widely-used forms of time series analysis. The DFT first appeared in the work of Carl Friedrich Gauss around 1805, whose motivation was to determine the orbits of asteroids by analysing the time series of their locations [79]. Although mathematical functions are typically written in a continuous form before being discretized, Gauss' research on the DFT actually predates the publication of Joseph Fourier's work on the continuous Fourier transform.

The Fourier transform can be more easily understood by first considering Fourier *series*. Fourier's goal was to express a periodic function  $f(t)$  (which can be anything from a polynomial curve to a time series) as an infinite series of sine and cosines,

$$f(t) = a_0 + \sum_{\omega=1}^{\infty} [a_{\omega} \cos(\omega t) + b_{\omega} \sin(\omega t)], \quad (20)$$

where  $\omega$  is the angular frequency and  $a_0$ ,  $a_{\omega}$  and  $b_{\omega}$  are constants known as Fourier coefficients. These coefficients will take certain values depending on the shape of the function, with the largest ones corresponding to components in the time series that have stationary frequencies (also referred to as *modes*).

Using this basis, the DFT of a data series  $f(n)$  can be defined as (see Appendix C.1):

$$F(\omega) = \sum_{n=0}^{N-1} f(n) e^{-\frac{2\pi i \omega n}{N}}. \quad (21)$$

This equation can be used to transform a time series from the time domain to the frequency domain, or from a dependence on  $t$  to one on  $\omega$ . This means that periodic terms in the time series will appear as peaks in the Fourier transform at the corresponding frequency. It is worth noting that although the frequency range is given by  $0 \leq \omega \leq N - 1$ , for real-valued time series the transform is reflected in the point  $\omega = N/2$ , with the coefficients corresponding to positive frequencies for  $\omega < N/2$  and negative frequencies for  $\omega > N/2$ . For this reason the negative-frequency components are usually not plotted, with the remaining components being multiplied by 2 to preserve the total amplitude.

Another common convention is to present the square of the Fourier transform rather than its amplitude. The reason is that this gives the *power* of the Fourier components, which is defined as the average squared variation from the equilibrium point per unit time,

$$P_\omega = \frac{1}{L} |F(\omega)|^2. \quad (22)$$

The units of  $P_\omega$  depend on those of the time series  $f(n)$  and its length  $L$ . This means if the time series was in volts and its length given in seconds then units of the power are simply  $V^2/s$ . Calculating the Fourier transform in terms of power is useful since the total power or *energy* of the time series can be computed in both the time and frequency domains, with

$$\sum_{\omega=0}^{N-1} P_\omega \equiv \frac{1}{N} \sum_{n=1}^N |f(n)|^2. \quad (23)$$

Problems with the Fourier transform begin when considering what happens when a time series contains frequencies that lie outside the detectable range. At the low end of the spectrum for frequencies  $< 1/L$  the components no longer appear periodic and instead seem more like a *trend*. The Fourier series of a non-periodic function is a sum of an infinite number of sines and cosines. Therefore, while most of the amplitudes of oscillations below the frequency limit will appear at  $\omega = 0$  in the Fourier transform, the rest are spread across higher frequencies. Hence, care must be taken to remove trends from a time series before calculating its Fourier transform, otherwise components which *can* be identified in the detectable frequency range will have unreliable amplitudes.

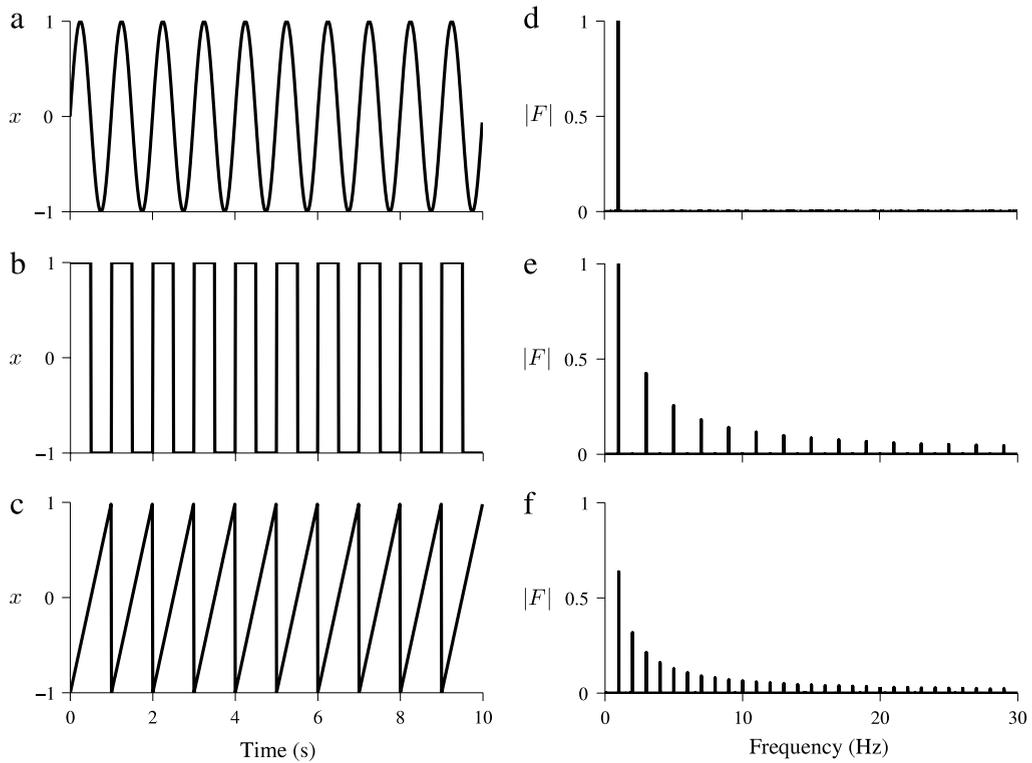
Even for frequencies that are inside the detectable range, difficulties still arise when using the Fourier transform as a frequency domain representation. Since the transform is based around detecting *sinusoidal* components in a time series, the amplitude of periodic components of other shapes is divided into modes that are spread across the frequency domain, as illustrated in Fig. 9. It is therefore possible that the higher frequency modes of components with pronounced amplitudes are mistakenly identified as separate components in the time series. Equally, measuring the amplitude of an oscillation by using the Fourier transform is not straightforward, as can be seen by comparing the height of the primary 1 Hz component in (d) and (f) in Fig. 9. This issue is particularly problematic for nonlinear dynamical systems because the nonlinear terms usually result in non-sinusoidal oscillations.

Another problem of the Fourier transform is the issue of nonstationary time series. The sines and cosines that form the basis of the transform have unchanging amplitudes and frequencies with time. Consequently, the Fourier transform can only be easily interpreted when applied to time series where these properties do not change with time. Fig. 10 gives an example of a chirp signal—a sine wave increasing linearly in frequency with time. As can be seen, this is represented by a broad range of much smaller peaks across the frequency interval of the chirp, which when added together would give the correct amplitude of the chirp. However, as soon as other components appear in the time series these nonstationary oscillations become very difficult to identify, despite their obvious presence in the time domain.

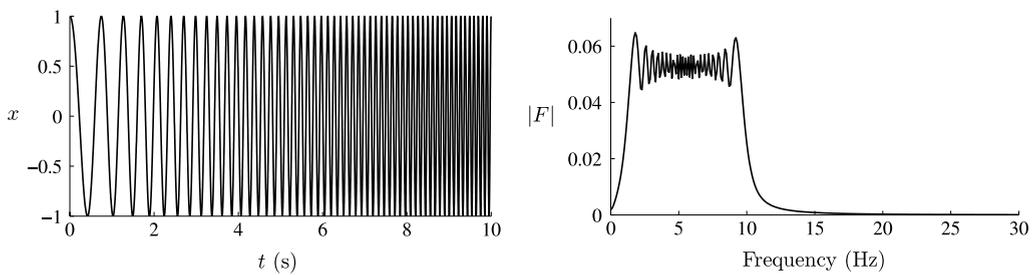
### 3.2. Analysis

Fig. 11 shows Fourier transforms of time series from the three main complex systems that are considered. Both the chaotic and stochastic systems exhibit a continuous distribution, which indicates a lack of periodicity in the time series. For the non-autonomous Poincaré system the Fourier transform very clearly detects the components of the quasi-periodic forcing, although these would be less prominent if they were non-sinusoidal. However, the Fourier components which correspond to the actual nonlinear dynamics of the system are spread across a wide range of frequencies in the same fashion as the chaotic and stochastic systems, despite the fact that these correspond to the dynamics of a simple limit cycle oscillator. This provides a very clear demonstration of how non-autonomy can often resemble chaotic or stochastic dynamics in the frequency domain.

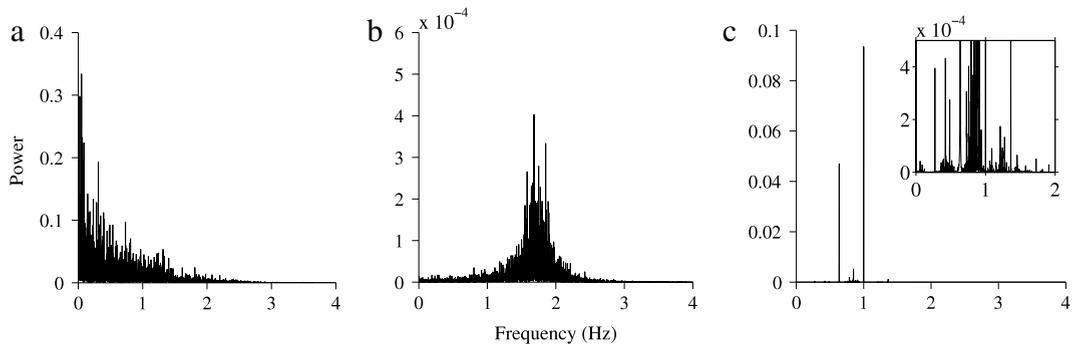
The stochastic Lorenz system has an interesting feature in the frequency domain, which can be seen in Fig. 12. By calculating linear fits of the data from log–log plots of the Fourier transform (not shown), it can be seen that at  $\sim 7$  Hz the power–frequency relation changes from power  $\propto 1/\omega^4$  to power  $\propto 1/\omega^2$ . This implies that the high frequency components observed above  $\sim 7$  Hz are simply the dynamics of the input noise, which as integrated (Brownian) noise should give a  $1/\omega^2$  relationship. For the analogue Lorenz system the amplitude–frequency relation also changes at  $\sim 7$  Hz, although from power  $\propto 1/\omega^4$  to a level distribution resembling white noise. This can be explained by the low power of these components,



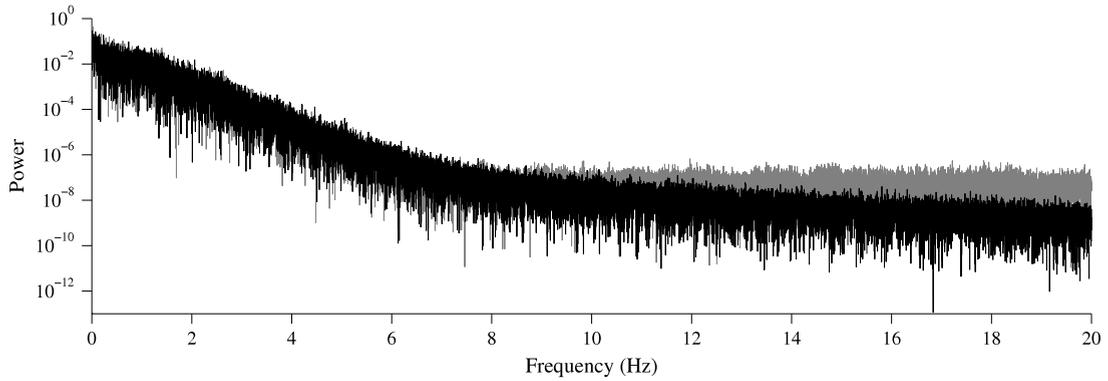
**Fig. 9.** Time series of three differently shaped periodic functions: (a) A sine wave; (c) A square wave; (e) A sawtooth wave; and the respective amplitudes of the one-sided Fourier transforms (b), (d) and (f). In all cases a sampling frequency of 1000 Hz was used.



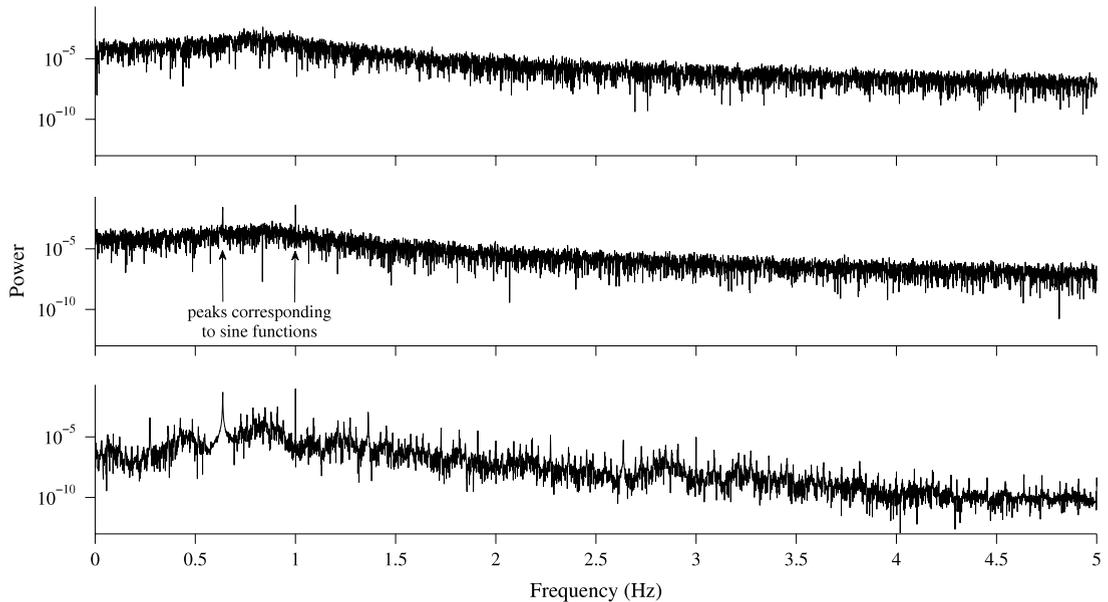
**Fig. 10.** A chirp signal ranging from 1–10 Hz generated over 10 s with a sampling frequency of 1000 Hz as seen in the time domain and its representation in the one-sided Fourier transform.



**Fig. 11.** Fourier transforms over the range 0–4 Hz for: (a) The Lorenz system; (b) The stochastic Duffing system; (c) The non-autonomous Poincaré system, where the inset reveals a broad spectrum of low-power modes.



**Fig. 12.** Fourier transforms over the range 0–20 Hz for the numerically-generated stochastic Lorenz system (17) for  $\xi = 2$  (black) and the analogue Lorenz system (grey).



**Fig. 13.** Fourier transforms over the range 0–5 Hz for the forced Poincaré system: (a) stochastic ( $\gamma = 0, \xi = 1$ ); (b) stochastic and non-autonomous ( $\gamma = 1, \xi = 1$ ); (c) non-autonomous ( $\gamma = 1, \xi = 0$ ).

which are  $\sim 10^{-7}$  and places them within the realm of errors due to the finite precision of the analogue setup. The dynamics observed at frequencies higher than 7 Hz is therefore noise associated with the recording equipment.

Fig. 13 shows the transition from stochastic to deterministic dynamics in the Poincaré system. In the purely stochastic case shown in (a), a single crest is observed in the range 0.5–1 Hz, with an otherwise continuous noise-like distribution spanning the rest of the spectrum. The addition of strong quasi-periodic forcing in (b) only results in two peaks at the corresponding frequencies of the two sine functions. It is only when the noise is completely removed in (c) that the Fourier transform undergoes a drastic change, resulting in a sequence of peaks spanning a range of frequencies much higher than the  $\sim 1$  Hz dynamics observed in the time domain. As in Fig. 9, these are caused by the shape of the oscillations being non-sinusoidal. In time series of dynamical systems, these are referred to as *harmonics*, which arise as a result of nonlinearity in the system altering the sinusoidal shape of the oscillation [29]. The observation of nonlinearity is then in turn a sign of the deterministic nature of a system, i.e. internal variables on which the dynamics depend on either quadratically, cubically, quartically, etc. The fact that these peaks in the Fourier transform can only be observed after the removal of the stochastic term demonstrates how the detection of nonlinearity is very sensitive to noise.

### 3.3. Summary

The Fourier transform is a good starting point in time series analysis. It primarily provides a compact, time-averaged view of the dynamics in a time series, making it useful for measuring the stationary properties of the system. The biggest issue is that its application is restricted to periodic time series. This means that pre-processing needs to occur to remove

low-frequency oscillations that lie below the detectable range (methods to do this are introduced in the next section). Non-sinusoidal oscillations are also not represented well, making the analysis of nonlinear systems difficult but equally causing the Fourier transform to be a good indicator for nonlinearity.

The interpretation of the Fourier transform becomes cryptic when it is applied to nonstationary time series. In this case the transform of a single component oscillating over some frequency range can often not be distinguished from that of band-limited noise over the same frequency range. Despite this, the Fourier transform can still provide a useful statistical interpretation of chaotic and stochastic systems over long time periods. However, such interpretations are unable to account for non-autonomous dynamics.

In terms of the application of the Fourier transform to the different types of complex systems, there appears to be some problems when stochastic dynamics is present. Even when the underlying system is the Duffing oscillator, the Fourier transform shows no indication of there being nonlinearity present. In addition, for deterministic systems there is still the issue of amplitude/power being divided between harmonics, which makes these components appear negligible next to the linear components as is seen in Fig. 11(c). Its usefulness in characterizing complex systems is therefore limited.

## 4. Time domain

The problems associated with analysing time series in the frequency domain suggest that the removal of time information should be avoided. Therefore, it is a logical step to see how time domain methods perform when applied to time series of complex systems. In addition, these methods are the easiest to visualize because their outputs can be compared directly with the original time series.

This section will initially cover the smoothing of time series data, which is an integral part of many time domain methods and is also required for the *preprocessing* of data. The following part of this section will look at splitting complex time series into simpler modes, which is more commonly known as decomposition.

### 4.1. Smoothing

Many time domain methods rely on windowing, which is where only a section of the time series is considered at any time. The process of windowing attempts to isolate the dynamics observed in the time series up to a certain timescale. However, this technique should be distinguished from the frequency domain operation of *filtering*, which involves the direct manipulation or removal of information from the data in the frequency domain. Instead, the effect of these time domain methods is best described as *smoothing* [80].

The advantage of using time domain smoothing over filtering is that time domain methods will always retain information by averaging at each point. Simple frequency domain filtering, whereby the Fourier components over a certain frequency range are removed, does not preserve information in this way and performs badly whenever non-sinusoidal components are present in the time series (i.e. the harmonics are not affected). Since smoothing acts in the time domain it also preserves all time-dependent information in the time series, which means it can be applied to nonstationary time series.

#### 4.1.1. Moving average

One of the simplest smoothing methods is the moving average [81]. This is defined by taking the statistical mean of the data within the window and setting the central point of the window to that value. The window is then shifted along by a single data point, removing a point from one side of the window and adding a new point at the opposite side. The mean is calculated and so on. The overall effect on the time series is that the amplitudes of rapidly changing fluctuations (high-frequency components) are averaged over and reduced while the slow trends (low-frequency components) remain largely unaffected. Hence, the output time series appears smoother and the key parameter in determining this smoothness is the window length,  $l$ . If the window is made larger then lower frequency components will be removed, in addition to the high frequency components that are removed for smaller windows (i.e. greater smoothing).

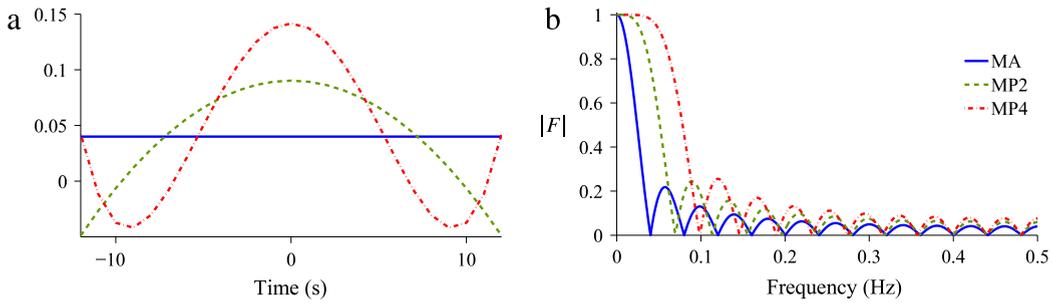
A common use of the moving average method is when downsampling a time series. The Nyquist–Shannon sampling theorem states that in order to prevent aliasing in the output time series, it is necessary to remove components with frequencies higher than  $\frac{1}{2}f_s$  where  $f_s$  is the desired sampling frequency [68,69]. The moving average is a simple way to achieve this and the process is very intuitive. For example, downsampling by a factor of 3 would require a moving average with  $l = 3$  points, with the value of each removed point being included in the average of a point in the downsampled time series. Another interpretation could be that the data are being squeezed into a smaller set with a minimal loss of information.

#### 4.1.2. Moving polynomial

An alternative method to the moving average is the moving polynomial. This again makes use of a window but rather than using a statistical average to find the slow trends in the time series, a least-squares polynomial fitting is used instead. The aim of this method is to find the coefficients  $a_0, a_1, a_2, \dots, a_m$  for the  $m$ th degree polynomial that best fits the points in a given window sample,

$$f(t) = a_0 + a_1t + a_2t^2 + \dots + a_mt^m, \quad (24)$$

where  $t$  is the time. The best-fitting polynomial can be found using the method of least squares (see Appendix C.2).



**Fig. 14.** Convolution functions with a 25-point window for a moving average (MA), quadratic moving polynomial (MP2) and quartic moving polynomial (MP4): (a) The window functions in the time domain; (b) The Fourier transforms of the window functions (using an arbitrary sampling frequency of 100 Hz).

An important difference to the moving average method is that in addition to the choice of  $l$ , there must also be a decision on what degree  $m$  of polynomial to use. This is a nontrivial task as the value of  $m$  has a huge impact on the output time series. Assuming the number of points in the time series  $N \gg m$ , higher degrees make better approximations of the data (including the high frequency fluctuations) while lower degrees follow only the trend of the data (the low frequency fluctuations). The smoothness of the output is therefore inversely dependent on  $m$ .

#### 4.1.3. Window functions

The description of the effects of smoothing methods on a time series has so far been fairly qualitative. However, another way to understand them is to look at the equivalent frequency-domain operations. The best way to do this is to visualize the methods acting as a convolution of a ‘window function’  $g$  and the time series  $f$ . Convolution essentially calculates the amount of overlap between the two functions as one is shifted over the other. For the moving average the corresponding window function is simply a rectangular pulse of width  $l$ , taking a value of  $1/l$  inside the window and 0 outside. The window function for the moving polynomial has a greater complexity and was derived by Savitzky and Golay [80]. The reason for transforming these methods into convolutions is to make use of the convolution theorem which directly equates an operation in the time domain to one in the frequency domain [82],

$$f * g = F^{-1} (F(f) \cdot F(g)), \quad (25)$$

where ‘\*’ denotes the convolution and  $F^{-1}$  is the inverse Fourier transform. Therefore, by taking the Fourier transform of the window function it is possible to see exactly how the amplitudes of components with different frequencies are affected by the method. This is commonly referred to as the *frequency response* of the function.

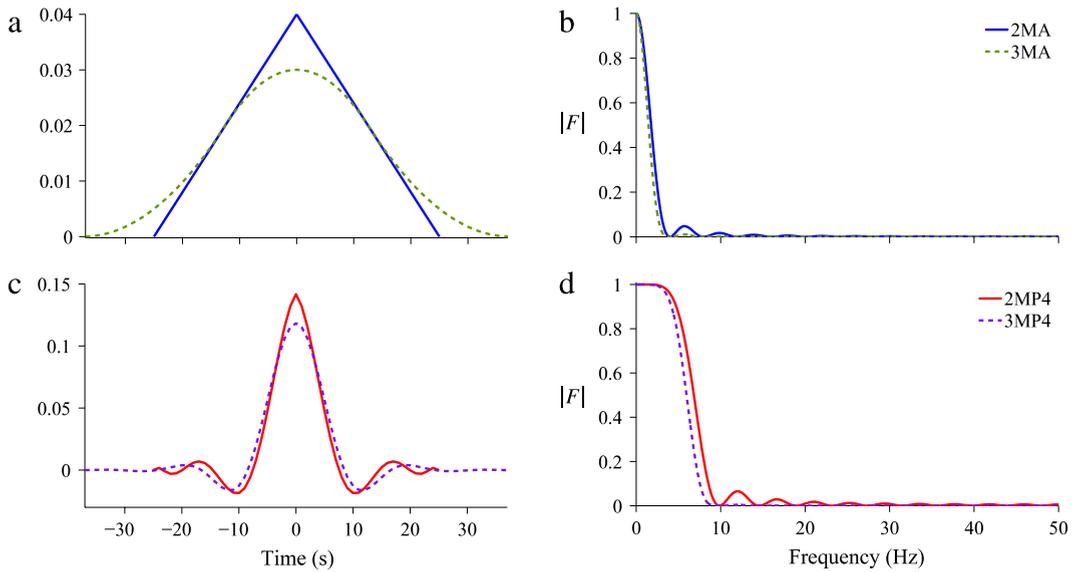
Since the same operation can be defined in the frequency domain, we could ask why we go to the trouble of sliding windows or, indeed, why not just define smoothing filters with ideal frequency responses and apply the convolution theorem? The answer is that while alterations in the amplitude of the frequency spectra are the same in both cases, the phases of the Fourier components can still be different. It is these phases that contain the information necessary to localize the window in time, otherwise if the operation is performed with filters in the frequency domain an entirely different time series can be generated.

Examples of corresponding window functions and their Fourier transforms for both the moving average and moving polynomial methods are shown in Fig. 14. It can be seen in Fig. 14(b) that both smoothing methods have a continuous frequency distribution and no sharp cut-off value as might be desired. As expected, the higher order polynomial preserves the amplitudes of higher-frequency components but there is also a plateau in the amplitude at the low frequency end which is not found for the moving average (i.e. the moving average decreases the amplitude of all the components by some amount). However, the moving average exhibits smaller peaks at frequencies higher than that of the first minimum.

It should also be noted that an additional weighting function can also be applied to the smoothing methods, which alters the distribution of the window function and changes what is seen as ‘local’ to a data point. Another technique used is multiple passing, where the smoothing method is applied at least one more time on the output time series. Mathematically, this can be expressed as the convolution of the convolution,

$$\begin{aligned} (f * g) * g &= F^{-1} \{ F [ F^{-1} (F(f) \cdot F(g)) ] \cdot F(g) \} \\ &= F^{-1} \{ F(f) \cdot F(g) \cdot F(g) \} \\ &= F^{-1} \{ F(f) \cdot F [ F^{-1} (F(g) \cdot F(g)) ] \} \\ &= f * (g * g), \end{aligned} \quad (26)$$

which means the convolution of the window function with itself gives the equivalent one-pass window function for the two-pass smoothing. This might seem to have important implications since, with each pass of smoothing, the method is unable to resolve a period of time equal to half the window length at each end of the time series (as the window runs out of data). The time series can therefore shrink by a considerable amount using multiple pass techniques. Unfortunately, the



**Fig. 15.** Convolution functions for a two (2MA) and three (3MA) pass of a 25-point moving average and a two (2MP4) and three (3MP4) pass of a 25-point quartic moving polynomial: (a), (c) The window functions in the time domain; (b), (d) The Fourier transforms of the window functions (using an arbitrary sampling frequency of 100 Hz).

same is still true even if the equivalent one-pass window function is used because the window length increases by the original length multiplied by the number of passes, leading to the same result. Using the one-pass equivalent though is still useful in reducing computation times, especially when the time series is stationary, in which case the Fast Fourier Transform algorithm can be applied [83]. It can also be seen in the second line of (26) that the effect in the frequency domain is given by the square of the Fourier transform of the window function. This provides a much cleaner smoothing effect, reducing the peaks at the higher frequencies and providing a better-defined cut-off frequency. Examples of multiple pass smoothing windows are shown in Fig. 15.

## 4.2. Decomposition

One of the most common endeavours in time domain analysis is the idea of decomposition. This refers to any transformation where a time series is divided into several new time series which can be summed together to reconstruct the original (see [84] and the references therein). The new time series can either be the oscillatory components, fluctuations corresponding to discrete frequency bands, or some other way of dividing the information present in the time series. For the systems investigated in this review, the number of non-zero components can be interpreted as an indicator of the complexity of the dynamics.

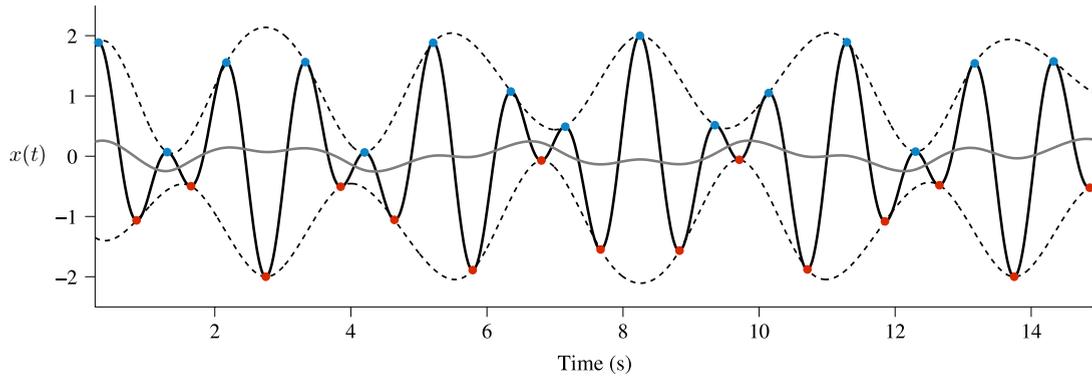
### 4.2.1. Empirical mode decomposition

One of the most intuitive methods for decomposing a time series is Empirical Mode Decomposition (EMD) [85]. The basis for this method is to create an upper and a lower envelope for the time series between which all of the data points are contained. The midpoint of these two envelopes is then used to define the trend, similar to the output of the smoothing methods in the previous section. By subtracting these trends a series of Intrinsic Mode Functions (IMFs) is produced, with the idea being that each one represents an oscillatory component of the time series.

The envelopes are defined by marking points wherever the data in the time series changes direction, as illustrated in Fig. 16. A cubic spline [86] is then used to interpolate between the peaks (the local maxima) to produce the upper envelope, and between the troughs (the local minima) to create the lower envelope. What this technique attempts to do is to approximate the peak-to-peak amplitude of the highest-frequency component in the time series, so the midpoint of these margins is the relative equilibrium point for this mode. When the midpoint is subtracted a new time series  $h_{11}(t)$  is produced, although this is not immediately an IMF since two criteria must also be satisfied. First,

$$N_p + N_t - N_z = 0 \quad \text{or} \quad \pm 1, \quad (27)$$

where  $N_p$ ,  $N_t$  and  $N_z$  are the number of peaks, troughs and zero-crossings respectively. If this criterion is not met then the process of calculating the envelopes and subtracting the trend is repeated on  $h_{1k}(t)$  until (27) is satisfied. This process, known as *sifting*, is still continued after this condition is met in order to remove any 'kinks' that may still be present. However, as will be seen in the examples, sifting introduces its own errors so there must be a condition to stop the process once a physically



**Fig. 16.** The sifting process as applied to the quasi-periodic function  $x(t) = \sin(2\pi t) + \sin(4t)$  (solid black line). The peaks (blue dots) and troughs (red dots) are first marked throughout the time series. The upper and lower margins (dotted lines) are then calculated using cubic spline interpolation between these points. The trend (grey line) is found by taking the average of the upper and lower envelopes.

meaningful mode is obtained. In the paper by Huang et al. [85] this is done by defining the standard deviation between two iterations,

$$\sigma = \sum_{x=0}^n \frac{(h_{1(k-1)}(x) - h_{1k}(x))^2}{h_{1(k-1)}^2(x)}, \quad (28)$$

where an appropriate minimum value of  $\sigma$  can be defined, typically between 0.2 and 0.3. If, on the  $k$ th iteration,  $\sigma$  is less than this value then  $h_{1(k-1)}$  becomes the first IMF. However, it was later noted that this ‘stopping criterion’ is a very arbitrary choice and is unrelated to the convergence to a defined IMF, so instead a criterion based on (27) was introduced [87,88]. In this way, the sifting is stopped when (27) is satisfied for a certain number of iterations  $S$  (typically  $3 \leq S \leq 5$ ). The IMFs produced for different values of  $S$  can then be used to give a mean result, with a confidence given by the standard deviation [88].

Once the first IMF is generated it can then be subtracted from the original time series. The process is then repeated on  $h_{2k}$  until the second IMF is found, and so on.

In ideal cases the resulting IMFs satisfy the conditions of completeness and orthogonality, meaning no information is lost and the original signal can be reconstructed from the decomposition [85]. However, in practice errors can easily arise that prevent this from being the case. The source of these errors is the continually repeated interpolation used to approximate the amplitude of the highest frequency component. This approximation also becomes a problem at the ends of the time series where *extrapolation* must instead be used after the last peak/trough. This often leads to large errors at the ends which become magnified with each IMF. Additionally, the sifting method does not behave well for nonstationary time series where the amplitude of an oscillatory component goes to 0. In such cases, the IMF immediately includes the peaks and troughs of the second highest frequency component, which not only makes the decomposition difficult to interpret but can also introduce errors due to sudden changes in the amplitude.

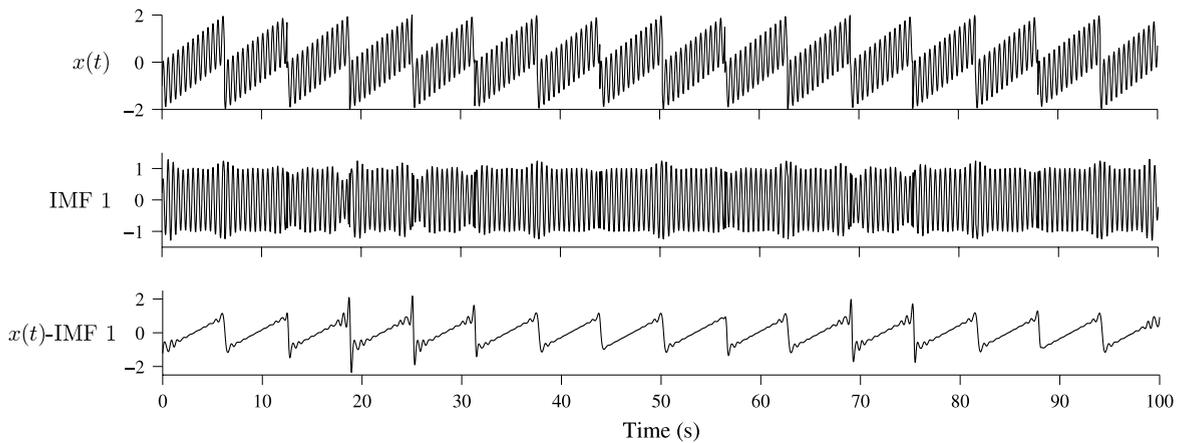
Despite these disadvantages, the sifting method of EMD is still a powerful tool as it provides a way of removing low frequency components without using a moving average or polynomial (in the time domain) or filter (in the frequency domain). For example, this makes it possible to extract a single component with a time-varying frequency without skewing the amplitude when it moves between high or low frequencies. The technique is therefore useful outside of EMD, so long as the propagation of errors is kept under control.

The absence of a mathematical basis such as the Fourier series can in some ways be seen as an advantage of EMD. Since the procedure is based only around peak/trough detection, the decomposition is less-affected by the actual shape of the oscillation. This means that harmonics do not arise, although smooth oscillations result in less errors than cases where there are sharp peaks as shown in Fig. 17. The only contradiction to this is when oscillations contain two or more peaks per cycle, in which case additional IMFs will be generated.

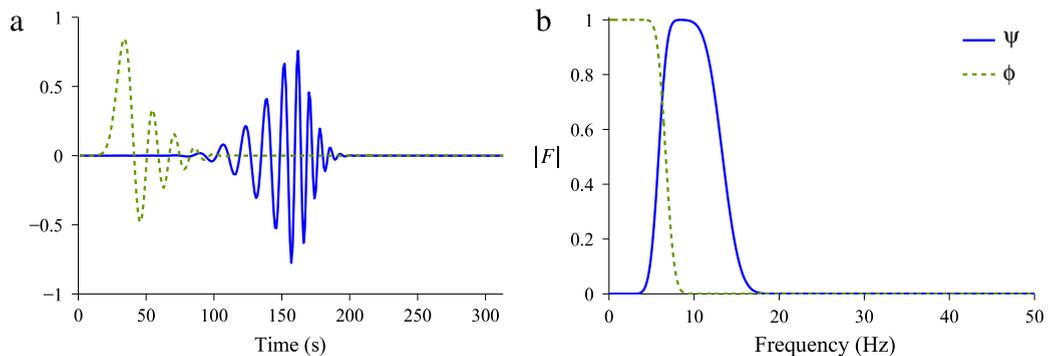
#### 4.2.2. Wavelets

A very different form of decomposition to EMD is possible with window functions known as *wavelets*. Wavelet analysis is discussed in detail in Section 5.4 but will be introduced briefly here. The essential idea behind wavelets is that they give an optimal time–frequency localization, meaning that they can identify the frequency of a component within a time series at a given time to a high accuracy. Starting with a window function known as the *mother wavelet*, this function is stretched or compressed so that the frequency response matches the desired range of frequencies [89].

The effect of convolution of a wavelet with a time series is not the same as smoothing since the frequency response is usually confined to a narrow band that does not extend to zero. In addition, the window size is adaptive so that it is only just large enough for the particular range of frequencies to still be detectable. A weighting function also ensures that the



**Fig. 17.** Empirical mode decomposition of the function  $x(t)$  which has been generated by the addition of a sawtooth wave with frequency  $\frac{1}{2\pi}$  Hz and a 2 Hz sine wave. The method recovers the general shape of the sawtooth where Fourier-based methods would fail, but it comes at the cost of mixing between the two oscillations.

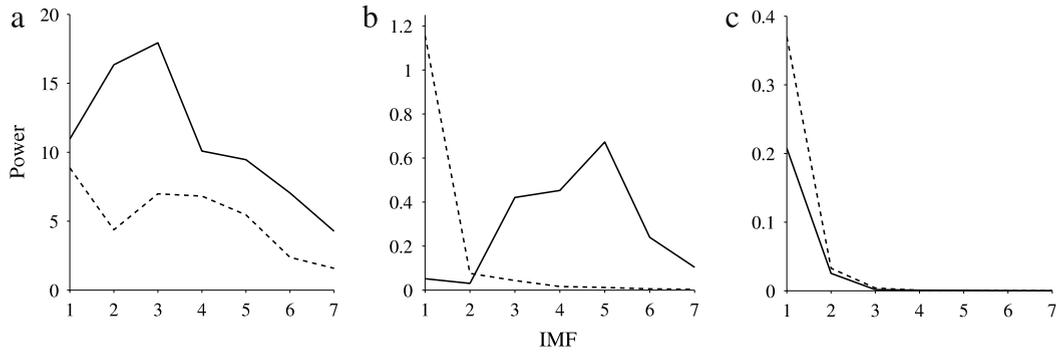


**Fig. 18.** The wavelet  $\psi$  and scaling function  $\phi$  for the Daubechies-20 wavelet [92] used to obtain the third-order detail and approximation: (a) The functions in the time domain; (b) The Fourier transforms of the functions (using an arbitrary sampling frequency of 100 Hz).

amplitude of the window function drops to zero at the edges of the window so that the points nearer to the centre contribute more. These two features give wavelets an optimal time–frequency resolution.

The way in which wavelets are used for decomposition is a method known as subband filtering or coding, or alternatively ‘multiresolution analysis’ [90,91,89]. The process starts by performing the convolution of the time series with the wavelet and a corresponding *scaling function*, examples of which are shown in Fig. 18. The scaling function is defined so that it preserves all of the frequency content below the range covered by the wavelet, meaning that it acts in a similar way to smoothing. The wavelet preserves all of the frequencies ranging from  $f_s/4$  up to the Nyquist frequency,  $f_s/2$ , which gives rise to a time series containing the finest ‘detail’ in the decomposition. In contrast, the scaling function preserves all frequencies ranging from 0 to  $f_s/4$  and produces a smoothed ‘approximation’ of the original time series. Since both the detail and approximation contain information only for frequencies below  $f_s/2$  they can then be downsampled by a factor of 2, so that the new Nyquist frequency becomes  $f_s/4$ . The procedure is then repeated on the approximation using the same wavelet and scaling function to produce a second-level detail and approximation. Since the sampling frequency is reduced to  $f_s/2$ , the detail contains frequencies in the range  $f_s/8$  to  $f_s/4$  while the approximation contains frequencies up to  $f_s/8$ . Downsampling occurs again and the method is continued until the length of the time series becomes a single data point.

The details produced at each level in the wavelet decomposition method (referred to as the *discrete* wavelet transform, to distinguish it from the *continuous* wavelet transform discussed in Section 5.4) are often presented in a single series of data called the *wavelet coefficients*, with the last coefficient being the final level approximation. These coefficients contain all of the information necessary to construct the original time series, which is achieved by carrying out the procedure in reverse and using functions that are the inverse to the wavelet and scaling function. However, their interpretation is not as straightforward as for the IMFs in EMD. The frequency ranges of the decomposed details are based solely on the sampling frequency of the original time series and bear no relation to the physical components that are present in the data. This poses a significant problem when dealing with nonstationary components which occupy a band of frequencies, where the band can be divided between the coefficients of two details. In such cases the choice of sampling frequency can have a huge impact on the identification and analysis of physical components.



**Fig. 19.** Total power of the first 7 intrinsic mode functions from empirical mode decomposition: (a) Lorenz system; (b) Stochastic Duffing system; (c) Non-autonomous Poincaré system. The dashed lines shows the IMFs for the same time series after applying a 0.5 s (2.5 s for the Duffing system) moving average to remove high-frequency components. To reduce the effect of the large errors at the edges, only the middle half of the time series (250 to 750 s) was used to compute the power of the IMFs.

#### 4.3. Analysis

Time domain smoothing is a useful tool which can be used to isolate all of the subsequent time series analysis to a predefined range of timescales and frequencies. This is important particularly when there are very low frequency dynamics and the time series is simply not long enough to be able to interpret these fluctuations. However, such slow oscillations and trends can still have an impact on the analysis of the dynamics within the observable frequency range, through phenomena such as high harmonics.

To prevent this, it is recommended to define the ‘trend’ of a time series using a smoothing method which removes components above a suitable timescale. This timescale is usually taken to be  $\sim 5f_s/L$ , where around five full cycles are observed over the length of the time series [93]. This is because at timescales larger than this the estimations of the statistical properties of the dynamics become unreliable. The trend defined by the smoothing process is then subtracted from the original time series, leaving only dynamics at shorter timescales.

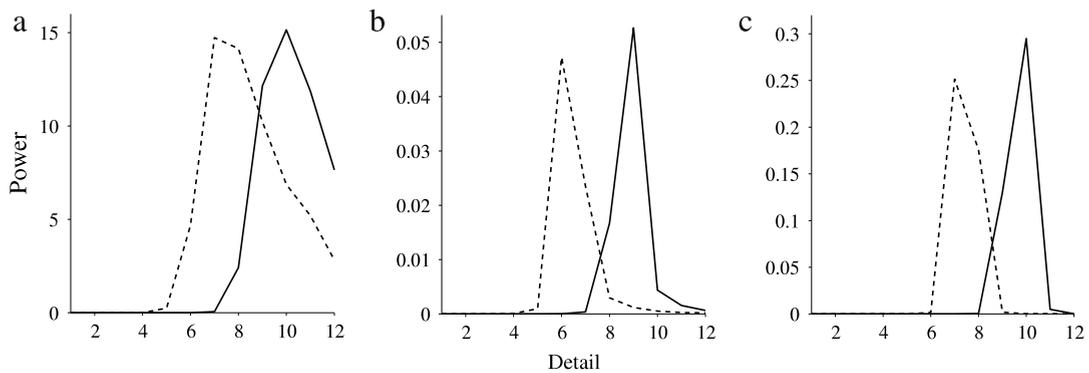
Equivalently, the accuracy of many time series analysis methods is reduced by high-frequency noise. Therefore, in addition to removing trends using a large window, high-frequency fluctuations can be smoothed over using a small window. The resulting time series is then said to be *bandpassed*, where the dynamics is preserved between both a high-frequency and a low-frequency cutoff.

The empirical mode decompositions of time series from our three complex systems are shown in Fig. 19. These plots give the same information as the Fourier power spectra shown in Fig. 11 but with lower resolution. Also, in contrast to Fourier series the lower-number modes correspond to high-frequency components. An advantage of this representation though is that the nonlinear frequency scale gives a more intuitive power spectrum, where stationary and nonstationary components of equal amplitude appear the same.

For the Lorenz system the distribution of power amongst the IMFs shows a kink which is not visible in the Fourier power spectrum. Here the large peak corresponds to the main orbits of the system, while the IMFs after this point correspond to the switching between the two fixed points. EMD also seems to work well for the non-autonomous Poincaré system where the dynamics is decomposed into a much smaller number of IMFs than for the two other systems.

Removing high-frequency components by smoothing appears to have an effect on the power distribution of the low-frequency IMFs. For the stochastic Duffing system the initial IMFs contain the low-amplitude stochastic dynamics, while the main oscillatory dynamics is spread across several IMFs. The decomposition of the unsmoothed time series therefore suffers greatly from error propagation due to the fact that the oscillations in the first IMFs are not continuous. The nature of stochastic dynamics means that the oscillations repeatedly drop to zero amplitude, causing mixing between IMFs. However, after smoothing the error propagation is reduced so that the main dynamics is confined to a single dominant IMF with a distribution similar to that of the non-autonomous Poincaré system.

Unlike other decomposition methods, EMD lacks a strong mathematical foundation so it is not possible to verify the completeness and orthogonality of the IMFs. A significant problem is that the method behaves differently depending on the nature of the time series it is used on. For the time series of the stochastic Duffing and Lorenz systems the dynamics is spread almost continuously across a wide frequency distribution (see Fig. 11), due to the high degree of nonstationary. This means that power is continuously vanishing and reappearing at different frequencies, which results in IMFs with a jumble of components at different frequencies and large errors propagating through the decomposition [94]. It is worth noting though that the problem of mixing between IMFs has recently been tackled in a modification to the method known as ensemble empirical mode decomposition (EEMD) [95]. This procedure adds white noise to the time series before applying EMD. This is then repeated using different realizations of the noise to produce an ensemble of decompositions, where the ‘physical’ IMFs are determined by the average. In the same paper the method is also used as a way to reduce the large



**Fig. 20.** Total power of wavelet decomposition details for the Daubechies-20 wavelet [92]: (a) Lorenz system; (b) Stochastic Duffing system; (c) Non-autonomous Poincaré system. Solid lines show the results for the fully-sampled time series with  $f_s = 1000$  Hz ( $f_s = 200$  Hz for the Duffing system) while the dotted lines show the decompositions when the time series were downsampled to  $f_s = 200$  Hz ( $f_s = 40$  Hz for the Duffing system).

errors caused by extrapolation at the edges. However, this modification can still not guarantee to completely eliminate error propagation.

The distributions of power in the wavelet decompositions of the three time series are shown in Fig. 20. As with EMD, the nonlinear frequency scale helps to simplify the power spectrum. However, altering the sampling frequency changes the frequency intervals which correspond to each detail of the decomposition. While this shifts the position of the details containing the main dynamics, more importantly it affects the shape of the power distribution. The most striking difference is seen for the Lorenz system where for the 1000 Hz time series a large part of the low frequency dynamics is contained within the final approximation of the decomposition (not included in the figure). In the case of the 200 Hz time series the peak in the distribution is also much flatter, which implies that the frequency range of the main oscillatory component is divided between two details.

This analysis shows that it is not guaranteed that wavelets will produce a unique decomposition. The use of wavelets to decompose complex time series in inverse problems is therefore not advised unless the frequency intervals of interest are known beforehand.

#### 4.4. Summary

Time domain methods are a useful first step in time series analysis, especially in the preprocessing stage of methods that act in other domains.

Time domain smoothing acts as a way of preserving phase information of oscillatory components while removing high frequency dynamics. This is by far preferable to frequency domain filtering which does not modify the phases of the Fourier components and can therefore not be applied to nonstationary time series.

Decomposition methods also provide ways of representing complex time series as a combination of simpler modes. The number of modes can be used as an indicator for the complexity of the system and can therefore offer some insight into the different types of dynamics, i.e. chaotic, stochastic or non-autonomous.

EMD is a very unique method and it has merits in that it can remove low-frequency components without the use of either smoothing or frequency filters. However, its iterative technique is not suited to time series where the dynamics is defined over a broadband spectrum, as is true in the case of complex dynamics. Such time series are susceptible to mixing between IMFs, which causes large errors to propagate through the decomposition. In particular, application to time series of stochastic systems is problematic due to the presence of oscillations with vanishing amplitudes.

Wavelet decomposition has become a popular method of isolating dynamics occurring at different timescales. However, the divisions in frequency between separate details is entirely arbitrary. A more sound approach would be to decide on the frequency intervals of the decomposition using *a priori* knowledge or otherwise choose empirically from the power distribution observed in the Fourier transform.

## 5. Time–frequency domain

The time domain decomposition methods offer advantages over the Fourier transform but it would still be useful to have spectral methods which could be applied effectively to the time series of complex systems.

While the frequency domain representation provided by the Fourier transform gives a great deal of information that would otherwise remain hidden in a complex time series, it is only suited to cases where periodic *and* stationary components exist. It is not able to provide unambiguous information about how the components change with time, as nonstationary

dynamics is simply represented by separate stationary components. What is needed is therefore a way of tracking the phase and frequency of different components with time.

In the context of fluid turbulence, the importance of tracking the frequency spectra in time arose after the discovery that energy does not transfer at a constant rate between scales but actually occurs intermittently [96]. Therefore, although time averages of the dynamics still agree with the stationary representation provided by Kolmogorov's original theory and the Fourier transform, other analytical tools are needed to understand the variation from the stationary solution [97,98].

The following section begins by detailing methods which are applicable to single oscillatory components with variable frequency, also known as the *instantaneous frequency*. It then extends to methods that have been developed to reveal the change in time of the frequency spectra of complex time series with more than one component.

### 5.1. Instantaneous frequency

The term 'instantaneous frequency' is somewhat of a misnomer as it implies that there is a general definition for the frequency of an oscillation at every point in time. In truth, frequency is defined as an integer number of cycles over a time interval, so it should not be possible to know the oscillation frequency at any arbitrary fraction of a cycle. However, this changes when we impose a model definition of the frequency where we assume that certain properties are preserved in each cycle. A common example of this is the assumption that the oscillation can be transformed to polar coordinates, where the angular rotation or *phase* is separable from the change in the radius (also known as the amplitude dynamics). A restriction in this case is that the change in frequency occurs at a much slower rate than the actual dynamics of the phase. The system can then be considered in the same way as a sinusoidal oscillation with a phase defined as  $\phi = \omega t + \theta$ . If we assume this phase can be defined for each cycle, the instantaneous angular frequency can then be found simply by taking the time derivative,

$$\omega = \frac{d\phi}{dt}. \quad (29)$$

#### 5.1.1. Marked events

The method of marked events calculates the frequency of an oscillation using a basic definition: the number of events per unit time. For the continuous oscillations observed in time series, an "event" can be any position between 0 and  $2\pi$  in the phase of a single cycle. The main choices are typically the peak, trough or a zero-crossing (for a mean-subtracted oscillation). Due to the considerations explained above it is important that the event is present in each cycle and also that its position in the phase is not altered. The instantaneous frequency can then be defined as the reciprocal of the period between subsequent events in different cycles.

The downside of this method is that the time resolution is equal only to the period of the oscillation. However, this resolution can be increased by using several marked events per cycle, so long as the positions of the events are well-defined. Still, the resulting time resolution will always be lower than that of the original time series.

The definition of an event needs to be decided with care, in particular when the oscillation is not symmetrical (i.e. not reflected across the mean as in the sine, square and sawtooth waves in Fig. 9). Rapidly-changing spikes are the preferred choice as they give the best time resolution, meaning the change in frequency can be tracked more precisely. In contrast, smoother features increase the difficulty of defining the exact time of an event and are more influenced by noise and other high frequency oscillations. In addition, zero-crossings can also be a poor choice as they are heavily influenced by drifts and other low frequency oscillations.

To deal with these problems, time domain smoothing (see 4.1) can sometimes be used to remove either the high frequency components (by using the smoothed time series) or the low frequency components (through subtraction of the smoothed time series).

While the marked events method can only track the frequency of a single component, the output can be treated as a completely new time series in its own right, giving the *variability* of the component's frequency with time. All of the methods discussed so far can then be reapplied to learn more about the more subtle dynamics of the oscillatory component.

The main limitation of the marked events method is that the sampling frequency of the resulting time series is variable over individual cycles of the oscillation, meaning that interpolation must always be performed before analysis. Since the frequency and phase cannot be defined between events this procedure is always an approximation.

#### 5.1.2. Hilbert transform

The Hilbert transform was devised by the mathematician David Hilbert to solve a problem relating to the study of differential equations in the complex plane [99]. It has since been widely adopted in time series analysis methods to convert a real signal into complex one known as an *analytic signal*.

The Hilbert transform  $H(t)$  of the time series  $f(t)$  takes the form:

$$H(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(u)}{u-t} du. \quad (30)$$

This is equivalent to the convolution of  $f(t)$  with  $\frac{1}{\pi t}$ . In either case, the function is non-integrable (since the integrand approaches infinity as  $u - t \rightarrow 0$ ) and is therefore defined mathematically by the Cauchy principal value.

It is not intuitively clear from (30) what the physical meaning of the Hilbert transform is. The best way to see this is to look at its relationship with the Fourier transform, where

$$F[H(t)](\omega) = \begin{cases} 2F[f(t)](\omega) & \text{for } \omega > 0 \\ F[f(t)] & \text{for } \omega = 0 \\ 0 & \text{for } \omega < 0. \end{cases} \quad (31)$$

Here it can be seen that  $H(t)$  removes the negative frequency components of the Fourier transform and compensates by doubling the amplitude of the positive frequency components. Since the negative frequencies are simply a reflection of the other half of the Fourier transform no information about the time series is lost.

The main advantage of the Hilbert transform is that it can find the phase of a single oscillation directly, without the need for methods such as time delay embedding which are discussed later in Section 6.1. The extracted phase can then be inserted into Eq. (29) to give the instantaneous frequency of the oscillation. In the discrete case this gives a time series of the instantaneous frequency with the same sampling frequency as the original time series.

The preservation of the time resolution makes the Hilbert transform far superior to the marked events method, although this comes at the cost of the type of time series it can be applied to. In particular, it is only suitable where the equilibrium point of the oscillation does not drift in the time domain; minima above the equilibrium point and maxima below it are interpreted as phase slips, meaning the Hilbert phase ceases to be monotonic and moves backwards. This restriction is due to the fact that the phase relates to harmonic oscillations and consequently applies only to 2D systems.

The method is extensively used in the Hilbert–Huang transform [85]. This is the process of taking the IMFs from EMD and applying the Hilbert transform to each to find the instantaneous frequencies. The amplitudes of the IMFs at each point in time are then added to the appropriate frequency bins to give a time–frequency representation of the decomposition. A significant advantage of this method is that, like EMD, it does not suffer from the effect of high harmonics. Also, since it uses all of the decomposed IMFs in a single representation, it is less affected by the mixing of components between the IMFs. However, the Hilbert–Huang transform also carries across all of the problems associated with EMD, including the propagation of errors through the IMFs.

## 5.2. Short-time Fourier transform

Other than by using a decomposition method beforehand, techniques to calculate the instantaneous frequency cannot be applied to the complex time series analysed in this review. However, direct transformations to the time–frequency domain also exist, with an applicability independent of the form of the time series.

The first of these is the short-time Fourier transform or STFT (also known as the windowed Fourier transform), which was developed as a solution to the shortcomings of the Fourier transform when dealing with nonstationary time series [100]. This was achieved by giving a *time–frequency* representation of a time series.

The STFT works in an intuitive way by computing a Fourier transform over a sliding window. As the window moves through the time series, the Fourier spectrum of the window is assigned to the central point. This gives a representation of the frequency distribution at each point in time. Using (C.10), the STFT is defined mathematically as:

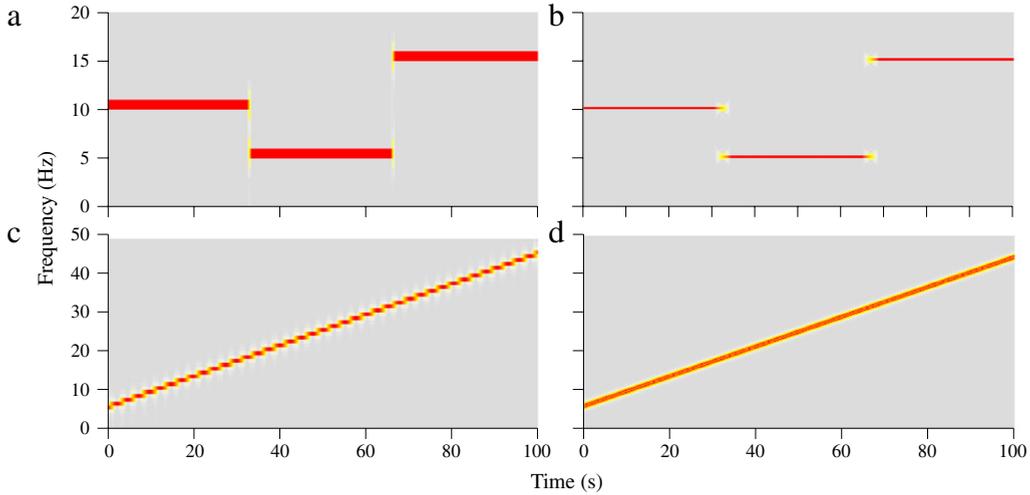
$$\tilde{f}(\omega, t) = \int_{-L/2}^{L/2} g(u - t)f(u)e^{-\frac{2\pi i\omega u}{l}} du, \quad (32)$$

where  $g(u)$  is a rectangular<sup>1</sup> function of length  $l$  that is zero outside of the interval  $-l/2 \leq u \leq l/2$ . The variable  $\omega$  is directly related to the frequency  $f_\omega$  by  $f_\omega = \frac{\omega}{l\Delta t}$ , while  $t$  is the time. In addition to the frequency spectrum, the complex coefficients of the STFT also provide information about the phase of the components (i.e. the position in the cycle with respect to time).

Unfortunately, at the same time as solving the problems caused by nonstationarity the STFT creates the new problem of time–frequency localization. The most significant change from (C.10) to (32) is that the length of the time series  $L$  has been replaced with the length of the window  $l$ . Since the basis of the original Fourier transform (i.e. the Fourier series (20)) is to represent a time series as a sum of sines and cosines with *integer* angular frequencies  $\omega = 1, 2, 3, \dots$ , the frequency step or *resolution* in (21) is given by  $1/L$ . When using the whole of the time series in the transform the effect of this resolution limit is negligible unless there are separate components very close in frequency. However, in the case of the short-time Fourier transform only part of the time series of length  $l$  is transformed at any time, meaning that both the frequency resolution and the lowest detectable frequency are instead given by the inverse of the window length,  $1/l$ .

This problem has in fact already been illustrated in Fig. 15, where a larger convolution function provides poorer resolution in time but a much sharper cut-off frequency due to a higher resolution in frequency. The relation between time and frequency resolution is known as the Gabor limit [100] and it in fact has many similarities to the Heisenberg uncertainty principle of quantum mechanics. If a small window is chosen then more information is available about changes in the time

<sup>1</sup> For optimal localization in time a Gaussian window can be used instead, in which case the operation is known as the Gabor transform [100].



**Fig. 21.** Short-time Fourier transforms with varying time and frequency resolutions. (a) and (b) are of a time series containing a sine wave with a frequency which starts at 10 Hz, then jumps to 5 Hz and finally rises to 15 Hz. (c) and (d) are of a time series containing a chirp signal which increases linearly in frequency from 5 to 45 Hz over 100 s. On the left hand side a short window of size  $l = 1$  s has been used, giving good time resolution but poor frequency resolution. On the right hand side a larger window of size  $l = 4$  s has been used, giving better frequency resolution but poorer time resolution (as indicated by gaps between the frequency jumps in (b)). In all cases a sampling frequency of 1000 Hz was used.

domain, although at the same time this sacrifices information from the frequency domain. On the other hand, a large window can be chosen to provide more information about the frequency domain while losing information from the time domain. These two extremes are illustrated in Fig. 21.

When there is a change in frequency of some oscillatory component, the corresponding Gabor limit of the chosen window size is put to the test. Although the changes in Fig. 21(a) and (b) occur instantaneously, the shorter window results in a stretching of the power along a large range of frequencies, while the longer window causes a division of the power into two separate frequencies for several seconds.

In Fig. 21(c) and (d) it is interesting to note that the width of the line corresponding to the chirp signal is the same regardless of the window size used. This indicates the finiteness of the Gabor limit and means that no matter how ‘optimal’ the window size is made, it is not possible to observe a precise change in frequency over a precise time interval.

### 5.3. Wigner–Ville transform

When looking for a method which is not as restricted by time–frequency resolution issues, the Wigner–Ville transform is an appropriate alternative. The original Wigner function was developed within the realm of quantum mechanics as a quasi-probability distribution [101]. However, years later the same function was derived independently by Jean Ville who applied it in the context of time–frequency analysis [102]. It is calculated using the *autocovariance*, which is the covariance between delayed versions of the same time series as is done when the autocorrelation is calculated (see Section 6.1.1). This gives a measure of local time–frequency symmetry in the time series, and since it does not rely on windowing the time–frequency resolution is not restricted. The Wigner–Ville transform (WVT) is given by:

$$W(\omega, t) = \int_{-\infty}^{\infty} f\left(t + \frac{\tau}{2}\right) f^*\left(t - \frac{\tau}{2}\right) e^{-2\pi i \omega \tau} d\tau. \quad (33)$$

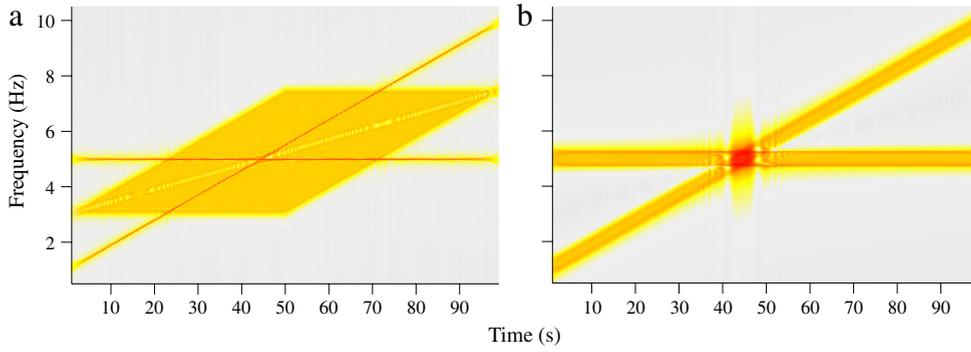
In reality the values of  $\tau$  are limited by the length and sampling frequency of the time series. For time series containing only real values, the Hilbert transform needs to be applied to  $f(t)$  beforehand (introduced in Section 5.1.2).

The advantage of the WVT over the STFT is that stationarity in the time series is not assumed over any time window. However, the WVT is still very difficult to interpret since the transform is non-additive [103], which means that for two components  $f_1(t)$  and  $f_2(t)$  of a time series,

$$W_{f_1+f_2} = W_{f_1} + W_{f_2} + W_{f_1, f_2} + W_{f_2, f_1}, \quad (34)$$

where  $W_{x,y} = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) y^*\left(t - \frac{\tau}{2}\right) d\tau$  is the *cross* Wigner–Ville transform. These interference or ‘‘cross terms’’ appear as peaks between two real components in the transform and are illustrated in Fig. 22(a). Various solutions have been proposed to remove these terms while preserving the real peaks. The most popular takes advantage of the fact that cross terms appear as oscillations in the transform (i.e. a series of peaks and troughs) which means they can be removed by appropriate smoothing along the time and frequency axes. The *smoothed* WVT is given by

$$W_s(\omega, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(v, u) W(\omega - v, t - u) dudv, \quad (35)$$



**Fig. 22.** A log-scale plot of the WVT of a time series composed of a sine wave with frequency at 5 Hz and a chirp wave increasing linearly in frequency from 1 to 10 Hz: (a) The standard WVT; (b) The smoothed WVT calculated using one-pass moving average windows of width 1 s and 1 Hz in the time and frequency domains respectively. In both cases a sampling frequency of 1000 Hz was used.

where  $\phi(v, u)$  is a kernel designed to filter out the cross terms. Unfortunately, this kernel introduces averaging over both time and frequency, which essentially reduces the resolution in each of these dimensions [104]. In Fig. 22, the moving averages remove the main effects of the cross terms but as a result increase the thickness of the lines corresponding to the two components. Overall, the advantages of the WVT over the STFT are therefore limited.

#### 5.4. Continuous wavelet transform

By far the best solution to the time–frequency resolution problem of the STFT is to move away from the Fourier basis of sines and cosines to a new basis of *wavelets*. As already discussed in Section 4.2.2, wavelets have been developed to give “optimal” time–frequency resolution which is achieved by using an *adaptive* window. The continuous wavelet transform (CWT) is found by converting (32) to this new basis,

$$W_T(s, t) = \int_{-L/2}^{L/2} \Psi(s, u - t) f(u) du, \quad (36)$$

where  $\Psi(s, t)$  is the mother wavelet, which is scaled according to the parameter  $s$  to change its frequency distribution and time-shifted according to  $t$ . This transform differs from the discrete wavelet transform discussed earlier because the frequency scale is continuous, meaning the wavelet components  $W_T(s, t)$  can be calculated for any arbitrary frequency.

Rather than computing a “stand-alone transform” for each time window, the CWT performs a different calculation depending on both time *and* frequency (or more specifically,  $s$ ). This means a new wavelet and window size is used to calculate each scale, with a small wavelet / window for high frequencies and larger ones for low frequencies. Hence, the time resolution at high frequencies is not degraded by the condition of needing a large window to detect low frequencies.

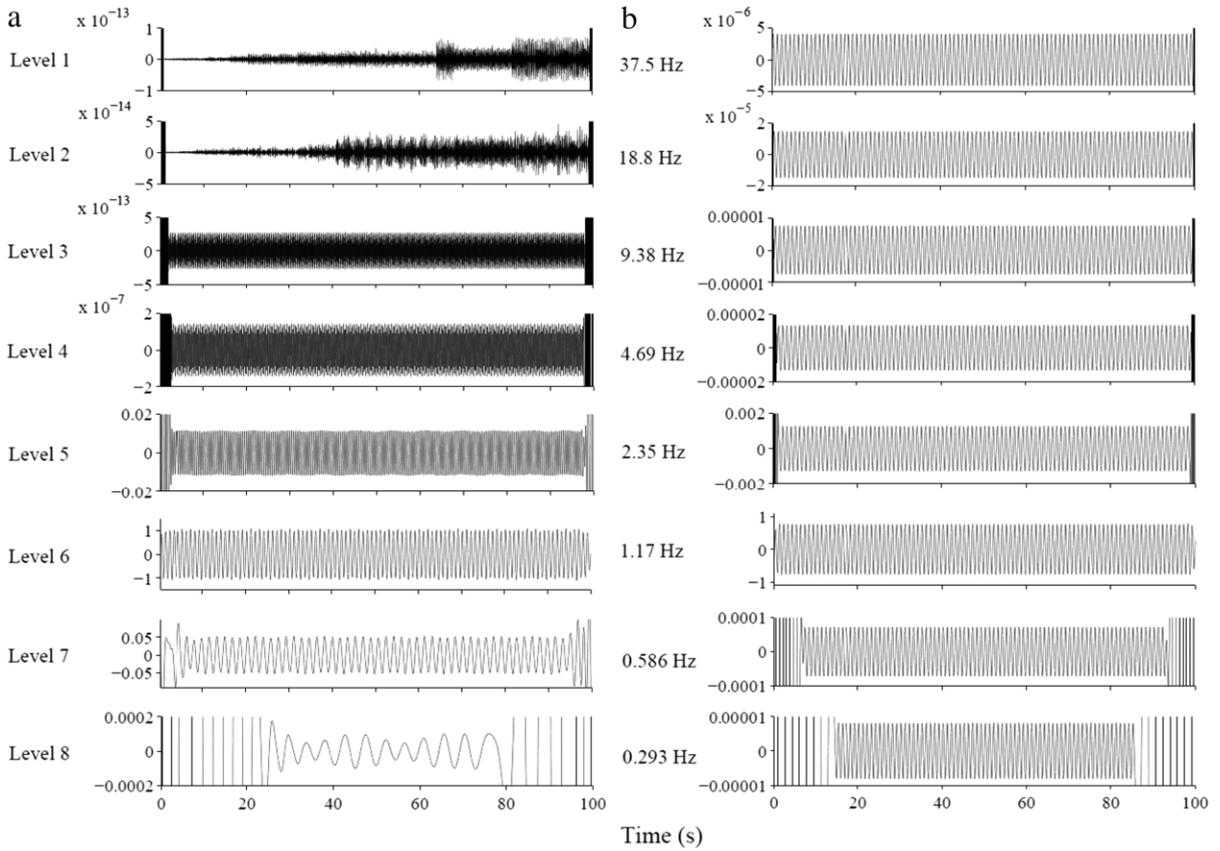
The main difference from the discrete wavelet transform (DWT) is that here the basis given by the mother wavelet is not necessarily orthogonal. This means that the wavelets have overlapping frequency bands and the components  $W_T(s, n)$  contain redundant information, as shown in Fig. 23. When applied to a single sine wave the details of the DWT contain completely independent phases, while the phase of the CWT is the same for all frequencies (neglecting edge effects). This means that a strong oscillatory component, harmonic or otherwise, can often dominate the CWT through a wide range of frequencies, eliminating any legitimate low-amplitude dynamics at those frequencies.

To increase time–frequency resolution to the very limit dictated by the uncertainty principle, the wavelet window needs to have a Gaussian shape [105]. Applying this principle and keeping the same sinusoidal basis as the Fourier transform gives rise to the Morlet wavelet [106],

$$\Psi(s, t) = \frac{1}{\sqrt{4\pi}} \left( e^{\frac{2\pi i \omega_0 t}{s}} - e^{-\frac{2\pi \omega_c^2}{2}} \right) e^{-\frac{t^2}{2s^2}}, \quad (37)$$

where  $s = 1/\omega$ . The parameter  $\omega_c$  is the central frequency, which determines the time–frequency resolution of the wavelet; high values of  $\omega_c$  give good frequency but poor time resolution while low values of  $\omega_c$  give good time but poor frequency resolution. The  $e^{-\frac{t^2}{2s^2}}$  part specifies the Gaussian shape of the window and stretching of the wavelet with respect to  $s$  while  $e^{-\frac{2\pi i \omega_0 t}{s}}$  is equivalent to the basis of the Fourier transform (C.8). In a sense, for each scale the Morlet wavelet transform can be thought of as the analytic signal of the bandpass-filtered dynamics around a specific frequency.

After optimally incorporating the time and frequency domains and removing many of the issues of dealing with nonstationary time series, the CWT can be interpreted much in the same way as the Fourier transform. While this is obviously an advantage when comparing the coefficients of the two transforms, it also means that the CWT inherits the same problem



**Fig. 23.** Comparison between the discrete and continuous wavelet transforms of a 1 Hz sine wave with a sampling frequency of 100 Hz: (a) DWT decomposition using Daubechies-20 wavelets [92], showing the detail at each level; (b) Real part of the CWT using the Morlet wavelet ( $\omega_c = 1$ ) for the frequencies corresponding to the centre of the levels in the DWT decomposition.

of high harmonics that was previously illustrated in Fig. 9. However, the fact that the wavelet transform obtains time-dependent phase information means that it is possible to detect relations between oscillations and identify the harmonics caused by nonlinearity. Therefore, despite the fact that the wavelet transform is a linear transformation, methods have now been developed that use it to find and extract nonlinear oscillations [107].

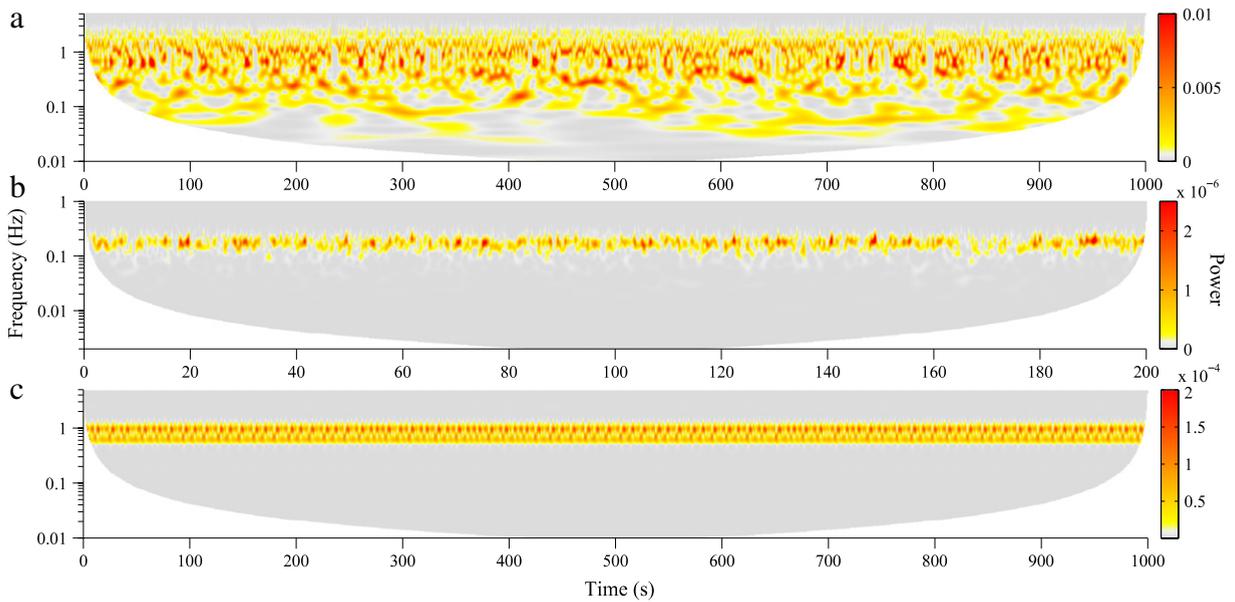
The main contrast to the Fourier transform is that the Morlet wavelet transform is continuous, which means that although the amplitude is analogous to the Fourier amplitude, the calculation of the power is not as simple as in (22). To find the wavelet power spectrum, the modulus square of the wavelet transform must be integrated over frequency,

$$P_W(\omega, t) = \int_{\omega - \frac{d\omega}{2}}^{\omega + \frac{d\omega}{2}} |W_T(\omega, t)|^2 d\omega. \tag{38}$$

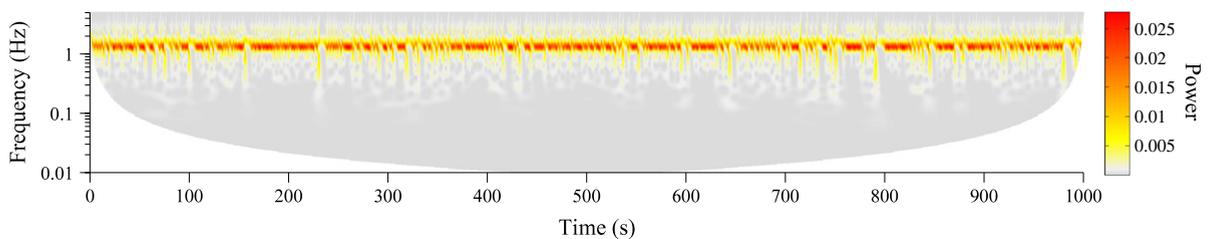
This means that values of either the power or energy of the wavelet transform for finite data (i.e. finite frequency resolution) can only ever be estimates.

When comparing the CWT with the STFT, a significant improvement lies in the frequency scale. While the STFT has a linear scale, the scale of the CWT is logarithmic. This means that at low frequencies the energy is divided into many more coefficients, making it easier to distinguish several low frequency nonstationary components which would otherwise be merged into the same frequency interval in the STFT.

It might seem that the time–frequency resolution limit can be bypassed by using wavelets to decompose a time series into separate components and then applying the Hilbert transform to obtain the instantaneous frequency of each. The same idea can actually be applied directly using only the CWT and is known as a *squeezed* wavelet transform. This approach, developed by Daubechies [108,109], uses the phase information in the wavelet transform to calculate the precise frequency, from which the amplitude of each point arises. Using this technique, it is indeed possible to determine the instantaneous frequency of several components in a time series with respect to time. However, the finite frequency resolution of the wavelets means that the components have to be well-separated in frequency for this to work. If this is not the case then the overlapping wavelets will detect elements from both components, meaning that the phase and instantaneous frequency cannot be defined. This results in the same time–frequency ‘blurring’ as seen in the standard CWT.



**Fig. 24.** Continuous Morlet wavelet transform of time series from: (a) Lorenz system; (b) Stochastic Duffing system; (c) Non-autonomous Poincaré system. The power was calculated by integrating with respect to frequency over the squared wavelet transform. Due to computer processing limitations, the time series were downsampled by a factor of 10 so that the sampling frequency was 100 Hz for the Lorenz and Poincaré systems and 20 Hz for the Duffing system. A central frequency  $\omega_c = 1$  was used for the wavelets.



**Fig. 25.** Continuous Morlet wavelet transform of the  $z$ -component for the Lorenz system for the same parameters as used in Fig. 24(a).

### 5.5. Analysis

After indicating its advantages over the STFT and WVT, the CWT is presented here as the optimal representation of the complex systems in the time–frequency domain.

Fig. 24 shows continuous wavelet transforms of the three complex systems. Viewing a time series in this way not only determines whether it is nonstationary, but it also shows the *degree* of nonstationarity. For the time series of the Lorenz and stochastic Duffing systems the frequency distribution changes continuously, while for the non-autonomous Poincaré system the dynamics is confined within the same frequency range for the entire duration. Therefore, while the frequency distribution is similarly broad and complex as in the case of the chaotic and stochastic systems, the time series of the non-autonomous system is actually *less* nonstationary. Therefore, while this might not be an intuitive result, it shows that a time-dependent output of a system is more closely related to the instability of that system than to an inherent non-autonomous component.

Perhaps the most noticeable observation in Fig. 24 is the very broad spectrum of the dynamics for the Lorenz system when compared to the others. This is due to the random switching between orbits around the two fixed points in phase space, which can be appreciated when it is compared to the wavelet transform of the  $z$ -component as shown in Fig. 25. The contrast between these two wavelet transforms is relevant to the limitations of the time delay embedding theorem as discussed next in Section 6.1.2. When only the  $z$ -component is considered the dynamics is confined to the higher frequencies  $\sim 1$  Hz, which means the dynamics occurring on longer timescales are neglected.

### 5.6. Summary

At the interface between the time and frequency domains a huge amount of extra information appears that was not present in either the time or frequency domains alone. By knowing the frequency of an oscillation at every point in time

we automatically obtain the complete phase dynamics of the oscillation as well. This makes the time–frequency domain extremely useful for looking at the interactions that take place between oscillations, as will be discussed later in Section 8.4.

The extension of the Fourier transform to the time–frequency domain removes the ambiguity when analysing nonstationary time series. This means that complex time series consisting of many oscillatory components can be distinguished from a single oscillation with a time-varying frequency. This is especially important for deterministic non-autonomous systems, which might otherwise be identified as being stochastic.

Despite these advantages, moving to the time–frequency domain means that a fundamental limit is placed on the resolution along both the time and frequency axes. Some methods are able to get around this limit but come with their own caveats such as the production of cross terms or the requirement that all oscillatory components are well-separated in frequency. Arguably though, the only method which improves the time–frequency resolution of the short-time Fourier transform without drawbacks is the continuous wavelet transform.

Despite having a linear basis, the wavelet transform also offers a way to deal with nonlinear systems by extracting phase information in time, which can then be used to piece together the harmonics of a base oscillation [107]. However, even without this further analysis the wavelet transform provides an informative representation of complex dynamics. The difference between the change in the frequency spectra for the more nonstationary chaotic stochastic time series and the time series from a stable non-autonomous system is clear in the time–frequency domain. This representation is therefore extremely useful for identifying non-autonomous dynamics.

## 6. Phase space

As seen in Section 2.3, a common way of analysing dynamical systems is to observe how they are represented in phase space. This is a very simple process for numerically generated systems where the raw dimensions  $x$ ,  $y$ ,  $z$  etc. are obtained directly. However, extracting the same phase information from a single time series poses a significant challenge.

This section will initially cover the steps needed to transform a single time series of a dynamical system to phase space. Later on, a method to determine the stability of trajectories in phase space is considered. Lastly, a method to extract the dimension of the attractor for each of the systems will be introduced.

### 6.1. Embedding

The process of transforming a time series  $x(t)$  into phase space is known as *embedding*. The framework of this transformation was developed by Floris Takens [110] and Ricardo Mañé [111] and involves the construction of an embedding vector for each point in time;

$$\mathbf{x}(t_i) = [x(t_i), x(t_i + l\Delta t), \dots, x(t_i + (d - 1)l\Delta t)], \quad (39)$$

where  $d$  is the *embedding dimension* and  $l$  is an integer, both of which must be chosen prior to embedding. The dimensions of the reconstructed attractor are therefore composed of time-delayed versions of the data in  $x(t)$ .

#### 6.1.1. Parameter estimation

For the choice of  $l$ , the embedding theorem specifies no conditions. Technically any delay time  $l\Delta t$  (so long as it is not exactly equal to the period of an oscillatory mode) should give a “correct” reconstruction of the attractor, preserving all of its local properties. However, in practice it is necessary to have a delay that maximizes the spread of the data in phase space and reduces the statistical errors of analytical phase space methods [112,113]. In general, the optimal time delays are neither extremely short, so that the values in the embedding vectors are essentially the same, or extremely long so that the values are uncorrelated variables [114].

Estimating the optimal time delay is simplest when the time series contains a dominant oscillation. In this case the delay time can be set as a fraction of the approximate period (ideally  $1/4$ ), allowing the coordinates to span the extrema of the oscillation [113]. However, when the dynamics contain several periodic components, statistical methods must be employed instead.

The autocorrelation function is a simple tool that can be used to find time delays that give the least correlation between points in a time series, hence causing points to spread out in phase space. It is defined as

$$r(l) = \frac{\sum_{i=1}^{N-l} (x(t_i + l\Delta t) - \bar{x})(x(t_i) - \bar{x})}{\sum_{i=1}^{N-l} (x(t_i) - \bar{x})^2}, \quad (40)$$

where  $\bar{x} = \frac{1}{N-l} \sum_{i=1}^{N-l} x(t_i)$  is the mean [82]. The appropriate time delay for embedding can be taken as the smallest value of  $l$  for which  $r(l) = 0$ . However, a significant problem with this method is that it looks for *linear* correlations, which makes the estimate of  $l$  dubious when it is applied to *nonlinear* systems. The autocorrelation function also neglects time-variability, which often results in overestimates of the time delay when applied to nonstationary time series. Another method was

therefore devised by Fraser and Swinney [115] that uses the idea of mutual information. This is a measure of how much information about one data series is shared with another, or in other words it is how much the uncertainty of one data series is reduced if you are given the other one. For the time series  $x(t)$  starting at  $t_0$  and the delayed time series  $x_d(t)$  starting at  $t_0 + l\Delta t$  the mutual information is given by

$$I(l\Delta t) = \sum_{i=1}^{N-l} P_{xx_d}(x(t_i), x_d(t_i)) \log_2 \left( \frac{P_{xx_d}(x(t_i), x_d(t_i))}{P_x(x(t_i))P_{x_d}(x_d(t_i))} \right), \quad (41)$$

where  $P_x$  and  $P_{x_d}$  are the probability distributions for  $x(t)$  and  $x_d(t)$  respectively and  $P_{xx_d}$  is the joint probability distribution (the probability of observing a value from one time series and a value from the other at the same time). In a similar way to the previous method, the first local minimum of  $I(l\Delta t)$  is used to determine a suitable value of  $l$ . However, the dependence on time series probability distributions needs careful consideration and a discussion on this is provided in Appendix D.1.

For the estimation of  $d$ , the embedding theorem states that in order to correctly reconstruct the system in phase space for any  $l$ , the following condition must be met:

$$d \geq 2D + 1, \quad (42)$$

where  $D$  is the smallest theoretical dimension of phase space for which the trajectories of the system will not overlap [116]. Of course, such a condition does not need to be strictly adhered to as if the overlapping of trajectories is minimal, the properties of an attractor reconstructed using a lower embedding dimension are almost identical. Unfortunately though, this condition has little meaning when examining observed time series where  $D$  is not known. For this purpose, the method of false nearest neighbours was developed as a way of estimating the dimension [117]. For a given  $d$ , each embedding vector is paired with its nearest neighbour, i.e. the one separated by the smallest Euclidean distance  $R_d$ :

$$R_d = ([x(t_i) - x(t_{NN})]^2 + [x(t_i + l\Delta t) - x(t_{NN} + l\Delta t)]^2 + \dots + [x(t_i + (d-1)l\Delta t) - x(t_{NN} + (d-1)l\Delta t)]^2)^{\frac{1}{2}}, \quad (43)$$

where  $t_{NN}$  is the position of the nearest neighbour in the time domain. A new embedding dimension  $d + 1$  is then used and  $R_{d+1}$  is now calculated for the same vectors that were identified as nearest neighbours in the previous embedding. This method works on a brute-force basis, increasing the embedding dimension and checking whether the change in distance between nearby points agrees with that expected for the correct dimension. If  $(R_{d+1}^2 - R_d^2)/R_d^2$  is greater than some threshold value  $R_T^2$  then the vector pair are labelled as *false* nearest neighbours. The choice of  $R_T$  is entirely subjective although others have stated that the algorithm performs well for  $10 \leq R_T \leq 30$  [117,118]. If the percentage of false nearest neighbours is high (typically  $> 1\%$  [117]) this implies that other dimensions have been projected onto the  $d$ -dimensional phase space, causing points that would otherwise be completely separate to appear close together. The procedure can then be repeated for  $d + 1, d + 2, \dots$  until the percentage of false nearest neighbours is sufficiently small so that the embedding dimension satisfies a similar condition to (42).

Estimating the required embedding dimension in this way might seem like unnecessary hassle, given that the condition in (42) does not state a maximum value for  $d$ . One might think that it is acceptable to simply make it arbitrarily large and this is true for completely deterministic systems. However, if there is even slight contamination with noise these higher dimensions can suddenly become populated with unrelated stochastic dynamics [116]. Any analysis on the reconstructed attractor of the system will then be susceptible to treating this noise as part of the deterministic dynamics.

### 6.1.2. Limitations

The theorem of Takens and Mañé can be applied to any data but care must always be taken. In particular, the procedure is quite arbitrary when constructing the embedding vectors, which can mean that a lot of redundant information is present when it is adapted for multivariate data. This causes problems when investigating causality between different processes and an alternative embedding procedure has therefore been devised to tackle this [119].

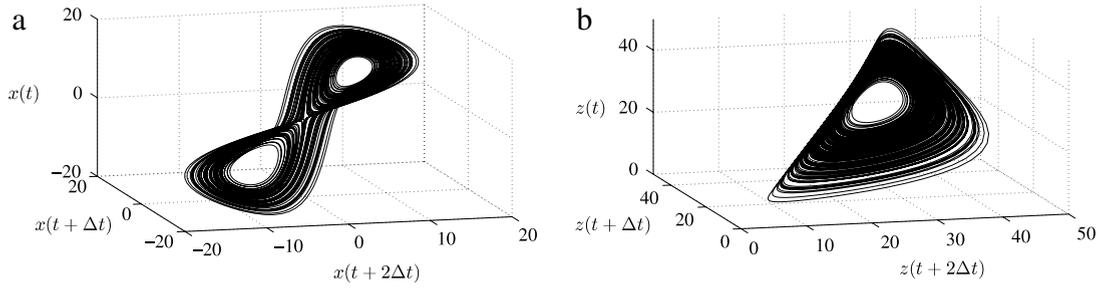
Another issue is that the theorem only assures that embedding preserves the local structure of the system's attractor. This means that for each point in the original phase space there will be points in the reconstructed phase space with a similar distribution of neighbours.

However, the theorem does not assure that the larger structure of the attractor is preserved. This is illustrated in Fig. 26 for the Lorenz system, where completely different shapes of the attractor are observed when time series from different system coordinates are used for the phase space reconstruction. As mentioned earlier, this is related to Fig. 25 where the dynamics corresponding to the movement between the two separate fixed points is represented by the low frequencies in the wavelet transform. Therefore, the issue of non-unique phase space embeddings is actually an issue of restricting the analysis of systems to certain timescales.

Despite this limitation, the preservation of local structure in embedding still provides a great deal of information about the stability and dimensionality of the system. The following methods use embedding as a tool to extract this information.

## 6.2. Lyapunov exponents

By reconstructing the phase space of a system, one can immediately observe features such as limit cycles and strange attractors. However, all representations of time are lost in phase space, which means it is not possible to answer questions



**Fig. 26.** Time delayed embedding of the Lorenz system with  $\Delta t = 0.06$  s and  $d = 3$ , using: (a) the x-coordinate; (b) the z-coordinate.

relating to stability or whether specific features observed in phase space relate to transient dynamics or if they are present throughout the entire time series. For example, observing a spiral in phase space means that there is an oscillation in the system that could either be gaining or losing amplitude over time.

The calculation of Lyapunov exponents is a technique designed to address this issue by providing a quantifiable measure of how trajectories in phase space change with time. Also known as *characteristic exponents*, they were first introduced by Aleksandr Lyapunov in 1892 [120] and are a measure of the rate of divergence (or convergence) between two infinitesimally close trajectories. If the initial separation between two trajectories is  $\delta\mathbf{x}(t_0)$  then at some future time  $t$  the separation will be given by

$$|\delta\mathbf{x}(t)| = |e^{\Lambda(t-t_0)} \delta\mathbf{x}(t_0)|, \tag{44}$$

where  $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_d]$  are the Lyapunov exponents. Although  $\mathbf{x}(t)$  relates to the trajectory of the system as a whole, an exponent can be calculated for each dimension of the phase space, leading to what is known as the *Lyapunov spectrum*.

The calculation of Lyapunov exponents requires numerical analysis even in the direct approach where the equations of a system are assumed to be known. The methods to do this were proposed by Eckmann and Ruelle [10] and also Sano and Sawada [121]. The time derivative or “flow” of the trajectories in phase space is first calculated,

$$\mathbf{f}(\mathbf{x}(t)) = \frac{d\mathbf{x}(t)}{dt}. \tag{45}$$

The elements of the Jacobi matrix  $J$  for a specific point in the flow,

$$J_{j,k}(\mathbf{x}(t)) = \frac{\partial f(\mathbf{x}(t))_j}{\partial x_k} \tag{46}$$

are then found (i.e. either estimated from trajectories embedded in phase space or calculated directly if the equations are known). Considering a small perturbation  $\delta\mathbf{x}(t_0)$  in the trajectory at  $t_0$ , the separation between the original and the perturbed trajectory after some time will be given by [122]:

$$\begin{aligned} \delta\mathbf{x}(t_0 + sz\Delta t) &= J(\mathbf{x}(t_0 + (s-1)z\Delta t)) \cdot J(\mathbf{x}(t_0 + (s-2)z\Delta t)) \cdots J(\mathbf{x}(t_0)) \cdot \delta\mathbf{x}(t_0) \\ &= Y(\mathbf{x}(t_0), s) \cdot \delta\mathbf{x}(t_0), \end{aligned} \tag{47}$$

where  $s$  is the number of steps in time and  $z$  is the number of points between each step. It was proven by Valery Oseledets [123] that the matrix

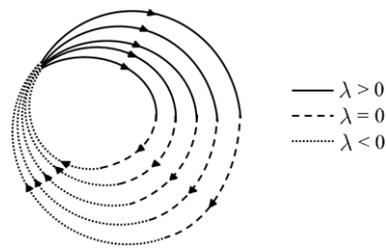
$$\Lambda = \lim_{s \rightarrow \infty} [Y(\mathbf{x}(t_0), s) \cdot Y(\mathbf{x}(t_0), s)^T]^{\frac{1}{2s}} \tag{48}$$

and its eigenvalues  $e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_d}$  can be computed, where  $\lambda_i$  are the Lyapunov exponents. The important feature to note, however, is that these are *global* exponents. This means that they will only give an average over the entire time interval. In addition, they can only ever be approximations since the limit  $s \rightarrow \infty$  cannot be fulfilled for a real time series. For these reasons, *local* Lyapunov exponents were introduced by Abarbanel et al. [116]. In this case the eigenvalues of the following matrix are used instead:

$$\Lambda(Y(\mathbf{x}(t_0), s)) = [Y(\mathbf{x}(t_0), s) \cdot Y(\mathbf{x}(t_0), s)^T]^{\frac{1}{2s}}, \tag{49}$$

where the matrix and derived exponents are now dependent on the initial point  $t_0$  and a time window given by  $sz\Delta t$ . The formulation of (48) and (49) would seem to suggest that maximizing the number of steps  $s$  should give the most accurate exponents. However, the size of the time window is in fact the only free parameter that can be arbitrarily set to increase the accuracy. Reducing  $z\Delta t$  instead leads to many more matrix multiplications over a shorter time interval which, rather than improving the results, actually increases the numerical error [124].

While the theoretical foundation for extracting Lyapunov exponents from time series seems straightforward, in practice it requires a series of complex calculations. Several methods have been proposed [114,124,121] and there exist many more



**Fig. 27.** A sketch of five trajectories moving in a clockwise direction within an attractor. The effective Lyapunov exponent  $\lambda$  dictates the structure of the attractor, with dense regions forming wherever a negative exponent dominates and sparse regions occurring when the prevailing exponent is positive. Note that an attractor such as the one shown here would need to be embedded in at least a 3-dimensional phase space but the 2-dimensional sketch is used here for illustrative purposes.

ways of implementing these in computer algorithms. However, a problem common to all of the methods is the huge number of free parameters that are involved, most of which depend entirely on the nature of the system under analysis. In addition to the parameters from the embedding, the estimate of  $Y(\mathbf{x}(t_0), s)$  requires a decision of what is ‘local’ to each point. This is important because in addition to there being a minimum so that the number of neighbouring points  $N \geq d$ , there also has to be a maximum  $N$  so that the local area does not span a wide space with respect to the size of the attractor [121]. To calculate the change in  $J(t)$  for each reference point there also needs to be a basis in phase space to approximate the system dynamics, which could be a least-squares fit [124], quadratic polynomials [122,114], high-order polynomials [125] or radial basis functions [122,125]. The remaining parameters,  $s$  and  $z$ , are also important choices since they dictate the timescales of dynamics that are analysed, introducing the same possibility of bias as the choice of  $l$  for embedding.

Important properties of the system can be determined by examining the signs of the Lyapunov exponents. If the exponent for some direction is positive then any two close trajectories moving in that direction will diverge exponentially. Conversely, if the exponent is negative then the trajectories will converge exponentially. A zero value exponent implies that the trajectories remain parallel, neither diverging nor converging. Ultimately, it is the combination of these exponents that results in the shape of the attractor, as illustrated in Fig. 27.

In a dissipative system the sum of all Lyapunov exponents is negative since such a system is defined by all trajectories converging to a fixed point. However, it is important to note that this is not a causal relation; a negative exponent also exists in the autonomous Poincaré oscillator because all trajectories converge to the same limit cycle. Additionally, if the largest exponent is positive then the system is chaotic.

The most significant problem of calculating the Lyapunov exponents using this method is that the results are parameter-dependent, especially for  $s$  and  $z$ . Unlike the embedding procedure, there is no direct way of estimating suitable values for these parameters. Although in theory the condition  $s \rightarrow \infty$  should hold, in practice the Jacobians  $J(\mathbf{x}(t))$  contain a finite amount of data and are therefore approximations of the actual flow of the system. Consequently, the long sequence of multiplications in (47) can actually result in large errors. This means QR decomposition has to be applied in order to prevent ill-conditioned matrices [114,124].

A simple solution to the parameter-selection problem was found by Parlitz [126]. In this case the exponents are simply recalculated after reversing the time direction so that the time series  $x(t)$  becomes:

$$x_{\text{rev}}(t_i) = x(t_{N-i}). \quad (50)$$

The new Lyapunov exponents calculated from  $x_{\text{rev}}(t)$  should then be equal to the original ones but with the opposite signs. Exponents that do not change upon time reversal are spurious, meaning they have not been accurately calculated. The parameters of the method therefore need to be adjusted until this criteria is met [126], leading to an ‘optimal’ values for  $s$  and  $z$ . The only remaining problem is that in some cases the range of parameter values for which the calculation works is quite small, which makes the possibility of temporal bias is even stronger, i.e. it is not possible to recalculate for other timescales where the values of the exponents may differ.

### 6.3. Rényi dimensions

It is possible to relate the idea of thermodynamic states to the phase space of a dynamical system. As mentioned previously, the connection between entropy and the distribution of states is an important concept in both cases, particularly when defining complexity. However, a more convenient way of quantifying this property is to define a minimal dimensional space for which the system attractor can be projected on and still retain the same entropy. Following from (5), Rényi dimensions are defined as [127–129]

$$D_\alpha = \lim_{\epsilon \rightarrow 0} \frac{\frac{1}{1-\alpha} \ln \sum_i p_i^\alpha}{\ln \frac{1}{\epsilon}}, \quad (51)$$

where in this context  $p_i$  are the probabilities of the system being at various areas of width  $\sim \epsilon$ . This concept is analogous to the minimum dimension  $D$  of phase space for which the trajectories of the system do not overlap (which is essentially the

minimum dimension for which the attractor retains its deterministic property). Lower values of  $D_\alpha$  correspond to systems where the dynamics do not contain much information, while the opposite is true for large values. It is also worth noting that  $D_\alpha$  is a decreasing function with  $D_0 \geq D_1 \geq D_2$ . These dimensions are also directly related to the *fractal* nature of the attractor, or in other words the extent to which the states of the system fill phase space. This will be returned to in Section 7.

One of the most widely used phase space measures is the correlation dimension  $D_2$ . The relatively high value of  $\alpha$  in this case means that this measure is influenced primarily by the dense regions of phase space, which contain information from many trajectories around the attractor. In order to estimate  $D_2$  an algorithm was created by Grassberger and Procaccia [130,131] which can be applied directly to an embedded time series. Takens also proposed a probabilistic way of estimating  $D_2$  [132] although it has not been as widely used.

The algorithm of Grassberger and Procaccia begins with the computation of the correlation integral,

$$C(\varepsilon) = \frac{1}{N(N-1)} \sum_{i=1}^{n_v} \sum_{\substack{j=1 \\ j \neq i}}^{n_v} H(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \tag{52}$$

where  $H$  is the Heaviside step function and  $n_v$  is the number of embedding vectors. This integral is simply a sum of the number of “neighbouring” embedding vectors, defined as those that have an absolute distance  $\leq \varepsilon$ . The correlation dimension is then defined as:

$$D_2 = \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{d(\ln C(\varepsilon))}{d(\ln \varepsilon)}. \tag{53}$$

Due to the finiteness of real-world time series, the limit  $N \rightarrow \infty$  is difficult to satisfy and causes large errors for the small values of  $\varepsilon$  required for calculating  $D_2$ . The true value of  $D_2$  must therefore be estimated from the plateau where  $\frac{d(\ln C(\varepsilon))}{d(\ln \varepsilon)}$  stays constant as  $\ln \varepsilon \rightarrow 0$ . An additional source of complication is that  $C(\varepsilon)$  will be different if other values of  $d$  and  $l$  are used for the embedding. However, this is not necessarily a disadvantage as the procedure can simply be repeated for a range of  $d$  and  $l$ , using the region of convergence to find the most reliable estimate of  $D_2$ . While this procedure is sensitive to noise, it is extremely useful in distinguishing genuinely deterministic fractal dynamics from stochastic dynamics (for which convergence does not occur) and was one of the first methods to provide such a test. The calculation of the correlation dimension therefore became an important link between the theory of dynamical systems and actual experimental data.

The Hausdorff dimension  $D_H$  is another measure used to describe systems in phase space. It can be thought of as a general measure of how properties such as the information content and space-filling characteristics are related to the scale [133]. The Hausdorff dimension is commonly estimated using the box-counting method, with

$$D_0 = \lim_{\epsilon \rightarrow 0} \frac{\log N_{\text{box}}(\epsilon)}{\log(1/\epsilon)}, \tag{54}$$

where  $N_{\text{box}}$  is the number of boxes of side length  $\epsilon$  required to cover the trajectory of a system in phase space. However, more computationally efficient algorithms have also been developed [134].

A quantity which is related to the Rényi dimensions can also be derived from the Lyapunov spectrum. The Kaplan–Yorke dimension is defined as [135,131,136]:

$$D_{KY} = j + \frac{\lambda_1 + \dots + \lambda_j}{|\lambda_{j+1}|}, \tag{55}$$

where the Lyapunov exponents  $\lambda_i$  are ordered from largest to smallest and  $j$  is the largest integer for which  $\lambda_1 + \dots + \lambda_j \geq 0$ . This dimension is equal to the information dimension,  $D_1$ . It was shown by Grassberger and Procaccia [131] that  $D_{KY}$  provides an upper limit of  $D_H$  for the condition  $D_H = D_1 = v$ , where  $v = \frac{\log C(\varepsilon)}{\log(\varepsilon)}$  is the estimated value of  $D_2$ .

One of the benefits of calculating Rényi dimensions is that, in a similar way to the false nearest neighbours method, they can be used to estimate the appropriate embedding dimension  $d$ . The reason for this is that if  $d$  is smaller than  $D_\alpha$  then the Rényi dimension will simply take the value of  $d$ .  $D_\alpha$  can then be recalculated for higher embedding dimensions until they converge to the same result, where the lowest embedding dimension to converge to this result can be used for other analysis. Additionally, this convergence also gives confidence in the estimate of  $D_\alpha$ .

#### 6.4. Eigenvector decomposition

Although the output of the previously discussed decomposition methods is in the time domain, some methods also make use of phase space to separate the components of a time series. Decomposition by eigenvectors refers to a large group of methods which rely on transforming a time series to a matrix format and from that creating a basis of orthogonal modes which make up the time series. The technique can be applied in cases where several variables are known (including the widely-used independent component analysis [137]), in addition to the single-variable case. Since the problem of multivariate time series are not dealt with here, only a single variable case will be considered and specifically the Karhunen–Loève expansion [138,139].

Also known as principal value decomposition, this method is the foundation for other popular methods including the more recent singular spectrum analysis [140] and it is also strongly related to the embedding methods discussed in Section 6.1. The following matrix is first constructed in the same way as the embedding vectors in (39),

$$\mathbf{A} = \begin{bmatrix} x(t_1) & x(t_2) & \dots & x(t_N - (d-1)l\Delta t) \\ x(t_1 + l\Delta t) & x(t_2 + l\Delta t) & \dots & x(t_N - (d-2)l\Delta t) \\ \vdots & \vdots & \ddots & \vdots \\ x(t_1 + (d-1)l\Delta t) & x(t_2 + (d-1)l\Delta t) & \dots & x(t_N) \end{bmatrix}, \quad (56)$$

where  $d$  and  $l$  must be chosen with the same considerations as discussed before [141]. Across the horizontal direction this matrix simply represents a trajectory in  $d$ -dimensional phase space. The next step is to compute the correlation matrix, which gives the strength of correlation between different dimensions of the embedded time series,

$$\mathbf{C} = \mathbf{A}^T \mathbf{A}, \quad (57)$$

with eigenvectors  $\mathbf{v}_i$  and eigenvalues  $\sigma_i$  given by:

$$\mathbf{C}\mathbf{v}_i = \sigma_i \mathbf{v}_i. \quad (58)$$

The eigenvectors are used to form the rows of a new matrix  $\mathbf{V}$  and are ordered with respect to the size of the corresponding eigenvalues, i.e. so that the eigenvector with the largest eigenvalue makes up the first row of  $\mathbf{V}$ , the one with the second largest makes up the second row, and so on. The  $i$ th mode of the time series decomposition is then given by  $\hat{\mathbf{A}}_i = \mathbf{h}_i \mathbf{v}_i^T$  where  $\mathbf{h}_i$  is the  $i$ th row of the matrix

$$\mathbf{H} = \mathbf{A}\mathbf{V}^T \quad (59)$$

and  $\mathbf{V}_i$  is the corresponding row of the matrix  $\mathbf{V}$ . The number of modes is equal to the embedding dimension  $d$ , with the first mode  $\hat{\mathbf{A}}_1$  representing the component with the most energy in the time series.

The set of decomposed modes makes up an orthogonal basis which can be summed together to obtain a representation of the original time series, much like with the Fourier series. The reconstruction is representative of the time series provided that the embedding dimension chosen is greater than or equal to the number of non-zero modes. For the simple case of a sine wave, the number of non-zero modes required to reconstruct the time series is 2 since in phase space a sinusoidal oscillation appears as a 2D circle [141]. For more complex dynamics a greater number of modes is required. However, even when the chosen embedding dimension is smaller than the number of non-zero modes, the main characteristic orthogonal components of the time series should still be preserved [141].

One problem of the Karhunen–Loève expansion is that it requires a stationary time series. The eigenvalues  $\sigma_i$  are not time-dependent, which means the method has the same limitations as the Fourier transform. In particular, this means that whenever nonstationary components are present, the number of modes required to reconstruct the time series correctly is artificially large. A solution to this problem is instead to calculate the modes over a local window by replacing  $N$  in (56) with the window length  $L$ . This method acts in a domain analogous to the time–frequency domain and the same solution is in fact used in the short-time Fourier transform (see Section 5.2). However, since all of the modes must be calculated from the same matrix the window length  $L$  is constant for all modes, regardless of the oscillating frequencies involved. This forces us to make a choice between a small window that will reduce the effects of a nonstationary time series or a large window that will be able to preserve low frequency dynamics, but not both.

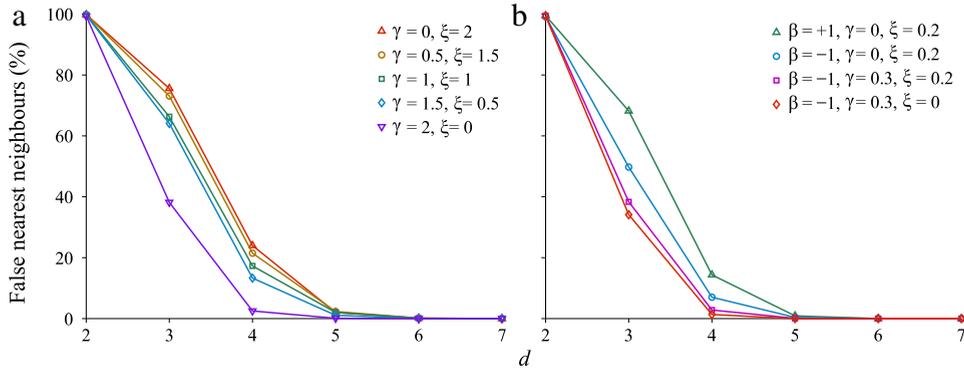
Another complication is that the number of modes is dependent on the embedding and, in particular, the embedding dimension must always be higher than the number of actual modes in the time series. For example, a stationary oscillation such as a sine wave appears as a limit cycle in phase space, meaning that 2 dimensions are required for the embedding. However, this will result in two sines in the final decomposition [141]. This means that the Karhunen–Loève decomposition is never a condensed transformation, but retains a significant amount of redundant information.

## 6.5. Analysis

### 6.5.1. Embedding

The embedding parameters were chosen for each time series using the first minimum of the mutual information for the time delay and the false nearest neighbours method for the dimension. In the latter case, the smallest dimension for which the percentage of false nearest neighbours is  $< 1\%$  was chosen.

Fig. 28(a) shows the false nearest neighbours for both the deterministic and stochastic non-autonomous Poincaré oscillator. When viewed at a specific point in time, as shown in Fig. 6(b), the non-autonomous attractor has a dimension  $\sim 1$ . This means that it would require an embedding dimension of  $d \approx 3$ . However, since the time delay embedding method removes all time-dependent information, the actual dimension estimated by false nearest neighbours is 5. The time-dependent components are therefore incorporated into the autonomous attractor in the same way as the stochastic dynamics in the other systems.



**Fig. 28.** Percentage of false nearest neighbours versus the embedding dimension for: (a) Poincaré oscillator under forcing as given in Eq. (19) with different ratios of the deterministic and stochastic non-autonomous terms; (b) Duffing oscillator as given in Eq. (18) with standard stochasticity ( $\beta = 1, \gamma = 0, \xi = 0.2$ ), stochastic resonance ( $\beta = -1, \gamma = 0, \xi = 0.2$ ), chaos ( $\beta = -1, \gamma = 0.3, \xi = 0$ ), chaos and stochasticity ( $\beta = -1, \gamma = 0.3, \xi = 0.2$ ). The calculation parameters used were  $R_T = 20$  and  $A_{tol} = 2$ .

The equivalent plot for the Duffing oscillator is shown in Fig. 28(b). Here the periodic forcing component actually reduces the required embedding dimension when the same stochastic forcing is added on top. This effect can be explained by the fact that points spread out more in phase space when the apparent dimension of the system is increased through the addition of a non-autonomous term. This means that  $R_d$ , as given in Eq. (43), has on average a larger value than when there is no periodic forcing. This effect can also be seen in the parameter  $\beta$ , where the existence of two potential wells for  $\beta = -1$  causes the points to spread out in the same way, so that the number of false nearest neighbours is smaller than the  $\beta = 1$  case for the same intensity of stochastic forcing. The fact that the simple limit cycle Duffing oscillator with added noise ( $\beta = +1, \gamma = 0, \xi = 0.2$ ) has the greatest percentage of false nearest neighbours shows that the required embedding dimension cannot be used to give an indication of the complexity of the system, or indeed of the dimension of the attractor.

### 6.5.2. Lyapunov exponents

Fig. 29 shows plots of the three Lyapunov exponents calculated from the attractors generated directly from the simulated data for the Lorenz system. Non-spurious exponents were found for  $z \approx 35$  ms with the number of steps  $s \geq 14$ . However, this value of  $s$  is much lower than what is required in practice for observed non-autonomous systems [122]. It is clear that the different numerical integration schemes do not affect the computation of the Lyapunov exponents. This reinforces the result of Zeni and Gallas [142] who investigated the value of the largest Lyapunov exponent of the chaotic Duffing oscillator integrated with the fourth-order Runge–Kutta scheme. In their analysis, a change in the integration step  $h$  had no effect on the exponents apart from the extreme case where  $h > 0.1s$ .

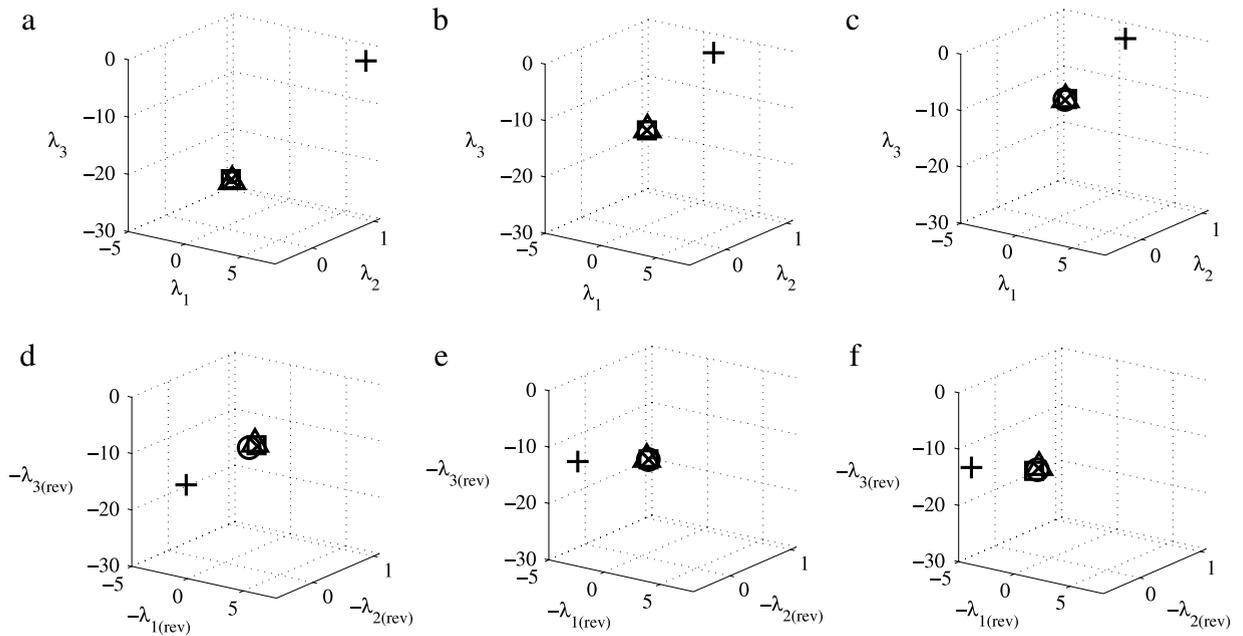
The Lyapunov exponents for the analogue Lorenz system in Fig. 29 appear far from the others because the system is contaminated with noise. In fact, for all the systems tested that contained a stochastic component, the exponents remained spurious for a wide range of parameter combinations. The reason for this is that the method relies strongly on the system being deterministic over short timescales. Even if chaos or a stable limit cycle is apparent over longer timescales, the method requires  $S \rightarrow \infty$  so that all of the Jacobians are calculated over small timesteps. This is where the effect of noise is most obvious so it becomes impossible to resolve the true Lyapunov spectrum.

Fig. 30 shows the evolution of the exponents in time for the non-autonomous Poincaré system for one set of parameters but it was found that for any choice of parameters the exponents would always be spurious at some periods in time. In this instance the method fails because the global exponent of the system is 0, which means the attractor is very thin when viewed locally. This makes the estimates of the Jacobians less accurate and also more susceptible to noise for measured time series [114]. The method is therefore biased against systems with either a zero or negative global Lyapunov exponent. This was clearly demonstrated when it was applied to the periodically-forced Duffing oscillator (Eq. (8)) which, despite having a non-autonomous component, worked well due to existence of a positive exponent and found three non-spurious exponents with values  $\lambda_1 = 1.57 \pm 0.14, \lambda_2 = 0.00 \pm 0.02$  and  $\lambda_3 = -3.49 \pm 0.11$ .

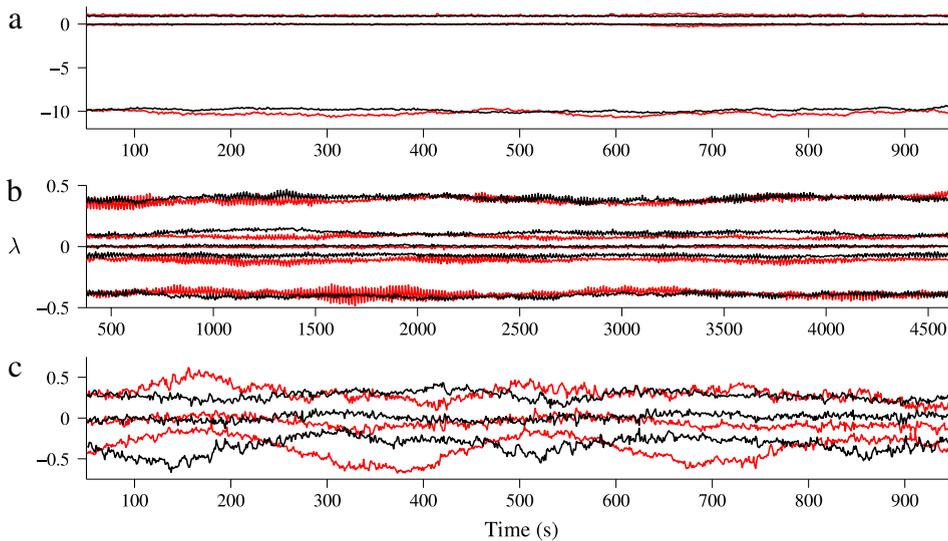
Surprisingly, Fig. 30 also shows that there is agreement between the forward-time and reverse-time Lyapunov exponents for the stochastic system, despite the fact that the method should not be able to calculate the exponents for noisy systems. However, this is more likely due to a problem with the testing procedure: stochastic systems are governed by a Markov process, which means that the same random dynamics is observed for whichever direction you travel in time. On the other hand, in the deterministic systems the direction of time does play a role in the dynamics, which could make it more difficult to find agreement between the forward-time and reverse-time Lyapunov exponents.

### 6.5.3. Correlation dimension

The correlation dimension  $D_2$  is usually estimated by calculation of  $\frac{d(\ln C(\epsilon))}{d(\ln \epsilon)}$  versus  $\ln \epsilon$  for increasing embedding dimension [143,144]. The plateau where these lines appear to converge is taken to be the estimate for  $D_2$ .

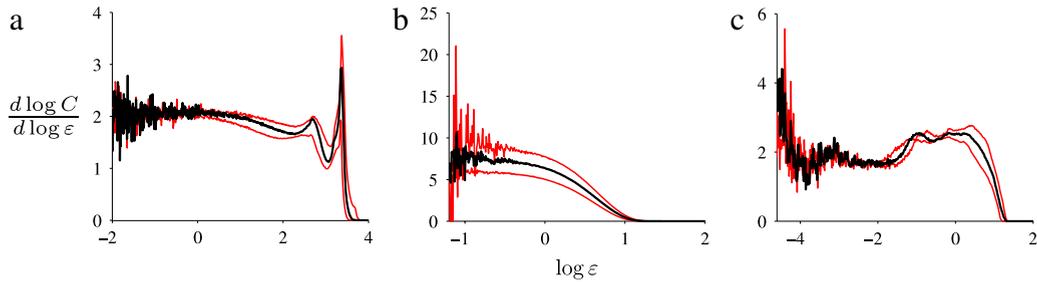


**Fig. 29.** Lyapunov exponents and reverse Lyapunov exponents for attractors reconstructed directly from the simulated trajectories for the Lorenz system with: (a,d)  $z = 0.1$  s; (b,e)  $z = 0.35$  s; (c,f)  $z = 0.5$  s. In each case  $s = 15/z$ . The integrated system using  $h = 10^{-3}$  s is given by O,  $h = 10^{-4}$  s by  $\square$ ,  $h = 10^{-5}$  s by  $\triangle$ , the Lie-integrated by  $\times$  and the analogue simulation by  $+$ . Noise contamination in the analogue case means that the exponents are spurious for each  $z$ , while for the other time series the exponents only match for  $z = 0.35$  s.

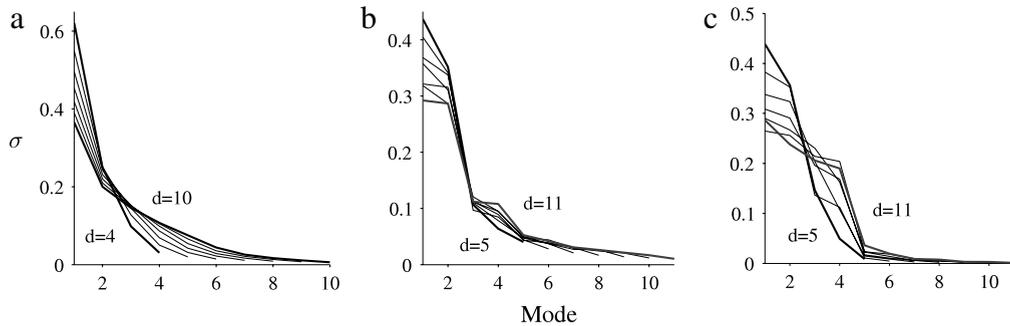


**Fig. 30.** Forward-time (black) and reverse-time (red) Lyapunov exponents for time delay embedded series: (a) Lorenz [ $s = 500$ ,  $z = 0.2$  s]; (b) Stochastic Duffing [ $s = 500$ ,  $z = 1.5$  s]; (c) Non-autonomous Poincaré [ $s = 200$ ,  $z = 0.5$  s]. The parameters  $s$  and  $z$  varied in each case due to the dependence on finding a fit between the forward-time and reverse-time exponents. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Fig. 31** shows that for the stochastic Duffing system the plateau in the correlation integral does not converge for different embedding dimensions. This is in contrast to the results of Osborne and Provenzale [145] which showed that stochastic systems have a finite and predictable value for the correlation dimension. However, this was later found by Theiler and others [146,147] to be a consequence of not having enough samples  $N$  in the time series and the analytical derivation shows that the correlation integral for pure  $1/f^\alpha$  noise scales as  $C(\varepsilon) \sim \varepsilon^d$ . This dependence on the embedding dimension occurs because the number of degrees of freedom in a stochastic time series is infinite, which means any finite embedding dimension is insufficient to reconstruct the true distribution of states and consequently will not give unique results.



**Fig. 31.** The derivative of the correlation integral for deterministic and stochastic test systems using the embedding dimension predicted by the false nearest neighbours algorithm (black lines) and  $\pm 1$  this dimension (red lines): (a) Lorenz; (b) Stochastic Duffing; (c) Non-autonomous Poincaré. The estimate for the Lorenz system is the most accurate, while for the stochastic system it is dependent on the embedding dimension. For the non-autonomous system it is significantly larger than the apparently one-dimensional attractor observed in Fig. 6(b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 32.** Eigenvalues from Karhunen–Loève decomposition for different embedding dimensions: (a) Lorenz system for  $d = 4, \dots, 10$  and  $\Delta t = 0.17$  s; (b) Stochastic Duffing system for  $d = 5, \dots, 11$  and  $\Delta t = 0.8$  s; (c) Non-autonomous Poincaré system for  $d = 5, \dots, 11$  and  $\Delta t = 0.31$  s. In each case the eigenvalues have been normalized by dividing by the sum over all of the eigenvalues.

It can also be seen in Fig. 31 that the plateau for the non-autonomous Poincaré system occurs at a value significantly greater than  $\sim 1$  (the approximate dimension of the time-dependent attractor). This shows that the embedding procedure has included the time-dependent components within the autonomous attractor, thereby increasing its apparent dimension.

### 6.6. Eigenvector decomposition

Despite the fact that Karhunen–Loève decomposition gives a final result which is represented in the time domain, the actual method still acts in phase space. This is clear from the plot shown in Fig. 32(a) where for the Lorenz system, which can be embedded in phase space without problems, there is very little dependence on the embedding dimension  $d$ . However, for the stochastic (b) and non-autonomous (c) systems, the distribution of eigenvalues changes in structure for each new embedding, resulting in kinks that cross over the previous distributions. It is therefore clear that both stochastic and non-autonomous components are included as higher-dimensional dynamics.

### 6.7. Summary

Phase space methods are an important tool in the time series analysis of classical complex systems which exhibit chaos. Indeed, there is no other domain which is able to provide more precise information about the dimension and (in)stability of a chaotic system. However, its application to other types of complex systems has significant drawbacks.

The fact that this type of analysis is based on deterministic systems makes it natural to expect that it would not perform well when applied to stochastic systems. If one only has finite data then it is not possible to reconstruct a phase space with the infinite amount of dimensions needed to analyse stochastic dynamics. In fact, for inverse problems there will almost always be some measured noise in the system, even for deterministic autonomous systems. This restricts the main use of phase space analysis to numerically-generated deterministic systems.

However, in addition to the problems encountered with stochastic systems, fully deterministic non-autonomous systems also appear very complex in phase space. Such systems acquire extra dimensions and the lack of a positive Lyapunov exponent means that the flow of trajectories is more subtle and difficult to capture statistically. Any analysis of non-autonomous systems in phase space should therefore cause some amount of distrust.

There are also more specific issues with phase space analysis. While not a direct form of bias against any of the complex systems, the method for calculating Lyapunov exponents from time series is still strongly geared towards chaotic systems.

In the case of non-chaotic systems, and in particular non-autonomous systems, the exponents become spurious and difficult to interpret. However, Lyapunov exponents are applicable to all types of deterministic systems. Despite this, the exponents of a non-chaotic, non-autonomous system are indistinguishable from those of a stochastic system.

Similarly, the decomposition of time series in phase space was shown to have a strong dependence on the embedding dimension for both stochastic and non-autonomous systems and it is consequently difficult to distinguish them from one another. Decomposition in phase space therefore offers no advantages over the time domain methods shown earlier.

## 7. Complexity analysis

Complexity, as defined here, differs from its previous usage which has been mainly applied when referring to the number of physical parameters and components required to define a system. Instead, complexity analysis deals with a quantity similar to the Rényi entropy, characterizing the amount of disorder or ‘roughness’ of a time series. As will be discussed later in this section, these two definitions of complexity are quite separate notions.

Two methods will be covered; firstly a statistical measure of complexity and afterwards a more in-depth form of complexity analysis which acts in the time domain by using windowing. The latter method has connections to the dimensionality analysis that is applied in phase space, so it will be possible to check for consistency between the different domains.

### 7.1. Hurst exponent

The Hurst exponent is a simple way of quantifying the complexity of a time series  $x$  by using statistical measures. It was introduced in 1951 by Harold Edwin Hurst [148] using the idea of the rescaled range,

$$R_{rs} = \frac{R}{S}, \quad (60)$$

where  $S$  and  $R$  are respectively the standard deviation of the time series and range of the integrated time series as given by:

$$S = \left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{1/2}, \quad (61)$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (62)$$

$$R = \max_{1 \leq i \leq n} y_i - \min_{1 \leq i \leq n} y_i, \quad (63)$$

$$y_i = \sum_{j=1}^i x_j - \bar{x}. \quad (64)$$

Hurst noted that as the segment length  $n$  of the time series increases,  $R_{rs}$  also tends to increase as a power law. Plotting  $\log(n)$  against  $\log(R_{rs})$ , the Hurst exponent is defined as the gradient of the line, or in equation form,

$$R_{rs}(n) = (cn)^H, \quad (65)$$

where  $c$  is a constant usually taken to be 0.5 [3]. Obviously, the estimated value of the Hurst exponent is dependent on the intervals of the time series used to define  $R_{rs}(n)$ . Before calculating the Hurst exponent, the value of  $R_{rs}(n)$  is therefore averaged from the intervals  $[1, n]$ ,  $[n + 1, 2n]$ ,  $\dots$ ,  $[N - n, N]$  across the whole time series [149].

A Hurst exponent in the range  $0.5 < H < 1$  indicates that the time series exhibits mostly persistent behaviour (if  $y$  has been increasing for some period it is likely to continue increasing) while if  $0 < H < 0.5$  this implies that the time series exhibits anti-persistent behaviour (if  $y$  has been increasing for some period it is likely to start decreasing). White noise, which is neither persistent nor anti-persistent, therefore has a value of  $H = 0.5$ . Note, however, that in many contexts it is common to perform analysis on the first time derivative of the data, which means an exponent of  $H = 0.5$  is sometimes attributed to Brownian motion as well [150].

### 7.2. Detrended fluctuation analysis

An area that is closely related to complexity analysis is *fractal* analysis. The idea of fractals is that the normalized small structure resembles the normalized large scale structure [151]. In the case of time series the scale refers to the size of the time period over which the data is observed. A high *fractal symmetry* means that the shape of the time series as defined by the high frequency fluctuations at small scales will appear similar to the shape as defined by the low frequency fluctuations at large scales. Simple time series such as a single sine wave have the least fractal symmetry, whereas noisier signals with many components have a high fractal symmetry, hence making it a good measure of complexity.

A widely used measure of fractal scaling is the fractal *dimension*,  $D_F$ , which is used in the same sense as the Rényi dimensions discussed earlier. It is defined as the exponent which dictates the number of similar structures  $n_s$  at a scale  $\epsilon$  contained within a larger similar structure,

$$n_s \propto \epsilon^{-D_F}. \tag{66}$$

High fractal dimensions therefore correspond to very ‘rough’ time series, where the repeated fractal structure contains a lot of detail.

The standard way of measuring fractal symmetry in a time series and to estimate  $D_F$  is with detrended fluctuation analysis (DFA) as introduced by Peng et al. [152]. This method has since been adapted to analyse the scaling of *cross*-correlations between different time series [153–155]. The goal of DFA is to find the self-similarity parameter  $\alpha$  as defined in the equation

$$y(t) \equiv a^\alpha y\left(\frac{t}{a}\right), \tag{67}$$

where  $y(t)$  is the time series integrated over time (calculated by the cumulative addition of successive points) and  $a$  is the scaling or magnification factor. This equation quantifies the fractal similarity by stating that the statistical properties for timescales equal to  $\frac{t}{a}$  can be made equal to those at the larger timescale  $t$  by simply multiplying with the factor  $a^\alpha$ . If no fractal symmetry exists within the time series then it is not possible to relate the amplitude scaling to time scaling as shown in this equation.

In order to estimate  $\alpha$  the time series is divided into sections of length  $n$ . For each section the local trend is removed either by subtracting a fitted polynomial or the trend from a moving average with window length  $n$  (in which case the method is referred to as detrended moving analysis (DMA)). As is always the case with polynomials, selecting the appropriate order is an issue although the convention has been to use a first order fit [152,3]. The root mean square *fluctuation* for the scale equal to  $n$  is then given by

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^N Y_n(t_i)^2}, \tag{68}$$

where  $Y(t)$  is the integrated and detrended time series. The fluctuation  $F(n)$  provides a measure of the amplitude at the corresponding scale which should follow the same scaling as in (67). By plotting  $\log F(n)$  against  $\log n$ , the self-similarity parameter  $\alpha$  is then given by the gradient of the line. However, this can only be calculated for straight sections of the line since curved sections indicate that fractal scaling is not present (i.e.  $\alpha$  is not constant but is in fact dependent on the scale).

The relationship between the DFA exponent and the fractal dimension of the time series is given by  $D_F = 3 - \alpha$  for  $2 \geq \alpha \geq 1$  [156]. Fractal scaling is present in so-called  $1/f$  noise, with Brownian ( $1/f^2$ ) noise having  $\alpha = 1.5$  and white noise ( $1/f^0$ ) having  $\alpha = 0.5$ .

The presence of fractal symmetry also results in an almost miraculous connection between the domains of analysis. In the frequency domain, as defined by the Fourier transform (21), the data scales according to the parameter  $\beta$  where  $F(\omega) \propto \omega^{-\beta}$ . The relation between  $\alpha$  and  $\beta$  is then simply  $\alpha = (1 + \beta)/2$  [157]. Therefore, despite the fact that  $1/f$  noise is defined by a variable with an infinite degree of freedom, the degree of freedom that relates to the actual invariant properties characterizing the time series in each of the domains is actually very small. The detection of fractal scaling is therefore very important in understanding the physics behind a time series and indicates a reduction in the number of physical parameters needed to describe the system.

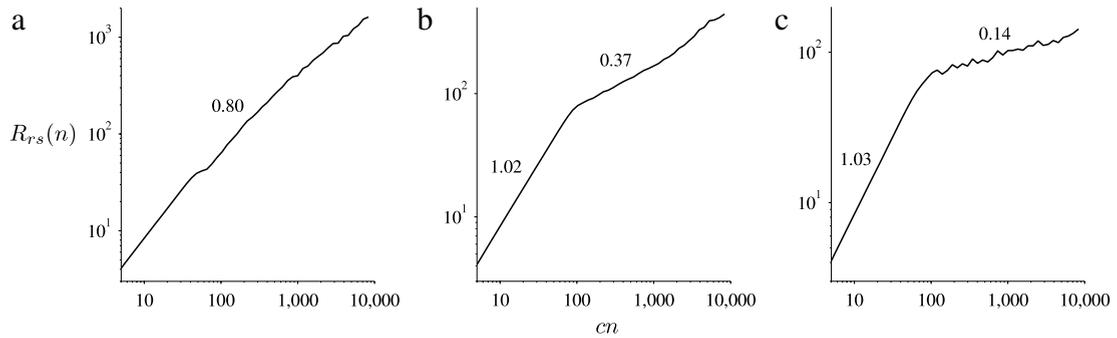
### 7.3. Multifractals

The methods discussed so far have been global measures of the complexity of the time series. This works well for monofractal cases where there is a single scaling law present throughout the time series. However, it is very rare to find a perfectly fractal time series, which was motivation for the development of a more general concept known as *multifractals* [158–161]. These describe cases where the scaling in the dynamics of the system is inhomogeneous, meaning that dynamics at different timescales follow different scaling laws so that

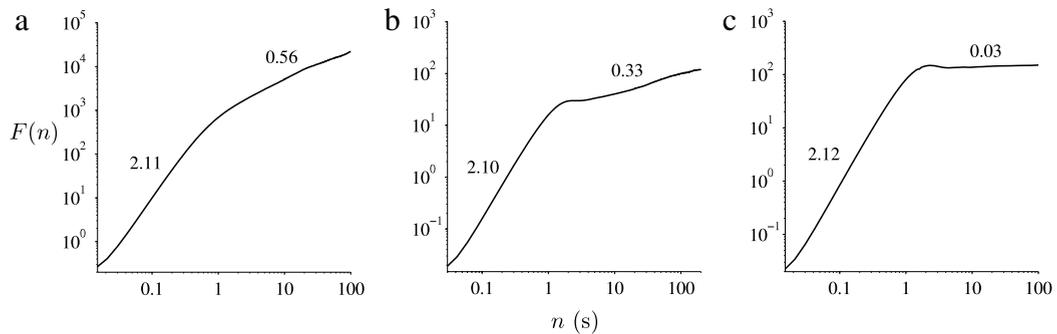
$$|x(t) - x(t_0)| \sim |t - t_0|^{\alpha_0}. \tag{69}$$

Such scaling laws have been used to help explain phenomena such as intermittency which occur at the smallest scales in turbulent fluid flows [97,98]. The scaling can also be measured in inverse problems using a generalized form of DFA [162]. Since the scaling exponents are localized in time, wavelets are ideal for analysis of multifractal time series in the time–frequency domain [163].

One problem with the multifractal framework is that dimensional estimates which give a broad range of values are taken to be an indication of multifractal dynamics [164]. However, this can also be attributed to the fact that some dimensional methods are unsuited to the system they are being applied to, as was found when determining the correlation dimension for the stochastic and non-autonomous systems. While such dynamics can be regarded as multifractal, a much more simple description of these systems exists.



**Fig. 33.** Calculation of Hurst exponents  $H$  for the time series from the three complex systems: (a) Lorenz system; (b) Stochastic Duffing system; (c) Non-autonomous Poincaré system. The numbers show the values of  $H$  calculated from linear fits of the slopes. The parameter  $c$  was taken to be 0.5.



**Fig. 34.** Detrended fluctuation analysis of time series from the three complex systems: (a) Lorenz system; (b) Stochastic Duffing system; (c) Non-autonomous Poincaré system. The numbers show the values for  $\alpha$  calculated from linear fits of the slopes ranging from 0.03–1 s and 3–100 s respectively.

#### 7.4. Analysis

The Hurst exponents of the Lorenz, stochastic Duffing and non-autonomous Poincaré systems were estimated from plots of the rescaled range  $R_{rs}$  against the segment length  $n$ , which are shown in Fig. 33. The line of the Lorenz system increases mainly linearly and the gradient corresponds to a Hurst exponent in the range  $0.5 < H < 1$ , indicating persistent dynamics. This is likely because the power spectra of the Lorenz time series contains a strong low-frequency contribution, which means it contains rises and falls over long timescales. However, for the stochastic and non-autonomous time series two distinct slopes can be seen. The slope occurring at small timescales corresponds to the main oscillatory component, which due to being smooth and persistent gives a similar value of  $H \sim 1$  in each case. The slopes at longer timescales relate to the actual complex dynamics of the systems and in both cases the exponents lie in the ‘anti-persistent’  $0 < H < 0.5$  range. Interestingly, the exponent of the non-autonomous system seems to indicate that the dynamics is more anti-persistent than those of the stochastic system, despite the fact that non-autonomous dynamics originate from a quasi-periodic function.

The DFA of the three time series is shown in Fig. 34. In this case two slopes are observed for all three of the systems. Again, the first slope that occurs at small timescales corresponds to the fundamental oscillatory motion of the system and hence gives a similar value of  $\alpha$  in each case. For the Lorenz system the value of  $\alpha$  for the second slope corresponds to a dimension higher than the one estimated using the correlation integral, since  $D_F \approx 2$  would imply  $\alpha \approx 1$ . For the stochastic Duffing system  $\alpha \ll 1$ , which means that the dimension is not defined since it lies outside the range  $1 \leq D_F \leq 2$ . However, for the non-autonomous Poincaré system  $\alpha \approx 0$ , which means the dimension certainly does not lie in the range  $1 \leq D_F \leq 2$ , despite the fact that the time-dependent attractor in Fig. 6(b) appears to have a dimension  $D_F \approx 1$ . This is due to the fact that, as with the phase space methods, DFA is unable to distinguish between dynamics resulting from purely time-dependent components in the system and those of the underlying low-dimensional system. A review of these effects is provided in [165].

#### 7.5. Summary

Complexity analysis shares many similarities to the determination of attractor dimensions in phase space. However, since it is performed in the time (and potentially frequency) domain this somewhat simplifies the analytical procedure, ensuring that there is no dependence on embedding parameters.

A problem with complexity analysis is that despite the fact that it uses the statistics of fluctuations in the time domain, it offers no time-dependent measures of the systems. Both the Hurst and DFA exponents for non-autonomous systems are still

influenced by the time-dependent components. Therefore, as with phase space and frequency domain analysis, complexity analysis is unable to offer a useful interpretation of a non-autonomous system.

For the chaotic and stochastic systems the information provides at least some insight into the dynamics. For the chaotic system DFA appears to be a more useful tool than the Hurst exponent as the detection of two scaling laws distinguishes the system from a purely random Markov process. The estimates of the fractal dimension offered by DFA is perhaps not as accurate as phase space estimators but it does seem to show that the chaotic system has a lower, finite dimension compared with the stochastic system.

## 8. Interactions

The methods up to this point have dealt with the analysis of one system with a single-variable time series. However, there exist a whole host of other time series analysis methods which are based around understanding the interactions between separate systems.

In this section the consequences of applying these methods to the various types of complex systems will be considered. However, note that the emphasis is still on single-variable time series, only that now there exists a separate time series for each of the interacting systems.

In the first part of this section the various observable interaction phenomena are defined. Phase dynamics methods are then introduced for the analysis of interacting systems where the phases can be defined. Following this, the recently developed Bayesian inference approach is discussed, in addition to time–frequency domain methods and an approach based on information theory.

### 8.1. Background

#### 8.1.1. Couplings

The precursor to any interaction phenomena is a transfer of information or *coupling* between two systems. From a modelling perspective, couplings can be defined as terms in the equations for one system which contain variables of another system. For example, two Poincaré oscillators can be coupled in the following way:

$$\begin{aligned}\frac{dx_1}{dt} &= -q_1x_1 - (\omega_1 + \gamma y_2)y_1 \\ \frac{dy_1}{dt} &= (\omega_1 + \gamma y_2)x_1 - q_1y_1 + \gamma(y_1 - y_2)^2 + \xi_1\eta_1(t) \\ \frac{dx_2}{dt} &= -q_2x_2 - \omega_2y_2 \\ \frac{dy_2}{dt} &= \omega_2x_2 - q_2y_2 + \xi_2\eta_2(t)\end{aligned}\tag{70}$$

where  $q_{1,2}$  are defined as in (10) and  $\eta_{1,2}$  are Gaussian white noises. Here, the coupling is given by a nontrivial quadratic term and is unidirectional. In addition, it is manifested in both the  $y$ -dimension and as a modulation of the oscillator frequency. The addition of noise terms makes this type of coupling very difficult to detect using the methods described in the previous sections. Instead, methods that are formulated specifically around detecting interactions are needed.

#### 8.1.2. Phases and protophases

Much of the analysis relies on the separation of ‘amplitude’ dynamics and ‘phase’ dynamics, as discussed previously in Section 5.1. Here, the definition of phase used is a generalization of 1-dimensional phase  $\phi = \omega t + \theta$  used in simple harmonic oscillators, where  $x(t) = A \cos(\phi(t))$ . In phase space this can be thought of as a transformation to polar coordinates, where one coordinate (the radius) corresponds to the amplitude dynamics, while another coordinate (the angle) corresponds to the phase dynamics.

However, this definition of phase assumes that the intra-cycle shape of the oscillations is harmonic. When such a model is applied to nonlinear dynamical systems the result is not the actual phase but rather a *protophase* [166,167]. Such protophases are not unique, which means that linear methods such as the Hilbert and wavelet transform can give different results.

The main condition for any protophase is that it grows monotonically, gaining  $2\pi$  with every cycle of the oscillator. In phase space this condition implies a transformation where all trajectories of the system orbit the fixed point in the same direction.

#### 8.1.3. Synchronization

One of the best-known phenomena observed in interacting systems is synchronization, which refers to when two oscillators are locked together in a relationship between their phases or amplitudes [1]. Synchronization does not necessarily occur whenever a coupling is present but it is a clear indicator of interaction. Its measures are also separate from those that only detect couplings.

One of the simplest forms is ‘complete’ synchronization where the coupled systems are identical but do not necessarily have the same initial conditions. In this case, the systems become synchronized when their states are the same, after which point the coupling terms vanish as the systems no longer need to be influenced by each other to remain synchronized. Another type is ‘generalized’ synchronization, which occurs when the state of one system can be determined directly via a functional relationship to the state of the other. This typically occurs when the influence of a unidirectional coupling from one system to the other causes the maximal Lyapunov exponent of the system to become negative. The other main type of synchronization is ‘phase’ synchronization, which is defined as the case where the phases of the two systems are locked in an  $n : m$  relationship so that the phase difference  $|n\phi_1 - m\phi_2|$  is constant, while the amplitudes may still vary independently.

It should be noted that the following will cover *phase synchronization* only. The motivation for considering only one scenario is because amplitude synchronization can often be reduced to phase synchronization in real world systems. In particular, many systems can be modelled as ensembles of interacting oscillators with negligible amplitude dynamics, where only phase dynamics is relevant [5]. Amplitude effects in these systems can be considered as a result of resonance when the oscillators become phase synchronized. In other cases, amplitude dynamics can sometimes be separated into additional oscillatory components, which can then be investigated individually for phase synchronization.

The concept of synchronization in complex systems might seem contradictory, but it is in fact well understood for coupled chaotic systems [168–170]. In such coupled systems it is also possible for synchronization to occur between individual oscillatory components (i.e. *locally* in frequency).

#### 8.1.4. Phase coherence

In light of analysis performed in the time–frequency domain, the concept of synchronization can be applied to phases extracted from the individual components in time series. However, these phases do not necessarily correspond to the actual phases of the oscillators, so the usage of the term ‘synchronization’ is not entirely correct. Instead, for a given frequency  $\omega$  the *phase coherence* is defined [17],

$$\Pi(\omega) = \frac{1}{N} \left| \sum_{n=1}^N e^{i(\phi_1(\omega, t_n) - \phi_2(\omega, t_n))} \right|, \quad (71)$$

where  $\phi_1(\omega, t_n) - \phi_2(\omega, t_n)$  is the phase difference between the oscillatory components of the same frequency from the two time series at the time  $t_j$ . When the phase difference of the oscillations in the time series remains constant for all time (i.e. the oscillations are coherent) then  $\Pi(\omega) = 1$ . Similarly, if  $\Pi(\omega) = 0$  this means that there is no tendency to preserve a particular phase difference.

At the lowest detectable frequencies  $\Pi(\omega) \rightarrow 1$ , regardless of the amount of coherence there might be, simply because the time average in (71) contains fewer observable oscillations. For example, if a single cycle of some oscillation was observed at the same time and frequency then the coherence would be high. It is only when subsequent cycles are observed that the phase coherence takes a meaningful value. It is therefore necessary to set the cutoff for the lowest frequency to be several times higher than the lowest *observable* frequency.

## 8.2. Phase dynamics methods

### 8.2.1. Synchrogram

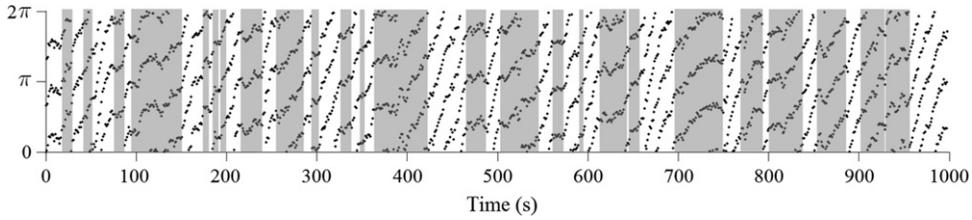
Synchrograms are visual representations of synchronization in its simplest sense. They involve extracting marked events for one oscillator and the continuous phase for the other oscillator. The extracted phase can be “wrapped” over finite cycles, so that it returns to 0 immediately after  $2\pi$ . The synchrogram is then a 2D scatter plot of this wrapped phase against the times of the marked events from the other oscillator.

Synchronization can be defined in the synchrogram as the periods when the phases of the two oscillators are stationary with respect to each other, which means that the marked events occur repeatedly at the same points in the wrapped prophase of the second oscillator. On the synchrogram these periods appear as horizontal lines, with the number of lines corresponding to the type of synchronization, e.g. for 1:1 synchronization one line would be seen per cycle, while two would be seen for 2:1 synchronization. Synchronization that is not 1:1 is more difficult to detect however, since it can only be observed as straight lines when the prophase is wrapped over a number of cycles equal to a multiple of the corresponding synchronization number, e.g. 10:9 synchronization would require wrapping the phase over multiples of 9 cycles (0 to  $18\pi$ , 0 to  $36\pi$ , 0 to  $54\pi$  etc.).

### 8.2.2. Synchronization index

While the synchrogram provides a simple and effective means of detecting synchronization, it does not provide a quantifiable measure and the times for the onset or loss of synchronization are purely subjective. Therefore, in order to carry out statistical hypothesis tests it is necessary to use a synchronization index, or more specifically, the *phase synchronization index* (to distinguish it from measures of amplitude synchronization).

The synchronization index was introduced by Tass et al. [171] and can be defined using the conditional probability for the phases of the two oscillators. The method involves dividing the cycle of one oscillator into  $N$  bins and integrating the



**Fig. 35.** Synchrogram of the coupled Poincaré oscillators given in (70). The grey shaded areas indicate periods of significant 3:1 synchronization, corresponding to a synchronization index calculated using a 20 s window with 50% overlap, 8 bins and 100 IID surrogates. The parameters used were  $\gamma = 1$ ,  $\omega_1 = 5\pi$  (2.5 Hz),  $\omega_2 = 5.3$  (0.84 Hz) and  $\xi_1 = \xi_2 = 0.2$ .

complex phase of the other oscillator over each bin, which in discrete form yields the following:

$$r_l = \frac{1}{M_l} \sum_n e^{i\phi(t_n)}, \quad (72)$$

where  $1 \leq l \leq N$  designates the bin,  $t_n$  are the times when the oscillator is in the part of the cycle allocated to the bin and  $M_l$  is the total number of points in the bin. If the same section of the phase  $\phi(t)$  for the other oscillator coincides with the same division of the cycle in each oscillation (i.e. the two oscillators are synchronized) then  $|r_l| = 1$ . The average over the bins is then calculated to give the synchronization index:

$$\gamma = \frac{1}{N} \sum_{l=1}^N |r_l|. \quad (73)$$

One of the disadvantages of  $\gamma$  is that it can only be used as a measure for one type of synchronization. To be able to quantify  $n : m$  synchronization, bins need to be allocated for parts of  $n$  sequential cycles for one oscillator while the phase of the other oscillator would need to be wrapped between 0 and  $2m\pi$ . As with the synchrogram, without *a priori* knowledge of  $n$  and  $m$  different combinations have to be tried separately.

To observe how the amount of synchronization changes with time a natural step is to calculate  $\gamma$  within a sliding window. The binning method can only detect synchronization if points from at least two 0 to  $2m\pi$  cycles of  $\phi(t)$  are present in each bin. This means that the window length needs to be a minimum of twice the period of the  $n : m$  cycle. However, larger windows also increase the number of points in the bins, which is important for reducing the effect of when the window length does not contain an integer number of  $n : m$  cycles. This is especially relevant to oscillators with characteristic frequencies that vary in time.

Since it is rare to obtain  $\gamma = 1$  due to the issues just mentioned, deciding what values constitute synchronization is a non-trivial task. A way to deal with this problem is to use *surrogate data* [3,171], which is covered in Appendix D.2. Using the two time series under investigation, two surrogate time series can be generated with uncorrelated phases but otherwise resembling the original ones. Calculating the synchronization index by using the same procedure for the surrogate time series provides a realization of the baseline for  $\gamma$ . This can be repeated for different surrogate time series to produce a distribution  $\gamma_{\text{sur}}$  of these baseline results. If this distribution is approximately Gaussian then  $\bar{\gamma}_{\text{sur}} + 2\sigma_{\text{sur}}$  provides a reasonable limit, where  $\bar{\gamma}_{\text{sur}}$  is the mean of  $\gamma_{\text{sur}}$  and  $\sigma_{\text{sur}}$  is the standard deviation, above which the phases can be said to be synchronized. However, it should be noted that surrogate data are only useful when oscillators have variable frequencies. For oscillators that are synchronized and have fixed frequencies for a significant period in time, no matter how the phases of the surrogates are generated, the probability of the surrogate phases being synchronized over a time interval will be much higher. The surrogate level also increases as the time window is made smaller and eventually reaches a point where  $\bar{\gamma}_{\text{sur}} + 2\sigma_{\text{sur}} > 1$ . The method is therefore only applicable when the oscillators are synchronized for intervals which allow for a long enough window to reduce the surrogate level but also when they are not synchronized at the same ratio for somewhat longer periods in time.

Fig. 35 shows a synchrogram and synchronization index for the case of coupled Poincaré oscillators. Due to the frequencies chosen, the most likely synchronization ratio is 3:1 and this can be seen in the synchrogram from the three lines that dominate throughout time. The synchrogram shows synchronization in intervals lasting from a few seconds to around 10 s. On the other hand, the synchronization index shows much longer periods of synchronization of up to around 60 s. In this case the horizontal lines that appear to wiggle in the synchrogram are counted as synchronization because of the discrete bin sizes used. In addition, despite the fairly small 20 s window used, the synchronization index cannot always keep up with the rapid changes between the synchronized and non-synchronized states and these instead appear as a longer epoch of continuous synchronization.

### 8.2.3. Direction of coupling

While the detection of a coupling between two systems is useful, the detection of the *direction* of that coupling is a more profound measurement. The following methods extract information from the phase of the two systems in order to find this

direction of coupling. In general, they rely on the principle that any coupling should be detectable as a causal relationship between the two phases  $\phi_1$  and  $\phi_2$ . This means that if the direction of coupling is  $\phi_1 \rightarrow \phi_2$  then the influence of  $\phi_1(t)$  should be observable in  $\phi_2$  at a future point in time, i.e. in  $\phi_2(t + \tau)$ .

The first method is known as the evolution map approach (EMA). It starts by defining a delayed time derivative of the unwrapped phase for both oscillators,

$$\Delta_{1,2}(t_k) = \phi_{1,2}(t_k + \tau) - \phi_{1,2}(t_k), \quad (74)$$

where  $\tau$  is a fixed time interval. Without any external influence or couplings,  $\Delta_{1,2}(t_k) = \omega\tau$  is expected, where  $\omega$  is the frequency of the oscillator. Of course in real systems this is rarely the case so further analysis that takes into account causality is needed. When the influence of the couplings is added the expression becomes:

$$\Delta_{1,2}(t_k) = \omega\tau + F_{1,2}(\phi_{2,1}(t_k), \phi_{1,2}(t_k)). \quad (75)$$

The term  $F_{1,2}$  that originates from the couplings can be estimated by fitting a Fourier series  $F_{1,2} = \sum_{m,l} A_{m,l} e^{im\phi_1 + il\phi_2}$  to  $\Delta_{1,2}(t_k) - \omega\tau$  using the least squares algorithm. Using a periodic basis for the fitting (rather than a polynomial or otherwise) is required because the influence from the oscillators is expected to be periodic in the phases.

If no dependence on  $\phi_2$  exists then the amplitudes  $A_{m,1}$  of  $F_1$  should be uncorrelated. Hence, if there is no coupling from  $\phi_2$  to  $\phi_1$  then there will be no significant change in  $F_1$  with respect to  $\phi_2$ . A coupling from  $\phi_2$  to  $\phi_1$  is therefore detected when  $\frac{\delta F_1}{\delta \phi_2}$  diverges significantly from 0. The coupling coefficients for each direction can therefore be defined as follows:

$$c_{1,2}^2 = \int \int_0^{2\pi} \left( \frac{\delta F_{1,2}}{\delta \phi_{2,1}} \right)^2 d\phi_1 d\phi_2. \quad (76)$$

The *directionality index* is then

$$d = \frac{c_2 - c_1}{c_1 + c_2}, \quad (77)$$

where  $d_1 = 1$  for  $\phi_1 \rightarrow \phi_2$  coupling,  $d_1 = -1$  for  $\phi_2 \rightarrow \phi_1$  coupling and  $d_1 = 0$  corresponds to symmetric bidirectional coupling.

Another way to quantify directionality is by taking the norm of the Fourier coefficients  $A_{m,l}$  directly [77,167]:

$$c = \frac{\sqrt{\sum_{m,l} |A_{m,l}|^2}}{A_{0,0}}, \quad (78)$$

where the summation contains all of the components apart from  $A_{0,0}$  (the frequency of the oscillator). The new coupling coefficients can then be inserted into (77) to give the directionality index. This somewhat simpler calculation will be adopted here for the calculation of directionality.

The instantaneous period approach (IPA) uses the same algorithm, except that the incremental phase is replaced by the Poincaré return times. The definition of a return to a reference phase  $\phi(t_0)$  is defined as when  $\phi(t_r) = \phi(t_0) + 2\pi$ , where  $t_r$  is the return time which can be estimated from the points  $\phi(t_r) - \phi(t_0) < 2\pi$  using interpolation [172]. Eq. (75) then becomes,

$$t_r = \bar{t}_r + \Theta(\phi_1, \phi_2), \quad (79)$$

where  $\bar{t}$  is the mean return time and  $\Theta(\phi_1, \phi_2)$  is the corresponding coupling function which can again be estimated using Fourier series. The advantage of IPA over EMA is that no parameter such as  $\tau$  needs to be defined, which means the end result is unique.

Another alternative is the mutual prediction approach (MPA). In this case a point  $\phi_1(t_k)$  in the unwrapped phase is chosen, and the information provided in  $\phi_1(t)$  for all  $t$  is then used to make a prediction of the value of  $\phi_1(t_k + \tau)$ . The error  $E_1(t_k) = |\phi_1'(t_k + \tau) - \phi_1(t_k + \tau)|$  is then calculated, where  $\phi_1'(t_k + \tau)$  is the prediction.

The idea of a mutual prediction is to predict the value of  $\phi_1(t_k + \tau)$  using both the information provided by  $\phi_1(t)$  and  $\phi_2(t)$ . The error for this prediction, denoted  $E_{1,2}(t_k)$ , can then be compared with  $E_1(t_k)$  to see if the prediction is different. If  $E_{1,2}(t_k) < E_1(t_k)$  then it is apparent that  $\phi_2(t_k)$  is providing information about the  $\phi_1(t_k + \tau)$ , implying that a  $\phi_2(t) \rightarrow \phi_1(t)$  coupling exists. Likewise, the condition  $E_{1,2}(t_k) < E_2(t_k)$  implies that a  $\phi_1(t) \rightarrow \phi_2(t)$  coupling exists.

As has been seen with other methods, the Fourier basis used for  $F_{1,2}$  does not include any time dependence, which inevitably means it will perform poorly when applied to time-varying couplings. In addition, while the choice of the parameter  $\tau$  has little effect on the value of  $d_{(1,2)}$  for the constant couplings and frequencies used in the original papers (its value is usually taken as the smallest period of the oscillations or the average of the periods [172]), time-varying parameters may make it difficult to choose an optimal value.

### 8.3. Bayesian inference approach

As just mentioned, a problem with the previous methods is that they are not formulated to account for time-dependent couplings. There is also no straightforward way to apply a moving time window because windowing means that a smaller data series goes into the estimate of the couplings, increasing the uncertainty.

The Bayesian theorem offers a solution to this windowing problem. It was formulated in the work of Thomas Bayes, published posthumously in 1763 [173]. It has since been widely applied to inverse problems where one would like to infer parameters related to the generation of a data set [174–178,77]. The theorem is summarized in

$$P(\mathcal{M}|\mathcal{X}) = \frac{P(\mathcal{X}|\mathcal{M})P_{\text{pr}}(\mathcal{M})}{P(\mathcal{X})}, \tag{80}$$

where  $P(\mathcal{X}|\mathcal{M})$  is the conditional probability of observing the data  $\mathcal{X}$  given the hypothesized parameters  $\mathcal{M}$ .  $P_{\text{pr}}(\mathcal{M})$  is the probability of  $\mathcal{M}$  before observing the data  $\mathcal{X}$  and

$$P(\mathcal{X}) = \int P(\mathcal{X}|\mathcal{M})P_{\text{pr}}(\mathcal{M})d\mathcal{M} \tag{81}$$

is the marginal probability of  $\mathcal{X}$ .  $P(\mathcal{M}|\mathcal{X})$  is known as the *posterior* probability and gives the probability that the hypothesized parameters are correct given  $\mathcal{X}$  and the *prior* probability  $P_{\text{pr}}(\mathcal{M})$ .

The most likely combination of values for the parameters for a single window of data is inferred by locating the stationary point in the negative-log likelihood function, known as maximum likelihood estimation. In this case the likelihood function is defined for the phases of two systems defined by the following stochastic differential equations [179,77],

$$\frac{d\phi_{1,2}}{dt} = \omega_{1,2} + F_{1,2}(\phi_{1,2}) + G_{1,2}(\phi_1, \phi_2) + \xi_{1,2}(t), \tag{82}$$

where  $F_{1,2}(\phi_{1,2})$  and  $G_{1,2}(\phi_1, \phi_2)$  are coupling functions which, as in the previous methods, are modelled using a Fourier basis. The parameters  $c_k$  for this basis are eventually inferred in a covariance matrix denoted  $\mathcal{E}$ . By making use of Bayes' theorem, the posterior covariance matrix for the previous window can be made the prior covariance matrix  $\mathcal{E}_{\text{prior}}$  for the current window. Hence, information is allowed to propagate between windows, allowing the inferred parameters to become more accurate with time [77].

However, the inference only improves if the parameters do not vary in time. To account for changes in the values of the parameters, the prior must take the form of a convolution between the posterior of the previous window and a diffusion matrix which describes the change in  $c_k$  [77]. The standard deviation corresponding to the diffusion of the parameters is assumed to be a known fraction of the parameters themselves,  $\sigma_k = pc_k$ , where  $p$  is known as the propagation constant. This modification allows the method to track the change in the couplings over time.

A tutorial for the implementation of this Bayesian-based approach is provided in [180]. A Matlab toolbox is also provided by the authors.

Fig. 36 shows the method applied to the two coupled Poincaré oscillators, although this time with time-varying parameters. For the times  $t = 125$  s and  $t = 575$  s when the coupling is strong the directionality indices were found to be  $d = -0.79$  and  $d = -0.64$  respectively, giving the correct direction of coupling. For the times  $t = 325$  s and  $t = 775$  s when the coupling is not present the indices were found to be  $d = 0.01$  and  $d = 0.21$  respectively. Notice how in the second case noise causes a hint of directionality due to the fact that the index used is only dependent on the ratio between  $c_1$  and  $c_2$ . It is therefore important to consider the amplitude of the coupling functions themselves, which is much smaller when there is no coupling. It is also interesting to note that the shape of the functions is spurious when no coupling exists but almost identical when the coupling is the same, despite being derived from completely separate time windows ( $F_1$  in (c) and (e)).

### 8.4. Time–frequency approach

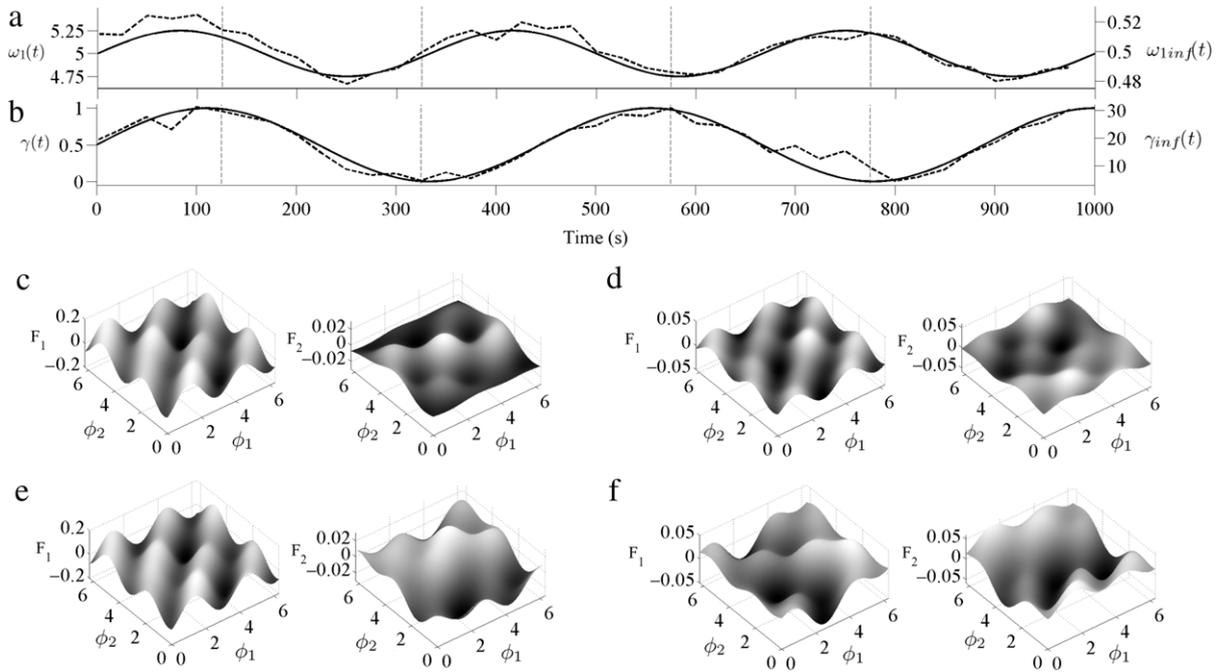
#### 8.4.1. Wavelet phase coherence

The two important quantities in (71) are the time and frequency that the phases are localized around. The problem of extracting these phases is therefore the same time–frequency resolution issue as discussed in Section 5. As before, the wavelet transform is the optimal solution which for the Morlet wavelet (Eq. (37)) has the complex form,

$$W_T(s, n) = a_{s,n} + ib_{s,n} = A_{s,n}e^{i\phi_{s,n}}, \tag{83}$$

where  $A_{s,n}$  is the amplitude of the wavelet transform and  $\phi_{s,n} \equiv \phi(\omega, t_n)$ . The method, first applied in the early 2000s [181–183], therefore involves taking the wavelet transform of each time series and calculating the coherence scale by scale. However, even when the phase coherence is calculated for two noise signals there is some level of coherence. Additionally, this baseline coherence is not constant for all scales but increases when moving lower in frequency.

The bias towards lower frequencies can be accounted for by using *surrogates* of the time series. These are formulated from the idea of hypothesis testing, where the surrogate is designed to preserve all of the properties of the original time series



**Fig. 36.** Detection of the direction of coupling for two Poincaré oscillators using Bayesian inference: (a) Variation of the frequency parameter of the first oscillator in time. (b) Variation of the coupling from the second oscillator to the first oscillator in time. In each case the dashed lines show the inferred parameters  $\omega_{inf}(t)$  and  $\gamma_{inf}(t)$ . (c–f) Coupling functions for the windows centred on  $t = 125$  s,  $t = 325$  s,  $t = 575$  s and  $t = 775$  s respectively. Parameters for the second oscillator were kept constant, with  $\omega_2 = 5.3$  and  $\varepsilon_2 = 0$ . The parameters for the method used a window length of 50s and propagation constant  $p = 0.8$ .

apart from the property that is being tested (see Appendix D.2). In the case of phase coherence, the desired null hypothesis is that the phases in the time series are independent for all frequencies, which means it is the time–phase information that needs to be randomized in the surrogates.

In the same way as is done for the synchronization index, the phase coherence can be calculated over time by using a sliding window along each scale. The time-localized phase coherence used by Sheppard et al. [184] provides a more normalized measure of coherence since the windows can be scaled appropriately with frequency so that the window will always contain the same number of cycles. Additionally, the work by Sheppard et al. has provided a rigorous testing procedure for *significant* phase coherence which uses a theoretical prediction of the bias mentioned as well as a test based on surrogate data.

A disadvantage of using wavelet phase coherence is that the low frequency limit is rather high. By starting with the continuous wavelet transform there is a reduction in the part of the time series that is observable at low frequencies because of the cone of influence. To add another complication, time domain smoothing is often required to remove the high harmonics of low frequencies that are not observable, which reduces the observable part of the time series even further as a result of losing data at the edges.

Another limitation of the method is that it can only be used to detect 1:1 synchronization between wavelet scales; there is no direct information about the interactions between nonlinear systems. The method is therefore mainly used for detecting similarities between two time series in the time–frequency domain, which can then be used as indirect evidence of interactions between the two systems that produced the time series.

#### 8.4.2. Wavelet bispectrum

If a synchronization-like measure can be defined locally in frequency then it is not a large leap to think that it should also be possible to detect couplings in a similar way. The corresponding measure is known as the *bispectrum*.

The bispectrum was introduced by Hasselmann et al. [185] to investigate nonlinear interactions in ocean waves. It arises from high-order statistics, which categorizes time series analysis methods with the corresponding measures such as the mean (first-order), variance (second-order) and skewness (third-order) [186]. For example, the autocorrelation function and Fourier transform are second-order measures in the time and frequency domain respectively. The bispectrum can be thought of as an estimate of the third-order in the frequency domain and describes the nonlinear (or more specifically, quadratic) properties of the time series. From the perspective of nonlinear dynamical systems, the bispectrum is able to provide a measure of the coupling between two oscillations in a time series.

The bispectrum can be derived from the Fourier spectra as an estimate of the statistic  $M_3$ , known as the third-order moment [187],

$$\hat{M}_3(f_1, f_2) = F(f_1)F(f_2)F^*(f_3), \tag{84}$$

where  $F$  corresponds to the Fourier transform of the time series and  $f_3 = f_1 + f_2$ . To reduce the impact of time-dependent effects (either stochastic or due to non-autonicity), the series is divided into  $K$  sections in time and the average moment is calculated [188]:

$$\hat{B}(f_1, f_2) = \frac{1}{K} \sum_{k=1}^K \hat{M}_3(f_1, f_2). \tag{85}$$

In the same way as the Fourier transform, the bispectrum is a complex quantity and can therefore be divided into a biphase  $\phi$  and biamplitude  $A$  where  $\hat{B} = Ae^{i\phi}$ . The biamplitude provides a direct measure of the quadratic coupling strength while the biphase carries little importance in the standard Fourier bispectrum.

Before moving further, it is important to note a consequence of dealing with the additional frequency  $f_3 = f_1 + f_2$ . As both  $f_1$  and  $f_2$  approach the Nyquist frequency  $f_s/2$  the bispectrum becomes meaningless because  $f_3$  starts to take amplitude and phase information at frequencies which are outside the observable range. Couplings at the highest frequencies are therefore not detectable, meaning an ‘effective’ Nyquist frequency for the bispectrum can be defined as the line from  $f_1 = f_s/4$ ,  $f_2 = f_s/4$  to  $f_1 = f_s/2, f_2 = 0$  (as well as the redundant case where  $f_1$  and  $f_2$  are switched).

A drawback of the bispectrum is that it is directly dependent on the amplitudes of the components in the time series, which means it does not give an accurate measure of the *relative* strengths of the couplings. This can be overcome by using the normalized version of the bispectrum [189,187],

$$b(f_1, f_2) = \frac{|\hat{M}_3(f_1, f_2)|}{\sqrt{|F(f_1)F(f_2)|^2|F(f_3)|^2}}, \tag{86}$$

where  $b(f_1, f_2)$  is known as the bicoherence and takes values between 0 and 1. As with many time series analysis techniques, positive identification of deterministic phenomena (i.e. couplings) will apparently occur when the bicoherence is applied to simple Gaussian white noise. These false positive detections are biased towards lower frequencies, with a chi-squared distribution [190–192].

As with any method that uses the standard Fourier basis, the Fourier bispectrum is ineffective for detecting couplings in nonstationary time series, which are most commonly associated with coupled nonlinear systems. Fortunately, wavelets can again be exploited for this task. The wavelet bispectrum follows directly from (84) by substituting the CWT for  $F$  and integrating over time [193,76],

$$B^W(s_1, s_2) = \int_L W(s_1, t)W(s_2, t)W^*(s_3, t)dt, \tag{87}$$

where  $s_3 = 1/(\frac{1}{s_1} + \frac{1}{s_2})$ . While the wavelet biamplitude gives the same information as the Fourier biamplitude, an instantaneous biamplitude  $A(s_1, s_2, n) = |W(s_1, n)W(s_2, n)W^*(s_3, n)|$  can now be defined to give a time-dependent bispectrum. More importantly though, an instantaneous biphase  $\phi(s_1, s_2, n) = \phi(s_1, n) + \phi(s_2, n) - \phi(s_3, n)$  can be defined which provides clear evidence for when a coupling between the two scales is present: when the biphase is constant a coupling exists [76].

The disadvantage of using the wavelet transform in this case is that new bias is introduced. Wherever the difference in the frequencies  $f_{\text{high}}$  and  $f_{\text{low}}$  is large, the adaptive frequency resolution means that the combined frequency  $f_{\text{high}} + f_{\text{low}} \approx f_{\text{high}}$ , making these couplings undetectable. On the other hand, the logarithmic frequency scale of the wavelet transform benefits the detection of couplings between two low frequencies.

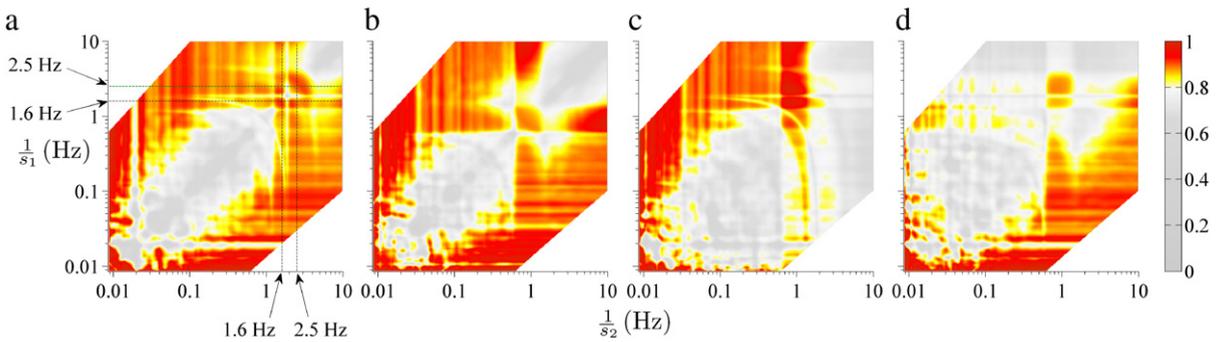
The same method can also be used to detect couplings between components from different time series. The *cross* bispectrum can be defined in several ways [194,193,76,195] using different combinations of the three wavelet components in (84) and (87), i.e.

$$B_{122}^W(s_1, s_2) = \int_L W_1(s_1, n)W_2(s_2, n)W_2^*(s_3, n)dn, \tag{88}$$

where  $W_1$  and  $W_2$  are the wavelet transforms of the corresponding time series. In addition, the wavelet bicoherence is given by [196]:

$$b_{122}^W(s_1, s_2) = \frac{|B_{122}^W(s_1, s_2)|}{\sqrt{\int_L |W_2(s_1, t)W_2(s_2, t)|^2 dt \int_L |W_1(s_3, t)|^2 dt}}. \tag{89}$$

By comparing the cross-bispectra from different combinations it is also possible to deduce some information about the direction of coupling between the oscillations in two separate time series.



**Fig. 37.** Time-averaged wavelet bicoherence for two coupled Poincaré oscillators, with unidirectional  $1 \rightarrow 2$  coupling. The auto-bicoherence for each oscillator,  $b_{111}$  and  $b_{222}$ , is shown in (a) and (b) respectively. For the cross-bicoherence,  $b_{211}$  is shown in (c) while  $b_{122}$  is shown in (d). The white parts correspond to wherever the third scale  $s_3 = 1 / \left( \frac{1}{s_1} + \frac{1}{s_2} \right)$  is closer to either  $s_1$  or  $s_2$  than the next smaller scale (as occurs between any pair of very small and very large scales). The bicoherence was calculated using the Morlet wavelet with  $\omega_c = 1$ .

Fig. 37 shows bicoherence plots for the quadratically coupled Poincaré oscillators. These plots highlight the sensitivity of the bispectrum to false positive detections due to the effect of noise. This can only be limited by performing additional averaging in time, such as splitting the time series into windows [193,76]. Despite these problems, prior knowledge of the main dynamics in the two time series can aid the interpretation of the bicoherence and is achieved easily using time–frequency analysis. Looking more closely at Fig. 37(a), the intersecting lines of low-bicoherence point to four peaks which can be seen at (1.6 Hz, 1.6 Hz), (1.6 Hz, 2.5 Hz), (2.5 Hz, 1.6 Hz) and (2.5 Hz, 2.5 Hz). The peak at (2.5 Hz, 2.5 Hz) corresponds to self-coupling and can be ignored. However, the other peaks relate to the coupling between the first oscillator and the intermediate frequency between the two oscillators, i.e.  $(5.3/2\pi + 2.5)/2$  Hz. These peaks are not visible in Fig. 37(b) in the quadrant of frequencies above  $1/s_1 \sim 1$  Hz,  $1/s_2 \sim 1$  Hz (which is devoid of intersecting lines) because the coupling is unidirectional, meaning that the time series of the second oscillator contains no interaction information. The same is seen in the cross-bicoherence which shows very strong peaks for the interacting frequencies at  $1/s_1 > 1$  Hz,  $1/s_2 \sim 1$  Hz in Fig. 37(c), where  $f_3$  comes from the time series of the first oscillator. On the other hand, these peaks are not observed in Fig. 37(d) where  $f_3$  comes from the time series of the second oscillator. Hence, wavelet bicoherence offers a means to detect both the existence of a coupling and its direction.

### 8.5. Information theoretic approach

Wavelets are not the only available tool for the analysis of interactions in complex systems. Another way to detect couplings is by using statistics based on information theory, such as transfer entropy [197] and Granger causality [198,199]. In the latter case, a coupling is said to exist if one system gives information about the state of the other system at some point in the future [76,200–202]. Starting with the probability distributions of the two time series,  $p(x_1(t))$  and  $p(x_2(t))$ , the Shannon entropy for each can be defined as

$$H(x_i) = - \sum p(x_i) \log p(x_i), \quad (90)$$

which gives a measure of the uncertainty or ‘randomness’ in  $x_i$ . The *joint* entropy can also be defined as

$$H(x_i, x_j) = - \sum \sum p(x_i, x_j) \log p(x_i, x_j), \quad (91)$$

where  $p(x_i, x_j)$  is the 2-dimensional joint probability distribution. The amount of common information contained in  $x_i$  and  $x_j$ , which is analogous to the inverse of the joint entropy, is given by the mutual information:

$$I(x_i; x_j) = H(x_i) + H(x_j) - H(x_i, x_j). \quad (92)$$

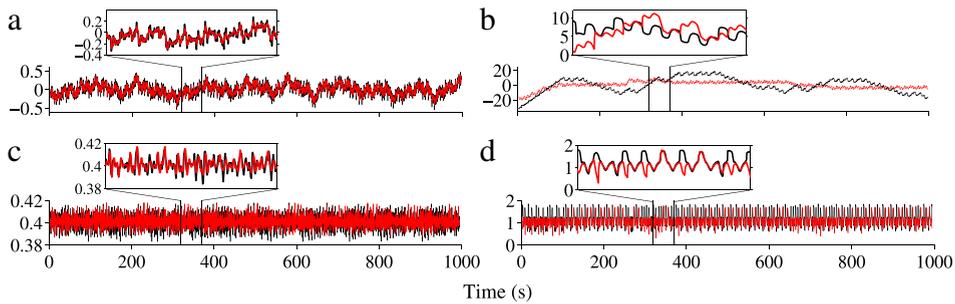
Finally, the conditional entropy is defined as

$$H(x_i|x_j) = - \sum \sum p(x_i, x_j) \log p(x_j|x_i), \quad (93)$$

where  $p(x_j|x_i)$  is the probability distribution for  $x_j$  if the value for  $x_i$  is given. The dependence between  $x_i$  and  $x_j$  without the possible influence of another variable  $x_3$  can then be defined using the *conditional* mutual information (CMI),

$$I(x_1; x_2|x_3) = H(x_1|x_3) + H(x_2|x_3) - H(x_1, x_2|x_3). \quad (94)$$

Consider now two time series  $x(t)$  and  $y(t)$ . The information flow from  $x$  to  $y$  is given by  $I(x; y_d|y)$ , where  $y_d$  is the delayed time series  $y(t + \tau)$  as used previously in (41) with  $\Delta t = \tau$ . This quantity excludes information from both the history of  $y(t)$  on itself and the common history of  $x(t)$  and  $y(t)$  [76]. Similarly, the information flow from  $y$  to  $x$  is given by  $I(y; x_d|x)$ .



**Fig. 38.** Comparison between the methods of extracting phase using the Hilbert transform (black) and marked events (red) in the case of unidirectional coupling from the non-autonomous Poincaré oscillator to the autonomous Poincaré oscillator. (Top) Variation in the unwrapped phase after removing the trend using a linear polynomial fit for: (a) The autonomous oscillator; (b) The non-autonomous oscillator. (Bottom) Variation in the return times for: (c) The autonomous oscillator; (d) The non-autonomous oscillator. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Therefore, the strength of coupling from one time series to another should be indicated in the amount of information flow in the corresponding direction.

In reality, there is always going to be some baseline mutual information contained within even two completely unrelated time series. This is why the method requires the use of surrogate data (see Appendix D.2) to determine whether there is a *significant* amount of information being transferred either from  $x_1 \rightarrow x_2$  or  $x_2 \rightarrow x_1$ . It is also dependent on estimates of the probability distributions of the time series, which require careful consideration as discussed in Appendix D.1. However, the main advantage of this approach is that it is not restricted by frequencies; the CMI gives a measure of the information transfer between two arbitrary sets of data, rather than being localized in any one domain. It is also worth noting that Granger causality can be calculated using other methods that do not rely on CMI [203].

## 8.6. Analysis

### 8.6.1. Phase of complex systems

The definition of phase used in the methods discussed in the previous sections has been based mostly around that of a limit cycle oscillator with a well-defined characteristic frequency. This means that the phase or return times can be extracted unambiguously by projecting the attractor of the system onto a 2-dimensional plane. Either Poincaré sections or the phase angle  $\phi = \arctan(x/y)$  in relation to the  $x$  and  $y$  dimensions of the plane can then be used [170,204]. Alternatively, the instantaneous phase can be found directly from the time domain by using the Hilbert transform as described in Section 5.1.2. In either case, the basic condition is that the chosen phase grows monotonically in time [167].

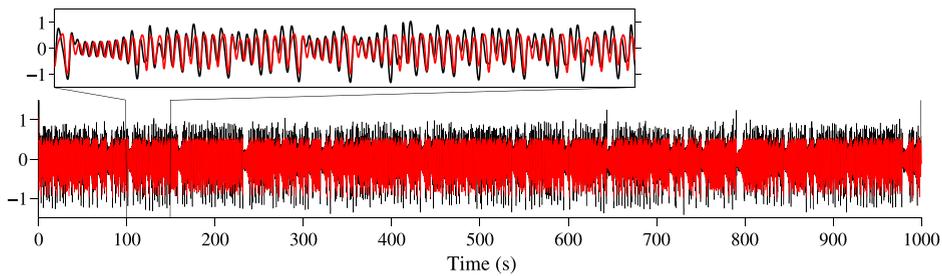
A comparison of two phase-extraction methods is shown in Fig. 38. When applied to the time series from a limit cycle oscillator, the extracted phase and return times seem to be in agreement. For the complex system there is some agreement between the two methods but this only ever occurs over timescales equal to a few cycles.

The problem of extracting the phase for a complex system occurs because phase is connected to fixed reference points. In limit cycle oscillators the system is confined to single orbits around the fixed points. However, in the case of complex systems there exist either different orbits around several fixed points (e.g. Lorenz system) or the fixed points with positions that change in time (e.g. stochastic Duffing and non-autonomous Poincaré systems). This means that the fixed point of the oscillation is continuously changing, so that there is no longer a static reference point for defining the phase.

Another complication with defining the phase of non-autonomous systems is that it is possible to extract two or more unique phases from the same system. This is because the phases corresponding to the non-autonomous terms are completely independent of the phase defined by the internal variables of the system. As mentioned in previous discussions, there is no way to separate these terms using time series analysis so the phases inevitably become mixed. Therefore, only with simulated data is it possible to extract a meaningful phase from a non-autonomous system.

In a sense, multiple phases can also be defined for autonomous systems like the Lorenz system. However, in these cases the phases are separated between timescales, which makes it possible to extract a meaningful phase. Indeed, it has recently been shown that the protophase found stroboscopically using Poincaré sections can be converted into an optimal ‘true’ phase [204]. However, this can only ever be done for the fast components of the system. For a slow component, such as the switching between orbits around the two fixed points in the Lorenz system, only one point in the phase can be defined because any continuous phase comprised of many Poincaré sections would also include the fast component. Information corresponding to the dynamics over longer timescales is therefore lost when extracting the phase of a complex system.

It seems that a possible solution to extracting the phase of complex systems might be to restrict the analysis to only the dynamics occurring over short timescales. This idea is applied to the protophases of  $x$  and  $z$  components of the Lorenz system in Fig. 39, where it can be seen that there is a great deal of agreement between the two estimates. However, the problem with this solution is that the phase extracted using the Hilbert transform is dominated by the effects of nonlinearity on these



**Fig. 39.** Detrended protophases of the Lorenz system. The black line shows the phase extracted from the  $x$ -component by calculating the Hilbert transform of the first IMF from empirical mode decomposition. The red line shows the phase obtained by using the Hilbert transform directly on the  $z$ -component. The lowest value of the instantaneous frequency was found to be  $\sim 1$  Hz, so a moving average of length 1 s was used to remove the components below this frequency in both cases. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

short timescales. In other words, the phase dynamics is mainly the result of the non-sinusoidal shape of single cycles of the oscillation, arising in the same way as the harmonics in the Fourier transforms in Fig. 9(b) and (c).

### 8.6.2. Pseudo-complexity

Since the phases of the complex systems cannot be defined from the time series under analysis, the methods from the phase dynamics and Bayesian inference approaches cannot be applied. However, to investigate the effects of chaos, stochasticity and non-autonomy on the performance of these methods, a new set of parameters is defined to generate time series for which the phase *can* be extracted. Changing the parameters in this way means that some properties of the dynamics may be altered but the three cases still fit the description of a chaotic, stochastic and non-autonomous system.

For the Lorenz system, the phase corresponding to the ‘fast’ component can be extracted directly from the  $z$ -component, as shown in Fig. 39. The parameters for the Lorenz system can therefore remain the same as given in Section 2.5.1.

In the case of the stochastic Duffing system, the damping term must be made negligible for the phase to be definable. However, setting  $\delta = 0.00001$  in Eq. (10) significantly increases the amplitude of the oscillations. To maintain a similar amplitude, the noise forcing term is also reduced by setting  $\xi = 0.1$ .

In a similar way, for the non-autonomous Poincaré system given in (12) the strength of the forcing term can be reduced to  $\gamma = 1$  in order to provide a more suitable time series for analysis of phase dynamics.

These adjustments are made strictly for the analysis of the phase dynamics methods. The other approaches based on the time–frequency domain and information theory do not require the systems to have a defined phase so in these cases the original parameters are still used.

### 8.6.3. Coupled system examples

To test the performance of the methods, each system is quadratically coupled to a Poincaré oscillator:

$$\begin{aligned} \frac{dx_1}{dt} &= -qx_1 - (\omega_1 + \gamma_f y_2) y_1 \\ \frac{dy_1}{dt} &= (\omega_1 + \gamma_f y_2) x_1 - qy_1 + \gamma_y y_2, \end{aligned} \quad (95)$$

where  $a = 1$ ,  $\alpha = 5$ ,  $\omega = 5\pi$  and  $y_2(t)$  is the  $y$ -component of the complex systems (6), (10) or (12). The parameters  $\gamma_f$  and  $\gamma_y$  determine the strength of parametric frequency modulation and  $y$ -dimension coupling respectively. For the case of bidirectional couplings the quadratic coupling term appeared in the  $y$ -component of the complex system, i.e.

$$\begin{aligned} \frac{dx_2}{dt} &= \sigma(y_2 - x_2), \\ \frac{dy_2}{dt} &= x_2(\rho - z_2) - y_2 + \gamma_t y_1, \\ \frac{dz_2}{dt} &= x_2 y_2 - \beta z_2, \end{aligned} \quad (96)$$

for the chaotic system,

$$\begin{aligned} \frac{dx_2}{dt} &= y_2, \\ \frac{dy_2}{dt} &= \xi \eta(t) - \delta y_2 - \beta x_2 - \alpha x_2^3 + \gamma_t y_1, \end{aligned} \quad (97)$$

**Table 1**  
Parameter values for Eqs. (95), (95), (96) and (98) for the seven cases considered.

		Coupling							
		None (noise)		Unidirectional			Bidirectional		
$\gamma_y$	Chaotic	0.1	0.1	0.07	0.07	$0.03 + 0.03 \sin(0.004\pi t)$		0.07	0.07
	Stochastic	0.1	0.1	1	1	$0.5 + 0.5 \sin(0.0008\pi t)$		1	1
	Non-autonomous	0.1	0.1	1	1	$0.5 + 0.5 \sin(0.004\pi t)$		1	1
$\gamma_f$	Chaotic	0	0.1	0	0.07	$0.03 + 0.03 \sin(0.004\pi t)$		0	0.07
	Stochastic	0	0.1	0	1	$0.5 + 0.5 \sin(0.0008\pi t)$		0	1
	Non-autonomous	0	0.1	0	1	$0.5 + 0.5 \sin(0.004\pi t)$		0	1
$\gamma_t$	Chaotic	0	0	0	0	0		7.5	7.5
	Stochastic	0	0	0	0	0		0.5	0.5
	Non-autonomous	0	0	0	0	0		0.5	0.5

for the stochastic system and

$$\begin{aligned} \frac{dx_2}{dt} &= -qx_2 - \omega y_2, \\ \frac{dy_2}{dt} &= \omega x_2 - qy_2 + \gamma_f f_y(t) + \gamma_t y_1, \\ f_y(t) &= \sin(2\pi t) + \sin(4t), \end{aligned} \tag{98}$$

for the non-autonomous system.

A total of seven cases were examined, which are outlined in Table 1. In particular, there are two variations of interaction for the unidirectional and bidirectional couplings, which gives a means to test the robustness of each of the methods. To judge how well the methods respond to time-dependent couplings, another set of data was simulated where  $\gamma_f$  and  $\gamma_y$  vary in time. Specifically, they were given the form  $\gamma_f(t) = \gamma_{f_0} + \gamma_{f_0} \sin(2\pi ft)$  and  $\gamma_y(t) = \gamma_{y_0} + \gamma_{y_0} \sin(2\pi ft)$ , where  $f = 0.002$  Hz for the coupled chaotic and non-autonomous systems and  $f = 0.0004$  Hz for the coupled stochastic system. However, to ensure that the methods are not simply detecting moments when there is no external influence to the Poincaré oscillator, the system was additionally coupled to a small noise term so that

$$\begin{aligned} \frac{dx_1}{dt} &= -qx_1 - (\omega_1 + \gamma_f(t)y_2)y_1 \\ \frac{dy_1}{dt} &= (\omega_1 + \gamma_f(t)y_2)x_1 - qy_1 + \gamma_y(t)(y_1 - y_2)^2 + \xi \eta(t), \end{aligned} \tag{99}$$

where  $\eta(t)$  is a normalized Gaussian white noise source and  $\xi = 0.1$ . Also, in addition to the coupled systems,  $y_2$  in (95) was replaced by a Gaussian white noise source to simulate both frequency modulation and additive forcing by noise. These time series were then paired with the uncoupled complex time series to test the methods when there is absolutely no coupling between the systems but the Poincaré system is still under forcing and frequency modulation.

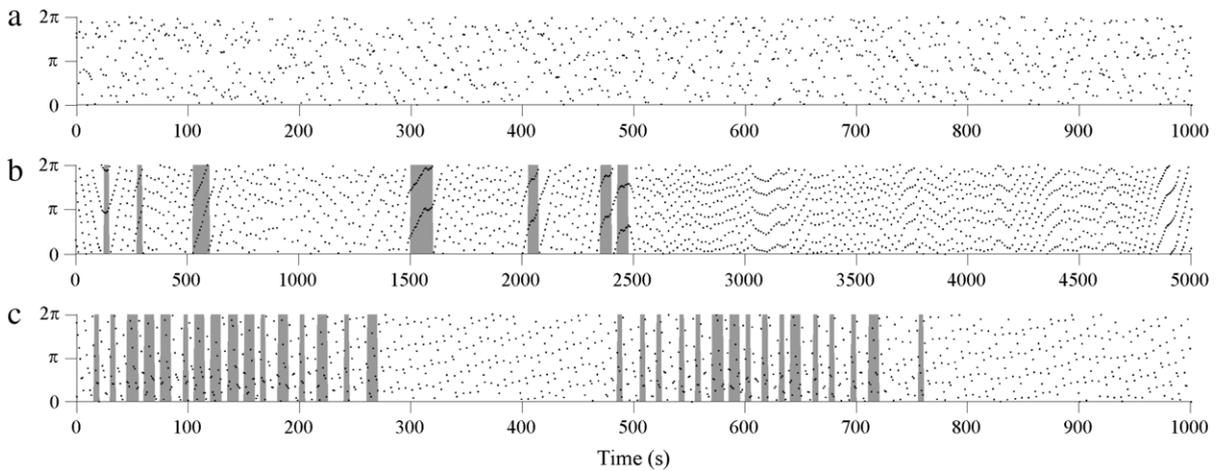
The values of the coupling parameters were chosen so that in each case the forcing from the complex systems had a maximum amplitude of  $\sim 1$ . For the coupling from the autonomous Poincaré oscillator to the complex systems the parameter was chosen so that the forcing would be equal to half the maximum amplitude of the complex system time series.

#### 8.6.4. Phase dynamics methods

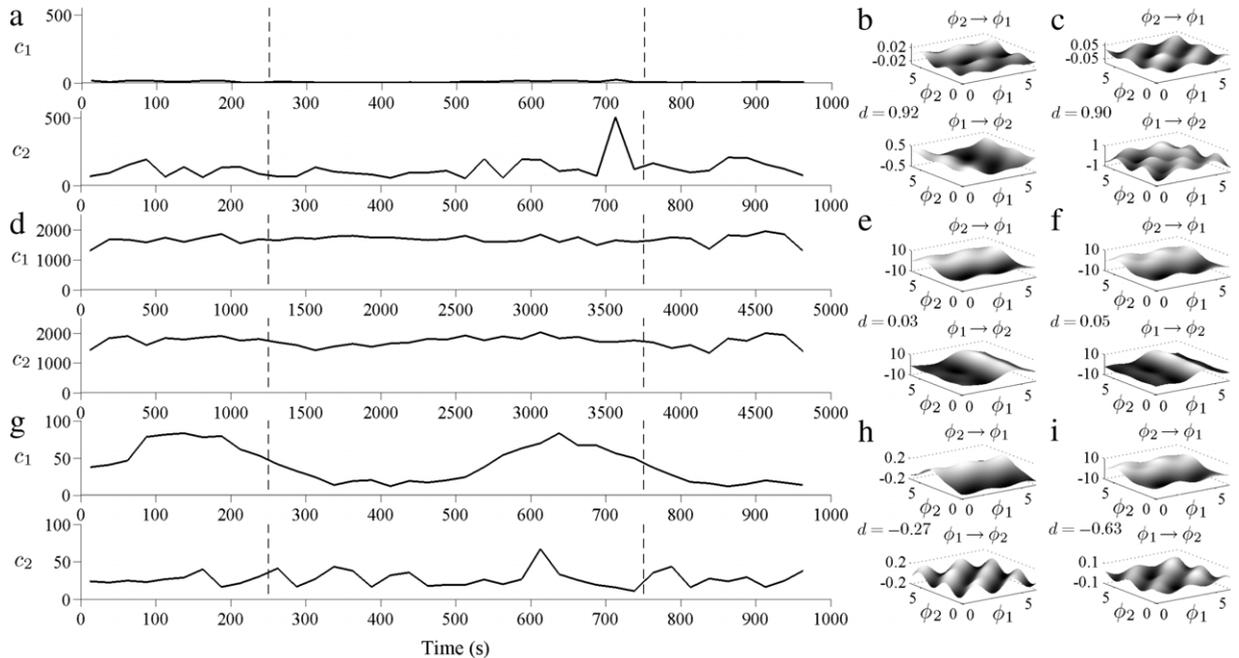
Protophases of the systems were extracted directly from the time series using the Hilbert transform. Estimates of the true phases were then calculated using the protophase to phase conversion as described in [167].

Analysis of the synchronization of the coupled systems is shown in Fig. 40. For the Lorenz system, no synchronization is detected in either the synchrogram or the synchronization index. For the stochastic system, periods of 1:2 synchronization were detected by the synchronization index and intervals of 1:5 synchronization can also be observed in the synchrogram. However, these do not seem to have any correlation with the points where the coupling is strongest because the amplitude of the Duffing oscillator, and hence the effective coupling strength, varies a great deal. This is in contrast to the non-autonomous system where the amplitude variation is kept small by the stable Poincaré limit cycle. Consequently, a pattern of 1:1 synchronization corresponding to the sinusoidal change in the coupling strength can be seen.

Fig. 41 shows the results of the EMA method when it was applied to the same time series. In the case of the chaotic system the detected coupling is in the wrong direction at all times, implying that the autonomous Poincaré oscillator was influencing the dynamics of the complex system. Here it is apparent that the phase has not been extracted accurately from the time series. In fact, it has recently been shown that systems such as the Lorenz oscillator require intensive treatment in phase space due to chaotic phase diffusion [204]. However, as has already been shown, the application of phase space methods to inverse problems is fraught with difficulties. The problem of applying phase dynamics methods to observed chaotic systems is therefore unresolved.



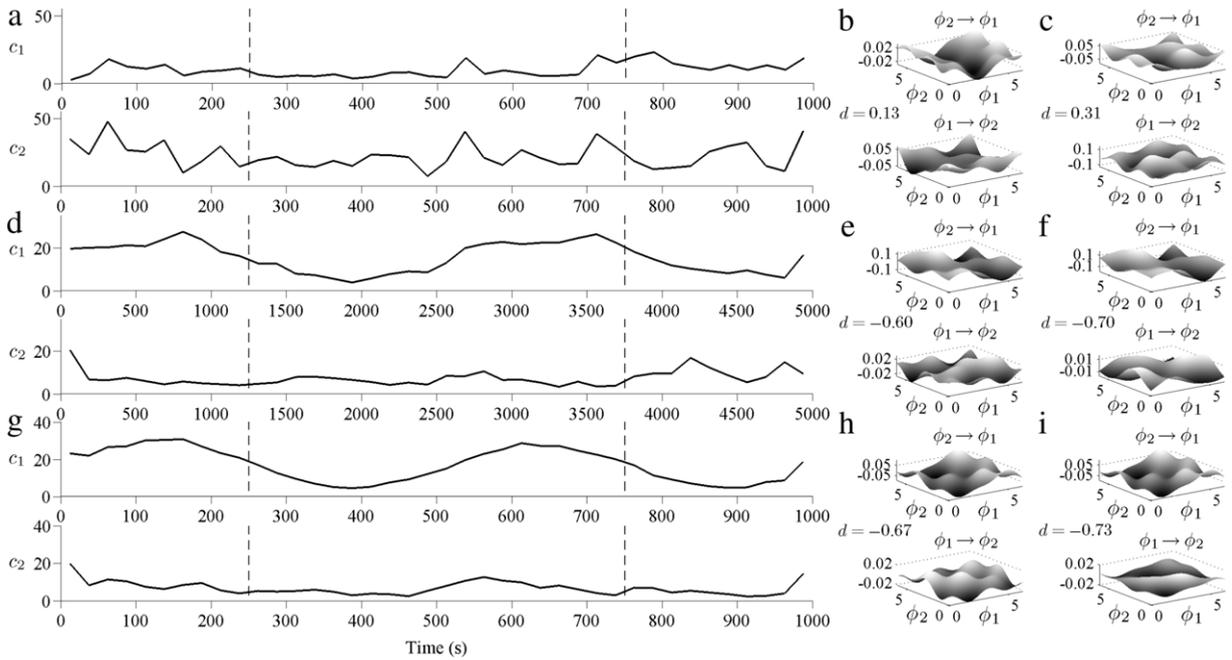
**Fig. 40.** Synchrograms calculated from the time series of the pseudo-complex systems for the case of the time-dependent coupling: (a) Chaotic system; (b) Stochastic system; (c) Non-autonomous system. In (b) intervals of significant 1:2 synchronization are shaded grey while in (c) intervals of 1:1 synchronization are shown. The synchronization index was determined using a 10 s sliding window (with 50 s for the stochastic Duffing system because of the longer periods of the oscillations). Significant synchronization was determined by points where the synchronization index exceeded the  $2\sigma$  level as calculated from 100 IID surrogates.



**Fig. 41.** The EMA method applied to the phases of the pseudo-complex systems for the case of the time-dependent coupling. The calculated coupling coefficients are shown in (a), (d) and (g) for the chaotic, stochastic and non-autonomous systems respectively. The coupling functions  $F_{1,2}$  are shown for the (b,c) chaotic, (e,f) stochastic and (h,i) non-autonomous systems at two separate points in time (indicated by the dashed lines) where the strength of the coupling is the same. Also shown are the values of the directionality index at each of these points. The functions were calculated using a 25 s (125 s for the stochastic Duffing system) sliding window with no overlap and using a time delay  $\tau = 0.5$  s (2.5 s for the stochastic Duffing system).

For the stochastic system (Fig. 41(d)), no obvious direction of coupling was detected using the EMA method. This is probably due to the presence of a strong noise influence, which is included into the least squares fitting procedure. However, for the non-autonomous system (Fig. 41(g)) the correct direction of coupling is found and the general shape of the coupling strength is seen in  $c_1$ .

The results of the Bayesian inference method are shown in Fig. 42. Again, for the chaotic system the spurious phase extraction results in an incorrectly detected direction of coupling, although the effect is not as severe as in the EMA method. However, for the stochastic Duffing system (Fig. 42(d)) the direction is correctly detected and the sinusoidal variation in the coupling strength can also be seen in  $c_1$ . The amplitude of both  $c_1$  and  $c_2$  is also far lower than what was calculated for the EMA method (Fig. 41(d)), indicating that the noise is no longer making the phases appear strongly dependent on the



**Fig. 42.** The Bayesian inference method applied to the phases of the pseudo-complex systems for the case of the time-dependent coupling. The calculated coupling coefficients are shown in (a), (d) and (g) for the chaotic, stochastic and non-autonomous systems respectively. The coupling functions are shown for the (b,c) chaotic, (e,f) stochastic and (h,i) non-autonomous systems at two separate points in time (indicated by the dashed lines) where the strength of the coupling is the same. Also shown are the values of the directionality index at each of these points. The functions were calculated using a 25 s (125 s for the stochastic Duffing system) sliding window and propagation constant  $p = 0.8$ .

Fourier components of the basis. The better performance when applied to the stochastic system is likely due to the fact that the Bayesian method infers the noise in addition to the other parameters. A stronger directionality is also detected for the non-autonomous system where the variation in the coupling strength is also much smoother.

The success of the Bayesian inference method, even when applied to the stochastic system, is because the method itself is based on stochastic dynamics [175–178]. This means that in addition to the deterministic parameters in the system, the noise is also inferred in a separate correlation matrix. The method is therefore well-suited to tracking time-varying parameters in systems that are influenced by noise.

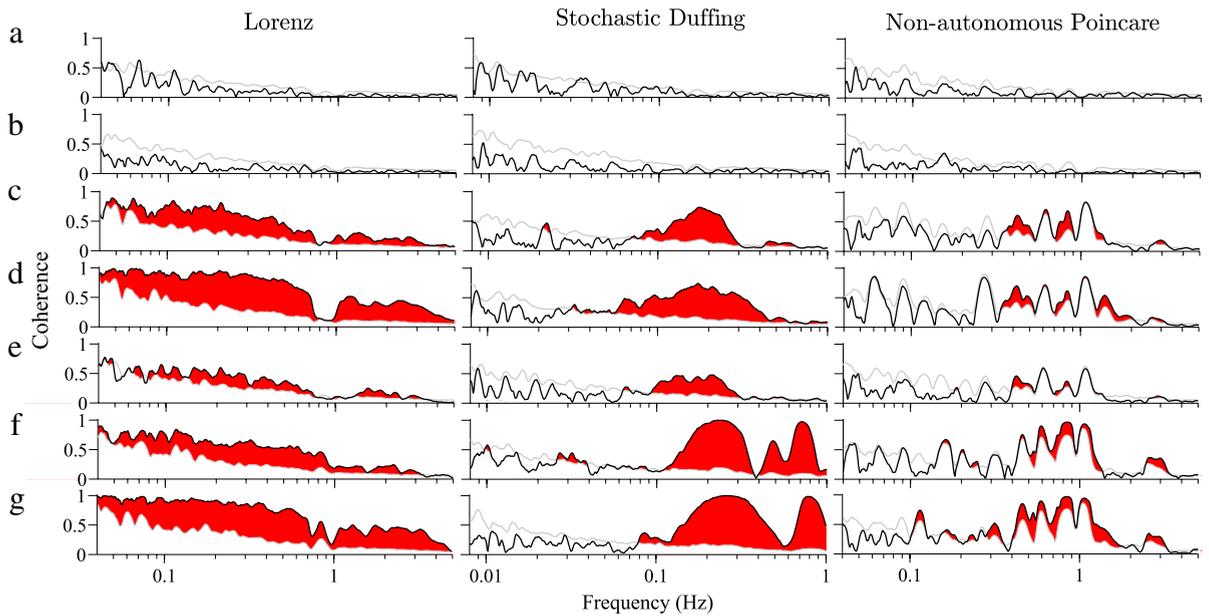
### 8.6.5. Wavelet phase coherence

Fig. 43 shows the time-averaged wavelet phase coherence for the seven pairs of test time series. Rather than directly detecting the couplings, this method looks for one of the possible effects of an interaction between two systems. In each case the method correctly finds almost no coherence between the time series of the complex systems and those of the noise-coupled Poincaré oscillator. In addition, significant coherence is always observed for the cases where couplings are present.

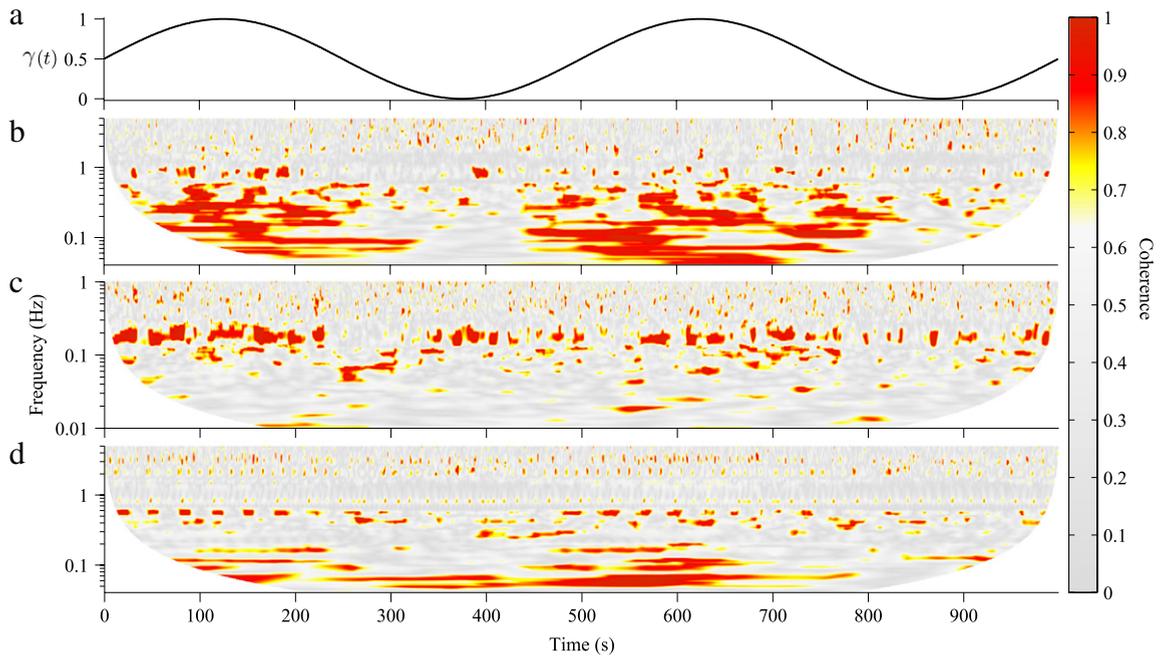
A difference between unidirectional and bidirectional coupling is only observed for the stochastic system. The reason for this is because the strong coupling has caused a loss of the complex dynamics in the Duffing system, which has led to much more coherence at the natural frequency of the oscillator and at the harmonics of this frequency.

For the non-autonomous system the coherence appears much less significant compared to the other complex systems. However, the peaks in the coherence are in fact similar in height; the main difference is actually the surrogate level, which reaches coherence values  $\sim 0.9$  for some of the peaks. This is because the dynamics of the non-autonomous system are influenced by stationary oscillations, which are much more confined in frequency. Detecting coherence between these components is therefore more difficult because variations in the frequency are required to check whether the phase difference remains constant. This effect is visible in the IAAFT surrogates which preserve the power spectrum of the time series and, hence, also contain components confined to the same frequencies.

The time-averaged coherence is less significant in each of the cases where the coupling is time-dependent. This is what might be expected due to the fact that the systems are not strongly coupled for much of time. Fig. 44 shows the same phase coherence but calculated using sliding windows of length proportional to the wavelet scale. When viewed in time, the correlation between the strength of the coupling and the amount of coherence between the time series is clear in the case of the chaotic system. It can also be observed as more subtle effects in the non-autonomous system, in particular at the frequencies around 0.5 and 1.5 Hz. However, the pattern is very difficult to see for the stochastic system, where the stochastic dynamics is perhaps difficult to distinguish from the noise added to the Poincaré oscillator.



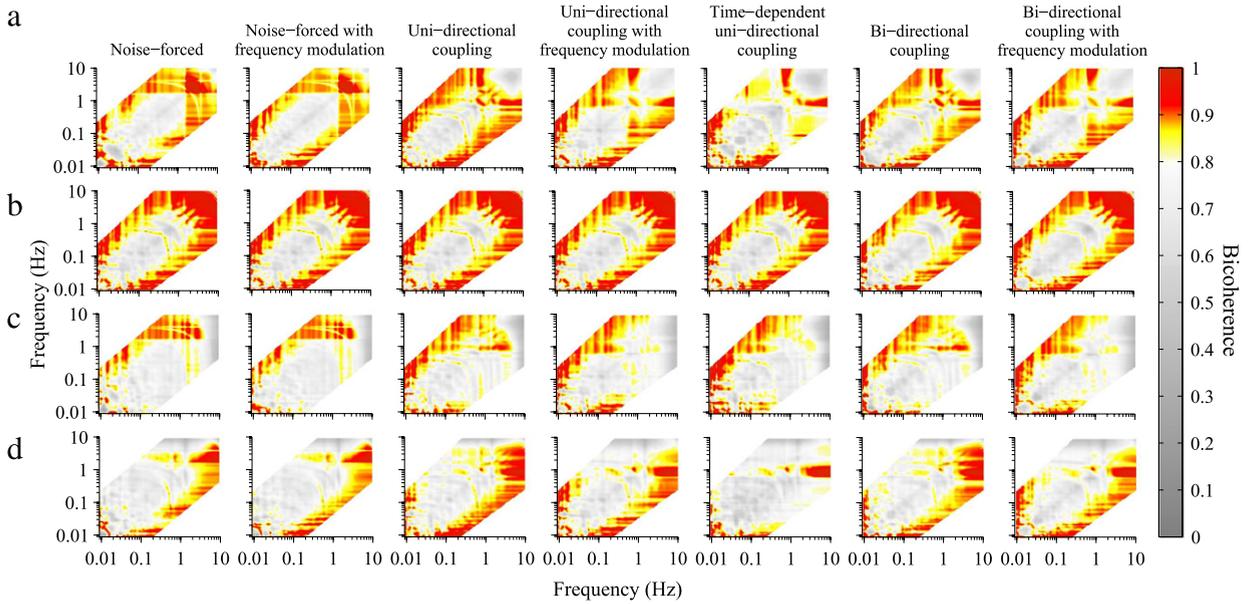
**Fig. 43.** Wavelet phase coherence between the time series of the Poincaré oscillator and the three complex systems for the seven cases outlined in Section 8.6.3: (a) Gaussian white noise coupling; (b) Gaussian white noise coupling with frequency modulation; (c) Unidirectional coupling; (d) Unidirectional coupling with frequency modulation; (e) Time-dependent unidirectional coupling; (f) Bidirectional coupling; (g) Bidirectional coupling with frequency modulation. Grey lines show the upper  $2\sigma$  levels from 100 IAAFT surrogates, with shaded areas indicating significant coherence.



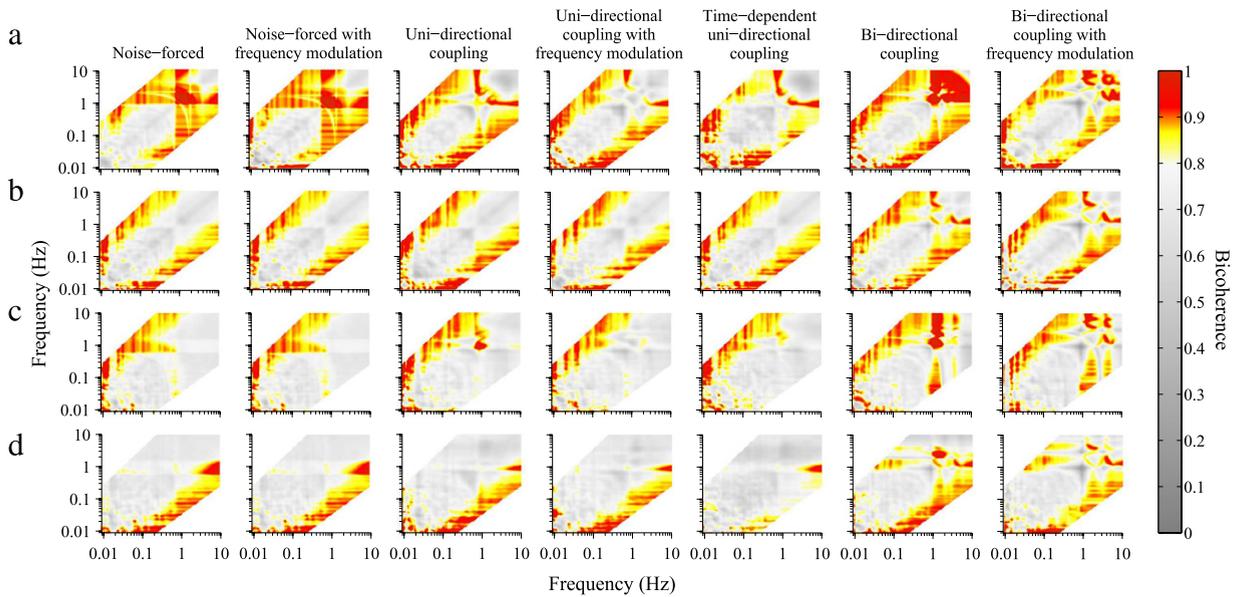
**Fig. 44.** Windowed wavelet phase coherence for the case of the time-dependent coupling described in Section 8.6.3: (a) Coupling used in simulations; (b) Chaotic system; (c) Stochastic system; (d) Non-autonomous system. The coherence was calculated using the Morlet wavelet with  $\omega_c = 1$  within windows of length  $6s$ , where  $s$  is the wavelet scale.

### 8.6.6. Wavelet bispectrum

Since complex systems exhibit a broadband frequency spectrum, information about couplings is spread across many bispectral components. This makes the interpretation of the time-averaged bispectrum difficult because a single coupling between two simple nonlinear oscillators results in several distinct peaks in the bispectrum [193,76], as observed in Fig. 37. Applying the bispectrum to complex systems with dynamics over a broad ranges of frequencies therefore results in a wide



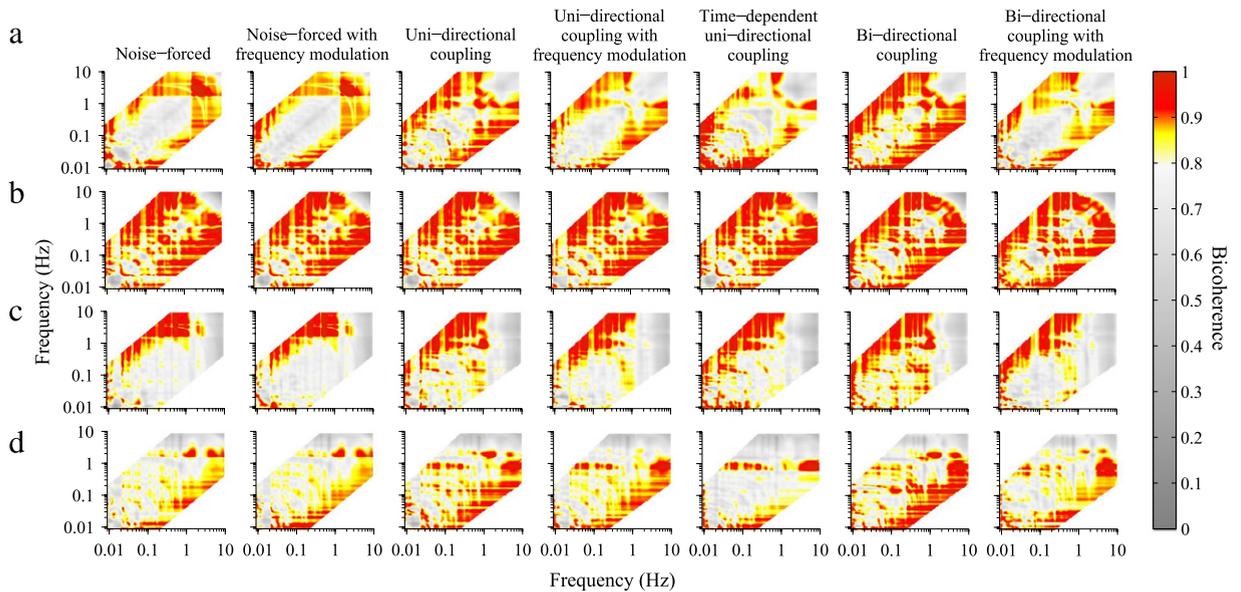
**Fig. 45.** Time-averaged bicoherence for the Poincaré oscillator (1) coupled to the chaotic system (2): (a)  $b_{111}$ ; (b)  $b_{222}$ ; (c)  $b_{211}$ ; (d)  $b_{122}$ . The bicoherence was calculated using the Morlet wavelet with  $\omega_c = 1$ . In this and the subsequent two figures, white parts were not calculated because the third scale  $s_3 = 1 / (\frac{1}{s_1} + \frac{1}{s_2})$  is closer to either  $s_1$  or  $s_2$  than the next smaller scale over these intervals.



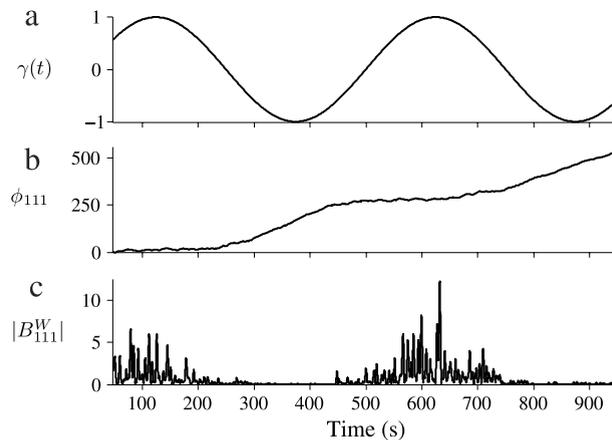
**Fig. 46.** Time-averaged bicoherence for the Poincaré oscillator (1) coupled to the stochastic system (2): (a)  $b_{111}$ ; (b)  $b_{222}$ ; (c)  $b_{211}$ ; (d)  $b_{122}$ . The bicoherence was calculated using the Morlet wavelet with  $\omega_c = 1$ .

distribution of peaks similar to what is observed in the bispectrum of two uncoupled, noisy systems. However, this does not exclude the possibility of observing changes in the bispectrum as a result of changes in the couplings.

The time-averaged bicoherence (the normalized bispectra) for each of the test time series pairs are shown in Figs. 45–47. Despite the complex patterns, the plots share many consistent features for the uncoupled cases (first two columns) and uncoupled cases (columns 3–7). It is therefore clear that the wavelet bispectrum is able to detect couplings in complex systems. The biggest problem is seen when comparing columns 3–5 with 6 and 7, where for a change in the *direction* of coupling it is very difficult to observe consistent differences in the bicoherence. This is because couplings to another time series cannot be detected when the dynamics in the uncoupled case are already spread out over a range of frequencies, as is the case in the time series of the complex systems. Essentially, the components and harmonics of the Poincaré oscillator are already present in the time series of the complex systems, so no difference can be seen in the bispectrum when they are coupled.



**Fig. 47.** Time-averaged bicoherence for the Poincaré oscillator (1) coupled to the non-autonomous system (2): (a)  $b_{111}$ ; (b)  $b_{222}$ ; (c)  $b_{211}$ ; (d)  $b_{122}$ . The bicoherence was calculated using the Morlet wavelet with  $\omega_c = 1$ .



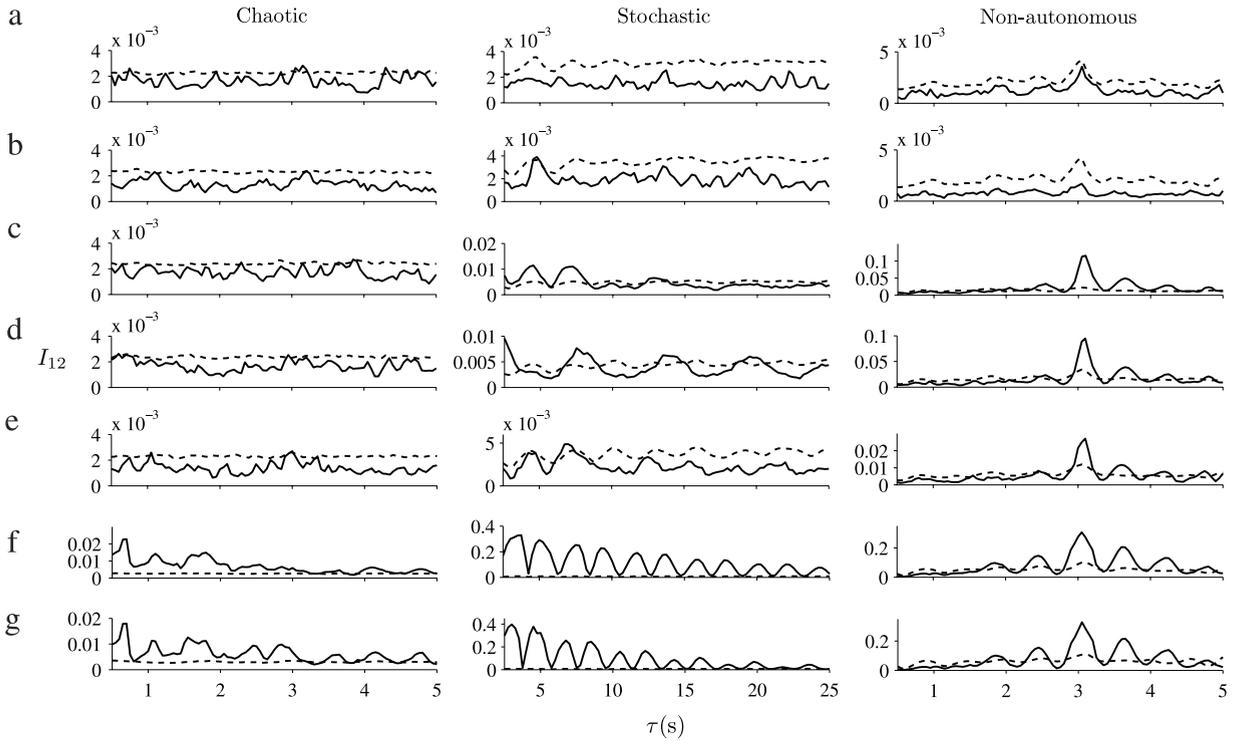
**Fig. 48.** Time-dependent bispectral analysis of the Poincaré / non-autonomous Poincaré system with time-dependent coupling. The strength of the coupling used in the numerical simulation is shown in (a). Below are the instantaneous biphase (b) and biamplitude (c) for the autobispectrum of the Poincaré oscillator, corresponding to the frequency pair 0.092–0.53 Hz. The wavelet transforms were calculated using the same parameters as before.

Again, the exception to this rule is the stochastic system where in the bidirectional cases the dynamics of the system are no longer complex. Consequently, changes in the bispectra can be seen for columns 3–5 and 6–7 in Fig. 46, which correspond to a change in the coupling direction.

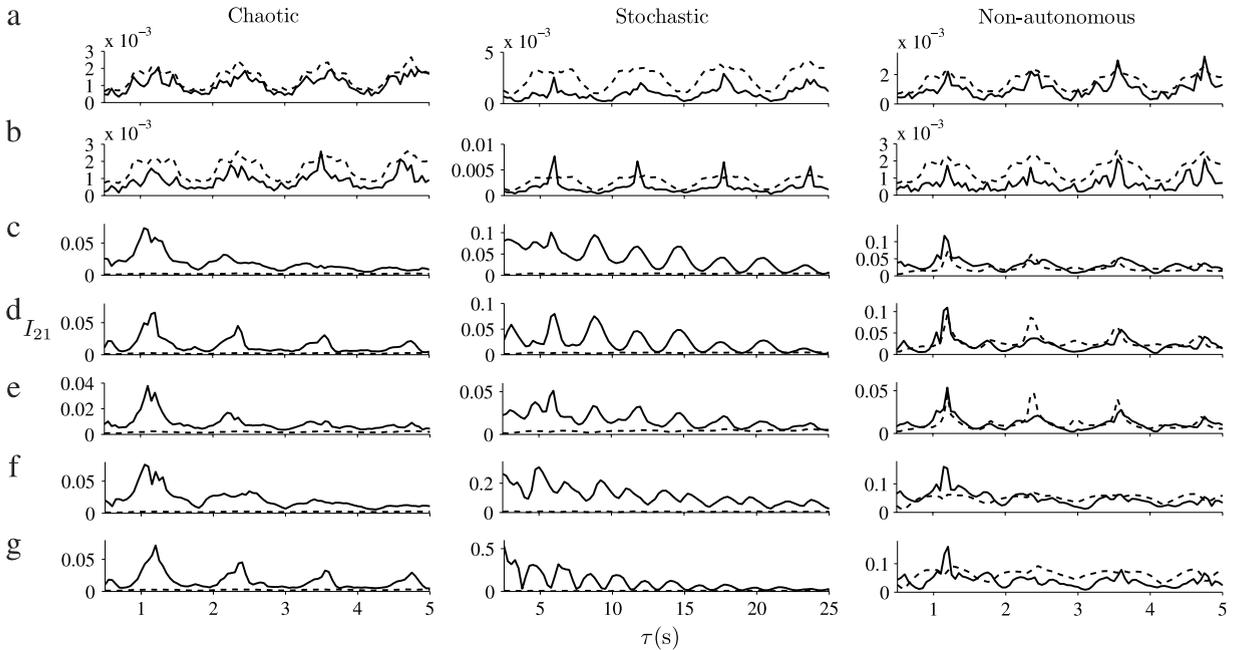
Fig. 48 shows the evolution of the biphase and biamplitude in time for the case of the time-dependent coupling with the non-autonomous system. The correlation between the biamplitude and the strength of coupling can be seen very clearly. Also, the biphase of the cross-bispectrum is observed to be constant when the coupling is strongest which, as mentioned previously, gives another indication that a coupling is present. Viewing the coefficients of the wavelet bispectrum in time therefore provides much more information about the coupling between the systems. However, it should be noted that in the case of complex systems this correlation can only be seen for very specific frequency pairs. In particular, the biphase gives spurious results for most other combinations.

### 8.6.7. Information theoretic approach

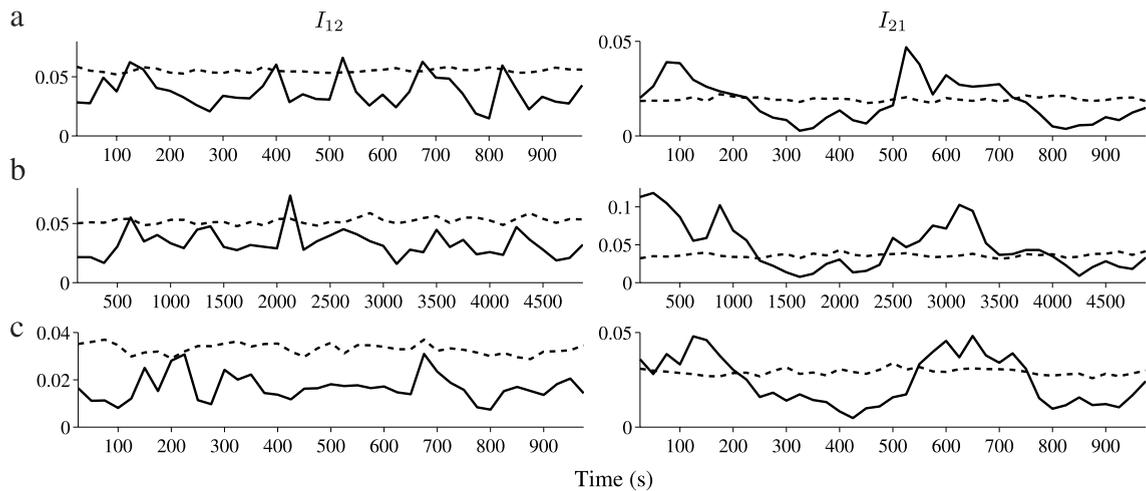
The results of the information theoretic approach are shown in Figs. 49–Fig. 50. For the cases where only noise coupling is present the CMI is below the surrogate level for almost all timescales, indicating no coupling between the systems, as expected.



**Fig. 49.** Conditional mutual information in the direction from the Poincaré oscillator to the complex systems: (a) Gaussian white noise coupling; (b) Gaussian white noise coupling with frequency modulation; (c) Unidirectional coupling; (d) Unidirectional coupling with frequency modulation; (e) Time-dependent unidirectional coupling; (f) Bidirectional coupling; (g) Bidirectional coupling with frequency modulation. Dashed lines in this and the following two figures show the upper  $2\sigma$  levels from IAAFT 100 surrogates and in all cases probability distributions were estimated using 4 bins (16 for joint-probability distributions).



**Fig. 50.** Conditional mutual information in the direction from the complex systems to the Poincaré oscillator: (a) Gaussian white noise coupling; (b) Gaussian white noise coupling with frequency modulation; (c) Unidirectional coupling; (d) Unidirectional coupling with frequency modulation; (e) Time-dependent unidirectional coupling; (f) Bidirectional coupling; (g) Bidirectional coupling with frequency modulation.



**Fig. 51.** Conditional mutual information for the time-dependent coupling from the complex systems to the Poincaré oscillator for specific time delays: (a) Chaotic system; (b) Stochastic system; (c) Non-autonomous system. The CMI was calculated in a moving window of length  $l = 50$  s using  $\tau = 0.5$  s for the chaotic and non-autonomous systems and  $l = 250$  s with  $\tau = 2.5$  s for the stochastic Duffing system.

For the chaotic system the method also appears to function as intended, showing significant CMI over all timescales when a coupling is present and no significant CMI when there is no coupling. In addition, the strength of the CMI does not change for  $I_{21}$  when the coupling in the opposite direction is added.

In the case of the stochastic system the method gives the correct direction of coupling over most timescales. However,  $I_{21}$  varies considerably in magnitude for the bidirectional coupling where the complex dynamics is lost. This shows how the value of the CMI is dependent on the frequency distribution of the dynamics as much as it is on the strength of the coupling.

The method performs worst for the non-autonomous system where significant CMI in the opposite direction to the coupling is observed over a range of timescales. This is likely due to the purely time-dependent components in the non-autonomous system, which will also appear in the dynamics of the Poincaré oscillator. These components provide information about the dynamics of the non-autonomous system, which results in significant CMI. This information is shown to be spurious in the case of the time-dependent coupling, where the sinusoidal components are not as prominent in the time series of the Poincaré oscillator. Consequently, the decrease in the CMI for  $I_{12}$  is much greater than that for  $I_{21}$ .

As with wavelet phase coherence and the wavelet bispectrum, the CMI can also be traced in time to reveal time-dependent couplings. Fig. 51 shows  $I_{12}$  and  $I_{21}$  calculated using a sliding window for the time series corresponding to the time-dependent coupling. As can be seen, the change in the coupling can clearly be seen in each case. However, it is also interesting to note that CMI drops below the significance level when there is only weak coupling. This demonstrates the limits of detecting couplings from probability distribution estimates of finite data, especially when using a smaller time window.

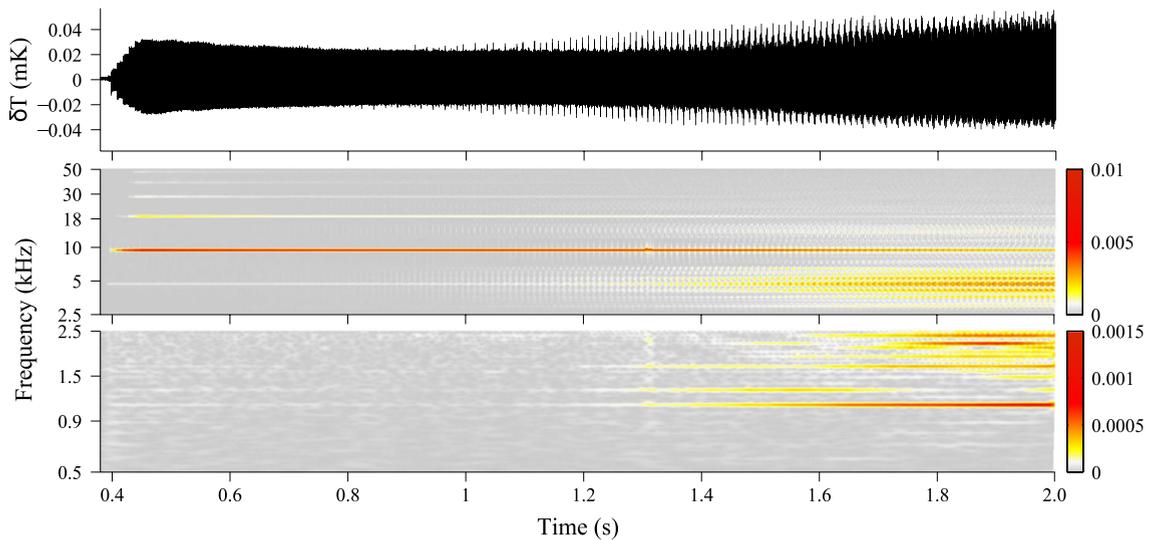
### 8.7. Summary

The analysis of interactions involving complex systems requires a markedly different approach to the typical framework that has been developed for coupled systems. The biggest obstacle is that there is no general way to define the phase of a complex system. Furthermore, attempts to extract the phase of these systems results in only the dynamics of the fast oscillatory components contributing to the phase. Dynamics over longer timescales and, hence, the slower effects of interactions may therefore be lost. The lack of a definition for the phase means that concepts such as synchronization, which is immediately apparent in coupled limit cycle oscillators, are difficult to apply to interactions involving complex dynamics.

Despite these problems, counterparts to the analysis of phases are still able to provide a great deal of useful interaction information directly from the time series of these systems. By exploiting the wealth of information in the time–frequency domain it has been possible to find coherence between two interacting systems as well as the strength and direction of the underlying coupling, even with time-varying parameters. In addition, methods based around information theory provide similar information without the need to transform to any analytical domain. Furthermore, these methods appear to work equally well for each type of complex system.

While they are not applicable to systems that are very complex, methods based around phase dynamics are still useful in cases where stochastic and non-autonomous dynamics are present in a weaker sense. However, at least in the system studied here, chaotic dynamics are too random to extract interaction information in this way.

The most promising of the phase dynamics methods is based on Bayesian inference. This is particularly useful for both stochastic and non-autonomous systems because it is able to effectively track both external noise and time-varying parameters, which was not possible in the EMA, IPA and MPA methods developed earlier. This makes the method very useful for inverse problems where both noise and time-variability are commonly encountered.



**Fig. 52.** The onset of acoustic wave turbulence in superfluid  $^4\text{He}$ . The measured time series is shown at the top, while the continuous wavelet transform of the time series is shown below in two parts using different scales to observe the more subtle inverse cascade. The calculation used the Morlet wavelet with a central frequency  $\omega_c = 5$  on the time series sampled at 100 kHz (data previously published in [25]).

## 9. Applications

Throughout this paper, the importance of using methods that detect variations in the properties of the system over time has been emphasized. For whatever time series analysis technique is applied, tracking the dynamics over time is essential for understanding non-autonomous systems, which are seen in much of the real world. The following provides examples of such systems and demonstrates how the methods described in this paper can be applied.

### 9.1. Acoustic turbulence

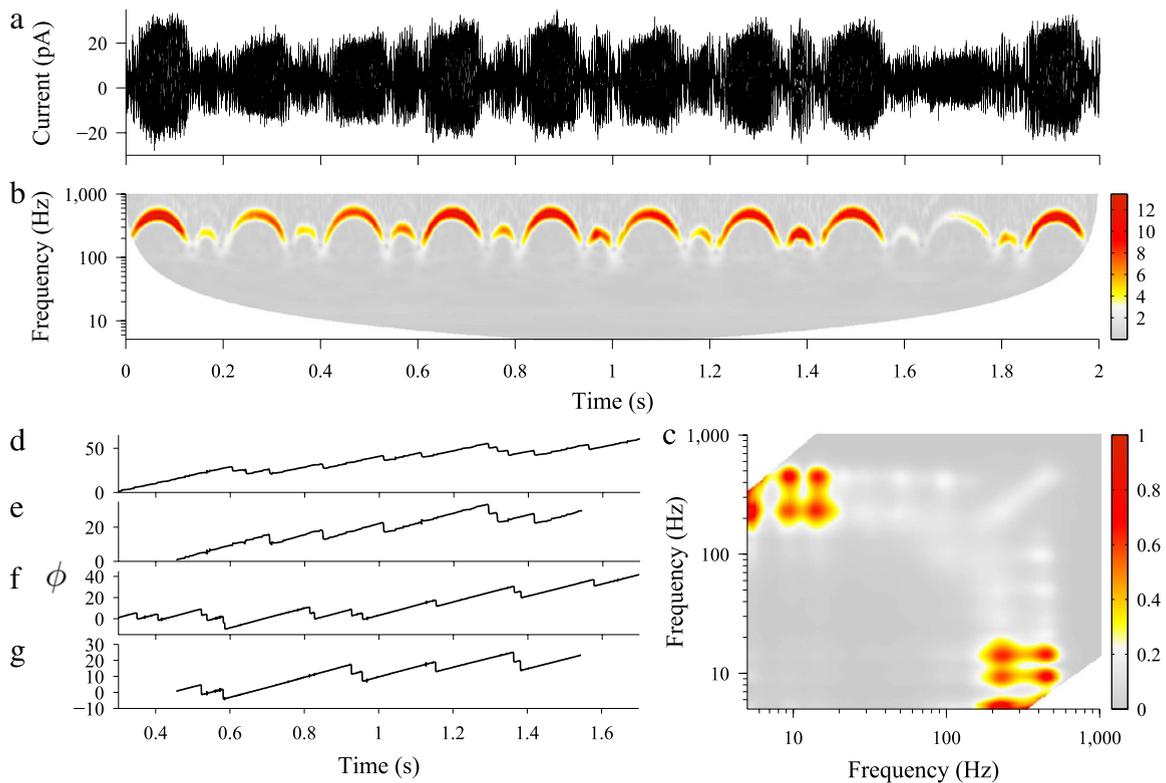
Even in the context of fluid dynamics and classical turbulence, non-autonomous dynamics can be important. Fig. 52 shows the onset of turbulence in cylindrical quartz tube filled with superfluid  $^4\text{He}$  (He II). The experiment involved heating of the He II at one end of the cylinder to produce a series of second sound waves, and measuring the resulting temperature changes at the other end using a superconducting bolometer (full experimental details can be found in [205,25,206,207]). The most obvious phenomenon seen in the wavelet transform is the Kolmogorov-like direct cascade, whereby the waves change shape and produce high harmonics from which energy is dissipated. However, a more subtle effect known as the *inverse cascade* is also seen to arise over time, where energy is being transferred from the second sound waves to low frequency subharmonics. The most important consequence of this inverse cascade is the spikes observed in the time series, which are representative of the rogue waves observed at sea.

The use of the wavelet transform allows the onset and gradual progression of turbulence to be tracked continuously. The full extent of the inverse cascade would be difficult to observe with other methods such as the short-time Fourier transform due to the poor relative frequency resolution at low frequencies. However, the adaptive resolution of the wavelet transform means that the precise times and frequencies at which subharmonics appear is revealed.

### 9.2. Electrons on helium

Data from another experiment are shown in Fig. 53. Here liquid  $^3\text{He}$  is condensed into a cell containing a tungsten filament which is heated briefly to produce electrons. The electrons are confined to surface states with a subband that allows them to be excited by microwave radiation, which enters the cell through a pair of sealed windows (full details of the experimental setup are given in [208]). The constant microwave radiation field induces the strong oscillatory current which is measured by the electrodes in the cell and is clearly seen in the continuous wavelet transform of the time series.

The oscillation appears to contain separate amplitude and phase dynamics: the frequency of the oscillations is being modulated by a periodic function but at the same time the amplitude of the oscillations changes non-periodically. As has been shown previously in the paper, such a mixture of amplitude and phase dynamics would be interpreted as a stochastic system in the frequency domain or phase space. It is only in the time–frequency domain that the dynamics corresponding to a simple non-autonomous oscillator become apparent.



**Fig. 53.** Resonant photovoltaic effect in surface state electrons on liquid  $^3\text{He}$ : (a) Time series of the current recorded by an electrode; (b) Continuous wavelet transform of the time series; (c) Wavelet autobispectrum of the time series; (d–g) The unwrapped wavelet biphase for the frequencies (234 Hz, 14.3 Hz), (234 Hz, 9.4 Hz), (440 Hz, 14.3 Hz) and (440 Hz, 9.4 Hz) respectively. The calculations used the Morlet wavelet with  $\omega_c = 5$  on the time series sampled at 4410 Hz.

The bispectrum shows other interesting features, in particular four peaks corresponding to coupling between the oscillations and two lower frequencies. In this case the oscillations have been divided into two frequencies which relate to the slow and fast components seen in the modulation. The biphase of the four frequency pairings confirms the existence of a coupling since the phases are not monotonically increasing. However, the biphase is not stationary either so the coupling is in some way intermittent. It also appears that the two low frequencies are in fact harmonics of the same oscillation due to the fact that the biphases appear to be correlated.

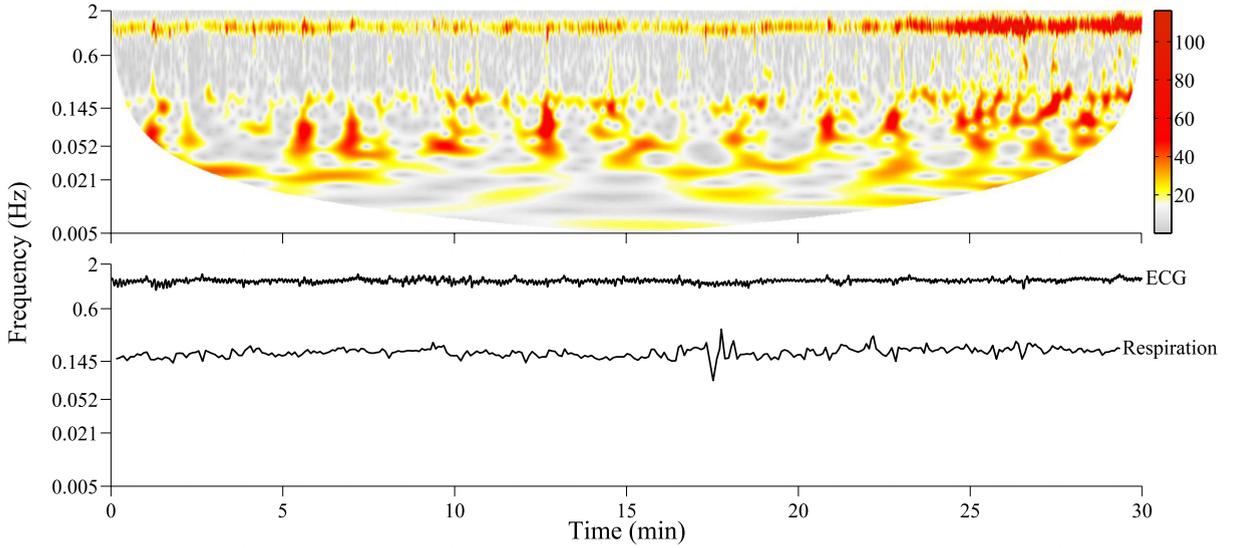
Overall, the sum of information from the time–frequency domain has identified the measured oscillation to have a time-dependent coupling to a low-frequency component. The mechanism behind these oscillations is still unknown, but this analysis has provided important insight by showing that they can be modelled as a deterministic non-autonomous system.

### 9.3. Cardiorespiratory system

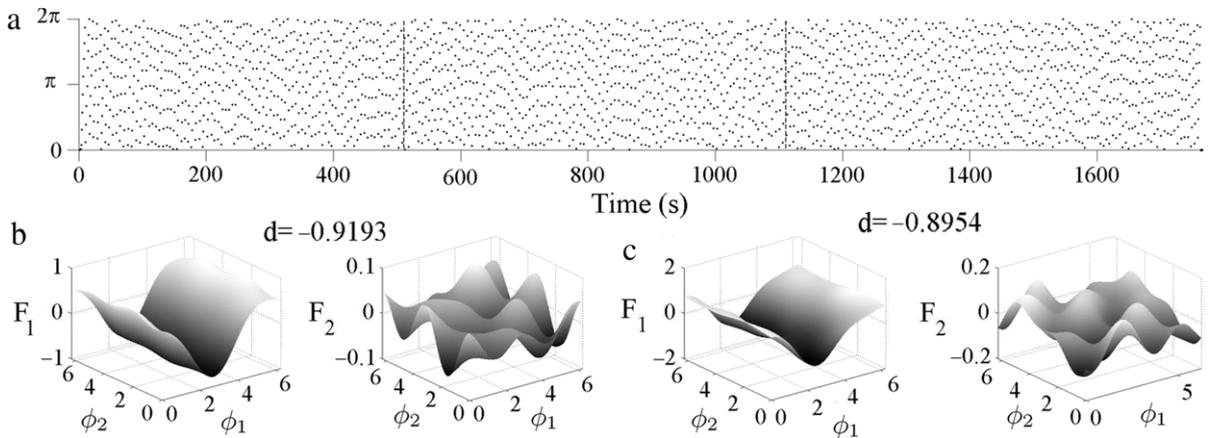
The methods are of course not restricted to fundamental Physics systems. Complex dynamics is also found in living systems, which can be thought of as non-autonomous systems in the strictest sense due to their openness and time-dependent nature. Fig. 54 shows time–frequency analysis for the peripheral blood flow in the skin, measured using laser Doppler flowmetry (LDF), an electrocardiogram (ECG) of the heart, and respiration, measured using a belt around the chest. The marked frequencies correspond to the following physiologically meaningful intervals [3]: nitric oxide (NO)-independent endothelial [0.005–0.0095 Hz], NO-dependent endothelial [0.0095–0.021 Hz], neurogenic [0.021–0.052 Hz], myogenic [0.052–0.145 Hz], respiration [0.145–0.6 Hz] and cardiac [0.6–2 Hz].

The advantage of viewing dynamics in the time–frequency domain is that it does not matter how they arise in the time series. In the LDF signal the oscillations of the heart and respiration are embedded in a complex array of other fluctuations. However, they are in agreement with the instantaneous frequencies extracted using an entirely different method from the ECG and respiration signals.

Fig. 55 shows the interaction analysis for the ECG and respiration signals of the same subject. The synchrogram indicates that there are no significant synchronization epochs in the recording. If this was the only analysis performed then one might argue that there is little interaction between the heart and respiration. However, the method of Bayesian inference is able to track the time-dependent coupling from the respiration to the heart. This can be observed directly from the coupling functions where the coupling function for the respiration to heart coupling maintains the same shape throughout the



**Fig. 54.** Time-frequency analysis of physiological signals from a single subject recorded over 30 min. The continuous wavelet transform of an LDF probe recording is plotted at the top, while instantaneous frequencies of the ECG and respiration signals calculated using marked events are shown below. The respiration signal was smoothed using a 0.2 s moving average prior to marking peaks. R-peaks were used as the marked events in the ECG signal. Both signals were sampled at 1200 Hz. The LDF time series was sampled at 40 Hz and the calculation used the Morlet wavelet with  $\omega_c = 1$ .



**Fig. 55.** Interactions between the heart ( $\phi_1$ ) and respiration ( $\phi_2$ ): (a) Synchrogram showing the position of the R-peaks in the ECG with respect to the respiration protophase; (b,c) Coupling functions calculated using Bayesian inference for the windows centred at 450 s and 1350 s respectively. The calculation used 50 s windows with propagation constant  $p = 0.8$ . Phases were extracted from the synchrosqueezed wavelet transform [109].

recording (which is in agreement with Fig. 54 where modulation of the heart rate by the respiration can be seen directly). Since the coupling is time-dependent, other techniques such as the EMA method are unable to reproduce these results.

### 10. Conclusion

Theories related to complex systems, such as the theory of turbulence, have been in development for well over 100 years but the analysis of inverse problems involving such systems has only followed after the much more recent advances in nonlinear dynamics. Previously, the framework of this analysis has been confined to specific types of complex systems, in particular those that are chaotic or stochastic. However, complex dynamics can additionally be generated in a third way through deterministic, non-autonomous forcing. Neither of the previous frameworks are compatible with the analysis of these systems, which means a different approach is needed.

This paper has provided a review of the applicability of time series analysis methods to different types of complex systems. Hence, this paper serves as a guide as to which are the most useful methods when tackling an inverse problem. A

main feature of these methods is the retention of information in the time domain. Even when these methods have a linear basis, such as the wavelet transform, the fact that the phases of the components can be tracked in time allows nonlinearity to be accounted for [107]. In addition, the time–frequency domain makes it possible to see the difference between chaotic or stochastic systems and a fairly stable non-autonomous system. In contrast, the traditional approaches to detect stability, offered by measures such as Lyapunov exponents, are unable to detect this difference.

Perhaps the greatest achievement of these methods is a way to uncover interactions between a complex system and a Poincaré oscillator. Despite the complex nature of this interaction, it is still possible to detect hints of synchronization, in addition to changes in the coupling strength and direction.

Lastly, the methods presented here have been applied to real data. The results provide a description of the dynamics of the underlying systems, which in turn aids the understanding of the processes involved. This is the crux of all inverse problems, regardless of the nature of the system. However, the extension to complex systems will surely open up more avenues of investigation into systems that may have previously seemed mysterious and incomprehensible.

## Acknowledgements

We are very grateful for the valuable discussions with Neil Drummond, Martin Horvat, Dmytro Iatsenko, Janez Jamšek, Jamie Luxmoore, Peter McClintock, Benjamin Pennington, Antonio Politi, Lawrence Sheppard, Tomislav Stankovski and Yevhen Suprunenko. We also thank Victor Efimov, Kimitoshi Kono, Denis Konstantinov and Masamitsu Watanabe for providing data presented in the example applications. This work was supported by the Engineering and Physical Sciences Research Council (UK) (Grant No. EP/100999X1) and in part by the Slovenian Research Agency (Program No. P20232).

## Appendix A. Matlab toolbox

The Matlab codes for most of the methods presented in this paper are provided in the TACTS toolbox. Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.physrep.2014.04.001>.

## Appendix B. Numerical integration

### B.1. Heun scheme

The Heun scheme is defined as:

$$\begin{aligned}\tilde{x}_{n+1} &= x_n + hf(t_n, x_n), \\ x_{n+1} &= x_n + \frac{h}{2} (f(t_n, x_n) + f(t_{n+1}, \tilde{x}_{n+1})), \\ \frac{dx}{dt} &= f(t, x).\end{aligned}\tag{B.1}$$

This is a second order integration scheme, calculating a prediction  $\tilde{x}_{n+1}$  of the next point in the trajectory using the current point  $x_n$  and then using that prediction to make a more accurate estimate of the gradient of the line connecting  $x_n$  to  $x_{n+1}$ . Higher order integration schemes also exist that are able to reduce the error further but with this they bring other limitations [209]. It should be noted that care must be taken when integrating systems with a noise term as the algorithm must be adjusted according to Honeycutt [60].

### B.2. Lie scheme

The Lie method is defined by the following differential operator:

$$L_o = f_1 \frac{\partial}{\partial x^1} + f_2 \frac{\partial}{\partial x^2} + \dots + f_d \frac{\partial}{\partial x^d},\tag{B.2}$$

where  $x^i$  are the dimensions of the system,  $f_i \equiv \frac{dx^i}{dt}(x_n^1, x_n^2, \dots, x_n^d)$  and  $d$  is the total number of dimensions. The integration scheme is then given by

$$\begin{bmatrix} x_{n+1}^1 \\ x_{n+1}^2 \\ \vdots \\ x_{n+1}^d \end{bmatrix} = e^{hL_o} \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^d \end{bmatrix},\tag{B.3}$$

which can be approximated by the Taylor series:

$$\begin{bmatrix} x_{n+1}^1 \\ x_{n+1}^2 \\ \vdots \\ x_{n+1}^d \end{bmatrix} \approx \sum_{k=0}^M \frac{h^k}{k!} L_0^k \begin{bmatrix} x_n^1 \\ x_n^2 \\ \vdots \\ x_n^d \end{bmatrix}, \tag{B.4}$$

where  $M$  is the order of the expansion (higher  $M$  corresponds to a greater accuracy of integration). The proof of (B.3) is given in [210].

### Appendix C. Derivations

#### C.1. Formulation of the Fourier transform

Consider integrating equation (20) over  $-\pi < n < \pi$ ,

$$\begin{aligned} \int_{-\pi}^{\pi} f(t) dt &= \int_{-\pi}^{\pi} a_0 dt + \int_{-\pi}^{\pi} \sum_{\omega=1}^{\infty} [a_{\omega} \cos(\omega t) + b_{\omega} \sin(\omega t)] dt \\ &= 2\pi a_0 + \sum_{\omega=1}^{\infty} a_{\omega} \int_{-\pi}^{\pi} \cos(\omega t) dt + \sum_{\omega=1}^{\infty} b_{\omega} \int_{-\pi}^{\pi} \sin(\omega t) dt \\ &= 2\pi a_0 + \sum_{\omega=1}^{\infty} \frac{a_{\omega}}{\omega} [\sin(\omega\pi) - \sin(-\omega\pi)] + \sum_{\omega=1}^{\infty} \frac{b_{\omega}}{\omega} [\cos(-\omega\pi) - \cos(\omega\pi)]. \end{aligned} \tag{C.1}$$

Since  $\omega$  is an integer, the two summations on the right hand side become 0. Rearranging, this gives an equation for  $a_0$ :

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) dt. \tag{C.2}$$

Similar derivations result in the following equations for  $a_{\omega}$  and  $b_{\omega}$ :

$$\begin{aligned} a_{\omega} &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(\omega t) dt, \\ b_{\omega} &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(\omega t) dt, \quad \omega = 1, 2, 3, \dots \end{aligned} \tag{C.3}$$

However, this still accounts only for the part of  $f(t)$  spanning  $-\pi < t < \pi$ . To find the coefficients for a function that spans some arbitrary period  $L$  there needs to be a change of variable. Setting  $t' = 2\pi t/L$ , another function can be defined as  $g(t') = f(t)$  and the new integration variable is given by  $dt' = 2\pi dt/L$ . Substituting into the previous equations gives:

$$g(t') = a_0 + \sum_{\omega=1}^{\infty} [a_{\omega} \cos(\omega t') + b_{\omega} \sin(\omega t')], \tag{C.4}$$

where

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} g(t') dt', \\ a_{\omega} &= \frac{1}{\pi} \int_{-\pi}^{\pi} g(t') \cos(\omega t') dt', \\ b_{\omega} &= \frac{1}{\pi} \int_{-\pi}^{\pi} g(t') \sin(\omega t') dt', \quad \omega = 1, 2, 3, \dots \end{aligned} \tag{C.5}$$

Then, returning to the original variable:

$$f(t) = a_0 + \sum_{\omega=1}^{\infty} \left[ a_{\omega} \cos\left(\frac{2\pi\omega t}{L}\right) + b_{\omega} \sin\left(\frac{2\pi\omega t}{L}\right) \right], \tag{C.6}$$

where

$$\begin{aligned} a_0 &= \frac{1}{L} \int_{-\frac{L}{2}}^{\frac{L}{2}} f(t) dt, \\ a_\omega &= \frac{2}{L} \int_{-\frac{L}{2}}^{\frac{L}{2}} f(t) \cos\left(\frac{2\pi\omega t}{L}\right) dt, \\ b_\omega &= \frac{2}{L} \int_{-\frac{L}{2}}^{\frac{L}{2}} f(t) \sin\left(\frac{2\pi\omega t}{L}\right) dt, \quad \omega = 1, 2, 3, \dots \end{aligned} \quad (\text{C.7})$$

These equations now define the Fourier series for a function over a finite period. Note that  $\omega$  now represents the standard frequency multiplied by  $L$  rather than the angular frequency. The step from here to the discrete Fourier transform is a very small one. The two sets of coefficients  $a_\omega$  and  $b_\omega$ , corresponding the sine and cosine terms can instead be replaced with a single complex coefficient  $A_\omega$  by using Euler's formula:

$$e^{\frac{2\pi i\omega t}{L}} = \cos\left(\frac{2\pi\omega t}{L}\right) + i \sin\left(\frac{2\pi\omega t}{L}\right). \quad (\text{C.8})$$

Using the same basis as before and setting  $a_\omega$  and  $b_\omega$  respectively as the real and imaginary parts of  $A_\omega$ , the complex Fourier series can then be written as:

$$f(t) = \sum_{\omega=-\infty}^{\infty} A_\omega e^{\frac{2\pi i\omega t}{L}}, \quad (\text{C.9})$$

where

$$A_\omega = \frac{1}{L} \int_{-\frac{L}{2}}^{\frac{L}{2}} f(t) e^{-\frac{2\pi i\omega t}{L}} dt. \quad (\text{C.10})$$

In the discrete case, integrals and summations spanning an infinite range cannot be used. The discrete Fourier transform is therefore defined as [211]:

$$F(\omega) = \sum_{n=0}^{N-1} f(n) e^{-\frac{2\pi i\omega n}{N}}. \quad (\text{C.11})$$

Additionally, the inverse transform is defined as:

$$f(n) = \frac{1}{N} \sum_{\omega=0}^{N-1} F_\omega e^{\frac{2\pi i\omega n}{N}}. \quad (\text{C.12})$$

## C.2. Least squares polynomial fitting

Starting with Eq. (24), the least squares error is defined as:

$$S = \sum_{i=1}^n [x_i - f(t_i)]^2 = \sum_{i=1}^n [x_i - (a_0 + a_1 t_i + a_2 t_i^2 + \dots + a_m t_i^m)]^2 = \min., \quad (\text{C.13})$$

where  $x_i$  are the data points corresponding to the times  $t_i$ . To find the least squares error approximation, the first derivatives with respect to the unknown coefficients are used:

$$\frac{\partial S}{\partial a_0} = 2 \sum_{i=1}^n [x_i - (a_0 + a_1 t_i + a_2 t_i^2 + \dots + a_m t_i^m)] \quad (\text{C.14})$$

$$\frac{\partial S}{\partial a_1} = 2 \sum_{i=1}^n t_i [x_i - (a_0 + a_1 t_i + a_2 t_i^2 + \dots + a_m t_i^m)] \quad (\text{C.15})$$

$$\frac{\partial S}{\partial a_2} = 2 \sum_{i=1}^n t_i^2 [x_i - (a_0 + a_1 t_i + a_2 t_i^2 + \dots + a_m t_i^m)] \quad (\text{C.16})$$

⋮

$$\frac{\partial S}{\partial a_m} = 2 \sum_{i=1}^n t_i^m [x_i - (a_0 + a_1 t_i + a_2 t_i^2 + \dots + a_m t_i^m)]. \quad (\text{C.17})$$

The condition where  $S = \min.$  is then found when  $\frac{\partial S}{\partial a_k} = 0$  for  $k = 0, 1, 2, \dots, m$ . The following linear equations are therefore derived:

$$\sum_{i=1}^n x_i = na_0 + a_1 \sum_{i=1}^n t_i + a_2 \sum_{i=1}^n t_i^2 + \dots + a_m \sum_{i=1}^n t_i^m \tag{C.18}$$

$$\sum_{i=1}^n x_i t_i = a_0 \sum_{i=1}^n t_i + a_1 \sum_{i=1}^n t_i^2 + a_2 \sum_{i=1}^n t_i^3 + \dots + a_m \sum_{i=1}^n t_i^{m+1} \tag{C.19}$$

$$\sum_{i=1}^n x_i t_i^2 = a_0 \sum_{i=1}^n t_i^2 + a_1 \sum_{i=1}^n t_i^3 + a_2 \sum_{i=1}^n t_i^4 + \dots + a_m \sum_{i=1}^n t_i^{m+2} \tag{C.20}$$

$$\vdots$$

$$\sum_{i=1}^n x_i t_i^m = a_0 \sum_{i=1}^n t_i^m + a_1 \sum_{i=1}^n t_i^{m+1} + a_2 \sum_{i=1}^n t_i^{m+2} + \dots + a_m \sum_{i=1}^n t_i^{2m}. \tag{C.21}$$

This can then be converted to matrix form  $\mathbf{x} = \mathcal{T} \mathbf{a}$  where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathcal{T} = \begin{bmatrix} 1 & t_1 & \dots & t_1^m \\ 1 & t_2 & \dots & t_2^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & \dots & t_n^m \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}. \tag{C.22}$$

The vector  $\mathbf{a}$  can then be found numerically using a matrix equation solver, after which the values can be substituted into (24) to obtain the best-fitting polynomial for the points contained within the window.

## Appendix D. Statistics

### D.1. Binning

Calculating probability distributions adds another level of complexity to time series analysis. This is because the accuracy and precision of the distribution is dependent on the type of *binning* used. To generate a probability distribution a set of evenly-spaced intervals (bins) needs to be defined between the minimum and maximum amplitude in the time series. An estimate of the probability distribution is then found simply by counting the number of points that fall into each bin.

If small bins are used then it is possible to see more details in the distribution but accuracy is lost due to the finiteness of the time series. If large bins are used then a greater number of points will be allocated to each bin, causing the accuracy to increase but the precision to be lost. The procedure is therefore a balancing act, with the best advice being to avoid either extreme and settle for an optimal bin size somewhere in-between. In particular, the number of bins should be dependent on the amount of data that is binned (i.e. the number of points in the time series,  $N$ ), which for a Gaussian distributions can be derived as [212]

$$\text{Number of bins} = e^{0.636} (N - 1)^{2/5}. \tag{D.1}$$

Other estimators also use adaptive bin sizes where the size is dependent on the location in the distribution [213].

### D.2. Surrogate data

The significance of certain features in time series can usually be tested by making repeat recordings of the system. However, even then it is possible for systematic bias to skew the interpretation of these features. An alternative approach to test for significance has therefore been developed and is known as surrogate data [201,214]. Starting with a single time series, this technique generates a new time series which preserves the properties that are consistent with a null hypothesis but randomizes all others. Reapplying the same analytical methods to the surrogate time series will therefore reveal the bias in the result.

One of the simplest types of surrogate data are the independent identically distributed (IID) surrogates [215,3]. These are used to test the null hypothesis that the dynamics of the time series can be modelled as any independent random process with a given probability distribution. They are generated by shuffling the order of the data points in the time series, moving each one to a random point in time. The result is a time series with the same distribution as the original but with none of its temporal structure. IID surrogates are most suitable for noisy time series with a flat frequency spectrum [215].

For time series which do contain structure in the frequency spectrum another type of surrogate data can be produced by using the Fourier transform. This is done by randomizing the complex phase of  $A_\omega$  from (C.10) so that each point

in the spectrum  $F$  has the same amplitude but different values for the real and imaginary parts. A real time series can then be reconstructed from the Fourier transform by making the second half of the transform (corresponding to negative frequencies) the complex conjugate of the first half. The surrogate time series is then given by the inverse Fourier transform:

$$f(n) = \frac{1}{L} \int_{-\frac{L}{2}}^{\frac{L}{2}} A_{\omega} e^{-\frac{2\pi i \omega n}{L}} d\omega. \quad (\text{D.2})$$

This procedure preserves the frequency spectrum of the original time series but no other properties. A significant problem is that the points in the surrogate time series are forced into a Gaussian distribution [214]. This has important repercussions if the surrogate time series is being used to test for significance in a method that involves probability distributions, such as the information theoretic approach. It also has an impact on phase space methods, since the embedded time series will also have a Gaussian distribution, altering the information about the actual distribution of states. However, for complexity analysis the distribution of states is often the feature that is being tested, which makes these surrogates a viable option.

If the distribution of the data was to be preserved in the surrogates it is important to consider how this will change the null hypothesis of the significance test. For the previous algorithm, the hypothesis is simply that the time series was generated by a linear Gaussian process. It is clear from the examples of stochastic systems in this paper that nonlinearities can drastically alter the distribution (take for example the shapes of the attractors in phase space). Therefore, a new null hypothesis can be formulated where the time series is generated from an invertible *nonlinear* transformation of a linear Gaussian process.

The following procedure generates surrogates to test this hypothesis using the amplitude adjusted Fourier transform (AAFT) [216,214]. To remove the effect of any nonlinear transformation, the time series is *rescaled* so that it has a Gaussian distribution. This simply means ordering the data from highest to lowest and substituting corresponding values from an equal-length time series of Gaussian white noise at the same positions in time. The Fourier transform is then applied as before, with the phases of  $A_{\omega}$  being randomized for every  $\omega$ . After using the inverse Fourier transform, the distribution is restored by rescaling the surrogate using the data from the original time series.

A problem with the AAFT procedure is that the resulting surrogates do not usually have the same frequency spectrum as the original time series. In order to preserve this spectrum, *iterative* amplitude adjusted Fourier transform (IAAFT) surrogates are needed [217,216]. In this case, the original time series data is first shuffled and then the Fourier transform of the new time series is calculated. The amplitudes of the Fourier components are adjusted to match the original, after which the data is transformed back to the time domain. As for the AAFT surrogates, the data is rescaled so that it contains the same values as the original time series. However, this process adjusts the frequency spectrum away from the desired distribution. The procedure is then iterated for a second time, transforming to the frequency domain, adjusting the amplitudes to match the original and then transforming back and rescaling. The iterations continue until the order of the data after rescaling converges to the order of the previous iteration.

## References

- [1] A. Pikovsky, M. Rosenblum, J. Kurths, *Synchronization – A Universal Concept in Nonlinear Sciences*, Cambridge University Press, Cambridge, 2001.
- [2] S.M. Reppert, D.R. Weaver, Coordination of circadian timing in mammals, *Nature* 418 (2002) 935–941.
- [3] Y. Shiozai, A. Stefanovska, P.V.E. McClintock, Nonlinear dynamics of cardiovascular ageing, *Phys. Rep.* 488 (2010) 51–110.
- [4] C.S. Bretherton, M. Widmann, V.P. Dymnikov, J.M. Wallace, I. Blade, The effective number of spatial degrees of freedom of a time-varying field, *J. Clim.* 12 (7) (1999) 1990–2009.
- [5] Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*, Springer-Verlag, Berlin, 1984.
- [6] S.H. Strogatz, Exploring complex networks, *Nature* 410 (2001) 268–276.
- [7] H. Haken, *Synergetics: An Introduction and Advanced Topics*, Springer, Berlin, 2004.
- [8] Y.F. Suprunenko, P.T. Clemson, A. Stefanovska, Chronotaxic systems: a new class of self-sustained nonautonomous oscillators, *Phys. Rev. Lett.* 111 (2) (2013) 024101.
- [9] Y.F. Suprunenko, P.T. Clemson, A. Stefanovska, Chronotaxic systems with separable amplitude and phase dynamics, *Phys. Rev. E* 89 (1) (2014) 012922.
- [10] J.P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors, *Rev. Modern Phys.* 57 (3) (1983) 617–656.
- [11] R. Friedrich, J. Peinke, M. Sahimi, M.R.R. Tabar, Approaching complexity by stochastic methods: from biological systems to turbulence, *Phys. Rep.* 506 (5) (2011) 87–162.
- [12] P.T. Clemson, Y.F. Suprunenko, T. Stankovski, A. Stefanovska, Inverse approach to chronotaxic systems for single-variable time series, *Phys. Rev. E* 89 (3) (2014) 032904.
- [13] L.J. Larson-Prior, R. Oostenveld, S.D. Penna, G. Michalareas, F. Prior, A. Babajani-Feremi, J.-M. Schoffelen, L. Marzetti, F. de Pasquale, F.D. Pompeo, J. Stout, M. Woolrich, Q. Luo, R. Bucholz, P. Fries, V. Pizzella, G. Romani, M. Corbetta, A. Snyder, for the WU-Minn HCP Consortium, adding dynamics to the human connectome project with MEG, *NeuroImage* 80 (2013) 190–201.
- [14] T.T. Nakagawa, V.K. Jirsa, A. Spiegler, A.R. McIntosh, G. Deco, Bottom up modeling of the connectome: linking structure and function in the resting brain and their changes in aging, *NeuroImage* 80 (2013) 318–329.
- [15] M. Brett, W. Penny, S. Kiebel, Introduction to random field theory, in: R.S. Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, C.J. Price, S. Zeki, J.T. Ashburner, W.D. Penny (Eds.), *Human Brain Function*, second ed., Academic Press, London, 2003, pp. 867–879.
- [16] A. Bernjak, P.B.M. Clarkson, P.V.E. McClintock, A. Stefanovska, Low-frequency blood flow oscillations in congestive heart failure and after  $\beta$ 1-blockade treatment, *Microvasc. Res.* 76 (3) (2008) 224–232.
- [17] L.W. Sheppard, V. Vuksanović, P.V.E. McClintock, A. Stefanovska, Oscillatory dynamics of vasoconstriction and vasodilation identified by time-localized phase coherence, *Phys. Med. Biol.* 56 (2011) 3583–3601.
- [18] P. Kvandal, L. Sheppard, S.A. Landsverk, A. Stefanovska, K.A. Kirkebøen, Impaired cerebrovascular reactivity after acute traumatic brain injury can be detected by wavelet phase coherence analysis of the intracranial and arterial blood pressure signals, *J. Clin. Monit. Comput.* 27 (4) (2013) 375–383.
- [19] D.G. Spiller, C.D. Wood, D.A. Rand, M.R.H. White, Measurement of single-cell dynamics, *Nature* 465 (7299) (2010) 736–745.
- [20] Q.A. Ni, A. Ganesan, N.N. Aye-Han, X.X. Gao, M.D. Allen, A. Levchenko, J. Zhang, Signaling diversity of PKA achieved via a  $\text{Ca}^{2+}$ -cAMP-PKA oscillatory circuit, *Nat. Chem. Biol.* 7 (1) (2011) 34–40.

- [21] M.B. Ahrens, M.B. Orger, D.N. Robson, J.M. Li, P.J. Keller, Whole-brain functional imaging at cellular resolution using light-sheet microscopy, *Nat. Methods* 10 (5) (2013) 413–420.
- [22] M.A. Aon, S. Cortassa, B. O'Rourke, Mitochondrial oscillations in physiology and pathophysiology, *Adv. Exp. Med. Biol.* 641 (2008) 98–117.
- [23] F.T. Kurz, M.A. Aon, B. O'Rourke, A.A. Armoundas, Wavelet analysis reveals heterogeneous time-dependent oscillations of individual mitochondria, *Am. J. Physiol. Heart Circ. Physiol.* 299 (2010) H1736–H1740.
- [24] V. Ruban, Y. Kodama, M. Ruderman, J. Dudley, R. Grimshaw, P.V.E. McClintock, M. Onorato, C. Kharif, E. Pelinovsky, T. Soomere, G. Lindgren, N. Akhmediev, A. Slunyaev, D. Solli, C. Ropers, B. Jalali, F. Dias, A. Osborne, Rogue waves – towards a unifying concept? Discussions and debates, *Eur. Phys. J. Spec. Top.* 185 (2010) 5–15.
- [25] V.B. Efimov, A.N. Ganshin, G.V. Kolmakov, P.V.E. McClintock, L.P. Mezhev-Deglin, Rogue waves in superfluid helium, *Eur. Phys. J. Spec. Top.* 185 (2010) 181–193.
- [26] S. Vaughan, Random time series in astronomy, *Phil. Trans. R. Soc. A* 371 (1984) (2013) 20110549.
- [27] N. Aoki, H. Hiraide, *Topological Theory of Dynamical Systems*, North-Holland, Amsterdam, Netherlands, 1994.
- [28] J.D. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications*, third ed., Springer, New York, 2003.
- [29] S. Strogatz, *Nonlinear Dynamics And Chaos*, Westview Press, Boulder, 2001.
- [30] O. Reynolds, An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels, *Phil. Trans. R. Soc.* 174 (1883) 935–982.
- [31] G. Stokes, On the effect of the internal friction of fluids on the motion of pendulums, *Trans. Camb. Phil. Soc.* 9 (1851) 8–106.
- [32] B. Eckhardt, E. Ott, S.H. Strogatz, D.M. Abrams, A. McRobie, Modeling walker synchronization on the millennium bridge, *Phys. Rev. E* 75 (2) (2007) 021110.
- [33] A. Kolmogorov, Local structure of turbulence in an incompressible fluid for very large Reynolds numbers, *Dokl. Akad. Nauk SSSR* 31 (1941) 301–305.
- [34] P.S. Landa, Turbulence in nonclosed fluid flows as a noise-induced phase transition, *Europhys. Lett.* 36 (6) (1996) 401–406.
- [35] L. Onsager, Statistical hydrodynamics, *Nuovo Cimento, Suppl.* 6 (2) (1949) 279–287.
- [36] M.-M. Mac Low, R.S. Klessen, Control of star formation by supersonic turbulence, *Rev. Modern Phys.* 76 (1) (2004) 125–194.
- [37] A.J. Turner, G. Gogoberidze, S.C. Chapman, Nonaxisymmetric anisotropy of solar wind turbulence as a direct test for models of magnetohydrodynamic turbulence, *Phys. Rev. Lett.* 108 (8) (2012) 085001.
- [38] I.S. Aranson, L. Kramer, The world of the complex Ginzburg–Landau equation, *Rev. Mod. Phys.* 74 (1) (2002) 99–143.
- [39] T. Bohr, M.H. Jensen, G. Paladin, A. Vulpiani, *Dynamical Systems Approach to Turbulence*, Cambridge University Press, Cambridge, 1998.
- [40] B.J. Jaworski, A.K. Kohli, Market orientation: antecedents and consequences, *J. Mark.* 57 (3) (1993) 53–70.
- [41] G.L. Mellor, T. Yamada, Development of a turbulence closure model for geophysical fluid problems, *Rev. Geophys.* 20 (4) (1982) 851–875.
- [42] J.-P. Eckmann, Roads to turbulence in dissipative dynamical systems, *Rev. Mod. Phys.* 53 (4) (1981) 643–654.
- [43] H. Haken, *Advanced Synergetics*, Springer-Verlag, Berlin, 1983.
- [44] P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, G. Vattay, *Chaos: Classical and Quantum*, ChaosBook.org, Niels Bohr Institute, Copenhagen, 2012.
- [45] W. Hillebrandt, F. Kupka, An introduction to turbulence, in: *Interdisciplinary Aspects of Turbulence*, in: *Lect. Notes Phys.*, vol. 756, Springer, Berlin, Heidelberg, 2009, pp. 1–20.
- [46] J. Uffink, Boltzmann's work in statistical physics, in: *The Stanford Encyclopedia of Philosophy*, Stanford University, Stanford, 2008.
- [47] W.M. Haddad, S.G. Nersesov, V. Chellaboina, Heat flow, work energy, chemical reactions, and thermodynamics: a dynamical systems perspective, in: *Thermodynamics*, InTech, Rijeka, 2011, pp. 51–72.
- [48] S. Carnot, *Réflexions sur la Puissance Motrice du Feu et sur les Machines Propres à Développer cette Puissance*, Bachelier, Bachelier, Paris, 1824.
- [49] I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes*, Interscience Publishers, New York, 1968.
- [50] A. Rényi, On measures of information and entropy, in: *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, University of California Press, Berkeley, 1961, pp. 547–561.
- [51] E.N. Lorenz, Deterministic non-periodic flow, *J. Atmospheric Sci.* 20 (2) (1963) 130–141.
- [52] J. Kurths, M.C. Romano, M. Thiel, G.V. Osipov, M.V. Ivanchenko, I.Z. Kiss, J.L. Hudson, Synchronization analysis of coupled noncoherent oscillators, *Nonlinear Dynam.* 44 (1–4) (2006) 135–149.
- [53] G. Duffing, Erzwungene schwingungen bei veränderlicher eigenfrequenz und ihre technische bedeutung, in: *Issues 41–42 of Sammlung Vieweg*, F. Vieweg & Sohn, 1918.
- [54] I. Kovacic, M.J. Brennan, *The Duffing Equation: Nonlinear Oscillators and Their Behaviour*, John Wiley & Sons, Chichester, 2011.
- [55] P. Holmes, A nonlinear oscillator with a strange attractor, *Phil. Trans. R. Soc. A* 292 (1394) (1979) 419–448.
- [56] F.C. Moon, P.J. Holmes, A magnetoelastic strange attractor, *J. Sound Vib.* 65 (2) (1979) 275–296.
- [57] P. Holmes, D. Rand, The bifurcations of Duffing's equation: an application of catastrophe theory, *J. Sound Vib.* 44 (2) (1976) 237–253.
- [58] M.I. Dykman, P.V.E. McClintock, R. Mannella, N. Stokes, Stochastic resonance in the linear and nonlinear responses of a bistable system to a periodic field, *JETP Lett.* 52 (3) (1990) 141–144.
- [59] M.I. Dykman, R. Mannella, P.V.E. McClintock, N.G. Stocks, Stochastic resonance in bistable systems – comment, *Phys. Rev. Lett.* 65 (20) (1990) 2606–2606.
- [60] R.L. Honeycutt, Stochastic Runge–Kutta algorithms. I. White noise, *Phys. Rev. A* 45 (1992) 600–603.
- [61] H. Risken, *The Fokker–Planck Equation*, second ed., Springer, Berlin, 1993.
- [62] R. Mannella, P.V.E. McClintock, Itô versus Stratonovich: 30 years later, *Fluct. Noise Lett.* 11 (1) (2012) 1240010.
- [63] A. Bahraminasab, F. Ghasemi, A. Stefanovska, P.V.E. McClintock, H. Kantz, Direction of coupling from phases of interacting oscillators: a permutation information approach, *Phys. Rev. Lett.* 100 (8) (2008) 084101.
- [64] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge Univ. Press, 2003.
- [65] P.E. Kloeden, M. Rasmussen, *Nonautonomous Dynamical Systems*, American Mathematical Soc, 2011.
- [66] A. Stefanovska, Coupled oscillators: complex but not complicated cardiovascular and brain interactions, *IEEE Eng. Med. Biol. Mag.* 26 (6) (2007) 25–29.
- [67] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, New York, 1982.
- [68] H. Nyquist, Certain topics in telegraph transmission theory, *Trans. AIEE* 47 (1928) 617–644.
- [69] C.E. Shannon, Communication in the presence of noise, *Proc. Inst. Rad. Eng.* 37 (1) (1949) 10–21.
- [70] C. Chatfield, *The Analysis of Time Series: an Introduction*, sixth ed., Chapman & Hall/CRC, Boca Raton, FL, 2004.
- [71] B. Kehlet, A. Logg, A Reference Solution for the Lorenz System on  $[0, 1000]$ , in: *AIP Proc.*, vol. 1281, American Institute of Physics, New York, 1981, pp. 1635–1638.
- [72] E. Sánchez, M.A. Matías, Experimental observation of a periodic rotating wave in rings of unidirectionally coupled analog Lorenz oscillators, *Phys. Rev. E* 57 (5) (1998) 6184–6186.
- [73] D.G. Luchinsky, P.V.E. McClintock, Irreversibility of classical fluctuations studied in analogue electrical circuits, *Nature* 389 (6650) (1997) 463–466.
- [74] D.G. Luchinsky, P.V.E. McClintock, M.I. Dykman, Analogue studies of nonlinear systems, *Rep. Progr. Phys.* 61 (8) (1998) 889–997.
- [75] C. Grebogi, S.M. Hammel, J.A. Yorke, T. Sauer, Shadowing of physical trajectories in chaotic dynamics: containment and refinement, *Phys. Rev. Lett.* 65 (13) (1990) 1527–1530.
- [76] J. Jamšek, M. Paluš, A. Stefanovska, Detecting couplings between interacting oscillators with time-varying basic frequencies: instantaneous wavelet bispectrum and information theoretic approach, *Phys. Rev. E* 81 (3) (2010) 036207.
- [77] T. Stankovski, A. Duggento, P.V.E. McClintock, A. Stefanovska, Inference of time-evolving coupled dynamical systems in the presence of noise, *Phys. Rev. Lett.* 109 (2012) 024101.
- [78] L. Biferale, Shell models of energy cascade in turbulence, *Annu. Rev. Fluid Mech.* 35 (2003) 441–468.
- [79] M.T. Heideman, D.H. Johnson, C.S. Burrus, Gauss and the history of the fast Fourier transform, *Rev. Modern Phys.* 1 (4) (1984) 14–21.

- [80] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (8) (1964) 1627–1639.
- [81] J.F. Kenney, E.S. Keeping, Moving averages, in: *Mathematics of Statistics*, Pt. 1, third ed., Van Nostrand, Princeton, 1962, pp. 221–223.
- [82] R. Bracewell, *The Fourier Transform & Its Applications*, third ed., McGraw Hill, New York, 2000.
- [83] J.W. Cooley, J.W. Tukey, An algorithm for the machine calculation of complex fourier series, *Math. Comp.* 19 (90) (1965) 297–301.
- [84] E.J. Godolphin, S.E. Johnson, Decomposition of time series dynamic linear models, *J. Time Ser. Anal.* 24 (5) (2003) 513–527.
- [85] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond. Ser. A* 454 (1998) 903–995.
- [86] J.H. Ahlberg, E.N. Nilson, J.L. Walsh, The theory of splines and their applications, in: *Mathematics in Science and Engineering*, Academic Press, New York, 1967.
- [87] N.E. Huang, Z. Shen, S.R. Long, A new view of nonlinear water waves: the Hilbert spectrum, *A. Rev. Fluid Mech.* 31 (1999) 417–457.
- [88] N.E. Huang, Z. Shen, S.R. Long, A confidence limit for the empirical mode decomposition and Hilbert spectral analysis, *Proc. R. Soc. Lond. Ser. A* 459 (2003) 2317–2345.
- [89] G. Kaiser, *A Friendly Guide to Wavelets*, Birkhäuser, Boston, 1994.
- [90] M. Smith, T. Barnwell, Exact reconstruction techniques for tree-structured subband coders, *IEEE Trans. Acoust. Speech Signal Process.* 34 (3) (1986) 434–441.
- [91] M. Vetterli, Filter banks allowing perfect reconstruction, *Signal Process.* 10 (3) (1986) 219–244.
- [92] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.
- [93] L. Keselbrenner, S. Akselrod, Selective discrete fourier transform algorithm for time–frequency analysis: method and application on simulated and cardiovascular signals, *IEEE Trans. Biol. Eng.* 43 (8) (1996) 789–802.
- [94] H. Li, L. Yang, D. Huang, The study of the intermittency test filtering character of Hilbert–Huang transform, *Math. Comput. Simulation* 70 (2005) 22–32.
- [95] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, *Adv. Adapt. Data Anal.* 1 (1) (2009) 1–41.
- [96] A. Kolmogorov, A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds number, *J. Fluid Mech.* 13 (1) (1961) 82–85.
- [97] C. Meneveau, Analysis of turbulence in the orthonormal wavelet representation, *J. Fluid Mech.* 232 (1991) 469–520.
- [98] K.R. Sreenivasan, R.A. Antonia, The phenomenology of small-scale turbulence, *Annu. Rev. Fluid Mech.* 29 (1997) 435–472.
- [99] D.V. Anosov, A.A. Bolibruch, *The Riemann–Hilbert Problem (Aspects of Mathematics)*, Friedrich Vieweg & Sohn Verlagsgesellschaft mbH, Wiesbaden, 1994.
- [100] D. Gabor, Theory of communication, *J. IEEE* 93 (1946) 429–457.
- [101] E.P. Wigner, On the quantum correction for thermodynamic equilibrium, *Phys. Rev.* 40 (5) (1932) 749–759.
- [102] J. Ville, Theory et application de la notion de signal analytique, *Cables et Transm.* 2A (1948) 61–77.
- [103] L. Galleania, L. Cohen, The Wigner distribution for classical systems, *Phys. Lett. A* 302 (4) (2002) 149–155.
- [104] M. Szmajda, K. Górecki, J. Mroczka, Gabor transform, SPWVD, Gabor–Wigner transform and wavelet transform – tools for power quality monitoring, *Metrol. Meas. Syst.* 17 (3) (2010) 383–396.
- [105] O. Rioul, M. Vetterli, Wavelets and signal processing, *IEEE Signal Proc. Mag.* 8 (4) (1991) 14–38.
- [106] P.S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, IOP Publishing, Bristol, 2002.
- [107] L.W. Sheppard, A. Stefanovska, P.V.E. McClintock, Detecting the harmonics of oscillations with time-variable frequencies, *Phys. Rev. E* 83 (2011) 016206.
- [108] I. Daubechies, S. Maes, A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models, in: *Wavelets in Medicine and Biology*, CRC Press, Bethesda, 1996, pp. 527–546.
- [109] I. Daubechies, J. Lu, H. Wu, Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool, *Appl. Comput. Harmon. Anal.* 30 (2) (2011) 243–261.
- [110] F. Takens, in: D.A. Rand, L.S. Young (Eds.), *Detecting Strange Attractors in Turbulence*, in: *Lecture notes in Mathematics*, vol. 898, Springer, New York, 1981.
- [111] R. Mañé, in: D.A. Rand, L.S. Young (Eds.), *On the Dimension of the Compact Invariant Sets of Certain Non-linear Maps*, in: *Lecture notes in Mathematics*, vol. 898, Springer, New York, 1981.
- [112] H.S. Kim, R. Eykholt, J.D. Salas, Nonlinear dynamics, delay times and embedding windows, *Physica D* 127 (1–2) (1999) 48–60.
- [113] M. Small, *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*, World Scientific, Singapore, 2005.
- [114] R. Brown, P. Bryant, H.D.I. Abarbanel, Computing the Lyapunov spectrum of a dynamical system from an observed time series, *Phys. Rev. A* 43 (6) (1991) 2787–2806.
- [115] A.M. Fraser, H.L. Swinney, Independent coordinates for strange attractors from mutual information, *Phys. Rev. A* 33 (2) (1986) 1134–1140.
- [116] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, L.S. Tsimring, The analysis of observed chaotic data in physical systems, *Rev. Modern Phys.* 65 (4) (1993) 1331–1392.
- [117] M.B. Kennel, R. Brown, H.D.I. Abarbanel, Determining embedding dimension for phase–space reconstruction using a geometrical construction, *Phys. Rev. A* 45 (6) (1992) 3403–3411.
- [118] C. Rhodes, M. Morari, False-nearest-neighbors algorithm and noise-corrupted time series, *Phys. Rev. E* 55 (5) (1997) 6162–6170.
- [119] L. Faes, G. Nollo, A. Porta, Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique, *Phys. Rev. E* 83 (5) (2011) 051112.
- [120] A.M. Lyapunov, The general problem of the stability of motion (translated into English by A. T. Fuller), *Internat. J. Control* 55 (1992) 531–773.
- [121] M. Sano, Y. Sawada, Measurement of the Lyapunov spectrum from a chaotic time series, *Phys. Rev. Lett.* 55 (10) (1985) 1082–1085.
- [122] M. Bračić, A. Stefanovska, Nonlinear dynamics of the blood flow studied by Lyapunov exponents, *Bull. Math. Biol.* 60 (3) (1998) 417–433.
- [123] V.I. Oseledets, A multiplicative ergodic theorem, *Trans. Moscow Math. Soc.* 19 (1968) 197–231.
- [124] J.-P. Eckmann, D. Ruelle, S. Ciliberto, Liapunov exponents from time series, *Phys. Rev. A* 34 (6) (1986) 4971–4979.
- [125] J. Holzfuss, U. Parlitz, Lyapunov exponents from time series, in: L. Arnold, J.-P. Eckmann (Eds.), *Lyapunov Exponents: Proceedings, Oberwolfach 1990*, in: *Lecture Notes in Mathematics*, vol. 1486, Springer-Verlag, Berlin, 1990, pp. 263–270.
- [126] U. Parlitz, Identification of true and spurious Lyapunov exponents from time series, *Int. J. Bifurcation Chaos* 2 (1) (1991) 155–165.
- [127] P. Grassberger, Generalized dimensions of strange attractors, *Phys. Lett. A* 97 (6) (1983) 227–230.
- [128] A. Renyi, On the dimension and entropy of probability distributions, *Acta Math. Acad. Sci. Hung.* 97 (1–2) (1959) 193–215.
- [129] P. Grassberger, R. Badii, A. Politi, Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors, *J. Stat. Phys.* 51 (1–2) (1988) 135–178.
- [130] P. Grassberger, I. Procaccia, Characterization of strange attractors, *Phys. Rev. Lett.* 50 (5) (1983) 346–349.
- [131] P. Grassberger, I. Procaccia, Measuring the strangeness of strange attractors, *Physica D* 9 (1983) 189–208.
- [132] F. Takens, in: B. Braaksma, H. Broer, F. Takens (Eds.), *On the Numerical Determination of the Dimension of an Attractor*, in: *Lecture Notes in Mathematics*, vol. 1125, Springer, Berlin, 1985, pp. 99–106.
- [133] J. Theiler, Estimating fractal dimension, *J. Opt. Soc. Amer. A* 7 (6) (1990) 1055–1073.
- [134] R. Badii, A. Politi, Hausdorff dimension and uniformity factor of strange attractors, *Phys. Rev. Lett.* 52 (19) (1984) 1661–1664.
- [135] P. Frederickson, J.L. Kaplan, E.D. Yorke, J.A. Yorke, The Liapunov dimension of strange attractors, *J. Differential Equations* 49 (2) (1983) 185–207.
- [136] R. Badii, A. Politi, Renyi dimensions from local expansion rates, *Phys. Rev. A* 35 (3) (1987) 1288–1293.
- [137] P. Comom, Independent component analysis, a new concept?, *Signal Process.* 36 (3) (1994) 287–314.
- [138] K. Karhunen, Zur spektraltheorie stochastischer prozesse, *Ann. Acad. Sci. Fenn. A1, Math. Phys.* 37.
- [139] M. Loève, *Fonctions Aleatoires De Second Ordre*, C. R. Acad. Sci., Paris, 1946.

- [140] R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D* 35 (3) (1989) 395–424.
- [141] M. Hožič, A. Stefanovska, Karhunen–Loève decomposition of peripheral blood flow, *Physica A* 281 (2000) 587–601.
- [142] A.R. Zeni, J.A.C. Gallas, Lyapunov exponents for a Duffing oscillator, *Physica D* 89 (1–2) (1995) 71–82.
- [143] A. Stefanovska, P. Krošelj, Correlation integral and frequency analysis of cardiovascular functions, *Open Syst. Inf. Dyn.* 4 (1997) 457–478.
- [144] A. Stefanovska, S. Strle, P. Krošelj, On the overestimation of the correlation dimension, *Phys. Lett. A* 235 (1) (1997) 24–30.
- [145] A.R. Osborne, A. Provenzale, Finite correlation dimension for stochastic systems with power-law spectra, *Physica D* 35 (3) (1989) 357–381.
- [146] C.D. Cutler, A theory of correlation dimension for stationary time series, *Phil. Trans. R. Soc. Lond.* 348 (1994) 343–355.
- [147] J. Theiler, Some comments on the correlation dimension of  $1/f^\alpha$  noise, *Phys. Lett. A* 155 (8–9) (1991) 480–493.
- [148] H.E. Hurst, Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.* 116 (1951) 770.
- [149] F. Serinaldi, Use and misuse of some Hurst parameter estimators applied to stationary and non-stationary financial time series, *Physica A* 389 (14) (2010) 2770–2781.
- [150] C.A. Los, *Financial Market Risk: Measurement and Analysis*, Routledge, London, 2003.
- [151] J.E. Hutchinson, Fractals and self-similarity, *J. Ind. Univ. Math.* 30 (1981) 713–747.
- [152] C. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organisation of DNA nucleotides, *Phys. Rev. E* 49 (2) (1994) 1685–1689.
- [153] B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series, *Phys. Rev. Lett.* 100 (8) (2008) 084102.
- [154] S. Arianos, A. Carbone, Cross-correlation of long-range correlated series, *J. Stat. Mech.* (2009) P03037.
- [155] B. Podobnik, I. Grosse, D. Horvatić, S. Ilic, P.Ch. Ivanov, H.E. Stanley, Quantifying cross-correlations using local and global detrending approaches, *Eur. Phys. J. B* 71 (2) (2009) 243–250.
- [156] M.V. Berry, Diffraction, *J. Phys. A* 12 (6) (1979) 781–797.
- [157] S. Havlin, R. Selinger, M. Schwartz, H.E. Stanley, A. Bunde, Random multiplicative processes and transport in structures with correlated spatial disorder, *Phys. Rev. Lett.* 61 (13) (1988) 1438–1441.
- [158] R. Benzi, On the multifractal nature of fully developed turbulence and chaotic systems, *J. Phys. A* 17 (18) (1984) 3521–3531.
- [159] U. Frisch, G. Parisi, *Turbulence and Predictability of Geophysical Flows and Climatic Dynamics*, North-Holland, Amsterdam, 1985, p. 84.
- [160] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, L. Procaccia, B.I. Shraiman, Fractal measures and their singularities: the characterization of strange sets, *Phys. Rev. A* 33 (2) (1986) 1141–1151.
- [161] M.H. Jensen, L.P. Kadanoff, A. Libchaber, I. Procaccia, I. Stavans, On the multifractal nature of fully developed turbulence and chaotic systems, *Phys. Rev. Lett.* 55 (25) (1985) 2798–2801.
- [162] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H.E. Stanley, Multifractal detrended fluctuation analysis of nonstationary time series, *Physica A* 316 (2002) 87–114.
- [163] J.F. Muzy, E. Bacry, A. Arneodo, The multifractal formalism revisited with wavelets, *Int. J. Bifurcation Chaos* 4 (2) (1994) 245–302.
- [164] P.C. Ivanov, L.A.N. Amaral, A.L. Goldberger, S. Havlin, M.G. Rosenblum, Z.R. Struzik, H.E. Stanley, Multifractality in human heartbeat dynamics, *Nature* 399 (6735) (1999) 461–465.
- [165] Z. Chen, P.C. Ivanov, K. Hu, H.E. Stanley, Effect of nonstationarities on detrended fluctuation analysis, *Phys. Rev. E* 65 (4) (2002) 041107.
- [166] B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, R. Mrowka, Uncovering interaction of coupled oscillators from data, *Phys. Rev. E* 76 (5) (2007) 055201.
- [167] B. Kralemann, L. Cimponeriu, M. Rosenblum, A. Pikovsky, R. Mrowka, Phase dynamics of coupled oscillators reconstructed from data, *Phys. Rev. E* 77 (6, Part 2) (2008) 066205.
- [168] S. Boccaletti, J. Kurths, G. Osipov, D.L. Valladares, C.S. Zhou, The synchronization of chaotic systems, *Phys. Rep.* 366 (1–2) (2002) 1–101.
- [169] M.G. Rosenblum, A.S. Pikovsky, J. Kurths, Phase synchronization of chaotic oscillators, *Phys. Rev. Lett.* 76 (11) (1996) 1804–1807.
- [170] A.S. Pikovsky, M.G. Rosenblum, G.V. Osipov, J. Kurths, Phase synchronization of chaotic oscillators by external driving, *Physica D* 104 (3–4) (1997) 219–238.
- [171] P. Tass, M.G. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkman, A. Schnitzler, H.-J. Freund, Detection of  $n : m$  phase locking from noisy data: application to magnetoencephalography, *Phys. Rev. Lett.* 81 (15) (1998) 3291–3294.
- [172] M.G. Rosenblum, L. Cimponeriu, A. Bezerianos, A. Patzak, R. Mrowka, Identification of coupling direction: application to cardiorespiratory interaction, *Phys. Rev. E* 65 (4) (2002) 041909.
- [173] T. Bayes, An essay towards solving a problem in the doctrine of chances, *Phil. Trans.* 53 (1763) 370–418.
- [174] V.N. Smelyanskiy, D.G. Luchinsky, A. Stefanovska, P.V.E. McClintock, Inference of a nonlinear stochastic model of the cardiorespiratory interaction, *Phys. Rev. Lett.* 94 (9) (2005) 098101.
- [175] A. Duggento, D.G. Luchinsky, V.N. Smelyanskiy, I. Khovanov, P.V.E. McClintock, Inferential framework for nonstationary dynamics. II. Application to a model of physiological signaling, *Phys. Rev. E* 77 (6) (2008) 061106.
- [176] A. Duggento, D.G. Luchinsky, V.N. Smelyanskiy, P.V.E. McClintock, Inferential framework for non-stationary dynamics: theory and applications, *J. Stat. Mech.* (2009) P01025.
- [177] A. Duggento, Analysis and applications of a Bayesian inference technique for dynamical systems (Ph.D. thesis), Lancaster University, 2009.
- [178] D.G. Luchinsky, V.N. Smelyanskiy, A. Duggento, P.V.E. McClintock, Inferential framework for nonstationary dynamics. I. Theory, *Phys. Rev. E* 77 (6) (2008) 061105.
- [179] A. Duggento, T. Stankovski, P.V.E. McClintock, A. Stefanovska, Dynamical Bayesian inference of time-evolving interactions: from a pair of coupled oscillators to networks of oscillators, *Phys. Rev. E* 86 (6) (2012) 061126.
- [180] T. Stankovski, A. Duggento, P.V.E. McClintock, A. Stefanovska, A tutorial on time-evolving dynamical Bayesian inference, submitted to *Eur. Phys. J. Spec. Top.*
- [181] M. Le Van Quyen, J. Foucher, J.P. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, F.J. Varela, Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony, *J. Neurosci. Methods* 111 (2) (2001) 83–98.
- [182] J.P. Lachaux, A. Lutz, D. Rudrauf, D. Cosmelli, M. Le Van Quyen, J. Martinerie, F. Varela, Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence, *Clin. Neurophysiol.* 32 (3) (2002) 157–174.
- [183] A. Bandrivskyy, A. Bernjak, P.V.E. McClintock, A. Stefanovska, Wavelet phase coherence analysis: application to skin temperature and blood flow, *Cardiovasc. Eng.* 4 (1) (2004) 89–93.
- [184] L.W. Sheppard, A. Stefanovska, P.V.E. McClintock, Testing for time-localised coherence in bivariate data, *Phys. Rev. E* 85 (2012) 046205.
- [185] K. Hasselmann, W. Munk, G. MacDonald, Bispectra of ocean waves, in: *Time Series Analysis*, Wiley, New York, 1963, pp. 125–139.
- [186] C.L. Nikias, A.P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice-Hall, Englewood Cliffs, 1993.
- [187] C.L. Nikias, M.R. Raghuveer, Bispectrum estimation: a digital signal processing framework, *IEEE Proc.* 75 (7) (1987) 869–891.
- [188] J. Jamšek, A. Stefanovska, P.V.E. McClintock, I.A. Khovanov, Time-phase bispectral analysis, *Phys. Rev. E* 68 (1) (2003) 016201.
- [189] H.C. Kim, E.J. Powers, Digital bispectral analysis and its applications to nonlinear wave interactions, *IEEE Trans. Plasma Sci.* 7 (2) (1979) 120–131.
- [190] J. Chung, E.J. Powers, The statistics of wavelet-based bicoherence, in: *Time-Frequency and Time-Scale Analysis*, IEEE, Pittsburgh, 1998, pp. 141–144.
- [191] S. Elgar, R.T. Guza, Statistics of bicoherence, *IEEE Trans. Acoust. Speech Signal Process.* 36 (10) (1988) 1666–1668.
- [192] R.A. Haubrich, Earth noise, 5 to 500 millicycles per second, *J. Geophys. Res.* 70 (6) (1965) 1415–1427.
- [193] J. Jamšek, A. Stefanovska, P.V.E. McClintock, Wavelet bispectral analysis for the study of interactions among oscillators whose basic frequencies are significantly time variable, *Phys. Rev. E* 76 (2007) 046221.
- [194] M.J. Hinich, G.R. Wilson, Time delay estimation using the cross bispectrum, *IEEE Trans. Signal Process.* 40 (1) (1992) 106–113.
- [195] K.S. Lii, K.N. Helland, Cross-bispectrum computation and variance estimation, *ACM Trans. Math. Software* 7 (3) (1981) 284–294.

- [196] B.P. van Milligen, E. Sánchez, T. Estrada, C. Hidalgo, B. Brañas, B. Carreras, L. García, Wavelet bicoherence—a new turbulence analysis tool, *Phys. Plasmas* 2 (8) (1995) 3017–3032.
- [197] T. Schreiber, Measuring information transfer, *Phys. Rev. Lett.* 85 (2) (2000) 461–464.
- [198] C. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* 37 (1969) 424–438.
- [199] L. Barnett, A.B. Barrett, A.K. Seth, Granger causality and transfer entropy are equivalent for Gaussian variables, *Phys. Rev. Lett.* 103 (23) (2009) 238701.
- [200] M. Paluš, A. Stefanovska, Direction of coupling from phases of interacting oscillators: an information-theoretic approach, *Phys. Rev. E* 67 (2003) 055201(R).
- [201] M. Paluš, From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics, *Contemp. Phys.* 48 (6) (2007) 307–348.
- [202] M. Vejmelka, M. Paluš, Inferring the directionality of coupling with conditional mutual information, *Phys. Rev. E* 77 (2) (2008) 026214.
- [203] D. Marinazzo, M. Pellicoro, M. Stramaglia, Kernel method for nonlinear Granger causality, *Phys. Rev. Lett.* 100 (14) (2008) 144103.
- [204] J.T.C. Schwabedal, A. Pikovsky, B. Kralemann, M. Rosenblum, Optimal phase description of chaotic oscillators, *Phys. Rev. E* 85 (2012) 026216.
- [205] V.B. Efimov, A.N. Ganshin, P.V.E. McClintock, G.V. Kolmakov, L.P. Mezhov-Deglin, Experimental study of the nonlinear second sound wave interaction in superfluid  $^4\text{He}$ , *J. Low Temp. Phys.* 145 (1–4) (2006) 155–164.
- [206] A.N. Ganshin, V.B. Efimov, G.V. Kolmakov, L.P. Mezhov-Deglin, P.V.E. McClintock, Experiments on wave turbulence: the evolution and growth of second sound acoustic turbulence in superfluid  $^4\text{He}$  confirm self-similarity, *New J. Phys.* 12 (2010) 083047.
- [207] G.V. Kolmakov, V.B. Efimov, A.N. Ganshin, P.V.E. McClintock, L.P. Mezhov-Deglin, Formation of a direct Kolmogorov-like cascade of second-sound waves in He II, *Phys. Rev. Lett.* 97 (2006) 155301.
- [208] D. Konstantinov, A. Chepelianskii, K. Kono, Resonant photovoltaic effect in surface state electrons on liquid helium, *J. Phys. Soc. Japan* 81 (2012) 093601.
- [209] A.R. Curtis, High-order explicit Runge–Kutta formulae, their uses and limitations, *J. Inst. Math. Appl.* 16 (1975) 35–55.
- [210] W. Gröbner, H. Knapp, Contributions to the Method of Lie-Series, Bibliographisches Institut, Mannheim, 1967.
- [211] J.O. Smith III, Mathematics of the Discrete Fourier Transform (DFT) With Audio Applications, second ed., W3K Publishing, 2007, <http://books.w3k.org/>.
- [212] R.K. Otnes, L. Enochson, Digital Time Series Analysis, Wiley, New York, 1972.
- [213] G.A. Darbellay, I. Vajda, Estimation of the information by an adaptive partitioning of the observation space, *IEEE Trans. Inform. Theory* 45 (1999) 1315–1321.
- [214] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J. Farmer, Testing for nonlinearity in time series: the method of surrogate data, *Physica D* 58 (1–4) (1992) 77–94.
- [215] M. Paluš, Detecting phase synchronization in noisy systems, *Phys. Lett. A* 235 (1997) 341–351.
- [216] T. Schreiber, A. Schmitz, Surrogate time series, *Physica D* 142 (3–4) (2000) 346–382.
- [217] T. Schreiber, H. Kantz, Predictability of complex dynamical systems, in: *Observing and Predicting Chaotic Signals*, Springer, New York, 2003.