

1

Introduction: a multiscale and turbulent-like world

1.1 Data from the real world

Data analysis is an essential part of scientific research and engineering applications. A proper data analysis method will provide a better understanding of the process under consideration, for example, extracting useful parameters to validate an existing theory or to inspire a new theory. However, data from the real world, such as field-observation data, well-controlled laboratory experiments, or numerical simulation, generally possess several problems. For instance, the data length may be too short to satisfy stationary or ergodic conditions, or the mechanism behind the data may be nonlinear, etc. In Figures 1.1 through 1.3 are several examples from the real world, demonstrating common problems of real datasets.

Figure 1.1a displays a collected oxygen saturation index obtained from a MAREL network (Automatic Monitoring Network for Littoral Environment, Ifremer, France) from the period January 1 to December 31, 2010. The large variation of the measured oxygen saturation index shows the nonstationarity of the data. For example, a high intensity of oxygen saturation index was observed at September 3, 2010. As shown in Section 1.3, to mimic this nonstationary event, high-order Fourier harmonic components are required. Moreover, sometimes the sensor fails to collect data, due to maintenance problems or failure of the system. This missing data problem is typical of field observation data. To emphasize this, we replotted a 10-day portion of the data in Figure 1.1b, which shows a discontinuous curve due to the problem of the data missing. This imposes a difficulty for the Fourier-based data analysis method, for which a uniform time step is often required. To see the irregular time step more clearly, the time interval δt has been shown between two successful measurements in Figure 1.1c. Visually, the time step δt demonstrates a strong intermittent distribution with large values of δt . Another aspect of this data set is that the underlying physical mechanism is nonlinear. It means that if one can write a governing equation for the oxygen

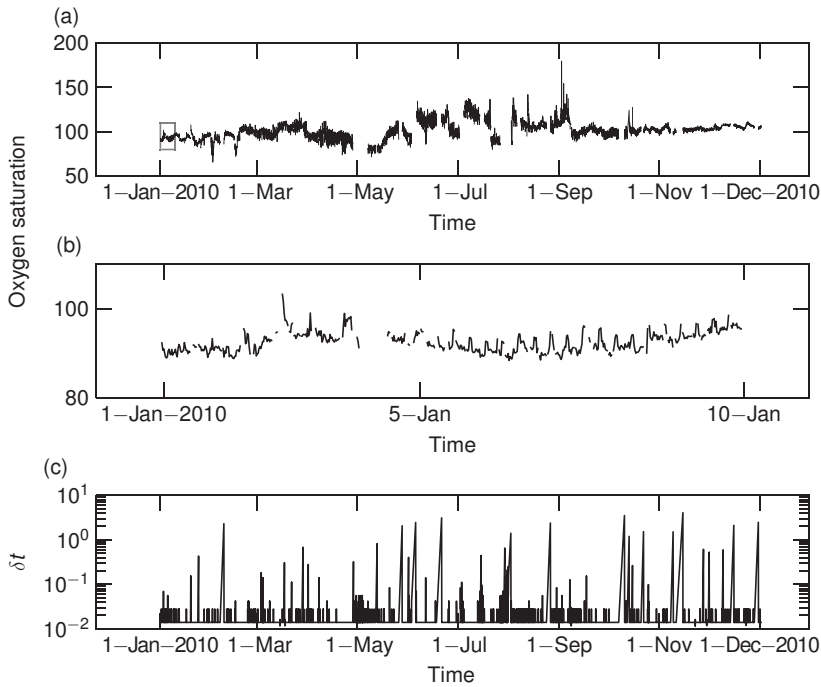


Figure 1.1 a) Collected oxygen saturation index on the time period January 1 to December 31, 2010, with a sampling time of 20 minutes. The data were collected by a sensor belonging to the MAREL network (Automatic monitoring network for littoral environment, Ifremer, France). Due to several reasons, e.g., failure of the sensor, there are several missing data points. A strong event was also observed around September 1, 2010, showing the nonstationarity of the process. b) An enlargement part on the time period January 1 to January 10, 2010. c) The time interval δt between two successful measurements. The intermittent distribution of δt demonstrates the problem of missing data or irregular time step.

saturation index, this equation must be nonlinear. Therefore, the difficulties of this type of data set are missing data/irregular time steps, both nonstationarity and nonlinearity.

Turning to another aspect of the real world data, namely varying sample size, considered here is an example from the global drifter program. Figure 1.2a illustrates several Lagrangian trajectories obtained from the global drifter program. Due to failures of the devices, battery life-time, etc., the life-time T of drifters varies from one to another. They provide measured physical quantities with different lengths. For example, the drifter collects data every six hours, corresponding to four data points per day. Therefore, a finite life-time T means a finite data sample, for example, $L = 4T$. Figure 1.2b shows the measured distribution of T . It has an exponential distribution, indicating a large variation of the data length.

1.1 Data from the real world

3

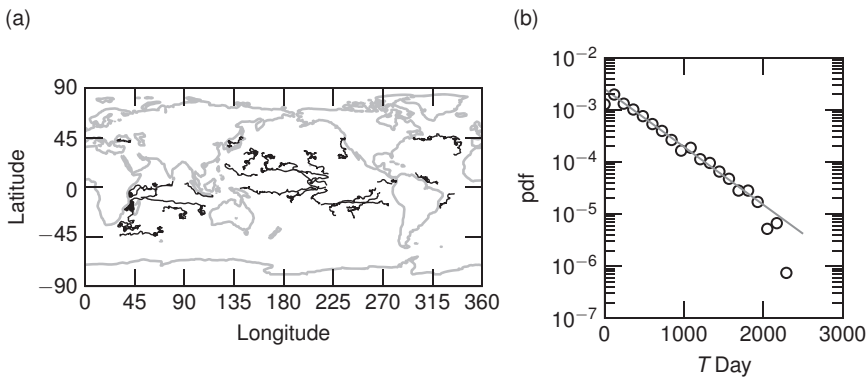


Figure 1.2 a) Illustration of the Lagrangian drifter trajectories obtained from the global drifter program. b) Measured pdf of the drifter life-time T . The measured pdf $p(T)$ has an exponential distribution. The drifter collected several physical quantities of the ocean with a sampling time of six hours. This provided a large variation of the data length.

More precisely, the mean and standard deviation of T are $T \simeq 370$ and $T \simeq 360$ days. This will impose difficulties when performing traditional spectrum analysis, since each drifter covers different time scales. This happens also in well-controlled laboratory experiments. For example, in Lagrangian turbulence experiments, the tracer particles are tracked experimentally in a turbulent flow with thousands of realizations (Toschi and Bodenschatz, 2009). Due to the finite measurement volume and the limitation of the particle tracking technique, the tracking period for each individual particle is different. We therefore have the same difficulty as the one for Lagrangian drifters.

The presence of a forcing scale, or strong deterministic forcing, is another important feature of the data sets from natural sciences, especially geophysics. For example, in geoscience, daily cycle and annual cycle are present in several collected data, and in marine data sets, the tidal cycle is also present for many series. This is typically the case for sea surface temperature, air temperature, river daily discharge, etc. Figure 1.3 displays a collected water temperature $\theta(t)$ °C from the MAREL network mentioned earlier in this chapter. The collected temperature possesses a strong annual cycle, a deterministic forcing from the solar-earth system. As shown in Chapter 4, the scaling behavior (see discussion in next subsection and more detail in Chapter 3), an important feature of complex systems, will be perturbed by such deterministic forcing. There exists a continuous range of excited time scales, at least between the annual cycle and the daily cycle, namely multiscale fluctuations. One of the objectives of this analysis is to try to better understand the statistical relationship between scales. Whether or not this relationship can be experimentally extracted from the data is one of the essential contents of this book. Some general

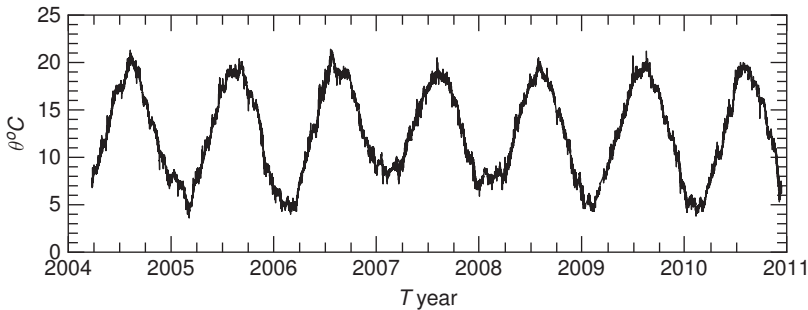


Figure 1.3 Collected water temperature data on the period March 24 to December 13, 2010. These data are automatically recorded every 20 minutes by the MAREL network. A strong annual cycle is observed. Note that this deterministic-like forcing is a typical structure in the geophysical data, which might induce difficulties in the scaling analysis.

comments on this issue are proposed in the next section, and more details are provided on this topic in the rest of this book.

Summarized in this list are the main properties of the data collected from the real world. They are:

1. finite sample size;
2. nonstationarity and nonlinearity;
3. presence of stochastic fluctuations, as well as deterministic forcing;
4. multiscale, or multiscaling statistical properties;
5. irregular time step, or missing data.

An ideal time-series analysis method should be able to handle all difficulties imposed by these the above listed properties to reveal the physics represented by the data.

1.2 Multiscale phenomena

As mentioned in Section 1.1 multiscale fluctuations are relevant in many complex systems in which many time or spatial degrees of freedom are present and may interact with each other. The most classical complex system is the case of turbulent flows (Frisch, 1995; Tsinober, 2009). In three-dimensional turbulent flows, a large range of time and spatial scales are involved and interact with each other, at least in the so-called inertial range. The famous Richardson-Kolmogorov energy cascade picture has been proposed to interpret the turbulent flow in a phenomenological way: the energy is transferred from large- to small-scale structures, until the viscosity scale, where the energy is converted into heat (see Kolmogorov theory in Chapter 2 or Frisch, 1995). Figure 1.4a displays a one second portion of the longitudinal

1.2 Multiscale phenomena

5

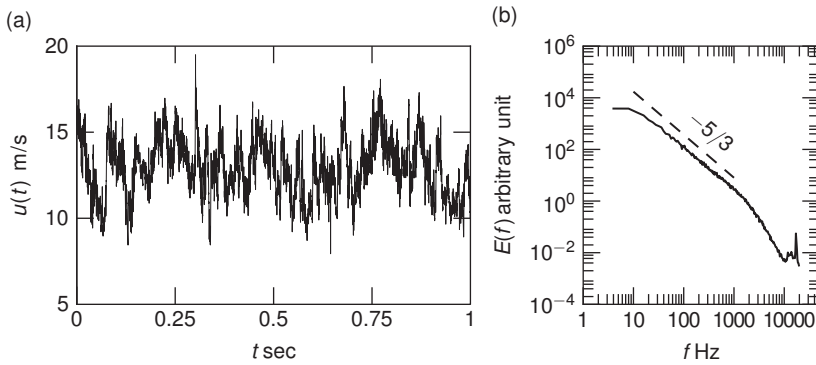


Figure 1.4 a) A one second portion of the measured longitudinal turbulent velocity in a wind tunnel experiment at Johns Hopkins University with a Taylor microscale based Reynolds number $Re_\lambda \simeq 720$. The measured velocity is fluctuating over a large range of time scales. b) The experimental Fourier power spectrum in a log-log plot. Power-law behavior $E(f) \sim f^{-\beta}$ is observed with a β close to the Kolmogorov 1941 theory prediction $5/3$.

velocity obtained from a wind tunnel experimental with a Taylor’s microscale based Reynolds number $Re_\lambda \simeq 720$. This experimental Eulerian velocity fluctuates over different time scales. High intensity events, known as nonstationary events, were observed. The corresponding Fourier power spectrum $E(f)$ is shown in Figure 1.4b in a log-log plot, in which a $5/3$ power-law relation from the Kolmogorov 1941 theory is illustrated by a dashed line. A power-law behavior is observed on the range $10 < f < 1000$ Hz. It corresponds with the Kolmogorov’s 1941 phenomenological theory (see more discussion in Chapter 2).

Another example is the Lagrangian velocity obtained from a high-resolution numerical simulation. Figure 1.5a shows the Lagrangian velocity along a Lagrangian trajectory and 1.5b the corresponding energy dissipation rate $\epsilon(t)/\langle\epsilon\rangle$. This trajectory was chosen on purpose to display a “vortex trapping” event around $t/\tau_\eta = 100$. The corresponding energy dissipation is fluctuating over a large range of amplitude. For the present trajectory, it is as high as 23 times the mean energy dissipation rate, showing a strong intermittent event. In fact, the highest energy dissipation rate could reach more than 100 times of the mean energy dissipation rate. Note that a continuous range of frequencies/scales is observed. The Richardson-Kolmogorov cascade picture is essentially a multiscale description of turbulent flows (Tsinober, 2009).

Hydrodynamic turbulent flows are governed by the Navier-Stokes equation (see Chapter 2). Unfortunately, a mathematical solution is unreachable for the moment. Therefore, a phenomenological theory, such as the Kolmogorov 1941 theory, must be verified experimentally or numerically. In the content of multiscale analysis, the

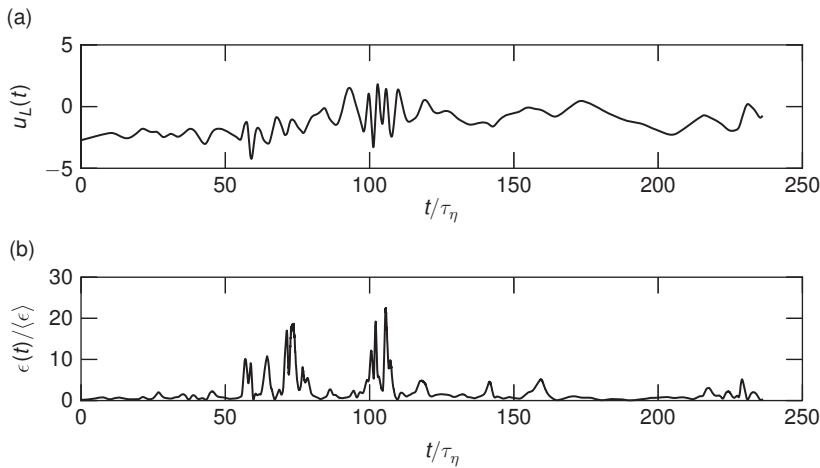


Figure 1.5 a) An example of turbulent velocity along a Lagrangian trajectory obtained from a high-resolution direct numerical simulation with a Taylor microscale-based Reynolds number $Re_\lambda \simeq 400$. A vortex trapping event is observed around $t/\tau_\eta = 100$. b) The measured energy dissipation rate $\epsilon(t)/\langle\epsilon\rangle$ along the same Lagrangian trajectory, showing a strong intermittent/nonstationary event around time $50 < t/\tau_\eta < 120$.

essential job of the data analysis is thus to reveal the statistical relations between different scales to verify theory predictions or to provide “seeds” for new theories.

There also exists numerous turbulent-like complex systems, such as financial activities (Ghashghaie et al., 1996; Mantegna and Stanley, 1996; Schmitt, Schertzer, and Lovejoy, 1999; Li and Huang, 2014); environmental variables in the sea (Schmitt et al., 2009; Zongo and Schmitt, 2011; Huang, Schmitt, and Gagne, 2014); daily river discharge (Tessier et al., 1996; Dahlstedt and Jensen, 2005; Huang et al., 2009a), etc., to name but a few. Unfortunately, an exactly governing equation for such a complex system cannot be written down. For now, various time series/data from field observations or well-controlled laboratory experiments have been obtained. A proper multiscale treatment of the data will provide a better understanding of the dynamics of such complex systems in order to extract the scale-dependent information/parameters. New theoretical considerations might be inspired by the data.

1.3 The Fourier-based methodology and its potential shortcomings

1.3.1 Linear asymptotic approximation

There are many time-frequency analysis methods (Cohen, 1995; Flandrin, 1998). Their basic idea originated in part from the Fourier analysis. It can be interpreted

1.3 The Fourier-based methodology and its potential shortcomings 7

as representing a given signal/function $x(t)$, by a given basis φ , i.e.,

$$x(t) = \int_{v \in R} \int_{t' \in R} \psi(t', v) \varphi(t, t', v) dv dt' \quad (1.1)$$

where ψ is a coefficient (function) that can be determined as:

$$\psi(t, v) = \int_{t' \in R} x(t) \varphi(t, t', v) dt' \quad (1.2)$$

Here, the basis function φ also can be interpreted as an integral kernel of Equation 1.2 (Cohen, 1995). It is an asymptotic approximation: the signal is asymptotically approximated by the chosen basis (function) φ . Usually, the property of the chosen basis is well known. Then the given signal is checked to see what it looks like with respect to the chosen basis (function) φ . For example, when the trigonometric function is chosen, the classical Fourier transform is obtained, for which ψ depends only on the frequency:

$$\psi(f) = \int_{-\infty}^{+\infty} f(x) e^{i2\pi fx} dx \quad (1.3)$$

in which $\psi(f)$ is the Fourier coefficient. $\psi(f)$ is independent with x , corresponding to a global property of x . Therefore, the Fourier analysis cannot identify a nonstationary event (Cohen, 1995; Flandrin, 1998; Huang et al., 1998; Huang, 2009).

Another example is the Wavelet transform, where ψ depends on time t and the scale a :

$$\psi(a, t) = \frac{1}{\sqrt{a}} \int_{t' \in R^n} x(t') \varphi\left(\frac{t' - t}{a}\right) dt' \quad (1.4)$$

where n is the dimension of the space, $\varphi(t)$ is the so-called mother wavelet and a is a dilatation parameter. To be a mother wavelet, $\varphi(t)$ should satisfy some conditions. (For details on wavelet theory, see Daubechies (1992); Meyer (1995); Mallat (1999)). The wavelet transform approach may also be considered as an adaptive-window Fourier transform (Huang et al., 1998). Note that the wavelet transform is local on different scales. Therefore, it is efficient for detecting and analyzing nonstationary events. However, as shown in Chapter 4, the wavelet-based method is still affected by the high-order harmonic problem. Indeed, this problem cannot be overcome if the basis is chosen *a priori*.

In general, the traditional approach for time-frequency analysis is to choose basis functions *a priori*. Once the basis (function) is fixed, the information that can be extracted from the data is determined. They are also energy-based approaches: only when the event contains enough energy can it then be detected by such methods (Huang et al., 1998; Huang and Wu, 2005).

1.3.2 High-order Fourier harmonic: mathematical or physical?

To show experimentally the problem of high-order harmonic components, here the Fourier analysis of the nonlinear Duffing equation is considered. It is shown that the high-order harmonic component is a requirement of the mathematical approach used here. They are not observed in the physical domain.

The nonlinear Duffing equation is written as

$$\frac{d^2x}{dt^2} + x(1 + \epsilon x^2) = b \cos(2\pi \Omega t) \quad (1.5)$$

in which ϵ is a nonlinear parameter. It can be considered as a pendulum with forcing function $b \cos(\Omega t)$, in which its pendulum length varies with the angle, showing a nonlinear mechanism in this system through the parameter ϵ . The parameters chosen in Equation (1.5) are $\epsilon = 1$, $b = 0.1$ and $\Omega = 0.04$. Note that $\epsilon = 1$ is finite, and thus Equation (1.5) presents a strong nonlinearity without analytical solution. A fifth-order Runge-Kutta algorithm is performed to solve Equation (1.5) with an initial condition $[x(0), x'(0)] = [1, 1]$ and a sampling frequency of 10 Hz. Figure 1.6a shows the numerical solution. Some comments on the numerical solution are provided in this section. Firstly, visually the wave profile of the numerical solution $x(t)$ is far from a sine or cosine wave; this is a nonlinear distortion (Huang et al., 1998, 2011a). This deviation is a result of the nonlinear mechanism. Secondly, the solution $x(t)$ is smooth with a mean period $T \simeq 9.524$, provided by the peak counting method. It corresponds to a dominant frequency $f_D \simeq 0.105$ Hz. Therefore, a

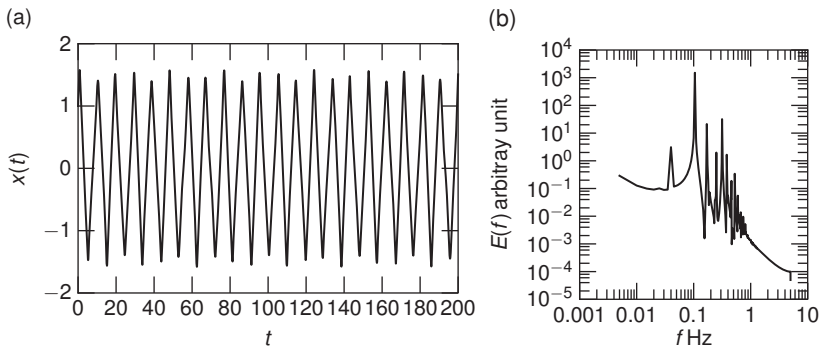


Figure 1.6 a) Numerical solution of the Duffing equation with $\epsilon = 1$, $b = 0.1$, $\Omega = 2\pi/25$ and initial value $[x(0), x'(0)] = 1$. Note that the solution $x(t)$ is smooth, with a mean period $T = 9.524$, corresponding to a domain frequency 0.105 Hz. Hence a frequency $f \gg 0.105$ is unphysical. b) The corresponding Fourier power spectrum. The forcing scale is observed at 0.04 Hz as the first peak in $E(f)$. The second peak is at 0.105 Hz for the domain frequency. High-order Fourier harmonic components are visible for $f \gg 0.105$ Hz. They are required by the linear approximation of the Fourier transform.

1.3 The Fourier-based methodology and its potential shortcomings

9

frequency much higher than this value is artificially found by the analysis method: this is a “high-order harmonic problem” (Cohen, 1995; Flandrin, 1998).

Figure 1.6b shows the measured Fourier power spectrum in a log-log plot. A strong peak is observed at $f \simeq 0.1$, which agrees well with the dominant frequency f_D in the previous paragraph. Also noted were the existence of several peaks. The first peak at $f \simeq 0.04$ Hz, corresponds to the external forcing Ω . The other peaks are larger than f_D and thus considered as high-order harmonic components. They are required by the Fourier analysis. This phenomenon is also observed for the Wavelet transform; for more details, see Huang et al. (2011a).

To show the high-order harmonic more clearly, $x(t)$ was reconstructed partially from several Fourier coefficients, i.e., $\tilde{x}(t) = \int_{f \in F} \mathcal{F}(f) \exp(-j2\pi ft) df$ for a given range of $f \in F$. Figure 1.7 shows (a) the reconstruction $\tilde{x}(t)$ from the domain

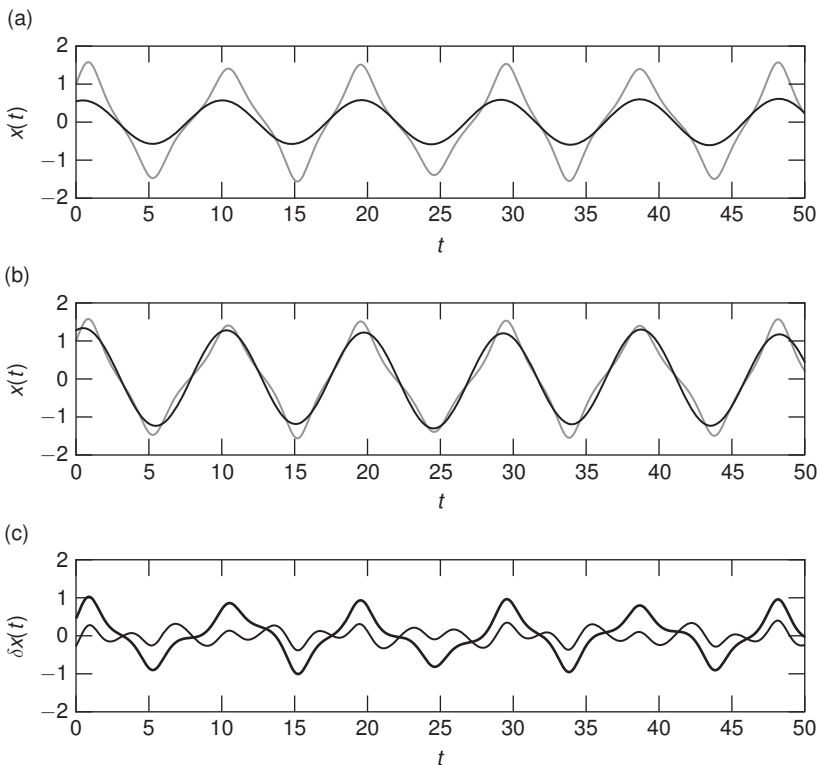


Figure 1.7 a) Reconstructed $\tilde{x}(t)$ from the domain frequency $f_D = 0.105$ Hz. b) Reconstructed $\tilde{x}(t)$ from the first 30 Fourier modes. c) Residue $\delta x(t) = x(t) - \tilde{x}(t)$, in which the thick solid line is for the domain frequency and the thin line is for the first 30 Fourier modes. Note that the high-order Fourier harmonic components are the result of the difference between the Duffing solution $x(t)$ and the sine wave profile.

frequency f_D , and (b) the reconstruction from the first 30 Fourier modes (resp. up to $f \simeq 0.15$ Hz, the first valley in between $0.1 < f < 0.2$ Hz). As mentioned here, the idea of the Fourier analysis is to approximate a given signal by using a series of sine or cosine waves. Therefore, with an increasing number of Fourier modes, the reconstructed signal becomes closer and closer to the original one. Figure 1.7c displays the residue $\delta x(t) = x(t) - \tilde{x}(t)$, in which the thick solid line is for the dominant Fourier mode and the thin solid line is for the first 30 Fourier modes. This clearly shows that due to the so-called nonlinear distortion, the high frequency oscillation (resp. the so-called high-order harmonics) is emerging in $\delta x(t)$. However, this high-frequency oscillation is not observed in the original signal $x(t)$. The high-order harmonic component is thus a requirement of the Fourier analysis, but is not a physical requirement. It will be shown in Chapter 4 that by introducing the idea of intrawave frequency modulation, the high-order harmonic component is then constrained by the use of a Hilbert-based method.

1.4 Conclusion and further remarks

In conclusion of this introductory chapter, for several reasons, the data obtained from the real world, natural sciences, and geosciences in general, are typically nonstationary and nonlinear. The data also display multiscale properties with fluctuations over a large range of scales. An ideal data processing method should handle all aspects mentioned in the previous section: nonstationary, nonlinear, multiscale, irregular time step/missing data, etc. In recent years, several methodologies have been proposed to handle such real data. These include the Wavelet transform, Detrended Fluctuation Analysis, Empirical Mode Decomposition, and the associated Hilbert-Huang Transform. They have been successful in some aspects, but in others, they have had difficulties. For example, the Wavelet transform is efficient for nonstationary processes, but it might be affected by nonlinear processes (Huang et al., 1998, 1999).

Therefore, before a method is applied to data from the real world, it is better to know whether this method is suitable for the given type of data. For example, a strong annual cycle is often observed in oceanic time series (see Figure 1.3). Due to the presence of the annual cycle, the structure function analysis is strongly influenced when the multiscaling/multifractal property of such data set is concerned. The result is then biased (Huang et al., 2011a); see also discussion in Chapter 4.

In this book, the theory of homogeneous and isotropic turbulence in Chapter 2 is first recalled as the most classical multiscale complex system. Here the ideas of energy cascade, intermittency, and multifractal are briefly described, along with discussions on several intermittency models. Chapter 3 presents several scaling stochastic processes, including self-similar processes and nonstationary