Chapter 4
Stochastic Models of Evolution

To continue the discussion of randomness given in Sect. 2.2.1, we briefly touch on *stochastic models* of temporal evolution (random processes). They can be specified either via explicit definition of their statistical properties (probability density functions, correlation functions, etc., Sects. 4.1, 4.2 and 4.3) or via stochastic difference or differential equations. Some of the most widely known equations, their properties and applications are discussed in Sects. 4.4 and 4.5.

4.1 Elements of the Theory of Random Processes

If, given initial conditions $\mathbf{x}(t_0)$ and fixed parameter values, a process demonstrates the same time realisation in each trial, then its natural description is deterministic (Chap. 3). However, such a situation is often not met in practice: different trials "under the same conditions" give different realisations of a process. One relates such non-uniqueness to influences from multiple uncontrolled factors, which are often present in the real world. Then, it is reasonable to refuse deterministic description and exploit an apparatus of the theory of probability and theory of random processes (see, e.g. Gihman and Skorohod, 1974; Kendall and Stuart, 1979; Malakhov, 1968; Rytov et al., 1978; Stratonovich, 1967; Volkov et al., 2000; Wentzel', 1975).

4.1.1 Concept of Random Process

Random process (random function of time) is a generalisation of the concept of random quantity to describe time-dependent variables. More precisely, its definition is given as follows. Firstly, *random function* is a random quantity depending not only on a random event ω but also on some parameter. If that parameter is time, then the random function is called random process and denoted $\xi(t, \omega)$. A quantity ξ may be both scalar (a scalar random process) and vector (a vector or a multidimensional random process). It may run either a discrete range of values (a process with discrete

states) or a continuous one. For the sake of definiteness, we further speak of the latter case. Studying and development of such models is the subject of the theory of random processes (Gihman and Skorohod, 1974; Stratonovich, 1967; Volkov et al., 2000; Wentzel', 1975). In the case of discrete time $t = 0, 1, 2, \ldots$, a random process is called a *random sequence*.

For a random process, an outcome of a single trial is not a single number (like for a random quantity) but a function $\xi(t, \omega_1)$, where ω_1 is a random event realised in a given trial. The random event can be interpreted as a collection of uncontrolled factors influencing a process during a trial. The function $\xi(t, \omega_1)$ is called a *realisation of a random process*. It is a deterministic (non-random) function of time, because a random event $\omega = \omega_1$ is fixed. In general, one gets different realisations as outcomes of different trials. A set of realisations obtained from various trials (i.e. for different ω) is called an *ensemble* of realisations (Fig. 4.1).
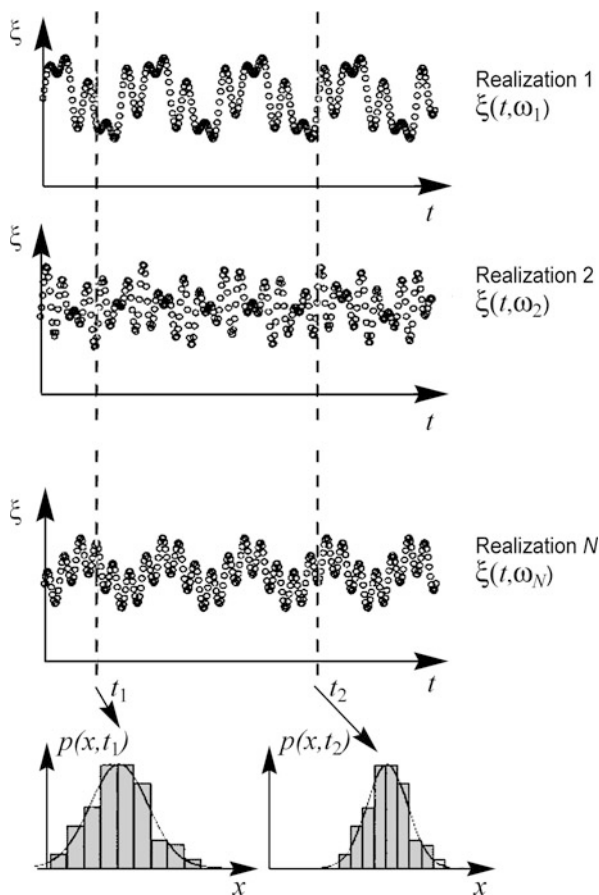

Fig. 4.1 An ensemble of N realisations (three of them are shown) and two sections of a random process

4.1.2 Characteristics of Random Process

At any fixed time instant t, a random process $\xi(t, \omega)$ is a random quantity. The latter is called a *section of a random process* at a time instant t and characterised by a probability density function $p(x, t)$. This distribution law is called *one-dimensional distribution* of the random process. It depends on time and may differ for two different time instants. Knowledge of one-dimensional distribution law $p(x, t)$ allows one to calculate expectation and variance of the process at any time instant t. If the distribution law varies in time, then the expectation

$$m(t) = E\left[\xi(t, \omega)\right] = \int_{-\infty}^{\infty} x p(x, t) \mathrm{d}x \qquad (4.1)$$

and the variance

$$\sigma_\xi^2(t) = E[\xi(t, \omega) - m(t)]^2 = \int_{-\infty}^{\infty} [x - m(t)]^2 p(x, t) \mathrm{d}x \qquad (4.2)$$

may vary in time as well. They are deterministic (non-random) functions of time, since dependence on random events is eliminated due to integration.

In general, sections $\xi(t, \omega)$ at different time instants t_1 and t_2 exhibit different probability density functions $p(x, t_1)$ and $p(x, t_2)$, Fig. 4.1. Joint behaviour of the sections is described by two-dimensional probability density function $p_2(x_1, t_1, x_2, t_2)$. One can define n-dimensional distribution laws p_n for any sets $t_1, t_2, \ldots, t_n$ in the same way. These laws constitute a collection of finite-dimensional distributions of a random process $\xi(t, \omega)$. Probabilistic properties of a process are fully defined only if the entire collection is given. However, since the latter represents an infinite number of distribution laws, one cannot in general fully describe a random process.

To be realistic, one must confine him/herself with the use of some characteristics, e.g. one- and two-dimensional distributions or low-order moments (expectation, variance, auto-covariance function). Thus, *auto-covariance function* depends on two arguments:

$$\begin{aligned} K(t_1, t_2) &= E\left[(\xi(t_1, \omega) - m(t_1))\,(\xi(t_2, \omega) - m(t_2))\right] = \\ &= \iint (x_1 - m(t_1))\,(x_2 - m(t_2))\,p_2(x_1, t_1, x_2, t_2)\mathrm{d}x_1\,\mathrm{d}x_2. \end{aligned} \qquad (4.3)$$

For fixed t_1 and t_2, the expression (4.3) defines *covariance* of the random quantities $\xi(t_1, \omega)$ and $\xi(t_2, \omega)$. If it is normalised by root-mean-squared deviations, one gets *autocorrelation function* $\rho(t_1, t_2) = K(t_1, t_2)/(\sigma_\xi(t_1)\sigma_\xi(t_2))$, i.e. correlation

coefficient between random quantities $\xi(t_1, \omega)$ and $\xi(t_2, \omega)$.[1] Autocorrelation function takes values ranging from -1 to 1. The value of $|\rho(t_1, t_2)| = 1$ corresponds to a deterministic linear dependence $\xi(t_1, \omega) = \text{const} \cdot \xi(t_2, \omega)$.

To characterise a process, one often uses conditional one-dimensional distribution $p_1(x, t | x_1, t_1)$, i.e. distribution of a section $\xi(t)$ under the condition that at a time instant t_1 the quantity ξ takes a value of $\xi(t_1) = x_1$. The function $p_1(x, t | x_1, t_1)$ is called probability density of the transition from a state x_1 at a time instant t_1 to a state x at a time instant t.

4.1.3 Stationarity *and* Ergodicity *of Random Processes*

An important property of a process is its stationarity or non-stationarity. A process is called *strongly stationary (stationary in a narrow sense)* if all its finite-dimensional distributions do not change under a time shift, i.e. $p_n(x_1, t_1, \ldots, x_n, t_n) = p_n(x_1, t_1+\tau, \ldots, x_n, t_n+\tau), \ \forall n, t_1, \ldots, t_n, \tau$. In other words, neither characteristic of a process changes under a time shift. A process is called *weakly stationary (stationary in a wide sense)* if its expectation, variance and autocorrelation function (i.e. moments up to the second order inclusively) do not change under a time shift. For a stationary process (in any of the two senses), one has $m(t) = \text{const}$, $\sigma_\xi^2(t) = \text{const}$ and $K(t_1, t_2) = k(\tau)$ with $\tau = t_2 - t_1$. The strong stationarity implies the weak one.[2]

In general, "stationarity" means invariance of some property under a time shift. If a property of interest (e.g. nth-order moment of the one-dimensional distribution) does not change in time, then a process is called stationary with regard to that property.

A process is called *ergodic* if all its characteristics can be determined from its single (infinitely long) realisation. For instance, the expectation is then determined as

$$ m = \lim_{T \to \infty} \frac{1}{T} \int_0^T \xi(t, \omega_1) dt $$

for almost any ω_1, i.e. temporal averaging and ensemble (state space) averaging give the same result. If one can get *all* characteristics of a process in such a way, the process is called *strictly ergodic*. If only certain characteristics can be restored from a single realisation, the process is called ergodic with regard to those characteristics. Thus, one introduces the concept of the first-order ergodicity, i.e. ergodicity with regard to the first-order moments, and so forth. Ergodic processes constitute an

---

[1] There is also another terminology where equation (4.3) is called an auto-correlation function, and after the normalisation, a normalised auto-correlation function. To avoid misunderstanding, we do not use it in this book.

[2] There is also somewhat different interpretation related to an additional requirement of a *finite* variance for a weakly stationary process. In such a case, the weak stationarity does not follow from the strong one.

important class, since in practice one often has a single realisation rather than a big ensemble of realisations. Only for an ergodic process one can restore its properties from such a data set. Therefore, one often assumes ergodicity of a process under investigation when time series analysis (Chaps. 5, 6, 7, 8, 9, 10, 11, 12 and 13) is performed. An ergodic process is stationary, while the reverse is not compulsorily true.

An example (a real-world analogue) of a random process is provided by almost any physical measurement. It may be measurements of a current in a non-linear circuit, exhibiting self-sustained oscillations. Measured time realisations differ for different trials due to thermal noise, various interference, etc. Moreover, having a realisation over a certain time interval, one cannot uniquely and precisely predict its future behaviour, since the latter is determined by random factors which will affect a process in the future.

A simpler example of a random process is a strongly simplified model representation of photon emission by an excited atom. An emission instant, initial phase, direction and polarisation are unpredictable. However, as soon as a photon is emitted and its starting behaviour gets known, the entire future is uniquely predictable. According to the representation considered, the process is described as sinusoidal function of time with a random initial phase (harmonic noise). Random processes of such type are sometimes called *quasi-deterministic*, since random factors determine only initial conditions, while further behaviour obeys a deterministic law.

4.1.4 Statistical Estimates of Random Process Characteristics

To get statistical estimates of a one-dimensional distribution law $p(x, t)$ and its moments, one can perform many trials and obtain a set of realisations $\xi(t, \omega_1)$, $\xi(t, \omega_2)$, ..., $\xi(t, \omega_n)$. The values of the realisations at a given time instant $t = t^*$ constitute a sample of size n from the distribution of a random quantity $\xi(t^*, \omega)$, Fig. 4.1. One can estimate a distribution law $p(x, t^*)$ and other characteristics based on that sample. It can be done for each time instant.

Multidimensional distribution laws can be estimated from an ensemble of realisations in an analogous way. However, the number of realisations for their reliable estimation must be much greater than that for the estimation of statistical moments or one-dimensional distributions.

A situation when one has only a single realisation is more complex. Only for an ergodic process and a sufficiently long realisation, one can estimate characteristics of interest by replacing ensemble averaging with temporal averaging (Sect. 4.1.3).

4.2 Basic Models of Random Processes

A random process can be specified via explicit description of its finite-dimensional probability distributions. In such a way, one introduces basic models in the theory of random processes. Below, we consider several of them (Gihman and Skorohod, 1974; Volkov et al., 2000; Wentzel', 1975).

(i) One of the most important models in the theory of random processes is the *normal (Gaussian) random process*. This is a process whose finite-dimensional distribution laws are all normal. Namely, an n-dimensional distribution law of this process reads as

$$p_n(x_1, t_1, \ldots, x_n, t_n) = \frac{1}{\sqrt{(2\pi)^n \, |\mathbf{V}_n|}} \exp\left(-\frac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}} \cdot \mathbf{V}_n^{-1} \cdot (\mathbf{x}_n - \mathbf{m}_n)\right)$$
(4.4)

for any n, where

$$\mathbf{x}_n = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}, \mathbf{m}_n = \begin{bmatrix} m(t_1) \\ m(t_2) \\ \cdots \\ m(t_n) \end{bmatrix}, \mathbf{V}_n = \begin{bmatrix} K(t_1, t_1) & K(t_1, t_2) & \ldots & K(t_1, t_n) \\ K(t_2, t_1) & K(t_2, t_2) & \ldots & K(t_2, t_n) \\ \cdots & \cdots & \cdots & \cdots \\ K(t_n, t_1) & K(t_n, t_2) & \ldots & K(t_n, t_n) \end{bmatrix},$$
(4.5)

$m(t)$ is the expectation, $K(t_1, t_2)$ is the auto-covariance function (4.3), T stands for transposition and $|\mathbf{V}_n|$ is a determinant of a matrix $\mathbf{V}_n$. Here, all the finite-dimensional distributions are known (a process is fully determined) if the expectation and the auto-covariance function are specified. A normal process remains normal under any linear transform.

(ii) A process *with independent increments*. This is a process for which the quantities $\xi(t_1, \omega)$, $\xi(t_2, \omega) - \xi(t_1, \omega)$, $\ldots$, $\xi(t_n, \omega) - \xi(t_{n-1}, \omega)$ (increments) are statistically independent for any n, $t_1, \ldots, t_n$, such that $n > 1$ and $t_1 < t_2 < \ldots < t_n$.

(iii) *Wiener's process*. This is an N-dimensional random process with independent increments for which a random vector $\xi(t_2, \omega) - \xi(t_1, \omega)$ for any $t_1 < t_2$ is distributed according to the normal law with zero mean and the covariance matrix $(t_2 - t_1)s^2 I_n$, where I_n is the nth-order unit matrix and $s = \mathrm{const}$. This is a non-stationary process. In the one-dimensional case, its variance linearly rises with time as $\sigma^2(t) = \sigma^2(t_0) + s^2 \cdot (t - t_0)$.

Wiener's process describes, for instance, a Brownian motion, i.e. movements of a Brownian particle under random independent shocks from molecules of a surrounding medium.

One can show that Wiener's process is a particular case of the normal process. Wiener's process with $s = 1$ is called standard.

(iv) A (first-order) *Markovian process* is a random process whose conditional probability density function for any $n, t_1, \ldots, t_n$, such that $t_1 < t_2 < \ldots < t_n$, satisfies the property $p_1(x_n, t_n | x_{n-1}, t_{n-1}, \ldots, x_1, t_1) = p_1(x_n, t_n | x_{n-1}, t_{n-1})$. This is also expressed as "the future depends on the past only via the present". Any finite-dimensional distribution law of this process is expressed via its one-dimensional and two-dimensional laws. One can show that Wiener's process is a particular case of a Markovian process.

One more important particular case is a Markovian process with a finite number K of possible states $S_1, \ldots, S_K$. Due to discreteness of the states, it is described in terms of probabilities rather than probability densities. Conditional probability $P\{\xi(t+\Delta t) = S_j | \xi(t) = S_i\}$ is called transition probability, since it describes the transition from the state i to the state j. A quantity $\lambda_{i,j}(t) = \lim_{\Delta t \to +0} P\{\xi(t + \Delta t) = S_j | \xi(t) = S_i\}/\Delta t$ is called transition probability density.

Markovian processes play an especial role in the theory of random processes. Multitude of investigations are devoted to them.

(v) *Poisson process* with a parameter $\lambda > 0$ is a scalar random process with discrete states possessing the following properties: (a) $\xi(0, \omega) = 0$; (b) increments of the process are independent; (c) for any $0 \leq t_1 < t_2$, a quantity $\xi(t_2, \omega) - \xi(t_1, \omega)$ is distributed according to the Poisson law with a parameter $\lambda(t_2 - t_1)$, i.e.

$$P\{\xi(t_2, \omega) - \xi(t_1, \omega) = k\} = \frac{\lambda(t_2 - t_1)^k}{k!} \exp(-\lambda(t_2 - t_1)),$$

where k is a non-negative integer. Poisson process is often used in applications, e.g. in the queueing theory.

(vi) *White noise* is a weakly stationary (according to one of the definitions of stationarity, see Sect. 4.1.3) random process whose values at different time instants are uncorrelated, i.e. its auto-covariance function is $k(\tau) = \text{const} \cdot \delta(\tau)$. It is called "white", since its power spectrum is a constant, i.e. all frequencies are equally presented in it. Here, one draws an analogy to the white light, which involves all frequencies (colours) of the visible part of spectrum. White noise variance is infinitely large: $\sigma_\xi^2 = k(0) = \infty$.

A widespread model is Gaussian white noise. This is a stationary process with a Gaussian one-dimensional distribution law and auto-covariance function $k(\tau) = \text{const} \cdot \delta(\tau)$. Strictly speaking, such a combination is contradictory, since white noise has infinite variance, while a normal random process has a finite variance. Yet, somewhat contradictory concept of Gaussian white noise is useful in practice and in investigations of stochastic differential equations (Sect. 4.5). Gaussian white noise can be interpreted as a process with a very large variance, while a time interval over which its auto-covariance function decreases down to zero is very small as compared with the other characteristic timescales of a problem under consideration.

(vii) A discrete-time analogue of white noise is a sequence of *independent identically distributed* random quantities. It is also often called white noise. Most often, one considers normal one-dimensional distribution, even though any other distribution is also possible. In case of discrete time, variance σ_ξ^2 is finite so that the process is weakly stationary, no matter what definition of the weak stationarity is used.

White noise is the "most unpredictable" process, since any interdependence between its successive values is absent. A sequence of independent

normally distributed quantities serves as a basic model in the construction of discrete-time stochastic models in the form of stochastic difference equations (Sect. 4.4).

(viii) *Markov chain* is a Markovian process with discrete states and discrete time. This simple model is widely used in practice. Its main characteristics are probabilities of transitions from one state to another one. Graph-theoretic tools are used for the analysis and representation of such models.

4.3 Evolutionary Equations for Probability Distribution Laws

Exemplary random processes derived from intuitive conceptual considerations are listed above. Thus, the normal random process can be obtained from the idea about big number of independent factors, white noise from independence of subsequent values and Poisson process from an assumption of rare events (Gihman and Skorohod 1974; Volkov et al., 2000; Wentzel', 1975). All "essential properties" of these three processes are known: finite-dimensional distribution laws, statistical moments, etc.

As for Markovian processes, they are based on the ideas about relationships between the future states and the previous ones. In general, a Markovian process may be non-stationary. Thus, one can ask how an initial probability distribution changes in time, whether it converges to some stationary one and what such a limit distribution looks like. Answers to those questions are not formulated explicitly in the definition of a Markovian process. However, to get the answers, one can derive evolutionary equations for a probability distribution law based directly on the definition. For a process with a finite number of states, they take the form of a set of ordinary differential equations (*Kolmogorov equations*):

$$
\begin{bmatrix} dp_1/dt \\ dp_2/dt \\ \ldots \\ dp_K/dt \end{bmatrix} = \begin{bmatrix} -(\lambda_{1,2}+\ldots+\lambda_{1,K}) & \lambda_{2,1} & \ldots & \lambda_{K,1} \\ \lambda_{1,2} & -(\lambda_{2,1}+\lambda_{2,3}+\ldots+\lambda_{2,K}) & \ldots & \lambda_{K,2} \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_{1,K} & \lambda_{2,K} & \ldots & -(\lambda_{K,1}+\ldots+\lambda_{K,K-1}) \end{bmatrix}
$$
$$
\cdot \begin{bmatrix} p_1 \\ p_2 \\ \ldots \\ p_K \end{bmatrix}, \tag{4.6}
$$

where $p_i(t)$ is a probability of a state S_i, $\lambda_{i,j}(t)$ are transition probability densities. If the functions $\lambda_{i,j}(t)$ are given, one can trace an evolution of the probability distribution starting from any initial distribution by integrating the Kolmogorov equations. In simple particular cases, e.g. for constant $\lambda_{i,j}$, a solution can be found analytically.

A problem is somewhat simpler in the case of Markov chains (at least, for numerical investigation), since an evolution of a probability vector is described with a Kth-order difference equation. For a vivid representation of Markovian processes with

discrete states, one often uses graphs where circles and arrows indicate different states and possible transitions between them.

In the case of a continuous-valued Markovian process, a state is to be described with a probability density function rather than a probability vector. Therefore, one derives partial differential equations for an evolution of the probability distribution law rather than ordinary differential equations (4.6). This is a *generalised Markov equation* (the other names are Kolmogorov – Chapman equation and direct Kolmogorov equation) for a conditional probability density function:

$$\frac{\partial p(x, t | x_0, t_0)}{\partial t} = \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \frac{\partial^k}{\partial x^k} \left[c_k(x, t) p(x, t | x_0, t_0) \right], \qquad (4.7)$$

where $c_k(x, t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x' - x)^k p(x', t + \Delta t | x, t) \mathrm{d}x'$ are coefficients related to "probabilities of change" of a state x and determining "smoothness" of the process realisations.

In an important particular case of a *diffusive Markovian process* (where $c_k = 0$ for any $k > 2$), the equation simplifies and reduces to

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x} \left(c_1(x, t) p(x, t) \right) + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left(c_2(x, t) p(x, t) \right), \qquad (4.8)$$

where c_1 is called the drift coefficient and c_2 is the diffusion coefficient. Equation (4.8) is also called *Fokker – Planck equation* (Wentzel', 1975; Risken, 1989). It is an equation of a parabolic type. Of the same form are diffusion and heat conduction equations in mathematical physics. The names of the coefficients originate from the same field. Relationships between parameters of stochastic differential equation specifying an original process and the drift and diffusion coefficients in the Fokker–Planck equation are considered in Sect. 4.5.

4.4 Autoregression and Moving Average Processes

A random process can be specified via a stochastic equation. Then, it is defined as a solution to a stochastic equation, i.e. its substitution into an equation makes the latter an identity. In particular, discrete-time stochastic equations which define random processes of "autoregression and moving average" (Box and Jenkins, 1970) are considered below. They are very often used in modelling from observed data.

Linear filter. As a basic model for the description of complex real-world processes, one often uses Gaussian white noise $\xi(t)$. Let it have zero mean and the variance σ_ξ^2. Properties of a real-world signal may differ from those of Gaussian white noise, e.g. an estimate of the autocorrelation function $\rho(\tau)$ may significantly differ from zero at non-zero time lags τ. Then, a fruitful approach is to construct

a model as Gaussian white noise transformed by a *linear filter*. In general, such a transform in discrete time is defined as

$$x_n = \xi_n + \sum_{i=1}^{\infty} \psi_i \xi_{n-i}. \tag{4.9}$$

For the variance of x_n to be finite (i.e. for x_n to be stationary), the weights ψ_i must satisfy $\sum_{i=1}^{\infty} \psi_i^2 \leq$ const. Linear transform (4.9) preserves normality of a process and introduces non-zero autocorrelations $\rho(\tau)$ at non-zero time lags.

Moving average processes. Using a model with an infinite number of weights is impossible in practice. However, one may reasonably assume that the value of ψ_i decreases quickly with i, i.e. the remote past weakly affects the present, and consider the model (4.9) containing only a finite number of weights q. Thereby, one gets a "moving average" process which is denoted as MA(q) and defined by the difference equation

$$x_n = a_n - \sum_{i=1}^{q} \theta_i \xi_{n-i} \tag{4.10}$$

involving $q + 1$ free parameters: θ_1, θ_2, ..., θ_q, and σ_ξ^2.

Autoregression processes. A general expression (4.9) can be equivalently rewritten in the form

$$x_n = \xi_n + \sum_{i=1}^{\infty} \pi_i x_{n-i}, \tag{4.11}$$

where the weights π_i are uniquely expressed via ψ_i. In more detail, conversion from (4.9) to (4.11) can be realised through subsequent exclusion of the quantities ξ_{n-1}, ξ_{n-2}, etc. from equation (4.9). For that, one expresses ξ_{n-1} via x_{n-1} and previous values of ξ as $\xi_{n-1} = x_{n-1} - \sum_{i=1}^{\infty} \psi_i \xi_{n-1-i}$. Then, one substitutes this expression into equation (4.9), thereby excluding ξ_{n-1} from the latter. Next, one excludes ξ_{n-2} and so on in the same manner. The process (4.11) involves an infinite number of parameters π_i. However, a fast decrease $\pi_i \to 0$ at $i \to \infty$ often takes place, i.e. the remote past weakly affects the present (already in terms of the values of x variable). Then, one may take into account only a finite number of terms in equation (4.11). As a result, one gets an "autoregression" process of an order p which is denoted as AR(p) and defined as

$$x_n = \xi_n + \sum_{i=1}^{p} \phi_i x_{n-i}. \tag{4.12}$$

This model contains $p + 1$ free parameters: ϕ_1, ϕ_2, ..., ϕ_p and σ_ξ^2. The values of weights must satisfy certain relationships (Box and Jenkins, 1970) for a process

to be stationary. Thus, in the case of $p = 1$, the variance of a process (4.12) is $\sigma_x^2 = \sigma_\xi^2 / \left(1 - \phi_1^2\right)$ so that one needs $|\phi_1| < 1$ to provide the stationarity. The term "autoregression" appears, since the sum in equation (4.12) determines regression of the current value of x on the previous values of the *same* process. The latter circumstance inspires the prefix "auto". The general concept of regression is described in Sect. 7.2.1.

AR processes represent an extremely popular class of models. One of the reasons is the simplicity of their parameter estimation (see Chaps. 7 and 8). Moreover, they are often readily interpretable from the physical viewpoint. In particular, the AR(2) process given by $x_n = \phi_1 x_{n-1} + \phi_2 x_{n-2} + \xi_n$ with appropriate values of the parameters ϕ_1 and ϕ_2 describes a stochastically perturbed linear damped oscillator, i.e. a generalization of the deterministic oscillator (3.2). Its characteristic period T and relaxation time τ are related to the parameters ϕ_1 and ϕ_2 as $\phi_1 = 2\cos(2\pi/T)\exp(-1/\tau)$ and $\phi_2 = -\exp(-2/\tau)$, see Timmer et al. (1998) for a further discussion and applications of the AR(2) process to empirical modelling of physiological tremor.

The same model equation was first used for the analysis of solar activity in the celebrated work (Yule, 1927), where parameters of an AR(2) process were estimated from the time series of annual sunspot numbers. It was shown that an obtained AR(2) model could reproduce 11-year cycle of solar activity and gave better predictions than a traditional description with explicit periodic functions of time, which had been used before. Since then, linear stochastic models have become a widely used tool in many fields of data analysis. As for the modelling of solar activity, it was considered in many works after 1927. In particular, non-linear improvements of AR models are discussed, e.g., in Judd and Mees (1995; 1998); Kugiumtzis et al. (1998). Additional results on the analysis of solar activity data are presented in Sect. 12.6.

Autoregression and moving average processes. To get a more efficient construction for the description of a wide range of processes, one can combine equations (4.10) and (4.12). Reasons for their combining are as follows. Let us assume that an observed time series is generated by an AR(1) process. If one tries to describe it as an MA process, then an infinite (or at least very large) order q is necessary. Estimation of a large number of parameters is less reliable that leads to an essential reduction of the model quality. Inversely, if a time series is generated with an MA(1) process, then an AR process of a very high order p is necessary for its description. Therefore, it is reasonable to combine equations (4.10) and (4.12) in a single model to describe an observed process most parsimoniously. Thus, one gets an autoregression and moving average process of an order (p, q) which is denoted ARMA (p, q) and defined as

$$x_n = \xi_n + \sum_{i=1}^{p} \phi_i x_{n-i} - \sum_{i=1}^{q} \theta_i \xi_{n-i}. \qquad (4.13)$$

It involves $p + q + 1$ free parameters.

Autoregression and integrated moving average processes. A stationary process (4.13) cannot be an adequate model for non-stationary processes with either

deterministic trends or stochastic ones. The term "stochastic trend" means *irregular* alternation of intervals, where a process follows almost deterministic law. However, an adequate model in the case of polynomial trends is a process whose finite difference is a stationary ARMA process. A finite difference of an order d is defined as $y_n = \nabla^d x_n$, where $\nabla x_n = x_n - x_{n-1}$ is the first difference (an analogue to differentiation) and ∇^d denotes d sequential applications of the operator ∇. Thus, one gets an autoregression and integrated moving average process of an order (p, d, q) denoted ARIMA(p, d, q) and defined via the set of difference equations

$$y_n = \xi_n + \mu + \sum_{i=1}^{p} \phi_i y_{n-i} - \sum_{i=1}^{q} \theta_i \xi_{n-i},$$
$$\nabla^d x_n = y_n. \tag{4.14}$$

An intercept μ determines a deterministic trend. To express x_n via the values of the ARMA process y_n, one should use summation operator (an analogue to integration), which is inverse to the operator ∇. It explains the word "integrated" in the title of an ARIMA process.

ARMA and ARIMA processes were the main tools to model and predict complex real-world processes for more than half a century (1920–1970s). They were widely used to control technological processes (Box and Jenkins, 1970, vol. 2). Their various modifications were developed, in particular, seasonal ARIMA models defined as ARIMA processes for a seasonal difference $\nabla_s x_n = x_n - x_{n-s}$ of the kind

$$y_n = \xi_n + \sum_{i=1}^{P} \Phi_i y_{n-is} - \sum_{i=1}^{Q} \Theta_i \xi_{n-is},$$
$$\nabla_s^D x_n = y_n, \tag{4.15}$$

where ξ_n is an ARIMA (p, d, q) process. A process (4.15) is called a seasonal ARIMA process of an order $(P, D, Q) \times (p, d, q)$. Such models are relevant to describe processes with seasonal trends (i.e. a characteristic timescale s).

Only during the last two decades due to the development of computers and concepts of non-linear dynamics, ARIMA models more and more "step back" in a competition with non-linear models (Chaps. 8, 9, 10, 11, 12 and 13), though they remain the main tool in many fields of knowledge and practice.

4.5 Stochastic Differential Equations and White Noise

4.5.1 The Concept of Stochastic Differential Equation

To describe continuous-time random processes, one uses *stochastic differential equations* (SDEs). The most well known is the first-order equation (so-called Langevin equation)

$$\mathrm{d}x(t)\big/\mathrm{d}t = F(x, t) + G(x, t) \cdot \xi(t), \tag{4.16}$$

where F and G are smooth functions of their arguments, $\xi(t)$ is zero-mean Gaussian white nose with the auto-covariance function $\langle \xi(t)\xi(t + \tau)\rangle = \delta(\tau)$.

Introducing a concept of SDE is by no means trivial since it includes a concept of the random process derivative $dx(t)/dt$. How should one understand such a derivative? The simplest way would be to assume that all the realisations of a process x are continuously differentiable and define the derivative at a point t as a random quantity, whose value is an ordinary derivative of a single realisation of x at t. However, this is possible only for processes $\xi(t)$ with sufficiently smooth realisations so that for each specific realisation of $\xi(t)$, equation (4.16) can be considered and solved as a usual ODE. However, white noise does not belong to such class of processes but reasonably describes multitude of practical situations (a series of independent shocks) and allows simplification of mathematical manipulations. To have an opportunity to analyse equation (4.16) with white noise $\xi(t)$, one generalises the concept of derivative $dx(t)/dt$ of a random process x at a point t. The derivative is defined as a random quantity

$$\frac{dx(t)}{dt} = \lim_{\Delta t \to 0} \frac{x(t + \Delta t) - x(t)}{\Delta t},$$

where the limit is taken in the root-mean-squared sense (see, e.g., Gihman and Skorohod, 1974; Volkov et al., 2000; Wentzel', 1975). However, even such a concept does not help much in practice. The point is that one should somehow integrate equation (4.16) to get a solution. Formally, one can write

$$x(t) - x(t_0) = \int_{t_0}^{t} F(x(t'), t')dt' + \int_{t_0}^{t} G(x(t'), t') \cdot \xi(t')dt' \qquad (4.17)$$

and estimate a solution over an interval $[t_0, t]$ via the estimation of the integrals. A stochastic integral is also defined via the limit in the root-mean-squared sense. However, its definition is not unique. There are two most popular forms of the stochastic integral: (i) Ito's integral is defined analogous to the usual Riemann's integral via the left rectangle formula and allows to get many analytic results (Oksendal, 1995); (ii) Stratonovich's integral is defined via the central rectangles formula (a symmetrised form of the stochastic integral) (Stratonovich, 1967); it is more readily interpreted from the physical viewpoint since it is symmetric with respect to time (Mannella, 1997). Moreover, one can define the generalised stochastic integral, whose particular cases are Ito's and Stratonovich's integrals (Gihman and Skorohod, 1974; Mannella, 1997; Volkov et al., 2000; Wentzel', 1975). Thus, the stochastic DE (4.16) gets an exact meaning if one indicates in which sense the stochastic integrals are to be understood.

If $G(x, t) = G_0 = $ const, all the above-mentioned forms of the stochastic integral $\int_{t_0}^{t} G(x(t'), t') \cdot \xi(t')dt'$ coincide. Below, we consider this simple case in more detail. One can show that a process x in equation (4.17) is Markovian. Thus, one

can write down the corresponding Fokker – Planck equation where the drift coefficient is $F(x, t)$ and the diffusion coefficient is G_0^2. Let us consider a particular case of $F = 0$:

$$\mathrm{d}x(t)\big/\mathrm{d}t = G_0\xi(t). \tag{4.18}$$

The solution to this equation can be written down formally as

$$x(t) - x(t_0) = G_0 \int_{t_0}^{t} \xi(t')\mathrm{d}t'. \tag{4.19}$$

One can show that the process (4.19) is Wiener's process. Its variance linearly depends on time as $\sigma_x^2(t) = \sigma_x^2(t_0) + G_0^2 \cdot (t - t_0)$. The variance of its increments over an interval Δt is equal to $G_0^2\Delta t$. It agrees well with known observations of Brownian particle motion, where mean square of the deviation from a starting point is also proportional to time.

A geophysical example. Equation (4.18) allows to derive an empirically established *Gutenberg – Richter law* for the repetition time of earthquakes depending on their intensity (Golitzin, 2003). Let x be a value of a mechanical tension (proportional to deformations) at a given domain of the earth's crust. Let us assume that it is accumulated due to different random factors (various shocks and so forth) described as white noise. On average, its square rises as $G_0^2(t - t_0)$ (4.18) starting from a certain zero time instant when the tension is weak. Earthquakes arise when the system accumulates sufficient elastic energy during a certain time interval and releases it in some way. If the release occurs when a fixed threshold E is reached, then a time interval necessary to accumulate such energy reads as $\tau = E/G_0^2$. From here, it follows that the frequency of occurrence of earthquakes with energy exceeding E is $\sim 1/\tau \sim G_0^2/E$, i.e. the frequency of occurrence is inversely proportional to energy. The Gutenberg – Richter law reduces to the same form under certain assumptions. Analogous laws describe appearance of tsunami, landslides and similar events (Golitzin, 2003).

An example from molecular physics. Under an assumption that independent shocks abruptly change a velocity of a particle rather than its coordinate, i.e. the white noise represents random forces acting on a particle, one gets the second-order SDE:

$$\mathrm{d}^2x(t)\big/\mathrm{d}t^2 = G_0\xi(t). \tag{4.20}$$

It allows to derive analytically the *Richardson – Obukhov law* stating that mean square of the displacement of a Brownian particle rises with time as $(t - t_0)^3$ under certain conditions. This law holds true for the sea surface within some range of scales (it is known as relative diffusion) (Golitzin, 2003).

4.5.2 Numerical Integration of Stochastic Differential Equations

The above examples allow an analytic solution but for non-linear F and/or G, one has to use numerical techniques, which differ from those for ODEs. For simplicity, we start again with equation (4.16) with $G(x, t) = G_0 = \text{const}$:

$$\mathrm{d}x\big/\mathrm{d}t = F(x(t)) + G_0\xi(t). \qquad (4.21)$$

At a given initial condition $x(t)$, an SDE determines an ensemble of possible future realisations rather than a single realisation. The function F uniquely determines only the conditional probability density functions $p(x(t + \Delta t)|x(t))$. If F is non-linear, one cannot derive analytic formulas for the conditional distributions. However, one can get those distributions numerically via the generation of an ensemble of the SDE realisations. For that, the noise term $\xi(t')$ over an interval $[t, \ t + \Delta t]$ is simulated with the aid of pseudo-random number generator and the SDE is numerically integrated step by step.

The simplest approach is to use the Euler technique with a small integration step h (see, e.g., Mannella, 1997; Nikitin and Razevig, 1978). The respective difference scheme for equation (4.21) reads as

$$x(t + h) - x(t) = F(x(t)) \cdot h + \varepsilon_0(t) \cdot G_0\sqrt{h}, \qquad (4.22)$$

where $\varepsilon_0(t), \varepsilon_0(t + h), \varepsilon_0(t + 2h), \ldots$ are independent identically distributed Gaussian random quantities with zero mean and unit variance. The second term in the right-hand side of equation (4.22) shows that the noise contribution to the difference scheme scales with the integration step as $\sqrt{h}$. This effect is not observed in ODEs where the contribution of the entire right-hand side is of the order of h or higher. For SDEs, such an effect takes place due to the integration of the white noise $\xi(t)$: the difference scheme includes the random term whose variance is proportional to the integration step. The random term dominates for very small integration steps h. The scheme (4.22) is characterised by an integration error of the order $h^{3/2}$, while for ODEs the Euler technique gives an error of the order of h^2.

Further, at a fixed step h, one can generate an ensemble of noise realisations $\varepsilon_0(t), \varepsilon_0(t + h), \varepsilon_0(t + 2h), \ldots$ and compute for each realisation the value of $x(t + \Delta t)$ at the end of the time interval of interest via the formula (4.22). From an obtained set of values of $x(t + \Delta t)$, one can construct a histogram, which is an estimate of the conditional probability density $p(x(t + \Delta t)|x(t))$. This estimate varies under the variation of h and tends to a true distribution only in the limit $h \to 0$ like an approximate solution of an ODE tends to a true one for $h \to 0$. In practice, one should specify so small integration step h that an approximate distribution would weakly change under further decrease in h. Typically, to get the same order of accuracy, one must use smaller steps for the integration of SDEs as compared with the corresponding ODEs to get similar convergence. This is due to the above-mentioned lower order of accuracy for the SDEs.

A process x in equation (4.16) or (4.21) may well be vector valued. Then, white noise $\xi(t)$ is also a multidimensional process. All the above considerations remain the same for vector processes. As an example, let us consider integration of stochastic equations of the van der Pol oscillator:

$$
\begin{aligned}
dx_1/dt &= x_2, \\
dx_2/dt &= \mu(1 - x_1^2)x_2 - x_1 + \xi(t),
\end{aligned}
\tag{4.23}
$$

with $\mu = 3$ (Timmer, 2000) and $G_0 = 1$. Estimates of the conditional distribution $p(x_1(t + \Delta t)|x(t))$ are shown in Fig. 4.2. We take the initial conditions $\mathbf{x}(t) = (0, -4)$ lying close to probable states of the system observed in a long numerically obtained orbit, $\Delta t = 0.5$ corresponding approximately to 1/18 of a basic period, and integration steps $h = 0.1, 0.01, 0.001$ and 0.0001. A distribution estimate stabilises at $h = 0.001$. Thus, an integration step should be small enough to give a good approximation to conditional distributions, often not more than about 0.0001 of a basic period. For a reasonable convergence of a numerical technique for the corresponding ODE, i.e. equation (4.23) without noise, a step 0.01 always suffices.

One more reason why dealing with SDEs is more complicated than numerical integration of ODEs is the above-mentioned circumstance that the integral of a random process $\xi(t)$ is an intricate concept. Let us now consider equation (4.16) with a non-constant function G. The corresponding Fokker–Planck equation takes different forms depending on the definition of the stochastic integrals. Namely, the drift coefficient reads as $F(x, t)$ under Ito's definition and as

$$
F(x, t) + \frac{1}{2}\frac{\partial G(x, t)}{\partial x}G(x, t)
$$

under Stratonovich's definition (Nikitin and Razevig, 1978; Mannella, 1997; Risken, 1989) (the diffusion coefficient is $G^2(x, t)$ in both cases). Accordingly, the
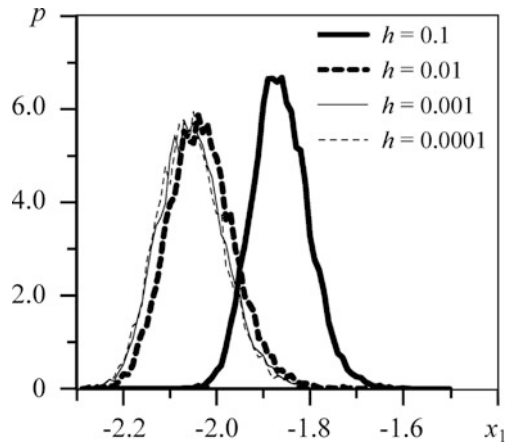
Fig. 4.2 Probability density estimates $p(x_1(t + \Delta t)|\mathbf{x}(t))$ for the system (4.23) at integration step $h = 0.1$, 0.01, 0.001, 0.0001 and initial conditions $\mathbf{x}(t) = (0, -4)$. Each histogram is constructed from an ensemble of 10,000 time realisations with a bin size of 0.01. Good convergence is observed at $h = 0.001$

Euler scheme, i.e. a scheme accurate up to the terms of the order $h^{3/2}$, depends on the meaning of the stochastic integrals. For Ito's integrals, the Euler scheme is similar to the above case (Nikitin and Razevig, 1978) and reads as

$$x(t+h) - x(t) = F(x(t)) \cdot h + \varepsilon_0(t) G(x(t)) \sqrt{h}. \qquad (4.24)$$

For Stratonovich's integrals often considered in physics (Risken, 1989; Mannella, 1997; Siegert et al., 1998), the Euler scheme takes the form

$$x(t+h) - x(t) = F(x(t)) \cdot h + \frac{1}{2} \frac{\partial G(x,t)}{\partial x} G(x(t)) \varepsilon_0^2(t) h + G(x(t)) \varepsilon_0(t) \sqrt{h}, \quad (4.25)$$

where an additional term of the order $O(h)$ appears. The latter is necessary to provide the integration error not greater than $O(h^{3/2})$ (Mannella, 1997; Nikitin and Razevig, 1978).

If one needs a higher order of accuracy, then the formula gets even more complicated, especially in the case of Stratonovich's integrals. It leads to several widespread pitfalls. In particular, a seemingly reasonable idea to integrate "deterministic" ($F(x, t)$) and "stochastic" ($G(x, t)\xi(t)$) parts of equation (4.16) separately, representing the "deterministic" term with the usual higher order Runge – Kutta formulas and the "stochastic" term in the form $\varepsilon_0(t) \cdot G(x(t)) \sqrt{h}$, is called an "exact propagator". However, it appears to give even worse accuracy than the simple Euler technique (4.25) since the integration of the "deterministic" and "stochastic" parts is unbalanced (Mannella, 1997). An interested reader can find correct formulas for integration of SDEs with higher orders of accuracy in Mannella (1997) and Nikitin and Razevig (1978).

4.5.3 Constructive Role of Noise

Noise (random perturbations of dynamics) is often thought of as an interference, an obstacle, something harmful for the functioning of a communication system, detection of an auditory signal and other tasks. However, it appears that noise in non-linear systems can often play a constructive role leading to enhancement of their performance. The most striking and widely studied phenomena of this type are called "stochastic resonance" and "coherence resonance".

The term *stochastic resonance* was introduced in Benzi et al. (1981) where the authors found an effect, which they tried to use for explanation of the ice age periodicity (Benzi et al., 1982). The same idea was developed independently in Nicolis (1981; 1982). The point is that the evolution of the global ice volume on the Earth during the last million years exhibits a kind of periodicity with an average period of about 10^5 years (a glaciation cycle). The only known similar timescale is observed for the variations in the eccentricity of the Earth's orbit around the Sun determined by influences of other bodies of the solar system. The perturbation in the total amount of solar energy received by the Earth due to this effect is about 0.1%. Then,

the question arises whether such a small periodic perturbation can be amplified so strongly to induce such a large-scale phenomenon as alternation of ice ages.

Benzi and co-authors considered an overdamped bistable oscillator driven by a Gaussian white noise ξ and a weak periodic signal $A\cos(\Omega t)$:

$$\mathrm{d}x \big/ \mathrm{d}t = x(a - x^2) + A\cos(\Omega t) + \xi(t) \qquad (4.26)$$

with the parameter $a > 0$. The corresponding autonomous system $\mathrm{d}x/\mathrm{d}t = x(a - x^2)$ has an unstable fixed point $x = 0$ and two stable fixed points $x = \pm\sqrt{a}$. In the presence of noise, an orbit spends a long time near one of the two stable states but sometimes jumps to another state due to the noise influence. Switching between the two states is quite irregular: only on average it exhibits a characteristics timescale, the so-called Kramers' rate.

In the presence of the periodic driving $A\cos(\Omega t)$, one can consider the system (4.26) as a transformation of an input signal $A\cos(\Omega t)$ into "output" signal $x(t)$. In other words, the system (4.26) performs signal detection. Performance of the system is better if $x(t)$ is closer to a harmonic one with the frequency Ω. Its closeness to periodicity can be quantified in different ways. In particular, a signal-to-noise ratio (SNR) can be introduced as the ratio of its power spectral density (Sect. 6.4.2) at the frequency Ω to its "noise-floor" spectral density. It appears that the dependence of SNR on the intensity of the noise ξ has a clear maximum at a non-zero noise level. The curve "SNR versus noise intensity" resembles the resonance curve of "output amplitude versus driving frequency". In the particular example of the system (4.26), the phenomenon can be explained by the dependence of Kramers' rate on the noise level so that the resonance takes place when Kramers' rate becomes equal to the driving frequency. Therefore, the phenomenon was called "stochastic *resonance*".

Thus, a weak periodic input may have stronger periodic output for some intermediate (non-zero) noise level. In other words, non-zero noise improves the system performance as compared with the noise-free case. Whether this phenomenon is appropriate to describe the glaciation cycles is still the matter of debate, but the effect was then observed in many non-linear systems of different origin (see the reviews Anishchenko et al., 1999; Bulsara et al., 1993; Ermentrout et al., 2008; Gammaitoni et al., 1998; McDonnell and Abbott, 2009; Moss et al., 2004; Nicolis, 1993; Wiesenfeldt and Moss, 1995). In particular, many works report stochastic resonance in neural systems such as mechanoreceptors of crayfish (Douglass et al., 1993), sensory neurons of paddlefish (Greenwood et al., 2000 ; Russell et al., 1999), other physiological systems (Cordo et al., 1996; Levin and Miller, 1996), different neuron models (Longtin, 1993; Volkov et al., 2003b, c) and so on. There appeared many extensions and reformulations of the concept such as aperiodic stochastic resonance (Collins et al., 1996), stochastic synchronisation (Silchenko and Hu, 2001; Silchenko et al., 1999; Neiman et al., 1998) and stochastic multiresonance (Volkov et al., 2005).

Currently, many researchers speak of a stochastic resonance in the following situation: (i) one can define input and output signals for a non-linear system;

(ii) performance of the system improves at some non-zero noise level as compared to the noise-free setting. Thus, the formulation is no longer restricted to weak and/or periodic input signals and bistable systems. The counterintuitive idea that noise can improve functioning of a system finds the following fundamental explanation (McDonnell and Abbott, 2009): The system is non-linear and its parameter values in a noise-free setting are suboptimal for the performance of a required task. Hypothetically, its performance could be improved by adjusting the parameters. The other way is the noise influence which may improve functioning of the system. Thus, stochastic resonance is a result of an interplay between noise and non-linearity. It cannot be observed in a linear system.

A similar phenomenon introduced in Pikovsky and Kurths (1997) is called *coherence resonance*. It is observed in excitable non-linear systems without input signals. Its essence is that an output signal of a system is most coherent (exhibits the sharpest peak in a power spectrum) at a non-zero noise level and becomes less regular both for weaker and stronger noises. The phenomenon was first observed in FitzHugh – Nagumo system which is sometimes used as a simple neuron model:

$$\varepsilon \, dx/dt = x - x^3/3 - y,$$
$$dy/dt = x + a + \xi(t). \tag{4.27}$$

The parameter $\varepsilon \ll 1$ determines the existence of fast motions (where only x changes) and slow motions (where $y \approx x - x^3/3$). The parameter $|a| > 1$ so that a stable fixed point is the only attractor of the noise-free system. The noise ξ is Gaussian and white. A stable limit cycle appears for $|a| < 1$. Thus, for $|a|$ slightly greater than 1, the system becomes excitable, i.e. a small but finite deviation from the fixed point (induced by the noise) can produce a large pulse (spike) in the realisation $x(t)$. These spikes are generated quite irregularly. The quantitative "degree of regularity" can be defined as the ratio of the mean interspike interval to the standard deviation of the interspike intervals. This quantity depends on the noise level: It is small for zero noise and strong noise and takes its maximum at an intermediate noise intensity. Again, the curve "degree of regularity versus noise intensity" looks like an oscillator resonance curve "output amplitude versus driving frequency". In the case of the system (4.27), the phenomenon can be explained by the coincidence of the two characteristic times: an activation time (the mean time needed to excite the system from the stable point, i.e. to get strong enough noise shock) and an excursion time (the mean time needed to return from an excited state to the stable state) (Pikovsky and Kurths, 1997). Therefore, the phenomenon is called "coherence *resonance*".

Thus, some non-zero noise may provide the most coherent output signal. It was quite unexpected finding which appeared fruitful to explain many observations. Similarly to stochastic resonance, the concept of coherence resonance was further extended, e.g., as doubly stochastic coherence (Zaikin et al., 2003), spatial (Sun et al., 2008b) and spatiotemporal (Sun et al., 2008a) coherence resonances. It is widely exploited in neuroscience, in particular, coherence resonance was observed

in neuron models (e.g. Lee et al., 1998; Volkov et al., 2003a) and central nervous system (e.g. Manjarrez et al., 2002).

To summarise, a possible constructive role of noise for functioning of natural non-linear systems and its exploitation in new technical devices is currently a widely debated topic in very different fields of research and applications.

References

Anishchenko, V.S., Neiman, A.B., Moss, F., Schimansky-Geier, L.: Stochastic resonance: Noise enhanced order. Physics – Uspekhi. **42**, 7–38 (1999)

Benzi, R., Parisi, G., Sutera, A., Vulpiani, A.: Stochastic resonance in climate change. Tellus. **34**, 10–16 (1982)

Benzi, R., Sutera, A., Vulpiani, A.: The mechanism of stochastic resonance. J. Phys. A. **14**, L453–L457 (1981)

Box, G.E.P., Jenkins, G.M.: Time series analysis. Forecasting and Control. Holden-Day, San Francisco (1970)

Bulsara, A., Hänggi, P., Marchesoni, F., Moss, F., Shlesinger, M. (eds.): Stochastic resonance in physics and biology. J. Stat. Phys. **70**, 1–512 (1993)

Collins, J.J., Chow, C.C., Capela, A.C., Imhoff, T.T.: Aperiodic stochastic resonance. Phys. Rev. E. **54**, 5575–5584 (1996)

Cordo, P., Inglis, J.T., Verschueren, S., Collins, J.J., Merfeld, D.M., et al.: Noise in human muscle spindles. Nature **383**, 769–770 (1996)

Douglass, J.K., Wilkens, L., Pantazelou, E., Moss, F.: Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. Nature 365, 337–339 (1993)

Ermentrout, G.B., Galan, R.F., Urban, N.N.: Reliability, synchrony and noise. Trends Neurosci. 31, 428–434 (2008)

Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F.: Stochastic Resonance. Rev. Mod. Phys. 70, 223–287 (1998)

Gihman, I.I., Skorohod, A.V.: The Theory of Stochastic Processes. Springer, Berlin (1974)

Golitsyn, G.S.: White noise as a base for explanation of many laws in nature. In: Gaponov-Grekhov, A.V., Nekorkin, V.I. (eds.) Nonlinear Waves – 2002. Institute of Applied Physics RAS, pp. 117–132. Nizhny Novgorod, (in Russian) (2003)

Greenwood, P.E., Ward, L.M., Russell, D.F., Neiman, A., Moss, F.: Stochastic resonance enhances the electrosensory information available to paddlefish for prey capture. Phys. Rev. Lett. **84**, 4773–4776 (2000)

Kendall, M.G., Stuart, A.: The Advanced Theory of Statistics,. vol. 2 and 3. Charles Griffin, London (1979)

Lee, S.-G., Neiman, A., Kim, S.: Coherence resonance in a Hodgkin-Huxley neuron. Phys. Rev. E. **57**, 3292–3297 (1998)

Levin, J.E., Miller, J.P.: Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. Nature **380**, 165–168 (1996)

Longtin, A.: Stochastic resonance in neuron models. J. Stat. Phys. **70**, 309–327 (1993)

Malakhov, A.N.: Fluctuations in self-oscillatory systems. Nauka, Moscow, (in Russian) (1968)

Mannella, R. Numerical integration of stochastic differential equations. In: Vazquez, L., Tirando, F., Martin, I. (eds.) Supercomputation in Nonlinear and Disordered Systems: Algorithms, Applications and Architectures, pp. 100–129. World Scientific, Singapore (1997)

McDonnell, M.D., Abbott, D.: What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. PLoS Comput. Biol. **5**, e1000348 (2009)

Moss, F., Ward, L., Sannita, W.: Stochastic resonance and sensory information processing: a tutorial and review of application. Clin. Neurophysiol. **115**, 267–281 (2004)

Neiman, A., Silchenko, A., Anishchenko, V., Schimansky-Geier, L.: Stochastic resonance: Noise-enhanced phase coherence. Phys. Rev. E. **58**, 7118–7125 (1998)

Nicolis, C.: Solar variability and stochastic effects on climate. Sol. Phys. **74**, 473–478 (1981)

Nicolis, C.: Stochastic aspects of climatic transitions-response to a periodic forcing. Tellus. **34**, 1–9 (1982)

Nicolis, C.: Long-term climatic transitions and stochastic resonance. J. Stat. Phys. **70**, 3–14 (1993)

Nikitin, N.N., Razevig, V.D.: Methods of digital modelling of stochastic differential equations and estimation of their errors. Comput. Math. Math. Physics. **18**(1), 106–117, (in Russian) (1978)

Oksendal, B.K.: Stochastic Differential Equations: An Introduction with Applications. Springer, Berlin (1995)

Pikovsky, A.S., Kurths, J.: Coherence resonance in a noise-driven excitable system. Phys. Rev. Lett. **78**, 775–778 (1997)

Risken, H.: The Fokker – Planck Equation. Springer, Berlin (1989)

Russell, D.F., Wilkens, L.A., Moss, F.: Use of behavioural stochastic resonance by paddlefish for feeding. Nature **402**, 291–294 (1999)

Rytov, S.M., Kravtsov Yu.A., Tatarskiy, V.I.: Introduction to statistical radiophysics (part II). Nauka, Moscow (1978). Translated into English: Springer, Berlin (1987)

Siegert, S., Friedrich, R., Peinke, J.: Analysis of data sets of stochastic systems. Phys. Lett. A. **243**, 275–280 (1998)

Silchenko, A., Hu, C.-K.: Multifractal characterization of stochastic resonance. Phys. Rev. E. **63**, 041105 (2001)

Silchenko, A., Kapitaniak, T., Anishchenko, V.: Noise-enhanced phase locking in a stochastic bistable system driven by a chaotic signal. Phys. Rev. E. **59**, 1593–1599 (1999)

Stratonovich, R.L.: Topics in the theory of random noise. Gordon and Breach, New York (1967)

Sun, X., Lu Q, Kurths, J.: Correlated noise induced spatiotemporal coherence resonance in a square lattice network. Phys. A. **387**, 6679–6685 (2008a)

Sun, X., Perc, M., Lu, Q., Kurths, J.: Spatial coherence resonance on diffusive and small-world networks of Hodgkin–Huxley neurons. Chaos. **18**, 023102 (2008b)

Timmer, J.: Parameter estimation in nonlinear stochastic differential equations. Chaos, Solitons Fractals. **11**, 2571–2578 (2000)

Volkov, E.I., Ullner, E., Kurths, J.: Stochastic multiresonance in the coupled relaxation oscillators. Chaos. **15**, 023105 (2005)

Volkov, E.I., Stolyarov, M.N., Zaikin, A.A., Kurths, J.: Coherence resonance and polymodality in inhibitory coupled excitable oscillators. Phys. Rev. E. **67**, 066202 (2003a)

Volkov, E.I., Ullner, E., Zaikin, A.A., Kurths, J.: Oscillatory amplification of stochastic resonance in excitable systems. Phys. Rev. E. **68**, 026214 (2003b)

Volkov, E.I., Ullner, E., Zaikin, A.A., Kurths, J.: Frequency-dependent stochastic resonance in inhibitory coupled excitable systems. Phys. Rev. E. **68**, 061112 (2003c)

Volkov, I.K., Zuev, S.M., Ztvetkova, G.M.: Random Processes. N.E. Bauman MGTU, Ìoscow, (in Russian) (2000)

Wentzel' A.D.: Course of the Theory of Random Processes. Nauka, Moscow (1975). Translated into English: A course in the theory of stochastic processes. McGraw-Hill Intern., New York (1981)

Wiesenfeldt, K., Moss, F.: Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. Nature **373**, 33–36 (1995)

Zaikin, A., Garcia-Ojalvo, J., Bascones, R., Ullner, E., Kurths, J. Doubly stochastic coherence via noise-induced symmetry in bistable neural models. Phys. Rev. Lett. **90**, 030601 (2003)