

---

# From Random Differential Equations to Structural Causal Models: the stochastic case

---

**Stephan Bongers**  
 Informatics Institute  
 University of Amsterdam  
 The Netherlands  
 S.R.Bongers@uva.nl

**Joris M. Mooij**  
 Informatics Institute  
 University of Amsterdam  
 The Netherlands  
 J.M.Mooij@uva.nl

## Abstract

Random Differential Equations provide a natural extension of Ordinary Differential Equations to the stochastic setting. We show how, and under which conditions, every equilibrium state of a Random Differential Equation (RDE) can be described by a Structural Causal Model (SCM), while pertaining the causal semantics. This provides an SCM that captures the stochastic and causal behavior of the RDE, which can model both cycles and confounders. This enables the study of the equilibrium states of the RDE by applying the theory and statistical tools available for SCMs, for example, marginalizations and Markov properties, as we illustrate by means of an example. Our work thus provides a direct connection between two fields that so far have been developing in isolation.

## 1 Introduction

Uncertainty and random fluctuations are a very common feature of real dynamical systems. For example, most physical, financial, biochemical and engineering systems are subjected to time-varying external or internal random disturbances. These complex disturbances and their associated responses are most naturally described in terms of stochastic processes. A more realistic formulation of a dynamical system in terms of differential equations should involve such stochastic processes. This led to the fields of stochastic and random differential equations, where the latter deals with processes that are sufficiently regular. Random differential equations (RDEs) provide the most natural extension of ordinary differential equations to the stochastic setting and have been widely accepted as an important mathematical tool in modeling

and analysis of numerous processes in physics and engineering systems (Bunke, 1972; Soong, 1973; Sobczyk, 1991; Rupp and Neckel, 2013).

These internal and external disturbances of RDEs are not only of stochastic nature, but they are also of causal nature. They are causal in the sense that the disturbance processes are affecting other processes of the system. This allows us to model interventions on RDEs by forcing certain processes to be of a certain form, e.g. moving an object to a fixed position. Perfect or surgical interventions break any other causal influences on the intervened processes, but other types of interventions also occur in practice.

Although at least in principle random differential equations could be used for modeling causal relationships between the processes, inferring such causal models from data is often difficult. A significant practical drawback of this modeling class is that obtaining time series data with sufficiently high temporal resolution is often costly, impractical or even impossible. Another issue is that if one has only access to a subset of the system's processes, for example due to practical limitations on the measurability of some of the processes, then in general there does not have to exist an RDE on this subset of processes that could be estimated. A similar issue arises when the RDE contains exogenous latent confounding processes.

Structural causal models (SCMs), also known as structural equation models, are another well-studied causal modeling tool and have been widely applied in the genetics, economics, engineering and social sciences (Pearl, 2009; Spirtes et al., 2000; Bollen, 1989). One of the advantages of SCMs over other causal modeling tools is that they have the ability to deal with cyclic causal relationships (Spirtes, 1995; Hyttinen et al., 2012; Mooij et al., 2011; Forré and Mooij, 2017; Bongers et al., 2016). In particular, recent work has shown how one can apply Markov properties (Forré and Mooij, 2017), how one can deal with marginalization and how one can causally in-

interpret these models in the cyclic setting (Bongers et al., 2016).

Over the years, several attempts have been made to interpret these structural causal models that include cyclic causal relationships. They can be derived from an underlying discrete-time or continuous-time dynamical system (Fisher, 1970; Iwasaki and Simon, 1994; Dash, 2005; Lacerda et al., 2008; Mooij et al., 2013). All these methods assume that the dynamical system under consideration converges to a single static equilibrium, with the exception of the analysis by Fisher (1970), who assumes that observations are time averages of a dynamical system. These assumptions give rise to a more parsimonious description of the causal relationships of the equilibrium states and ignore the complicated but decaying transient dynamics of the dynamical system. The assumption that the system has to equilibrate to a single static equilibrium is rather strong and limits the applicability of the theory, as many dynamical systems have multiple equilibrium states.

In this paper, we relax this condition and capture, under certain convergence assumptions, every random equilibrium state of the RDE in an SCM. Conversely, we show that under suitable conditions, every solution of the SCM corresponds to a sample-path solution of the RDE. Intuitively, the idea is that in the limit when time tends to infinity the random differential equations converge exactly to the structural equations of the SCM. Moreover, we show that this construction is compatible with interventions under similar convergence assumptions. We like to stress that our construction automatically captures the stochastic behavior of the RDE in the associated SCM. It can deal with randomness in the initial conditions, the coefficients and via the random inhomogenous part (captured as additive noise in the SCM), thereby significantly extending the work by Mooij et al. (2013) who only considers the deterministic setting.

The advantage of SCMs over RDEs is that by not modeling the transient random dynamics of the RDE, one arrives at a more compact representation for learning and prediction purposes of random systems that have reached equilibrium. Another advantage is that one can marginalize over a subset of the systems variables and get a more parsimonious representation that preserves the causal semantics (Bongers et al., 2016). Yet another advantage is that it is easier to deal with confounders within the framework of SCMs, as we only need to model the equilibrium distribution of these confounders, and don't need to model their dynamics.

The remainder of the paper is organized as follows: In Section 2 we review the necessary theory about stochastic processes in order to describe RDEs. In Section 3 we

introduce random dynamical models, that define RDEs together with interventions, and we discuss convergence properties of those models. In Section 4 we introduce structural causal models. In Section 5 we present our main result, which builds the bridge between RDEs and SCMs. In Section 6 we give an example from chemical kinetics and Section 7 contains a discussion including some open problems. Proofs are provided in the Supplementary Material.

## 2 Preliminaries

Let  $T$  be a set. A *stochastic process* is an  $\mathbb{R}^n$ -valued function  $\mathbf{X} : T \times \Omega \rightarrow \mathbb{R}^n$  such that  $\mathbf{X}_t$  (which denotes  $\mathbf{X}(t, \cdot)$ ) is for each  $t \in T$  an  $\mathcal{F}$ -measurable function<sup>1</sup> (i.e. a random variable) on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . We will always assume that there exists some background probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  on which the random variables and processes live. Furthermore,  $T$  will always denote an interval in  $\mathbb{R}$  and has the meaning of time, if not stated otherwise. For each  $\omega \in \Omega$  we have an  $\mathbb{R}^n$ -valued function  $\mathbf{X}_t(\omega) : T \rightarrow \mathbb{R}^n$  from  $t$  to  $\mathbf{X}_t(\omega)$ , which is called a *sample path* or *realization* of  $\mathbf{X}$ .

Let two such processes  $\mathbf{X}$  and  $\mathbf{Y}$  be *equivalent*, i.e., for all  $t \in T$  we have  $\mathbb{P}(\mathbf{X}_t = \mathbf{Y}_t) = 1$ , then for all  $t \in T$  there are sets  $\Omega_t \in \mathcal{F}$  such that  $\mathbb{P}(\Omega_t) = 1$  and  $\mathbf{X}_t(\omega) = \mathbf{Y}_t(\omega)$  holds for all  $(t, \omega) \in \cup_{t \in T} (\{t\} \times \Omega_t)$ . If, moreover, one can choose the sets  $\Omega_t$  independently of  $t$ , that is  $\Omega_t = \Omega^*$  such that  $\mathbf{X}_t(\omega) = \mathbf{Y}_t(\omega)$  holds for all  $(t, \omega) \in T \times \Omega^*$ , then we denote such an equivalence between the processes  $\mathbf{X}$  and  $\mathbf{Y}$  by  $\mathbf{X} \stackrel{T}{=} \mathbf{Y}$ .

If the sample paths of a stochastic process  $\mathbf{X}$  are continuous on  $T$  for almost all  $\omega \in \Omega$ , then the process  $\mathbf{X}$  is called *sample-path continuous* on  $T$ . For a sample-path continuous process  $\mathbf{X}$  and  $t \in T$  the function defined, with probability one, by

$$\lim_{s \rightarrow t} \mathbf{X}_s$$

is a random variable, if it exists (Doob, 1953; Loève, 1977). Moreover, a stochastic process  $\mathbf{X}$  is called *sample-path differentiable* on  $T$ , if there exists a set  $\Omega^* \in \mathcal{F}$  such that  $\mathbb{P}(\Omega^*) = 1$  and that for all  $\omega \in \Omega^*$  the derivative

$$\dot{\mathbf{X}}_t(\omega) := \frac{d}{dt} \mathbf{X}_t(\omega) := \lim_{h \rightarrow 0} \frac{\mathbf{X}_{t+h}(\omega) - \mathbf{X}_t(\omega)}{h}, \quad t \in T,$$

exists. The mapping  $\dot{\mathbf{X}} : T \times \Omega^* \rightarrow \mathbb{R}^n$  is called the *sample-path derivative* of  $\mathbf{X}$ . Note that there is always a stochastic process  $\mathbf{Y}$  such that  $\mathbf{Y} \stackrel{T}{=} \dot{\mathbf{X}}$ . Similarly

<sup>1</sup>Assuming the Borel  $\sigma$ -algebra  $\mathfrak{B}^n$  on  $\mathbb{R}^n$ , that is the smallest  $\sigma$ -algebra on  $\mathbb{R}^n$  which contains all open  $n$ -balls.

the *sample-path integral* is defined, that is, a stochastic process  $\mathbf{X}$  is called *sample-path integrable* on  $T$ , if the integral<sup>2</sup>  $\int_T \mathbf{X}_t(\omega) dt$  exists for almost all  $\omega \in \Omega$ .

A random vector  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  can itself be seen as a stochastic process, that is the process  $\mathbf{X} : T \times \Omega \rightarrow \mathbb{R}^n$  defined by  $\mathbf{X}(t, \omega) := \mathbf{X}(\omega)$ . This stochastic process is by definition sample-path continuous. Moreover, if a process  $\mathbf{X} : T \times \Omega \rightarrow \mathbb{R}^n$  is sample-path continuous on  $T = \mathbb{R}$  and there exists a random variable  $\mathbf{X}^* : \Omega \rightarrow \mathbb{R}^n$  such that

$$\lim_{t \rightarrow \infty} \mathbf{X}_t = \mathbf{X}^*$$

almost surely, then we say that  $\mathbf{X}$  *converges to*  $\mathbf{X}^*$  and we will call the process  $\mathbf{X}$  *convergent*.

### 3 Random Differential Equations

Ordinary differential equations, which have the general form

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

provide a simple deterministic description of the dynamics of dynamical systems. The solution  $x(t)$  of an initial value problem consisting of differential equation (1) together with an initial value

$$x(t_0) = x_0 \quad (2)$$

represents the state of such a system at time  $t$ , given that the state (2) was attained at time  $t_0$ . The inclusion of random effects in the dynamical system leads to a number of modifications that can be made to the formulation of the initial value problem (1), (2) (Gard, 1988). The first, and simpler, case arises when the initial value is replaced by a random variable  $X_0$ . The second case arises when the deterministic function  $f(t, x)$  has random coefficients, i.e. it is replaced by a random function  $F(t, X, E)$ , where  $E$  is a stochastic process uncoupled with the solution process  $X$ . As a special case,  $f(t, x)$  may be replaced by a random function with a random inhomogenous part (i.e., additive noise), that is, it is replaced by a random function  $F(t, X) + E$ . Of course, a combination of these cases could hold.

The inclusion of random effects in differential equations leads to two distinct classes of equations, for which the random processes have differentiable and non-differentiable sample paths, respectively. If the random processes occurring in a differential equation (for example  $X$  and  $E$ ) are sufficiently regular, i.e. have differentiable sample paths, then the majority of problems can be

<sup>2</sup>We use here Lebesgue integration, hence we assume the Lebesgue measure on the Lebesgue  $\sigma$ -algebra of Lebesgue measurable sets of  $T$ .

analyzed by use of methods which are analogous to those in deterministic theory of differential equations; such equations are called *random differential equations*. The second class occurs when the inhomogenous part is an irregular stochastic process such as Gaussian white noise. The equations are then written symbolically as stochastic differentials, but are interpreted as integral equations with Ito or Stratonovich stochastic integrals. These differential equations are called *stochastic differential equations*. In this paper, we will focus on random differential equations.

#### 3.1 Observational random dynamical models

We will define a random differential equation in terms of an observational random dynamical model:

**Definition 1.** An observational random dynamical model (oRDM) is a tuple

$$\mathcal{R}^o := \langle T, \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, \mathbf{F}, \mathbf{E} \rangle$$

where

- $T \subseteq \mathbb{R}$  is a time interval,
- $\mathcal{I}$  is a finite index set of endogenous processes,
- $\mathcal{J}$  is a finite index set of exogenous processes,
- $\mathcal{X} = \prod_{i \in \mathcal{I}} \mathcal{X}_i$  is the product of the codomains of the endogenous processes, where each codomain  $\mathcal{X}_i = \mathbb{R}^{d_i}$ ,
- $\mathcal{E} = \prod_{j \in \mathcal{J}} \mathcal{E}_j$  is the product of the codomains of the exogenous processes, where each codomain  $\mathcal{E}_j = \mathbb{R}^{e_j}$ ,
- $\mathbf{F} : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{X}$  is a function that specifies the dynamics,
- $\mathbf{E} : T \times \Omega \rightarrow \mathcal{E}$  is an exogenous stochastic process.

The oRDM gives the observational random dynamics of the random dynamical system, without any intervention from outside. The random dynamics are described in terms of random differential equations (Bunke, 1972):

**Definition 2.** A stochastic process  $\mathbf{X} : T \times \Omega \rightarrow \mathcal{X}$  is a sample-path solution of the random differential equations (RDE) associated to  $\mathcal{R}^o$ ,

$$\dot{\mathbf{X}} = \mathbf{F}(\mathbf{X}, \mathbf{E}), \quad (3)$$

if the ordinary differential equations (ODE)

$$\dot{\mathbf{X}}(t, \omega) = \mathbf{F}(\mathbf{X}(t, \omega), \mathbf{E}(t, \omega))$$

are satisfied for almost all  $\omega \in \Omega$ .

Let  $t_0 \in T$  and let  $\mathbf{X}_0$  be a  $d$ -dimensional random variable, where  $d = \sum_{i \in \mathcal{I}} d_i$ , such that  $\mathbf{X}_{t_0} = \mathbf{X}_0$  a.s., then

$\mathbf{X}$  is called a sample-path solution of (3) with respect to the initial condition  $(\mathbf{X}_0, t_0)$ .

The sample-path solution of (3) with respect to the initial condition  $(\mathbf{X}_0, t_0)$  is called unique on  $T$  if for an arbitrary pair  $\mathbf{X}, \tilde{\mathbf{X}}$  of sample-path solutions with respect to the initial conditions  $(\mathbf{X}_0, t_0)$  we have  $\mathbf{X} \stackrel{T}{=} \tilde{\mathbf{X}}$ .

We associated an ordinary differential equation to any specific sample path  $\mathbf{X}_t(\omega)$ . The solutions of these ordinary differential equations are the sample paths of a stochastic process  $\mathbf{X}$ , which is the sample-path solution of the random differential equation.

In particular, an oRDM is a deterministic dynamical model, if the background probability space is  $(\{1\}, \{\emptyset, \{1\}\}, \mathbb{P}_{\{1\}})$ . In this setting, the associated RDE is just a single ODE.

**Example 1 (Damped coupled harmonic oscillator).** Consider the well-known damped coupled harmonic oscillator, consisting of a one-dimensional system of  $d$  point masses  $m_i \in \mathbb{R}$  ( $i = 1, \dots, d$ ) with positions  $Q_i$  and momenta  $P_i$ . They are coupled by springs, with spring constants  $k_i \in \mathbb{R}$  and equilibrium lengths  $\ell_i \in \mathbb{R}$  ( $i = 0, \dots, d$ ), under influence of friction with friction coefficients  $b_i \in \mathbb{R}$  and with fixed end-points  $Q_0 = 0$  and  $Q_{d+1} = L$  (see Figure 1).

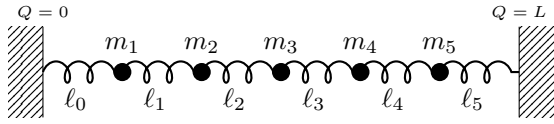


Figure 1: Damped coupled harmonic oscillator for  $d = 5$ .

The equations of motion of this system are provided by the ODE:

$$\begin{aligned} \dot{P}_i &= k_i(Q_{i+1} - Q_i - \ell_i) \\ &\quad - k_{i-1}(Q_i - Q_{i-1} - \ell_{i-1}) - \frac{b_i}{m_1} P_i \\ \dot{Q}_i &= P_i/m_i. \end{aligned}$$

Suppose that the lengths  $\ell_i$  are not constant, but are independent normally distributed around  $L/(d+1)$  with a certain variance. Then this ODE with random coefficients is actually a random differential equation modeled by an oRDM.

If the oRDM is sufficiently regular, then the majority of problems for such models can be analyzed by use of methods which are analogous to those in the theory of ordinary differential equations (Bunke, 1972; Sobczyk, 1991; Rupp and Neckel, 2013).

**Definition 3.** An oRDM  $\mathcal{R}^\circ$  is called regular if  $\mathbf{F} : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{X}$  is continuous and  $\mathbf{E}$  is sample-path continuous.

For ordinary differential equations a sufficient condition, for the existence and uniqueness of a solution with respect to an initial value, is the Lipschitz condition. Similarly, one can prove, by using results from the theory of ordinary differential equations, that there exists a similar sufficient condition for random differential equations of regular oRDMs.

**Theorem 1.** Consider a regular oRDM  $\mathcal{R}^\circ$ . If for almost all  $\omega \in \Omega$  there exists a continuous function  $L_\omega : T \rightarrow \mathbb{R}$  such that for each  $t \in T$  and  $x_1, x_2 \in \mathcal{X}$  the condition

$$\|\mathbf{F}(x_1, \mathbf{E}(t, \omega)) - \mathbf{F}(x_2, \mathbf{E}(t, \omega))\| \leq L_\omega(t) \|x_1 - x_2\|$$

is satisfied, where  $\|\cdot\|$  means the Euclidean norm in  $\mathcal{X}$ , then for any initial condition  $(\mathbf{X}_0, t_0)$  there exists a unique sample-path solution of the RDE (3) w.r.t.  $(\mathbf{X}_0, t_0)$ .

### 3.2 Intervened random dynamical models

Interventions on an observational random dynamical model can be modeled in different ways. Here we will consider interventions on the endogenous processes. We model an intervention on a subset  $K \subseteq \mathcal{I}$  of the endogenous processes by forcing those processes to be  $\eta_K : T \times \Omega \rightarrow \mathcal{X}_K$ . This can be seen as a ‘‘surgical’’ intervention, since they break the causal influences on the intervened processes (Eberhardt, 2014). The random dynamics of the other processes  $\mathcal{I} \setminus K$  are still untouched and are described in terms of the RDE associated to those processes, that is<sup>3</sup>

$$\dot{\mathbf{X}}_{\setminus K} = \mathbf{F}_{\setminus K}(\mathbf{X}_{\setminus K}, \eta_K, \mathbf{E}).$$

This yields the following random dynamical model including interventions.

**Definition 4.** A random dynamical model (RDM) is a tuple

$$\mathcal{R} := \langle T, \mathcal{I}, \mathcal{J}, K, \mathcal{X}, \mathcal{E}, \mathbf{F}_{\setminus K}, \eta_K, \mathbf{E} \rangle$$

where

- $T \subseteq \mathbb{R}$  is a time interval,
- $\mathcal{I}$  is a finite index set of endogenous processes,
- $\mathcal{J}$  is a finite index set of exogenous processes,
- $K \subseteq \mathcal{I}$  is a subset of intervened processes,
- $\mathcal{X} = \prod_{i \in \mathcal{I}} \mathcal{X}_i$  is the product of the codomains of the endogenous processes, where each codomain  $\mathcal{X}_i = \mathbb{R}^{d_i}$ ,
- $\mathcal{E} = \prod_{j \in \mathcal{J}} \mathcal{E}_j$  is the product of the codomains of the exogenous processes, where each codomain  $\mathcal{E}_j = \mathbb{R}^{e_j}$ ,

<sup>3</sup>For  $K \subseteq \mathcal{I}$ , we adopt the notation  $\setminus K$  for  $\mathcal{I} \setminus K$ .

- $F_{\setminus K} : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{X}_{\setminus K}$  is a function that specifies the dynamics of the  $\mathcal{I} \setminus K$  processes,
- $\eta_K : T \times \Omega \rightarrow \mathcal{X}_K$  is an intervened stochastic process,
- $E : T \times \Omega \rightarrow \mathcal{E}$  is an exogenous stochastic process.

If  $K = \emptyset$ , then we call  $\mathcal{R}$  also a non-intervened random dynamical model, otherwise we will call it an intervened random dynamical model.

The (intervened) random dynamical model gives the (intervened) random dynamics of the random dynamical system, where the random dynamics are described by the following set of equations:

**Definition 5.** A stochastic process  $\mathbf{X} : T \times \Omega \rightarrow \mathcal{X}$  is a sample-path solution of the (intervened) random differential equations associated to the (intervened) RDM  $\mathcal{R}$ ,

$$\begin{aligned} \dot{\mathbf{X}}_{\setminus K} &= F_{\setminus K}(\mathbf{X}, \mathbf{E}) \\ \mathbf{X}_K &= \eta_K, \end{aligned} \quad (4)$$

if  $\mathbf{X}_K \stackrel{T}{=} \eta_K$  and the ordinary differential equations

$$\dot{\mathbf{X}}_{\setminus K}(t, \omega) = F_{\setminus K}(\mathbf{X}(t, \omega), \mathbf{E}(t, \omega))$$

are satisfied for almost all  $\omega \in \Omega$ .

Let  $t_0 \in T$  and let  $\mathbf{X}_0$  be a  $d$ -dimensional random variable, where  $d = \sum_{i \in \mathcal{I}} d_i$ , such that  $\mathbf{X}_{t_0} = \mathbf{X}_0$  a.s., then  $\mathbf{X}$  is called a sample-path solution of the (intervened) RDE (4) with respect to the initial condition  $(\mathbf{X}_0, t_0)$ .

The sample-path solution of (4) with respect to the initial condition  $(\mathbf{X}_0, t_0)$  is called unique on  $T$  if for an arbitrary pair  $\mathbf{X}, \tilde{\mathbf{X}}$  of sample-path solutions with respect to the initial conditions  $(\mathbf{X}_0, t_0)$  we have  $\mathbf{X} \stackrel{T}{=} \tilde{\mathbf{X}}$ .

In particular, the non-intervened model  $\langle T, \mathcal{I}, \mathcal{J}, \emptyset, \mathcal{X}, \mathcal{E}, F, *, \mathbf{E} \rangle$ , where  $*$  is the terminal process  $* : T \times \Omega \rightarrow \{1\}$ , yields the same sample-path solutions as the observational random dynamical model  $\langle T, \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, F, \mathbf{E} \rangle$ . They describe the same random dynamics and in this sense the class of observational random dynamical models can be seen as a subclass of the class of random dynamical models.

**Definition 6.** We call an RDM  $\mathcal{R}$  linear, if the function  $F_{\setminus K} : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{X}_{\setminus K}$  is given by

$$F_{\setminus K}(\mathbf{x}, e) := \mathbf{B}\mathbf{x} + \Gamma e,$$

where  $\mathbf{B} \in \mathbb{R}^{K \times \mathcal{I}}$  and  $\Gamma \in \mathbb{R}^{K \times \mathcal{J}}$  are matrices.

The function  $F_{\setminus K}$  that defines the dynamics of the RDM encodes a functional structure that can be represented by a directed mixed graph.

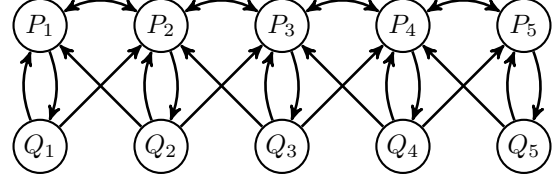


Figure 2: Functional graph of the RDM for the damped coupled harmonic oscillator of Example 1 for  $d = 5$ .

**Definition 7.** We define the functional graph  $\mathcal{G}(\mathcal{R})$  of an RDM  $\mathcal{R}$  as the directed mixed graph with nodes  $\mathcal{I}$ , directed edges  $i \rightarrow j$  if and only if  $i$  is a functional parent<sup>4</sup> of  $j$  w.r.t.  $F_{\setminus K}$  and bidirected edges  $i \leftrightarrow j$  if and only if there exists a  $k \in \mathcal{J}$  such that  $k$  is a functional parent of both  $i$  and  $j$  w.r.t.  $F_{\setminus K}$ .

For a linear RDM one would draw an edge  $i \rightarrow j$  if  $B_{ji}$  is non-zero and  $i \leftrightarrow j$  if both  $\Gamma_{ik}$  and  $\Gamma_{jk}$  are non-zero for some  $k$ .

The causal semantics of a random dynamical model can be modeled using interventions:

**Definition 8.** Given an RDM  $\mathcal{R} = \langle T, \mathcal{I}, \mathcal{J}, K, \mathcal{X}, \mathcal{E}, F_{\setminus K}, \eta_K, \mathbf{E} \rangle$ , a subset  $I \subseteq \mathcal{I}$  and a stochastic process  $\xi_I : T \times \Omega \rightarrow \mathcal{X}_I$ , the intervention  $\text{do}(I, \xi_I)$  maps  $\mathcal{R}$  to the intervened RDM  $\mathcal{R}_{\text{do}(I, \xi_I)} = \langle T, \mathcal{I}, \mathcal{J}, I \cup K, \mathcal{X}, \mathcal{E}, F_{\setminus (I \cup K)}, (\eta_{K \setminus I}, \xi_I), \mathbf{E} \rangle$ .

Note that interventions on disjoint subsets of the endogenous processes commute.

**Example 2 (Damped coupled harmonic oscillator).** Consider the damped coupled harmonic oscillator of Example 1. Its functional graph is depicted in Figure 2. We can perform an intervention on  $Q_i$  by moving the position of the  $i^{\text{th}}$  mass to a fixed position  $\xi_i^* \in \mathbb{R}$ . This is modeled by replacing the equation of motion of the  $i^{\text{th}}$  position by the process  $\xi_i$ , that defines the motion of moving the  $i^{\text{th}}$  mass to the fixed position  $\xi_i^* \in \mathbb{R}$ . Performing a similar intervention on the momentum  $P_i$  usually does not lead to an RDM with sample-path solutions that converge to a certain random variable.

We can define a regularity condition for the RDM  $\mathcal{R}$  similar to the one for oRDMs.

**Definition 9.** An RDM  $\mathcal{R}$  is called regular if  $F_{\setminus K} : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{X}_{\setminus K}$  is continuous and both  $\mathbf{E}$  and  $\eta_K$  are sample-path continuous.

The existence and uniqueness Theorem 1 generalizes to the RDM  $\mathcal{R}$ .

<sup>4</sup>Let  $\mathcal{X} := \prod_{i \in \mathcal{I}} \mathcal{X}_i$  and  $\mathcal{Y} := \prod_{j \in \mathcal{J}} \mathcal{Y}_j$  and consider a function  $f : \mathcal{X} \rightarrow \mathcal{Y}$ . We call an  $i \in \mathcal{I}$  a functional parent of  $j \in \mathcal{J}$  w.r.t.  $f$ , if there does not exist a function  $\tilde{f}_j : \mathcal{X}_{\setminus \{i\}} \rightarrow \mathcal{Y}_j$  such that  $\tilde{f}_j = f_j$ .

**Corollary 1.** Consider a regular RDM  $\mathcal{R}$ . If for almost all  $\omega \in \Omega$  there exists a continuous function  $L_\omega : T \rightarrow \mathbb{R}$  such that for each  $t \in T$  and  $x_1, x_2 \in \mathcal{X}_{\setminus I}$  the condition

$$\begin{aligned} & \|F_{\setminus K}(x_1, \eta_K(t, \omega), \mathbf{E}(t, \omega)) \\ & \quad - F_{\setminus K}(x_2, \eta_K(t, \omega), \mathbf{E}(t, \omega))\| \\ & \leq L_\omega(t) \|x_1 - x_2\| \end{aligned}$$

is satisfied, then for any initial condition  $(\mathbf{X}_0, t_0)$  such that  $(\mathbf{X}_0)_K = (\eta_K)_{t_0}$  a.s. there exists a unique sample-path solution of the RDE (4) w.r.t.  $(\mathbf{X}_0, t_0)$ .

Every linear RDM for which  $\mathbf{E}$  and  $\eta_K$  are sample-path continuous is regular. Moreover, there always exists an  $L_\omega \in \mathbb{R}$ , which is independent of  $\omega$ , such that the condition of Corollary 1 holds, hence for every regular linear RDM there exists a unique sample-path solution for any initial condition  $(\mathbf{X}_0, t_0)$  (Bunke, 1972).

### 3.3 Steady random dynamical models

Here we consider an important subclass of regular RDMs that satisfies certain convergence properties.

**Definition 10.** We call an RDM  $\mathcal{R}$  steady, if  $\mathcal{R}$  is regular,  $T = \mathbb{R}$ , the process  $\mathbf{E}$  converges to a random variable  $\mathbf{E}^*$  and the process  $\eta_K$  converges to a random variable  $\eta_K^* \in \mathcal{X}_K$ .

The class of steady RDMs is not stable under arbitrary interventions, that is a steady RDM  $\mathcal{R}$  does not have to stay steady under an intervention, however it is stable under the following class of interventions:

**Definition 11.** We call an intervention  $\text{do}(I, \xi_I)$  a perfect intervention if the process  $\xi_I$  converges to a random variable  $\xi_I^* \in \mathcal{X}_I$ .

Note that for any perfect intervention  $\text{do}(I, \xi_I)$ , the process  $\xi_I$  is sample-path continuous by definition.

Although steadiness of an RDM guarantees that the exogenous and intervened processes converge, it does, in general, not guarantee that any of its sample-path solutions converges. However:

**Definition 12.** Given a steady RDM  $\mathcal{R}$ . If a sample-path solution  $\mathbf{X}$  converges to a random variable  $\mathbf{X}^*$ , then we say that the sample-path solution equilibrates and we call  $\mathbf{X}^*$  an equilibrium variable of the sample-path solution  $\mathbf{X}$ .

If a sample-path solution  $\mathbf{X}$ , that describes the behaviour of the system, equilibrates, then in particular we have

$$\lim_{t \rightarrow \infty} \dot{\mathbf{X}}_t = \mathbf{0}.$$

almost surely.

## 4 Structural Causal Models

Structural causal models (SCMs), also known as structural equation models, provide a probabilistic description of the causal semantics of a system. They are widely used for causal modeling purposes (Pearl, 2009; Spirtes et al., 2000; Bollen, 1989). In this paper, we will follow the terminology of Bongers et al. (2016).

**Definition 13.** A structural causal model (SCM) is a tuple

$$\mathcal{M} := \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, \mathbf{f}, \mathbf{E} \rangle$$

where

- $\mathcal{I}$  is a finite index set of endogenous variables,
- $\mathcal{J}$  is a finite index set of exogenous variables,
- $\mathcal{X} = \prod_{i \in \mathcal{I}} \mathcal{X}_i$  is the product of the codomains of the endogenous variables, where each codomain  $\mathcal{X}_i = \mathbb{R}^{c_i}$ ,
- $\mathcal{E} = \prod_{j \in \mathcal{J}} \mathcal{E}_j$  is the product of the codomains of the exogenous variables, where each codomain  $\mathcal{E}_j = \mathbb{R}^{d_j}$ ,
- $\mathbf{f} : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{X}$  is a measurable function that specifies the causal mechanism,
- $\mathbf{E} : \Omega \rightarrow \mathcal{E}$  is a random variable.<sup>5</sup>

The solutions are described in terms of structural equations.

**Definition 14.** A random variable  $\mathbf{X} : \Omega \rightarrow \mathcal{X}$  is a solution of the SCM  $\mathcal{M}$  if the structural equations

$$\mathbf{X} = \mathbf{f}(\mathbf{X}, \mathbf{E})$$

are satisfied almost surely.

The causal mechanism encodes a functional structure that can be represented by a directed mixed graph.

**Definition 15.** We define the functional graph  $\mathcal{G}(\mathcal{M})$  of an SCM  $\mathcal{M}$  as the directed mixed graph with nodes  $\mathcal{I}$ , directed edges  $i \rightarrow j$  if and only if  $i$  is a functional parent of  $j$  w.r.t.  $\mathbf{f}$  and bidirected edges  $i \leftrightarrow j$  if and only if there exists a  $k \in \mathcal{J}$  such that  $k$  is a functional parent of both  $i$  and  $j$  w.r.t.  $\mathbf{f}$ .

### 4.1 Intervened structural causal models

The causal semantics of a structural causal model can be modeled using perfect interventions (Pearl, 2009).

<sup>5</sup>We slightly deviate from Bongers et al. (2016), where instead they take an exogenous probability measure on  $\mathcal{E}$ .

**Definition 16.** Given an SCM  $\mathcal{M}$ , a subset  $I \subseteq \mathcal{I}$  and an endogenous variable  $\xi_I : \Omega \rightarrow \mathcal{X}_I$ , the intervention  $\text{do}(I, \xi_I)$  maps  $\mathcal{M}$  to the intervened model  $\mathcal{M}_{\text{do}(I, \xi_I)} = \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, \tilde{\mathbf{f}}, \mathbb{P}_{\mathcal{E}} \rangle$  where the intervened causal mechanism  $\tilde{\mathbf{f}}$  is defined by:

$$\tilde{f}_i(\mathbf{x}, \mathbf{e}) := \begin{cases} \xi_i & i \in I \\ f_i(\mathbf{x}, \mathbf{e}) & i \in \mathcal{I} \setminus I. \end{cases}$$

We call an intervention  $\text{do}(I, \xi_I)$  a perfect intervention if  $\xi_I \in \mathcal{X}_I$ .

Note that interventions on disjoint subsets of endogenous variables commute.

## 5 From Steady RDMs to SCMs

We now have set the stage for constructing an SCM from an RDM under some convergence properties. Here, we will consider steady RDMs, as discussed in Section 3.3, for which the exogenous and intervened processes are well-behaved as time tends to infinity. For this class of RDMs we will see that the random differential equations, that determine the sample-path solutions of the RDM, play an analogous role to the structural equations, that determine the solutions of the SCM.

**Definition 17.** Given a steady RDM  $\mathcal{R} = \langle T, \mathcal{I}, \mathcal{J}, K, \mathcal{X}, \mathcal{E}, \mathbf{F}_{\setminus K}, \boldsymbol{\eta}_K, \mathbf{E} \rangle$ . Define the SCM  $\mathcal{M}_{\mathcal{R}}$  associated to  $\mathcal{R}$  to be  $\langle T, \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, \mathbf{f}^*, \mathbf{E}^* \rangle$  where the associated causal mechanism  $\mathbf{f}^* : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{X}$  is defined by

$$\mathbf{f}^*(\mathbf{x}, \mathbf{e}) := (\mathbf{x}_{\setminus K} + \mathbf{F}_{\setminus K}(\mathbf{x}, \mathbf{e}), \boldsymbol{\eta}_K^*)$$

with

$$\boldsymbol{\eta}_K^* := \lim_{t \rightarrow \infty} (\boldsymbol{\eta}_K)_t$$

and

$$\mathbf{E}^* := \lim_{t \rightarrow \infty} \mathbf{E}_t.$$

Note that the steadiness of  $\mathcal{R}$  implies the measurability of  $\mathbf{f}^*$ . This leads to our first main result:

**Theorem 2.** Given a steady RDM  $\mathcal{R}$ . If there exists a sample-path solution  $\mathbf{X}$  of  $\mathcal{R}$  that equilibrates to  $\mathbf{X}^*$ , then  $\mathbf{X}^*$  is a solution of the associated SCM  $\mathcal{M}_{\mathcal{R}}$ .

The converse does not hold in general, however we have the following sufficient condition:

**Proposition 1.** Consider a steady RDM  $\mathcal{R}$  such that  $\boldsymbol{\eta}_K \in \mathcal{X}_K$  (i.e.,  $\boldsymbol{\eta}_K$  is constant in time). If  $\mathbf{X}^*$  is a solution for the associated SCM  $\mathcal{M}_{\mathcal{R}}$ , then there exists a sample-path solution  $\mathbf{X}$  of  $\mathcal{R}$  that equilibrates to  $\mathbf{X}^*$ .

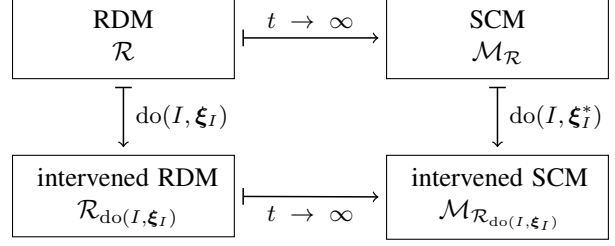


Figure 3: This diagram shows that perfect intervention commutes with the mapping from steady RDM to SCM as is made explicit in Theorem 3.

We can weaken the condition that  $\boldsymbol{\eta}_K$  has to be constant over time by imposing the following additional assumption on the model.

**Proposition 2.** Consider a steady RDM  $\mathcal{R}$  for which (i) there exists an  $\boldsymbol{\eta}_K^* \in \mathcal{X}_K$  and a  $t_0 \in T$  such that  $(\boldsymbol{\eta}_K)_t = \boldsymbol{\eta}_K^*$  for all  $t \geq t_0$  and (ii) for almost all  $\omega \in \Omega$  there exists a continuous function  $L_\omega : T \rightarrow \mathbb{R}$  such that for each  $t \in T$  and  $x_1, x_2 \in \mathcal{X}_{\setminus K}$  the condition

$$\begin{aligned} & \| \mathbf{F}_{\setminus K}(x_1, \boldsymbol{\eta}_K(t, \omega), \mathbf{E}(t, \omega)) \\ & \quad - \mathbf{F}_{\setminus K}(x_2, \boldsymbol{\eta}_K(t, \omega), \mathbf{E}(t, \omega)) \| \\ & \leq L_\omega(t) \|x_1 - x_2\| \end{aligned}$$

is satisfied. If  $\mathbf{X}^*$  is a solution for the associated SCM  $\mathcal{M}_{\mathcal{R}}$ , then there exists a unique sample-path solution  $\mathbf{X}$  of  $\mathcal{R}$  that equilibrates to  $\mathbf{X}^*$ .

Consider the diagram in Figure 3. So far, we have defined each mapping in this diagram separately (see Definition 8, 16 and 17). The next result shows that this diagram commutes:

**Theorem 3.** Given a steady RDM  $\mathcal{R} = \langle T, \mathcal{I}, \mathcal{J}, K, \mathcal{X}, \mathcal{E}, \mathbf{F}_{\setminus K}, \boldsymbol{\eta}_K, \mathbf{E} \rangle$ , a subset  $I \subseteq \mathcal{I}$  and a process  $\xi_I : T \times \Omega \rightarrow \mathcal{X}_I$  such that  $\boldsymbol{\eta}_K$  equilibrates to  $\boldsymbol{\eta}_K^* \in \mathcal{X}_K$  and  $\xi_I$  equilibrates to  $\xi_I^* \in \mathcal{X}_I$ . Then:

$$(\mathcal{M}_{\mathcal{R}})_{\text{do}(I, \xi_I^*)} = \mathcal{M}_{\mathcal{R}_{\text{do}(I, \xi_I)}}.$$

In other words, perfect intervention commutes with the mapping from steady RDM to SCM.

**Example 3.** Consider a linear RDM  $\mathcal{R} = \langle T, \mathcal{I}, \mathcal{J}, \emptyset, \mathcal{X}, \mathcal{E}, \mathbf{F}, *, \mathbf{E} \rangle$  where  $\mathbf{F}$  is of the form as in Definition 6 and  $\mathbf{E}$  is a random variable, that is a stochastic process that is constant in time. Then the associated SCM is  $\mathcal{M}_{\mathcal{R}} = \langle T, \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, \mathbf{f}^*, \mathbf{E} \rangle$  where the causal mechanism is defined by

$$\mathbf{f}^*(\mathbf{x}, \mathbf{e}) = \mathbf{A}\mathbf{x} + \boldsymbol{\Gamma}\mathbf{e}$$

where  $\mathbf{A} := \mathbf{I} + \mathbf{B}$ .

**Example 4 (Damped coupled harmonic oscillator).** Consider again the damped coupled harmonic oscillator of Example 1. The structural equations of the associated SCM are given by

$$\begin{aligned} P_i^* &= k_i(Q_{i+1}^* - Q_i^* - \ell_i) \\ &\quad - k_{i-1}(Q_i^* - Q_{i-1}^* - \ell_{i-1}) + (1 - \frac{b_i}{m_1})P_i^* \\ Q_i^* &= Q_i^* + P_i^*/m_i. \end{aligned}$$

These equations describe the equilibria of the positions and momenta. Figure 2 reflects the intuition that at equilibrium the position of each mass has a direct causal effect on the position of its neighbors. This can be seen more clearly by marginalizing over the momentum variables. Observing that the momentum variables always vanish at equilibrium, we can focus on the position variables as the variables of interest. We can marginalize over the momentum variables by solving each equation of  $P_i^*$  w.r.t. itself and then substituting these in the equations of  $Q_i^*$  (Bongers et al., 2016). This yields the marginal model with the following structural equations

$$Q_i^* = Q_i^* + \frac{k_i}{b_i}(Q_{i+1}^* - Q_i^* - \ell_i) - \frac{k_{i-1}}{b_i}(Q_i^* - Q_{i-1}^* - \ell_{i-1})$$

Resolving the self-loops of this marginal model by solving each equation w.r.t. itself gives the structural equations

$$Q_i^* = \frac{k_i(Q_{i+1}^* - \ell_i) + k_{i-1}(Q_{i-1}^* - \ell_{i-1})}{k_i + k_{i-1}}$$

and this model yields the same causal semantics for the position variables as the original model (Bongers et al., 2016). The functional graph associated to this model is depicted in Figure 4a. If we now perform a perfect intervention on  $Q_3$  by moving the 3<sup>th</sup> mass to a fixed position  $\xi_3^* \in \mathbb{R}$ , then we get the graph as depicted in Figure 4(b). Because these models are uniquely solvable and linear we can perform  $d$ -separation w.r.t. both graphs and conclude that  $Q_1^* \perp\!\!\!\perp Q_5^* | Q_3^*$  holds in the intervened model but not in the observational model (Forré and Mooij, 2017).

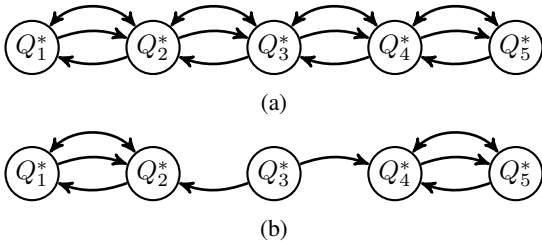


Figure 4: Functional graph of the marginal SCM associated to the damped coupled harmonic oscillator of Example 1 for  $d = 5$  after resolving the self-loops, under (a) no intervention and (b) perfect intervention on  $Q_3^*$ .

This example demonstrates that the equilibrium variables of the RDM can be studied by statistical tools applicable to SCMs. This sheds some new light on the concept of causality as expressed within the framework of structural causal models.

## 6 Application: Chemical Kinetics

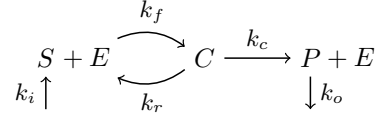


Figure 6: Basic enzyme reaction

Chemical kinetics is the study of rates of chemical processes. The chemical processes are described by the chemical reactions which are often modeled through ordinary differential equations. A well-known chemical reaction is the basic enzyme reaction which is schematically represented in Figure 6 (Murray, 2002).

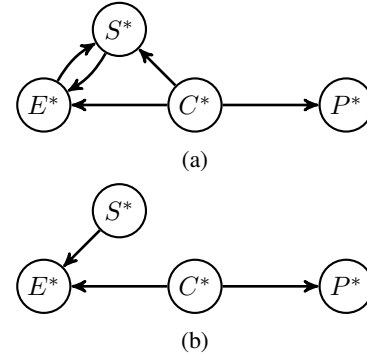


Figure 7: The functional graph of the SCM associated to the basic enzyme reaction under (a) perfect intervention on  $C^*$  and (b) perfect intervention on  $C^*$  and  $S^*$ .

It describes an enzyme  $E$ , binding to a substrate  $S$ , to form a complex  $C$ , which in turn releases a product  $P$  while regenerating the original enzyme. The  $k$ 's, called the rate constants, quantify the rate of a chemical reaction. These chemical reactions satisfy the law of mass action, which states that the rate of a reaction is proportional to the product of the concentrations of the reactants. Applying this to the concentration processes  $S$ ,  $E$ ,  $C$  and  $P$  of the basic enzyme reaction, gives the RDE:

$$\begin{aligned} \dot{S} &= k_i - k_f ES + k_r C \\ \dot{E} &= -k_f ES + (k_r + k_c) C \\ \dot{C} &= k_f ES - (k_r + k_c) C \\ \dot{P} &= k_c C - k_o P. \end{aligned}$$



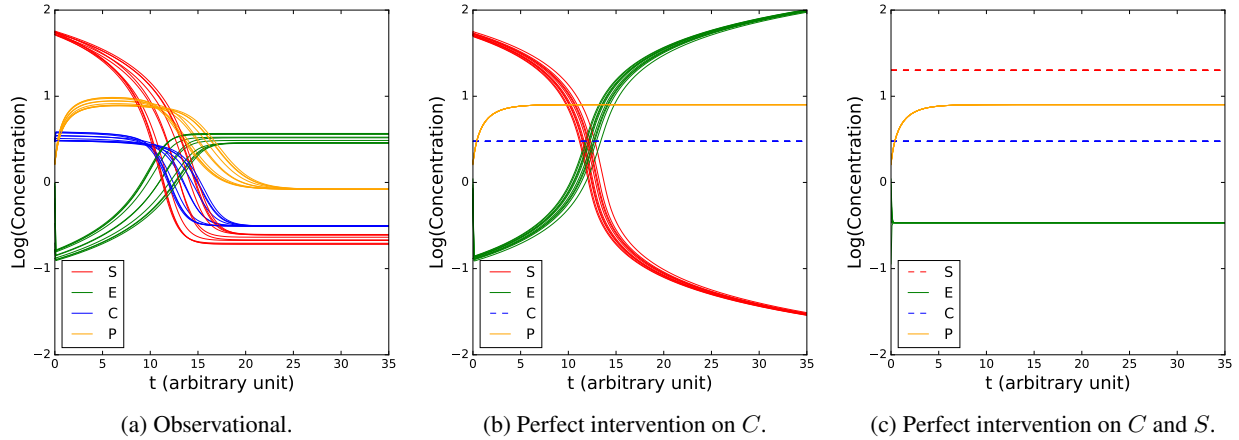


Figure 5: Simulation of the RDE associated to the basic enzyme reaction with random initial conditions under different interventions.

Although this RDE has no random coefficients (or random inhomogenous part), randomness can enter the RDE via the initial conditions. In Figure 5a we simulated the RDE with rate constants  $(k_i, k_f, k_r, k_c, k_o) = (0.5, 1.1, 0.9, 1.6, 0.6)$  and random initial conditions. The randomness of the initial conditions evolves over time and is captured in the associated SCM at equilibrium. That is, they are described by the associated SCM:

$$\begin{aligned}
 S^* &= k_i k_f^{-1} E^{*-1} - k_r k_f^{-1} E^{*-1} C^* \\
 E^* &= k_f^{-1} (k_r + k_c) S^{*-1} C^* \\
 C^* &= k_f (k_r + k_c)^{-1} E^* S^* \\
 P^* &= k_c k_o^{-1} C^*,
 \end{aligned}$$

where we removed the self-loops for convenience. This is an example of an SCM that is not uniquely solvable, which is illustrated in Figure 5a by the dispersion of the concentration  $S$  and  $E$  at large  $t$ , hence this example cannot be treated with the theory of Mooij et al. (2013) which assumes no dependence on initial conditions.

Let us for the moment fix the concentration of the complex by performing a perfect intervention on  $C$  as illustrated in Figure 5b. From the functional graph of the associated intervened SCM in Figure 7a we can read off that performing another perfect intervention on the substrate  $S$  should have no effect on the product  $P$ , as it would lead to the functional graph in Figure 7b where there is no directed path from  $S$  to  $P$ . This prediction, based on the functional graph of the SCM associated to the RDM, is indeed verified by the simulations in Figures 5b–5c. Intuitively, this is also what one would expect, since the complex is the only element in the system that is capable of releasing the product.

This illustrates that random differential equations are capable of modeling randomness through the initial con-

ditions, while the causal semantics at equilibrium of the dynamical system are parsimoniously described by the associated SCM.

## 7 Discussion

In this paper we built a bridge between the world of random differential equations and the world of structural causal models. This allows us to study a plethora of physical and engineering systems subject to time-varying random disturbances within the framework of structural causal models. We naturally extend the work of Mooij et al. (2013) to the stochastic setting, which allows us to address both cycles and confounders. In particular, we relaxed the condition that the dynamical system has to equilibrate to a single static equilibrium, and show that if an RDE is sufficiently regular all equilibrium sample-path solutions of the RDE are described by the solutions of the associated SCM, while pertaining the causal semantics.

There are two possible interesting directions for future research. The first is relaxing the regularity assumption. Earlier work has shown that SCMs can be derived from stochastic differential equations (Hansen and Sokol, 2014), however they restrict to the acyclic case. The second is relaxing the convergence assumption. Although the convergence assumption is a convenient and simplifying assumption, convergence of the stochastic processes is not always satisfied in practice. Recent work has shown that dynamic asymptotic behaviour of ordinary differential equations can be captured by dynamic structural causal models (Rubenstein et al., 2016). Other related work on discrete-time dynamical system and causality which does not require a single static equilibrium assumption is (Voortman et al., 2010).

## Acknowledgements

Stephan Bongers and Joris Mooij were supported by NWO, the Netherlands Organization for Scientific Research (VIDI grant 639.072.410).

## References

- Bollen, K. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons, New York, USA.
- Bongers, S., Peters, J., Schölkopf, B., and Mooij, J. M. (2016). Structural causal models: Cycles, marginalizations, exogenous reparametrizations and reductions. *arXiv.org preprint*, arXiv:1611.06221 [stat.ME].
- Bunke, H. (1972). *Gewöhnliche Differentialgleichungen mit Zufälligen parametern*. Akademie, Berlin, DE.
- Dash, D. (2005). Restructuring dynamic causal systems in equilibrium. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS 2005)*.
- Doob, J. (1953). *Stochastic Processes*. Wiley, New York, USA.
- Eberhardt, F. (2014). Direct causes and the trouble with soft interventions. *Erkenntnis*, 79(4):755–777.
- Fisher, F. M. (1970). A correspondence principle for simultaneous equation models. *Econometrica*, 38(1):73–92.
- Forré, P. and Mooij, J. M. (2017). Markov properties for graphical models with cycles and latent variables. *arXiv.org preprint*, arXiv:1710.08775 [math.ST].
- Gard, T. (1988). *Introduction to Stochastic Differential Equations*, volume 114 of *Monographs and textbooks in pure and applied mathematics*. Marcel Dekker, New York, USA.
- Hansen, N. and Sokol, A. (2014). Causal interpretation of stochastic differential equations. *Electron. J. Probab.*, 19.
- Hytinen, A., Eberhardt, F., and Hoyer, P. O. (2012). Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, 13(1):3387–3439.
- Iwasaki, Y. and Simon, H. A. (1994). Causality and model abstraction. *Artificial Intelligence*, 67(1):143–194.
- Lacerda, G., Spirtes, P. L., Ramsey, J., and Hoyer, P. O. (2008). Discovering cyclic causal models by independent components analysis. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI 2008)*.
- Loève, M. (1977). *Probability Theory II*, volume 46 of *Graduate Texts in Mathematics*. Springer, New York, USA, 4th edition.
- Mooij, J. M., Janzing, D., Heskes, T., and Schölkopf, B. (2011). On causal discovery with cyclic additive noise models. In *Advances in Neural Information Processing Systems (NIPS 2011)*, pages 639–647.
- Mooij, J. M., Janzing, D., and Schölkopf, B. (2013). From ordinary differential equations to structural causal models: the deterministic case. In Nicholson, A. and Smyth, P., editors, *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, pages 440–448. AUAI Press.
- Murray, J. (2002). *Mathematical Biology*. Springer, New York, USA, 3th edition.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, USA, 2nd edition.
- Rubenstein, P. K., Bongers, S., Mooij, J. M., and Schölkopf, B. (2016). From deterministic ODEs to dynamic structural causal models. *arXiv.org preprint*, arXiv:1608.08028 [cs.AI].
- Rupp, F. and Neckel, T. (2013). *Random Differential Equations in Scientific Computing*. Versita, London, GB.
- Sobczyk, K. (1991). *Stochastic Differential Equations*, volume 40 of *Mathematics and its Applications*. Springer.
- Soong, T. (1973). *Random Differential Equations in Science and Engineering*, volume 103 of *Mathematics in Science and Engineering*. Academic Press, New York, USA.
- Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. In *Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence (UAI 1995)*, pages 491–498.
- Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. The MIT Press, Cambridge, Massachusetts, 2nd edition.
- Voortman, M., Dash, D., and Druzdzel, M. J. (2010). Learning why things change: the difference-based causality learner. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010)*.

## Supplementary Material

### Proofs

#### Proof of Theorem 1

*Proof.* Continuity of  $F : \mathcal{X} \times \mathcal{E} \rightarrow \mathcal{X}$  and sample-path continuity of  $E : T \times \Omega \rightarrow \mathcal{E}$  implies that for almost all  $\omega \in \Omega$  the function  $F(x, E(t, \omega))$  is continuous on  $T \times \mathcal{X}$ . Moreover, continuity of  $F$  implies separate continuity of  $F$ . That is, for each  $x \in \mathcal{X}$  the function  $F(x, e)$  is continuous in  $e$  and in particular measurable in  $e$ . Hence for all  $(t, x) \in T \times \mathcal{X}$  the function  $F(x, E(t, \omega))$  is  $\mathcal{F}$ -measurable. Applying theorem 1.2 in Bunke (1972) proves the result.  $\square$

#### Proof of Corollary 1

*Proof.* Apply Theorem 1 to the regular oRDM  $\langle T, \setminus K, K \cup \mathcal{J}, \mathcal{X}_{\setminus K}, \mathcal{X}_K \times \mathcal{E}, F_{\setminus K}, (\eta_K, E) \rangle$ .  $\square$

#### Proof of Theorem 2

*Proof.* Let  $X$  be a sample-path solution that equilibrates to  $X^*$ . Then

$$\lim_{t \rightarrow \infty} \dot{X}_t = \mathbf{0} \quad \text{a.s.},$$

which gives

$$\begin{aligned} \mathbf{0} &= \lim_{t \rightarrow \infty} F_{\setminus K}(X_t, E_t) = F_{\setminus K}(\lim_{t \rightarrow \infty} X_t, \lim_{t \rightarrow \infty} E_t) \\ &= F_{\setminus K}(X^*, E^*) \quad \text{a.s.}, \end{aligned}$$

where we used continuity of  $F$  in the second equality, and steadiness in the last equality. This gives

$$\begin{aligned} X_{\setminus K}^* &= X_{\setminus K}^* + F_{\setminus K}(X^*, E^*) \\ X_K^* &= X_K^* \quad \text{a.s.}, \end{aligned}$$

and hence

$$X^* = f(X^*, E^*) \quad \text{a.s.}$$

$\square$

#### Proof of Proposition 1

*Proof.* Let  $X^*$  be a solution of  $\mathcal{M}_{\mathcal{R}}$ . Then the stochastic process  $X : T \times \Omega \rightarrow \mathcal{X}$  defined by  $X(t, \omega) := X^*$  is a sample-path solution of  $\mathcal{R}$  that equilibrates to  $X^*$ .  $\square$

#### Proof of Proposition 2

*Proof.* Let  $X^*$  be a solution of  $\mathcal{M}_{\mathcal{R}}$ . Then by Corollary 1 there exists a unique sample-path solution  $X$  w.r.t. the initial condition  $(X^*, t_0)$ . Hence  $X$  is the unique sample-path solution that equilibrates to  $X^*$ .  $\square$

#### Proof of Theorem 3

*Proof.* Applying the perfect intervention  $\text{do}(I, \xi_I^*)$  (by Definition 16) to  $\mathcal{M}_{\mathcal{R}}$  yields the SCM  $(\mathcal{M}_{\mathcal{R}})_{\text{do}(I, \xi_I^*)} := \langle \mathcal{I}, \mathcal{J}, \mathcal{X}, \mathcal{E}, f^*, E \rangle$  where  $f^*$  is defined by

$$f^*(x, e) := (x_{\setminus (I \cup K)} + F_{\setminus (I \cup K)}(x, e), \eta_{K \setminus I}^*, \xi_I^*).$$

Applying Definition 17 to the RDM  $\mathcal{R}_{\text{do}(I, \xi_I^*)} := \langle T, \mathcal{I}, \mathcal{J}, I \cup K, \mathcal{X}, \mathcal{E}, F_{\setminus (I \cup K)}, (\eta_{K \setminus I}, \xi_I), E \rangle$  yields the same SCM.  $\square$