Chapter 2

The overlap – a new approach for model discrimination

Introductory comments. In this chapter, the model–data–overlap concept for model discrimination is demonstrated directly after introducing some notation. The chapter is closed by an illustrative example. The differentiation from existing model discrimination concepts as well as the explanation of and justification for the overlap approach is given later in chapter 4, after reviewing existing concepts in model discrimination and parameter estimations in chapter 3.

Notation and setting. The type of models considered in this chapter are $D-$dimensional initial value problems, characterized by a set of ordinary differential equations (ODE) with parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_P)$ and initial values $\mathbf{y}_0$

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{y}(t) = f(\mathbf{y}(t); \boldsymbol{\theta}), \quad \mathbf{y}(t_0) = \mathbf{y}_0. \tag{2.1}$$

Their solutions at time t are denoted by $\Phi^t\,\mathbf{y}_0$. That model type is frequently used in chemical reaction kinetics, biokinetics, systems biology or polymerization processes and is going to resurface in chapter 5 and 6.

Assume that experimental data, associated to the system under consideration, is available at times $t = (t_1, \ldots, t_N)$. The measured data $\mathbf{d} = (\mathbf{d}(t_1), \ldots, \mathbf{d}(t_N))$ in general corresponds to measurements of some model sensors $\mathbf{d} = G(\mathbf{y})$. For simplicity, one may assume that $\mathbf{d}$ means some (if not all) of the components of $\mathbf{y}$. If the right side f of the ODE, the initial values $\mathbf{y}_0$ and the experimental data $\mathbf{d}$ are given, one additionally needs a concept of measuring the deviation between model and data, for example, by means of a functional

$$\mathcal{F}(\theta) \;=\; \text{deviation between}\,(\mathbf{d}(t_1), \ldots, \mathbf{d}(t_N)) \;\text{ and }\; \left(\Phi^{t_1}\mathbf{y}_0, \ldots, \Phi^{t_N}\mathbf{y}_0\right),$$

where a deviation can be understood in a very broad sense, for example, ranging from weighted residua to overlaps of probability distributions, as it will be illustrated in the following.

In order to distinguish between matrices, vectors and scalars, the following calligraphic notation is chosen. Matrices are abbreviated by capital and bold letters, like $\mathbf{M}$, vectors by small and bold letters, like $\mathbf{v}$, and scales by small and normal letters like n. When a component of a vector is abbreviated by some index, the letter appear in normal fonts.

Model variability. The model–data–overlap approach directly incorporates the sensitivity of the parameters as well as of the initial values in case of ordinary differential equations like (2.1). The expression *model variability* shall reflect this very ability of the model's trajectory to change by means of parameter as well as initial values perturbations. For reasons of simplicity, only model variability due to parameter sensitivity is considered in the following. By letting the initial values becomes parameters themselves, both scenarios can be unified.

As a consequence, one introduces a distribution π_θ governing the statistics of the parameters in subsequent realizations. In each single realization, θ is selected due to π_θ resulting in a single trajectory $\Phi_\theta^t \mathbf{y}_0$. Thus, the parameter distribution π_θ induces a distribution of trajectories $\Phi_\theta^t \mathbf{y}_0$ in the state space, developing simultaneously from the joint initial state $\mathbf{y}_0$. This model variability, denoted $\mathcal{M}_t$ in the following, is compared to the variability $\mathcal{D}_t$ of the measured data.

The model variability $\mathcal{M}_t$ is a positive measure defined as (c.f. [149])

$$
\begin{aligned}
\mathcal{M} &: \Gamma \times \mathbb{R} \to [0, a] \qquad a \in \mathbb{R}^+ \setminus \{0\} \\
\mathcal{M}_t(A) &= \frac{1}{C(t)} \int_\Theta \mathbf{1}_A(\Phi_\theta^t \mathbf{y}_0) \pi_\theta \mathrm{d}\,\theta
\end{aligned}
\tag{2.2}
$$

for any set $A \subset \Gamma$, where Γ denotes the entire state space, π_θ the parameter distribution. The characteristic function $\mathbf{1}_A(x)$ is given by

$$
\mathbf{1}_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases} \tag{2.3}
$$

In order to interpret (2.2) as a distribution later, one normalizes the model variability by a function $C(t) \in L^\infty$, which is specified later in section 4.2.

Through the construction in (2.2), one also attaches a stochastic interpretation to the purely deterministic setting of (2.1). The definition of $\mathcal{D}_t$ is analogously to that of $\mathcal{M}_t$, but realized by means of data variances. The specifications are also given later in section 4.2.

With these definitions, one can illustrate the model–data–overlap as seen in figure 2.1 and 2.2.

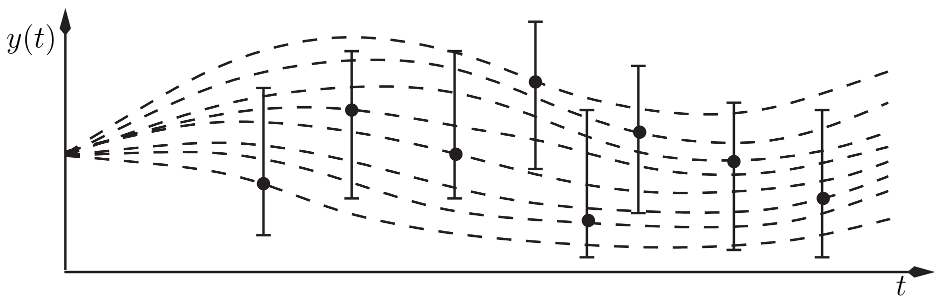

Figure 2.1: Model variability validates model–data–reproducibility: The black dots are measured data $d(t)$ with attached error bars representing some confidence interval of the data variability $\mathcal{D}_t$. Each measured data point can be explained by a single trajectory, representing a realization for θ from π_θ. Additionally, these trajectories also "pass" through confidence intervals of other data points and therefore validate the corresponding data also.

Model–data–overlap. Matching model and data variability reveals the information on the model–data–fit. In other words, this idea of assessment and validation allows for a new approach in model discrimination as the *overlap* of the model variability $\mathcal{M}_t$ and the data variability $\mathcal{D}_t$ describes the data–model–reproducibility. Having a pool of candidate models, one can discriminate between them by ranking and picking the one with the highest overlap value. As a result, one chooses the very model with the highest probability that the data distribution can be explained by the model variability with respect to its parameter distribution.

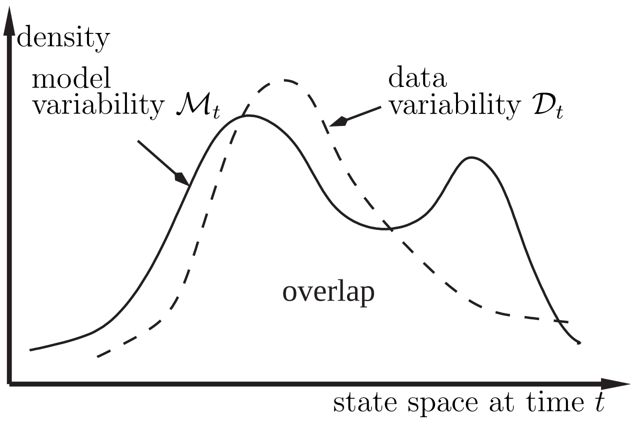

Figure 2.2: Overlap of model variability distribution $\mathcal{M}_t$ and data distribution $\mathcal{D}_t$ of measured data at a single time t.

Embedded parameter estimation. The overlap approach embeds the parameter estimation in the model validation and discrimination process. The parameters are chosen, more exactly the distributions π of parameters, such that the overlap between data and model $\mathcal{F}_{\mathcal{O}}$

$$\mathcal{F}_{\mathcal{O}} = \text{overlap of data and model variability} \tag{2.4}$$

is maximal.

Example. In the following, model discrimination by means of the overlap is demonstrated for a special class of systems (2.1), namely linear initial value problems,

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\,\mathbf{y}(t,\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})\,\mathbf{y}(t) + \mathbf{b}(\theta) \quad \text{with} \quad \mathbf{y}(0) = \mathbf{y}_0, \tag{2.5}$$

with the analytical solution

$$\Phi^t\,\mathbf{y}_0 = \exp\left(t\mathbf{J}(\boldsymbol{\theta})\right)\mathbf{y}_0 + \mathbf{J}(\boldsymbol{\theta})^{-1}\left(\exp\left(t\mathbf{J}(\boldsymbol{\theta})\right) - \mathbf{1}\right)\mathbf{b}(\boldsymbol{\theta}). \tag{2.6}$$

The two candidate models M_1

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} -2\theta & 2 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -6 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} x(t_0) \\ y(t_0) \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}, \tag{2.7}$$

and model M_2

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} -2 & 2\theta \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -6 \\ \theta \end{pmatrix}, \quad \begin{pmatrix} x(t_0) \\ y(t_0) \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}, \tag{2.8}$$

are to be discriminated. Both models coincide for $\theta = 1$ (see figure 2.3).

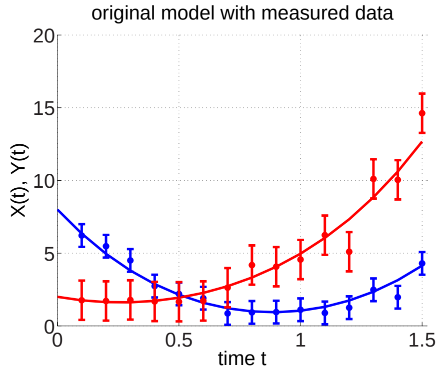

Figure 2.3: Model and data plot: For $\theta = 1$, the trajectories of M_1 and M_2 coincide. These trajectory values were taken and then perturbed to produce the data. The data, used for illustrations later, is symbolized by points with attached error bar.

As mentioned before, the parameter θ is interpreted as a distribution within the overlap concept. They can for example be modelled as a normal distribution with its hyperparameters $\theta_{\mathcal{O}}$ and $\Delta\theta_{\mathcal{O}}$, namely symbolizing the mean and the standard deviation:

$$\theta \sim \mathcal{N}\left(\theta_{\mathcal{O}}, \Delta\theta_{\mathcal{O}}\right).$$

Due to the different parameter–model–structures of M_1 and M_2 with respect to its parameter θ, the model variability might take different shapes during its time propagation.

Looking at figure 2.4, the following scenario is possiblewithin the overlap framework: To favor a model, where the distance between the model and data is larger but there model and data variabilities match better. On the contrary, there might be cases, where model and data coincide completely, but are rejected, since for example the model variability is much larger than the data one.

The phenomena described above results in a different model–data–reproducibility. The results of the overlap optimization for the model M_1 and M_2 are shown in table 2.1. Whereas the overlap in the y-component is roughly the same for both models, the x-component clearly favors the model M_1 to M_2.

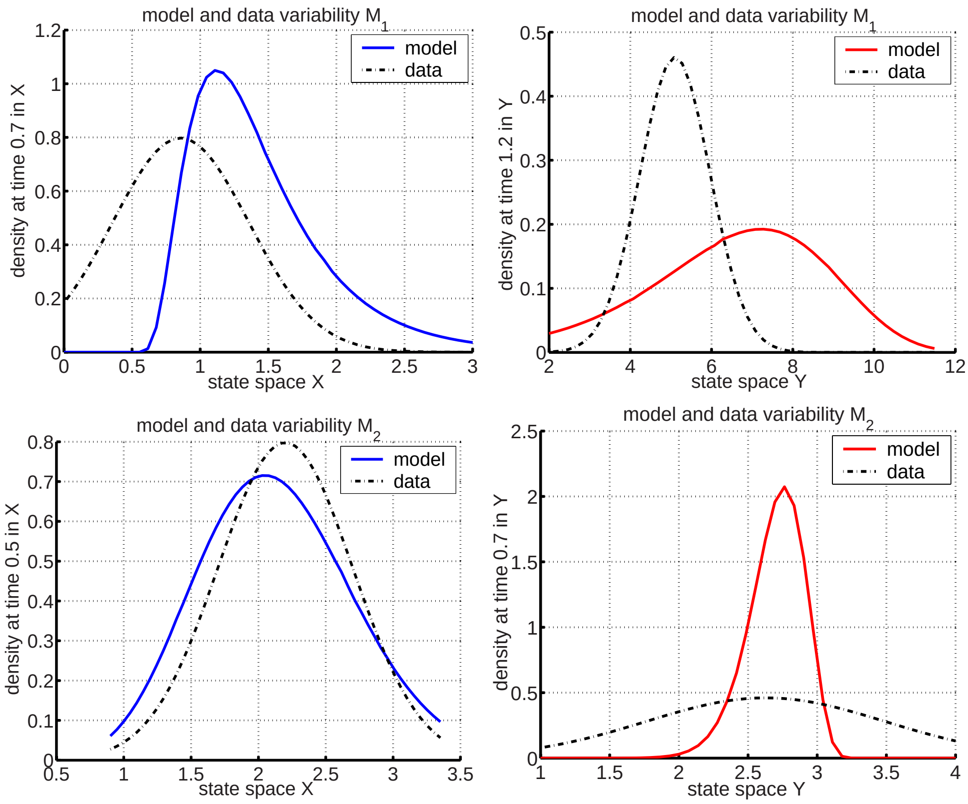

Figure 2.4: The shapes of the model variability can take different forms during a propagation, even though the initial parameter distribution is normal. Therefore, the overlap also incorporates linear and non–linear parameter sensitivity effects on the model.

	$\theta_{\mathcal{O}}$	$\Delta\theta_{\mathcal{O}}$	$\mathcal{F}_{\mathcal{O}}$ in x	$\mathcal{F}_{\mathcal{O}}$ in y	$\mathcal{F}_{\mathcal{O}}$ total
M_1	0.916	0.312	80.0 %	73.2 %	76.8 %
M_2	0.978	0.540	47.9 %	74.2 %	61.3 %

Table 2.1: results of overlap $\mathcal{F}_{\mathcal{O}}$ for models M_1 and M_2.

A more detailed analysis of the model along with a comparison of the overlap approach to traditional ones, is shown later in section 6.1, after reviewing other concepts and introducing implementation concepts.