Chapter 17
Prediction Uncertainty Estimation Despite Unidentifiability: An Overview of Recent Developments

Gunnar Cedersund

Abstract One of the most important properties of a mathematical model is the ability to make predictions: to predict that which has not yet been measured. Such predictions can sometimes be obtained from a simple simulation, but that requires that the parameters in the model are known from before. In biology, the parameters are usually both not known from before and not identifiable, i.e. the parameter values cannot be determined uniquely from available data. In such cases of unidentifiability, the space of acceptable parameters is large, often infinite in certain directions. For such large spaces, sampling-based approaches that try to characterize the entire space have difficulties. Recently, a new type of alternative approaches that circumvent this characterization problem has been proposed: where one only searches those directions in the space of acceptable parameters that are relevant for the uncertainty of a particular prediction. In this review chapter, these recently proposed methods are compared and contrasted, both regarding theoretical properties, and regarding user experience. The focus is on methods from the field of systems biology, but also methods from biostatistics, pharmacodynamics, and biochemometrics are discussed. The hope is that this review will increase the usefulness and understanding of already proposed methods, and thereby help foster a tradition where predictions only are deemed interesting if their uncertainties have been determined.

G. Cedersund (✉)
Department of Clinical and Experimental Medicine, IMT, Linköping University,
58185 Linköping, Sweden
e-mail: gunnar.cedersund@liu.se; gunnar.cedersund@gmail.com; +46(0)702-512323

G. Cedersund
Department of Biomedical Engineering, IMT, Linköping University,
58185 Linköping, Sweden

17.1 Introduction

Mathematical modelling has been a part of modern science already since its early foundation. In those early days, different hypotheses concerning the movements of the planets were incorporated in mathematical models, and the ability of these hypotheses to describe available data and to produce reliable predictions were used as the basis for choosing between the hypotheses. Those two components—rejections and predictions—are still the two corner stones of mathematical modelling as it now at last is getting more widely used also within biology (Fig. 17.1a). In this chapter we will deal with the latter of these two components, predictions. More specifically, we
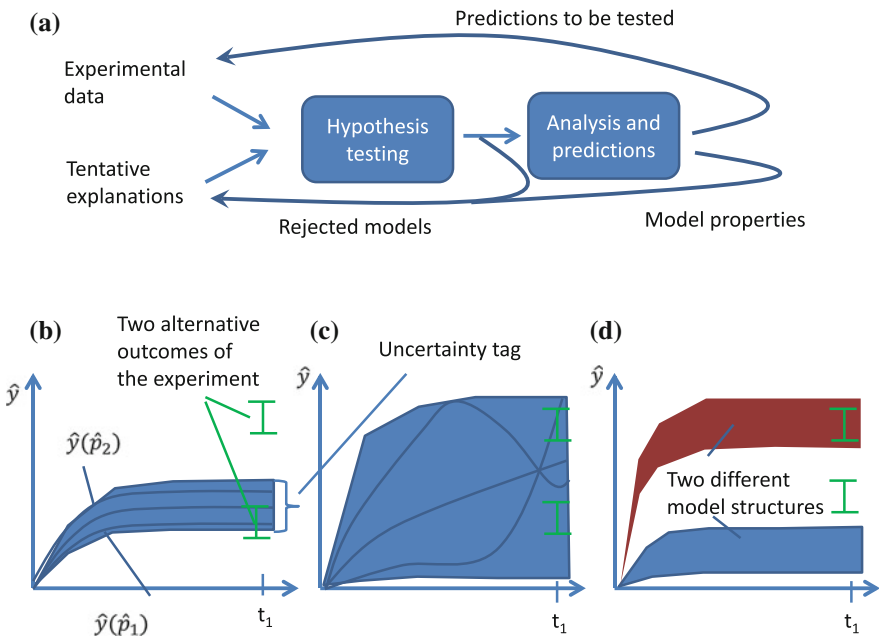
Fig. 17.1 a The two main steps of model-based data analysis: hypothesis testing and prediction analysis. The first step has been dealt with in previous chapters (e.g. [26]), and leads to rejections or tentative acceptances of models. The second step is dealt with in this chapter, and identified predictions feeds back to the design of new experiments, or in general to new knowledge regarding the original hypotheses. **b–d** If you know the uncertainty of the prediction, you can in before-hand guarantee that the experiment will give you something, independently of the outcome. In (**b**), the experiment is done to test the model. Only if the experimental data (*green*) lies within the prediction's uncertainty tag (*blue area*) will the model be accepted. In (**c**), the experiment is done to further determine the parameters: the more uncertain the prediction, the more the space of the acceptable parameters will be constrained by the data. In (**d**), the experiment is done to distinguish between the two models: independently of the experimental outcome, at the most one of the models will remain non-rejected after the new data point has been collected. The main result reviewed in this chapter regards how to go beyond characterization of all individual parameter trajectories (such as $\widehat{y}(\widehat{p}_1)$ and $\widehat{y}(\widehat{p}_2)$ in (**b**)) to methods that only looks for those parameters that give extreme predictions

will make an overview of recent advances in methods that allow us to produce useful predictions also in biology: predictions that take into account the known uncertainties in data and prior knowledge, and convert these to predictions with uncertainty tags (Fig. 17.1b). Importantly, these new methods work and may identify well-determined predictions also in the case of unidentifiable parameters.

Several of the herein presented methods are only a few years old, and it is therefore important to emphasise how important a development these methods are. This importance comes from the fact that the uncertainties in biology and medicine are so much bigger than corresponding uncertainties within physics, and that the classical methods for generating predictions with uncertainty tags therefore no longer are useful. These uncertainties are of at least two types: structural and parametrical. The structural uncertainties are bigger in biology, because there are more unknown details that make up each hypothesis. In physics, traditional hypotheses have been about the laws of nature, and e.g. whether these follow relativistic or Newtonian mechanics. In contrast, in biology, a hypothesis is often concerning the structure of a part of a biological network, and most meaningful hypotheses regarding such a network are in themselves containing many structural uncertainties. For instance, a hypothesis may be that a certain type of feedback generates a specific behaviour, such as an overshoot [4]. This hypothesis has many potential implementations, and is thus corresponding to many sets of equations, e.g. corresponding to different assumptions regarding the kinetic laws describing the involved reactions. The second type of uncertainty concerns the values of the parameters. In other words, even if one has made all structural decisions, e.g., concerning whether to use mass action or Michaelis-Menten rate law expressions, the values of the kinetic parameters that are to be used in those rate expression are still unknown. This again stands in stark contrast to the situation in physics, where most parameters are natural constants, which can be determined once and for all, and which already have been calculated prior to the modelling.

This second problem, that of uncertain parameters, is often referred to as unidentifiability, and it has been extensively studied in the literature. Identifiability is often considered to exist in two types: structural and practical. Structural identifiability has to do with the structure of the equations, and asks the question whether the parameters in principle can be obtained from those equations, if measurement uncertainty and unsufficient excitation not would be an issue. Although much beautiful theory has been produced for structural identifiability examinations, e.g. based on differential geometry [18, 25], this question is not too important for biological applications, where measurement uncertainty and insufficient excitation often are major problems. Therefore, the second type, practical unidentifiability analysis, which takes into account the specific data set at hand, is much more relevant. Much early analysis of practical identifiability has been concerned with the eigenvalues of the Hessian (i.e. the second-derivative) of the cost function. If the predicted output is linearly dependent on the parameters, this Hessian can be proportional to the inverse of the covariance matrix of the parameter uncertainties [17]. Unfortunately, for most biological problems, the assumption of linearity is not fulfilled, and the Hessian then only gives a local measure of parameters, at a probably non-unique point in the parameter space. The result of such a Hessian-based analysis is therefore hard to use.

For this reason, methods that make global assessments are more interest. There are attempts to produce global versions of the Hessian approach (e.g. [24]), but these do not solve the fundamental problem of linearity. A more promising approach, which also is intuitive and easy to implement, is the method based on the profile of the likelihood, simply called profile likelihood (PLH). This approach has been known for decades [21], but has not been used within the systems biology community until recently [23].

Several of the chapters in this book [11] have dealt with various ways to identify and handle uncertainty (e.g. [14]), but this is the only one that exclusively deals with predictions. Predictions are important to characterise, with uncertainty tags, for a wide variety of reasons. First and most importantly, predictions are what feeds back to the experiments, and therefore closes the experimental/theory loop (Fig. 17.1a), which produces an ever-evolving understanding of the system. Here it should be stated that both well-determined predictions (herein called core predictions) and poorly determined predictions (herein called suggestions or beach statements) are of interest: core predictions allow you to test the quality of the model (Fig. 17.1b), and beach statements are important because measuring them will produce a big improvement in the overall well-determination of the parameters in the model (Fig. 17.1c). Note that an important special case of the first usage occurs when you have two different core predictions from two different models (Fig. 17.1d); then the experiment ensures that at least one of the model structures will be rejected, independently of the outcome of the experiment. Second, even those predictions that cannot, or will not, directly be tested experimentally are interesting because they do provide a deeper understanding of the studied model: what are the key properties that have to be fulfilled in this system in order for it to produce the data, how is the model actually producing the observed behaviour? Third, reliable predictions are useful in a medical context. If some specific prediction of the model has been shown to be useful as clinical markers, e.g. to correctly diagnose a patient, it is important that that prediction is obtained with a correct degree of certainty. Fourth, well-determined predictions in a model means that it is useful for interacting with other models, in a supermodel. Say for instance that the first model describes the internal dynamics of a specific organ, and that its input-output profile is well-determined from experimental data. Then this organ model can be used in a hierarchical model, incorporating several organs and their cross-talk, even though some of the internal predictions in the model may be uncertain [22]. Finally, note that prediction uncertainty is not the same as parameter uncertainty: all parameters may be undetermined in the model, even though many well-determined predictions exists. For all these reasons, methods that deals with prediction uncertainty, and that can handle the in biology common situation of unidentiability, are of utmost importance.

All in all, to find predictions with uncertainty is therefore an important subject, and it is therefore interesting to look at some of the recent developments that have been done within the field. The rest of the chapter is structured as follows. First some basic notations regarding ODEs are introduced. Second, the most straightforward and simple approaches are introduced, which simply collects all found parameters that seem acceptable. Third, the recent developments regarding methods based on profiling the likelihood and modified optimization problems are introduced, and the

different variants are contrasted and discussed. These new methods are then also compared to three other important related approaches: Bayesian methods, cluster Newton, and neutral parameters. The chapter ends with a summary, and with a discussion of what the price is of not bothering with prediction uncertainty.

17.2 Basic Notations

The following presentation is done in the framework of ordinary differential equations (ODEs), but most concepts and methods hold equally well for all systems for which there exists a predictor, and the ability to form a likelihood function. Now follow quite a few notations. An ODE-based model is henceforth described by the following notations:

$$\dot{x} = f(x, p_x, u) \tag{17.1a}$$
$$x(0) = x_0 \tag{17.1b}$$
$$\widehat{y} = g(x, p_x, p_y, u) \tag{17.1c}$$

where x are the states, usually describing the concentration or amount of various substances; where $\dot{x}$ represents the derivative of the states with respect to time, and where f is the non-linear smooth function used to calculate these derivatives; where p_x are the parameters used to calculate f, usually corresponding to kinetic rate constants; where u is the input, which may depend on time, and which is usually known and controlled by the experimentalist; where $x(0)$ contains the values of the states at time $t = 0$, and where these values are described by the parameters x_0; where $\widehat{y}$ are the simulated model outputs corresponding to the measured experimental signals; where g is a smooth nonlinear function; and where p_y are parameters only appearing in the measurement equations. Note that x, u and y may depend on t, but that the notation is dropped unless the time-dependence needs to be especially stated, as in Eq. (17.1b). Note that at this stage there are three types of parameters, with potentially unknown values, and that all unknown parameters are collected in the parameter p, i.e. currently

$$p = (p_x, x_0, p_y) \tag{17.2}$$

Finally, the term parameter requires some further comments, since its usage opens for two interpretations. In some cases, a parameter means a point in the parameter space; in such cases the terms parameter point or parameter value will typically be used. Similarly, to refer to an individual element p_i in the parameter vector p, the term individual parameter will typically be used.

The three equations (17.1) form a model structure, which is denoted $\mathcal{M}$, and this model structure is turned into a specific model, denoted $\mathcal{M}(p)$, if the parameters are set to specific values. Now, assume that the N measurements are collected in a set Z,

$$Z = \{y(t_i)\}_{i=1}^{N} \qquad (17.3)$$

where $y(t_i)$, describes the measurement vector at time t_i. Unless otherwise stated, the null hypothesis is that the considered model structure is true, i.e. that

$$y(t_i) = \widehat{y}(t_i, p^0) + \nu \quad \forall\, i = 1...N \qquad (17.4)$$

where p^0 denotes the"true" parameter values (which exists if the null hypothesis is true), and where ν is a noisy signal that follows some distribution denoted D. In this chapter, the term 'explanation' is used to describe a hypothesis that not only should produce correct predictions mathematically, but also do this using mechanisms that make sense biologically.

17.3 Simple Approaches to Identify Predictions with Uncertainty

The underlying reason why predictions are uncertain is that there are many parameter points that all correspond to an acceptable agreement with available data and prior knowledge, and therefore must be considered as non-rejected. The formal decision of whether a parameter is to be considered as rejected or not is taken using a cost function, denoted V(p), and a cut-off, denoted δ (aspects of choosing V(p) and δ are discussed in section "The cut-off level" below). In other words, if the set of non-rejected parameters is denoted $\mathcal{A}$, it can formally be defined as

$$\mathcal{A} := \{p : V(p) \leq \delta\} \qquad (17.5)$$

where ":=" means that the left hand side is defined by the right hand side. With this set $\mathcal{A}$ formally defined, a basic principle of identifying parameter uncertainties presents itself: simply find the parameter points within $\mathcal{A}$ that gives a maximal and minimal value of the considered prediction. This is the principle of frequentist approaches, and it is contrasted to the Bayesian approaches later in this chapter. In the most simple approaches, the search for maximal and minimal values of the considered prediction within $\mathcal{A}$ is simplified to the search for maximal and minimal values within an approximation of $\mathcal{A}$ (Step 1 in Fig. 17.2). This approximation can be obtained by traversing $\mathcal{A}$ in different ways, and then saving the encountered parameter points. In [4], this traversing was done using a modified optimization approach, which uses multiple simplexes instead of one, to more fully cover the parameter space. Various alternatives to ordinary optimization and parameter estimation approaches that can be used for this are reviewed e.g. in [8]. Another approach sometimes used is based on the conventional PLH [23]. As the profiles are traversed to search for parameter uncertainties, the parameter points are saved, and then analysed with respect to the considered prediction. However, the limitation of these approaches is that there is no guarantee that the most extreme predictions have been found. In contrast, such a
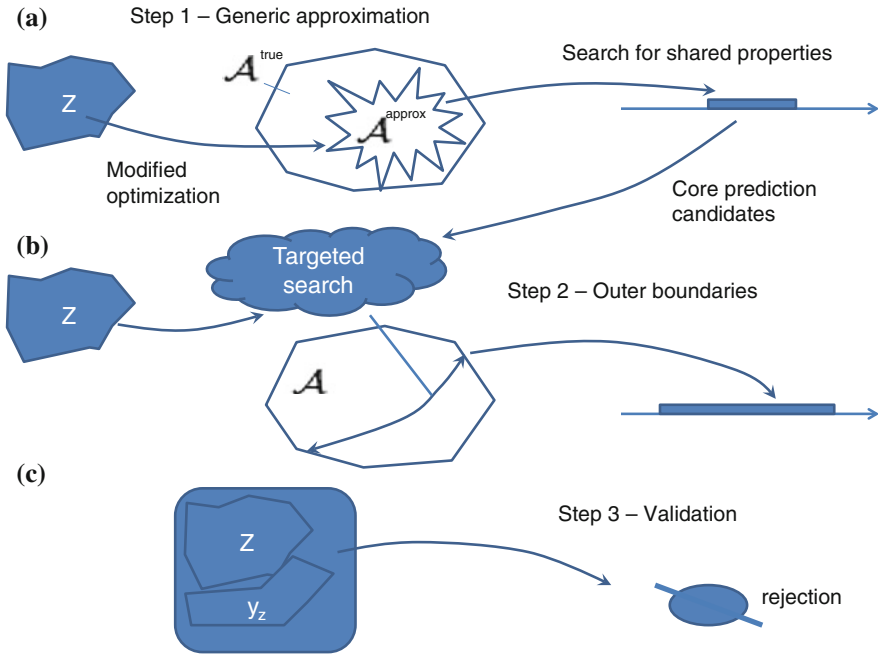
Fig. 17.2 The three main steps involved in prediction uncertainty assessments. **a** *Step 1* is simply to make a point-wise approximation of the true space of acceptable parameters, $\mathcal{A}$. In Bayesian approaches, this is the only step, and then the space is approximated so that the parameter points are most dense where the most likely parameters are. **b** *Step 2* lies at the heart of the most recent developments, which primarily are in focus in this chapter. On has then picked a specific prediction, probably because *Step 1* judged it to be well-determined, and then does a targeted approximation for that particular prediction. The most important things are to find the max and min values that the prediction can obtain, while still remaining in $\mathcal{A}$. **c** *Step 3* is in the core prediction approach simply a validation that the prediction rejection really holds, if the measured value lies outside the prediction uncertainty. In PPL, this step also involves assessments of profiles. Figure adopted from [6]

guarantee is obtained, at least in principle, by the newly developed approaches which modify the PLH to instead deal with prediction uncertainty, using various modified optimization problems.

17.4 PLH-Inspired Methods Based on Modified Optimization Problems

The now presented approach has recently been presented in different forms, based on different theoretical traditions, but ending up in similar but not identical methods. The following presentation is an attempt to bring these approaches together into a joint description, where similarities and differences more easily can be highlighted.

17.4.1 The Idea

Consider a specific prediction denoted z. This prediction can be anything that can be calculated using the given model structure and a specific parameter point, p. In other words, z must be a function that depends solely on the parameters, $z = z(p)$, but it can otherwise be anything of interest: e.g. a value of a state at a specific time-point, a value of a reaction rate at a specific time point, a single parameter value, or some other arbitrary function of these values, but that still depends on a given time point. The property z can sometimes also be considered as an entire time-series, but here we have encountered the first difference: that the property can only be a time-series in one of the settings, that based on core predictions.

In fact, the basic idea behind the core-prediction based approach presented in [7] was presented in the more general sense of z being a time-series. Nevertheless, when the full 3-step implementation of the core prediction approach was presented for specific examples in [6], the complications of studying a time-series was not considered. The inclusion of time-series are mainly concerned with the issue of dependencies and statistical interpretations of the results, and this whole issue is further discussed in a separate section below, entitled "Dependencies and the handling of predictions of time-series". For now, we consider the more specific case of z being a single scalar valued function of parameters and states at a given time point, and consider the general idea as presented in [6]. This idea is given by the following equation

$$\Delta_z^{\max} = z(p_{\max}) \text{ where} \tag{17.6a}$$

$$p_{\max} = \underset{p}{\arg\max}\{z\} \text{ subject to } V(p) < \delta \tag{17.6b}$$

$$\Delta_z^{\min} = z(p_{\min}) \text{ where} \tag{17.6c}$$

$$p_{\min} = \underset{p}{\arg\min}\{z\} \text{ subject to } V(p) < \delta \tag{17.6d}$$

where $\Delta_z^{\max}$ is the upper boundary of the prediction, and $\Delta_z^{\min}$ is the lower boundary. In words, the idea behind this approach is thus simply the one already stated: to find the maximal and minimal values of z that exists within $\mathcal{A}$.

The basic idea behind the other approach is based on a theory known as prediction profile likelihood (PPL). The concept of PPL was first mentioned in 1956 [10], and the concept of prediction profile likelihood was first introduced in 1979 [20]. Since then, quite a few various alternatives to this approach have been introduced, as reviewed e.g. in [2]. A first version that is relevant to this discussion was uploaded on the public article database arXiv in [15], and then as a standard paper in [16]. The basic idea behind PPL is to extend the data series Z with a new data point, z^*, which is a measurement of the property z. Often the uncertainty of z^* is zero, and then the maximum likelihood approach simply boils down to maximizing the likelihood, or minimizing the cost V(p), while fulfilling $z(p) = z^*$. This additional data point, z^*,

then enters as an additional parameter, and one can fix this parameter to different values, while optimizing over the others, as in a traditional PLH approach.

Let us now see how these two ideas may be implemented in practice.

17.4.2 The Stepping

The core-prediction approach is formally a standard, albeit difficult, constrained optimization problem, and may therefore in principle be implemented using any method that can deal with such problems. However, since many of the most commonly used implementations of optimization algorithms cannot deal with such difficult constraints, the following simple implementation was suggested in [6]:

1. Initiate $w > 0$, $f_{incr} > 1$
2. $\widehat{p} = \arg_p \min V_C(p)$, where $V_C(p) = V(p) + w \cdot C(z(p))$
3. if $V(p) > \delta$, return $z(\widehat{p})$; else $w = w \cdot f_{incr}$, go to 2
4. *Output:* the maximum or minimum value that z can take while still agreeing with data

where $C(z)$ is chosen as a function that grows with z if you are seeking $\Delta_z^{\min}$ (e.g. $C(z) = z$ or $C(z) = \log(z)$) and where $C(z)$ is chosen as a diminishing function if you are seeking $\Delta_{prop}^{\max}$ (e.g. $C(z) = 1/z$ or $C(z) = -z$).

The central idea behind this little algorithm is that you increase the contribution from the maximisation (or minimisation) of the new cost function $C(\cdot)$ compared with the component from the old cost function $V(p)$, until the maximisation (or minimisation) cannot proceed without violating $V(p) < \delta$; the optimised parameter values where you stop would then in theory and for sufficiently small step-sizes, f_{incr}, lie at the border of the core prediction uncertainty, i.e. at $\Delta_z^{\max}$ (or $\Delta_z^{\min}$). In other words, if there is an upper boundary for the prediction z, it is because that property is observable (i.e. can be identified from data [6]), i.e. because there is a tradeoff for the model between fitting to the data and maximizing that value.

In the other approach, based on PPL, one formulates a similar but not identical constrained optimization problem as in (17.6)

$$\widehat{p} = \arg \min_p V(p) \ \text{ where } z(p) = z^* \qquad (17.7)$$

This constrained optimization problem corresponds to the case where the added data point, z^*, has zero uncertainty. In practice they solve this equation by optimizing

$$\widehat{p} = \arg \min_p V_C(p) \ \text{ where } V_C(p) = V(p) + \eta(z - z^*) \qquad (17.8)$$

where η is a Lagrange multiplier, which is increased until the resulting optimization yields a sufficient fulfillment of the constraint $z(p) = z^*$. Once that fulfillment is

achieved, z^* is increased (or decreased) until the maximum (or minimum) border, where $V(p) = \delta$, has been reached.

There are some subtle but crucial differences between these two approaches to the stepping. First, for the core predictions approach, the stepping is done with the Lagrange multiplier w, which should lead to corresponding steps in the property $z(p)$. In contrast, in PPL, the stepping is done with z^* directly, and steps in their Lagrange multiplier η are only taken to make $z(p) - z^*$ closer to zero. On the plus side for the core prediction approach, this difference means that the core prediction approach only needs to perform one constrained optimization problems, whereas the PPL approach solves many such problems. Since constrained optimizations are highly difficult, this may be a big advantage, and something that opens up for usage of more advanced global methods for the single constrained optimization problem (17.6). Nevertheless, the difficulty of the many constrained optimizations in PPL is relieved by the fact that they are similar, and that the end result of the previous optimization, for a slightly different value of z^*, probably is a good start guess to the new optimization, with the new value of z^*. However, in the experience of the author, the most important difference between the two approaches to stepping is another: that the core prediction approach seems to be more inviting to jumps between different qualitative behaviours, which means that incremental changes in w not always lead to incremental changes in $z(p)$. For this reason, if one wants highly resolved profiles, the PPL approach to stepping is probably to be preferred compared to the stepping done in the simple algorithm solving Eq. (17.6). Note, however, that these methods are rapidly evolving.

17.4.3 The Cut-Off Level

The cut-off level is another issue where there are similarities and differences between the two approaches. In short, using the core prediction approach, one can use any cut-off, which opens up for more pragmatic choices, but in PPL they use the same theory as for the parametric PPL, which provides for a stronger theoretical underpinning, but also makes the options fewer.

In other words, for the core prediction approach, one can use any choice of cost function, and any reasoning behind the choice of δ. Nevertheless, the standard choice for $V(p)$ is the traditional least square

$$V(p) = \sum_{t=0}^{N} \sum_{i} \frac{(y_i(t) - \widehat{y}_i(t, p))^2}{\sigma_i(t)} + \text{additional terms} \qquad (17.9)$$

where i sums over the different measurement signals, σ is the standard deviation of the measurement noise, and the additional terms are optional. If the additional terms are included, one no longer has the traditional least square, and its theoretically sound properties. Nevertheless, often such additional terms are anyway warranted

and useful to include, e.g. to aid the search, or to require subjectively important qualitative behaviours. The most common way to chose a cut-off between acceptable and rejected costs is probably to use the above $V(p)$ without additional terms, and then use the inverse of the cumulative chi-square distribution, where the degrees of freedom equals the number of data points. This is often referred to as the chi-square test. More advanced approaches compensates for the number of identifiable parameters, uses empirical distributions from bootstrapping, or other principles altogether, such as model discrimination or residual correlation-based tests. These cut-off methods, and more, are reviewed in the same paper that published the first version of the core prediction approach, [7].

For the PPL cut-off, the value is closely associated to the standard PLH approach. This means two things. First, the cost function must be equal to the likelihood function, which is maximized as described in earlier chapters. This maximum likelihood approach is the same as minimizing the chi-squre cost above (Eq. (17.9) without the additional terms), under the assumption of additive and normally distributed measurement noise. Second, the cut-off is chosen as a certain addition, δ_{PPL} to the cost at the optimal parameters, $V(\widehat{p})$. This addition, δ_{PPL}, is equal to the inverse of the cumulative chi-square distribution for 1 degree of freedom.

17.4.4 Validation of the Obtained Boundaries

Both of the new approaches have a variation step, or a voluntary final step, which is called something with the word validation in it, but again the two versions are only related and not exactly the same.

For the core prediction approach, the validation step is simply a check that the prior steps have not led to a misconclusion, and that the corresponding rejection really holds true. In other words, Steps 1 and 2 (Fig. 17.2) have led to a prediction with outer boundaries, and data values outside of these boundaries will probably lead to a rejection. However, even if the experimentally measured data point would lie outside of the obtained uncertainty, that is not a guarantee that the combined dataset, including the original data and the new data point, would lead to a rejection of the model, since that depends on the uncertainty of the new data point, and on how far away from the rejection boundary the model was based on the original data set. In other words, the question of whether the predicted rejection leads to an ultimate rejection needs to be tested, or validated. This validation step, Step 3, thus validates the rejction. This step was introduced already in [4], which also introduced the concept of core predictions, and then the step followed directly on Step 1, which made the validation even more important. Note that also here, the validation of a predicted rejection is made using standard rejection methods, such as the chi-square test, and that no new theory is needed.

For PPL, the validation step is more a variation of the original PPL. In this variation, the added data point z^* has an uncertainty, and is thus added on an equal footing to the other data points. In other words, no constrained optimization is needed to

ensure that $z(p) = z^*$. One instead simply optimizes over the normal parameters, obtains a cost, modifies z^*, does a new optimization, and repeats until an unacceptable agreement with the data is obtained. Note that this method of course has the limitation that it depends on what the uncertainty of z^* is assumed to be, and that the assumed uncertainty may differ from the actual uncertainty of a subsequently collected actual data point.

17.4.5 Dependencies and the Handling of Predictions of Time-Series

A final important difference between the two approaches is that already hinted at: how they relate to predictions of entire time-series.

In the core prediction approach, one can in principle expand the present framework to account for timeseries using only a minor modification, but one need to be careful about the interpretation of the results. The key modification that needs to be done is to introduce a distance measure between two time-series, so that the time-series can be projected down to a scalar, which can be added as a constraint. A standard way to introduce such a distance measure is to take the difference between the two time-series at each time point, and integrate over time. In other words, for two time-series denoted d_1 and d_2 their distance, denoted $D(d_1, d_2)$, is given by

$$D(d_1, d_2) := \int d_1(t) - d_2(t)dt \qquad (17.10)$$

With these definitions in place, consider a specific predicted time-series $z = d(\widehat{p})$. The max and min values of the uncertainty region around this time series is given by the same algorithm as before, where C(p) is a growing function of $D(d(p),z)$ for finding $\Delta_z^{\min}$, and a diminishing function for finding $\Delta_z^{\min}$. Note that this is a generalization of the original algorithm: if the time-series is collapsed to a single time-point, the time-series approach finds the same boundaries as the original algorithm. Note also that this time-series formulation was the first way this core prediction approach was introduced, and that [6] therefore only presents an implementation and testing of that approach, using a particular solution to the general constrained optimization problem. We will now turn to the PPL viewpoint of this generalization, and then discuss the differences in interpretations and possibilities.

When considering time-series in the PPL framework, statistics enters the picture in another way, which makes things more complicated. In PPL, the statistical measure enters in the calculation of the term δ_{PPL}, which for a single additional data point z^* is calculated as the inverse of the cumulative chi-square distribution, with one degree of freedom, and at the desired significance level. In other words, if the significance level is 0.05, the cut-off is $\delta = V(\widehat{p})+ \sim 3.8$, and the prediction should lie within the specified boundaries in 95 % of the cases. In [15], they tested this claim, by generating many artificial datasets, with different noise realizations, and saw that at

least for the therein tested example, the fraction of successful prediction uncertainties converges to a reasonable vicinity of the fraction predicted by the significance level. This theory comes from the standard PLH approach, and depends on the assumption that the individual parameters, p_i, are uncorrelated, and that the new additional data point is considered as such an individual parameter. This uncorrelation assumption would have problems if one would switch from a single data point to an entire time-series, comprising many or an infinite number of data points. This is the case because if a datapoint at a specific time-point is predicted to be high, neighbouring datapoints will also be predicted to be high. In other words, for time-series, the statistical basis for PPL breaks down, and they can no longer predict the cut-off using that theory. That problem does not disappear in the validation version of PPL.

Let us now compare to the statistical viewpoint of the core prediction approach. In that approach, the cut-off is decided *a priori*, using any statistical measure, such as the chi-square value. Then, with this specification of $\mathcal{A}$ in place, the remaining exercise is to find the extreme predictions, the max and the min, that lie within this space, and this is purely an optimization problem. In this viewpoint, there is no difference between looking at the most extreme timeseries or the most extreme time-point, because no assumptions of uncorrelation appears, as it does in the PPL approach. However, one needs to keep that in mind when interpreting a predicted time-series: it is the extreme time-series considered as a whole that is predicted. In other words, other acceptable time-series may contain values at specific time-points that are more extreme (higher or lower) than the most extreme time-series. This problem is to some extent illustrated in Fig. 17.1c. Finally, note that if one wants to use the prediction for rejections based on new data-points at various points in this predicted time-series, the validation step in the core prediction approach ensures that the rejection holds statistically: Step 3 is a standard rejection formulation, even though more than one data point has been added. Therefore, no rejections will erroneously be done because of this described problem with correlations.

17.5 Related Approaches

17.5.1 Bayesian Approaches

The most important and state-of-the-art alternative to those presented herein are probably those developed in a Bayesian setting (reviewed in e.g. [14]). These theories are in practice implemented using Markov Chain Monte Carlo (MCMC) simulations [3, 12, 26, 29]. In this setting, one does not exclusively characterise $\mathcal{A}$ and ignores the rest of the parameter space, but one instead gives all parameter points a probability $P(p|Z)$, which says how likely the parameter points are, given the available data. These probabilities can then be used to calculate the corresponding probability of general model properties $P(z|Z)$, which can then be combined with a cut-off to obtain an uncertainty of the prediction, with a corresponding significance. However, despite

the conceptual appeal of this approach, and the existence of rather general methods, there are important limitations. For instance, the method requires the specification of a prior distribution of the parameter values, even if one has no prior knowledge on these distributions; the choice of this prior will determine the outcome. The short-coming of this assumption in terms of achieving conclusive statements is discussed in the next section. The most important limitation, however, is that these methods only converge if the system is identifiable, or only mildly unidentifiable; such limitations do not apply to the herein presented approach. In cases of non-identifiability it appears that an important difference between the two approaches is revealed. In frequentist approaches such as those presented herein one considers the value of the cost function (or likelihood), whereas in MCMC approaches one looks at the density, and in cases of unidentifiability these do not necessarily coincide. To sort out what is the correct approach, likelihood or density, is an important task for future research comparing and combining frequentist and Bayesian approaches. Finally, just as Step 1 in the frequentist approaches, MCMC approaches suffer the problem of the curse of dimensionality, since they try to approximate a multi-dimensional distribution using points. This problem is circumvented using PPL or Step 2 in the core prediction approach. In fact, Step 2 is per definition better than MCMC at finding extreme points, since you can always use the outcome of MCMC as a start guess for the constrained optimization, which per definition will find something equally good or better.

17.5.2 Other Related Approaches and the Relation to Other Fields

A relatively extensive overview of related methods—such as interval analysis based methods [13, 28], sloppy modelling [5, 19], etc.—is available in [6]. Nevertheless, there are some important subsequent developments, and developments in other fields, which are not mentioned therein. The most important such method is PPL, and it has been described in detail above. One important neighboring research field is known as pharmacokinetics. This is a field that has done modelling of biological systems since the 70s, and they have a relatively well-developed methodological toolbox. For instance, as mentioned above, the method known as profile likelihood (for parameter uncertainty), has been used in the pharmacokinetics community for several decades, but has only recently been discovered in the systems biology community. Similarly, the problem of unidentifiability, and the need to consider multiple parameter sets, has been described also within the pharmacokinetics community, and methods have been proposed to characterize such sets. One such method is cluster Newton [1], where a cluster of parameter sets are considered together, used to approximate a surface, and where this surface then is used in the optimization. However, although cluster Newton has shown some strengths, e.g. regarding speed compared to certain other approaches, cluster Newton should primarily be considered as a new alternative to

Step 1 (Fig. 17.2), and there does not yet seem to exist a correspondence to Step 2 in the pharmacokinetics community. Another important neighboring community is known as chemometrics. Also they have now acknowledged the need to identify the well-determined predictions, and to account for the effect of unidentifiability. In the chemometrics community, the space of acceptable parameters is referred to as "neutral parameters", and there are recent papers (e.g. [27]), where they analyse this space to find the well-determined properties (the core predictions). A recent and related method to identify the well-determined properties has been published also in the systems biology community [9]. This method starts by the identification of subsets of parameters for which the so-called Hessian matrix has full rank, except for one parameter. Within this subset, one can then numerically determine the inter-relations among the parameters, from ordinary profile-likelihood plots. Finally, fits of simple analytical expressions to these numerically determined interrelations can semi-automatically give the expressions for the core predictions.

In summary, the field of handling unidentifiability is rapidly gaining in attention, not only in the field of systems biology but in several other related fields. However, there does not exist a correspondence to Step 2 in any other field; their methods are still only based on various ways of approximating the space of acceptable parameters (Step 1).

17.6 Summary and Discussion

In this chapter we have considered the important topic of obtaining predictions with uncertainty. The focus has been on frequentist approaches, which considers the value of the likelihood function, or some other cost function, as the prime decider of whether a parameter is acceptable or not. This decision is determined by a cut-off value, δ, and with this cut-off decided, the space of acceptable parameters $\mathcal{A}$ is defined. In this chapter, two recently presented approaches to analyse this space have been reviewed and contrasted: the core prediction approach, and PPL. In Step 1 of the core prediction approach, $\mathcal{A}$ is approximated through a point cloud, and the extreme values of the prediction z are approximated by the extreme values in this point-cloud. In Step 2, these approximate boundaries are improved upon for a specific prediction, by formulating the constrained optimization problem (17.6), which can be solved directly using an optimization algorithm that can handle nonlinear constraints, or by using the simple algorithm mentioned herein. As is shown in [6], Step 2 clearly improves upon the boundaries for a complicated model for insulin signalling. In the core-prediction approach, there is a final and optional Step 3, which validates that the potential rejection suggested by a new experimental data point really holds. In PPL, the new data point is added as a parameter, whereafter its uncertainty is characterised using a normal PLH. If the added data point contains an uncertainty, the PPL is referred to as validation PPL. The stepping and cut-off in PPL are different, and these differences have been explained. For instance, it seems like the stepping strategy described for PPL is more stable. Also, because of the differences in where

and how the statistical considerations enters the picture, the extension to time-series is handled differently in the two approaches. In the core prediction approach, Step 2 is merely an optimization exercise, and time-series can just as easily be optimized for, as individual data points. However, even though potential rejections are checked in Step 3 one should be careful with the interpretation, as it is only the time-series as a whole that is extreme, not necessarily individual time-points. In contrast, for PPL, time-series cannot be considered, since the theory behind the cut-off is based on the independence of the parameters. Finally, even though this chapter provides a merging of these two approaches, it will be an important challenge for future research to also merge and relate these methods with the three other related alternatives: cluster Newton, Bayesian, and neutral parameters.

17.6.1 What Is the Price of Considering Predictions Without Uncertainty?

To end this chapter, we want to remind the reader of the conceptual and epistemological breakthrough that it means to now have predictions with uncertainty, and of the high price that is paid by not considering this uncertainty. For many years, systems biology modelling has been done with guessed parameter values, or using parameters from the literature that are not appropriate, or based on models based on earlier guesses. If the parameters are wrong, and a single parameter point is considered, predictions from such a model is of limited, if any, interest. However, even if the parameter point could be right, i.e. it is not unrealistic, it is still uncertain, and only a single point in the space of acceptable parameter points. Since the size and uncertainty of $\mathcal{A}$ typically is big in biology, this means that the resulting prediction uncertainty ($\Delta_z^{\min}$, $\Delta_z^{\max}$) also could be big, and probably would be, at least for some predictions. If one does not know which predictions are well-determined, and which are not, one has to assume that all predictions are highly undetermined. Such undetermined predictions can only make statements of the character "it could be like this, or it could be in some other way". Such statements are in [6] defined as suggestions, or beach statements, and they are weak statements. However, if one knows that a certain prediction must lie within certain narrow boundaries, that prediction is of the same epistemological level as a rejection, since the prediction is an implicit rejection: if the prediction is not sufficiently fulfilled when tested experimentally, the model will be rejected. Since a rejection is the strongest epistemological statement available in science, a core prediction is a statement that is final This finality holds since the outer boundaries of a core prediction will not be refuted by the collection of more data as long as that data do not show that previous data was erroneous. Core predictions are therefore well-determined, and of the strongest possible epistemological character [6]. Finally, even if one knows that a prediction is highly uncertain, i.e. that each possible behaviour within that uncertainty just is a suggestion, that is still important knowledge. One reason for that is that such highly uncertain predictions are those

that most likely will be most beneficial to measure, if one wants to further determine the parameters in the model. In that way, predictions with uncertainty, but not predictions without uncertainty, feed back to the user in a powerful way, which allows you to know both how the model works, what you know and what you guess, and which allows you to close the experiment/modelling cycle in the most powerful way. When you do a new experiment, you can know exactly why what the new data will provide: a test of the entire model structure (Fig. 17.1b), a characterization of undetermined aspects of the model (Fig. 17.1c), or a discrimination between two competing model structures (Fig. 17.1d).

Conflict of Interest

The author declares that he has no conflict of interest.

References

1. Aoki, Y., Hayami, K., de Sterck, H., Konagaya, A.: Cluster Newton method for sampling multiple solutions of underdetermined inverse problems: applications to a parameter identification problem in pharmacokinetics. SIAM J. Sci. Comput. **36**(1), B14–B44 (2013)
2. Bjørnstad, JF.: (1990) Predictive likelihood: a review. Stat. Sci. **5**, 242–265
3. Box, G., Tiao, G.: (1973) Bayesian inference in statistical analysis, Wiley Online Library
4. Brännmark, C., Palmér, R., Glad, T., Cedersund, G., Strålfors, P.: Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework. J. Biol. Chem. **94**, 121–163 (2010)
5. Brown, K.S., Sethna, J.P.: Statistical mechanical approaches to models with many poorly known parameters. Phys. Rev. E Stat. Nonlinear Soft Matter Phys. **68**, 021904 (2003)
6. Cedersund, G.: Conclusions via unique predictions obtained despite unidentifiability—new definitions and a general method. FEBS J. **279**, 3513–3527 (2012)
7. Cedersund, G., Roll, J.: Systems biology: model based evaluation and comparison of potential explanations for given biological data. FEBS J. **276**, 22–903 (2009)
8. Cedersund, G., Samuelsson, O., Ball, G., Tegnér, J., Gomez-Cabrero, D.: Optimization in biology parameter estimation and the associated optimization problem. In: Geris, L., Gomez-Cabrero, D. (eds.) Uncertainty in Biology, A Computational Modeling Approach. Springer, Chem (2016, this volume)
9. Eisenberg, M.C., Hayashi, M.A.L.: Determining identifiable parameter combinations using subset profiling. Math. Biosci. **256**, 116–126 (2014)
10. Fisher, R.A.: Statistical Methods and Scientific Inference. Oliver and Boyd, London (1956)
11. Geris, L., Gomez-Cabrero, D.: Uncertainty in Biology, A Computational Modeling Approach. Springer, Chem (2016, this volume)
12. Geyer, C.: (1992) Practical Markov Chain Monte Carlo. Stat. Sci. 473–483
13. Jaulin, l., Kieffer, M., Didrit, O., Walter, E.: Applied interval analysis: with examples in parameter and state estimation, robust control and robotics. Springer, Heidelberg (2001)
14. Kirk, P., Silk, D., Stumpf, M.P.H.: Reverse Engineering under uncertainty. In: Geris, L., Gomez-Cabrero, D. (eds.) Uncertainty in Biology, A Computational Modeling Approach. Springer, Chem (2016, this volume)

15. Kreutz, C., Raue, A., Timmer, J.: (2012) Likelihood based observability analysis and confidence intervals for predictions of dynamic models. http://arxiv.org/abs/1107.0013
16. Kreutz, C., Raue, A., Timmer, J.: Likelihood based observability analysis and confidence intervals for predictions of dynamic models. BMC Syst. Biol. **6**, 120 (2012)
17. Ljung, L.: System Identification—Theory for the User, 2nd edn. PTR Prentice Hall, Upper Saddle River (1999)
18. Ljung, L., Glad, T.: On global identifiability of arbitrary model parameterization. Automatica **30**, 265–237 (1994)
19. Mannakee, BK., Ragsdale, AP., Transtrum, M., Gutenkunst, RN.: Sloppiness and the geometry of parameter space. In: Geris, L., Gomez-Cabrero, D. (eds.) Uncertainty in Biology, Springer, Heidelberg (2015)
20. Mathiasen, P.E.: Prediction functions. Scand. J. Stat. **6**, 1–21 (1979)
21. Meeker, W., Escobar, L.: Teaching about approximate confidence regions based on maximum likelihood estimation. Am. Stat. **49**, 48–53 (1995)
22. Nyman, E., Brännmark, C., Palmér, R., Brugård, J., Nyström, F.H., Strålfors, P., Cedersund, G.: A hierarchical whole-body modeling approach elucidates the link between in Vitro insulin signaling and in Vivo glucose homeostasis. J. Biol. Chem. **286**, 26028–26041 (2011)
23. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J.: Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics **25**, 1923–9 (2009)
24. Sahle, S., Mendes, P., Hoops, S., Kummer, U.: A new strategy for assessing sensitivities in biochemical models. Philos. Trans. A Math. phys. Eng. Sci. **366**, 3619–3631 (2008)
25. Sedoglavic, A.: A probabilistic algorithm to test local algebraic observability in polynomial time. J. Symbolic Comput. **33**, 735–755 (2002)
26. Sunnåker, M., Stelling, J.: Model extension and model selection. In: Geris, L., Gomez-Cabrero, D. (eds.) Uncertainty in Biology, A Computational Modeling Approach. Springer, Chem (2016, this volume)
27. Tafintseva, V., Tøndel, K., Ponosov, A., Martens, H.: J. Chemometr. **28**, 645–655 (2014)
28. Tucker, W.: Interval methods. In: Geris, L., Gomez-Cabrero, D. (eds.) Uncertainty in Biology, Springer, Heidelberg (2015)
29. Vanlier, J., Tiemann, C., Hilbers, P., van Riel, N.: An integrated strategy for prediction uncertainty analysis, Bioinformatics, In press (2012)