

INVARIANT IMAGE REPARAMETERISATION: A UNIFIED APPROACH TO STRUCTURAL AND PRACTICAL IDENTIFIABILITY AND MODEL REDUCTION

Oliver J. Maclaren¹, Ruanui Nicholson¹, Joel A. Trent¹, Joshua
Rottenberry², and Matthew J. Simpson^{2,3}

¹Department of Engineering Science, University of Auckland, Auckland
1142, New Zealand.

²School of Mathematical Sciences, Queensland University of Technology
(QUT), Brisbane, Australia.

³ARC Centre of Excellence for the Mathematical Analysis of Cellular
Systems, QUT, Brisbane, Australia.

Abstract

Parameter non-identifiability – both structural and practical – presents a fundamental challenge when using mathematical models to interpret experimental or observational data and perform reliable uncertainty quantification. This issue is particularly acute in highly complex, applied areas such as the life sciences or engineering, where determining appropriate model complexity is complicated by experimental constraints and noisy data. While several approaches exist for diagnosing and resolving parameter non-identifiability, including symbolic/algebraic methods, profile likelihood analysis, and sloppiness analysis, these approaches have distinct limitations and are rarely combined. We present a novel integrated approach called Invariant Image Reparameterisation (IIR) that combines key elements of these methods to carry out both structural and practical identifiability analysis. Our approach enables the discovery and ordering of common classes of practically identifiable nonlinear parameter combinations without

requiring symbolic computation. This goal is achieved through a sloppiness-inspired log coordinate transformation that enables the replacement of symbolic reparameterisation computations with numerical calculation at a single reference estimate, and an invariance condition determines when this local calculation holds globally. The parameter combinations determined by this method are also naturally ordered by degree of identifiability. Our analysis supports model reduction by replacing a practically non-identified model by a structurally-non-identified approximate model that can be further reparameterised in terms of identified parameters only. By treating parameter combinations determined by our approach as interest parameters within our established likelihood-based Profile-Wise Analysis framework, we enable a unified analysis of nonlinear practical identifiability, model reduction, and model-based prediction that incorporates uncertainty quantification in the form of likelihood profiles and confidence sets. To facilitate practical adoption, we provide a Julia library on GitHub demonstrating our methodology across diverse mathematical models spanning data science, physics, engineering, chemistry, and biology.

Short title: Invariant Image Reparameterisation

Keywords: Mathematical model; Identifiability; Parameter estimation; Uncertainty quantification; Model reduction; Prediction.

1 Introduction

Mechanistic mathematical models encoding specific assumed or conjectured mechanisms are widely used for interpreting experimental and observational data and, hence, learning about underlying causes of the observed phenomena. Mathematical models are routinely used to provide insight into a broad range of applications including engineering (e.g. material science [1, 2], fluid dynamics and transport [3, 4]), ecology and population biology [5], disease transmission dynamics [6] as well as chemical reactions [7] and econometrics [8], among many other application areas. Regardless of the application area, an essential challenge in this undertaking is determining whether a particular measurement design and corresponding datasets contain enough information to give unique or sufficiently precise model parameter estimates. Parameter estimates are of interest both as representations of underlying mechanisms and as an intermediate step in constructing further predictions beyond the sample data.

Establishing whether model parameters are uniquely determined is often referred to as *structural identifiability analysis* in the modelling literature [9, 10, 11, 12, 13, 14], and is also simply referred to as identifiability analysis in the statistical literature [15]. In contrast, learning whether sufficiently precise parameter estimates are possible when working with finite noisy observations corresponds to the notion of *practical identifiability* (or estimability) analysis [16, 15, 17, 18]. Parameter non-identifiability essentially means that several distinct combinations of parameters can replicate the same model outputs or observations, meaning that the mechanisms encoded in the mathematical model have not been uniquely determined by the observations [15, 19]. From an applications point of view, parameter non-identifiability can mean that different combinations of mechanisms in the mathematical model can explain the same data. The common situation of having parameter non-identifiability means that it is difficult to draw conclusions about which particular mechanisms explain observational phenomena.

When model parameters are not identifiable, we may wish to explore whether some particular combinations of parameters are nevertheless identified [15, 20, 21], as well as the related possibility of carrying out *model reduction* to simplify the model in some sense so that the resulting model is identified (i.e., has identifiable parameters) [12, 13]. We consider this task and its relation to identifiability analysis in the present work. A straightforward example of non-identifiability arises when we consider the spatial motion of invasion fronts of a biological population modelled by the well-known Fisher-Kolmogorov model [6]. In this reaction-diffusion model, individuals are assumed to undergo random migration with diffusivity D

and logistic proliferation with rate λ . Making the most straightforward observations of the long-time front velocity c enables us to estimate the product $D\lambda$ [22], since we have $c = 2\sqrt{D\lambda}$ for initial conditions with compact support [6, 23]. This data, however, provides no insight into estimates of D or λ individually since there are infinitely many choices of D and λ that give rise to the same long-time speed of invasion. This non-identifiability means that making observations of the travelling wave speed c does not provide specific mechanistic insight into the rate at which cells move or rate at which cell proliferate.

Many approaches have been used to determine the identifiability of parameter combinations in the presence of non-identifiability (whether practical or structural) of the complete model parameter set. In the literature the main methods include: (i) symbolic/algebraic structural identifiability methods [24, 20, 25]; (ii) profile likelihood practical identifiability methods [26, 27, 28, 29, 30, 31, 32, 33]; and, (iii) model sloppiness analysis [34, 35, 36, 37]. Each approach brings certain strengths and limitations. For example, symbolic methods excel at automatically suggesting identifiable parameter combinations [20, 21] but typically directly apply only to structural identifiability problems and idealised infinite noise-free data which means that the results are not always applicable to real-world scenarios with sparse, imperfect data. Furthermore, when dealing with e.g. continuous state-space models, these methods typically require the use of Taylor or lie derivative series to generate exhaustive summaries of the model output (see [20, 21] for unified summaries of symbolic/algebraic methods and the role of different methods of generating the required exhaustive summaries). Alternatively, statistical likelihood-based methods excel at handling practical identifiability and providing estimates of uncertainty [38] but usually require user choices of interest parameter combinations to analyse, which means that automation is difficult. Finally, sloppiness analysis [34, 35, 36, 37] suggests identifiable parameter combinations in the practical identifiability setting but typically uses inaccurate and non-parameterisation-invariant approximations of cost or likelihood functions for uncertainty estimates. Furthermore, the relationship between sloppiness and identifiability is not always clear, as sloppiness is a relative notion quantified in terms of ratios of eigenvalues (or singular values) and, e.g., we may have a sloppy yet identifiable model [10].

A lack of full identifiability in a model generates a need and desire to reduce or reparameterise models using identifiable parameter combinations. These tasks – determining identifiable combinations of model parameters and model reduction – are thus closely related. In applications of sloppiness and profile likelihood analysis to model reduction, users often simply set poorly-identified parameters to zero or some arbitrary value [39, 40, 41]. However, unless done carefully, this can lead to setting individually poorly-identified parameters to values

that jointly violate the requirements on possible values for the well-identified parameter combinations. For example, in the case of $c = 2\sqrt{D\lambda}$ discussed above, although D and λ are not identified from c , we cannot simply set one to e.g. zero unless this is consistent with what we know about c ; the parameters appearing in a given combination must still satisfy some relationships jointly. Symbolic methods aim to systematically provide a model reduction based on a lower-dimensional reparametrisation [20, 21]. However, this requires explicit symbolic expressions for the model and the reparameterised reduced model, which may not be feasible for many mathematical or simulation models, and only covers structural non-identifiability for infinite noise-free data.

In this work we propose a new approach that: (i) automatically suggests well-identified parameter combinations without symbolic computation; (ii) implies reparameterisations suited to model reduction that does not require an explicit reduced form model; and (iii) collectively sets poorly-identified parameters in the original model consistently with well-identified parameter combinations, and provides interval estimates for both parameters and predictions with good statistical guarantees in the non-ideal data (practical identifiability/statistical analysis) setting. To achieve this, we present an approach for discovering practically identifiable parameter combinations consistent with the symbolic/algebraic literature but using parameter transformations inspired by the sloppiness literature to convert the general symbolic problem to a numerical problem that can be solved at a single reference estimate. We then use these parameter combinations as the interest parameters in our recently developed *Profile-Wise Analysis* (PWA) framework for practical identifiability, uncertainty analysis and model reduction based on profile likelihood [19, 38]. For certain general classes of identified parameter combinations, we can recover the information provided by structural methods up to essentially arbitrary accuracy. Outside of this exact class of parameter combinations, we can still discover approximate parameter combinations that facilitate more efficient prediction and estimation.

2 Methods

Here, we outline the technical foundations of our ‘Invariant Image Reparameterisation’ approach, which builds on and extends the frameworks reviewed in [20], [21], [38], and indirectly [15, 42]. Though we develop the details here and in the supplemental material TODO, a simple, practical algorithm for implementation is provided at the end of the section. We apply these methods to examples in the Results section.

Our method consists of four key components that work together to determine identifiable (and nonidentifiable) parameter combinations: (1) an auxiliary mapping that connects model parameters to observable quantities, (2) a general reparameterisation framework, (3) a coordinate transformation that simplifies the analysis and enables numerical computation, and (4) a singular value decomposition (SVD) for implementing the reparameterisation. Once identifiable (and nonidentifiable) parameter combinations are established, they can serve as so-called target or interest parameters in our Profile-Wise Analysis (PWA) workflow [38]. The last step could also be replaced by alternative approaches such as Bayesian inference, which can also benefit from reparameterisation methods CITE TODO.

Before presenting each component of our approach in detail, we first provide some essential mathematical background and notation.

2.1 Terminology

Throughout this section, we work relatively informally with (assumed-to-be) smooth mappings between finite-dimensional spaces and so recall some standard terminology (see, e.g., [43]). We assume the terms injective/one-to-one and surjective/onto mapping are known. In addition, a mapping is called a *submersion* if its derivative has full rank or, equivalently, if its derivative is an onto mapping into its target space. A mapping is called an *immersion* if its derivative is injective/one-to-one. Additionally, the *observed Fisher information*, which we use extensively, measures the amount of information data carries about parameters and is defined as the negative Hessian of the log-likelihood function [44, 32, 45]. The term *identifiable* refers to parameters or parameter combinations that can be uniquely determined from ‘perfect data’. In the context of a model mapping parameters to data distributions, $\theta \mapsto p(y; \theta)$, this is equivalent to the mapping, or more generally an induced relation in terms of parameter combinations, being injective [15].

2.2 Model representation: the auxiliary mapping, exhaustive summaries, and data distribution parameters

We begin by representing the model structure through what we call [38] the auxiliary mapping ϕ , which connects model parameters to data distribution parameters that are assumed to be in one-to-one correspondence with the model data distribution. Though we use slightly different terminology, this is essentially identical to the ‘exhaustive summary’ approach re-

viewed by [20, 21]. This means we have a mapping of the form

$$\theta \mapsto \phi(\theta), \quad (1)$$

where θ are the mechanistic model parameters of the full model, along with a density function of the form

$$p(y; \phi), \quad (2)$$

where ϕ here provides a one-to-one parameterisation of possible density functions for observable data y . Together, these induce a probability model for data y given mechanistic model parameters θ :

$$p(y; \theta) = p(y; \phi(\theta)). \quad (3)$$

Here we are using the same symbol ϕ for both the mapping and the resulting data distribution parameter values. We assume that ϕ is known in terms of its role in parameterising data distributions, but the mapping $\phi(\theta)$ may only be known indirectly, e.g. through solving a differential equation or other mathematical model. For example, ϕ may represent the mean of a normal distribution, where this is, in turn, determined by solving a differential equation involving parameters θ . In the case where the data distribution parameters are formally infinite-dimensional, such as the solution to an ODE or PDE treated as the mean or related parameter of an observation model [46, 47], we consider ϕ to correspond to a finite-dimensional numerical approximation of the solution, such as a vector defined on a fine grid obtained from a numerical solver or simulation model. As we define our models in Julia, we can, relatively generally, evaluate any required derivatives using automatic differentiation/algorithmic differentiation at the source code level [48]. This is demonstrated in our two ODE examples. We thus avoid the need for alternative exhaustive summary methods such as Taylor series or generating series TODO REFs.

We also make use of the observed Fisher information (in various coordinates), which, in θ coordinates, is given by

$$\mathcal{J}(\theta) = -H_l(\theta; y) = -\nabla_\theta(D_\theta l(\theta; y))^T = -\frac{\partial^2 l(\theta; y)}{\partial \theta \partial \theta^T} \quad (4)$$

where $l(\theta; y)$ is the log-likelihood of the model, i.e. $\log p(y; \theta)$ treated as a function of θ for fixed data y . Here, $D_\theta f = \frac{\partial f}{\partial \theta^T}$ denotes the Jacobian derivative of a function f with respect to θ , $\nabla_\theta f = (D_\theta f)^T$ denotes the gradient, and $H_f = \frac{\partial^2 f}{\partial \theta \partial \theta^T}$ denotes the Hessian [49]. We also use subscripts to denote differentiation when convenient, e.g. f_θ or $f_{\theta\theta}$, etc.

A central component of our approach is the observation that the above basic auxiliary mapping can always be decomposed as

$$\phi(\theta) = \tilde{\phi}(\psi(\theta)), \quad (5)$$

where:

- θ represents the original model parameters;
- $\psi(\theta)$ represents identifiable parameter combinations;
- $\tilde{\phi}$ provides a one-to-one mapping from identifiable parameter combinations to data distribution parameters;

We use a ‘tilde’ to denote the mapping ϕ written in terms of the (potentially reduced) identifiable parameter combinations. This function can be thought of as a reduced model taking just the identified parameter combinations. Our decomposition is illustrated in Figure 1.

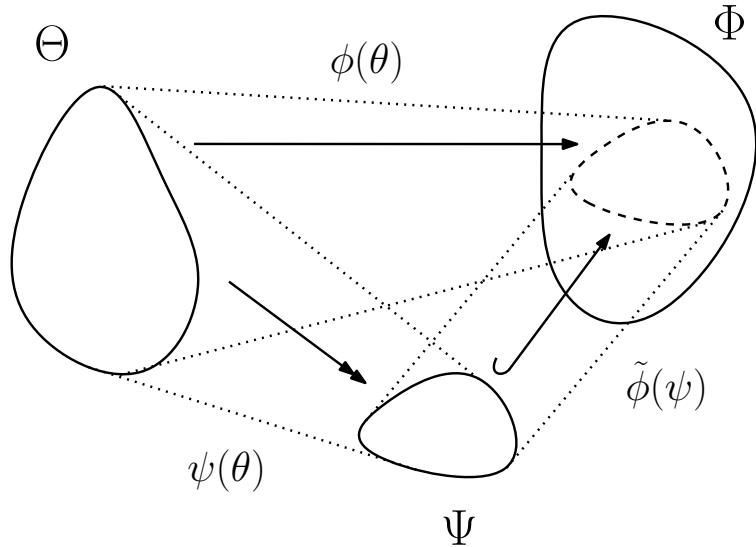


Figure 1: Illustration of the general decomposition of the auxiliary mapping into an identifiable reparameterisation followed by a reduced model mapping. Here θ represent the original model parameters, $\psi(\theta)$ represents identifiable parameter combinations, and $\tilde{\phi}$ provides a one-to-one mapping from identifiable parameter combinations to data distribution parameters. We use a double-headed arrow to indicate an onto mapping and an arrow with a hooked tail to represent a one-to-one mapping.

That such a decomposition is guaranteed to exist in principle, if not in practice, by the epimono factorisation property of the general category of sets and mappings [50], which states that any mapping $f : A \rightarrow B$ between sets A and B can be factored as the composition

$$f = m \circ e, \quad (6)$$

where $e : A \rightarrow C$ is onto (epimorphic), $m : C \rightarrow B$ is one-to-one (monomorphic), and C is some intermediate space, often called the *image*, $\text{im}(f)$, in the context of category theory. That is, we have a factorisation of the form $A \rightarrow \text{im}(f) \rightarrow B$. In standard set-theoretic definitions, the concrete subset of B determined as the output of $\text{im}(f) \rightarrow B$ is often referred to as the image; here the intermediate space represents the basic ‘abstract’ image and the concrete realisation in B is determined up to a one-to-one embedding of this abstract image. This abstract factorisation can be thought of as a nonlinear generalisation of the rank factorisation in linear algebra CITE TODO; as discussed below, these ideas are even more closely related in the context of smooth, constant rank mappings (see, e.g. TODO REF). Such a decomposition is also at the heart of the prior approach of [20] and [21], though they do not state the result in this form and where what we call data distribution parameters they call exhaustive summaries of the model. The main contribution of our approach is an alternative, numerical method for constructing such decompositions. We consider this next.

2.3 Symbolic/algebraic reparameterisation condition

Here we provide our own, though essentially equivalent, presentation of the reparameterisation ideas discussed in [20, 21] TODO CATCHPOLE ETC TOO, where we emphasise that this approach provides a practical implementation of the general epi-mono factorisation ideas considered above. In particular, we can make progress beyond the general epi-mono factorisation structure by further assuming constant rank of ϕ and sufficient smoothness on all mappings involved.

Concretely, we assume that the epi-mono composition structure can be constructed such that $\psi(\theta)$ is a smooth submersion onto some intermediate manifold for all θ , and so $D_\theta\psi(\theta)$ has maximal rank for all θ and is surjective, and that $\tilde{\phi}(\psi)$ is a smooth immersion into the space of data distribution parameters, and so $D_\psi\tilde{\phi}(\psi)$ is injective for all ψ [43]. As above, D_θ denotes the Jacobian derivative with respect to θ , and similarly for ψ and so on.

Practically, these assumptions mean that the epi-mono factorisation structure is assumed to transfer to the corresponding composition of the derivatives arising from the chain rule. That is, given the composition structure above, we have, by the chain rule,

$$D_\theta\phi(\theta) = D_\psi\tilde{\phi}(\psi(\theta))D_\theta\psi(\theta). \quad (7)$$

Under the above assumptions, the rank of $D_\psi\tilde{\phi}(\psi(\theta))$ and $D_\theta\phi(\theta)$ are the same as $D_\psi\tilde{\phi}(\psi(\theta))$ is one-to-one (see supplementary material TODO). Hence, when the overall ϕ mapping is

not one-to-one, we must have that the corresponding derivatives share the same null space (kernel), i.e.

$$\ker D_\theta \psi(\theta) = \ker D_\theta \phi(\theta). \quad (8)$$

Now, given explicit knowledge of $\phi(\theta)$ and $D_\theta \phi(\theta)$, this defines a differential equation to solve for $\psi(\theta)$. In particular, given null space vectors $\alpha(\theta)$ obtained from $D_\theta \phi(\theta)$, where $\alpha(\theta)$ are given as an explicit function of θ , we then solve

$$D_\theta \psi(\theta) \alpha(\theta) = 0, \quad (9)$$

for $\psi(\theta)$. Enforcing this for all and only the null-space vectors $\alpha(\theta)$, and requiring that the image of ψ is of dimension given by the rank of $D_\theta \phi(\theta)$, gives solution(s), not necessarily unique, for $\psi(\theta)$ defining an identifiable reparameterisation. This is equivalent to the conditions given by [20, 21] TODO CITE EARLIER WORK. It's also equivalent to equating the rows of $D_\theta \psi(\theta)$ to the vectors spanning the row space of $D_\theta \phi(\theta)$. The image of these vectors under ψ gives the intermediate ‘image’ component C of the epi-mono factorisation.

2.3.1 Fisher information formulation

The above is given in terms of the auxiliary mapping; as shown in detail in the supplementary material TODO REF, we can also work directly with the observed Fisher information, which provides an even closer link to practical identifiability concerns. This same point is discussed in [21]. Briefly, we assume that, as a function of the data distribution parameters, ϕ , with fixed data y , the log-likelihood $l(\phi; y)$ has a unique maximum in the interior of the mechanistically parameterised range, $\phi[\Theta]$, determined as the solution $\hat{\phi}$ to the first-order score equations $l_\phi(\phi; y) = 0$. We then call any mechanistic model parameter θ for which $\phi(\theta) = \hat{\phi}$ a maximum likelihood estimate of θ and denote it by $\hat{\theta}$, giving the following transformation law for the observed Fisher information:

$$\mathcal{J}(\hat{\theta}) = \phi_\theta(\hat{\theta})^T \mathcal{J}(\hat{\phi}) \phi_\theta(\hat{\theta}). \quad (10)$$

This transformation law of the observed Fisher information evaluated at the maximum likelihood is a well-known result $\phi(\theta)$ defines a one-to-one reparameterisation TODO REF but, as shown in the supplementary material, continues to hold even when ϕ is not one-to-one (as long as, essentially, the true data distribution can be obtained by some mechanistic model, i.e. $\phi_{\text{true}} \in \phi[\Theta]$).

Assuming the observed Fisher information, as considered in terms of data distribution parameters but not necessarily ‘mechanistic’ model parameters, is non-singular, the nullspace

and row space of the observed Fisher information are the same as that of the auxiliary mapping (supplementary material). This is a slightly stronger condition than assuming that the distribution parameters give a one-to-one indexing of the associated distributions but is a relatively weak assumption in general. We can thus alternatively determine the identifiable parameters by solving a nullspace constraint in terms of the Fisher information rather than the Jacobian derivative of the auxiliary mapping. We return to this below.

2.4 Parameter space linearisation and sloppiness

While the above condition provides a general approach to reparameterisation, it requires ‘symbolic’ or ‘algebraic’ knowledge of the Jacobian (or, equivalently, of the associated Fisher information) as a function of the mechanistic model parameter, and the null space we find is, hence, a function of the mechanistic model parameter. This is done by solving a partial differential equation. In contrast, numerical methods typically base analysis at a single point and only provide numerical values for null space vectors. This limits the reparameterisation information available to local reparameterisation or global but linear reparameterisations. Thus, numerical approaches are generally thought to be unable to provide meaningful non-linear reparameterisations of the type provided by symbolic methods [21]. However, here we show that we can apply an initial coordinate transformation to bridge these approaches. Importantly, as the likelihood function is invariant to reparameterisation [44, 32, 45], this approach loses no information; rather it aims to ‘present’ the information contained in the likelihood in a clearer form.

The parameter transformation we primarily focus on is the componentwise log transformation of all parameters, $\log \theta = [\log \theta_1, \log \theta_2, \dots, \log \theta_p]^T$. However, our basic approach applies if, e.g. only some components are logged or even if both logged and unlogged parameters are considered simultaneously. In the context where each parameter is included only once in either logged or unlogged form, we can consider the parameters that aren’t logged to be ‘pre-logged’ and the implicit ‘original’ parameters to be the exponentials of the logged parameters. The key assumption is that the coordinate transformation should be an immersion (i.e. have injective derivative). The motivation for focusing primarily on log transformations is that they are particularly natural for many physical and mechanistic models, where parameter combinations frequently appear as monomial dimensionless groups (so-called π terms) due to the Buckingham Pi theorem of dimensional analysis [51, 52]. Furthermore, dimensionless groups often play a key role in asymptotic approximations and model reduction [52] and here we are interested connections between identifiability and model reduction.

While log-parameter transformations are commonly used in the sloppiness literature REF TODO, and this inspired our present approach, the motivation for this choice is often not discussed in much detail in this literature beyond general reference to the idea that parameters in models typically span different orders of magnitude and that log transformations are related to working with dimensionless (and positive) parameters. Here, our motivation is explicitly that taking the log ‘linearises’ monomial terms in parameter space, though the overall function will still typically be a nonlinear function of the resulting linear combinations of log parameters. This turns out to be enough to reduce the reparameterisation condition to a single-point evaluation when the reduced parameter combinations are monomials.

Specifically, when parameters appear in the model in combinations of the form

$$\pi_i = \theta_1^{r_{i1}} \theta_2^{r_{i2}} \cdots \theta_p^{r_{ip}}, \quad (11)$$

where r_{ij} are typically integers in dimensional analysis, taking logarithms converts these nonlinear combinations into linear combinations:

$$\log(\pi_i) = r_{i1} \log(\theta_1) + r_{i2} \log(\theta_2) + \cdots + r_{ip} \log(\theta_p). \quad (12)$$

This means that if we define $\theta^* = \log(\theta)$ to be the *component-wise* log of the parameter vector, a model expressed in terms of parameter combinations such as dimensionless π groups can be written in the general form

$$\phi(\theta) = \tilde{\phi}(\exp(A\theta^*)), \quad (13)$$

where A is a matrix whose rows contain the coefficients r_{ij} and \exp is the exponential function, similarly applied componentwise. Here $\tilde{\phi}$ is the one-to-one mapping appearing in the mono-epi factorisation introduced above. Note that $\exp A$ generates a vector of parameter combinations, each in the form of a monomial in the original parameters. Furthermore, if A has a null space, then the resulting vector $\exp(A\theta^*)$ can be of reduced dimension compared to θ^* and θ . While the coefficients in A are traditionally restricted to integers in dimensional analysis, here we allow them to be arbitrary real numbers to facilitate numerical computation and provide more flexibility in discovering practically useful parameter combinations. We also do not require the combinations to be dimensionless *a priori*. These relaxations come at a potential cost to interpretability, which we address later by considering scaling and rounding operations that can convert these real-valued coefficients to integers while preserving key properties of the approximation.

Writing the auxiliary mapping in terms of log parameters, i.e. defining $\phi_*(\theta^*) = \phi(\theta)$ and $\psi_*(\theta^*) = \psi(\theta)$, where we use an upper asterisk to denote logged variables, $\theta^* = \log \theta$, and we

use a subscript asterisk to indicate functions of the logged variables, the basic factorisation becomes:

$$\phi_*(\theta^*) = \tilde{\phi}(\psi_*(\theta^*)) \quad (14)$$

This idea is illustrated by the alternative pathway highlighted in Figure 2. We show a generic parameter transformation a in the figure to emphasise that while log transformation is the typical case, the idea applies more generally in principle.

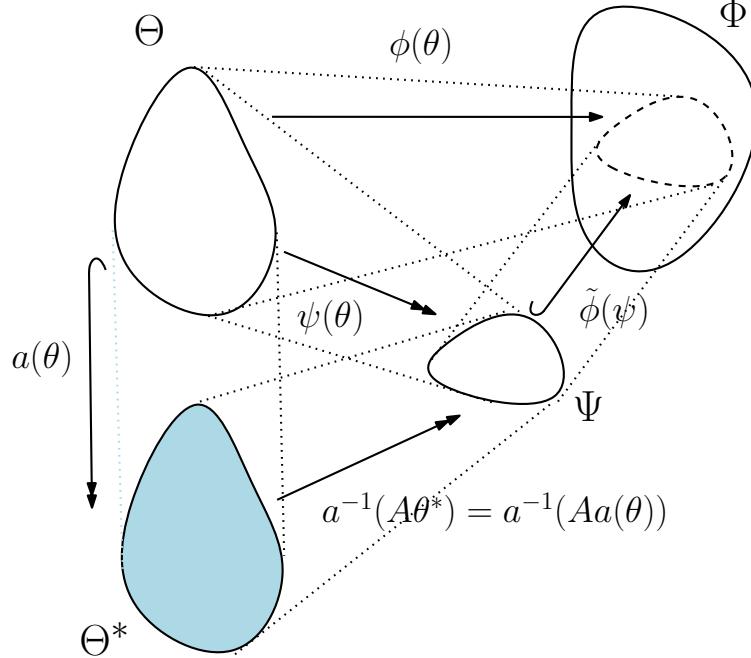


Figure 2: Illustration of the use of an initial componentwise transformation $a(\theta)$, typically log, followed by a linear transformation and a componentwise inverse, to generate identifiable parameter combinations. The chain rule applied on the path from Θ^* to Φ allows us to equate the overall null space to that of the constant linear transformation A , the key unknown of the parameter transformation.

The chain rule expression then becomes:

$$D_{\theta^*} \phi_*(\theta^*) = D_{\psi_*} \tilde{\phi}(\psi_*(\theta^*)) D_{\theta^*} \psi_*(\theta^*), \quad (15)$$

where we have $\psi_*(\theta^*) = \exp(A\theta^*)$, and hence $D_{\theta^*} \psi_*(\theta^*) = \text{diag}(\exp(A\theta^*))A$. Thus we have

$$D_{\theta^*} \phi_*(\theta^*) = D_{\psi_*} \tilde{\phi}(\psi_*(\theta^*)) \text{diag}(\exp(A\theta^*))A. \quad (16)$$

As $\tilde{\phi}$ is an immersion by our assumptions above, and the componentwise transformation a (typically \exp) is also an immersion, their composition preserves the key properties of the

factorisation. In particular (see supplementary material for proofs), the first two terms in the chain rule equation compose to give an injective mapping, while the matrix A represents the potentially non-one-to-one, surjective linear mapping. This leads to our key condition defining the reparameterisation:

$$\ker A = \ker D_{\theta^*} \phi_*(\theta^*). \quad (17)$$

Here A is a (to be determined) constant matrix defining appropriate linear combinations of log parameters to generate monomials. Crucially, as the left-hand side is independent of the parameter value, the right-hand side must be as well; we can, therefore, evaluate the derivative at any convenient point and know the associated nullspace is the same globally. This reduces what would be a PDE in the original coordinates to an algebraic computation in logged coordinates. Once A is determined based on the above, following the procedure in the following subsection, we have determined our transformation given by equation 13. We note that the above does not uniquely determine A , but establishes the existence of such a matrix capturing the identifiable parameter combinations. As below, we choose a particular singular value decomposition approach to determine A , with further desirable properties, but other methods could be used.

If the model assumed non-identifiable, the point used will generally be, e.g., any (non-unique) maximum likelihood estimate, and the Jacobian will be singular. This is expected and causes no issues in general. We can evaluate the required derivative by finite differences or, in Julia in the cases we consider, by automatic/algorithms differentiation of the source code [48]. Note that automatic differentiation at the source code level is distinct from symbolic differentiation and only evaluates the numerical value of the derivative at a fixed input value [53]. We also consider the case of near non-identifiability (i.e. practical non-identifiability) in the following subsections.

2.5 Singular value decomposition for nonlinear parameter combinations

The last main ingredient is a singular value decomposition of the Jacobian derivative in log-parameter coordinates to construct A and, hence, identify and separate parameter combinations based on their identifiability.

Specifically, given the Jacobian in log-parameters $D_{\theta^*} \phi_*$ at a reference estimate $\hat{\theta}^*$, we compute its SVD:

$$D_{\theta^*} \phi_*(\hat{\theta}^*) = U \Sigma V^T = U_r \Sigma_r V_r^T \quad (18)$$

where the second expression involving U_r , Σ_r and V_r is the reduced (cf. full) SVD, i.e. involving only non-zero singular values. The right singular vectors (columns of V) corresponding to non-zero singular values, i.e. those in V_r , span the space of locally identifiable parameter combinations, and the remainder span the non-identifiable parameter space. In particular, taking $A = V_r^T$, or any row-scaled version of this, ensures A has the same kernel (nullspace) as $D_{\theta^*} \phi_*(\hat{\theta}^*)$, and gives a reduced reparameterisation.

The case of practical identifiability, or near non-identifiability, is more subtle, but also handled by this approach. In this case, there will be singular values close to zero, and the corresponding singular vectors will be elements of the row space rather than the null space of the Jacobian. As the SVD leads to an orthogonal basis these singular vectors can still be considered to span a self-contained ‘practically non-identifiable’ parameter space, and the remaining singular vectors will still span the identifiable parameter space (there may also be a true null space if there are exactly zero singular values). The singular values thus provide a natural ranking of the identifiable parameter combinations by their ‘degree of identifiability’, including the possibilities of structural non-identifiability (zero singular value) and practical non-identifiability (small singular value). Treating the singular vectors as corresponding to parameter combinations with different degrees of identifiability rather than imposing a hard threshold somewhat avoids the issue of choosing a threshold and also leads to the idea of using a full reparameterisation, as discussed below.

2.5.1 Full reparameterisation

In the truly singular case, in addition to a reduced reparameterisation, we can also partition the full θ vector into identifiable combinations $\psi(\theta)$ and non-identifiable combinations $\lambda(\theta)$ using the full SVD via:

$$\theta \mapsto \begin{bmatrix} \psi(\theta) \\ \lambda(\theta) \end{bmatrix} = \exp(V^T \log(\theta)) \quad (19)$$

where $\phi(\theta)$ is the subvector corresponding to the rows of V^T associated with non-zero singular values and $\lambda(\theta)$ corresponds to the rows associated with zero singular values. This is useful because, although a model may in principle only depend on a reduced set of parameters, many simulation models are based on complex codes that take the full set of parameters as arguments and would have to be rewritten as functions of only the reduced set in order to run. By reparameterising the full set, including the non-identified parameters, the original codes can be used. Furthermore, as discussed above, in the case of near but not exact non-identifiability, i.e. practical non-identifiability, one would not want to accidentally

exclude potentially relevant variables based on some threshold. Given a log transformation, the combinations generated by the above take the form of products and ratios of original parameters to various powers.

2.5.2 Fisher information formulation

As mentioned above, since the nullspace and row space of the observed Fisher information are the same as that of the auxiliary mapping, we can also extend this approach to work directly with the observed Fisher information. This may be more natural for many models. Applying the same idea of an initial parameter-space linearising transformation, we can determine the identifiable parameters by solving a nullspace constraint, or, by simply considering the singular vectors (eigenvectors) of the Fisher information with non-zero singular values (eigenvalues). This is often the simplest approach in practice. If the parameters appear in monomial combinations in the nonlinear model, then an initial component-wise log transformation will enable us to find the identifiable parameters almost exactly, up to finite discretisation errors. Furthermore, the nullspace vectors define non-identifiable parameter combinations orthogonal to the identifiable combinations. Orthogonal parameters are particularly useful for likelihood-based statistical inference [54, 55], and we consider this aspect in more detail in the context of our results.

Following this approach, we again get a transformation (of the full parameter vector) of the form

$$\theta \mapsto \begin{bmatrix} \psi(\theta) \\ \lambda(\theta) \end{bmatrix} = \exp(\bar{V}^T \log(\theta)), \quad (20)$$

where \bar{V}^T is now the matrix of singular vectors from the SVD of the observed Fisher information $\mathcal{J}(\theta^*)$ expressed in log-transformed coordinates θ^* .

2.6 Practical vs structural identifiability and observation operators

That the observed Fisher information and the auxiliary mapping have the same rank and null space may seem surprising at first, as the former is generally associated with *practical* identifiability analysis, while the latter is essentially a *structural* identifiability tool. However, recalling that we are taking the data distribution parameters ϕ to be given by the output of, e.g., a numerical simulation model, this is resolved by clarifying whether a given analysis is in terms of the solution grid resolution or the measurement grid resolution. That is, if

we consider our analysis at the level of our observation grid, then our mapping to data distribution parameters is actually given by, e.g.,

$$\phi = B_{\text{obs}}\phi_{\text{fine}} \quad (21)$$

where ϕ_{fine} is our fine-scale solution and B_{obs} is, e.g., a linear operator mapping the fine-scale vector to the vector of values at given observation points. This potentially introduces additional null space or at least additional poorly identified components. We can thus consider both structural and practical identifiability questions within the same framework, depending on whether we consider fine-scale discretisation or coarse-scale discretisation. If we consider our analysis on the ideal solution grid level, then, to numerical approximation, we will expect to obtain the same results as a structural identifiability analysis, though for ill-posed problems, it may be difficult or even impossible to distinguish effectively (i.e. practically) non-identified and truly (structurally) non-identified scenarios [15] TODO MORE REFs. In such cases we believe it is more appropriate to report the practical identifiability results, which is what the present approach provides.

A more subtle issue occurs when the ‘poor identifiability’ is not due to finite data issues as such, but rather due to the existence of a limiting approximate model involving a reduced set of parameters. That is, the existence of a structurally non-identifiable model in a particular parameter limit that is not exactly reached in reality. In this case, the auxiliary mapping (even on a ‘fine grid’), can have arbitrarily small singular values depending on the region of parameter space considered. This lack of clean division between practical and structural identifiability is discussed further in [15]. In the present work, our ‘parameterised normal approximations to the binomial and Poisson models’ example provides a simple illustration of this issue, where the limiting Poisson model is structurally non-identifiable, but the binomial model is only practically non-identifiable.

In practice, if we choose to normalise the rows of V^T to have integer entries where possible (by scaling and rounding rows appropriately) while maintaining orthogonality, the resulting parameter combinations should, in principle, correspond exactly to structurally identifiable (for non-zero singular values) and structurally non-identifiable (for zero singular values) combinations that would be found using symbolic methods. Furthermore, due to continuity of the singular vectors with respect to the singular values (REF TODO), in the case of near non-identifiability, the resulting parameter combinations should, depending on the rounding procedure used, typically correspond to those that are structurally identifiable and non-identifiable in the limit model where small singular values become zero. This is a key advantage of our approach, as it provides a bridge between numerical and symbolic methods

for structural identifiability analysis.

Although there exist various formal algorithms for determining integer or rational approximations to real-valued vectors (TODO REFs), here we use a simple scaling and rounding heuristic. Different scale factors for a given row reflect higher or lower powers of the corresponding basic combination of parameters and preserve orthogonality relationships and the key identifiability information. Rounding may break this slightly in some cases, but the effects are likely small in many cases, and attempts can be made to keep resulting vectors orthogonal. Scaling and rounding can be done by inspection or via automated heuristics (e.g. as in our Julia code), though we leave the best way to fully automate this to future work. In the case of structural non-identifiability, the effects of scaling and rounding can be reduced, as discussed above, by taking the grid considered to be the solution rather than the observation grid. The question of how reliably an approximating structurally non-identifiable model can be found in the practically non-identifiable case is a more open question. Though we consider this question in the context of our examples, we leave a more general analysis to future work.

Next, we describe a series of examples of varying complexity to which we will apply our methodology.

3 Example models

Here, we describe three examples of varying areas of application and complexity, covering statistics and data science (parameterised normal distributions based on approximations to the binomial and Poisson models) [56, 54, 57], biochemistry and bioengineering (Monod/Michaelis-Menten kinetics) [58, 59], and physics and engineering (groundwater flow, heat conduction, diffusive and other transport modelling) [Crank1975, 3, 60, 61, 62, 63, 64, 65].

3.1 Statistics and data science: Parameterised normal approximations to the binomial and Poisson models

Our first example is a ‘toy’ model motivated by the (seemingly) simple problem of simultaneously estimating the number of trials n and success probability p in a binomial model given a collection of independent success counts from repeating the n -trial experiment k of times,

where k is known. This problem is reviewed in [56], where the authors note that the ‘near Poisson’ regime of large n and small p lead to unstable maximum likelihood estimates. This corresponds to a non-identifiable limit and hence (as also shown here) a case of practical non-identifiability close to this limit. A likelihood-based analysis of the problem is given by [54], further discussed in [57], and our analysis of this example is inspired by theirs. To simplify the example, however, and to enable the use of derivative-based methods, we treat the ‘number of trials’ as a continuous parameter (similar ideas are used in modern higher-order likelihood theory [66]). In fact, although the binomial (and Poisson) models provide motivation for this example, here we treat the n and p parameters as simply providing an abstract parameterisation of the normal distribution model with a similar structure to more realistic mechanistic models. To further simplify the details of the example, we also consider a model analogous to the normal approximation of the binomial model, along with the associated Poisson limit.

With these caveats, the model we consider is, for a single experiment (each analogous to an ‘ n -trial’ Binomial experiment):

$$Y \sim \mathcal{N}(np, np(1 - np)), \quad (22)$$

along with the ‘Poisson limit’ case:

$$Y \sim \mathcal{N}(np, np). \quad (23)$$

In our framework, this model is produced in a structured manner by combining the normal data distribution model, uniquely defined by the mean and variance parameters:

$$\begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix} \leftrightarrow \mathcal{N}(\mu, \sigma^2), \quad (24)$$

with (for the non limiting model) the auxiliary mapping:

$$\phi : \begin{bmatrix} n \\ p \end{bmatrix} \mapsto \begin{bmatrix} np \\ np(1 - p) \end{bmatrix} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}. \quad (25)$$

That is, the underlying (or ‘mechanistic’) parameter is given by vectors of the form

$$\begin{bmatrix} n \\ p \end{bmatrix}, \quad (26)$$

while the ‘data distribution’ parameters are the usual mean and variance parameters in one-to-one correspondence with a given normal distribution as in equation 24. The auxiliary

mapping defines the dependence of μ and σ^2 on the underlying parameters n, p . Although this is not really a ‘mechanistic’ model in the usual sense, one could consider an underlying binomial model is a lower-level mechanism that ‘generates’ the higher-level (approximately) normally distributed observations. The problem is to determine both n and p given k observations from the single experiment model given by equation 22 or 23.

3.2 Biochemistry and bioengineering: Michaelis-Menten/Monod kinetics

The Michaelis-Menten equation [59] was originally developed in the context of enzyme kinetics to represent the overall rate of an enzyme-mediated reaction as a function of the concentration of the substrate, but is now widely used in many areas of biology and bioengineering a simple model involving saturating kinetics. Monod [58] used Michaelis-Menten style kinetics in microbiology to model the growth rate parameter of microorganisms as a function of substrate concentration. In this context, as with other more complex models in which the Michaelis-Menten equation appears as a component, the equations are coupled with other equations and variables of the system. Holmberg [67] carried out an early identifiability analysis of Michaelis-Menten parameters in the context of Monod-style microbial growth models, showing that the parameters are structurally identified but not practically identified, particularly in the regime where the initial substrate concentration, $S(0)$ is low relative to the half-saturation constant, K .

Here we consider a simplified, substrate-only Michaelis-Menten/Monod model for the depletion of a substrate S in the form:

$$\frac{dS}{dt} = -\frac{\nu S}{K + S}, \quad (27)$$

where S is the substrate concentration, ν is the maximum growth rate, and K is called the half-saturation, half-velocity, or Michaelis-Menten constant, depending on context. A natural limit model, in the substrate concentration (relative to K) regime, is then:

$$\frac{dS}{dt} = -\frac{\nu}{K}S. \quad (28)$$

For simplicity, we assume a normal distribution error model with constant standard deviation $\sigma = 0.05$, though e.g. log-normal errors or other error models could also be considered easily within our framework [46]. If we denote the model solution at time t for parameter vector $\theta = [\nu, K]^T$ by $S(t; \nu, K) = S(t; \theta)$, we can consider the ‘full model solution’ for a given

parameter vector to be represented approximately by its evaluation on a fine time grid as a *vector* of values $s(\nu, K) = s(\theta) = S(t_1; \theta), S(t_2; \theta), \dots, S(t_k; \theta)$, where t_1, t_2, \dots, t_k are the fine time-grid points. Our *observed* concentrations, s_{obs} are then defined on the observation grid as

$$s_{\text{obs}} \sim \mathcal{N}(B_{\text{obs}}s(\theta), \sigma^2 I), \quad (29)$$

where our observation model has standard deviation $\sigma = 0.05$, and B_{obs} is a linear mapping of the fine-scale solution to the observation grid. Our ‘structural’ auxiliary mapping is hence

$$\phi : \begin{bmatrix} \nu \\ K \end{bmatrix} \mapsto s(\nu, K), \quad (30)$$

while we can also consider a ‘practical’ auxiliary mapping

$$\phi : \begin{bmatrix} \nu \\ K \end{bmatrix} \mapsto B_{\text{obs}}s(\nu, K), \quad (31)$$

where the latter incorporates the resolution of the observation grid but not the measurement noise. In our present work we use the structural auxiliary mapping to determine the reparameterisation, but the practical auxiliary mapping could also be used, and also enters into the analysis via the point estimate determination.

3.3 Physics and engineering: Flow in a composite medium

Our final example is a simple model appearing in groundwater modelling [3, 60, 61] and analogous areas such as heat flow [62, 63], diffusive transport! [68], and many other transport phenomena models [65, 64]. For concreteness, we consider the problem in the context of groundwater modelling, specifically steady flow of groundwater between two water bodies in a one-dimensional, two-part heterogeneous confined aquifer with two different diffusivity coefficients and a recharge term (e.g. rainfall). Figure 3 shows a schematic of the model.

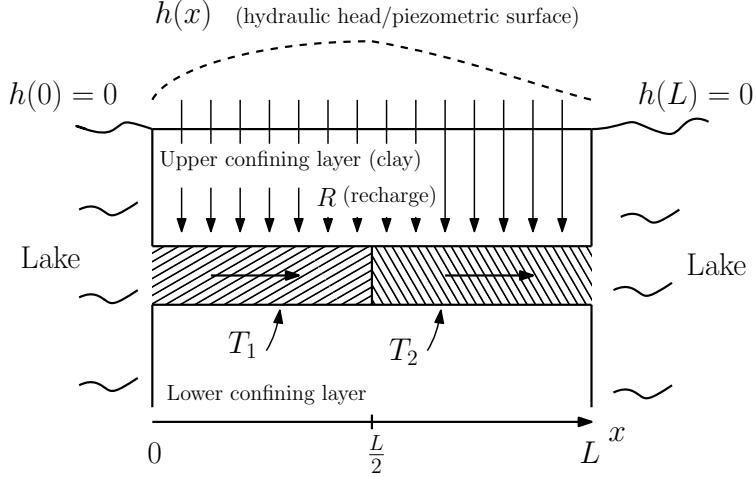


Figure 3: Schematic of the groundwater flow model.

The model is governed by the equation (see e.g. [60, 61]):

$$T_i \frac{d^2 h}{dx^2} + R = 0, \quad (32)$$

where h is the hydraulic head, $x \in [0, L]$ is the spatial coordinate, R is the (assumed constant) source term representing rainfall infiltration, and T_i is the transmissivity in region $i = 1, 2$. We use fixed head boundary conditions at the two ends of the aquifer, and continuity of flow and head at the interface between the two regions (assumed to be at $L/2$):

$$\begin{aligned} h_1(0) &= 0 \\ h_2(L) &= 0 \\ h_1\left(\frac{L}{2}\right) &= h_2\left(\frac{L}{2}\right) \\ -T_1 \frac{dh_1}{dx}\left(\frac{L}{2}\right) &= -T_2 \frac{dh_2}{dx}\left(\frac{L}{2}\right), \end{aligned} \quad (33)$$

where L is the length of the aquifer, $h_1(x)$ and $h_2(x)$ are the hydraulic heads in regions 1 and 2 respectively, and the last condition represents continuity of the fluid flux at the interface. Equivalent models appear in heat conduction or diffusive transport problems in composite media (see references above). This particular model is sufficiently simple that it can be solved analytically, and the solution to this equation is piecewise quadratic, given by

$$h(x) = \begin{cases} -\frac{1}{2} \left(\frac{R}{T_1} \right) x^2 + \alpha_1 x + \beta_1, & x < \frac{L}{2} \\ -\frac{1}{2} \left(\frac{R}{T_2} \right) x^2 + \alpha_2 x + \beta_2, & x > \frac{L}{2} \end{cases} \quad (34)$$

where α_i and β_i are constants determined by the boundary and interface conditions. From $h(0) = 0$ we have $\beta_1 = 0$, and from the flux continuity condition at $x = L/2$ we find $\alpha_1 = \alpha_2 = \alpha$. The remaining conditions, $h(L) = 0$ and continuity of h at $x = L/2$, give

$$\begin{aligned}\alpha &= \frac{L}{2} \left(\frac{R}{T_2} \right) + \frac{L}{8} \left[\left(\frac{R}{T_1} \right) - \left(\frac{R}{T_2} \right) \right], \\ \beta_2 &= -\frac{L^2}{8} \left[\left(\frac{R}{T_1} \right) - \left(\frac{R}{T_2} \right) \right].\end{aligned}\tag{35}$$

In terms of unknown parameters, these expressions only depend on the ratios of the recharge rate and transmissivity coefficients, i.e. $R_1 = R/T_1$ and $R_2 = T_1/T_2$. Thus we expect that the model is structurally non-identifiable, and confirm our approach discovers this in the Results section.

As in the Michaelis-Menten example, we complete the model by assuming a simple normal error model with constant standard deviation, here $\sigma = 0.2$ on the observed hydraulic head values. Furthermore, we again use the fine-scale solution as the ‘full model solution’ and map to the observation grid via a linear observation operator B_{obs} . The structural auxiliary mapping is then given by the fine-scale solution, and the practical auxiliary mapping is given by the observation grid values.

Thus we have

$$\phi : \begin{bmatrix} T_1 \\ T_2 \\ R \end{bmatrix} \mapsto h(T_1, T_2, R),\tag{36}$$

where $h(T_1, T_2, R) = h(\theta)$ is the fine-scale solution to the model given by equation 32 for parameter vector $\theta = [T_1, T_2, R]^T$. The statistical model is then given by

$$h_{\text{obs}} \sim \mathcal{N}(B_{\text{obs}}h(\theta), \sigma^2 I),\tag{37}$$

where h_{obs} is the observed hydraulic head values.

4 Results and Discussion

4.1 Parameterised Normal Models

These examples are sufficiently simple that we can illustrate the core ideas analytically, and so we first consider some analytical results. These results are somewhat ‘obvious’ but we

derive them according to our general framework to illustrate the method. Similarly, the numerical results we then discuss were produced without assuming these analytical results were available, and used our generic model-agnostic code.

We focus our analytical illustration on the Poisson limit model for simplicity, defined by equation 23. In this model we expect only np to be identifiable, as the variance is equal to the mean. We will show that both symbolic (original parameters) and numerical (log-transformed parameters) approaches obtain this result.

We also present the results of numerical, likelihood-based analysis of both parameterised normal models. We consider the likelihood functions in both original and our reparameterisations. Although the likelihood is invariant to reparameterisation, we see how the reparameterisation ‘brings out’ the separation of information concerning the identifiable and non-identifiable parameters. Furthermore, we see that approximate, non-parameterisation-invariant likelihood procedures such as quadratic log-likelihood approximations and likelihood ‘slices’ are more reasonable in the reparameterised space.

4.1.1 Analytical results: Poisson limit model, original parameterisation

To first obtain the symbolic result we note, as above, that the auxiliary mapping in terms of untransformed parameters is given by

$$\phi(n, p) = \phi\left(\begin{bmatrix} n \\ p \end{bmatrix}\right) = \begin{bmatrix} np \\ np \end{bmatrix}.$$

The Jacobian of this mapping is given by

$$D\phi(n, p) = \begin{bmatrix} p & n \\ p & n \end{bmatrix}. \quad (38)$$

This is a function of n and p and clearly has a non-trivial null space for all n and p . As the Jacobian and the associated null space depend on (are given as a function of) n and p we refer to this as the ‘symbolic’ null space of the model. We can determine this by inspection, noting that vectors of the form

$$\alpha(n, p) = \begin{bmatrix} n \\ -p \end{bmatrix} \quad (39)$$

satisfy $D\phi(n, p)\alpha(n, p) = 0$ for all n and p . Thus the symbolic null space has dimension one and is spanned by the vector $\alpha(n, p)$; furthermore, the row space of $D\phi(n, p)$ hence has dimension one and is spanned by a symbolic vector orthogonal to $\alpha(n, p)$. This implies that

$\psi(n, p)$ is scalar-valued and can be determined by equating the (unknown) single row of $D\psi(n, p)$ to the (transpose of the) vector spanning the (known) row space of $D\phi(n, p)$. By inspection we see that the row space of $D\phi(n, p)$ is spanned by the vector

$$\begin{bmatrix} p \\ n \end{bmatrix}, \quad (40)$$

which is orthogonal to $\alpha(n, p) = [n, -p]^T$. These properties are also evident from the ‘symbolic’ rank factorisation of $D\phi(n, p)$:

$$D\phi(n, p) = \begin{bmatrix} p & n \\ p & n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} p & n \end{bmatrix}. \quad (41)$$

Thus we can determine $\psi(n, p)$ by solving the equations in the row space form

$$\begin{aligned} \frac{\partial \psi}{\partial n} &= p \\ \frac{\partial \psi}{\partial p} &= n, \end{aligned} \quad (42)$$

or solving the partial differential equation

$$\frac{\partial \psi}{\partial n} n + \frac{\partial \psi}{\partial p} (-p) \quad (43)$$

based on the orthogonality of the rows of $D\psi(n, p)$ and the vectors spanning the null space of $D\phi(n, p)$.

By inspection we see that $\psi(n, p) = np$ is a solution to these equations, and hence the identifiable parameterisation is given by $\psi(n, p) = np$, as expected.

This approach is equivalent to that of [20, 21] but, as with their approach, requires symbolic determination of the Jacobian and the solution of partial differential equations. We now show how our numerical approach can determine the same result without symbolic manipulation. We can again carry out the key analysis analytically in this simple case.

4.1.2 Analytical results: Poisson limit model, log-transformed parameterisation

We now consider the Poisson limit model in log-transformed parameters. By definition (specifically, from parameterisation invariance) the auxiliary mapping in terms of log-transformed parameters satisfies

$$\phi_*(n^*, p^*) = \phi_*\left(\begin{bmatrix} n^* \\ p^* \end{bmatrix}\right) = \phi\left(\begin{bmatrix} n \\ p \end{bmatrix}\right) = \phi\left(\exp\left(\begin{bmatrix} n^* \\ p^* \end{bmatrix}\right)\right), \quad (44)$$

where $n^* = \log n$ and $p^* = \log p$ and where \exp is the componentwise exponential function so that

$$\begin{bmatrix} n \\ p \end{bmatrix} = \exp\left(\begin{bmatrix} n^* \\ p^* \end{bmatrix}\right).$$

The right-hand side of the above equation is obtained from the definition of the auxiliary mapping in terms of untransformed parameters:

$$\phi\left(\begin{bmatrix} n \\ p \end{bmatrix}\right) = \begin{bmatrix} np \\ np \end{bmatrix} = \begin{bmatrix} \exp(n^*) \exp(p^*) \\ \exp(n^*) \exp(p^*) \end{bmatrix} = \begin{bmatrix} \exp(n^* + p^*) \\ \exp(n^* + p^*) \end{bmatrix}. \quad (45)$$

Thus the Jacobian of the auxiliary mapping in terms of log-transformed parameters is given by

$$D\phi_*(n^*, p^*) = \begin{bmatrix} \exp(n^* + p^*) & \exp(n^* + p^*) \\ \exp(n^* + p^*) & \exp(n^* + p^*) \end{bmatrix} = \exp(n^* + p^*) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad (46)$$

As expected, this takes the SVD form (up to normalisation of the singular vectors):

$$D\phi_*(n^*, p^*) = \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{D_{\psi_*} \tilde{\phi}(\psi_*(\theta^*))} \underbrace{\exp(n^* + p^*)}_{\text{diag}(\exp(A\theta^*))} \underbrace{\begin{bmatrix} 1 & 1 \end{bmatrix}}_A, \quad (47)$$

where $\theta^* = [n^*, p^*]^T$, A is the matrix of the constant linear transformation component of the reparameterisation and $D_{\psi_*} \tilde{\phi}(\psi_*(\theta^*))$ is the Jacobian of the auxiliary mapping in terms of the transformed, identifiable parameters. The ‘diagonal’ matrix is a scalar in this case as the intermediate image is represented by a single parameter. Since $\psi_*(n^*, p^*) = \exp(n^* + p^*)$, and the output of the auxiliary mapping is

$$\phi_*(n^*, p^*) = \begin{bmatrix} \exp(n^* + p^*) \\ \exp(n^* + p^*) \end{bmatrix} = \begin{bmatrix} \psi_*(n^*, p^*) \\ \psi_*(n^*, p^*) \end{bmatrix}, \quad (48)$$

we have that $D_{\psi_*} \tilde{\phi}(\psi_*(\theta^*))$ is also constant in this case. In more general classes of models, the A matrix will remain constant, but the Jacobian of the auxiliary mapping in terms of the identifiable parameters will depend on the parameters.

Thus we conclude that the identifiable parameterisation is given by

$$\psi_*(n^*, p^*) = \exp(A\theta^*) = \exp\left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} n^* \\ p^* \end{bmatrix}\right) = \exp(n^* + p^*) = \exp(\log n + \log p) = np. \quad (49)$$

This is the same result as obtained in the symbolic analysis above. However, as the A matrix here is constant, and corresponds (up to normalisation) to the singular vector of the Jacobian

of the auxiliary mapping in log-transformed parameters, we can determine it with a single numerical evaluation of the Jacobian. This is the key advantage of our numerical approach.

We can also determine the non-identifiable parameters by considering the full SVD of the Jacobian (or by finding orthogonal vectors to the identifiable parameter combination). In this case, there is a single singular vector, given by (up to normalisation)

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad (50)$$

which spans the null space of the Jacobian and is orthogonal to the row space giving the identifiable parameter combination. This implies that the non-identifiable parameterisation is given by

$$\lambda_*(n^*, p^*) = \exp\left(\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} n^* \\ p^* \end{bmatrix}\right) = \exp(n^* - p^*) = \frac{n}{p}. \quad (51)$$

4.1.3 Numerical results: Poisson limit model

Below we present the results of our numerical analysis of the Poisson limit model. We consider the likelihood functions in both the original parameterisation, Figure 4, and the ‘invariant image’ reparameterisation, Figure 5. These results illustrate the separation of information concerning the identifiable and non-identifiable parameters in the reparameterised space, and the clearer suitability of approximate, non-parameterisation-invariant likelihood procedures in the reparameterised space.

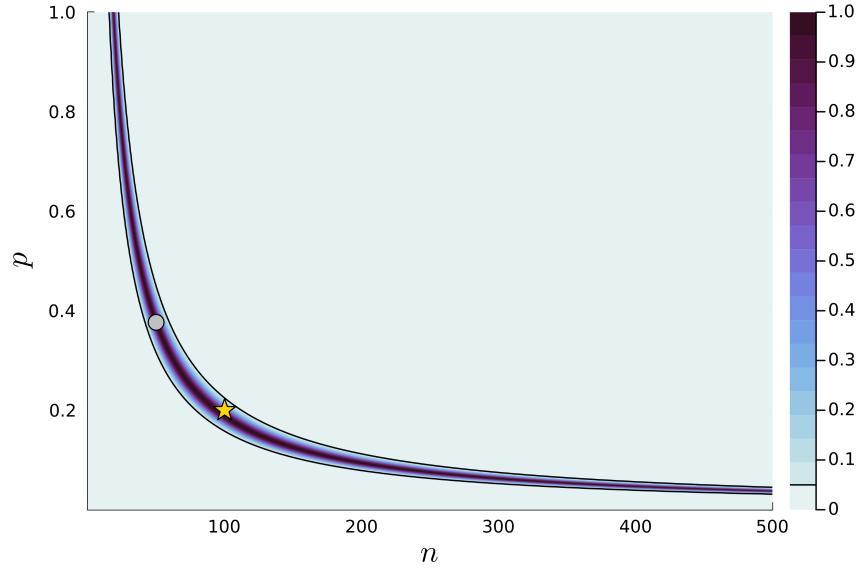
For these results we used true parameter values of $n = 100$ and $p = 0.2$, and a sample size of 10. For reproducibility, the data realisation used was [21.9, 22.3, 12.8, 16.4, 16.4, 20.3, 16.2, 20.0, 19.7, 24.4]. This was generated from the *non-limit* model, i.e. treating the limit model as an approximation at the analysis rather than data generation stage. The analysis can be found in the repository at `examples/stat_model.jl`, setting the `poisson_limit` variable to `true`.

In Figure 4, showing the likelihood in the original parameterisation, we see a long ‘banana’ shaped likelihood contour in the (n, p) plane, illustrating the existence of (approximately) equivalent parameter combinations lying along a curved relationship. The profile likelihoods for each parameter, shown in the bottom row of Figure 4, illustrate the individual non-identifiability of the parameters. The likelihood is completely flat for both parameters, other than where the bounds of the parameter space are reached. Together these results imply

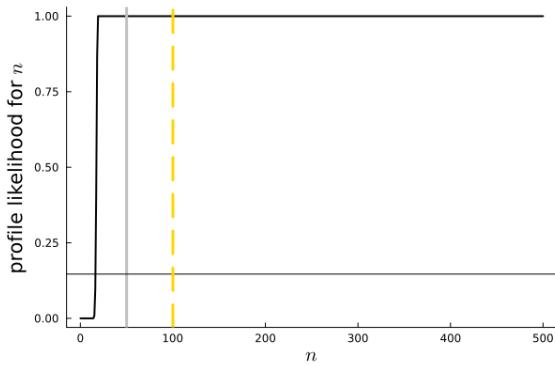
that the two parameters are not individually identifiable in this model, though there may be some combination of parameters that is identifiable (i.e. np , as we have shown analytically).

In Figure 5, showing the likelihood in the reparameterised space, we see the likelihood is constant in the vertical, $\frac{n}{p}$, direction and varies only in the horizontal, np , direction. This reflects the identifiability of the parameter combination np , and the non-identifiability of the parameter combination n/p . Furthermore this structure means the information concerning the identifiable and non-identifiable parameters is completely separated, as the likelihood factors into a product of a function of np and a (constant) function of n/p (see e.g. the discussion in [54, 55]). Thus we can, in principle, construct an autonomous reduced model based on the identifiable parameter combination np (as can be seen analytically).

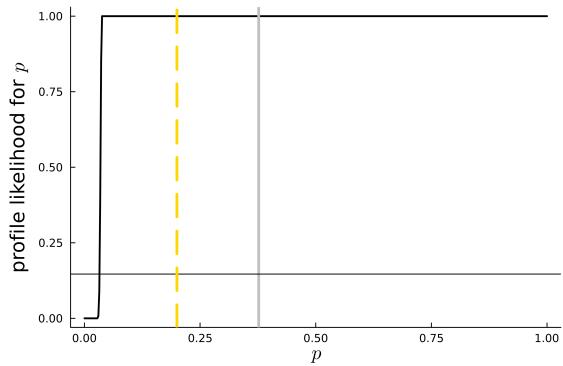
The profile likelihoods for each parameter, shown in the bottom row of Figure 5, further illustrate the distinct identifiability properties of the parameter combinations. The profile likelihood for np is a simple Gaussian-like function, showing good estimability, while the profile likelihood for n/p is completely flat, demonstrating non-identifiability. This is consistent with our analytical results.



(a) Joint likelihood contour plot in the (n, p) plane. The gold star indicates the true parameter value, and the silver circle indicates the maximum likelihood estimate. The colour bar represents the relative likelihood, with a black contour line indicating the 95% confidence interval threshold.

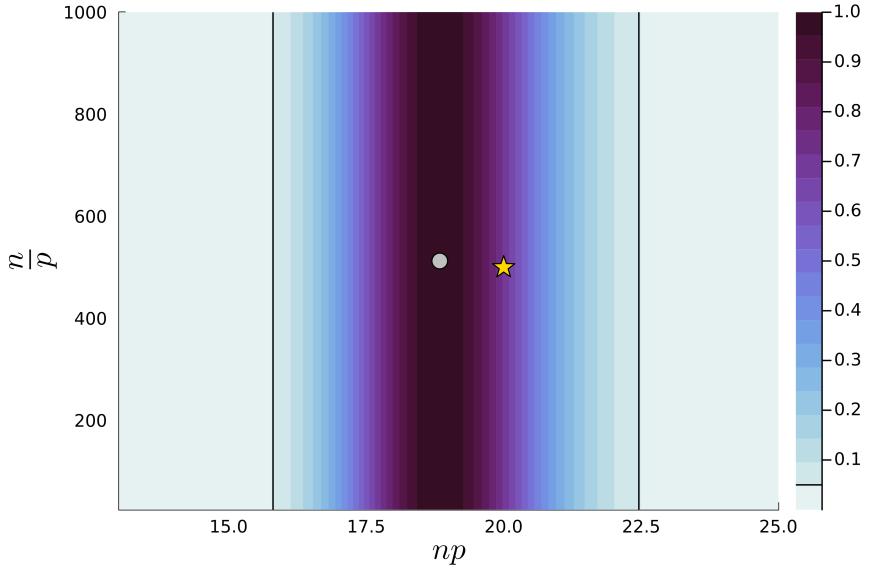


(b) Profile likelihood for n . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

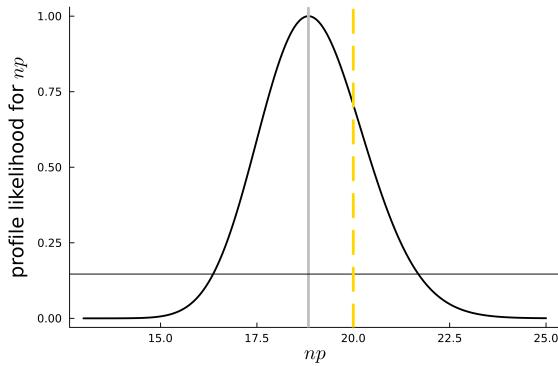


(c) Profile likelihood for p . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

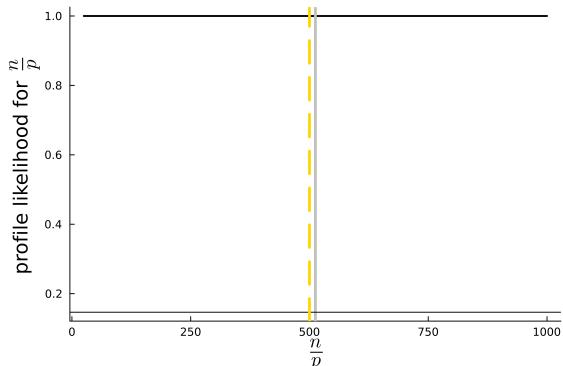
Figure 4: Poisson-limit example: (Top) joint contour in the (n, p) plane. (Bottom) profile likelihoods.



(a) Joint likelihood contour plot in the $(np, \frac{n}{p})$ plane. The gold star indicates the true parameter value, and the silver circle indicates the maximum likelihood estimate. The colour bar represents the relative likelihood, with a black contour line indicating the 95% confidence interval threshold.



(b) Profile likelihood for np . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.



(c) Profile likelihood for $\frac{n}{p}$. The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

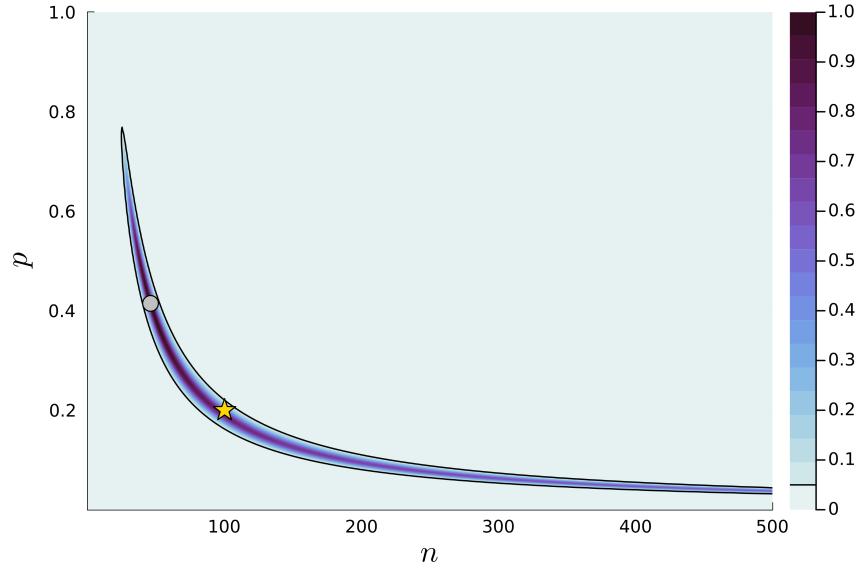
Figure 5: Reparameterised Poisson-limit example: (Top) joint contour in the $(np, \frac{n}{p})$ plane. (Bottom) profile likelihoods.

4.1.4 Numerical results: non-limit model

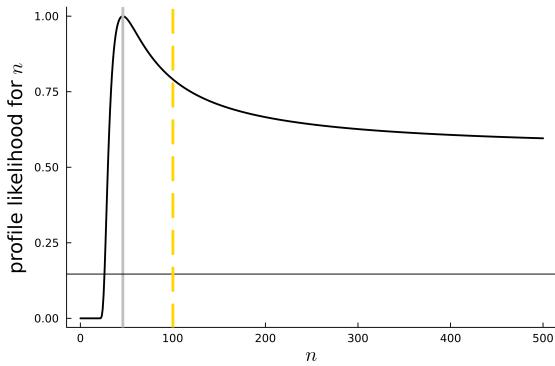
We now consider the non-limit model, defined by equation 22. As can be verified analytically, this model is now technically identifiable. For example, the mean and variance are now distinct as $np \neq np(1 - p)$ for $p \neq 0$, and we can solve for n and p separately from these two quantities. However, near the Poisson limit we also expect the parameters to be practically non-identifiable, i.e. poorly-identified. In this setting the well-identified parameter determined from the SVD will be close but not exactly equal to np , while there will be a poorly identified parameter close to but not exactly equal to n/p . We can either work with the exact SVD singular vectors obtained e.g. for our point estimate or use the heuristic rounding approach to obtain the approximate reparameterisation. Any reparameterisation in which the associated vectors are linearly independent is valid and the associated (parameterisation-independent) likelihood analysis losing no information. The main difference is the extent to which exact or approximate separation of information is obtained and the extent to which the well-identified parameter is (likelihood or observed information) ‘orthogonal’ to the poorly-identified parameter (see e.g. [54, 55]). Here we find that our rounding heuristic (divding by smallest non-zero entry and rounding to nearest 0.5) indeed recovers the previous identifiable and non-identifiable parameter combinations, but these now correspond to ‘well-identified’ and ‘poorly but structurally identified’ parameters, respectively.

For these results again we used true parameter values of $n = 100$ and $p = 0.2$, and a sample size of 10. For reproducibility, the data realisation used was (again) [21.9, 22.3, 12.8, 16.4, 16.4, 20.3, 16.2, 20.0, 19.7, 24.4]. The analysis can be found in the repository at `examples/stat_model.jl`, setting the `poisson_limit` variable to `false`.

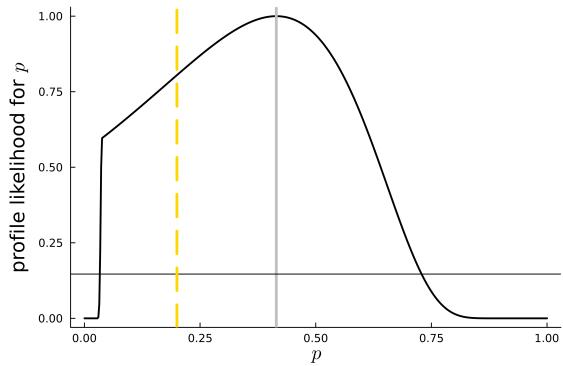
These results can be seen in Figures 6 and 7, which largely mirror those of the Poisson limit model. The well-identified parameter np and the poorly identified parameter n/p are again clearly separated in the reparameterised space. However, in this case the separation is not exact, and the poorly identified parameter is not completely flat in the profile likelihood. We also see *one-sided identifiability* in both the original parameterisation and in the poorly identified parameter in the reparameterised model – the likelihood is much flatter on one side of the maximum likelihood estimate than the other. We can attribute this to the model becoming poorly identified in the Poisson limit, which occurs asymmetrically in parameter space. These results illustrate the similar, but more complex structure of practical non-identifiability compared to structural non-identifiability.



(a) Joint likelihood contour plot in the (n, p) plane. The gold star indicates the true parameter value, and the silver circle indicates the maximum likelihood estimate. The colour bar represents the relative likelihood, with a black contour line indicating the 95% confidence interval threshold.

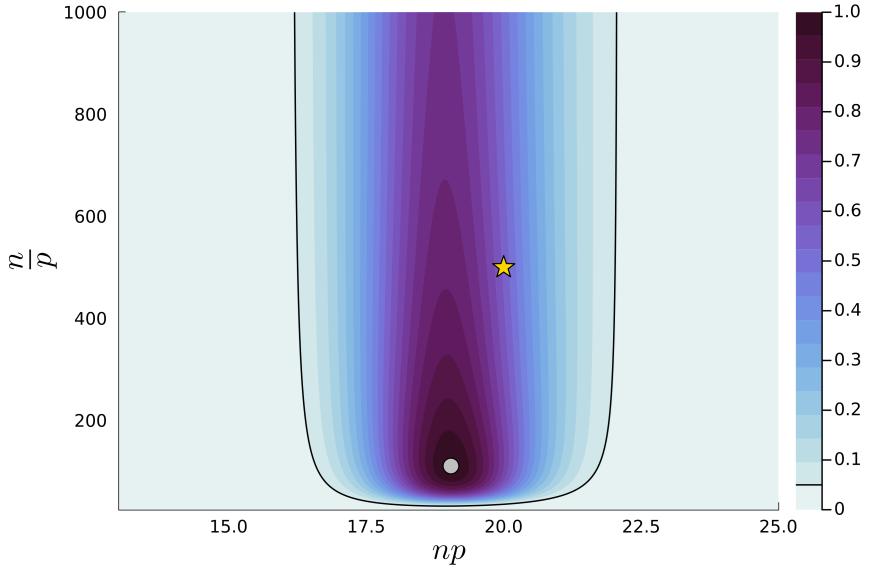


(b) Profile likelihood for n . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

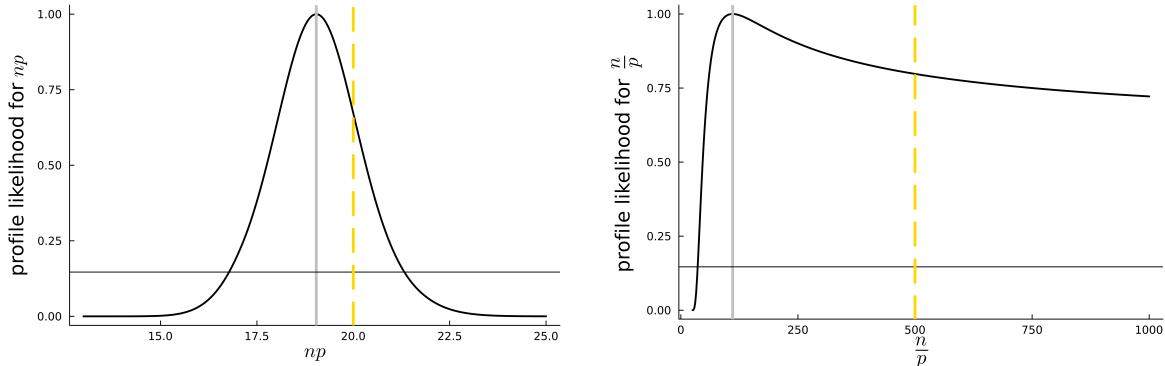


(c) Profile likelihood for p . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

Figure 6: Non-limit (approximate Binomial) example: (Top) joint contour in the (n, p) plane. (Bottom) profile likelihoods.



(a) Joint likelihood contour plot in the $(np, \frac{n}{p})$ plane. The gold star indicates the true parameter value, and the silver circle indicates the maximum likelihood estimate. The colour bar represents the relative likelihood, with a black contour line indicating the 95% confidence interval threshold.



(b) Profile likelihood for np . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

(c) Profile likelihood for $\frac{n}{p}$. The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

Figure 7: Reparameterised non-limit (approximate Binomial) example: (Top) joint contour in the $(np, \frac{n}{p})$ plane. (Bottom) profile likelihoods.

4.2 Michaelis-Menten/Monod kinetics

We now consider the Michaelis-Menten/Monod kinetics model, defined by equation 27. This model is structurally identifiable, but practically non-identifiable in the limit regime where the initial substrate concentration, $S(0)$ is low relative to the half-saturation constant, K . Rather than consider the limit model first, we first present the results from the non-reduced model. We then note the reduction suggested by the analysis corresponds to the limit model mentioned above, and proceed to present the results of this limit model. This reflects the more natural order of analysis, where identifiability analysis suggests the possibility of a reduced, structurally non-identified model, which can be investigated subsequently. We first present and compare parameter space results for each model, and then present prediction results for each model.

For these results we used an initial concentration of $S(0) = 1$, a true half-saturation constant of $K = 5.0$, and a maximum growth rate of $\nu = 1.0$. We used a normal observation model with standard deviation $\sigma = 0.05$, and observed the concentrations at 11 equally spaced time points between 0 and 20. We used a fine time grid of 201 points between 0 and 20 for the model solution in the auxiliary mapping computations. This regime is relatively close to the limit regime, but not so close that the effects are negligible. We expect the parameters to be potentially practically non-identifiable in this regime, but structurally identifiable. Our analysis can be found in the repository at `examples/mm_model.jl`.

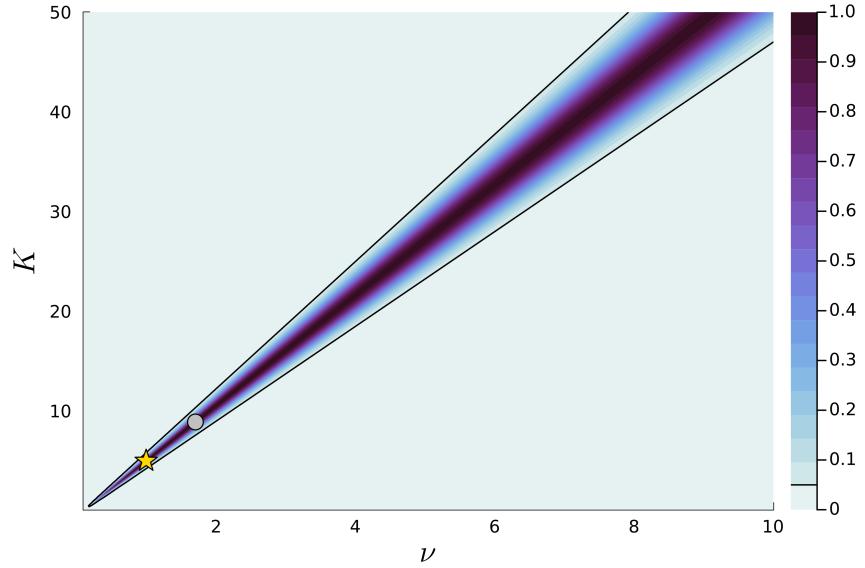
Figure 8 shows the results of the non-reduced model analysis in the original parameterisation. We see that the joint likelihood contours in the (ν, K) plane consist of ‘fanning’ lines, indicating the existence of parameter combinations having similar likelihood values. The profile likelihoods for each parameter show that the maximum likelihood estimate is close to the true parameter value in this particular case, but that the likelihood is very flat to one side (excluding the effects of parameter bounds) and show characteristic ‘one-sided identifiability’. This indicates that the parameters are practically non-identifiable in this regime, as expected.

Figure 9 shows the results of the non-reduced model analysis in reparameterised space. Here the parameter combinations identified are $\frac{K}{\nu}$ (well-identified) and νK (poorly identified). The first parameter represents the time scale associated with the limit model. We see a similar likelihood contour shape to the previous model analysis, with an apparent limiting regime where the parameters are decoupled and one parameter is well-identified while the other is poorly identified. Outside of the limit regime the relationship between the parameters is more complex, and the parameters are not as clearly separated. In this case the true

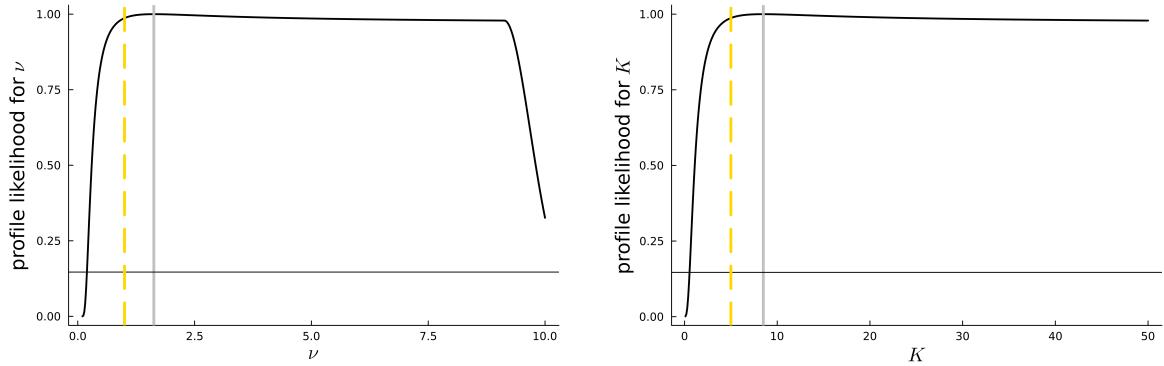
parameter lies in-between the limiting and non-limiting regimes, representing a non-trivial case where using the reduced model would incur some approximation error. Again, however, the likelihood is invariant to reparameterisation, so analysis of the non-reduced model in either coordinate system is equally valid. The role of reparameterisation is to bring out the separation of information concerning the potentially identifiable and non-identifiable parameters, and the suitability of approximate, non-parameterisation-invariant likelihood procedures in the reparameterised space or model reductions.

Although only approximate, these results do suggest a structurally non-identified reduced model may be useful, in the limit as $K\nu$ becomes large while $\frac{K}{\nu}$ stays constant. This corresponds to the limit model mentioned in the model description section above (which can be written equivalently in terms of $\frac{K}{\nu}$ or $\frac{\nu}{K}$). Figure 10 shows the results of the reduced model analysis in the reparameterised space. We see the likelihood contours agree well in the limit regime, while keeping the separation of parameters in the non-limit regime where the approximation technically breaks down. The profile likelihoods show full non-identifiability of the poorly identified parameter, and good identifiability of the well-identified parameter. This is consistent with the expected behaviour of the reduced, structurally non-identifiable model.

Finally, we consider the prediction results for the full model. For both the original and reparameterised model we show both the profile-wise prediction intervals [38] capturing the uncertainty in the prediction driven approximately solely by the target parameter, and the joint prediction intervals, capturing the uncertainty in the prediction due to both parameters. For both profile-wise and joint prediction intervals we use the 95% confidence level at 2 degrees of freedom in order to compare the ability of individual parameters to capture the joint uncertainty. The choice of degrees of freedom for prediction is investigated in more detail in [69]. These results are shown in Figure 11 and 12. We see that in the reparameterised model the profile-wise prediction interval driven by the *well-identified* parameter captures more of the uncertainty in the prediction, while the poorly identified parameter has less effect, while in the original model both parameters contribute more equally to the prediction uncertainty. This effect becomes more pronounced the more non-identified the poorly-identified parameter becomes (see, e.g., the next example). Thus the reparameterised model supports model reduction in the predictive sense as well, at least approximately. Importantly, however, this is ‘within sample’ prediction, and the uncertainty in genuinely new domains may depend more strongly on the poorly identified parameter. This is an important caveat to the use of reduced models, and highlights the importance of understanding the presence of non-identifiability in the full model.



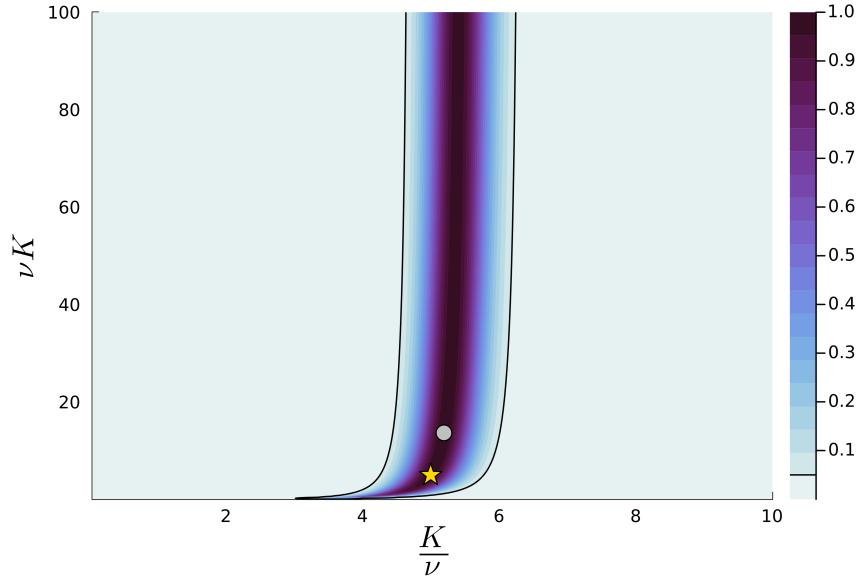
(a) Joint likelihood contour plot in the (ν, K) plane. The gold star indicates the true parameter value, and the silver circle indicates the maximum likelihood estimate. The colour bar represents the relative likelihood, with a black contour line indicating the 95% confidence interval threshold.



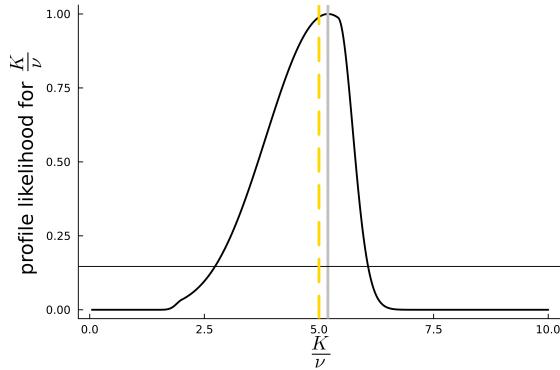
(b) Profile likelihood for ν . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

(c) Profile likelihood for K . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

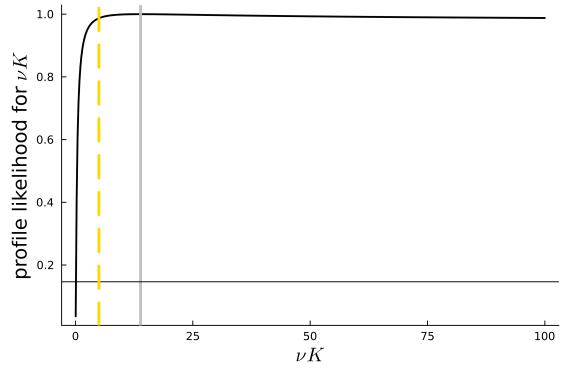
Figure 8: Michaelis-Menten/Monod example: (Top) joint contour in the (ν, K) plane. (Bottom) profile likelihoods.



(a) Joint likelihood contour plot in the $(\frac{K}{\nu}, \nu K)$ plane. The gold star indicates the true parameter value, and the silver circle indicates the maximum likelihood estimate. The colour bar represents the relative likelihood, with a black contour line indicating the 95% confidence interval threshold.

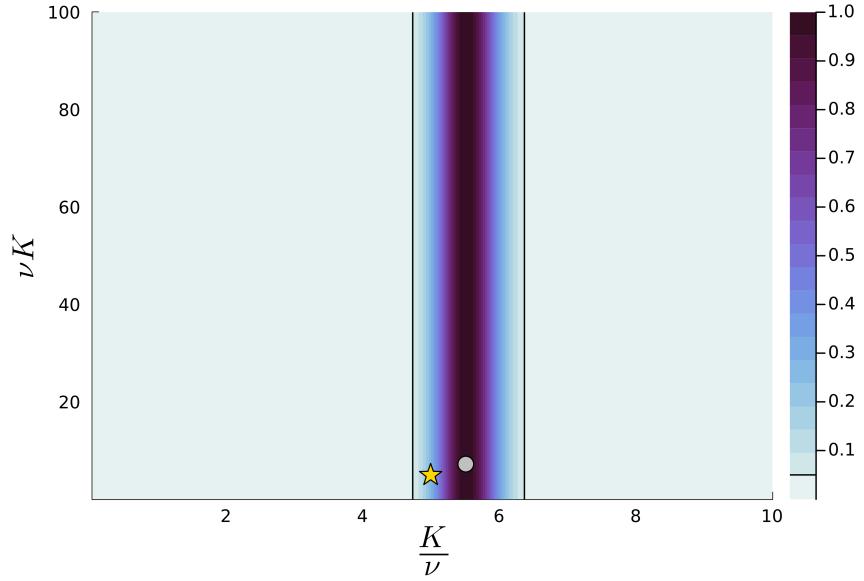


(b) Profile likelihood for $\frac{K}{\nu}$. The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

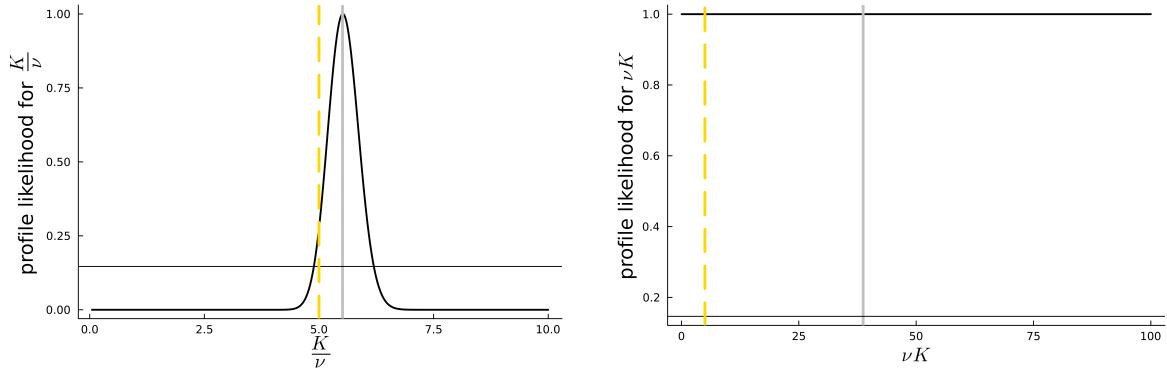


(c) Profile likelihood for νK . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

Figure 9: Reparameterised Michaelis-Menten/Monod example: (Top) joint contour in the $(\frac{K}{\nu}, \nu K)$ plane. (Bottom) profile likelihoods.



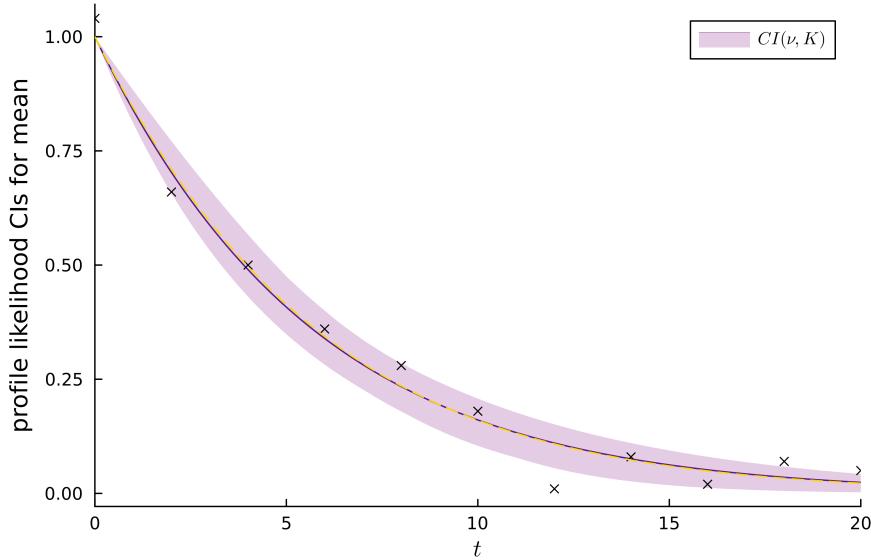
(a) Joint likelihood contour plot in the $(\frac{K}{\nu}, \nu K)$ plane. The gold star indicates the true parameter value, and the silver circle indicates the maximum likelihood estimate. The colour bar represents the relative likelihood, with a black contour line indicating the 95% confidence interval threshold.



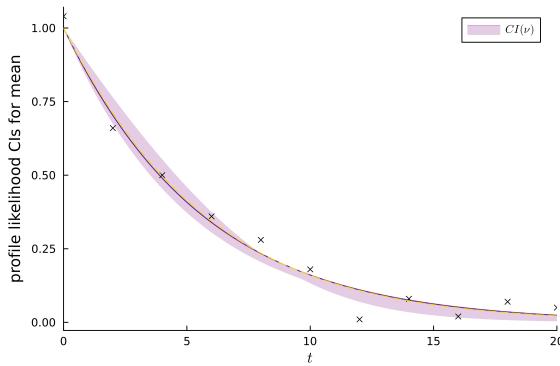
(b) Profile likelihood for $\frac{K}{\nu}$. The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

(c) Profile likelihood for νK . The solid silver vertical line indicates the maximum likelihood estimate, and the dashed gold vertical line indicates the true parameter value. The horizontal black line indicates the threshold for the 95% confidence interval.

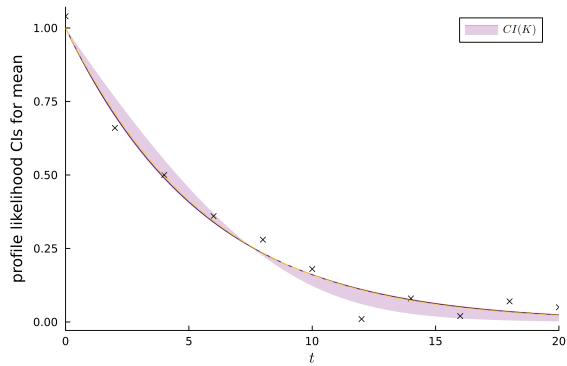
Figure 10: Reparameterised Michaelis-Menten/Monod limit example: (Top) joint contour in the $(\frac{K}{\nu}, \nu K)$ plane. (Bottom) profile likelihoods.



(a) Joint prediction interval for the full model based on original (ν, K) parameterisation. The gold trajectory indicates the true mean function, and the dark purple trajectory indicates the maximum likelihood estimate. The purple shaded region represents the 95% (predictive) confidence interval for the mean.

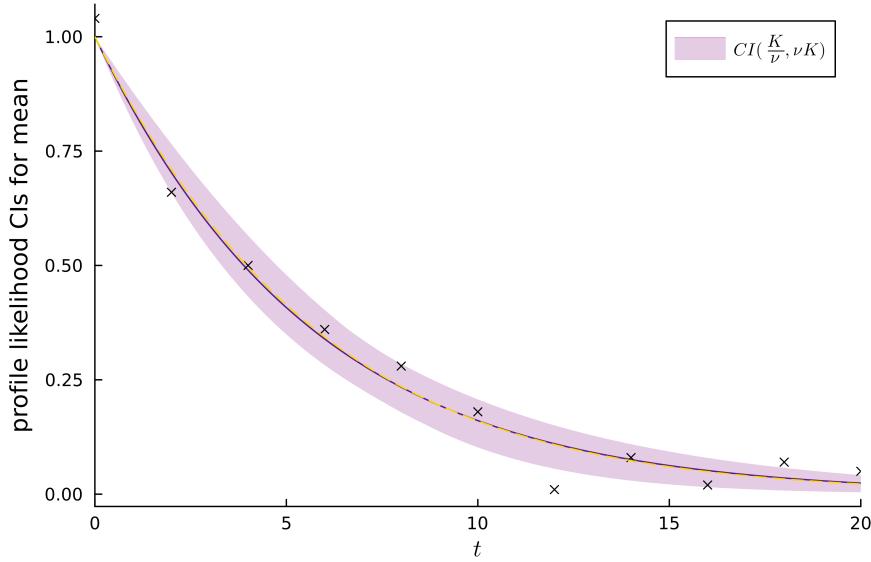


(b) Profile-wise prediction interval based on ν . The gold trajectory indicates the true mean function, and the dark purple trajectory indicates the maximum likelihood estimate. The shaded purple region represents the approximate (profile-wise) 95% (predictive) confidence interval for the mean.

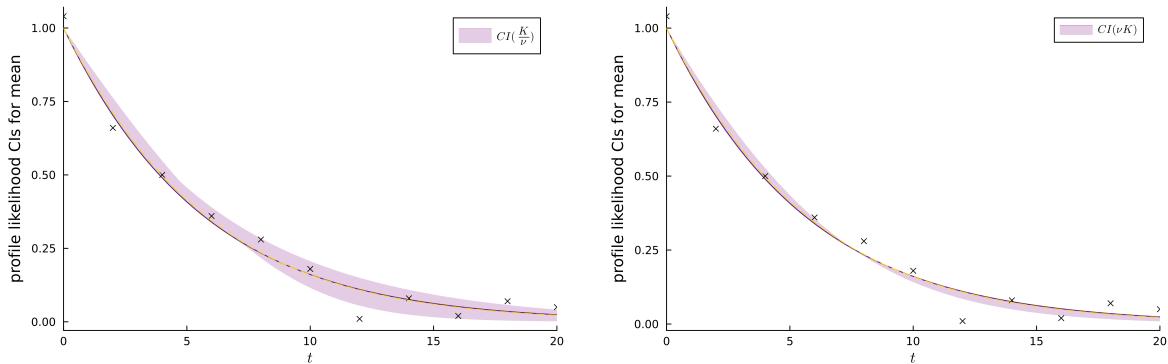


(c) Profile-wise prediction interval based on K . The gold trajectory indicates the true mean function, and the dark purple trajectory indicates the maximum likelihood estimate. The shaded purple region represents the approximate (profile-wise) 95% (predictive) confidence interval for the mean.

Figure 11: Predictive confidence intervals for the mean in the Michaelis-Menten/Monod example: (Top) joint predictive confidence interval for the mean in terms of the original (ν, K) parameterisation. (Bottom) profile-wise prediction intervals for each parameter.



(a) Joint prediction interval for the full model based on the $(\frac{K}{\nu}, \nu K)$ parameterisation. The gold trajectory indicates the true mean function, and the dark purple trajectory indicates the maximum likelihood estimate. The shaded purple region represents the 95% (predictive) confidence interval for the mean.



(b) Profile-wise prediction interval based on $\frac{K}{\nu}$. The gold trajectory indicates the true mean function, and The dark purple indicates the maximum likelihood estimate. The shaded purple region represents the approximate (profile-wise) 95% (predictive) confidence interval for the mean.

(c) Profile-wise prediction interval based on νK . The gold trajectory indicates the true mean function, and The dark purple indicates the maximum likelihood estimate. The shaded purple region represents the approximate (profile-wise) 95% (predictive) confidence interval for the mean.

Figure 12: Predictive confidence intervals for the mean in the reparameterised Michaelis-Menten/Monod example: (Top) joint predictive confidence interval for the mean in terms of the $(\frac{K}{\nu}, \nu K)$ reparameterisation. (Bottom) profile-wise prediction intervals for each parameter.

4.3 Flow in composite media

Here we consider the groundwater model, representing flow in composite media, defined by equation 32. As shown in the model section, this model is structurally non-identifiable, and can be written in terms of a reduced set of monomial functions of the parameters. Rather than consider further analytical results, we proceed directly to the numerical analysis of the model and verify that our numerical approach can uncover the non-identifiability of the model and suggest a reduced parameterisation.

For these results we used a true parameter value of $T_1 = 3.0$, $T_2 = 1.0$, and $R = 1.0$. We considered a fine grid of 201 points between $x = 0$ and $x = L = 100$, and an observation grid of 19 equally-spaced points, excluding the endpoints. We used a normal observation model with standard deviation $\sigma = 0.2$ and a single, spatial sample (i.e. a single observation of the model solution at the observation points, giving a data vector of length 19). For reproducibility, the data realisation used was [186.4, 402.6, 505.2, 756.1, 1144.1, 790.9, 1283.5, 1647.6, 872.3, 1144.4, 1691.2, 1352.7, 1519.9, 1316.0, 1437.1, 726.7, 952.3, 759.4, 272.0]. repository at `examples/transport_model.jl`.

Figure 13 shows the results of the analysis in the original parameterisation. For visualisation we consider each of the three joint two-dimensional likelihood profiles with the third parameter ‘profiled out’. We see that the joint profile likelihood contours in the (T_1, T_2) plane ..., the joint profile likelihood contours in the (T_1, R) plane ..., and the joint profile likelihood contours in the (T_2, R) plane We also show the one-dimensional profile likelihoods for each parameter, which show ...

Figure 14 shows the results of the analysis in the reparameterised space. Interestingly, the reparameterisation suggested by the SVD is $(\frac{T_2}{R}, \frac{T_1}{\sqrt{T_2}R}, T_1 T_2 R)$, with the first having a singular value two-three orders of magnitude larger than the second, and the third being numerically equivalent to zero (order 10^{-14}). The first two combinations can be written as a one-to-one mapping to the reduced parameter pair that appear in the analytical results, i.e. to $(\frac{T_1}{R}, \frac{T_1}{R})$, but the SVD prefers the alternative reparameterisation. This reflects the fact that the SVD aims to find an orthogonal (here in log-parameter space) reparameterisation in which the information on each parameter combination is maximally separated. Both pairs of parameter combinations are orthogonal to the non-identified parameter combination $T_1 T_2 R$ in log-space, and are linearly independent, and hence provide a valid basis for the identified subspace. The SVD parameterisation further gives orthgonality of the combinations spanning the two-dimensional identified subspace and provides the maximal separation of information in this sense. In the present case we find that the two parameterisations lead

to visually indistinguishable results, and so only show the results of the SVD reparameterisation. This also allows us to assign a ‘degree of identifiability’ to the parameters, finding that $\frac{T_2}{R}$ is much more well-identified than $\frac{T_1}{\sqrt{T_2 R}}$ (and $\frac{T_1}{R}$), and that $T_1 T_2 R$ is completely non-identified. This quantification is not obvious just from the analytical model solution.

Finally, we show the profile-wise prediction intervals for the full model in the original parameterisation in Figure 15, and in the reparameterised space in Figure 16. We see that ...

We see that the joint profile likelihood contours in the ... plane ..., the joint profile likelihood contours in the ... plane ..., and the joint profile likelihood contours in the ... plane We also show the one-dimensional profile likelihoods for each parameter, which show ...

Finally, we show the profile-wise prediction intervals for ... in the original parameterisation in Figure 15, and in the reparameterised space in Figure 16. We see that ...

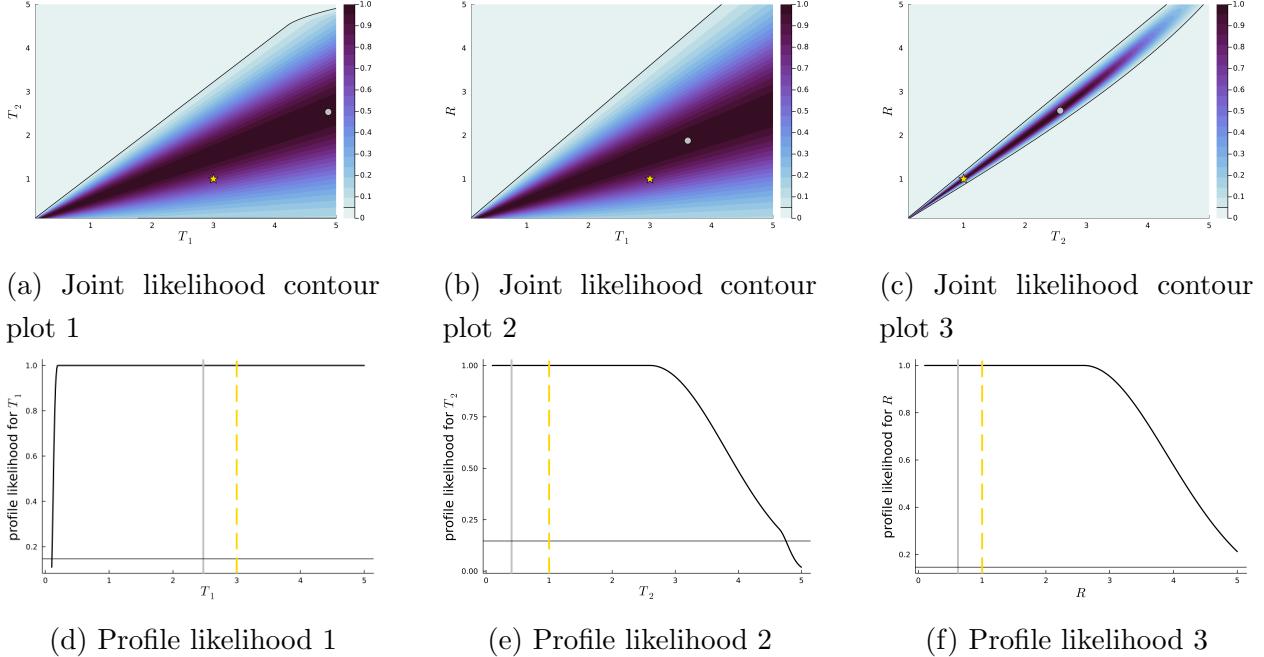


Figure 13: Caption describing all six subfigures

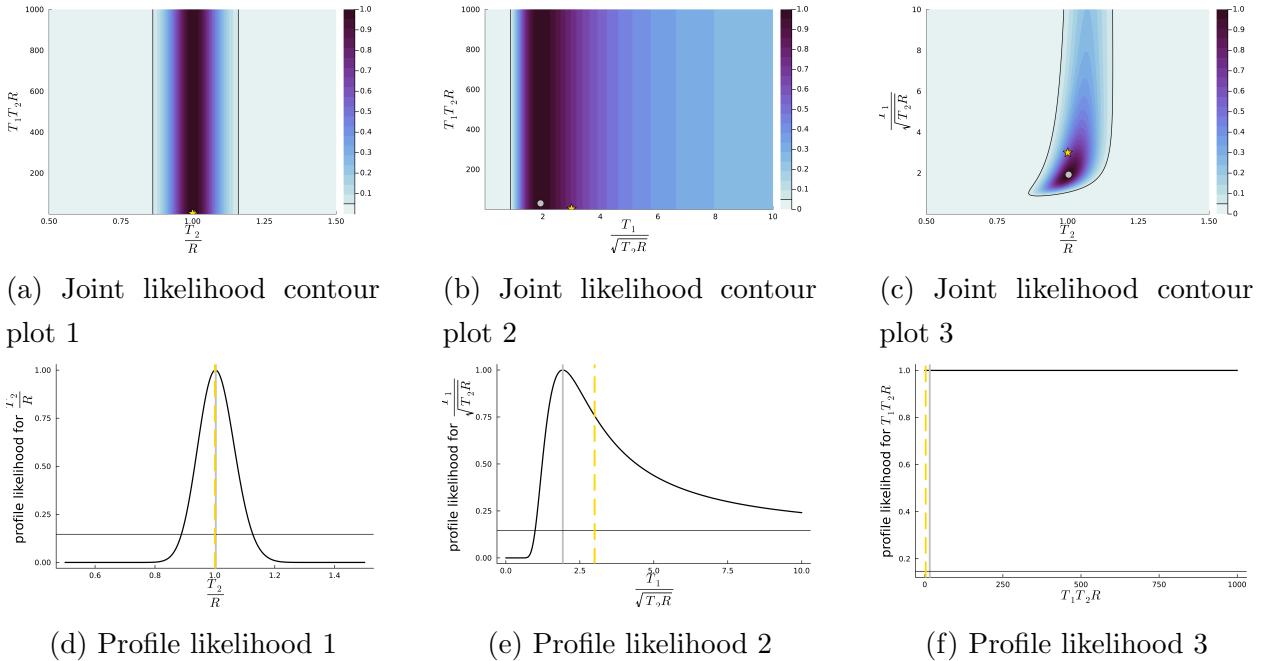


Figure 14: Caption describing all six subfigures

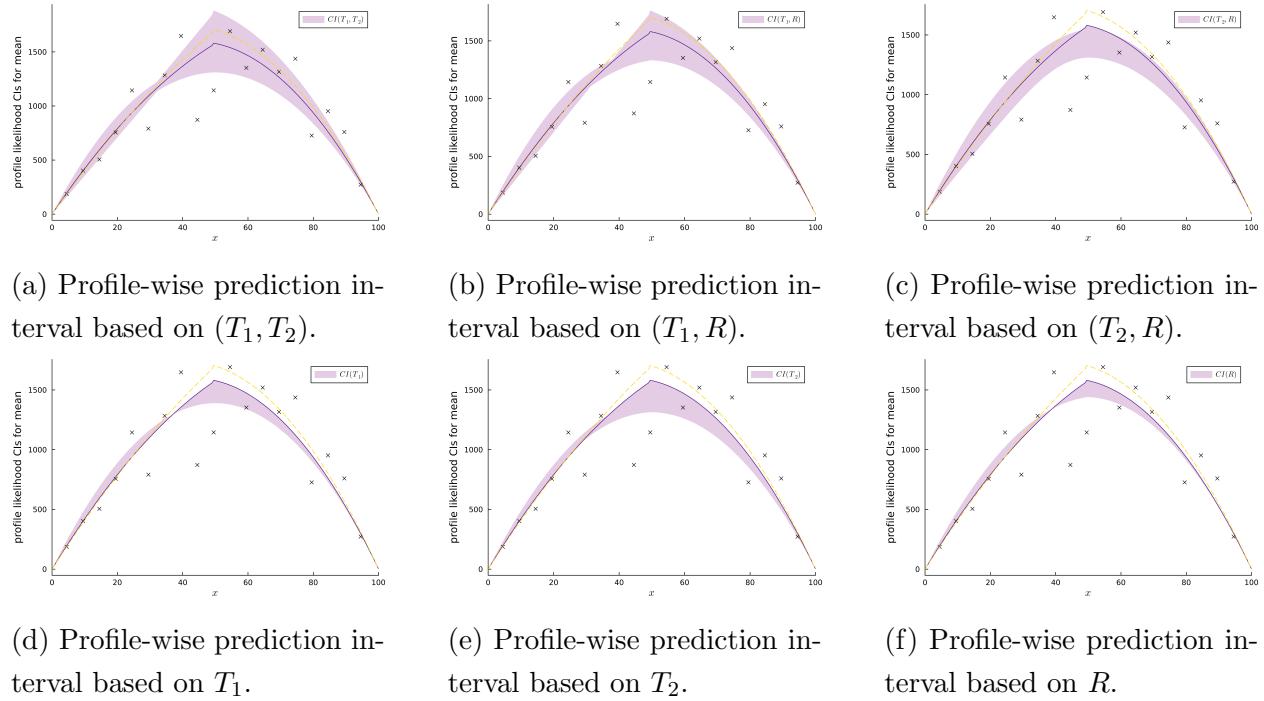


Figure 15: Caption describing all six subfigures

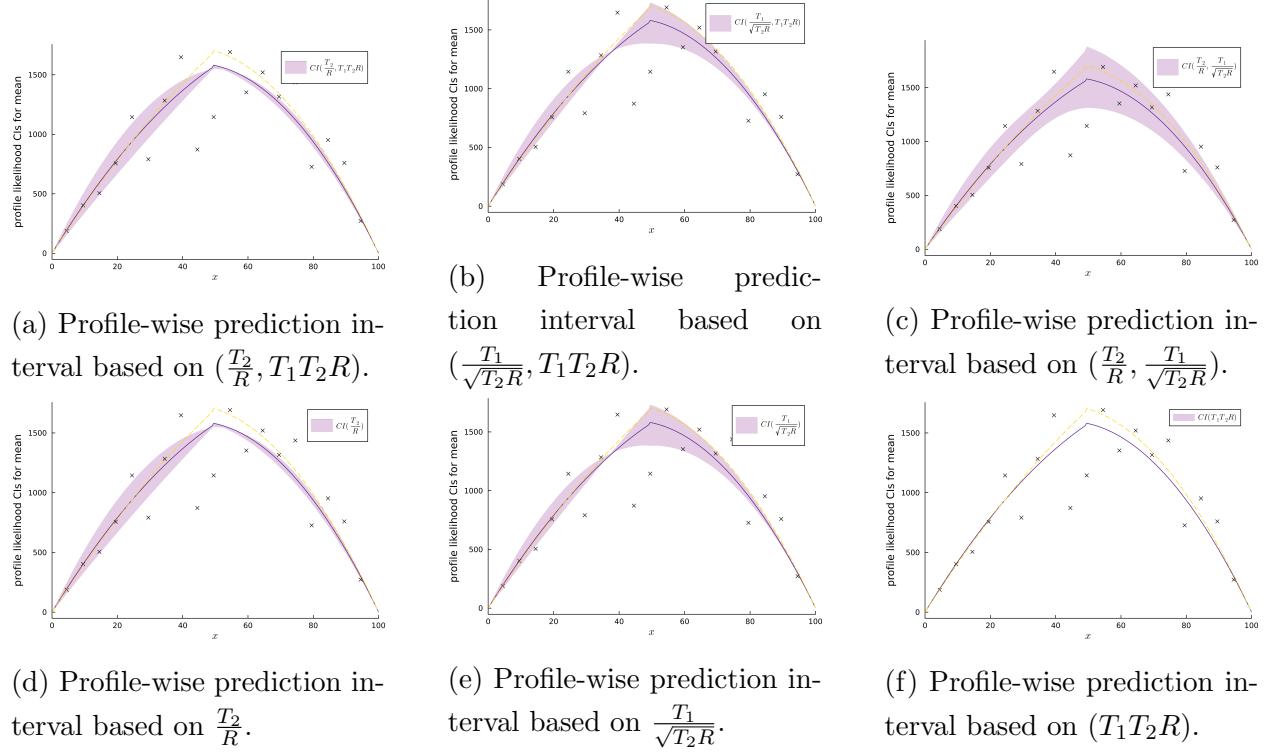


Figure 16: Caption describing all six subfigures

5 Conclusion and Future Work

- Here we deal with ODE/PDE type models, but the concepts apply to SDE and simulation-based stochastic models too
- Here we deal with all parameters at once, but we can deal with subsets of parameters depending on our problem of interest.

6 Acknowledgements

Funding: This work is supported by the Australian Research Council Centre of Excellence for Mathematical Analysis of Cellular Systems grant no. CE230100001 (MJS).

7 Author Contributions

All authors contributed to the conceptualisation, review and editing of the manuscript. OJM wrote the original draft with contributions from MJS. Both OJM and MJS contributed to the choice of examples and design of numerical experiments. OJM developed the method details, proofs, wrote the code, and performed the numerical experiments, with discussion and input from the other authors.

8 Competing interests

The authors declare that they have no competing interests.

9 Data and materials availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary material. Julia code associated with this paper is available from GitHub at <https://github.com/omaclaren/reparam>.

References

1. Gould PL and Feng Y. Introduction to Linear Elasticity. Springer, 2018
2. Howell P, Kozyreff G, and Ockendon J. Applied Solid Mechanics. Cambridge University Press, 2008
3. Bear J. Dynamics of Fluids in Porous Media. American Elsevier Publishing Company, 1972
4. Krantz WB. Scaling Analysis in Modeling Transport and Reaction Processes. Wiley Interscience, 2007
5. Kot M. Elements of Mathematical Ecology. Cambridge University Press, 2001
6. Murray J. Mathematical Biology: I An Introduction. Springer, 2002
7. Horn F and Jackson R. General mass action kinetics. Archive for Rational Mechanics and Analysis 1972; 47:81–116. DOI: 10.1007/BF00251225

8. Hendry DF and Morgan MS. The foundations of econometric analysis. Cambridge University Press, 1997
9. Bellu G, Saccomani MP, Audoly S, and D'Angiò L. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine* 2007; 88:52–61. DOI: [10.1016/j.cmpb.2007.07.002](https://doi.org/10.1016/j.cmpb.2007.07.002)
10. Chiş O, Villaverde AF, Banga JR, and Balsa-Canto E. On the relationship between sloppiness and identifiability. *Mathematical Biosciences* 2016; 282:147–61. DOI: [10.1016/j.mbs.2016.10.009](https://doi.org/10.1016/j.mbs.2016.10.009)
11. Ligon TS, Frölich F, Chiş O, Banga JR, Balsa-Canto E, and Hasenauer J. GenSSI 2.0: multi-experimental structural identifiability analysis of SBML models. *Bioinformatics* 2018; 34:1421–3. DOI: [10.1093/bioinformatics/btx735](https://doi.org/10.1093/bioinformatics/btx735)
12. Meshkat N, Eisenberg M, and DiStefano III JJ. An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner Bases. *Mathematical Biosciences* 2009; 222:61–72. DOI: <https://doi.org/10.1016/j.mbs.2009.08.010>
13. Meshkat N, Sullivant S, and Eisenberg M. Identifiability results for several classes of linear compartment models. *Bulletin of Mathematical Biology* 2015; 77:1620–51. DOI: [10.1007/s11538-015-0098-0](https://doi.org/10.1007/s11538-015-0098-0)
14. Díaz-Seoane S, Rey Barreiro X, and Villaverde AF. STRIKE-GOLDD 4.0: User-friendly, efficient analysis of structural identifiability and observability. *Bioinformatics* 2022; 39:btac748. DOI: [10.1093/bioinformatics/btac748](https://doi.org/10.1093/bioinformatics/btac748)
15. Maclarens OJ and Nicholson R. What can be estimated? Identifiability, estimability, causal inference and ill-posed inverse problems. *arxiv preprint* 2019; 1904.02826. DOI: [10.48550/arXiv.1904.02826](https://doi.org/10.48550/arXiv.1904.02826)
16. Hines KE, Middendorf TR, and Aldrich RW. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. *Journal of General Physiology* 2014; 143:401. DOI: [10.1085/jgp.201311116](https://doi.org/10.1085/jgp.201311116)
17. Wieland FG, Hauber AL, Rosenblatt M, C T, and Timmer J. On structural and practical identifiability. *Current Opinion in Systems Biology* 2021; 25. DOI: [10.1016/j.coisb.2021.03.00](https://doi.org/10.1016/j.coisb.2021.03.00)
18. Kreutz C, Raue A, and J T. Likelihood-based observability analysis and confidence intervals for predictions of dynamics models. *BMC Systems Biology* 2012; 6. DOI: [10.1186/1752-0509-6-120](https://doi.org/10.1186/1752-0509-6-120)

19. Simpson MJ and Maclaren OJ. Making predictions from nonidentifiable models. *Bulletin of Mathematical Biology* 2024; 86:80. doi: [10.1007/s11538-024-01294-0](https://doi.org/10.1007/s11538-024-01294-0)
20. Cole DJ, Morgan BJT, and Titterington DM. Determining the parameteric structure of models. *Mathematical Biosciences* 2010; 228:16–30. doi: [10.1016/j.mbs.2010.08.004](https://doi.org/10.1016/j.mbs.2010.08.004)
21. Cole D. Parameter redundancy and identifiability. CRC Press, 2020
22. Maini PK, McElwain DLS, and Leavesley DI. Traveling wave model to interpret a wound-healing cell migration assay for Human Peritoneal Mesothelial cells. *Tissue Engineering* 2004; 10:475–82. doi: [10.1089/107632704323061834](https://doi.org/10.1089/107632704323061834)
23. Simpson MJ and McCue SW. Fisher–KPP-type models of biological invasion: open source computational tools, key concepts and analysis. *Proceedings of the Royal Society A: Mathematics, Physics and Engineering Sciences* 2024; 480:20240186. doi: [10.1098/rspa.2024.0186](https://doi.org/10.1098/rspa.2024.0186)
24. Chappell MJ and Gunn RN. A procedure for generating locally identifiable reparameterisations of unidentifiable non-linear systems by the similarity transform approach. *Mathematical Biosciences* 1998; 148:21–41. doi: [10.1016/S0025-5564\(97\)10004-9](https://doi.org/10.1016/S0025-5564(97)10004-9)
25. Chiş O, Banga JR, and Balsa-Canto E. Structural identifiability of systems biology models: a critical comparison of methods. *Plos One* 2011; 6:e27755. doi: [10.1371/journal.pone.0027755](https://doi.org/10.1371/journal.pone.0027755)
26. Kreutz C, Raue A, Kaschek D, and Timmer J. Profile likelihood in systems biology. *The FEBS Journal* 2013; 280:2564–71. doi: [10.1111/febs.12276](https://doi.org/10.1111/febs.12276)
27. Raue A, Kreutz C, Theis FJ, and Timmer J. Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2013; 271:20110544. doi: [10.1098/rsta.2011.0544](https://doi.org/10.1098/rsta.2011.0544)
28. Raue A, Karlsson J, Saccomani MP, Jirstrand M, and Timmer J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 2014; 30:1440–8. doi: [10.1093/bioinformatics/btu006](https://doi.org/10.1093/bioinformatics/btu006)
29. Villaverde AF, Raimúndez E, Hasenauer J, and Banga JR. Assessment of prediction uncertainty quantification methods in systems biology. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2023; 30:1725–36. doi: [10.1109/TCBB.2022.3213914](https://doi.org/10.1109/TCBB.2022.3213914)

30. Simpson MJ, Baker RE, Vittadello ST, and Maclare OJ. Parameter identifiability analysis for spatiotemporal models of cell invasion. *Journal of the Royal Society Interface* 2020; 17. DOI: [10.1098/rsif.2020.0055](https://doi.org/10.1098/rsif.2020.0055)
31. Simpson MJ, Browning AP, Warne DJ, Maclare OJ, and Baker RE. Parameter identifiability and model selection for sigmoid population growth models. *Journal of Theoretical Biology* 2022; 535:1100998. DOI: [10.1016/j.jtbi.2021.110998](https://doi.org/10.1016/j.jtbi.2021.110998)
32. Pawitan Y. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press, 2001
33. Maiwald T, Hass H, Steiert B, Vanlier J, Engesser R, Raue A, Kipkeew F, Bock HH, Kaschek D, Kreutz C, and Timmer J. Driving the model to its limit: Profile likelihood based model reduction. *Plos One* 2016; 11:e0162366. DOI: [10.1371/journal.pone.0162366](https://doi.org/10.1371/journal.pone.0162366)
34. Brown KS and Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E* 2003; 68:021904. DOI: [10.1103/PhysRevE.68.021904](https://doi.org/10.1103/PhysRevE.68.021904)
35. Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, and Cerione RA. The statistical mechanics of complex signaling networks: nerve growth factor signalling. *Physical Biology* 2004; 1:184. DOI: [10.1088/1478-3967/1/3/006](https://doi.org/10.1088/1478-3967/1/3/006)
36. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, and Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLOS Computational Biology* 2007; 3:e189. DOI: [10.1371/journal.pcbi.0030189](https://doi.org/10.1371/journal.pcbi.0030189)
37. Monsalve-Bravo GM, Lawson BAJ, Drovandi C, Burrage K, Brown KS, Baker CM, Vollert SA, Mengersen K, McDonald-Madden E, and Adams MP. Analysis of sloppiness in model simulations: unveiling parameter uncertainty when mathematical models are fitted to data. *Science Advances* 2022; 8:eabm5952. DOI: [10.1126/sciadv.abm5952](https://doi.org/10.1126/sciadv.abm5952)
38. Simpson MJ and Maclare OJ. Profile-Wise Analysis: A profile likelihood-based workflow for identifiability analysis, estimation, and prediction with mechanistic mathematical models. *PLOS Computational Biology* 2023; 19:e1011515. DOI: [10.1371/journal.pcbi.1011515](https://doi.org/10.1371/journal.pcbi.1011515)
39. Lawrie J and Hearne J. Reducing model complexity via output sensitivity. *Ecological Modelling* 2007; 207:137–44. DOI: [10.1016/j.ecolmodel.2007.04.013](https://doi.org/10.1016/j.ecolmodel.2007.04.013)

40. Elevitch CR and Johnson Jr CR. A procedure for ranking parameter importance for estimation in predictive mechanistic models. *Ecological Modelling* 2020; 419:108948. DOI: 10.1016/j.ecolmodel.2020.108948
41. Vollert SA, Drovandi C, Monsalve-Bravo GM, and Adams MP. Strategic model reduction by analysing model sloppiness: A case study in coral calcification. *Environmental Modelling & Software* 2023; 159:105578. doi: 10.1016/j.envsoft.2022.105578
42. Maclaren OJ and Nicholson R. Models, identifiability, and estimability in causal inference. *Workshop on the Neglected Assumptions in Causal Inference (NACI) at the 38 th International Conference on Machine Learning, 2021.* 2021
43. Lee J. Introduction to Smooth Manifolds. Graduate Texts in Mathematics. Springer New York, 2013. Available from: <https://books.google.co.nz/books?id=w4bhBwAAQBAJ>
44. Pace L and Salvan A. Principles of Statistical Inference from a Neo-Fisherian Perspective. World Scientific, 1997
45. Cox DR. Principles of statistical inference. Cambridge University Press, 2006
46. Murphy RJ, Maclaren OJ, and Simpson MJ. Implementing measurement error models with mechanistic mathematical models in a likelihood-based framework for estimation, identifiability analysis and prediction in the life sciences. *Journal of the Royal Society Interface* 2024; 21:20230402. doi: 10.1098/rsif.2023.0402
47. Simpson MJ, Murphy RJ, and Maclaren OJ. Modelling count data with partial differential equation models in biology. *Journal of Theoretical Biology* 2024; 580:111732. DOI: 10.1016/j.jtbi.2024.111732
48. Revels J, Lubin M, and Papamarkou T. Forward-mode automatic differentiation in Julia. arXiv preprint arXiv:1607.07892 2016
49. Magnus JR and Neudecker H. Matrix differential calculus with applications in statistics and econometrics. John Wiley & Sons, 2019
50. Lawvere FW and Rosebrugh R. Sets for mathematics. Cambridge University Press, 2003
51. Buckingham E. On physically similar systems; illustrations of the use of dimensional equations. *Physical review* 1914; 4:345
52. Barenblatt GI. Scaling. Cambridge University Press, 2003

53. Griewank A and Walther A. Evaluating derivatives: principles and techniques of algorithmic differentiation. SIAM, 2008
54. Aitkin M and Stasinopoulos M. Likelihood analysis of a binomial sample size problem. Contributions to probability and statistics: Essays in honor of Ingram Olkin 1989 :496–9
55. Cox DR and Reid N. Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society: Series B (Methodological) 1987; 49:1–18
56. Olkin I, Petkau AJ, and Zidek JV. A comparison of n estimators for the binomial distribution. Journal of the American Statistical Association 1981; 76:637–42
57. Aitkin M. Statistical inference: an integrated Bayesian/likelihood approach. CRC Press, 2010
58. Monod J. The growth of bacterial cultures. Annual Review of Microbiology 1949; 3:371–94. DOI: <https://doi.org/10.1146/annurev.mi.03.100149.002103>
59. Johnson KA and Goody RS. The original Michaelis constant: Translation of the 1913 Michaelis–Menten Paper. Biochemistry 2011; 50:8264–9. DOI: 10.1021/bi201284u
60. Fitts CR. Groundwater science. Elsevier, 2002
61. Rushton KR. Groundwater hydrology: conceptual and computational models. John Wiley & Sons, 2004
62. Carslaw H and Jaeger J. Conduction of Heat in Solids. Oxford science publications. Clarendon Press, 1959
63. Hahn DW and Özisik MN. Heat conduction. John Wiley & Sons, 2012
64. Stakgold I. Boundary value problems of mathematical physics: volume 1. SIAM, 2000
65. Bird R, Stewart W, and Lightfoot E. Transport Phenomena. Transport Phenomena v. 1. Wiley, 2006
66. Davison AC, Fraser DA, and Reid N. Improved likelihood inference for discrete data. Journal of the Royal Statistical Society Series B: Statistical Methodology 2006; 68:495–508
67. Holmberg A. On the practical identifiability of microbial growth models incorporating Michaelis–Menten type nonlinearities. Mathematical Biosciences 1982; 62:23–43
68. Crank J. The mathematics of diffusion. Oxford university press, 1979

69. Trent J. Likelihood-based computational analysis and uncertainty quantification for mechanistic models. MA thesis. University of Auckland, 2024. Available from: <https://github.com/JoeTrent/LikelihoodBasedProfileWiseAnalysis.jl>