



OPEN

Estimating global identifiability using conditional mutual information in a Bayesian framework

Sahil Bhola & Karthik Duraisamy

A novel information-theoretic approach is proposed to assess the global practical identifiability of Bayesian statistical models. Based on the concept of conditional mutual information, an estimate of information gained for each model parameter is used to quantify the identifiability with practical considerations. No assumptions are made about the structure of the statistical model or the prior distribution while constructing the estimator. The estimator has the following notable advantages: first, no controlled experiment or data is required to conduct the practical identifiability analysis; second, unlike popular variance-based global sensitivity analysis methods, different forms of uncertainties, such as model-form, parameter, or measurement can be taken into account; third, the identifiability analysis is global, and therefore independent of a realization of the parameters. If an individual parameter has low identifiability, it can belong to an identifiable subset such that parameters within the subset have a functional relationship and thus have a combined effect on the statistical model. The practical identifiability framework is extended to highlight the dependencies between parameter pairs that emerge a posteriori to find identifiable parameter subsets. The applicability of the proposed approach is demonstrated using a linear Gaussian model and a non-linear methane-air reduced kinetics model. It is shown that by examining the information gained for each model parameter along with its dependencies with other parameters, a subset of parameters that can be estimated with high posterior certainty can be found.

With the growth in computational capabilities, statistical models are becoming increasingly complex to make predictions under various design conditions. These models often contain uncertain parameters which must be estimated using data obtained from controlled experiments. While methods for parameter estimation have matured significantly, there remain notable challenges for a statistical model and estimated parameters to be considered reliable. One such challenge is the *practical identifiability* of model parameters which is defined as the possibility of estimating each parameter with high confidence given different forms of uncertainties, such as parameter, model-form, or measurement are present^{1–3}. Low practical identifiability of the statistical model can lead to an ill-posed estimation problem which becomes a critical issue when the parameters have a physical interpretation and decisions are to be made using their estimated values^{4,5}. Further, such identifiability deficit can also lead to an unreliable model prediction, and therefore such statistical models are not suitable for practical applications^{6,7}. Therefore, for a reliable parameter estimation process and model prediction, it is of significant interest that the practical identifiability is evaluated before any controlled experiment or parameter estimation studies are conducted^{8,9}.

In frequentist statistics, the problem of practical identifiability is to examine the possibility of *unique* estimation of model parameters θ ⁸. Under such considerations, methods examining identifiability are broadly classified into *local* and *global* identifiability methods. While the former examines the possibility that $\theta = \theta^k$ is a unique parameter estimate within its neighborhood $N(\theta^k)$ in the parameter space, the latter is concerned with the uniqueness of θ^k when considering the entire parameter space. Local sensitivity analysis has been widely used to find parameters that produce large variability in the model response^{10–13}. In such an analysis, parameters resulting in large variability are considered relevant and therefore assumed to be identifiable for parameter estimation. However, the parameters associated with large model sensitivities could still have poor identifiability characteristics¹⁴. Another class of frequentist identification methods is based on the analysis of the properties

Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, USA. email: sbhola@umich.edu

of the Fisher information matrix (FIM). Staley et al.¹⁵ proposed that the positive definiteness of the FIM is the necessary and sufficient condition for the parameters to be considered practically identifiable. Similarly, Rothenberg¹⁶ showed that the identifiability of the parameters is equivalent to the non-singularity of the FIM. However, subsequent findings^{17,18} have reported that models with singular FIM could also be identifiable. Weijers et al.¹⁹ extended the classical FIM analysis and showed that even if an individual parameter has low identifiability it can belong to an identifiable subset, such that the subset is practically identifiable. Parameters within such subsets have functional relationships with each other, thus resulting in a combined effect on the model response. It has been shown that such identifiable subsets can be found by examining the condition number (E -criterion) and determinant (D -criterion), and selecting parameter pairs with the smallest condition number and largest determinant. Similarly, Machado et al.²⁰ considered the D to E ratio to examine practical identifiability and to find the identifiable subsets. Another popular identification technique is likelihood profiling^{4,21–23}. The method is based on finding the likelihood profile of a parameter by maximizing the likelihood with respect to the rest of the parameters. Parameters for which its likelihood profile is shallow are deemed to have low practical identifiability. In addition to evaluating practical identifiability, likelihood profiling could also be used to find functional relationships between parameters, which is helpful for model reparameterization^{24,25}. However, due to the several re-optimizations required to obtain the likelihood profiles, the method does not scale well with parameter space and could quickly become computationally intractable. While methods based on FIM or likelihood profiling have gained significant popularity they only examine local identifiability. This means that the estimate of practical identifiability is dependent on θ^k for which the analysis is conducted and is only valid within its neighborhood $N(\theta^k)$. To overcome the limitations of local identifiability, global identifiability methods using Kullback-Leibler divergence²⁶ and identifying functions²⁷ have been proposed. However, such methods are computationally complex and not suitable for practical problems. Moreover, since such methods are based on frequentist statistics, they are unable to account for parametric uncertainty and therefore unable to provide an honest representation of global practical identifiability.

There have been few studies examining global practical identifiability in a Bayesian framework. Early attempts were based on global sensitivity analysis (GSA) that apportions the variability (either by derivatives or variance) of the model output due to the uncertainty in each parameter^{6,28–30}. Unlike local sensitivity analysis, GSA-based methods simultaneously vary model parameters according to their distributions, thus providing a measure of global sensitivity that is independent of a particular parameter realization. However, global parameter sensitivity does not guarantee global practical identifiability³¹. Pant et al.³² and Capellari et al.³³ formulated the problem of practical identifiability as gaining sufficient information about each model parameter from data. An information-theoretic approach was used to quantify the information gained, such that larger information gain would mean larger practical identifiability. However, assumptions about the structure of parameter-data joint distribution were made when developing the estimator. A similar approach was used by Ebrahimian et al.³⁴ where the change in parameter uncertainty moving from the prior distribution to the estimated posterior distribution was used to quantify information gained. Pant³⁵ proposed information sensitivity functions by combining information theory and sensitivity analysis to quantify information gain. However, the joint distribution between the parameters and the data was assumed to be Gaussian.

Framed in a Bayesian setting, the information-theoretic approach to identifiability provides a natural extension to include different forms of uncertainties that are present in practical problems. In this work, a novel estimator is developed from an information-theoretic perspective to examine the practical identifiability of a statistical model. The expected information gained from the data for each model parameter is used as a metric to quantify practical identifiability. In contrast to the aforementioned methods based on information theory, the proposed approach has the following novel advantages: first, the estimator for information gain can be used for an a priori analysis, that is, no data is required to evaluate practical identifiability; second, the framework can account for different forms of uncertainty, such as model-form, parameter, and measurement; third, the framework does not make assumptions about the joint distribution between the data and parameters as in the previous methods; fourth, the identifiability analysis is global, rather than being dependent on a particular realization of model parameters. Another contribution of this work is an information-theoretic estimator to highlight dependencies between parameter pairs that emerge a posteriori, however, in an a priori manner. Combining the knowledge about information gained about each parameter and parameter dependencies using the proposed approach, it is possible to find parameter subsets that can be estimated with high posterior certainty before any controlled experiment is performed. Broadly, this can dramatically reduce the cost of parameter estimation, inform model-form selection or refinement, and associate a degree of reliability to the parameter estimation.

The manuscript is organized as follows. In "[Bayesian parameter inference](#)" the Bayesian paradigm for parameter estimation is presented. In "[Quantifying information gain](#)" differential entropy and mutual information are presented as information-theoretic tools to quantify the uncertainty associated with random variables and information gain, respectively. In "[Estimating practical identifiability](#)" an a priori estimator is developed to quantify global practical identifiability in a Bayesian construct. In "[Estimating parameter dependence](#)" the problem of estimating parameter dependencies is addressed. An a priori estimator is developed to quantify parameter dependencies developed a posteriori. The practical identifiability framework is applied to a linear Gaussian statistical model and methane-air reduced kinetics model; results are presented in "[Numerical experiments](#)". Concluding remarks are presented in "[Concluding remarks and perspectives](#)".

Quantifying practical identifiability in a Bayesian setting

In this section, we first present the Bayesian framework for parameter estimation. Next, we utilize the concepts of differential entropy and mutual information from information theory to quantify information contained in the data about uncertain parameters of the statistical model. Thereafter, we extend the idea of mutual information to

develop an a priori estimator to quantify practical identifiability in a Bayesian setting. While in most statistical models low practical identifiability is due to insufficient information about model parameters, it may often be the case that identifiable subsets exist. Parameters within such subsets have functional relations and exhibit a combined effect on the statistical model, such that the subset is practically identifiable. To find such identifiable subsets, we develop an estimator to highlight dependencies between parameter pairs that emerge a posteriori.

Bayesian parameter inference

Consider the observation/data $y \in \mathcal{Y}$ of a physical system which is a realization of a random variable $Y : \Omega \rightarrow \mathbb{R}^n$ distributed as $p(y)$, where \mathcal{Y} is the set of all possible realizations of the random variable. Herein, we will use the same lower-case, upper-case, and symbol notation to represent a realization, random variable, and the set of all possible observations, respectively. Consider another real-valued random variable $\Theta : \Omega \rightarrow \mathbb{R}^m$ distributed as $p(\theta) : \mathbb{R}^m \rightarrow \mathbb{R}^+$ which denotes the uncertain parameters of the model. The data is assumed to be generated by the statistical model given as

$$y \triangleq \mathcal{F}(\theta, d) + \xi, \quad (1)$$

where $\mathcal{F}(\theta, d) : \mathbb{R}^m \times \mathbb{R}^\ell \rightarrow \mathbb{R}^n$ is the *forward model* which maps the parameters and model inputs $d \in \mathbb{R}^\ell$ to the prediction space. For simplicity, consider the input of the model d as known. The random variable ξ is the additive measurement noise or uncertainty in our measurement. Once the observations are collected using controlled experiments, the *prior* belief of the parameter distribution $p(\theta|d)$ can be updated to obtain the *posterior* distribution $p(\theta|y, d)$ via the Bayes' rule

$$p(\theta|y, d) = \frac{p(y|\theta, d)p(\theta|d)}{p(y|d)}, \quad (2)$$

where $p(y|\theta, d)$ is called the model likelihood and $p(y|d)$ is called the evidence.

Quantifying information gain

Updating parameter belief from the prior to posterior in (2) is associated with a gain in information from the data. This gain can be quantified as the change in the uncertainty of the parameters Θ . As an example, consider a 1D Gaussian prior and posterior distribution such that the information gain can be quantified as a change in variance (a measure of uncertainty) of the parameter distribution. *A greater reduction in parameter uncertainty is a consequence of more information gained from the data.*

In general, the change in parameter uncertainty between the prior and posterior distributions for a given input of the model $d \in \mathcal{D}$ is defined as

$$\Delta \mathcal{U}(d, y) \triangleq \mathcal{U}(p(\theta|d)) - \mathcal{U}(p(\theta|y, d)), \quad (3)$$

where \mathcal{U} is an operator quantifying the amount of uncertainty or the lack of information for a given probability distribution. Thus, the expected information gained about the parameters is defined as

$$\Delta \mathcal{U}(d) \triangleq \mathcal{U}(p(\theta|d)) - \int_{\mathcal{Y}} \mathcal{U}(p(\theta|y, d)) dy. \quad (4)$$

One popular choice for the operator \mathcal{U} is the differential entropy^{32,33,36} which is defined as the average Shannon information³⁷ for a given probability distribution. Mathematically, for a continuous random variable $Z : \Omega \rightarrow \mathbb{R}^f$ with distribution $p(z) : \mathbb{R}^f \rightarrow \mathbb{R}^+$ and support \mathcal{Z} , the differential entropy is defined as

$$H(p(z)) = H(Z) \triangleq - \int_{\mathcal{Z}} p(z) \log p(z) dz. \quad (5)$$

Using differential entropy to quantify the uncertainty of a probability distribution, the change in uncertainty (or expected information gain) of Θ can be evaluated as

$$\Delta \mathcal{U}(d) = H(\Theta|d) - H(\Theta|Y, d), \quad (6a)$$

$$= H(\Theta|d) + H(Y|d) - H(\Theta, Y|d), \quad (6b)$$

$$= - \int_{\Theta} p(\theta|d) \log p(\theta|d) d\theta + \int_{\Theta, \mathcal{Y}} p(\theta, y|d) \log p(\theta|y, d) d\theta dy, \quad (6c)$$

$$= \int_{\Theta, \mathcal{Y}} p(\theta, y|d) \log \frac{p(\theta, y|d)}{p(\theta|d)p(y|d)} d\theta dy, \quad (6d)$$

$$\triangleq I(\Theta; Y|d). \quad (6e)$$

The quantity $I(\Theta; Y | d)$ is called mutual information between the random variables Θ and Y given the model inputs $\mathcal{D} = d$ ³⁸. In the case of discrete random variables the mutual information is measured in *bits*, whereas in the case of continuous variables the units are *nats*.

Remark 1 The mutual information $I(\Theta; Y | d)$ is always non-negative³⁸. This means that updating the parameter belief from the prior to the posterior cannot increase parameter uncertainty.

Estimating practical identifiability

In a Bayesian framework where the parameters are treated as random variables, practical identifiability can be determined by examining information gained about each model parameter^{32,35}. Parameters for which the data is uninformative cannot be estimated with a high degree of confidence and therefore are practically unidentifiable. While mutual information in (6e) is a useful quantity to study information gained from data about the entire parameter set, it does not apportion information gained about each parameter. Therefore, to examine practical identifiability, we define a conditional mutual information

$$I(\Theta_i; Y | \Theta_{\sim i}, d) \triangleq \mathbb{E}_{\Theta_{\sim i}}[I(\Theta_i; Y | \Theta_{\sim i} = \theta_{\sim i}, d)], \quad (7)$$

where $\Theta_{\sim i}$ are all parameters except Θ_i and $\mathbb{E}_{\Theta_{\sim i}}[\cdot]$ denotes the expectation over $p(\theta_{\sim i} | d)$. Using such conditional mutual information for practical identifiability is based on the intuition that on average high information gained about Θ_i means high practical identifiability. We can thus present the following definitions for identifiability in a Bayesian setting.

Definition 1 (Local identifiability) Given a statistical model with parameters Θ , a parameter $\Theta_i \in \Theta$ is said to be locally identifiable if sufficient information is gained about it for a particular realization $\theta_{\sim i}$ of $\Theta_{\sim i}$.

Definition 2 (Global identifiability) Given a statistical model with parameters Θ , a parameter $\Theta_i \in \Theta$ is said to be globally identifiable if sufficient information is gained about it on average with respect to the distribution $p(\theta_{\sim i} | d)$.

The expectation over possible realizations of $\Theta_{\sim i}$ in (7) therefore provides a statistical measure of global practical identifiability⁸. On the contrary, evaluating (7) at a fixed $\theta_{\sim i}$ will result in a local identifiability measure, which means that the information gained about Θ_i will implicitly depend on $\theta_{\sim i}$.

Typically, (7) does not have a closed-form expression and must be estimated numerically. Using the definition of differential entropy in (5) the conditional mutual information can be written as

$$I(\Theta_i; Y | \Theta_{\sim i}, d) = \int_{\Theta_i, \Theta_{\sim i}, \mathcal{Y}} p(\theta_i, \theta_{\sim i}, y | d) \log \frac{p(\theta_i, y | \theta_{\sim i}, d)}{p(\theta_i | \theta_{\sim i}, d)p(y | \theta_{\sim i}, d)} d\theta_i d\theta_{\sim i} dy, \quad (8a)$$

$$= \int_{\Theta_i, \Theta_{\sim i}, \mathcal{Y}} p(\theta_i, \theta_{\sim i}, y | d) \log \frac{p(y | \theta_i, \theta_{\sim i}, d)p(\theta_i | \theta_{\sim i}, d)}{p(\theta_i | \theta_{\sim i}, d)p(y | \theta_{\sim i}, d)} d\theta_i d\theta_{\sim i} dy, \quad (8b)$$

$$= \int_{\Theta_i, \Theta_{\sim i}, \mathcal{Y}} p(\theta_i, \theta_{\sim i}, y | d) \log \frac{p(y | \theta_i, \theta_{\sim i}, d)}{p(y | \theta_{\sim i}, d)} d\theta_i d\theta_{\sim i} dy. \quad (8c)$$

Remark 2 In terms of differential entropy, the conditional mutual information in (7) can be defined as

$$I(\Theta_i; Y | \Theta_{\sim i}, d) \triangleq H(\Theta_i | \Theta_{\sim i}, d) - H(\Theta_i | \Theta_{\sim i}, Y, d), \quad (9a)$$

$$= H(\Theta_i, \Theta_{\sim i} | d) + H(\Theta_{\sim i}, Y | d) - H(\Theta_{\sim i} | d) - H(\Theta_i, \Theta_{\sim i}, Y | d). \quad (9b)$$

In case the parameters are uncorrelated,

$$I(\Theta_i; Y | \Theta_{\sim i}, d) = H(\Theta_i | d) + H(\Theta_{\sim i}, Y | d) - H(\Theta_i, \Theta_{\sim i}, Y | d). \quad (10)$$

While this formulation does not involve any conditional distributions involving the parameters or data, it requires joint distributions, namely, $p(\theta_i, \theta_{\sim i} | d)$, $p(\theta_{\sim i}, y | d)$, $p(\theta_i, \theta_{\sim i}, y | d)$. Typically, such joint distributions do not have a closed-form expression and must be approximated.

In the special case where Θ_i perfectly correlates with $\Theta_{\sim i}$ such that the realization of $\theta_{\sim i}$ provides sufficient information about θ_i , the term inside the logarithm in (8c) becomes identically unity. For such a case, the data is not informative about Θ_i and the effective parameter dimensionality m_{eff} becomes less than m . For a more general case, Monte-Carlo integration can be used to approximate the high dimensional integral as

$$I(\Theta_i; Y | \Theta_{\sim i}, d) \approx \hat{I}(\Theta_i; Y | \Theta_{\sim i}, d) = \sum_{k=1}^{n_{\text{outer}}} \log \frac{p(y^k | \theta_i^k, \theta_{\sim i}^k, d)}{p(y^k | \theta_{\sim i}^k, d)}, \tag{11}$$

where $(\theta_i^k, \theta_{\sim i}^k)$ is drawn from the distribution $p(\theta_i, \theta_{\sim i} | d)$; y^k is drawn from the likelihood distribution $p(y | \theta_i^k, \theta_{\sim i}^k, d)$; and n_{outer} is the number of Monte-Carlo samples. Typically, conditional evidence $p(y | \theta_{\sim i}, d)$ does not have a closed-form expression, and therefore $p(y^k | \theta_{\sim i}^k, d)$ must be numerically approximated. One approach is to rewrite the conditional evidence $p(y^k | \theta_{\sim i}^k, d)$ by means of marginalization as

$$p(y^k | \theta_{\sim i}^k, d) \triangleq \int_{\Theta_i} p(y^k, \theta_i | \theta_{\sim i}^k, d) d\theta_i = \int_{\Theta_i} p(y^k | \theta_i, \theta_{\sim i}^k, d) p(\theta_i | \theta_{\sim i}^k, d) d\theta_i. \tag{12}$$

For simplicity, assume that the parameters are uncorrelated prior to observing the data, and are also independent of the model inputs d . As a result, (12) can be re-written as

$$p(y^k | \theta_{\sim i}^k, d) = \int_{\Theta_i} p(y^k | \theta_i, \theta_{\sim i}^k, d) p(\theta_i) d\theta_i. \tag{13}$$

This results in a low-dimensional integral over a univariate prior distribution $p(\theta_i)$. Evaluating (13) using the classical Monte-Carlo integration can dramatically increase the overall cost of estimating the conditional mutual information in (11), especially if the likelihood evaluation is computationally expensive. In the special case where the priors are normally distributed, this cost can be reduced by considering a ζ -point Gaussian quadrature rule. Using the quadrature approximation in (13) gives

$$p(y^k | \theta_{\sim i}^k, d) \approx \hat{p}(y^k | \theta_{\sim i}^k, d) = \sum_{\zeta=1}^{\zeta=n_{\text{inner}}} [p(y^k | \theta_i^\zeta, \theta_{\sim i}^k, d)] \gamma^\zeta, \tag{14}$$

where θ_i^ζ and γ^ζ are the ζ^{th} quadrature point and weight, respectively; n_{inner} is the number of quadrature points. Here, we use the Gauss-Hermite quadrature rule, which uses the t^{th} order Hermite polynomial and will be exact for polynomials up to order $2t - 1$ ³⁹. In a much more general case where the prior distributions can be non-Gaussian (however, can still be evaluated), the cost of estimating (13) can be reduced by using importance sampling with a proposal distribution $q(\theta_i)$. Using importance sampling we can rewrite (13) as

$$p(y^k | \theta_{\sim i}^k, d) = \int_{\Theta_i} [p(y^k | \theta_i, \theta_{\sim i}^k, d) w(\theta_i)] q(\theta_i) d\theta_i, \tag{15}$$

where $w(\theta_i) = p(\theta_i)/q(\theta_i)$ are the importance sampling weights. In the case where the proposal distribution $q(\theta_i)$ is Gaussian, the quadrature rule can be applied to (15) as

$$p(y^k | \theta_{\sim i}^k, d) \approx \hat{p}(y^k | \theta_{\sim i}^k, d) = \sum_{\zeta=1}^{\zeta=n_{\text{inner}}} [p(y^k | \theta_i^\zeta, \theta_{\sim i}^k, d) w(\theta_i^\zeta)] \gamma^\zeta. \tag{16}$$

Combining the estimator for conditional evidence ((14) or (16)) with (11) results in a biased estimator for conditional mutual information^{40,41}. While the variance is controlled by the numerical accuracy of estimating the high-dimensional integral in (11), the bias is governed by the accuracy of approximating the conditional evidence in (12). This means that the variance is controlled by n_{outer} Monte-Carlo samples and bias by n_{inner} quadrature points.

In practice, estimating conditional evidence can become computationally expensive, especially when the variability in the output of the forward model is high with respect to Θ_i given $\Theta_{\sim i} = \theta_{\sim i}$, that is, large $\nabla_{\theta_i} \mathcal{F}(\theta, d)|_{\theta_{\sim i} = \theta_{\sim i}}$. For such statistical models, conditional evidence can become near zero such that numerical approximation by means of vanilla Monte-Carlo integration or Gaussian quadrature in (14) can be challenging⁴¹. Using an estimator based on importance sampling for conditional evidence as shown in (16) can alleviate this problem by carefully choosing the density of the proposal $q(\theta_i)$. As an example, consider the case where the additive measurement noise ξ is normally distributed as $\mathcal{N}(0, \Gamma)$ such that the likelihood of the model is distributed as $p(y | \theta) = \mathcal{N}(\mathcal{F}(\theta, d), \Gamma)$, and y^k is sampled according to $\mathcal{N}(\mathcal{F}(\theta_i^k, \theta_{\sim i}^k, d), \Gamma)$. In the case where model predictions have large variability with respect to the parameter Θ_i for a given $\Theta_{\sim i} = \theta_{\sim i}$ the model likelihood can become small. For such a case, the importance-sampling-based estimator given in (16) can be used by constructing a proposal around the sample θ_i^k , such as $q(\theta_i) = \mathcal{N}(\theta_i^k, \sigma_{\text{proposal}}^2)$ where $\sigma_{\text{proposal}}^2$ is the variance of the proposal distribution. This results in a robust estimation of conditional evidence and prevents infinite values for conditional mutual information. Here, we consider (16) to estimate conditional evidence.

Remark 3 Assessing the practical identifiability in a Bayesian framework is dependent on the prior distribution. Although the framework presented in this article is entirely an a priori analysis of practical identifiability, prior selection can affect estimated identifiability. Prior selection in itself is an extensive area of research and is not considered a part of this work.

Physical interpretation of identifiability in an information-theoretic framework

Assessing practical identifiability using the conditional mutual information described in (7) provides a relative measure of *how many bits (or nats) of information is gained for a particular parameter*. In practical applications where this information gain can vary on disparate scales, it is useful to associate a physical interpretation to identifiability. Following Pant et al.³², consider a hypothetical direct observation statistical model given as $\psi \triangleq \theta_i + \Lambda$, where $\Lambda \sim \mathcal{N}(0, \sigma_\Lambda^2)$ is the additive measurement noise. Given this observation model, we can define an information gain equivalent variance $\mathcal{C}(\Theta_i)$ as the *measurement uncertainty in the direct observation model given* $I(\Theta_i; \Psi) = \hat{I}(\Theta_i; Y | \Theta_{\sim i}, d)$. Large $\mathcal{C}(\Theta_i)$ would mean that the information gained about Θ_i (using (7)) for the statistical model (1) would lead to higher measurement uncertainty if the parameter is observed directly.

If the prior distribution $p(\theta_i)$ can be approximated by means of an equivalent normal distribution $\mathcal{N}(\mu_e, \sigma_e^2)$ then $I(\Theta_i; \Psi)$ is given as

$$I(\Theta_i; \Psi) \triangleq \frac{1}{2} \log \left(1 + \frac{\sigma_e^2}{\sigma_\Lambda^2} \right), \tag{17}$$

such that

$$\mathcal{C}(\Theta_i) \triangleq \sigma_\Lambda^2 = \sigma_e^2 (\exp\{2\hat{I}(\Theta_i; Y | \Theta_{\sim i}, d)\} - 1)^{-1}. \tag{18}$$

This information gain equivalent variance only depends on the information gained for the model parameter, and thus, can be used as a metric to compare different model parameters.

Estimating parameter dependence

In most statistical models, unknown functional relationships or dependencies may be present between parameters such that multiple parameters have a combined effect on the statistical model. Such parameters can form an identifiable subset where an individual parameter will exhibit low identifiability, however, the subset is collectively identifiable. This means that the data is uninformative or weakly informative about an individual parameter within the subset, whereas it is informative about the entire subset. As an example, consider the statistical model: $y = \theta_1 \theta_2 * d + \xi$ for which individually identifying Θ_1 or Θ_2 is not possible as they have a combined effect on the statistical model. However, it is clear that Θ_1 and Θ_2 belong to an identifiable subset such that the pair (Θ_1, Θ_2) is identifiable. Thus, considering the statistical model given by $y = \theta_3 * d + \xi$ where $\theta_3 = \theta_1 * \theta_2$ will have better identifiability characteristics. For such statistical models, the traditional method of examining correlations between parameters is often insufficient, as it only reveals linear functional relations between random variables.

To highlight the parameter dependencies, consider the statistical model given in (1) such that we are interested in examining the relations between $\Theta_i \in \Theta$ and $\Theta_j \in \Theta$ that emerge a posteriori. While the conditional mutual information presented in "Estimating practical identifiability" provides information on the practical identifiability of an individual parameter, it does not provide information about dependencies developed between pairs of parameters. To quantify such dependencies, we define a conditional mutual information between parameter pairs

$$I(\Theta_i; \Theta_j | Y, \Theta_{\sim ij}, d) \triangleq \mathbb{E}_{\Theta_{\sim ij}} [\mathbb{E}_Y [I(\Theta_i; \Theta_j | Y = y, \Theta_{\sim ij} = \theta_{\sim ij}, d)]], \tag{19}$$

which evaluates the average information between the variables Θ_i and Θ_j that is obtained a posteriori. Here, $\Theta_{\sim ij}$ is defined as all the parameters of the statistical model except Θ_i and Θ_j .

A closed-form expression for (19) is typically not available, and therefore a numerical approximation is required. In integral form, (19) is given as

$$I(\Theta_i; \Theta_j | Y, \Theta_{\sim ij}, d) \triangleq \int_{\Theta_i, \Theta_j, \Theta_{\sim ij}, \mathcal{Y}} p(\theta_i, \theta_j, \theta_{\sim ij}, y | d) \left[\log [p(\theta_i, \theta_j | y, \theta_{\sim ij}, d)] - \log [p(\theta_i | y, \theta_{\sim ij}, d)p(\theta_j | y, \theta_{\sim ij}, d)] \right] d\theta_i d\theta_j d\theta_{\sim ij} dy, \tag{20}$$

where

$$p(\theta_i, \theta_j | y, \theta_{\sim ij}, d) \triangleq \frac{p(y | \theta_i, \theta_j, \theta_{\sim ij}, d)p(\theta_i, \theta_j, | \theta_{\sim ij}, d)}{p(y | \theta_{\sim ij}, d)}, \tag{21a}$$

$$p(\theta_i | y, \theta_{\sim ij}, d) \triangleq \frac{p(y | \theta_i, \theta_{\sim ij}, d)p(\theta_i | \theta_{\sim ij}, d)}{p(y | \theta_{\sim ij}, d)}, \tag{21b}$$

$$p(\theta_j | y, \theta_{\sim ij}, d) \triangleq \frac{p(y | \theta_j, \theta_{\sim ij}, d)p(\theta_j | \theta_{\sim ij}, d)}{p(y | \theta_{\sim ij}, d)}, \tag{21c}$$

via Bayes' theorem.

Remark 4 In terms of differential entropy, the conditional mutual information in (19) can be defined as

$$I(\Theta_i; \Theta_j | Y, \Theta_{\sim ij}, d) \triangleq H(\Theta_i | Y, \Theta_{\sim ij}, d) - H(\Theta_i | \Theta_j, Y, \Theta_{\sim ij}, d), \tag{22a}$$

$$= H(\Theta_i, \Theta_{\sim i,j}, Y | d) + H(\Theta_j, \Theta_{\sim i,j}, Y | d) - H(\Theta_{\sim i,j}, Y | d) - H(\Theta_i, \Theta_j, \Theta_{\sim i,j}, Y | d). \quad (22b)$$

Such a formulation requires evaluating joint distributions, namely, $p(\theta_i, \theta_{\sim i,j}, y | d)$, $p(\theta_j, \theta_{\sim i,j}, y | d)$, $p(\theta_{\sim i,j}, y | d)$, and $p(\theta_i, \theta_j, \theta_{\sim i,j}, y | d)$. Typically, such joint distributions do not have a closed-form expression and must be approximated.

For the sake of illustration, assume that the parameters are uncorrelated with each other *prior to observing the data*. As a consequence of this assumption, any relations developed between Θ_i and Θ_j are discovered solely from data. Furthermore, it is also reasonable to assume that prior knowledge of the parameters is independent of the input of the model d . Substituting (21a) through (21c) into (20) we obtain

$$I(\Theta_i; \Theta_j | Y, \Theta_{\sim i,j}, d) = \int_{\Theta_i, \Theta_j, \Theta_{\sim i,j}, \mathcal{Y}} p(\theta_i, \theta_j, \theta_{\sim i,j}, y | d) \left[\log [p(y | \theta_i, \theta_j, \theta_{\sim i,j}, d)] + \log [p(y | \theta_{\sim i,j}, d)] - \log [p(y | \theta_i, \theta_{\sim i,j}, d)] - \log [p(y | \theta_j, \theta_{\sim i,j}, d)] \right] d\theta_i d\theta_j d\theta_{\sim i,j} dy. \quad (23)$$

Similar to "Estimating practical identifiability" we can estimate the conditional mutual information in (23) using Monte-Carlo integration as $\hat{I}(\Theta_i; \Theta_j | Y, \Theta_{\sim i,j}, d) \approx I(\Theta_i; \Theta_j | Y, \Theta_{\sim i,j}, d)$ where

$$\hat{I}(\Theta_i; \Theta_j | Y, \Theta_{\sim i,j}, d) = \sum_{k=1}^{k=n_{\text{outer}}} \log \frac{p(y^k | \theta_i^k, \theta_j^k, \theta_{\sim i,j}^k, d) p(y^k | \theta_{\sim i,j}^k, d)}{p(y^k | \theta_i^k, \theta_{\sim i,j}^k, d) p(y^k | \theta_j^k, \theta_{\sim i,j}^k, d)}, \quad (24)$$

where θ_i^k , θ_j^k , and $\theta_{\sim i,j}^k$ are drawn from the prior distributions $p(\theta_i)$, $p(\theta_j)$, and $p(\theta_{\sim i,j})$, respectively; y^k is drawn from the likelihood distribution $p(y | \theta_i^k, \theta_j^k, \theta_{\sim i,j}^k, d)$. The conditional evidence in (24) can be obtained by means of marginalization

$$p(y^k | \theta_{\sim i,j}^k, d) \triangleq \int_{\Theta_i, \Theta_j} p(y^k | \theta_i, \theta_j, \theta_{\sim i,j}^k, d) p(\theta_i, \theta_j) d\theta_i d\theta_j, \quad (25a)$$

$$p(y^k | \theta_i^k, \theta_{\sim i,j}^k, d) \triangleq \int_{\Theta_j} p(y^k | \theta_j, \theta_i^k, \theta_{\sim i,j}^k, d) p(\theta_j) d\theta_j, \quad (25b)$$

$$p(y^k | \theta_j^k, \theta_{\sim i,j}^k, d) \triangleq \int_{\Theta_i} p(y^k | \theta_i, \theta_j^k, \theta_{\sim i,j}^k, d) p(\theta_i) d\theta_i. \quad (25c)$$

Similar to "Estimating practical identifiability" the conditional evidence in (25a) through (25c) can be efficiently estimated using importance sampling along with Gaussian quadrature rules. However, it should be noted that (25a) is an integral over a two-dimensional space, and therefore requires n_{inner}^2 quadrature points.

Numerical experiments

This section presents numerical experiments to validate the information-theoretic approach to examine practical identifiability. The estimate obtained for global identifiability is compared with the variance-based global sensitivity analysis by means of first-order Sobol indices computed using SALib^{42,43} (see S1 in the supplementary material). First, a linear Gaussian statistical model is considered for which practical identifiability can be analytically examined through the proposed information-theoretic approach. This model is computationally efficient and is therefore ideal for conducting estimator convergence studies. Next, the practical identifiability of a reduced kinetics model for methane-air combustion is considered. Reduced kinetics models are widely used in the numerical analysis of chemically reactive flows since embedding detailed chemistry of combustion is often infeasible. Such reduced kinetic models are often parameterized such that constructing models with practically identifiable parameters is desirable to improve confidence in the model prediction.

Application to a linear Gaussian model

The identifiability framework is now applied to a linear Gaussian problem for which closed-form expressions are available for the conditional mutual information in (7) and (19) (see S2 in the supplementary material). Consider the statistical model

$$y = \mathcal{F}(\theta, d) + \xi \quad ; \quad \xi \sim \mathcal{N}(0, \Gamma), \quad (26)$$

where $\mathcal{F}(\theta, d) = \mathbf{A}\theta$ and $\mathbf{A} \in \mathbb{R}^{n \times m}$ is called the feature matrix. The prior distribution is given by $p(\theta) = \mathcal{N}(\mu_\theta, \Sigma_\theta)$ where $\mu_\theta \in \mathbb{R}^m$ and $\Sigma_\theta \in \mathbb{R}^{m \times m}$. Model likelihood is therefore given by

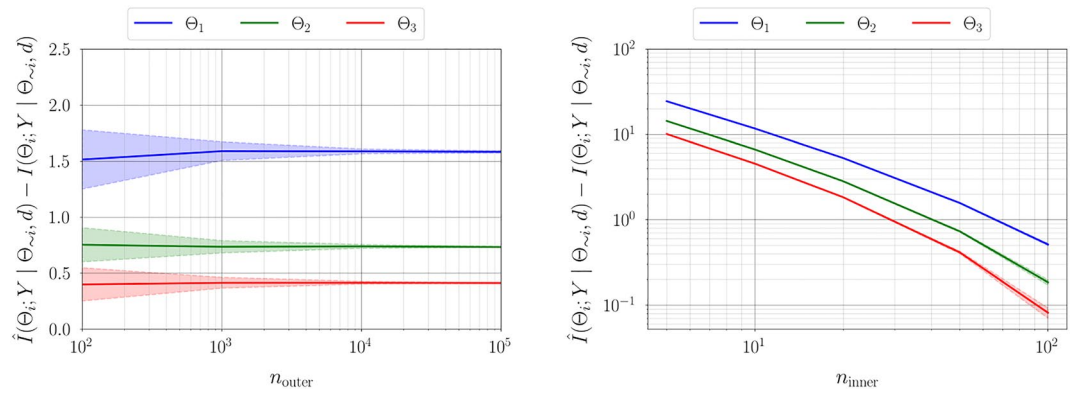


Figure 1. Convergence of the variance in practical identifiability estimator of a linear Gaussian statistical model (left); the number of quadrature points $n_{\text{inner}} = 50$ is considered and the number of Monte-Carlo integration samples n_{outer} is varied. Bias convergence for practical identifiability estimator in case of a linear Gaussian statistical model (right); $n_{\text{outer}} = 10^4$ is considered and n_{inner} is varied. For a given n_{inner} increasing n_{outer} reduces the variance, whereas increasing n_{inner} for a given n_{outer} decreases the bias in the estimate.

$p(y | \theta) = \mathcal{N}(\mathbf{A}\theta, \Gamma)$ where $\Gamma \in \mathbb{R}^{n \times n}$. Here, μ_{Θ} , Σ_{Θ} , and Γ are all constants and are considered known. The evidence distribution for this model is given by $p(y) \triangleq \mathcal{N}(\mu_Y, \Sigma_Y) = \mathcal{N}(\mathbf{A}\mu_{\Theta}, \mathbf{A}\Sigma_{\Theta}\mathbf{A}^T + \Gamma)$, such that no model-form error exists. Consider a feature matrix

$$\mathbf{A} = \begin{pmatrix} d_1 & d_1^2 & \dots & d_1^m \\ d_2 & d_2^2 & \dots & d_2^m \\ \vdots & \vdots & \ddots & \vdots \\ d_n & d_n^2 & \dots & d_n^m \end{pmatrix}, \tag{27}$$

where $d_{i=1}^n$ are n linearly-spaced points in an interval $[-1, 1]$, and $m = 3$, which means that the statistical model has 3 uncertain parameters. Assume an uncorrelated measurement noise $\Gamma = \sigma_{\xi}^2 \mathbb{I}$ with $\sigma_{\xi}^2 = 0.1$. For the purpose of parameter estimation, synthetic data is generated using (26) assuming $\theta^* = [1, 2, 3]^T$ and $n = 100$.

Parameter identifiability

The goal of the framework developed in "Estimating practical identifiability" is to assess the practical identifiability of the statistical model in (26) before any controlled experiment is conducted. Consider $\mu_{\Theta} = \mathbf{0}$ and $\Sigma_{\Theta} = \mathbb{I}$. Using such an uncorrelated prior distribution for the identifiability study ensures that the information obtained is only due to the observation of the data (as discussed in "Estimating parameter dependence"). Using historical parameter estimates can improve the prior (Remark 1) which can affect the identifiability analysis. However, we have not considered any such prior refinement.

Figure 1 illustrates the convergence of error in estimating information gain for each parameter using the estimator developed in "Estimating practical identifiability". As expected, for a fixed number of quadrature points, increasing the number of Monte-Carlo integration points decreases the variance in estimation. However, for a fixed n_{outer} increasing the number of quadrature points reduces the bias in the estimate. Figure 2 illustrates the variance and bias convergence of error in estimating parameter dependencies as described in "Estimating parameter dependence". As expected and observed, the variance in error is controlled by the accuracy of Monte-Carlo integration, that is, by n_{outer} , and the bias is controlled by the quadrature approximation, that is, through n_{inner} .

The first-order Sobol indices, estimated information gain, and information gain equivalent variance $\mathcal{C}(\Theta_i)$ are shown in Figure 3. The estimated first-order Sobol indices (see S3 in the supplementary material for convergence study) show that the considered linear Gaussian forward model has the largest output variability due to uncertainty in Θ_1 , followed by Θ_2 and Θ_3 . This implies that the forward model is most sensitive to the parameter θ_1 , followed by θ_2 and then θ_3 . This is not surprising since $d_{i=1}^n$ are points in the interval $[-1, 1]$. Thus, according to the first-order Sobol indices, the relevance of the parameters follows the order: θ_1, θ_2 , and θ_3 . The estimated information gained agrees well with the truth. Further, the obtained trend suggests that the data is most informative about Θ_1 , followed by Θ_2 , and then Θ_3 . As discussed in "Estimating practical identifiability", practical identifiability follows the same trend. Furthermore, as reported in previous work³¹, it can be seen that parameters with good identifiability characteristics also exhibit high model sensitivity. Using the hypothetical direct observation model described in "Estimating practical identifiability", the smallest measurement uncertainty is obtained for the variable Θ_1 , followed by Θ_2 and Θ_3 . That is, parameters with high practical identifiability are associated with low measurement uncertainty in a direct observation model.

Figure 4 illustrates the variability of the first-order Sobol indices and the estimated information gain with measurement noise variance σ_{ξ}^2 . The first-order Sobol indices only account for the parameter uncertainty, and therefore remain unchanged with an increase in measurement noise. However, the estimated information gain

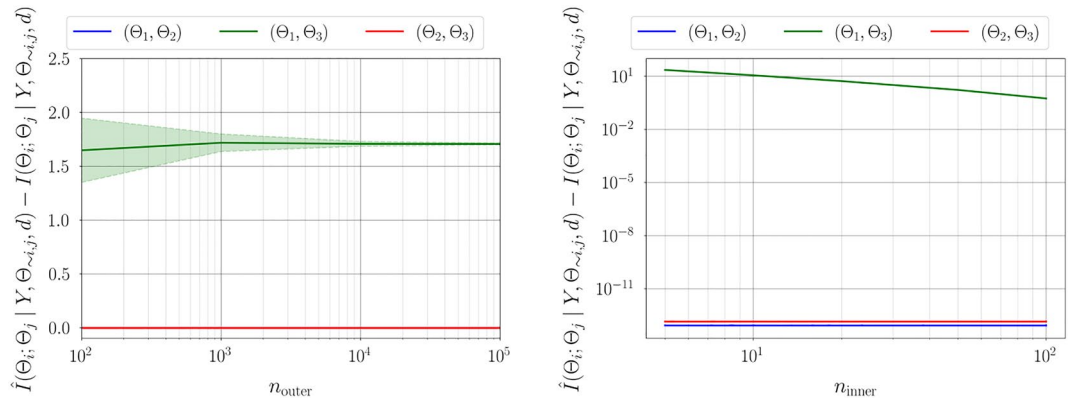


Figure 2. Convergence of the variance in practical identifiability for estimating parameter dependencies of a linear Gaussian statistical model (left); the number of quadrature points $n_{\text{inner}} = 50$ is considered and the number of Monte-Carlo integration samples n_{outer} is varied. Bias convergence for estimating parameter dependencies in the case of a linear Gaussian statistical model (right); $n_{\text{outer}} = 10^4$ is considered and n_{inner} is varied. For a given n_{inner} increasing n_{outer} reduces the variance, whereas increasing n_{inner} for a given n_{outer} decreases the bias in estimating the parameter dependencies.

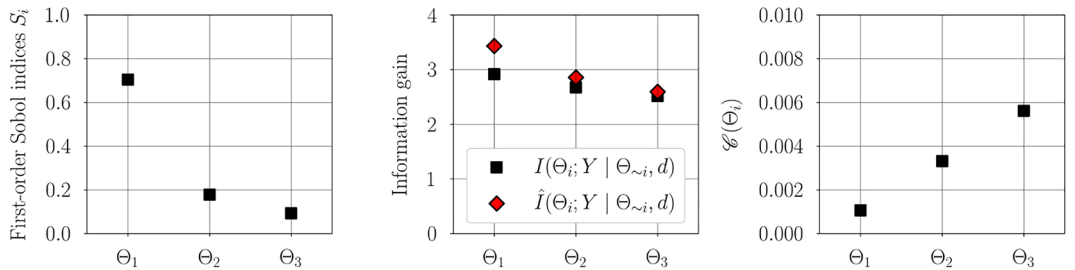


Figure 3. First-order Sobol indices (left), information gain (center), and information gain equivalent variance $\mathcal{S}(\Theta_i)$ (right) for linear Gaussian model. Sobol indices show that the output of the statistical model has the largest variability due to uncertainty in Θ_1 , followed by Θ_2 and Θ_3 . Variable Θ_1 exhibits the largest gain in information and therefore the highest practical identifiability, followed by Θ_2 and then Θ_3 . For a direct observation model, the variable Θ_1 has the lowest measurement uncertainty, followed by Θ_2 and Θ_3 .

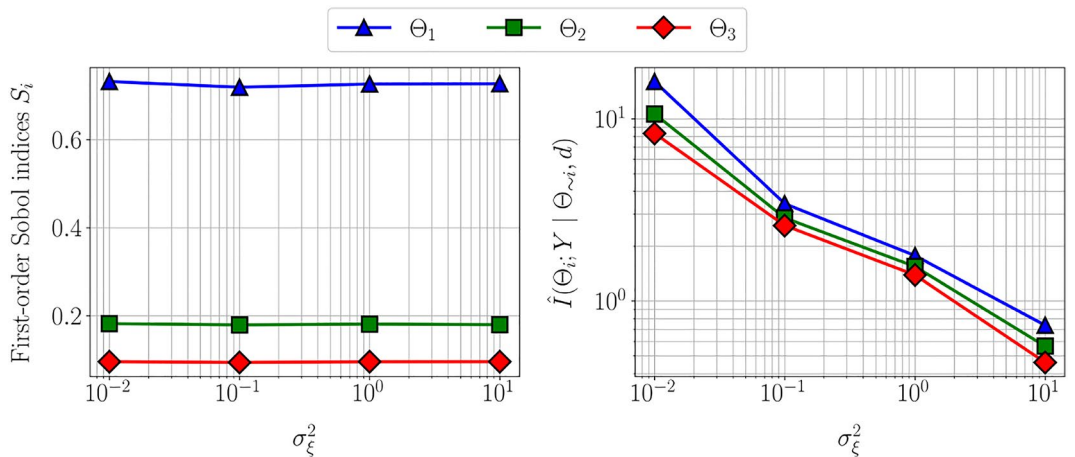


Figure 4. First-order Sobol indices (left) and estimated information gain (right) vs. measurement noise variance σ_{ξ}^2 for linear Gaussian model. Increasing measurement noise covariance does not affect the variability of the output with respect to the parameters and therefore the first-order Sobol indices remain unchanged. However, the information gain decreases with increasing measurement noise covariance.

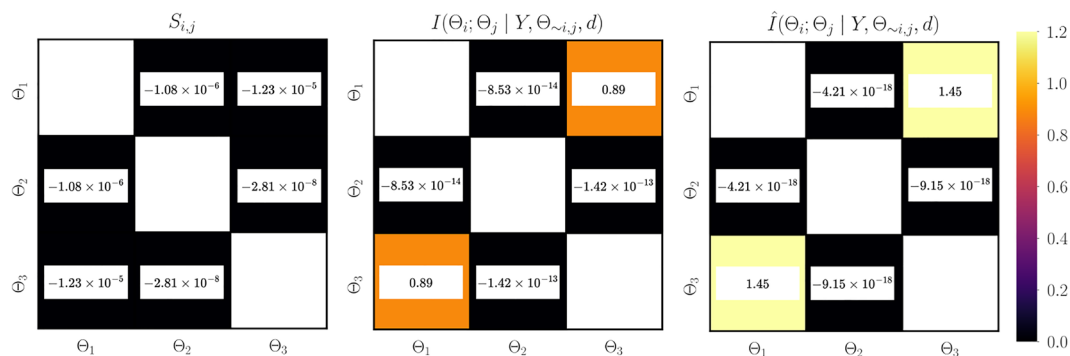


Figure 5. Second-order Sobol indices (left), true parameter dependencies (center), and estimated parameter dependencies (right) for linear Gaussian model. The second-order Sobol indices show negligible interactions between parameter pairs. The obtained estimate of parameter dependency agrees well with their true values; the trend is preserved. Θ_1 and Θ_3 have the largest dependency on one another, and therefore are expected to have a combined effect on the output of the statistical model.

and thereby the practical identifiability decreases with measurement noise. This observation corroborates the intuition that large measurement uncertainty will lead to large uncertainty in the parameter estimation.

Figure 5 shows the second-order Sobol indices and the true and estimated dependencies between the parameter pairs for the linear Gaussian model. Examining the second-order Sobol indices (see S3 in the supplementary material for convergence study) shows that there are negligible interactions between parameter pairs. Estimated parameter dependencies agree well with the truth; the trend is preserved. The bias observed is due to the error in approximating the conditional evidence, as shown in Figure 2. It can be clearly seen that the parameters Θ_1 and Θ_3 have high dependencies. This means that these parameters compensate for one another such that they will have a combined effect on the output of the statistical model. These parameters are associated with the features $d_{i=1}^n$ and $d_{i=1}^3$ which, in fact, have a similar effect on the statistical model for $d_{i=1}^n \in [-1, 1]$. This observation also shows that the low practical identifiability of Θ_3 is mainly due to the underlying dependency with Θ_1 such that the pair (Θ_1, Θ_3) has a combined effect in the statistical model.

Parameter estimation

For the linear Gaussian model, the joint distribution $p(\theta, y)$ can be written as

$$p(\theta, y) \triangleq \mathcal{N}(\mu_{\Theta, Y}, \Sigma_{\Theta, Y}) = \mathcal{N}\left(\begin{bmatrix} \mu_{\Theta} \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{\Theta} & \Sigma_{\Theta} \mathbf{A}^T \\ \mathbf{A} \Sigma_{\Theta} & \Sigma_Y \end{bmatrix}\right), \quad (28)$$

such that the analytical posterior distribution is given as $p(\theta | y) = \mathcal{N}(\mu_{\Theta_{post}}, \Sigma_{\Theta_{post}})$, where $\mu_{\Theta_{post}} = \mu_{\Theta} + \Sigma_{\Theta} \mathbf{A}^T \Sigma_Y^{-1} (y - \mu_Y)$ and $\Sigma_{\Theta_{post}} = \Sigma_{\Theta} - \Sigma_{\Theta} \mathbf{A}^T \Sigma_Y^{-1} \mathbf{A} \Sigma_{\Theta}$ using Gaussian conditioning.

Samples from the posterior distribution and the aggregate posterior prediction are shown in Figure 6. Variables Θ_1 and Θ_3 have a negative correlation, whereas Θ_2 is uncorrelated with other parameters. This means that the parameter variables Θ_1 and Θ_3 have (linear) dependencies on each other, and Θ_2 does not have such dependencies. These dependencies were suggested during the a priori analysis conducted on the statistical model as illustrated in Figure 5. Aggregate posterior prediction agrees well with the data and exhibits high certainty.

Figure 7 illustrates the change in variance of the parameter Θ_i defined as $\Delta(\sigma_{\Theta_i}^2) \triangleq \sigma_{\Theta_i}^2 - \sigma_{\Theta_i, post}^2$ versus σ_{ξ}^2 . Parameter Θ_2 exhibits the smallest posterior uncertainty, followed by Θ_1 and Θ_3 for all σ_{ξ}^2 . While Θ_1 has the largest estimated information gain (Figure 3(center)), it exhibits dependencies with Θ_3 (Figure 5(right)), thereby resulting in larger posterior uncertainty in comparison to Θ_2 . In practical applications, where model selection or parameter selection is critical, examining the information gain and parameter dependencies can therefore aid in finding parameters that can be estimated with high certainty. Increasing the measurement noise results in a smaller change in parameter variance, that is, the parameters exhibit larger posterior uncertainty. This is also shown by the variation of estimated information gain with measurement noise (Figure 4(right)). On the contrary, the first-order Sobol indices remain unchanged with measurement noise (Figure 4(left)).

Application to methane chemical kinetics

Accurate characterization of chemical kinetics is critical in the numerical prediction of reacting flows. Although there have been significant advancements in computational architectures and numerical methods, embedding the full chemical kinetics in numerical simulations is almost always infeasible. This is primarily because of the high-dimensional coupled ordinary differential equations that have to be solved to obtain concentrations of a large number of involved species. As a result, significant efforts have been made to develop reduced chemical kinetics models that seek to capture features such as ignition delay, adiabatic flame temperature, or flame speed observed using the true chemical kinetics^{44–47}. These reduced mechanisms are typically formulated using a combination of theory and intuition, leaving unresolved chemistry, resulting in uncertainties in the relevant

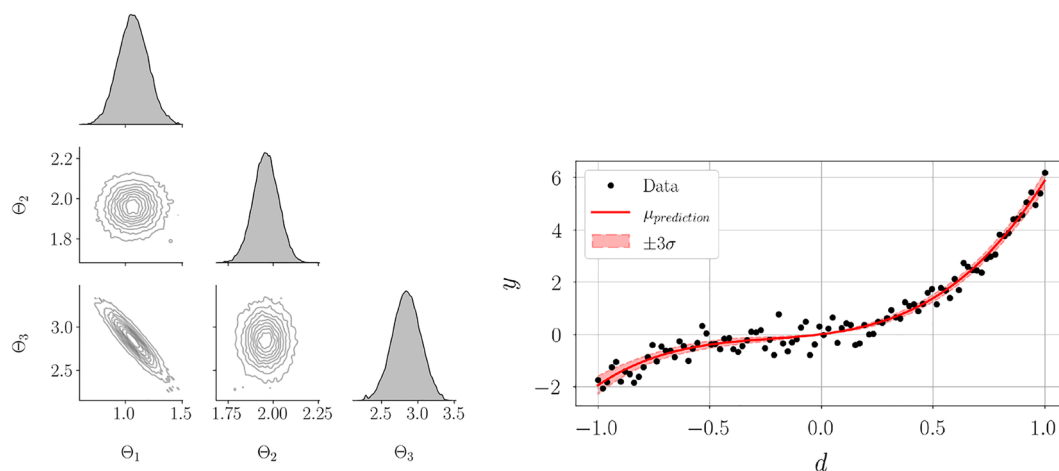


Figure 6. Correlation plot for samples obtained from the true posterior distribution (left) and the obtained aggregate posterior prediction (right) for linear Gaussian model. A negative correlation is observed between Θ_1 and Θ_3 , whereas Θ_2 is uncorrelated from other parameters. Aggregate posterior prediction agrees well with the data and exhibits high certainty.

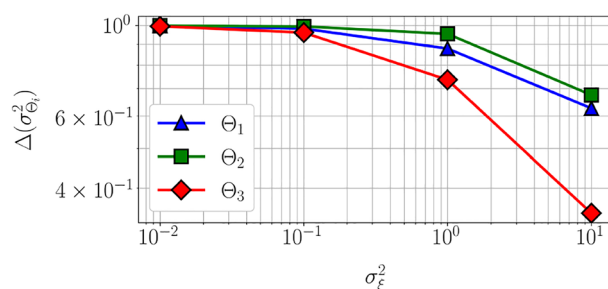


Figure 7. Change in parameter variance $\Delta(\sigma_{\Theta_i}^2)$ vs. measurement noise covariance σ_{ξ}^2 for linear Gaussian model. Increasing measurement noise results in a smaller change in parameter variance from the prior to the posterior. Largest reduction in variance is observed for Θ_2 , followed by Θ_1 and Θ_3 .

rate parameters⁴⁸. Selecting a functional form of the modeled reaction rate terms that lead to reliable parameter estimation is highly desirable^{7,49}. This means that for high confidence in parameter estimation and thereby model prediction, the underlying parameterization of reaction rate terms must exhibit high practical identifiability.

Shock tube ignition is a canonical experiment used to develop and validate combustion reaction mechanism⁵⁰. In such experiments, the reactant mixture behind the reflected shock experiences elevated temperature and pressure, followed by mixture combustion. An important quantity of interest in such experiments is the time difference between the onset of the reflected shock and the ignition of the reactant mixture, defined as the ignition delay t_{ign} ⁵¹. Ignition delay is characterized as the time of maximum heat release or steepest change in reactant temperature and is therefore a key physio-chemical property for combustion systems.

To illustrate the practical identifiability framework we will consider stoichiometric methane-air combustion in a shock tube under an adiabatic, ideal-gas constant pressure ignition assumption. Typically, the chemical kinetics capturing detailed chemistry of methane-air ignition is computationally expensive due to hundreds of associated reactions. To model the reaction chemistry, consider the classical 2-step mechanism proposed by Westbrook et al.⁵² that accounts for the incomplete oxidation of methane. This reduced mechanism consists of a total of 6 species (5 reacting and 1 inert species, namely, N_2) and 2 reactions (1 reversible), thus drastically reducing the cost of evaluating the chemical kinetics. The reactions involved in this reduced chemical kinetics model are



where the overall reaction rates are temperature-dependent and are modeled using the Arrhenius rate equation as

$$k_1 \triangleq A e^{\frac{-48400}{RT}} [\text{CH}_4]^{-0.3} [\text{O}_2]^{1.3}, \quad (31)$$

$$k_{2f} \triangleq 3.98 \times 10^{14} e^{\frac{-40000}{RT}} [\text{CO}][\text{H}_2\text{O}]^{0.5} [\text{O}_2]^{0.25}, \quad (32)$$

$$k_{2b} \triangleq 5 \times 10^8 e^{\frac{-40000}{RT}} [\text{CO}_2], \quad (33)$$

where $A = 2.8 \times 10^9$ is the pre-exponential factor, R is the ideal gas constant, and T is the temperature in Kelvin. To solve the resulting reaction equations `CANTERA v2.6.0`⁵³ is used. Figure 8 illustrates the temperature evolution using the 2-step mechanism and GRI-Mech 3.0⁵⁴ for an initial temperature $T_o = 1500\text{K}$, initial pressure $P_o = 100\text{kPa}$, and at a stoichiometric ratio $\phi = 1$. The GRI-Mech 3.0 mechanism consists of detailed chemical kinetics with 53 species and 325 reactions. As noticed, the 2-step mechanism under-predicts the ignition delay by nearly an order of magnitude. To improve the predictive capabilities of the 2-step mechanism a functional dependency for the pre-exponential factor can be introduced as $\log A = \mathcal{G}(T_o, \phi)$, where

$$\mathcal{G}(T_o, \phi) \triangleq 18 + \theta_1 + \tanh(\theta_2 + \theta_3 * \phi) \frac{T_o}{1000}. \quad (34)$$

Here, θ_1, θ_2 , and θ_3 are the uncertain model parameters. Similar parameterization has been used for *n*-dodecane reduced chemical kinetics⁴⁸. It should be noted that while a more expressive functional form for the pre-exponential factor can be chosen in (34), the goal of the framework is to ascertain practical identifiability. For parameter estimation, consider the detailed GRI-Mech 3.0 to be the 'exact solution' to the combustion problem which can then be used to generate the data. Consider logarithm of ignition temperature at $T_o = 1100, 1400, 1700$ and 2000K at $\phi = 1.0$ and $P_o = 100\text{kPa}$ as the available data for model calibration. Assume an uncorrelated measurement noise $\Gamma = \sigma_\xi^2 \mathbb{I}$ with $\sigma_\xi^2 = 0.1$.

Parameter identifiability

The practical identifiability framework is now applied to the methane-air combustion problem to examine the identifiability of the model parameters in (34) before any controlled experiments are conducted. Consider an uncorrelated prior distribution for the model parameters as $\theta_1 \sim \mathcal{N}(0, 1)$; $\theta_2 \sim \mathcal{N}(0, 1)$; $\theta_3 \sim \mathcal{N}(0, 1)$. Such priors result in pre-exponential factors in an order similar to those previously reported⁵², and are therefore considered suitable for the study. Similar to "Application to a linear Gaussian model" historical estimates of the model parameters are not considered for examining identifiability.

The first-order Sobol indices, estimated information gain, and information gain equivalent variance $\mathcal{C}(\Theta_i)$ are shown in Figure 9. The information gain is estimated using $n_{\text{outer}} = 12000$ Monte-Carlo samples, and $n_{\text{inner}} = 5$ quadrature points. Examining the first-order Sobol indices (see S3 in the supplementary material for convergence study), the output of the forward model exhibits the largest variability due to uncertainty in the variable Θ_1 . Followed by similar variability in the model output with respect to Θ_2 and Θ_3 . The largest information gain is observed for the variable Θ_1 , followed by similar gains for Θ_2 and Θ_3 . This means that Θ_1 will have the highest practical identifiability, followed by a much lower identifiability for Θ_2 and Θ_3 . Using the hypothetical direct observation model as described in "Estimating practical identifiability", the variable Θ_1 with the largest practical identifiability exhibits the lowest measurement uncertainty, followed by similar uncertainty for Θ_2 and Θ_3 .

Figure 10 shows the second-order Sobol indices and estimated parameter dependencies. The second-order Sobol indices (see S3 in the supplementary material for convergence study) follow the trend $S_{2,3} > S_{1,2} \approx S_{1,3}$, suggesting that there are underlying interactions between the parameters Θ_2 and Θ_3 . As observed, the low identifiability of Θ_2 and Θ_3 suggested in Figure 9 is primarily due to the underlying dependencies between pairs (Θ_1, Θ_2) and (Θ_1, Θ_3) . To estimate the parameter dependencies $n_{\text{outer}} = 12000$ Monte-Carlo samples, and

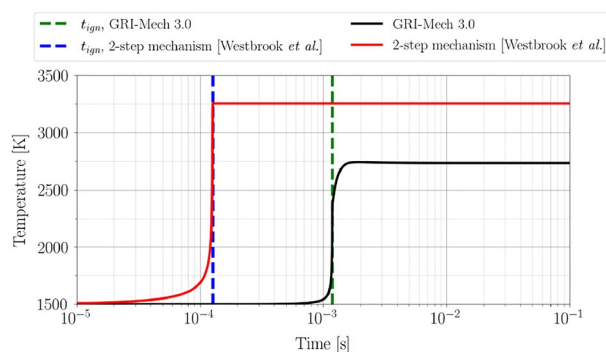


Figure 8. Temperature evolution for stoichiometric methane-air combustion at $T_o = 1500\text{K}$, $P_o = 100\text{kPa}$ by means of 2-step mechanism⁵² in comparison with GRI-Mech 3.0. Ignition delay time t_{ign} at which mixture releases maximum heat is under-predicted by nearly an order of magnitude by the 2-step mechanism.

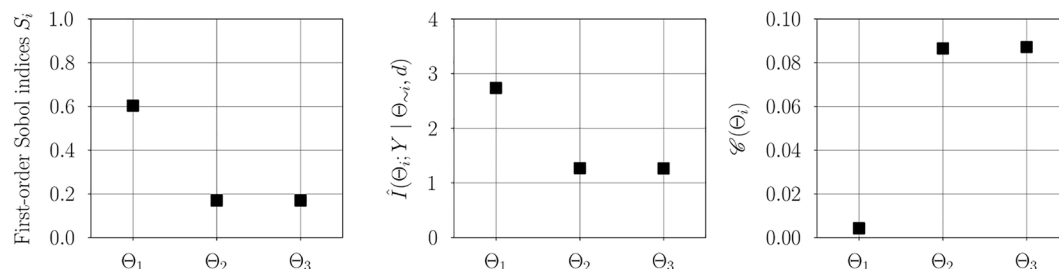


Figure 9. First-order Sobol effect indices (left), information gain (center), and information gain equivalent variance $\mathcal{V}(\Theta_i)$ (right) for methane-air combustion model. Sobol indices show that the largest variability in the output of the statistical model is to uncertainty in Θ_1 ; Θ_2 and Θ_3 exhibit similar variabilities. Variable Θ_1 exhibits the information gain and therefore highest practical identifiability; Θ_2 and Θ_3 have similar information gain. Variable Θ_1 exhibits the lowest measurement uncertainty for the direct observation model, followed by similar uncertainty for Θ_2 and Θ_3 .

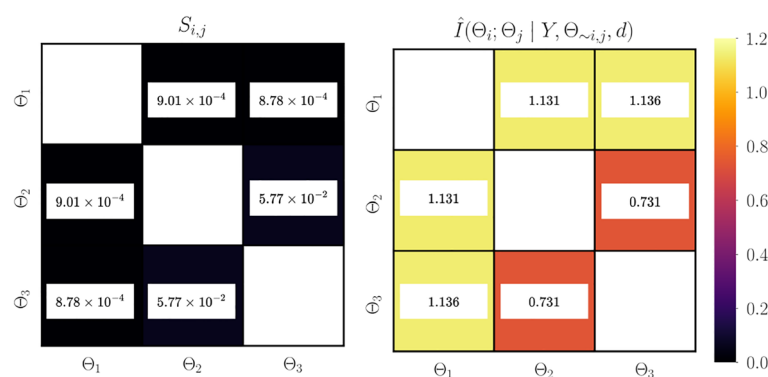


Figure 10. Second-order Sobol indices (left) and estimated parameter dependencies (right) for methane-air combustion model. The trend $S_{2,3} > S_{1,2} \approx S_{1,3}$ suggests that there are underlying interactions between the parameters Θ_2 and Θ_3 . Pairs (Θ_1, Θ_2) and (Θ_1, Θ_3) have nearly the same dependencies on one another. Pair (Θ_2, Θ_3) exhibit low dependencies.

$n_{\text{inner}} = 5$ and 10 quadrature points are used for single and two-dimensional integration space, respectively. Similar magnitude of parameter dependencies obtained for the pairs (Θ_1, Θ_2) and (Θ_1, Θ_3) in addition to similar information gain for Θ_2 and Θ_3 also suggest underlying symmetry with respect to Θ_1 . This means that the interchange of Θ_2 and Θ_3 will not affect the output of the statistical model, which can be clearly seen in (34) for $\phi = 1$. This is also evident from the second-order Sobol indices which suggest that there is a combined effect on the output of the statistical model due to interactions between Θ_2 and Θ_3 .

Parameter estimation

Now, let us consider the parameter estimation problem which seeks $p(\theta | y)$, that is the posterior distribution. Typically, a closed-form expression for the posterior distribution is not available due to the non-linearities in the forward model or the chosen family of the prior distribution. As an alternative, sampling-based methods such as Markov Chain Monte Carlo (MCMC) that seek samples from an unnormalized posterior have gained significant attention. These methods construct Markov chains for which the stationary distribution is the posterior distribution. The Metropolis-Hastings algorithm is an MCMC method that can be used to generate a sequence of samples from any given probability distribution⁵⁵. The adaptive Metropolis algorithm is a powerful modification to the Metropolis-Hastings algorithm and is used here to sample from the posterior distribution⁵⁶.

Figure 11 illustrates the correlation between samples obtained using the Adaptive Metropolis algorithm and the obtained aggregate posterior prediction for ignition delay time. Any (linear) correlation is not observed between the variables; however, the joint distribution between pairs (Θ_1, Θ_2) and (Θ_1, Θ_3) show similarities. These similarities were also observed during the a priori analysis quantifying parameter dependencies as shown in Figure 10.

The obtained aggregate prediction shows dramatic improvement over the 2-step mechanism in predicting ignition delay time over a wide range of temperatures. Using a functional form as (34) for the pre-exponential factor also improved the mixture temperature evolution, as shown in Figure 12. However, the adiabatic flame temperature, which is defined as the mixture temperature upon reaching equilibrium, is still being over-predicted.

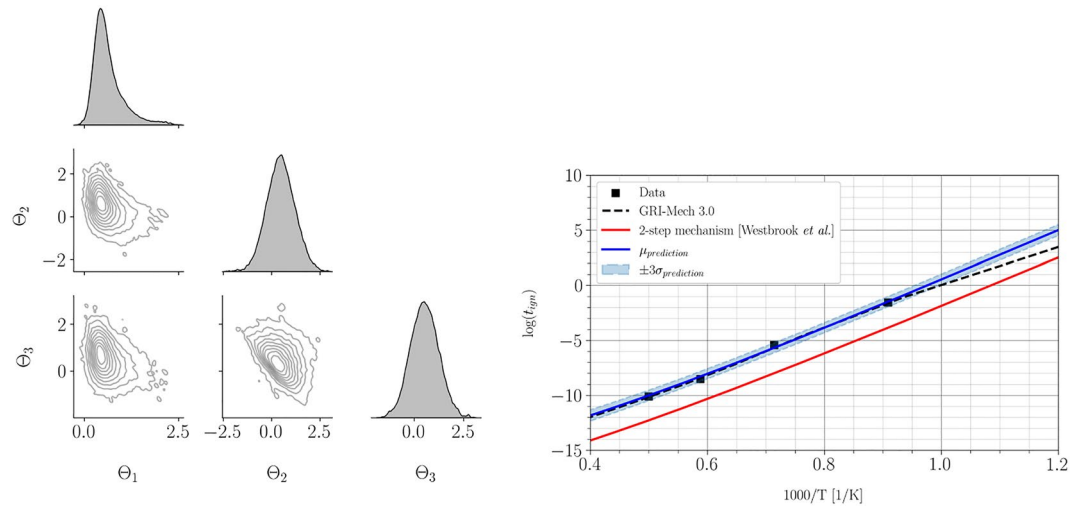


Figure 11. Correlation plot for samples obtained from the posterior distribution (left) and the obtained aggregate posterior prediction (right) for the methane-air combustion model. Correlation plots do not reveal any relations among variables. Aggregate posterior prediction agrees well with the data and exhibits high certainty.

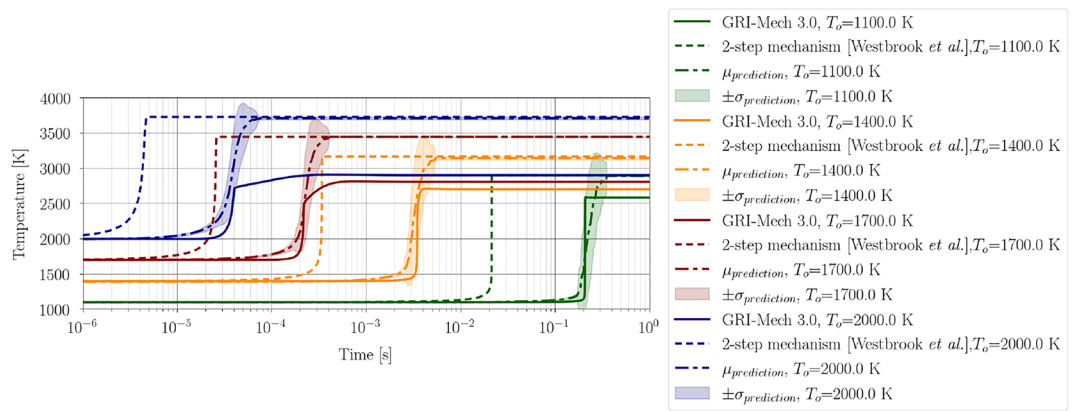


Figure 12. Aggregate temperature evolution for methane-air combustion model. Aggregate prediction agrees well with the GRI-Mech 3.0 detailed mechanism.

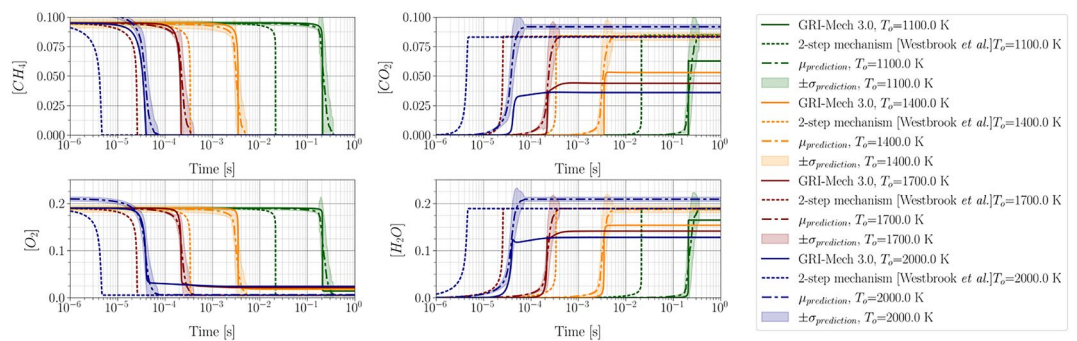


Figure 13. Aggregate species concentration evolution for methane-air combustion model. Aggregate prediction agrees well with the GRI-Mech 3.0 detailed mechanism.

An improvement in the prediction of the evolution of species concentration over time is also noticed, as shown in Figure 13.

Concluding remarks and perspectives

Examining the practical identifiability of statistical models is useful in many applications, such as parameter estimation, model-form development, and model selection. Estimating practical identifiability prior to conducting controlled experiments or parameter estimation studies can assist in a choice of parametrization that can be associated with a high degree posterior certainty, thus improving confidence in estimation and model prediction.

In this work, a novel information-theoretic approach based on conditional mutual information is presented to assess global practical identifiability of a statistical model in a Bayesian framework. The proposed framework examines the expected information gain for each parameter from the data before performing controlled experiments. Parameters with higher information gain are characterized by having higher posterior certainty, and thereby have higher practical identifiability. The adopted viewpoint is that the practical identifiability of a parameter does not have a binary answer, rather it is the relative practical identifiability among parameters that is useful in practice.

In contrast to previous numerical approaches used to study practical identifiability, the proposed approach has the following notable advantages: first, no controlled experiment or data is required to conduct the practical identifiability analysis; second, different forms of uncertainties, such as model-form, parameter, or measurement can be taken into account; third, the framework does not make assumptions about the distribution of the data and parameters as in the previous methods; fourth, the estimator provides knowledge about global identifiability and is therefore not dependent on a particular realization of the parameters. To provide a physical interpretation to practical identifiability in the context of examining information gain for each parameter, an information gain equivalent variance for a direct observation model is also presented. The practical identifiability framework is then extended to examine dependencies among parameter pairs. Even if an individual parameter exhibits poor practical identifiability characteristics, it can belong to an identifiable subset such that parameters within the subset have functional relationships with one another. Parameters within such an identifiable subset have a combined effect on the statistical model and can be collectively identified. To find such subsets, a novel a priori estimator is proposed to quantify the expected dependencies between parameter pairs that emerge a posteriori.

To illustrate the framework, two statistical models are considered: (a) a linear Gaussian model and (b) a non-linear methane-air reduced kinetics model. For the linear Gaussian model, it is shown that parameters with large information gain and low parameter dependencies can be estimated with high confidence. The variance-based global sensitivity analysis (GSA) also illustrates that parameter sensitivity is necessary for identifiability. However, as conclusively shown, the inability of variance-based GSA to capture different forms of uncertainties can lead to unreliable estimates for practical identifiability. The information gain equivalent variance obtained using a direct observation model shows that parameters with high practical identifiability will be associated with low measurement uncertainty if observed directly. In the case of the methane-air reduced kinetics model, it is shown that parameters with large dependencies can have low information gain and therefore low practical identifiability. Further, the proposed estimator can capture non-linear dependencies and reveal structures within the parameter space before performing controlled experiments. Such non-linear dependencies cannot be observed when considering a posteriori parameter correlations, as only linear relations can be well understood.

Data availability

All data generated or analyzed during this study are included in this published article.

Code availability

Code related to the experiments performed in this article is available at <https://github.com/sahilbhola14/GIM>.

Received: 11 April 2023; Accepted: 10 October 2023

Published online: 26 October 2023

References

- Bellman, R. & Åström, K. J. On structural identifiability. *Math. Biosci.* **7**(3–4), 329–339 (1970).
- Cobelli, C. & Distefano, J. J. 3rd. Parameter and structural identifiability concepts and ambiguities: A critical review and analysis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **239**(1), 7–24 (1980).
- Paulino, C. D. M. & Bragança Pereira, C. A. On identifiability of parametric statistical models. *J. Ital. Stat. Soc.* **3**(1), 125–151 (1994).
- Raue, A., Kreutz, C., Theis, F. J. & Timmer, J. Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **371**(1984), 20110544 (2013).
- Lam, N., Docherty, P. & Murray, R. Practical identifiability of parametrised models: A review of benefits and limitations of various approaches. *Math. Comput. Simul.* **199**, 202–216 (2022).
- Ramancha, M. K., Astroza, R., Madarshahian, R. & Conte, J. P. Bayesian updating and identifiability assessment of nonlinear finite element models. *Mech. Syst. Signal Process.* **167**, 108517 (2022).
- Deussen, P. & Galvanin, F. A model-based experimental design approach to assess the identifiability of kinetic models of hydroxymethylfurfural hydrogenation in batch reaction systems. *Chem. Eng. Res. Des.* **178**, 609–622 (2022).
- Ran, Z.-Y. & Hu, B. G. Parameter identifiability in statistical machine learning: A review. *Neural Comput.* **29**(5), 1151–1203 (2017).
- Qian, G. & Mahdi, A. Sensitivity analysis methods in the biomedical sciences. *Math. Biosci.* **323**, 108306 (2020).
- Tomovic, R., Parezanovic, N. S. & Merritt, M. J. Sensitivity of dynamic systems to parameters which increase the order of mathematical models. *IEEE Trans. Electron. Comput.* **6**, 890–897 (1965).
- Meissinger, H. F. & Bekey, G. A. An analysis of continuous parameter identification methods. *Simulation* **6**(2), 94–102 (1966).
- Tortorelli, D. A. & Michaleris, P. Design sensitivity analysis: Overview and review. *Inverse Probl. Eng.* **1**(1), 71–105 (1994).
- Saltelli, A. *et al.* *Global Sensitivity Analysis: The Primer* (Wiley, 2008).

14. Dobre, S., Bastogne, T. & Richard, A. Global sensitivity and identifiability implications in systems biology. *IFAC Proc. Vol.* **43**(6), 54–59 (2010).
15. Staley, R. M. & Yue, P. C. On system parameter identifiability. *Inf. Sci.* **2**(2), 127–138 (1970).
16. Rothenberg, T. J. Identification in parametric models. *Econom. J. Econom. Soc.* **39**, 577–591 (1971).
17. Stoica, P. & Söderström, T. On non-singular information matrices and local identifiability. *Int. J. Control* **36**(2), 323–329 (1982).
18. Petersen, B., Gernaey, K. & Vanrolleghem, P. A. Practical identifiability of model parameters by combined respirometric-titrimetric measurements. *Water Sci. Technol.* **43**(7), 347–355 (2001).
19. Weijers, S. R. & Vanrolleghem, P. A. A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. *Water Sci. Technol.* **36**(5), 69–79 (1997).
20. Machado, V. C., Tapia, G., Gabriel, D., Lafuente, J. & Baeza, J. A. Systematic identifiability study based on the fisher information matrix for reducing the number of parameters calibration of an activated sludge model. *Environ. Model. Softw.* **24**(11), 1274–1284 (2009).
21. Raue, A. *et al.* Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**(15), 1923–1929 (2009).
22. Raue, A., Kreutz, C., Maiwald, T., Klingmüller, U. & Timmer, J. Addressing parameter identifiability by model-based experimentation. *IET Syst. Biol.* **5**(2), 120–130 (2011).
23. Simpson, M. J., Baker, R. E., Vittadello, S. T. & Maclaren, O. J. Practical parameter identifiability for spatio-temporal models of cell invasion. *J. R. Soc. Interface* **17**(164), 20200055 (2020).
24. Eisenberg, M. C. & Hayashi, M. A. Determining identifiable parameter combinations using subset profiling. *Math. Biosci.* **256**, 116–126 (2014).
25. Maiwald, T. *et al.* Driving the model to its limit: Profile likelihood based model reduction. *PLoS One* **11**(9), 0162366 (2016).
26. Ran, Z. Y. & Hu, B. G. Determining parameter identifiability from the optimization theory framework: A Kullback-Leibler divergence approach. *Neurocomputing* **142**, 307–317 (2014).
27. Ran, Z. Y. & Hu, B. G. An identifying function approach for determining parameter structure of statistical learning machines. *Neurocomputing* **162**, 209–217 (2015).
28. Archer, G. E. B., Saltelli, A. & Sobol, I. M. Sensitivity measures, anova-like techniques and the use of bootstrap. *J. Stat. Comput. Simul.* **58**(2), 99–120 (1997).
29. Saltelli, A., Tarantola, S. & Campolongo, F. Sensitivity analysis as an ingredient of modeling. *Stat. Sci.* **15**, 377–395 (2000).
30. Kucherenko, S., Feil, B., Shah, N. & Mauntz, W. The identification of model effective dimensions using global sensitivity analysis. *Reliability Engineering & System Safety* **96**(4), 440–449 (2011).
31. Wu, X., Shirvan, K. & Kozłowski, T. Demonstration of the relationship between sensitivity and identifiability for inverse uncertainty quantification. *J. Comput. Phys.* **396**, 12–30 (2019).
32. Pant, S. & Lombardi, D. An information-theoretic approach to assess practical identifiability of parametric dynamical systems. *Math. Biosci.* **268**, 66–79 (2015).
33. Capellari, G., Chatzi, E., Mariani, S., *et al.* Parameter identifiability through information theory. In Proc. of the 2nd ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP), 15–17 (2017).
34. Ebrahimiyan, H., Astroza, R., Conte, J. P. & Bitmead, R. R. Information-theoretic approach for identifiability assessment of nonlinear structural finite-element models. *J. Eng. Mech.* **145**(7), 04019039 (2019).
35. Pant, S. Information sensitivity functions to assess parameter information gain and identifiability of dynamical systems. *J. R. Soc. Interface* **15**(142), 20170871 (2018).
36. Aggarwal, A., Lombardi, D. & Pant, S. An information-theoretic framework for optimal design: Analysis of protocols for estimating soft tissue parameters in biaxial experiments. *Axioms* **10**(2), 79 (2021).
37. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948).
38. Cover, T. M. *Elements of Information Theory* (Wiley, 1999).
39. Särkkä, S. *Bayesian Filtering and Smoothing* (Cambridge University Press, 2013).
40. Ryan, K. J. Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *J. Comput. Graph. Stat.* **12**(3), 585–603 (2003).
41. Huan, X. & Marzouk, Y. M. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. Comput. Phys.* **232**(1), 288–317 (2013).
42. Herman, J. & Usher, W. SALib: An open-source python library for sensitivity analysis. *J. Open Sour. Softw.* <https://doi.org/10.21105/joss.00097> (2017).
43. Iwanaga, T., Usher, W. & Herman, J. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. *Socio-Environ. Syst. Model.* **4**, 18155. <https://doi.org/10.18174/sesmo.18155> (2022).
44. Bhattacharjee, B., Schwer, D. A., Barton, P. I. & Green, W. H. Optimally-reduced kinetic models: Reaction elimination in large-scale kinetic mechanisms. *Combust. Flame* **135**(3), 191–208 (2003).
45. Peters, N. & Rogg, B. *Reduced Kinetic Mechanisms for Applications in Combustion Systems* (Springer, 2008).
46. Pepiot-Desjardins, P. & Pitsch, H. An efficient error-propagation-based reduction method for large chemical kinetic mechanisms. *Combust. Flame* **154**(1–2), 67–81 (2008).
47. Kelly, M., Dooley, S. & Bourque, G. Toward machine learned highly reduced kinetic models for methane/air combustion. *Turbo Expo Power Land Sea Air* **84942**, 03–04005 (2021).
48. Hakim, L., Lacaze, G., Khalil, M., Najm, H. N. & Oefelein, J. C. Modeling auto-ignition transients in reacting diesel jets. *J. Eng. Gas Turbines Power* **138**(11), 112806 (2016).
49. Vajda, S., Rabitz, H., Walter, E. & Lecourtier, Y. Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. *Chem. Eng. Commun.* **83**(1), 191–219 (1989).
50. Davidson, D. F. & Hanson, R. K. Interpreting shock tube ignition data. *Int. J. Chem. Kinet.* **36**(9), 510–523 (2004).
51. Chaos, M. & Dryer, F. L. Chemical-kinetic modeling of ignition delay: Considerations in interpreting shock tube data. *Int. J. Chem. Kinet.* **42**(3), 143–150 (2010).
52. Westbrook, C. K. & Dryer, F. L. Simplified reaction mechanisms for the oxidation of hydrocarbon fuels in flames. *Combust. Sci. Technol.* **27**(1–2), 31–43 (1981).
53. Goodwin, D.G., Moffat, H.K., Schoegl, I., Speth, R.L., Weber, B.W. Cantera: An object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes. <https://www.cantera.org>. Version 2.6.0 (2022). <https://doi.org/10.5281/zenodo.6387882>
54. Smith, G.P., Golden, D.M., Frenklach, M., Moriarty, N.W., Eiteneer, B., Goldenberg, M., Bowman, C.T., Hanson, R.K., Song, S., Gardiner Jr., W.C., Lissianski, V.V., Zhiwei: GRI-Mech 3.0 (1999). http://www.me.berkeley.edu/gri_mech/
55. Tierney, L. Markov chains for exploring posterior distributions. *Ann. Stat.* **22**, 1701–1728 (1994).
56. Haario, H., Saksman, E. & Tamminen, J. An adaptive metropolis algorithm. *Bernoulli* **7**, 223–242 (2001).

Author contributions

S.B. primarily developed the methodology and generated numerical results. K.D. formulated the problem statement and provided guidance for the work. All authors reviewed the manuscript.

Funding

The authors disclose support for the research of this work from OUSD(RE) Grant No: N00014-21-1-295.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-44589-3>.

Correspondence and requests for materials should be addressed to S.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023