

---

# A GUIDE TO BAYESIAN INFERENCE FOR REGRESSION PROBLEMS

---

Clemens Elster<sup>1</sup>, Katy Klauenberg<sup>1</sup>, Monika Walzel<sup>1</sup>, Gerd Wübbeler<sup>1</sup>, Peter Harris<sup>2</sup>, Maurice Cox<sup>2</sup>, Clare Matthews<sup>2</sup>, Ian Smith<sup>2</sup>, Louise Wright<sup>2</sup>, Alexandre Allard<sup>3</sup>, Nicolas Fischer<sup>3</sup>, Simon Cowen<sup>4</sup>, Steve Ellison<sup>4</sup>, Philip Wilson<sup>4</sup>, Francesca Pennecchi<sup>5</sup>, Gertjan Kok<sup>6</sup>, Adriaan van der Veen<sup>6</sup>, Leslie Pendrill<sup>7</sup>

<sup>1</sup>Physikalisch-Technische Bundesanstalt (PTB), Germany

<sup>2</sup>National Physical Laboratory (NPL), UK

<sup>3</sup>Laboratoire National de métrologie et d'essais (LNE), France

<sup>4</sup>LGC Ltd., UK

<sup>5</sup>Istituto Nazionale Di Ricerca Metrologica (INRIM), Italy

<sup>6</sup>VSL – Dutch Metrology Institute, the Netherlands

<sup>7</sup>SP Technical Research Institute of Sweden, Sweden

This document<sup>a</sup> is a deliverable of project NEW04 “Novel mathematical and statistical approaches to uncertainty evaluation” (08/2012-07/2015) funded by the European Metrology Research Programme (EMRP).

The document may not be copied or published for resale.

---

<sup>a</sup> To be cited as: C. Elster, K. Klauenberg, M. Walzel, G. Wübbeler, P. Harris, M. Cox, C. Matthews, I. Smith, L. Wright, A. Allard, N. Fischer, S. Cowen, S. Ellison, P. Wilson, F. Pennecchi, G. Kok, A. van der Veen, L. Pendrill, *A Guide to Bayesian Inference for Regression Problems*, Deliverable of EMRP project NEW04 “Novel mathematical and statistical approaches to uncertainty evaluation”, 2015.

# Contents

<b>Preface</b>	<b>I</b>
<b>Glossary and notation</b>	<b>II</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Regression model . . . . .	1
1.2 Bayesian inference . . . . .	1
1.3 Methods from classical statistics . . . . .	2
1.4 Model checking . . . . .	3
1.5 Case studies . . . . .	4
<b>2 Steps in a Bayesian inference</b>	<b>5</b>
2.1 Statistical modelling . . . . .	5
2.2 Prior distributions . . . . .	6
2.3 Numerical methods . . . . .	9
2.4 Sensitivity analysis . . . . .	12
2.5 Further aspects . . . . .	13
<b>3 Straight line fit – Normal linear regression using conjugate priors</b>	<b>14</b>
3.1 Statistical modelling . . . . .	14
3.2 Prior distributions . . . . .	15
3.3 Numerical methods . . . . .	16
3.4 Sensitivity analysis . . . . .	18
<b>4 Flow meter calibration – Normal linear regression with constraints</b>	<b>21</b>
4.1 Statistical modelling . . . . .	21
4.2 Prior distributions . . . . .	22
4.3 Numerical methods . . . . .	24
4.4 Sensitivity analysis . . . . .	26
<b>5 Inference of thermophysical properties – use of hierarchical priors</b>	<b>29</b>
5.1 Statistical modelling . . . . .	29
5.2 Prior distributions . . . . .	30
5.3 Numerical methods . . . . .	31
5.4 Sensitivity analysis . . . . .	33
<b>6 Analysis of immunoassay data – a statistical calibration problem</b>	<b>35</b>
6.1 Statistical modelling . . . . .	35
6.2 Prior distributions . . . . .	36
6.3 Numerical methods . . . . .	38
6.4 Sensitivity analysis . . . . .	41
<b>7 Analysis of digital PCR data – a parametric inverse problem</b>	<b>43</b>
7.1 Statistical modelling . . . . .	43
7.2 Prior distributions . . . . .	44
7.3 Numerical methods . . . . .	46
7.4 Sensitivity analysis . . . . .	47
<b>8 Acknowledgements</b>	<b>49</b>
<b>References</b>	<b>50</b>

<b>A</b>	<b>Algorithms and source code</b>	<b>54</b>
A.1	MATLAB <sup>®</sup> software for <i>straight line fit</i> . . . . .	54
A.2	MATLAB <sup>®</sup> software for <i>flow meter calibration</i> . . . . .	58
A.3	Metropolis-Hastings algorithm for <i>inference of thermophysical properties</i> . .	63
A.4	WinBUGS software for <i>analysis of immunoassay data</i> . . . . .	64
A.5	Metropolis-Hastings algorithm for <i>analysis of digital PCR data</i> . . . . .	66
<b>B</b>	<b>Data for case studies</b>	<b>67</b>
B.1	Data for <i>straight line fit</i> . . . . .	67
B.2	Data for <i>flow meter calibration</i> . . . . .	68
B.3	Data for the <i>analysis of immunoassay data</i> . . . . .	69

## Preface

This *Guide* provides practical guidance on Bayesian inference for regression problems. In order to benefit from this *Guide*, the reader should be familiar with probability theory, statistics and mathematical calculus at least to a level that gives an understanding of (the principles of) the “GUM” [9] and its supplements, which are the primary documents regarding measurement uncertainty evaluation in metrology. However, even without following this document in all its detail, the presented real-life case studies illustrate the potential of a Bayesian inference. The provided software and algorithms can serve as template solutions for treating similar problems. Albeit the types of regression problems in this *Guide* originate from the considered case studies and do not cover all possible regression scenarios, the guidance given here should find broad applicability.

One advantage of Bayesian inference is the possibility to account for available prior knowledge. Prior information can, for example, be gained from previous experiments, through expert opinion or from available knowledge about the underlying physics, chemistry or biology, such as non-negativity of concentrations. Thorough elicitation of prior knowledge can be challenging. However, including prior information into the analysis often leads to more reliable estimates and smaller uncertainties, and in some cases is essential to obtain meaningful results at all. The real-life case studies presented in this *Guide* exemplify how prior knowledge can be elicited and utilized. Another advantage of the Bayesian approach is that the information gained in one experiment can be taken into account completely in the analysis of a subsequent, related experiment. This is particularly important for reliable uncertainty propagation. A relevant example is when a calibration curve – estimated from one set of measurements – is used for inferences in subsequent, related measurements, and one of the case studies will address this task.

Practical challenges in the application of Bayesian methods often are the selection of a prior distribution and the computation of numerical results. We provide an overview about the selection of a prior distribution that reflects available prior knowledge, and the case studies exemplify this process in several practical situations. Detailed guidance on computational methods is also given, and specific algorithms and software are presented for the case studies.

This *Guide* is one deliverable of EMRP<sup>1</sup> project *NEW04* [18] concerned with the development and application of statistical methods for uncertainty evaluation in metrology. One motivation is that the “GUM” provides very little guidance on the treatment of regression problems. The “GUM” contains elements from both classical and Bayesian statistics, and generally it leads to different results than a Bayesian inference [17]. Since the “GUM” is currently being revised with the intention to align it with the Bayesian point of view [8], and as neither the “GUM” nor its current supplements deal with Bayesian regression, there is a need for corresponding guidance in metrology. This document is intended to provide a basis for future guidelines in this direction. In addition to its potential relevance for metrology, this *Guide* may also be beneficial for scientists in other fields.

The structure of this *Guide* is as follows. In the introduction we specify the types of regression problems considered, outline a Bayesian inference in general terms, and briefly describe the case studies. The following chapter describes in detail the steps in a Bayesian inference, namely the specification of the statistical model, the choice of a prior distribution, the numerical calculation of results, and the analysis of their sensitivity. Each of the subsequent chapters then presents a single case study exemplifying all these steps. The Appendices provide software source code as well as data for the considered case studies.

---

<sup>1</sup> European Metrology Research Programme (EMRP) ([www.emrponline.eu](http://www.emrponline.eu), accessed on 16 February 2015).

## Glossary and notation

The following symbols are used in this *Guide*.

$p$	number of unknown parameters in regression function
$n$	number of observations
$q$	number of unknown additional parameters in likelihood function
$x$	independent or regressor variable
$y$	dependent or response variable
$a \propto b$	$a$ is proportional to $b$
$\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$	unknown parameters in regression function
$\mathbf{y} = (y_1, \dots, y_n)^\top$	observations
$\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^\top$	unknown additional parameters in likelihood function
$\mathbf{X}$	design matrix for linear regression problem
$\mathbf{I}$	identity matrix of appropriate dimension
$f_{\boldsymbol{\theta}}(x)$	regression function with unknown parameters $\boldsymbol{\theta}$ and independent variable $x$
$I_P(\theta)$	credible interval for $\theta$ with credibility $P$
$p(\mathbf{y} \boldsymbol{\theta}, \boldsymbol{\delta})$	sampling distribution for $\mathbf{y}$ given $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$
$l(\boldsymbol{\theta}, \boldsymbol{\delta}; \mathbf{y})$	likelihood function for $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ given the observation $\mathbf{y}$
$\pi(\boldsymbol{\theta}, \boldsymbol{\delta})$	prior distribution for $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$
$\pi(\boldsymbol{\theta}, \boldsymbol{\delta} \mathbf{y})$	posterior distribution for $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ given $\mathbf{y}$
$U(a, b)$	rectangular distribution with probability density $g(\theta) = 1/(b - a), \quad a \leq \theta \leq b$
$N(\theta_0, \sigma^2)$	Normal distribution with probability density $g(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\theta - \theta_0)^2}$
$N(\boldsymbol{\theta}_0, \mathbf{V})$	multivariate Normal distribution with probability density $g(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{p/2} \sqrt{\det(\mathbf{V})}} e^{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{V}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}$
$t_\nu(\theta_0, \sigma^2)$	scaled and shifted $t$ -distribution with $\nu$ degrees of freedom and probability density $g(\theta) = \frac{1}{\sigma} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{1}{\nu} \{(\theta - \theta_0)/\sigma\}^2\right)^{-\frac{\nu+1}{2}}$
$t_\nu(\boldsymbol{\theta}_0, \mathbf{V})$	scaled and shifted multivariate $t$ -distribution with $\nu$ degrees of freedom and probability density $g(\boldsymbol{\theta}) = \frac{\Gamma[(\nu+p)/2]}{\Gamma(\nu/2)\nu^{p/2}\pi^{p/2}(\det(\mathbf{V}))^{1/2}} \left(1 + \frac{1}{\nu}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{V}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right)^{-(\nu+p)/2}$
$\Gamma(\alpha, \beta)$	Gamma distribution with probability density $g(\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \theta^{\alpha-1} e^{-\theta/\beta}, \quad \theta > 0$
$\text{IG}(\alpha, \beta)$	inverse Gamma distribution with probability density $g(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-\alpha-1} e^{-\beta/\theta}, \quad \theta > 0$

NIG( $\boldsymbol{\theta}_0, \mathbf{V}, \alpha, \beta$ )	Normal inverse Gamma distribution where NIG( $\boldsymbol{\theta}_0, \mathbf{V}, \alpha, \beta$ ) = N( $\boldsymbol{\theta}_0, \sigma^2 \mathbf{V}$ )IG( $\alpha, \beta$ ), with probability density $g(\boldsymbol{\theta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{p/2} \sqrt{\det(\mathbf{V})}} e^{-\frac{1}{2\sigma^2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{V}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)} \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} e^{-\beta/\sigma^2}$
Beta( $\alpha, \beta$ )	Beta distribution with probability density $g(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}, \quad 0 \leq \theta \leq 1$
$\Phi(\theta)$	distribution function for standard Normal distribution $\Phi(\theta) = \int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi}} e^{-\zeta^2/2} d\zeta$
$\Phi(\theta; \theta_0, \sigma^2)$	distribution function for Normal distribution with mean $\theta_0$ and variance $\sigma^2$ $\Phi(\theta; \theta_0, \sigma^2) = \Phi((\theta - \theta_0)/\sigma)$
CMP( $x; \lambda, \nu$ )	Conway-Maxwell Poisson probability mass function $\text{CMP}(x; \lambda, \nu) = \frac{\lambda^x}{(x!)^\nu} \frac{1}{Z(\lambda, \nu)}, \quad x = 0, 1, 2, \dots, \lambda, \nu > 0,$ where $Z(\lambda, \nu) = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^\nu}$

Degree of belief or state of knowledge probability density functions are denoted by  $\pi()$ , and often loosely termed ‘distributions’. The random variables to which these distributions belong are identified through the arguments given to  $\pi()$ . For example,  $\pi(\theta, \sigma^2)$  is the distribution for the random variables  $\theta$  and  $\sigma^2$ . The same symbol ( $\theta$ , for example) is used to denote both the random variable and possible values a variable can take. The intended meaning should follow from the context. Distributions for observable random variables are denoted by  $p()$ . Again, the function  $p()$  is identified through its arguments. For example,  $p(y|\theta)$  denotes the sampling distribution for an observable  $y$  given  $\theta$ . Furthermore,  $y$  denotes the random variable, an observed realization of it, or possible values the variable can take.

An expression of the form  $\pi(\theta|\theta_0, \sigma^2) = \text{N}(\theta_0, \sigma^2)$  is understood in the sense of  $\theta|\theta_0, \sigma^2 \sim \text{N}(\theta_0, \sigma^2)$ , where ‘ $\sim$ ’ means ‘is distributed as’. That is,  $\theta$  is normally distributed with mean  $\theta_0$  and variance  $\sigma^2$ . The expression

$$\epsilon_i \stackrel{iid}{\sim} \text{N}(0, \sigma^2), \quad i = 1, \dots, n, \quad (1)$$

indicates that  $\epsilon_1, \dots, \epsilon_n$  are independent and identically distributed according to a Normal distribution with zero mean and variance  $\sigma^2$ .

Whenever bounds are omitted in an integral, the domain extends to the support of the integrand. For example, the integral  $\int \pi(\theta, \sigma) d\sigma$  is understood as  $\int_0^\infty \pi(\theta, \sigma) d\sigma$ , since  $\pi(\theta, \sigma) = 0$  for  $\sigma < 0$ .

In the descriptions of the steps in a Bayesian inference for a general regression problem, a generic notation is used. For example,  $x$  is used for the independent variable,  $\boldsymbol{\theta}$  for the unknown parameters of the regression function, etc. However, in the description of each case study, a more suggestive notation is often used that is more natural for the application. For example, in the case study concerned with the inference of thermophysical properties,  $t$  denoting time is used for the independent variable, and  $(\tau, B_i)^\top$  is used for the unknown parameters of the regression function, where  $\tau$  denotes a characteristic time and  $B_i$  a biot number.

# 1 Introduction

In this chapter the considered type of regression problems is specified and a brief introduction to Bayesian inference is given. We also mention alternative methods from classical statistics and point towards tools for model checking. Finally, the case studies are briefly introduced.

## 1.1 Regression model

We consider regression models of the form

$$y_i = f_{\boldsymbol{\theta}}(x_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (2)$$

where  $\mathbf{y} = (y_1, \dots, y_n)^\top$  denotes the observed data corresponding to known values  $\mathbf{x} = (x_1, \dots, x_n)^\top$ ,  $f_{\boldsymbol{\theta}}(x)$  is a given function with unknown parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ , and the errors  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$  follow a specified distribution

$$p(\boldsymbol{\epsilon} | \boldsymbol{\theta}, \boldsymbol{\delta}). \quad (3)$$

Typically, the unknown additional parameters  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^\top$  in the distribution (3) are variance parameters, but the distribution of  $\boldsymbol{\epsilon}$  may in general also depend on  $\boldsymbol{\theta}$ , for example, in the case of a constant relative error model such as

$$\epsilon_i \sim N(0, (\sigma f_{\boldsymbol{\theta}}(x_i))^2), \quad i = 1, \dots, n. \quad (4)$$

Throughout this *Guide* we assume that  $x_1, \dots, x_n$  are known exactly. If this is not the case, the methods presented in this *Guide* can still be applied as long as the uncertainties about the  $x_i$  are negligible compared to those about the  $y_i$ . For ease of notation we often denote the data simply by  $\mathbf{y}$  instead of writing  $(\mathbf{x}, \mathbf{y})$ . Principally,  $x$  and  $y$  might be univariate or multivariate, albeit in this *Guide* they are throughout univariate. We refer to  $f_{\boldsymbol{\theta}}(x)$  as the regression function, to  $y$  as the dependent or response variable, and to  $x$  as the independent or regressor variable.

A simple example is the Normal straight line regression model

$$y_i = \theta_1 + \theta_2 x_i + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), \quad i = 1, \dots, n, \quad (5)$$

which may be used to describe the relationship between a traceable, highly-accurate reference device with values denoted by  $x$  and a device to be calibrated with values denoted by  $y$ . The pairs  $(x_i, y_i)$  then denote simultaneous measurements made by the two devices of the same measurand such as, for example, temperature.

The basic goal in the treatment of a regression model is to *estimate* the unknown parameters  $\boldsymbol{\theta}$  of the regression function, and possibly also the unknown variance parameters  $\boldsymbol{\delta}$ . In some cases, the estimated regression model is subsequently used for the *prediction* of the independent variable given one or several future observations of the dependent variable. The combination of both tasks is also known as *statistical calibration* (see, e.g., [13]). In other cases, the regression model is used to estimate a quantity that depends on the regression parameters. Examples include the value of the regression function and a derivative of the function for a specified value of the independent variable, or the integral of the function between two values of the independent variable.

## 1.2 Bayesian inference

In a Bayesian inference probability distributions are used to encode one's prior knowledge about  $\boldsymbol{\theta}$  and  $\boldsymbol{\delta}$ . Then the data  $\mathbf{y}$  are taken into account to update the prior belief about these

parameters. Technically, this is done through Bayes' theorem:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\delta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}, \boldsymbol{\delta}) \times l(\boldsymbol{\theta}, \boldsymbol{\delta}; \mathbf{y}). \quad (6)$$

$\pi(\boldsymbol{\theta}, \boldsymbol{\delta})$  is the prior distribution expressing the prior knowledge about the parameters,  $l(\boldsymbol{\theta}, \boldsymbol{\delta}; \mathbf{y})$  denotes the likelihood function, and  $\pi(\boldsymbol{\theta}, \boldsymbol{\delta} | \mathbf{y})$  is the posterior distribution that combines the prior belief about  $\boldsymbol{\theta}$  and  $\boldsymbol{\delta}$  with the information contained in the data. The likelihood  $l(\boldsymbol{\theta}, \boldsymbol{\delta}; \mathbf{y})$  equals the sampling distribution  $p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\delta})$ , viewed as a function of the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\delta}$  for the given data  $\mathbf{y}$ . It is thus determined by the statistical model (3) for the errors and the regression model (2) together with the observed data  $\mathbf{y}$ .

The posterior  $\pi(\boldsymbol{\theta}, \boldsymbol{\delta} | \mathbf{y})$  is a conditional distribution, and it can be used to make probability statements about the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\delta}$  after the data have been observed. Application of probability calculus allows the determination of marginal distributions such as

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \int \pi(\boldsymbol{\theta}, \boldsymbol{\delta} | \mathbf{y}) \, d\boldsymbol{\delta}. \quad (7)$$

The marginal posterior distribution (7) expresses our complete knowledge about  $\boldsymbol{\theta}$ . Occasionally, one may want to summarize this knowledge in terms of an estimate such as, for example, the posterior mean

$$E(\boldsymbol{\theta} | \mathbf{y}) = \int \tilde{\boldsymbol{\theta}} \pi(\tilde{\boldsymbol{\theta}} | \mathbf{y}) \, d\tilde{\boldsymbol{\theta}}, \quad (8)$$

together with a measure of the spread of the distribution such as the posterior covariance matrix

$$\text{Cov}(\boldsymbol{\theta} | \mathbf{y}) = \int (\tilde{\boldsymbol{\theta}} - E(\boldsymbol{\theta} | \mathbf{y}))(\tilde{\boldsymbol{\theta}} - E(\boldsymbol{\theta} | \mathbf{y}))^\top \pi(\tilde{\boldsymbol{\theta}} | \mathbf{y}) \, d\tilde{\boldsymbol{\theta}}. \quad (9)$$

Alternatively, one may want to determine a credible region  $\boldsymbol{\Omega}$  that contains  $\boldsymbol{\theta}$  with a high probability  $P$  (e.g., 0.95), satisfying

$$\int_{\boldsymbol{\Omega}} \pi(\boldsymbol{\theta} | \mathbf{y}) \, d\boldsymbol{\theta} = P. \quad (10)$$

For a single parameter  $\theta$  one often quotes a 95 % credible interval  $I_{0.95}(\theta) = [\underline{\theta}, \bar{\theta}]$ , where

$$\int_{\underline{\theta}}^{\bar{\theta}} \pi(\theta | \mathbf{y}) \, d\theta = 0.95. \quad (11)$$

In general, condition (11) does not determine a credible interval uniquely, and an appropriate choice of the credible interval will depend on the application. For example, for unimodal distributions with support  $(-\infty, \infty)$  the shortest interval (called highest posterior density (HPD) interval) may be recommended. Often also probabilistically symmetric intervals are reported, for which the probability of  $\theta$  exceeding  $\bar{\theta}$  equals the probability that  $\theta$  is smaller than  $\underline{\theta}$ . For unimodal, symmetric distributions the two intervals coincide.

For a general introduction to Bayesian inference and further reading we refer to, for example, [4, 11, 21, 55].

### 1.3 Methods from classical statistics

Regression problems are an important topic also in classical statistics, and corresponding methods are available. Albeit this *Guide* focuses on Bayesian methods for regression problems, we briefly mention two other methods.



Maximum likelihood estimation is one popular technique for parameter estimation in classical statistics and it is often used for solving regression problems. The method estimates the model parameters by maximizing the likelihood function. Confidence intervals for the resulting parameter estimates may then be derived on the basis of asymptotic results [54] or by bootstrap methods [16]. Software for maximum likelihood estimation is widely available, for instance in R [48] and MATLAB® [32].

Least-squares methods (see, e.g., [19]) provide another popular tool for the treatment of regression problems. In their simplest form these methods determine estimates of the parameters by minimizing

$$\sum_{i=1}^n (y_i - f_{\theta}(x_i))^2 \quad (12)$$

with respect to the parameters. For regression problems with errors that are Normally distributed with constant variance (e.g., as in (5)) the resulting parameter estimates are equivalent to those obtained by maximum likelihood estimation. In general, however, least-squares methods lead to estimates that are different from those obtained from maximum likelihood estimation. Note that maximum likelihood estimation is also capable of providing estimates of (unknown) variance parameters.

One advantage of least-squares methods is their simplicity. For example, for Normal linear regression models their numerical implementations are essentially based on numerical linear algebra. Software for least-squares estimation is widely available (e.g., in R [48] and MATLAB® [32]). We refer in particular to [47] for software developed within the *NEW04* project for the least-squares treatment of calibration problems.

## 1.4 Model checking

Any statistical inference is conditional on the statistical model, and any Bayesian inference depends also on the chosen prior. It is therefore desirable to assess the plausibility of both in the light of the data, which in turn may motivate some refinement of them.

When fitting a regression function to data one usually considers the (standardized) residuals, i.e., the differences of the data and the estimated regression function (divided by the estimated standard deviations of the data). For example, when the errors  $\epsilon_1, \dots, \epsilon_n$  in (2) are assumed to be independently and identically distributed, a plot of the residuals should essentially look ‘random’, provided that the estimate of the regression function is ‘fairly good’. A deficiency in the chosen model for the regression function or in the employed distribution of the errors can often be detected visually as a systematic pattern in the residuals. We provide residual plots in our treatment of the case studies. Alternatively, formal tests can be applied in a residual analysis, including goodness-of-fit tests from classical statistics such as a Kolmogorov Smirnov test or a  $\chi^2$ -test (e.g., [54]), to test the conformity of the residuals with the assumed distribution of the errors.

In recent years also advanced Bayesian tools have been developed for formal model checking (e.g., [3, 21, 22, 23, 58]). These tools appear to be similar to those developed in classical statistics, but they have a different methodological basis and are more flexible, for example, in handling unknown parameters, missing data, etc. One tool that is used in this context is the so-called posterior predictive distribution. This distribution depends on the employed statistical model and the Bayesian posterior. Therefore it also accounts for the chosen prior distribution. The posterior predictive distribution can be used to produce extra data that should look similar to the given data if the model is adequate. Depending on the purpose

of the model, different discrepancy measures can be defined to reflect this similarity for some aspect of the data. These test quantities can then be displayed graphically or tested by calculating tail area probabilities, so-called posterior predictive  $p$ -values. The posterior predictive distribution can also be used in connection with cross validation, where one can select a subset of the observations, and compare this subset with the corresponding posterior predictive distribution when the latter is determined by all observations excluding those of the selected subset (cf. [3, 22, 58]).

We note that model checking can only address single characteristics in the conformity of model and data, rather than confirming the adequacy of the whole model [28, p. 192], and ideally should be accompanied with a critical assessment of the results by experts in the field.

## 1.5 Case studies

The following five case studies serve as examples in this *Guide* to illustrate the above steps of Bayesian inference. Each case study highlights a different characteristic with respect to the model, the prior distribution, or the numerical methods.

1. *Straight line fit – Normal linear regression using conjugate priors.* This case study demonstrates the use of conjugate prior distributions that leads to an analytic expression for the posterior distribution. The underlying Normal linear model is a simple yet very common model.
2. *Flow meter calibration – Normal linear regression with constraints.* This case study illustrates a Normal linear regression problem in which prior knowledge is available in terms of constraints on the regression function. This requires numerical methods, and we present a simple Monte Carlo procedure that can be applied here.
3. *Inference of thermophysical properties – use of hierarchical priors.* This case study illustrates the use of hierarchical prior distributions. Furthermore, the regression function is not known explicitly but is determined through the numerical solution of a partial differential equation.
4. *Analysis of immunoassay data – a statistical calibration problem.* In this case study the variances of the observations are not constant, and we also illustrate how prior distributions can be derived from historical measurements. The task is also one of statistical calibration, and we demonstrate how the posterior distribution determined in the calibration stage can be used as (part of) the prior distribution for the subsequent prediction.
5. *Analysis of digital PCR data – a parametric inverse problem.* This case study represents a parametric inverse-problem rather than a regression task. Here we demonstrate that such problems can be treated by the very same Bayesian methodology as the preceding regression tasks.

## 2 Steps in a Bayesian inference

This chapter describes the steps carried out in a Bayesian inference. These steps comprise the specification of the statistical model, the selection of the prior distribution, the numerical calculation of the results and the determination of their sensitivity. We also explain the particular features of the case studies with respect to each step.

### 2.1 Statistical modelling

The specification of the statistical model is a first step in a Bayesian inference. The statistical model describes the relationship between the observed data  $\mathbf{y}$  and all relevant influencing parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\delta}$  in a probabilistic sense. More precisely, the data are considered to be a sample randomly drawn from the sampling distribution  $p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\delta})$ . The likelihood function  $l(\boldsymbol{\theta}, \boldsymbol{\delta}; \mathbf{y})$  required for the application of Bayes' theorem (6) equals the sampling distribution, viewed as a function of the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\delta}$  for the given data  $\mathbf{y}$ .

In order to specify the statistical model for the regression problem (2) the regression function  $f_{\boldsymbol{\theta}}(x)$  and the distribution of the error  $p(\boldsymbol{\epsilon}|\boldsymbol{\theta}, \boldsymbol{\delta})$  need to be stated. We note that the regression function  $f_{\boldsymbol{\theta}}(x)$  and the distribution of the errors  $\boldsymbol{\epsilon}$  ought to be considered as a pair in the sense that the distribution for the errors models (in a probabilistic sense) the deviations of the data from the chosen regression function.

For example, for the Normal straight line regression model (5) the sampling distribution is given by

$$p(\mathbf{y}|\theta_1, \theta_2, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta_1 - \theta_2 x_i)^2}, \quad (13)$$

which illustrates how the deterministic regression function together with the probabilistic model for the errors in the data jointly determine the statistical model. In this case the regression function is given by a straight line,  $f_{\boldsymbol{\theta}}(x) = \theta_1 + \theta_2 x$ , and the distribution of the errors is taken as a Normal distribution with zero mean and constant variance  $\sigma^2$ , i.e.,  $p(\boldsymbol{\epsilon}|\boldsymbol{\theta}, \boldsymbol{\delta}) = p(\boldsymbol{\epsilon}|\sigma^2) = N(\mathbf{0}, \sigma^2 \mathbf{I})$ .

The specification of the regression function may be based on a theoretical or a phenomenological relationship. The former requires understanding of the underpinning (physical) laws, while the latter calls for knowledge about the shape of the regression function. Specific aspects depend on the particular application.

Similarly, the specification of the distribution of the errors may be based on an understanding of the process of measuring, or is determined through phenomenological modelling. Aspects that should be considered in this context are, for example, whether observations display non-Gaussian behaviour of the errors, such as asymmetry (favouring negative or positive errors), or an increased possibility of observations in the tails of the distribution (so called outlying data). Further aspects are whether different observations are dependent (rather than independent), or whether observations exhibit heteroscedasticity, i.e., their variances depend on the response or regressor variables or even on unobserved variables (rather than being constant).

A close collaboration between statisticians and experimenters or experts of the field is recommended to investigate these aspects.

## Case studies

The case studies illustrate the following features of statistical modelling:

- The *straight line fit* described in Section 3.1 illustrates the specification of the statistical model for a general linear regression model and a Normal sampling distribution with unknown but constant variance.
- The *flow meter calibration* described in Section 4.1 illustrates a particular instance of a linear regression model in the form of a calibration curve and a Normal sampling distribution with unknown but constant variance.
- The *inference of thermophysical properties* described in Section 5.1 illustrates a non-linear regression model defined implicitly by a computational procedure and a Normal sampling distribution with unknown but constant variance.
- The *analysis of immunoassay data* described in Section 6.1 illustrates a non-linear regression model defined explicitly and a Normal sampling distribution with a variance that depends on the value of the independent variable.
- The *analysis of digital PCR data* described in Section 7.1 illustrates the specification of the statistical model for a parametric inverse problem.

## 2.2 Prior distributions

The prior distribution reflects one's state of knowledge before the data are taken into account. Usually, prior knowledge will be available and ought to be included into the analysis in the form of an informative prior. Prior knowledge can result, for instance, from the elicitation of expert knowledge [20, 44], or through the re-use of a posterior distribution obtained from the analysis of a previous experiment. Sometimes, however, prior knowledge may be vague or even missing, or one may want to disregard it. In such cases noninformative prior distributions [36] can be employed. In this section we introduce the specification of an informative or a noninformative prior, and we describe features of the choice of priors in the case studies.

### Informative priors

Informative priors are distributions that express the available prior knowledge. In the simplest case the prior knowledge may be provided by a Bayesian inference of a related previous experiment. In that case the posterior distribution from the previous experiment can be used as the prior for some or, ideally, all unknown parameters in the current analysis. In other cases the available prior knowledge may come from various sources of information, for example, physical constraints, historical data, or assessments of one or several experts in the field, which are then used to construct a prior distribution. The process of capturing available prior knowledge and its formal transformation into a prior distribution is known as elicitation, and statistical techniques have been developed for this purpose (e.g., [20, 44, 49]).

In order to derive, for example, a prior distribution for a single parameter one may identify intervals that include that parameter with specified probabilities (e.g., 50 %, 90 %, 95 %, etc.), and then fit a distribution taken from a family of distributions to that information. The choice of the family of distributions may depend on the expected range of the parameter, the assumed symmetry of the distribution, and whether heavy tails are likely. For example,

if a (possibly scaled) parameter varies between zero and one, the Beta distribution provides a flexible parametric family of distributions, or if a parameter is known to be positive, a Gamma distribution may be chosen. Inverse Gamma distributions, on the other hand, are often employed as priors for variance parameters. When the range of the parameter is the whole real line, and when the distribution can be assumed symmetric, the family of scaled and shifted  $t$ -distributions could be used.

Instead of selecting a prior from a single family of distributions one can also use hierarchical priors. For example, a Normal distribution may be deemed appropriate to express one's prior knowledge about an unknown parameter, but there remains some uncertainty about the mean and the variance of the Normal distribution. In that case a hierarchical model could be chosen in which the mean and the variance of the Normal distribution are themselves modelled as random variables, and where so-called hyperpriors are used to express the uncertainty about them.

It is important to clarify whether the unknown parameters may be treated as independent *a priori*, or whether the available knowledge implies dependencies between them. In the latter case a joint distribution needs to be selected.

Conjugate prior distributions are also an attractive choice. A prior is conjugate when both prior and posterior belong to the same family of distributions. The main advantage of conjugate priors is that the posterior is known analytically, and hence extensive numerical calculations can be avoided. For example, the Normal inverse Gamma prior is conjugate for the Normal linear regression problem, see Section 3.2. Whether the use of a conjugate prior is appropriate depends on whether the available prior knowledge can be expressed in terms of the corresponding family of distributions.

To fully specify an informative prior, typically an infinite number of its features (e.g., quantiles) need to be elicited [5]. Therefore, usually no single prior distribution is the 'right' one, but rather a number of different distributions can be considered as expressing the available prior knowledge. This reflects the uncertainty in the process of eliciting and quantifying one's prior knowledge, and it lends the Bayesian inference naturally to a sensitivity analysis in which the variability of the results arising from different choices of the prior distribution is investigated. If the results vary appreciably, either additional prior knowledge or a more rigorous elicitation process is required, or further data need to be gathered.

### Noninformative and vague priors

Bayesian inferences are sometimes carried out using so-called noninformative priors. Examples are applications in which the available prior information is very ambiguous or is to be ignored. Other examples include problems for which an elicitation process runs the risk of underestimating the *a priori* uncertainty or is too difficult, for instance in multivariate problems with many parameters where the assumption of *a priori* independence of the parameters is not adequate. One might also be interested in using a noninformative prior as a reference to determine the impact that available prior knowledge has on the results.

The choice of a noninformative prior is not straightforward as there is no 'universal' noninformative prior that applies in all cases. Many different principles have been proposed over the years for the selection of such a prior, and include Jeffreys' prior [35] that leads to results that are invariant under a reparametrisation, the so-called reference prior [6] that determines a noninformative prior in such a way that the gain in information for the resulting posterior is maximized in a certain sense, cf. also [10], and a probability matching prior [14] that provides

credible intervals which can (approximately) serve at the same time as confidence intervals. Furthermore, simple constant priors are often employed.

Instead of using one of the above principles to select a noninformative prior one may alternatively resort to a chosen family of priors (e.g., conjugate priors), and let the priors become more and more vague (e.g., by letting the variance of the prior tend to infinity). Controlling the ‘vagueness’ of the prior in this way then also yields information on the sensitivity of the results.

In contrast to informative prior distributions, noninformative priors are often improper (i.e. cannot be normalized), and propriety of the posterior is then not necessarily guaranteed. When using improper priors it is important to check the propriety of the resulting posterior, but this can prove challenging and samples drawn from the posterior may not give an indication about propriety. Note that when a particular improper prior (e.g., a constant prior) leads to an improper posterior, use of a corresponding proper vague prior (e.g., a rectangular prior over a large, but finite, interval) does not help. The reason is that while such a proper vague prior produces a proper posterior, this posterior will depend strongly on the choice of the vagueness of the prior (e.g., on the width of the interval chosen for a rectangular prior).

One result in asymptotic statistics is that under certain regularity conditions a Bayesian posterior becomes independent of the chosen prior as the number of data tends to infinity [57]. A sensitivity analysis may reveal whether the asymptotic case is reached in practice, and if so, the actual choice of a (noninformative) prior does not really matter. Unfortunately, reaching the asymptotic case can require quite a large amount of data (see, e.g., [38] for the case of inferring the mean of multivariate Gaussian observations).

The use of noninformative priors is not undisputed among Bayesians, see, for example, the two discussion papers [5, 29] and the various comments given therein. Nonetheless, they are often employed and can prove useful in practice.

## Case studies

The case studies illustrate the following features of the selection of a prior:

- The *straight line fit* described in Section 3.2 illustrates the use of conjugate priors for which the posterior is given in analytical form.
- The *flow meter calibration* described in Section 4.2 illustrates the selection of a prior distribution in a Normal linear regression problem when prior knowledge is given in the form of constraints on the values of the regression function.
- The *inference of thermophysical properties* described in Section 5.2 illustrates the use of hierarchical priors and how the hyperpriors can be chosen on the basis of available prior knowledge.
- The *analysis of immunoassay data* described in Section 6.2 illustrates how historical data can be used to determine an informative prior distribution, as well as the construction of a prior distribution for a subsequent prediction problem.
- The *analysis of digital PCR data* described in Section 7.2 illustrates the selection of different families of prior distributions to reflect knowledge about the likely interval of values for each parameter and other available prior knowledge.

## 2.3 Numerical methods

Once the statistical model for the data and the prior distribution have been specified, the posterior distribution is, in principle, determined. However, in most cases one needs to resort to numerical techniques to calculate it. One reason is that the posterior distribution<sup>2</sup>  $\pi(\boldsymbol{\theta}|\mathbf{y}) = C(\mathbf{y})\pi(\boldsymbol{\theta})l(\boldsymbol{\theta}; \mathbf{y})$  is known only up to the normalization constant

$$C(\mathbf{y}) = \left( \int \pi(\boldsymbol{\theta})l(\boldsymbol{\theta}; \mathbf{y})d\boldsymbol{\theta} \right)^{-1}, \quad (14)$$

and, although the prior and likelihood are usually given in explicit form, in most cases the integral (14) cannot be evaluated analytically. Furthermore, in order to determine marginal posterior distributions such as

$$\pi(\theta_1|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y}) d\theta_2 \cdots d\theta_p, \quad (15)$$

or expectations and covariances of these distributions, the evaluation of the necessary integrals again requires usually numerical methods.

Straightforward numerical quadrature applied to integrals such as (14) or (15) in several (and often many) dimensions is usually intractable (cf. [28]). Analytic approximations such as the Laplace method (see [56] and more recently [52]) are sometimes used, but quantifying their degree of approximation can prove difficult. In this *Guide* we concentrate on Monte Carlo procedures, i.e., stochastic methods, which are also state of the art in the evaluation of high-dimensional integrals [50]. Markov chain Monte Carlo (MCMC) methods represent a flexible class of stochastic methods that are well-established in Bayesian data analysis [21], and the numerical calculations presented in this *Guide* are mainly based on variants of these methods.

MCMC methods provide means to draw samples from a specified probability distribution. These methods generate a Markov chain whose stationary distribution equals the probability distribution of interest. As Monte Carlo methods in general, MCMC methods thus allow for (approximate) evaluation of possibly high-dimensional integrals by averaging over the generated sample (or functions thereof). In contrast to standard Monte Carlo methods, MCMC methods allow a sample to be obtained from a target distribution without drawing single independent realizations from that distribution, and this is particularly useful for the general applicability of this sampling method.

A generic MCMC algorithm is the Metropolis-Hastings algorithm, which generates a sequence  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots$  from a target distribution  $\pi(\boldsymbol{\theta})$  as follows. At each iteration  $t$  a new candidate sample, say  $\tilde{\boldsymbol{\theta}}$ , is drawn from a proposal distribution  $q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}_t)$ . This candidate is then accepted with probability

$$\min \left( 1, \frac{\pi(\tilde{\boldsymbol{\theta}})q(\boldsymbol{\theta}_t|\tilde{\boldsymbol{\theta}})}{\pi(\boldsymbol{\theta}_t)q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}_t)} \right) \quad (16)$$

as the sample  $\boldsymbol{\theta}_{t+1}$ . If  $\tilde{\boldsymbol{\theta}}$  is not accepted, then  $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t$  is used. The proposal distribution  $q(\cdot|\cdot)$  may be of quite general form, and the sequence produced in this way is under fairly general conditions a Markov chain whose stationary distribution equals the target distribution, denoted in (16) by  $\pi(\cdot)$ . One usually discards the initial samples during the so-called ‘burn in’ phase to ensure that the chain has approached its stationary state. However, an

---

<sup>2</sup>For ease of presentation in this subsection  $\boldsymbol{\theta}$  contains all unknown parameters. In subsequent subsections, dealing with the example applications, we distinguish again between the parameters of the regression function and additional parameters  $\boldsymbol{\delta}$ .

unsuitable choice of the proposal distribution may prevent that even after a long burn in phase. A suitable choice of  $q(\cdot|\cdot)$  should allow the full exploration of the parameter space (or the support of the target distribution) and, at the same time, enable a reasonable acceptance rate. We refer to, for example, [28] and [50] for more details on convergence and other theoretical considerations.

It is essential that in expression (16) the target distribution enters in both the numerator and the denominator. This implies that normalization constants (such as (14)) need not be known. Hence, in order to sample from the posterior one does not have to insert the normalized distribution  $\pi(\boldsymbol{\theta}|\mathbf{y})$  for the target distribution  $\pi(\cdot)$  in (16), but using the (non-normalized) product of prior and likelihood is sufficient. Hence, MCMC procedures can produce a sample from the posterior distribution without determining the normalization constant (14). We note, however, that for model selection or Bayesian model averaging, for example, the normalization constant (14) (also called evidence) needs to be determined.

A popular variant of the Metropolis-Hastings algorithm is the random walk Metropolis-Hastings algorithm which determines at iteration  $t$  a candidate  $\tilde{\boldsymbol{\theta}}$  according to

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_t + \boldsymbol{\epsilon}, \quad (17)$$

where  $\boldsymbol{\epsilon}$  follows a zero mean symmetric distribution that is independent of  $\boldsymbol{\theta}_t$ , for example a multivariate zero mean Normal distribution with suitably chosen covariance matrix. The candidate  $\tilde{\boldsymbol{\theta}}$  is then accepted with probability

$$\min\left(1, \frac{\pi(\tilde{\boldsymbol{\theta}})}{\pi(\boldsymbol{\theta}_t)}\right), \quad (18)$$

which immediately follows from (16) since the proposal distribution satisfies in this case  $q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}_t) = q(\boldsymbol{\theta}_t|\tilde{\boldsymbol{\theta}})$ . Note that in this way a sequence from any (reasonable) distribution can be produced from a sequence of, for example, Normal deviates. Good acceptance rates for random walk Metropolis-Hastings algorithms have been suggested to be close to 0.2 to 0.3 [51].

Another popular example of a Metropolis-Hastings algorithm is the Gibbs sampler, which for each component (or block of components)  $\theta_i$  of  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$  applies the proposal distribution  $\pi(\theta_i|\boldsymbol{\theta}_{-i})$ , i.e., the conditional distribution

$$\pi(\theta_i|\boldsymbol{\theta}_{-i}) = \frac{\pi(\theta_i, \boldsymbol{\theta}_{-i})}{\int \pi(\theta_i, \boldsymbol{\theta}_{-i}) d\theta_i} \quad (19)$$

of the target distribution  $\pi(\boldsymbol{\theta})$ , where  $\boldsymbol{\theta}_{-i} := (\theta_1 \dots \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)^\top$ . The candidate sample produced in this way differs from  $\boldsymbol{\theta}_t$  only in the  $i$ -th component (or block of components). One can show that the Gibbs sampler always accepts the generated candidate points. While the Gibbs sampler may be slow to explore the whole target distribution, especially for highly correlated or high-dimensional random variables, it has the advantage of utilizing a proposal distribution that does not reject any drawn candidate (thus avoiding useless simulations), it is widely applicable (e.g., even if the full conditional distribution does not have an analytically closed form), and the Gibbs sampler is implemented in the powerful and free software package BUGS (Bayesian inference Using Gibbs Sampling) [42].

When using a MCMC procedure results are of a stochastic nature and, strictly speaking, call for a (further) inference. When the sample produced is sufficiently large, and convergence to the stationary distribution of the chain has already been reached, the posterior, or results derived thereof, can be approximated accordingly well. For example, in order to approximate



the marginal posterior distribution for  $\theta_1$  say, one simply uses the first component of the produced Markov chain, and the posterior mean for  $\theta_1$  can be approximated by the average of the values of the first component of the chain.

The samples of a Markov chain are correlated, and one usually uses thinning, i.e. takes a subsample (such as every 10<sup>th</sup> or 100<sup>th</sup> sample) from the chain which reduces this correlation. To better assess convergence and stability of the results, one typically considers several independent chains. The dependence on initial values (which should be sufficiently dispersed) is reduced by discarding the first samples of a chain (in the burn in phase). Depending on the context, a typical burn in might be in the range of 1 % to 50 % of the length of the whole chain (see, e.g., [27] and [21]). Monitoring convergence criteria allows a decision on whether (or when) the size of a sample is sufficiently large. In special cases, e.g. when the posterior is known to be well-behaved and unimodal, it can be sufficient to use a single long Markov chain and to start from a typical value, even without discarding initial values of the chain.

Although assured theoretically under mild assumptions, ‘convergence’ of Markov chain simulations to the target distribution is not guaranteed for any finite sample. Monitoring and diagnosing convergence is thus necessary to decide whether the simulations provide sufficiently correct results. However, convergence assessment strategies are mainly empirical methods suffering from the defect that “you’ve only seen where you’ve been” [50, p. 464]. We therefore recommend a critical assessment of the results and, whenever possible, to exploit in addition alternative procedures.

One aspect of MCMC convergence is the behaviour of averages, which allows an assessment of whether a chain has explored all the complexity of the target distribution (e.g., all modes). For an overview of the diverse measures available to monitor the convergence of averages, see [50]. One criterion, for example, is based on potential scale reduction factors as introduced in [24] and generalized in [12]. This approach compares some measure of the posterior scale (i.e., an estimate using all samples) with an underestimate of the true scale measure – usually the average over the scale estimate for each chain separately. If an MCMC simulation has converged, the posterior scale estimate as well as the within chain scale estimate should stabilize as a function of the number of iterations, and the ratio of both, the so called potential scale reduction factor, should approach one [12]. While these three quantities give information on whether convergence is possibly reached when calculated after all iterations, observing them also graphically after blocks of, say, 50 iterations provides information on how the simulation is progressing. One example of such a scale measure (and the derived potential scale reduction factor) is the variance between and within chains as originally proposed in [24] for approximately Gaussian posterior distributions. More generally, these scale measures can be adapted to the particular inference at hand. One may look at the length of the empirical 80 % interval of all samples and of each chain, as well as their ratio to diagnose convergence. These convergence statistics are also implemented in BUGS [42].

Non-convergence of Markov chains is of great concern for MCMC techniques, because estimates resulting from an inference may be inaccurate or simply wrong. After diagnosing non-convergence, various strategies exist to improve the MCMC algorithm. The simplest is to increase the number of iterations. If this solution is prohibitive, for example, when many data sets or many models are fitted, the MCMC algorithm itself needs to be improved.

When convergence problems are due to correlation, a reparametrisation might reduce this correlation and improve the MCMC algorithm. However, for non-linear models no universal rules for ‘good’ reparametrisations exist [28, pp. 96f]. Potentially, a signed root transformation may improve convergence when heavy tailed distributions are involved [28, p. 98].

Alternatively, convergence can be improved by implementing a different MCMC method altogether. Also the target distribution itself can be modified to improve mixing, cf. [28, Section 6.4]. Examples are importance sampling or auxiliary variables [21, Chapter 13].

Improper posterior distributions can arise when improper prior distributions are employed. However, whether a posterior is proper may not always be visible from the corresponding MCMC chains, cf. [30] for a discussion of such problems observed for the Gibbs sampler.

Cases with analytical solutions, as in the example of a Normal linear regression model when using conjugate priors, are highly welcome. And they can be used to check or validate numerical procedures. Nonetheless, MCMC methods have the great advantage of flexibility in that prior distributions or likelihood functions can easily be exchanged.

## Case studies

The case studies illustrate the following features of numerical methods:

- The *straight line fit* described in Section 3.3 illustrates a problem for which the results are given analytically.
- The *flow meter calibration* described in Section 4.3 illustrates a simple Monte Carlo rejection algorithm to treat prior knowledge in the form of constraints on the values of the regression function in a Normal linear regression model.
- The *inference of thermophysical properties* described in Section 5.3 illustrates the use of a Metropolis-Hastings algorithm.
- The *analysis of immunoassay data* described in Section 6.3 illustrates the use of Gibbs sampling and its implementation in BUGS.
- The *analysis of digital PCR data* described in Section 7.3 illustrates the use of a Metropolis-Hastings algorithm together with a variable transformation to enhance convergence.

## 2.4 Sensitivity analysis

A Bayesian inference requires the specification of a statistical model and a prior distribution. The choice of each may be associated with some uncertainty and therefore calls for sensitivity analyses (cf. [21, Section 6.6]). The goal is to investigate whether obtained results are sensitive within reasonable variations of the selected prior, of the statistical model, or both. If the results appear to be sensitive, either further data or additional prior knowledge ought to be gathered.

A chosen prior distribution is neither ‘right’ or ‘wrong’. It should simply summarize the prior knowledge of the analyst. However, the elicitation of prior knowledge and its formalization in terms of a proper, informative prior can be challenging, and often includes some degree of approximation (cf. [49, Section 3]). For example, when choosing an inverse Gamma prior for a variance parameter  $\sigma^2$ , one selects the two parameters of that distribution to reflect one’s prior assessment about  $\sigma^2$ , either formally, e.g. by fitting the distribution to specified quantiles, or informally. The actual choice of an informative prior may in any case be seen as an approximation of the available prior knowledge. And this holds also for the choice of a noninformative prior as there are several ways to construct such a prior (cf. Section 2.2).

It is good practice to include into a sensitivity analysis also a variation of the statistical model since that is often chosen on pragmatic grounds. Reasonable alternative models ought to be considered, which may be based on competing physical explanations or on more robust distributional assumptions, for example, sampling distributions with heavier tails.

## Case studies

The case studies illustrate the following features of sensitivity analysis:

- The *straight line fit* described in Section 3.4 illustrates a sensitivity analysis with respect to the variation of the parameters in a chosen family of priors.
- The *flow meter calibration* described in Section 4.4 illustrates a sensitivity analysis with respect to the variation of a prior expressed in the form of constraints on the regression function.
- The *inference of thermophysical properties* described in Section 5.4 illustrates a sensitivity analysis with respect to variations of a hierarchical prior.
- The *analysis of immunoassay data* described in Section 6.4 illustrates a sensitivity analysis with respect to the variation of the form of the prior distributions and of the statistical model.
- The *analysis of digital PCR data* described in Section 7.4 illustrates a sensitivity analysis with respect to the variation of the prior for each unknown in turn

## 2.5 Further aspects

There are further aspects that can be relevant to a Bayesian treatment of regression problems, such as Bayesian model selection [7, 25] or Bayesian model averaging [31, 33, 41]. For example, in a polynomial model the polynomial's degree might not be known exactly *a priori* and ought to be inferred in the light of the data, or perhaps the results for several models should be merged. Bayesian inference is well suited to address such issues, yet in this *Guide* we assume that a single, well-characterized model is available. Another aspect is the use of transformations of the data [11, Chapter 10]. One aim of applying such data transformations is to reach a desired distribution for the errors, for example, a Normal distribution. Another aim is to stabilize the numerical calculations, an aspect on which we briefly comment in the descriptions of the numerical methods of Sections 3.3, 6.3 and 7.3.

### 3 Straight line fit – Normal linear regression using conjugate priors

This case study presents the task of fitting a straight line to data that are Normally distributed with unknown, constant variance. A conjugate prior distribution is used which leads to an analytic expression for the posterior distribution. While in this particular example a straight line is considered, the formalism carries over to general linear regression models and our presentation captures that.

#### 3.1 Statistical modelling

Figure 1 shows example data that are modelled by the Normal straight line regression model (5). In dealing with this model, however, we use a more general (matrix) notation than that used in (5), namely

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (20)$$

where  $\mathbf{I}$  denotes the identity matrix of dimension  $n$ , and  $\mathbf{X}$  the  $n \times 2$  design matrix given by

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}. \quad (21)$$

The reason for using this notation is that the subsequent treatment, including the selection of conjugate prior distributions and guidance on the choice of their parameters as well as the analytic calculation of the posterior distribution, carries over to the Normal linear regression model with general design matrix  $\mathbf{X}$  of dimension  $n \times p$  where  $n \geq p$ . The  $p$  columns of  $\mathbf{X}$  may be determined by any chosen set of basis functions, for example, defining a polynomial of general degree or a Fourier series expansion. The design matrix, however, should have rank  $p$ .

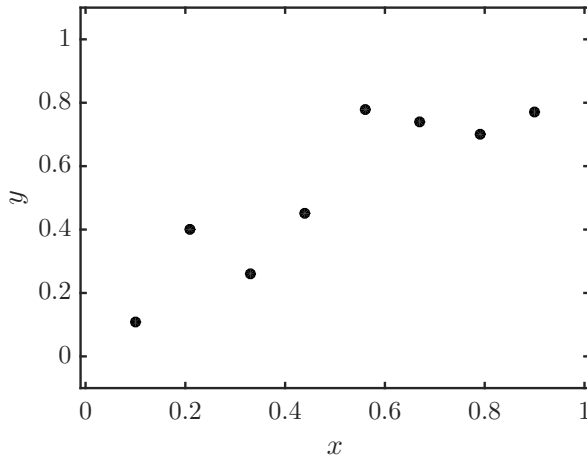


Figure 1: Example data for the Normal linear regression problem. The statistical model is given in (5) (and (20)).

Two properties of the statistical model (20) are important for its subsequent treatment: the linearity of the regression function with respect to the parameters  $\boldsymbol{\theta}$ , and the fact that the observations are independent and Normally distributed with constant variance. The likelihood function is given by

$$l(\boldsymbol{\theta}, \sigma^2; \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})}. \quad (22)$$

### 3.2 Prior distributions

We illustrate the use of conjugate priors for which the posterior is given in analytical form, and demonstrate how the parameters of the conjugate prior can be selected to reflect the prior knowledge about the variance of the observations and about the linear regression function.

**Choice of family of priors.** The Normal inverse Gamma (NIG) prior for  $(\boldsymbol{\theta}, \sigma^2)$  given by

$$\boldsymbol{\theta}|\sigma^2 \sim \mathcal{N}(\boldsymbol{\theta}_0, \sigma^2 \mathbf{V}_0) \quad \text{and} \quad \sigma^2 \sim \text{IG}(\alpha_0, \beta_0), \quad (23)$$

is conjugate for the Normal linear regression model (20), cf., for example, [45]. The conditional prior for  $\boldsymbol{\theta}$  is a multivariate Normal distribution with mean  $\boldsymbol{\theta}_0$  and covariance matrix  $\sigma^2 \mathbf{V}_0$ , and the prior for  $\sigma^2$  is an inverse Gamma distribution with shape  $\alpha_0$  and scale  $\beta_0$ . Subsequently, we abbreviate the prior (23) for  $(\boldsymbol{\theta}, \sigma^2)$  by  $\text{NIG}(\boldsymbol{\theta}_0, \mathbf{V}_0, \alpha_0, \beta_0)$ .

**Informative prior for  $\sigma^2$ .** We start by selecting the parameters  $\alpha_0$  and  $\beta_0$  of the inverse Gamma prior distribution for  $\sigma^2$  such that the resulting distribution reflects our expectations about the variance. One could formalize such a process by, for example, specifying a most probable value of  $\sigma^2$ , together with one or several percentiles, and then choose  $\alpha_0$  and  $\beta_0$  such that this information is best reflected in the resulting inverse Gamma distribution. Less formally, one can plot the distribution for  $\sigma^2$ , or for  $\sigma$ , and select  $\alpha_0$  and  $\beta_0$  such that the resulting distribution reflects one's *a priori* assessment. Figure 2 illustrates various distributions for  $\sigma$  for different choices of  $\alpha_0$  and  $\beta_0$ . We assume that the distribution obtained for  $\alpha_0 = 0.4$  and  $\beta_0 = 0.004$  roughly models our prior knowledge about  $\sigma^2$  and we use it as the prior for the variance (see Figure 3).

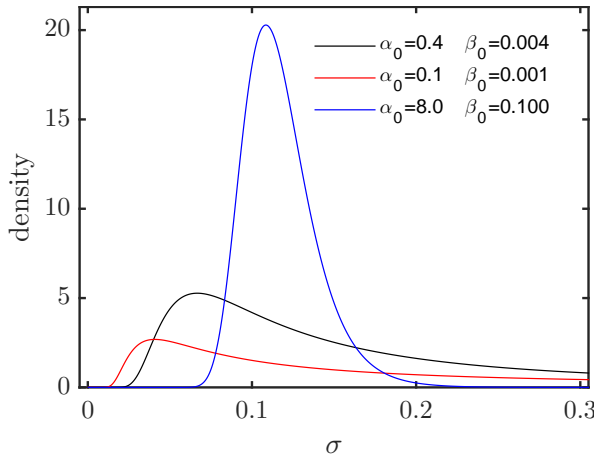


Figure 2: Visualization of the prior distribution  $\pi(\sigma) = 2\sigma\pi(\sigma^2)$  in the Normal linear regression problem (5) for different choices of  $\alpha_0$  and  $\beta_0$ . The prior for  $\sigma^2$  is given in (23).

**Informative prior for  $\boldsymbol{\theta}$ .** In the next step we specify the parameters in the Gaussian prior for  $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$ . We assume, for given  $\sigma^2$ , that  $\theta_1$  and  $\theta_2$  can be treated as independent *a priori*, so that  $\mathbf{V}_0 = \text{diag}(v_{11}, v_{22})$ . According to our prior knowledge we expect  $\theta_1$  to be about zero with an associated standard uncertainty of 0.2. Since  $\sigma$  is mainly concentrated at 0.1 (cf. Figure 3) we chose  $v_{11} = 0.2^2/0.1^2$ . For  $\theta_2$  we assume a value around one with an associated standard uncertainty of 0.2, and hence we chose  $v_{22} = 0.2^2/0.1^2$ .

Table 1 summarizes the values of the parameters selected to define the informative prior distribution for this case study. In order to illustrate, or check, our choice of prior for the parameters of the straight line we draw randomly a large sample of values  $(\theta_1, \theta_2)$  from the distribution (23). For each pair  $(\theta_1, \theta_2)$  the resulting straight line can be drawn, and Figure 3 shows a band which contains (at each point) 95 % of these curves<sup>3</sup>. Figure 3 thus illustrates

<sup>3</sup>These curves can also be obtained analytically, using the fact that *a priori*  $\theta_1 + \theta_2 x$  follows a suitably scaled and shifted *t*-distribution.

the elicited prior knowledge about the unknown regression curve, and possible adjustments of the parameters of the priors could be made in view of this figure.

$\alpha_0$	$\beta_0$	$(\boldsymbol{\theta}_0)_1$	$(\boldsymbol{\theta}_0)_2$	$(\mathbf{V}_0)_{11}$	$(\mathbf{V}_0)_{22}$	$(\mathbf{V}_0)_{12}$
$4 \times 10^{-1}$	$4 \times 10^{-3}$	0	1	4	4	0

Table 1: Parameters of the NIG prior distribution (23) for the Normal linear regression problem (5).

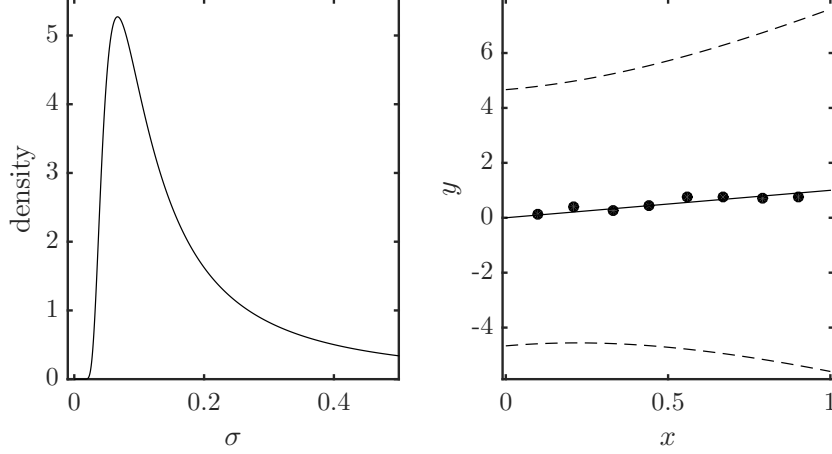


Figure 3: Visualization of the NIG prior distribution (23) (with the parameters specified in Table 1) for the Normal linear regression problem (5). Left: Prior distribution  $\pi(\sigma) = 2\sigma\pi(\sigma^2)$  for the standard deviation of observations. Right: Mean regression curve (solid line) and pointwise 95 % intervals (dashed lines). The dotted points show the data.

**Noninformative prior.** We note that the standard noninformative prior usually applied for the Normal linear regression model is given by

$$\pi(\boldsymbol{\theta}, \sigma^2) \propto 1/\sigma^2, \quad (24)$$

cf., for example, [21], which can formally be obtained from the conjugate prior (23) with  $\alpha_0 = -p/2$ ,  $\beta_0 = 0$  and  $\mathbf{V}_0^{-1} \rightarrow \mathbf{0}$  [45]. (Note that for  $\alpha_0 < 0$  the NIG prior is no longer a proper distribution.) The resulting marginal posterior  $\pi(\boldsymbol{\theta}|\mathbf{y})$  is a scaled and shifted multivariate  $t$ -distribution with  $\nu = n - p$  degrees of freedom, and mean and covariance matrix given by

$$\mathbb{E}(\boldsymbol{\theta}|\mathbf{y}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}, \quad (25)$$

$$\text{Cov}(\boldsymbol{\theta}|\mathbf{y}) = \frac{\nu}{\nu - 2} s^2 (\mathbf{X}^\top \mathbf{X})^{-1}, \quad (26)$$

where  $s^2 = \min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) / (n - p)$ .

The posterior obtained when using the noninformative prior (24) may be compared with the posterior obtained for the informative prior (cf. Section 3.3), for example, to assess the amount of information that is contained in the informative prior. And it may even be used as the final result in cases where the prior knowledge is very vague or shall be ignored.

### 3.3 Numerical methods

Since conjugate priors are used in this example, the posterior can be calculated analytically. Specifically, for the NIG prior (23) the joint posterior is also a NIG distribution,

$$\boldsymbol{\theta}, \sigma^2 | \mathbf{y} \sim \text{NIG}(\boldsymbol{\theta}_1, \mathbf{V}_1, \alpha_1, \beta_1), \quad (27)$$

where

$$\boldsymbol{\theta}_1 = \left( \mathbf{V}_0^{-1} + \mathbf{X}^\top \mathbf{X} \right)^{-1} \left( \mathbf{V}_0^{-1} \boldsymbol{\theta}_0 + \mathbf{X}^\top \mathbf{y} \right), \quad (28)$$

$$\mathbf{V}_1 = \left( \mathbf{V}_0^{-1} + \mathbf{X}^\top \mathbf{X} \right)^{-1}, \quad (29)$$

$$\alpha_1 = \alpha_0 + \frac{1}{2}n, \quad (30)$$

$$\beta_1 = \beta_0 + \frac{1}{2} \left( \boldsymbol{\theta}_0^\top \mathbf{V}_0^{-1} \boldsymbol{\theta}_0 + \mathbf{y}^\top \mathbf{y} - \boldsymbol{\theta}_1^\top \mathbf{V}_1^{-1} \boldsymbol{\theta}_1 \right), \quad (31)$$

and  $n$  denotes the number of data. The marginal posterior  $\pi(\boldsymbol{\theta}|\mathbf{y})$  is a multivariate (scaled and shifted) t-distribution,  $t_\nu(\boldsymbol{\theta}_1, \mathbf{V}_2)$ , i.e.,

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \left( 1 + \frac{(\boldsymbol{\theta} - \boldsymbol{\theta}_1)^\top \mathbf{V}_2^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_1)}{\nu} \right)^{-(\nu+p)/2}, \quad (32)$$

with  $\nu = 2\alpha_1$  and  $\mathbf{V}_2 = \beta_1/\alpha_1 \mathbf{V}_1$ . The mean and covariance of  $\boldsymbol{\theta}|\mathbf{y}$  are given by

$$\mathbb{E}(\boldsymbol{\theta}|\mathbf{y}) = \boldsymbol{\theta}_1 \quad (\text{for } \nu > 1), \quad (33)$$

$$\text{Cov}(\boldsymbol{\theta}|\mathbf{y}) = \frac{\nu}{\nu - 2} \mathbf{V}_2 \quad (\text{for } \nu > 2). \quad (34)$$

In order to ensure that numerical calculations are stable, the columns of the matrix  $\mathbf{X}$  ought to be of similar norm, which can usually be achieved by an appropriate scaling (and centering) of the regressor variable(s). Note that scaling (and centring) of the regressor variable(s) can sometimes also be useful in the interpretation or specification of a prior distribution for  $\boldsymbol{\theta}$ .

In Appendix A.1 we provide MATLAB<sup>®</sup> source to undertake the necessary calculations to provide graphical and numerical results. Application of the software to the data from Figure 1 yields the results shown in Figures 4 and 5, and Table 2. The comparison of the results obtained for the informative prior and the noninformative prior shows that the former is indeed informative and leads to an improved inference, i.e., one arriving at smaller uncertainties. The plot of the residuals does not indicate an inconsistency of data and derived model.

$\alpha_1$	$\beta_1$	$(\boldsymbol{\theta}_1)_1$	$(\boldsymbol{\theta}_1)_2$	$(\mathbf{V}_1)_{11}$	$(\mathbf{V}_1)_{22}$	$(\mathbf{V}_1)_{12}$
4.400	0.056	0.080	0.887	0.393	1.158	-0.561

$\hat{\theta}_1$	$\hat{\theta}_2$	$I_{0.95}(\theta_1)$	$I_{0.95}(\theta_2)$	$\hat{\sigma}^2$	$I_{0.95}(\sigma^2)$
0.080	0.887	[-0.080, 0.240]	[0.613, 1.161]	0.017	[0.006, 0.043]

Table 2: Results for the Normal linear regression problem obtained by fitting model (5) and the NIG prior (23) (with the parameters specified in Table 1) to the data given in Figure 1. Upper part: Parameters of the NIG posterior (27). Lower part: Posterior means and standard deviations as well as 95 % (probabilistically symmetric) credible intervals for the regression and additional parameters.

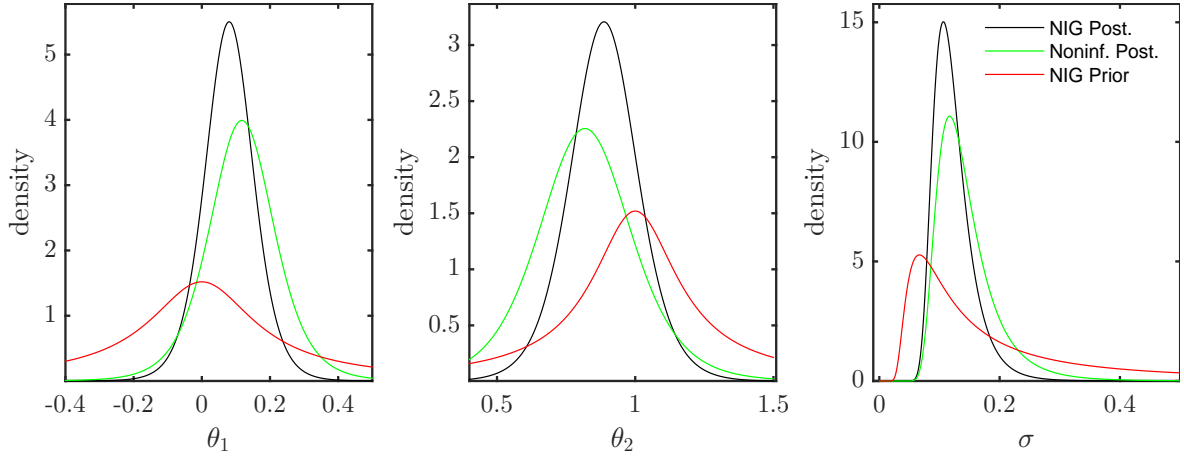


Figure 4: Posterior distributions of  $\theta_1$  (left),  $\theta_2$  (middle) and  $\sigma$  (right) for the Normal linear regression problem, obtained by fitting model (5) and the NIG prior (23) (with the parameters specified in Table 1) to the data given in Figure 1 (black lines). The green lines show the according results for the noninformative prior (24). The red lines illustrate the NIG prior.

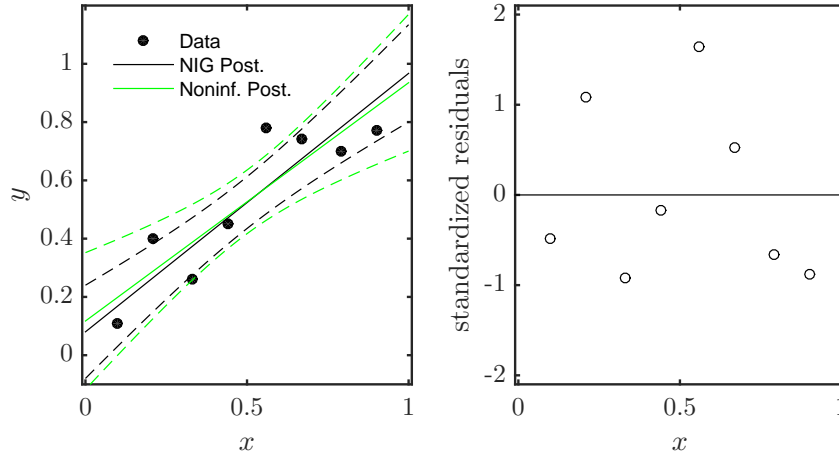


Figure 5: Results for the Normal linear regression problem obtained by fitting model (5) and the NIG prior (23) (with the parameters specified in Table 1) to the data given in Figure 1. Left: Estimated regression function  $y = \hat{\theta}_1 + \hat{\theta}_2 x$  and pointwise 95 % credible intervals (black lines). The green lines show the corresponding results for the noninformative prior (24). Right: Standardized residuals  $(y_i - (\hat{\theta}_1 + \hat{\theta}_2 x_i)) / \hat{\sigma}$ , where  $\hat{\theta}_1, \hat{\theta}_2$  and  $\hat{\sigma}$  denote posterior means of  $\theta_1, \theta_2$  and  $\sigma$ .

### 3.4 Sensitivity analysis

In this example we illustrate a sensitivity analysis in which the parameters of the employed family of conjugate prior distributions are varied, while the form of the prior and also the statistical model are kept fixed. The different parameter settings for the NIG prior studied during the sensitivity analysis are listed in Table 3. Note that Prior A is identical to the prior used in Section 3.3. In addition to these NIG priors the posterior results corresponding to the noninformative prior (24) described in Section 3.2 are also considered in the sensitivity analysis.

Figure 2 in Section 3.2 shows the variations in the prior for  $\pi(\sigma) = 2\sigma\pi(\sigma^2)$  when using the different settings for  $\alpha_0$  and  $\beta_0$  listed in Table 3. By using the priors from Table 3 to analyse the data from Figure 1 we obtain the results shown in Figures 6 and 7, and Table 4. In Figure 6 the posterior distributions for  $\sigma$  are depicted where the posteriors A to C correspond



	$\alpha_0$	$\beta_0$	$\theta_1$	$\theta_2$	$(\mathbf{V})_{11}$	$(\mathbf{V})_{22}$	$(\mathbf{V})_{12}$
Prior A	0.400	0.004	0.000	1.000	4.000	4.000	0.000
Prior B	0.100	0.001	0.000	1.000	2.000	2.000	0.000
Prior C	8.000	0.100	0.100	1.100	10.000	10.000	0.000

Table 3: Settings of the sensitivity analysis for the Normal linear regression problem under model (5). Given are the parameters for the NIG prior.

to the priors A to C in Table 3. As can be seen, the different priors result in different posterior distributions for  $\sigma$  which, however, exhibit a large overlap. The noninformative prior leads to the broadest of these posterior distributions.

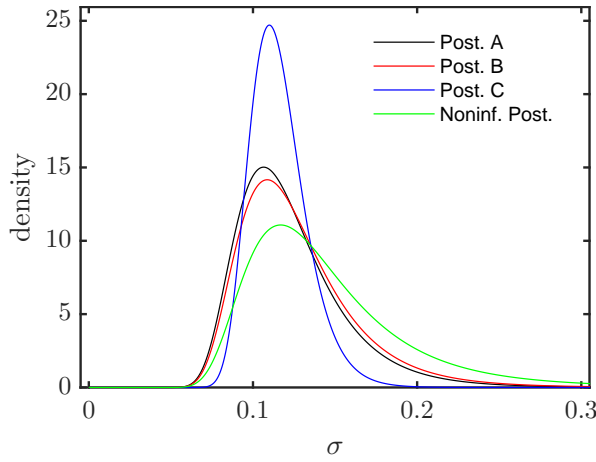


Figure 6: Results of the sensitivity analysis for the Normal linear regression problem. Posterior distributions for the standard deviation  $\sigma$  of observations obtained by fitting model (5) and the NIG prior (23) (with the parameters specified in Table 3) or the noninformative prior (24) to the data given in Figure 1.

In Figure 7 the different posterior results are compared in terms of the resulting regression functions as well as in terms of the resulting pointwise 95 % credible intervals. The variations in these results are minor. We note that this does not imply that the prior information has no impact; when using the noninformative prior (24) instead, different results, and in particular larger uncertainties, are obtained (see also Table 4).

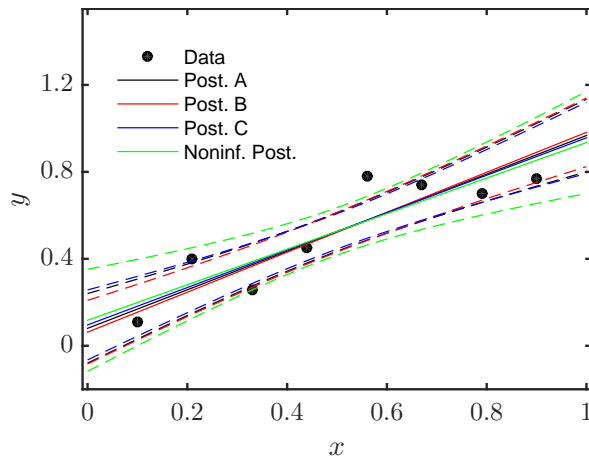


Figure 7: Results of the sensitivity analysis for the Normal linear regression problem. Estimated regression functions and 95 % (pointwise) credible intervals obtained by fitting model (5) and the NIG prior (23) (with the parameters specified in Table 3) or the noninformative prior (24) to the data shown by dotted points.

	$\hat{\theta}_1$	$I_{0.95}(\theta_1)$	$\hat{\theta}_2$	$I_{0.95}(\theta_2)$
Prior A	0.080	$[-0.080, 0.240]$	0.887	$[0.613, 1.161]$
Prior B	0.063	$[-0.084, 0.209]$	0.919	$[0.675, 1.163]$
Prior C	0.096	$[-0.065, 0.257]$	0.861	$[0.579, 1.142]$
Noninf. Prior	0.117	$[-0.117, 0.352]$	0.818	$[0.402, 1.233]$

Table 4: Results of the sensitivity analysis for the Normal linear regression problem. Posterior means and 95 % (probabilistically symmetric) credible intervals for the regression parameters obtained by fitting model (5) and the NIG prior (23) (with the parameters specified in Table 3) or the noninformative prior (24) to the data given in Figure 1.

## 4 Flow meter calibration – Normal linear regression with constraints

This case study concerns the determination of a calibration curve for a flow meter. The example is one of a Normal linear regression model (cf. Section 3), but the available prior knowledge includes a constraint on the values of the calibration curve. A numerical method is required to obtain the results of a Bayesian inference, which for the particular problem considered here takes the form of a simple Monte Carlo procedure combined with an accept/reject algorithm. A companion paper [40] presents a more detailed analysis of the calibration problem, including a comparison of the results from a Bayesian analysis with those provided by an ordinary least-squares analysis, which is conventionally applied but does not allow consideration of prior knowledge.

### 4.1 Statistical modelling

A turbine flow meter is a measuring device that indicates the volume of fluid flowing through the device per unit of time. A high-quality meter usually has an electrical pulse output such that each pulse corresponds to a fixed volume of fluid passing through the meter. It follows that the frequency of the pulse output is proportional to the flow rate  $q$ , and the proportionality factor is called the K-factor  $k$ . The manufacturer of the flow meter usually specifies a constant K-factor for the entire measurement range  $[q_{\min}, q_{\max}]$ . In a calibration of the meter, estimates of the K-factor are determined at a number of known flow rates, and a calibration curve is fitted to the calibration data. The calibration curve can then be used to provide an estimate of the K-factor at any flow rate within the measurement range [2, 34, 40].

Figure 8 shows calibration data for a master turbine flow meter from VSL's calibration facility for water flow meters, comprising  $n = 55$  measured values  $k_i$  of the K-factor (reported in  $\text{L}^{-1}$ ) corresponding to known values  $q_i$  of flow rate (reported in  $\text{L min}^{-1}$ ) provided by a compact prover reference device. Also indicated is the constant K-factor  $k_{\text{spec}} = 13.163 \text{ L}^{-1}$  specified by the manufacturer of the meter.

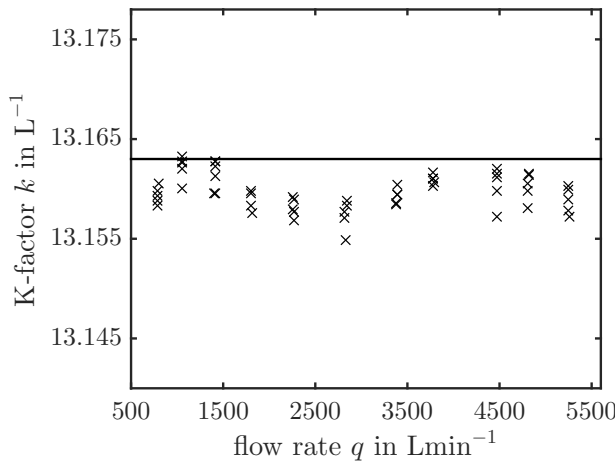


Figure 8: Example data for the calibration of a flow meter (crosses) and the constant K-factor specified by the manufacturer (horizontal line).

The statistical model for the calibration problem is a particular instance of the Normal linear regression model (cf. Section 3):

$$k_i = f_{\boldsymbol{\theta}}(q_i) + \epsilon_i, \quad f_{\boldsymbol{\theta}}(q) = \boldsymbol{\psi}^{\top}(q)\boldsymbol{\theta}, \quad \epsilon_i \stackrel{\text{iid}}{\sim} \text{N}(0, \sigma^2), \quad i = 1, \dots, n, \quad (35)$$

where

$$\boldsymbol{\psi}^{\top}(q) = (q^{r_1}, \dots, q^{r_p}). \quad (36)$$

The exponents  $r_1, \dots, r_p$ , with  $p < n$ , are specified in advance. In this example,  $p = 5$  and the exponents take the values 0, -1, 1, 2 and 3, so that

$$\boldsymbol{\psi}^\top(q) = \left(1, \frac{1}{q}, q, q^2, q^3\right), \quad (37)$$

and the calibration curve takes the particular form

$$f_{\boldsymbol{\theta}}(q) = \theta_1 + \frac{\theta_2}{q} + \theta_3 q + \theta_4 q^2 + \theta_5 q^3. \quad (38)$$

The likelihood function (cf. expression (22)) is given by

$$l(\boldsymbol{\theta}, \sigma^2; k_1, \dots, k_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (k_i - \boldsymbol{\psi}^\top(q_i)\boldsymbol{\theta})^2}. \quad (39)$$

Given observations  $\mathbf{k} = (k_1, \dots, k_n)^\top$  of the K-factor corresponding to known flow rates  $\mathbf{q} = (q_1, \dots, q_n)^\top$ , the goal is to estimate the parameters  $\boldsymbol{\theta}$  of the calibration curve  $f_{\boldsymbol{\theta}}(q)$  together with the standard deviation  $\sigma$  of the observations.

## 4.2 Prior distributions

Prior knowledge about  $\sigma^2$  is expressed as follows: the repeatability of the flow meter is specified by the manufacturer as a standard deviation  $\sigma_{0,\text{rel}}$  relative to  $k_{\text{spec}}$  with  $\sigma_{0,\text{rel}} = 0.025\%$ . The prior knowledge gives information about the repeatability standard deviation  $\sigma$ .

Prior knowledge about  $\boldsymbol{\theta}$  is expressed as follows: the deviation of the calibration curve from the curve obtained at the previous calibration is less than  $\delta$  of  $k_{\text{spec}}$  with maximum relative deviation  $\delta = 0.075\%$ . The prior knowledge gives information about the reproducibility of the calibration procedure and is assumed to be valid when performing two calibrations shortly after one another. Whereas prior knowledge in a Bayesian analysis is usually expressed explicitly in terms of the unknown parameters  $\boldsymbol{\theta}$ , the prior knowledge here is formulated in terms of a constraint on the values of the curve  $f_{\boldsymbol{\theta}}(q)$ . The formulation of the prior knowledge in this way complicates the Bayesian analysis to some extent, because prior knowledge about  $\boldsymbol{\theta}$  is expressed only implicitly by the constraint and to check whether a value of  $\boldsymbol{\theta}$  complies with the prior knowledge it is necessary to consider properties of the whole calibration curve.

**Choice of family of priors.** The parameters  $\boldsymbol{\theta}$  and  $\sigma^2$  are modelled as being independent *a priori* because the prior knowledge about  $\boldsymbol{\theta}$  relates to the calibration procedure and that about  $\sigma^2$  to the flow meter and information provided by the manufacturer. Consequently,

$$\pi(\boldsymbol{\theta}, \sigma^2) = \pi(\boldsymbol{\theta})\pi(\sigma^2). \quad (40)$$

The following family of prior distributions is considered for  $\sigma^2$ :

$$\sigma^2 \sim \text{IG}(\alpha_0, \beta_0) \quad \text{with} \quad \alpha_0 = \nu_0/2, \quad \beta_0 = \nu_0\sigma_0^2/2. \quad (41)$$

The following family of prior distributions is considered for  $\boldsymbol{\theta}$ :

$$\pi(\boldsymbol{\theta}) \propto \begin{cases} 1, & \text{if } \boldsymbol{\theta} \in \Omega_{\boldsymbol{\theta}_0, \gamma}, \\ 0, & \text{otherwise,} \end{cases} \quad (42)$$

where  $\Omega_{\boldsymbol{\theta}_0, \delta}$  is the set

$$\Omega_{\boldsymbol{\theta}_0, \delta}(\boldsymbol{\theta}) = \{\boldsymbol{\theta} \mid |f_{\boldsymbol{\theta}}(q) - f_{\boldsymbol{\theta}_0}(q)| \leq \delta \times k_{\text{spec}} \text{ for } q_{\min} \leq q \leq q_{\max}\}. \quad (43)$$

In other words, the prior  $\pi(\boldsymbol{\theta})$  is constant for all those  $\boldsymbol{\theta}$  that lead to a calibration curve whose deviation from a prescribed (calibration) curve is bounded in absolute value over the measurement range by  $\delta \times k_{\text{spec}}$ . The parameters used to specify the prior distribution for  $\boldsymbol{\theta}$  are  $\boldsymbol{\theta}_0$  and  $\delta$ . In practice, the prior knowledge is implemented by calculating discrete values of the calibration curve for a given  $\boldsymbol{\theta}$  for a large number of uniformly-spaced flow rates chosen to span the measurement range.

**Prior for  $\sigma^2$ .** Based on the prior knowledge about the flow meter specified by the manufacturer,  $\sigma_0 = 0.0033 \text{ L}^{-1}$  was chosen. Furthermore, we set  $\nu_0 = 1$  and  $\nu_0 = n$  to represent different degrees of belief in the information about the flow meter repeatability provided by the manufacturer. The values of  $\alpha_0$  and  $\beta_0$  determined by the prior knowledge are given in Table 5, and the corresponding prior distributions for  $\sigma$  are shown in Figure 9. Note that in Figure 9 (and elsewhere) probability distributions are calculated for  $\sigma$  in  $\text{L}^{-1}$  but displayed with  $\sigma$  expressed as a fraction of  $k_{\text{spec}}$  in %.

$\sigma_{0,\text{rel}}$ in %	$\nu_0$	$\alpha_0$	$\beta_0$ in $\text{L}^{-2}$
0.025	1	0.5	$5.4 \cdot 10^{-6}$
0.025	$n$	27.5	$3.0 \cdot 10^{-4}$

Table 5: Parameters of the prior distribution (41) for  $\sigma^2$  chosen to represent different degrees of belief in the information about the flow meter repeatability standard deviation provided by the manufacturer.

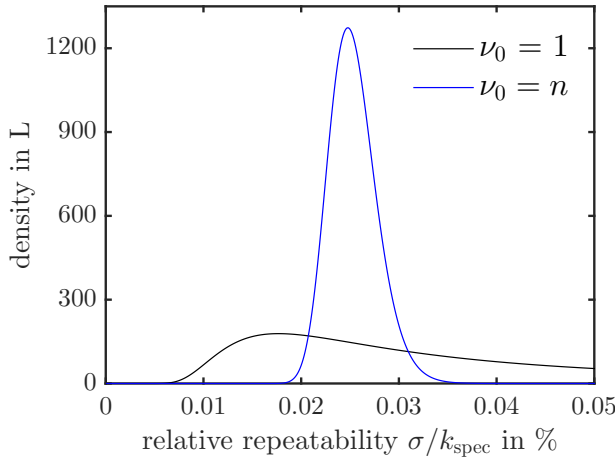


Figure 9: Prior distributions  $\pi(\sigma) = 2\sigma\pi(\sigma^2)$  for the standard deviation of observations defined by the parameters given in Table 5.

**Prior for  $\boldsymbol{\theta}$ .** For the purpose of this case study, the prior distribution for  $\boldsymbol{\theta}$  is based on a calibration curve determined as the ordinary least-squares solution to the calibration data from a previous calibration of the meter. Figure 10 shows the new calibration data (from Figure 8), the calibration curve from that previous calibration, and lower and upper bounds on the values of a calibration curve for the new calibration corresponding to the information  $\delta = 0.075 \%$ .

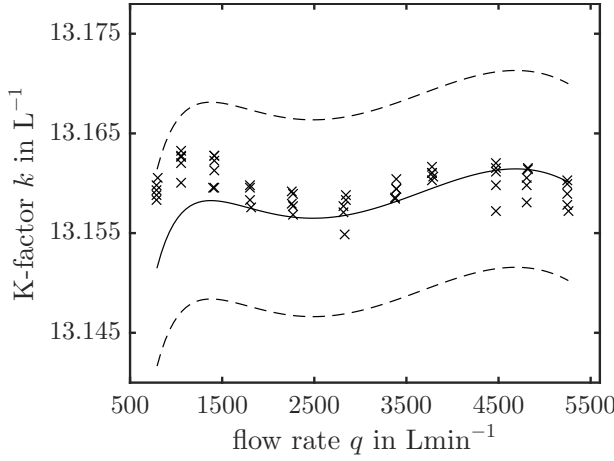


Figure 10: Visualization of the prior (42) for  $\theta$  in the Normal linear regression problem (35). The dashed lines indicate lower and upper bounds on the calibration curve corresponding to the prior information (43) based on a previous calibration (solid curve) and the assumed maximum relative deviation  $\delta = 0.075\%$  from it. The crosses show the data.

### 4.3 Numerical methods

When an analytical expression for the posterior distribution  $\pi(\theta, \sigma^2 | \mathbf{k})$  is not available, Markov Chain Monte Carlo methods are usually applied to carry out the numerical calculation of the posterior distribution (cf. Section 2.3). However, in this case study, which is an example of a Normal linear regression model with prior knowledge that determines whether values of  $\theta$  are feasible or infeasible, the approach proposed in [21, p. 385] can be applied. The approach uses a simple Monte Carlo procedure combined with an accept/reject algorithm to obtain independent samples from the posterior distribution, and comprises the following steps:

1. Determine the posterior distribution for the ‘unconstrained problem’ in which the prior knowledge in the form of the constraint on the values of the calibration curve is ignored and the prior distribution takes the form

$$\pi(\theta, \sigma^2) \propto \text{IG}(\nu_0/2, \nu_0\sigma_0^2/2);$$

2. Sample  $(\theta, \sigma^2)$  from the posterior distribution for the unconstrained problem and discard those samples for which  $f_\theta(q)$  contradicts the constraint used to specify the prior distribution  $\pi(\theta)$ .

The advantage of this approach is that, because the prior distribution for the unconstrained problem is conjugate for the Normal linear regression model, the unconstrained problem can be solved analytically, and samples from the corresponding posterior can be drawn easily.

The analytical solution to the unconstrained problem is the Normal inverse Gamma distribution  $\text{NIG}(\theta_1, \mathbf{V}_1, \alpha_1, \beta_1)$  and can be written as

$$\pi_0(\theta, \sigma^2 | \mathbf{k}) = \pi_0(\sigma^2 | \mathbf{k}) \pi_0(\theta | \sigma^2, \mathbf{k}), \quad (44)$$

where

$$\pi_0(\sigma^2 | \mathbf{k}) = \text{IG}(\alpha_1, \beta_1), \quad \pi_0(\theta | \sigma^2, \mathbf{k}) = \text{N}(\theta_1, \sigma^2 \mathbf{V}_1), \quad (45)$$

with

$$\boldsymbol{\theta}_1 = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{k}, \quad (46)$$

$$\mathbf{V}_1 = (\mathbf{X}^\top \mathbf{X})^{-1}, \quad (47)$$

$$\alpha_1 = \alpha_0 + \frac{1}{2}(n - p), \quad (48)$$

$$\beta_1 = \beta_0 + \frac{1}{2}S_{\text{OLS}}. \quad (49)$$

In the above,  $\mathbf{X}$  is the design matrix with elements  $X_{ij} = q_i^{r_j}$ ,  $\boldsymbol{\theta}_1$  is the ordinary least-squares solution, and

$$S_{\text{OLS}} = (\mathbf{k} - \mathbf{X}\boldsymbol{\theta}_1)^\top (\mathbf{k} - \mathbf{X}\boldsymbol{\theta}_1). \quad (50)$$

In order to sample from the posterior distribution  $\pi(\boldsymbol{\theta}, \sigma^2 | \mathbf{k})$ , the following procedure can be used. First, sample  $\sigma_{(s)}^2$  from  $\pi_0(\sigma^2 | \mathbf{k})$ , and then  $\boldsymbol{\theta}_{(s)}$  from  $\pi_0(\boldsymbol{\theta} | \sigma_{(s)}^2, \mathbf{k})$ . Those pairs  $\{(\boldsymbol{\theta}_{(s)}, \sigma_{(s)}^2)\}$  for which  $\pi(\boldsymbol{\theta}_{(s)}, \sigma_{(s)}^2) = 0$  are rejected. The retained pairs  $\{(\boldsymbol{\theta}_{(r)}, \sigma_{(r)}^2)\}$  are independent samples from the required posterior distribution, and they can be used to approximate the posterior and thereby results derived from it. For example, for a given value  $q$  of flow rate, the values  $\{\boldsymbol{\psi}^\top(q)\boldsymbol{\theta}_{(r)}\}$  constitute independent samples from the posterior distribution of  $f_{\boldsymbol{\theta}}(q)$ .

The described procedure can be generalized to other regression problems with constraints, cf. [21]. Since the samples are independent, and as the posterior exists, convergence assessment is much easier. No burn in phase or thinning is required (as for Markov chains), and the accuracy of results can be expected to be of the order of  $1/\sqrt{N}$ , where  $N$  is the number of accepted samples. In general applications with constraints, however, one has to ensure that the unconstrained posterior exists. Furthermore, if the constraints define only a (very) small portion of the region in which the unconstrained posterior is concentrated, then the procedure can become very inefficient.

A MATLAB<sup>®</sup> implementation of the above approach is provided and described in Appendix A.2. Application of that software using the prior distributions shown in Figures 9 and 10 to the flow meter calibration data in Figure 8 yielded the results given in Table 6 and Figures 11 and 12, where  $q_{\text{mid}}$  is the midpoint of the minimum and maximum flow rates in the measurement range. The results are based on  $10^6$  Monte Carlo trials and, respectively,  $N = 999\,230$  and  $N = 960\,116$  samples accepted from the results of those trials.

$\nu_0$	$f_{\boldsymbol{\theta}}(q_{\text{min}})$ in L <sup>-1</sup>	$f_{\boldsymbol{\theta}}(q_{\text{mid}})$ in L <sup>-1</sup>	$f_{\boldsymbol{\theta}}(q_{\text{max}})$ in L <sup>-1</sup>
1	13.15947 (59)	13.15812 (35)	13.15841 (57)
$n$	13.15937 (101)	13.15811 (65)	13.15841 (107)

Table 6: Estimates and standard uncertainties for the K-factor at three flow rates for prior distributions with  $\nu_0 = 1$  and  $\nu_0 = n$  chosen to represent different degrees of belief in the information about the flow meter repeatability standard deviation provided by the manufacturer.

A comparison of the prior and posterior distributions for  $\sigma$  for the case  $\nu_0 = 1$  (Figure 11) indicates that the prior knowledge about the repeatability of the flow meter specified by the manufacturer is rather conservative compared with the repeatability observed during the calibration. It follows that the standard uncertainties for the K-factor at the three flow rates are smaller, in this case study, when the prior knowledge about  $\sigma$  is less informative. The estimates of the K-factors are generally comparable, except at the minimum flow rate  $q_{\text{min}}$  where the calibration data is closest to violating the prior knowledge about the values of the calibration curve.

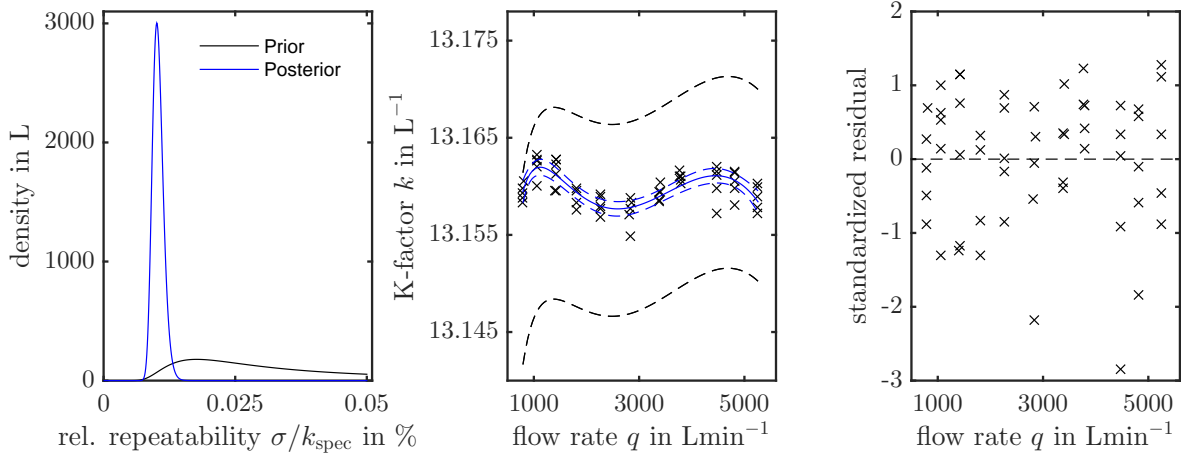


Figure 11: Results for the calibration of a flow meter for a prior distribution with  $\nu_0 = 1$  and  $\delta = 0.075\%$ . Left: prior and posterior distributions for the repeatability standard deviation  $\sigma$ . Middle: calibration data (black crosses), lower and upper bounds on the values of a calibration curve (black broken lines), the estimated calibration curve  $f_{\hat{\theta}}$  (blue solid line), and point-wise credible intervals for probability  $P = 0.95$ . Right: standardized residuals calculated as  $(k_i - f_{\hat{\theta}}(q_i))/\hat{\sigma}$ .

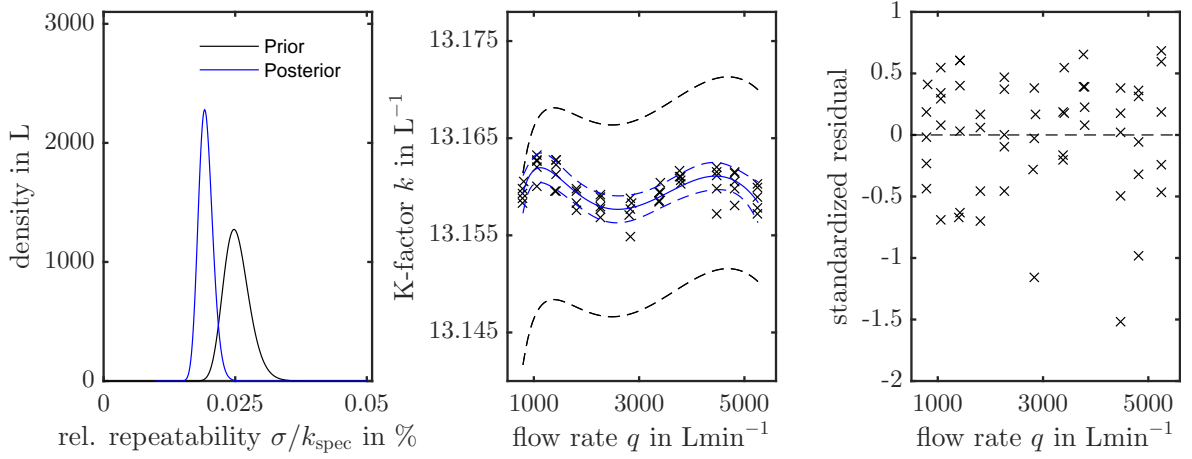


Figure 12: As Figure 11, but for  $\nu_0 = n$  and  $\delta = 0.075\%$ .

The middle graph of Figure 11, for the case  $\nu_0 = 1$  corresponding to when posterior information about  $\sigma$  comes predominantly from the data, shows that the widths of the point-wise credible intervals are generally smaller than the dispersions of the measured data for each nominal flow rate. This property follows since the credible intervals provide information about the estimates of the K-factor obtained by evaluating the calibration curve determined from all the calibration data rather than about an individual measured value of the K-factor.

The treatment described in this case study is not intended to answer questions about whether the flow meter has changed between calibrations. To do so the data obtained from the previous calibration would not be used to provide prior knowledge for the new calibration, but to establish a calibration curve to be compared with that determined by the new calibration.

#### 4.4 Sensitivity analysis

The results given in Section 4.3 illustrate how the posterior information about the flow meter is influenced by the prior knowledge about the flow meter repeatability described by the parameter  $\nu_0$ , which represents the ‘degree-of-belief’ in the value  $\sigma_{0,\text{rel}}$  specified by



the manufacturer. In this section the sensitivity of the posterior information about the K-factor evaluated at the flow rates  $q_{\min}$ ,  $q_{\text{mid}}$  and  $q_{\max}$  on the parameter  $\delta$ , which controls the constraint on the values of the calibration curve, is investigated. Table 7 gives the estimates and standard uncertainties for the K-factor at the three flow rates for prior distributions with  $\nu_0 = n$  and different values of  $\delta$ . (The central value  $\delta = 0.075\%$  corresponds to that used to specify the prior knowledge in Section 4.2.) Figure 13 shows the posterior distributions for the K-factor at the minimum flow rate  $q_{\min}$  and the cases  $\delta = 0.090\%$ ,  $0.075\%$  and  $0.060\%$ , respectively.

The results show that the estimates and standard uncertainties for the K-factor at  $q_{\text{mid}}$  and  $q_{\max}$  are insensitive to the value of  $\delta$ . The estimate and standard uncertainty for the K-factor at  $q_{\min}$ , where the calibration data is closest to violating the prior knowledge about the values of the calibration curve, are also quite insensitive to increasing  $\delta$  from 0.075, but they change appreciably when  $\delta$  is reduced. The graphs in Figure 13 show the considerable influence of  $\delta$  on (the shape of) the posterior distribution for the K-factor at  $q_{\min}$ .

$\delta$ in %	$f_{\theta}(q_{\min})$ in $\text{L}^{-1}$	$f_{\theta}(q_{\text{mid}})$ in $\text{L}^{-1}$	$f_{\theta}(q_{\max})$ in $\text{L}^{-1}$	$N$
0.090	13.15946 (111)	13.15812 (65)	13.15841 (107)	999 733
0.080	13.15944 (107)	13.15811 (65)	13.15841 (107)	990 303
0.075	13.15937 (101)	13.15811 (65)	13.15841 (107)	960 116
0.070	13.15921 (91)	13.15810 (65)	13.15840 (107)	877 376
0.060	13.15856 (66)	13.15805 (65)	13.15836 (107)	485 998

Table 7: Estimates and standard uncertainties for the K-factor at three flow rates for prior distributions with  $\nu_0 = n$  and different values of  $\delta$ .

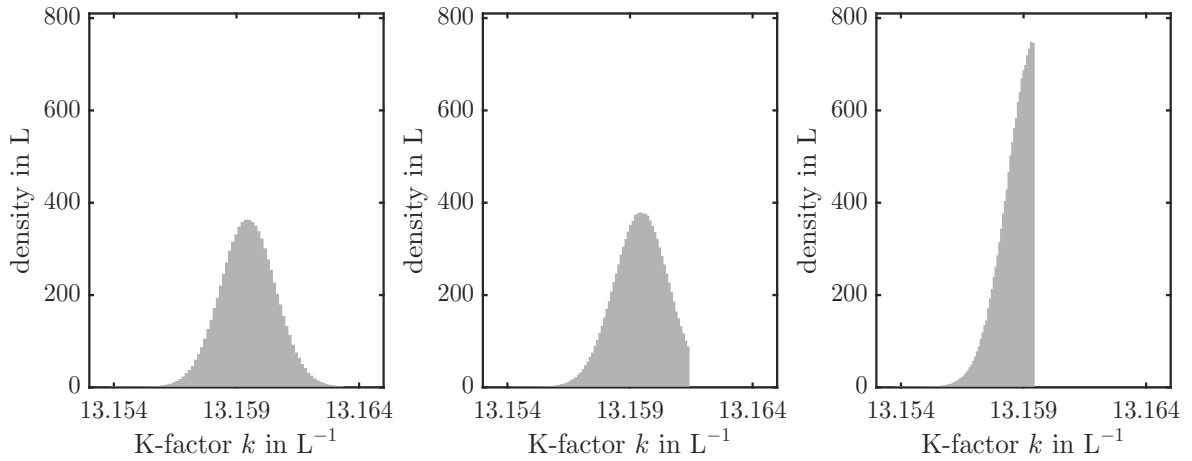


Figure 13: Posterior probability distributions for the value of the K-factor provided by the calibration curve at the minimum flow rate  $q_{\min}$  when  $\nu_0 = n$ . Left:  $\delta = 0.090\%$ . Middle:  $\delta = 0.075\%$ . Right:  $\delta = 0.060\%$ .

The behaviour at the flow rate  $q_{\min}$  illustrated in Figure 13 is a direct consequence of expressing prior knowledge about the parameters  $\theta$  in the form of ‘hard’ constraints on the values of the calibration curve. An advantage of representing the prior knowledge in this way is that the constraints can be arbitrarily complex, since all that is needed for the numerical calculations is to have a way to decide whether the calibration curve defined by a value of  $\theta$  satisfies or not those constraints.

An alternative form of prior distribution might associate small, but non-zero, probabilities

with calibration curves that violate the constraints. But constructing such a prior distribution is not straightforward, and the numerical methods needed to determine the posterior distribution become more sophisticated.

## 5 Inference of thermophysical properties – use of hierarchical priors

This case study presents the determination of the thermophysical properties of a material from transient temperature measurements. The example constitutes a nonlinear regression model, in which we illustrate the use of hierarchical prior distributions. A further feature of this example is that the regression function is not known explicitly but needs to be determined through the numerical solution of a partial differential equation. It is considered as an inverse problem in the sense that the quantity of interest, the thermal diffusivity, is not measured directly but is a characteristic of the material that is responsible for the observed temperature variations. The case study also illustrates the numerical calculation of results by a Metropolis-Hastings algorithm.

### 5.1 Statistical modelling

The goal in this example is to infer the thermal diffusivity of the specimen under test. Data for a thermogram obtained by the so-called FLASH technique [46] are represented in Figure 14. The thermogram shows the (normalized) temperature rise on the back face of a thin disk induced by a short energy pulse applied to the front face. The observed temperature variations can be summarized with two parameters: the characteristic time  $\tau$  which controls the increasing part of the curve and a so-called biot number  $Bi$  which controls the decreasing part of the curve. The inference is performed for both parameters. Then, the thermal diffusivity  $a$  is obtained from the thickness  $e$  of the specimen under test and the characteristic time  $\tau$  according to

$$a = \frac{e^2}{\tau}. \quad (51)$$

The statistical model is given by

$$y_i = f(t_i; \tau, Bi) + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), \quad i = 1, \dots, n, \quad (52)$$

where  $y_i$  is a measured value taken from the normalized thermogram at time  $t_i$ , and the regression function  $f(t; \tau, Bi)$  is defined implicitly as the solution to the one-dimensional heat equation with appropriate boundary conditions. For more details on this application we refer to [1].

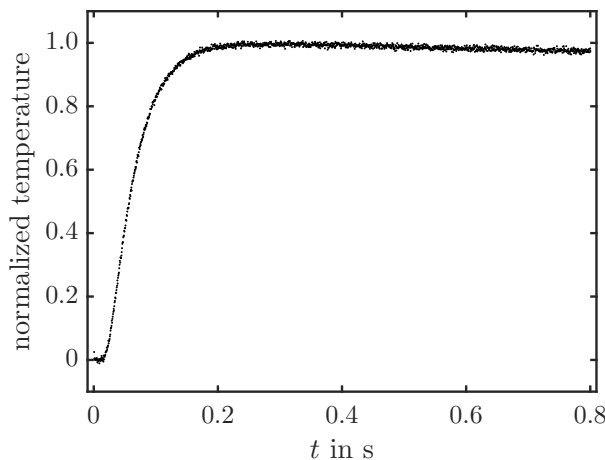


Figure 14: Example data for the problem of inferring thermophysical properties. Displayed is a normalized thermogram obtained from a FLASH experiment for ARMCO iron measurements. The statistical model is given in (52).

The likelihood function is given by

$$l(\tau, Bi, \sigma^2; \mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(t_i; \tau, Bi))^2}. \quad (53)$$

Similarly to the two previous case studies in Sections 3 and 4, a homoscedastic model is considered, i.e., the variance of the observations is constant. Furthermore, the data are assumed to be independently and Normally distributed. However, one particular challenge in this case study is that the regression function is not given in analytic form, but is determined numerically by solving a partial differential equation. As a consequence, calculation of the posterior distribution can only be done numerically, which requires repeated (potentially time-consuming) evaluations of the regression function.

## 5.2 Prior distributions

Prior knowledge about a quantity of interest may be given by experts of the domain in different ways. The knowledge available may be more or less informative, and this can be modelled using either informative or non-informative prior probability distribution [1]. In this section, it is considered that the experts have good confidence about a Gaussian distribution for the parameters  $\tau$  and  $Bi$ . Nevertheless, the associated means and variances may not be reliably known. For example, the mean of  $\tau$  (respectively  $Bi$ ), denoted as  $\mu_\tau$  (respectively  $\mu_{Bi}$ ) may be known to lie within a stipulated interval. It may be difficult also to give a reliable estimate of the variance, denoted as  $\varphi_\tau$  (respectively  $\varphi_{Bi}$ ). In order to deal with such kind of knowledge, we use hierarchical priors for the parameters  $\tau$  and  $Bi$  in the statistical model (52), and we select the hyperpriors in these hierarchical priors in the light of the available prior knowledge.

**Choice of family of priors.** The hierarchical priors used for  $\tau$  and  $Bi$  are given by

$$\begin{aligned}\tau | \mu_\tau, \varphi_\tau &\sim \text{N}(\mu_\tau, \varphi_\tau), \\ \mu_\tau &\sim \text{U}(\tau_1, \tau_2), \\ \varphi_\tau &\sim \text{IG}(\alpha_\tau, \beta_\tau),\end{aligned}\tag{54}$$

and

$$\begin{aligned}Bi | \mu_{Bi}, \varphi_{Bi} &\sim \text{N}(\mu_{Bi}, \varphi_{Bi}), \\ \mu_{Bi} &\sim \text{U}(Bi_1, Bi_2), \\ \varphi_{Bi} &\sim \text{IG}(\alpha_{Bi}, \beta_{Bi}).\end{aligned}\tag{55}$$

The prior knowledge about the variance  $\sigma^2$  is expressed in terms of an inverse Gamma distribution:

$$\sigma^2 \sim \text{IG}(\alpha_0, \beta_0).\tag{56}$$

All parameters are treated as being independent *a priori*. The motivation for using the particular priors (54) and (55) were discussions with experts.

**Informative prior for  $\sigma^2$ .** The prior knowledge about the variance of observations from thermograms, cf. Figure 14, is based on recent experimental data. Since thermograms are normalized, it is reasonable to expect that the variability of the observations in one thermogram is similar to that in previous ones provided by the same device. For the example introduced in Section 5.1 this implies that the variance is expected to be around  $10^{-3}$ , which we express in terms of an inverse Gamma distribution  $\text{IG}(\alpha_0, \beta_0)$  with parameters  $\alpha_0 = 3$  and  $\beta_0 = 0.01$  (see Figure 15).

**Informative prior for  $\tau$  and  $Bi$ .** A discussion with experts about the ARMCO iron measurements at ambient temperature (cf. Section 5.1) led to the particular choice of parameters for the hierarchical priors given in Table 8 (see also Figure 15).

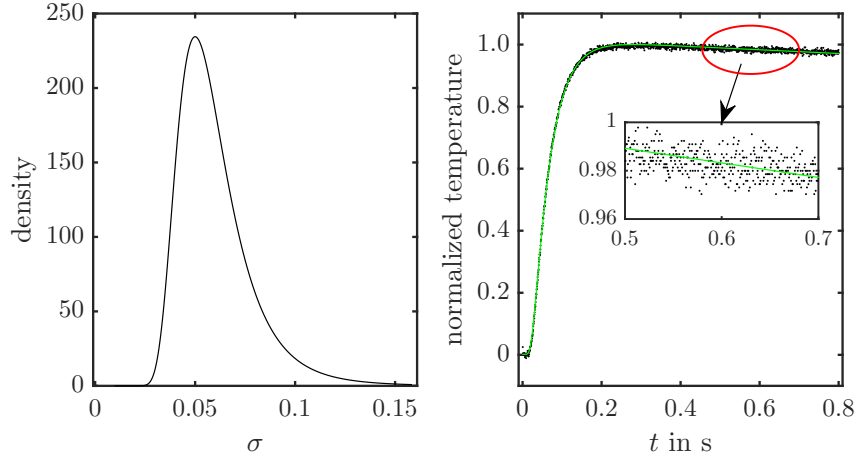


Figure 15: Visualization of the prior distributions (54) - (56) (with the parameters specified in Table 8) for the problem of inferring thermophysical properties under model (52). Left: Prior distribution  $\pi(\sigma) = 2\sigma\pi(\sigma^2)$  for the standard deviation of observations. Right: Mean regression curve (solid line). The dotted points show the data.

$\tau_1$ in s	$\tau_2$ in s	$\alpha_\tau$	$\beta_\tau$ in s <sup>2</sup>	$Bi_1$	$Bi_2$	$\alpha_{Bi}$	$\beta_{Bi}$	$\alpha_0$	$\beta_0$
0.4	0.5	4	$3 \times 10^{-5}$	0.001	1	4	$3 \times 10^{-2}$	3	0.01

Table 8: Parameters of the prior distributions (54), (55) and (56) for  $\tau$ ,  $b$  and  $\sigma^2$  for the problem of inferring thermophysical properties under model (52).

### 5.3 Numerical methods

This example illustrates the use of a particular Metropolis-Hastings algorithm [60] for the calculation of results by a Monte Carlo method. The proposed symmetric single-site Metropolis-Hastings algorithm samples and tests proposal values for each parameter individually, and it is described in Appendix A.3. Its application to the data from Figure 14 produced the results given in Table 9 and Figure 16, where a chain of length  $10^4$  has been used.

Quantity	Estimate	Standard uncertainty	95 % coverage interval
$\tau$ in s	0.4194	0.0004	[0.4187, 0.4201]
$Bi$	0.0129	0.0002	[0.0126, 0.0132]
$\sigma$	0.0065	0.0001	[0.0063, 0.0068]

Table 9: Results for the problem of inferring thermophysical properties obtained by fitting model (52) and the priors (54) - (56) (with the parameters specified in Table 8) to the data given in Figure 14. Given are the posterior means and standard deviations, as well as 95 % (probabilistically symmetric) credible intervals for the regression and additional parameters.

The posterior distribution associated with the thermal diffusivity is determined using a Monte Carlo simulation based on equation (51), with a Gaussian distribution for the thickness  $e$  of the specimen and the posterior sample for  $\tau$ . The corresponding results are given in Table 10 and Figure 17.

In order to assess convergence of the Markov chains the procedures of Gelman-Rubin [24] and Brooks-Gelman [12] have been applied. Figure 18 shows corresponding results which indicate convergence of the Markov chains. In this case study, one can reasonably conclude

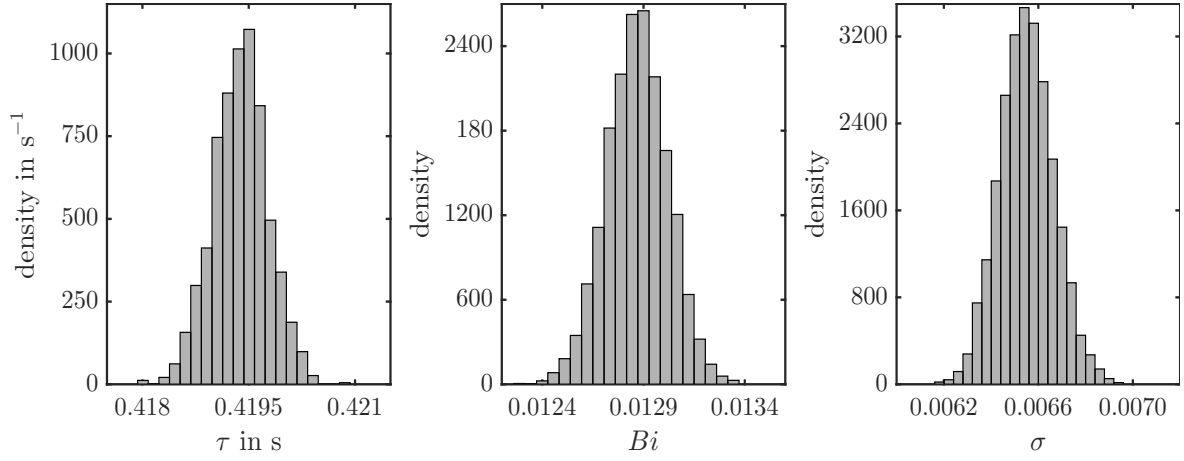


Figure 16: Posterior distributions of  $\tau$  (left),  $Bi$  (middle) and  $\sigma$  (right) for the problem of inferring thermophysical properties obtained by fitting model (52) to the data given in Figure 14, using the priors (54) – (56) (with the parameters specified in Table 8).

Quantity	Estimate	Standard uncertainty	95 % coverage interval
$a \cdot 10^{-6}$ in $\text{m}^2\text{s}^{-1}$	20.23	0.03	[20.18, 20.29]

Table 10: Results for the thermal diffusivity determined from the posterior distribution of  $\tau$  and the thickness of the sample  $e$ , using a Monte Carlo simulation. Given are the associated mean and standard deviation, as well as 95 % (probabilistically symmetric) coverage interval

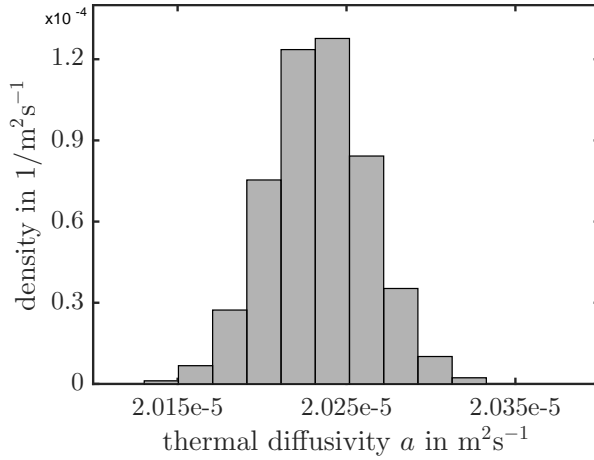


Figure 17: Results for the thermal diffusivity  $a$  determined from the posterior distribution of  $\tau$  and the thickness of the sample  $e$ , using a Monte Carlo simulation.

that convergence is reached for 4000 iterations.

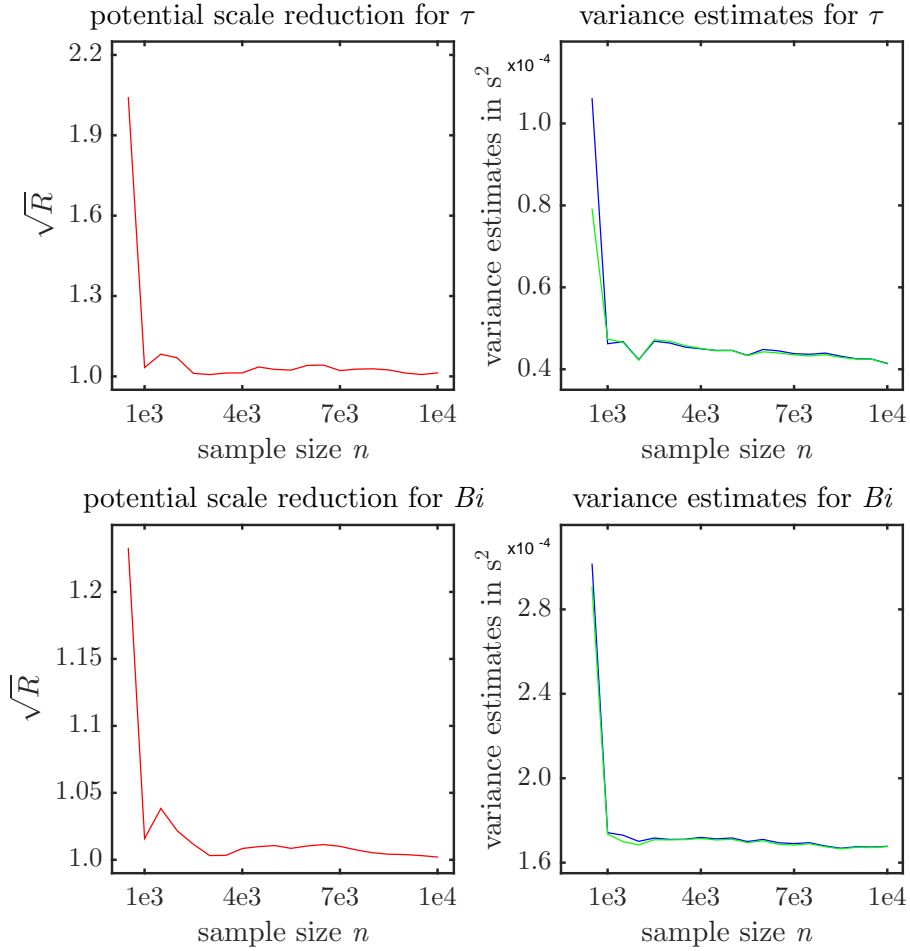


Figure 18: Gelman-Rubin convergence statistics for the problem of inferring thermophysical properties. On the left are represented the potential scale reduction  $\sqrt{R}$  for both  $\tau$  (top) and  $Bi$  (bottom). These curves should approach 1. On the right are represented the pooled (in blue) and within-sequence (in green) variances for  $\tau$  (top) and  $Bi$  (bottom). Each of these curves should approach a constant value. According to [12], convergence is assessed if both conditions are fulfilled.

## 5.4 Sensitivity analysis

This example illustrates the effect of a different choice for the prior distributions. In the first place, we evaluate whether a different choice for the values of the hyperparameters related to the characteristic time  $\tau$  ( $\tau_1$ ,  $\tau_2$ ,  $\alpha_\tau$  and  $\beta_\tau$ ) has an effect on the inference of thermophysical properties. The results are gathered in Table 11. The inference appears very robust to any change in the hyper parameters. In the second place, the Gaussian prior distribution with hyperparameters is changed to a rectangular prior with bounds  $\tau_{\inf}$  and  $\tau_{\sup}$ . The corresponding results are summarized in Table 12.

With a rectangular prior distribution, the Metropolis-Hastings algorithm leads to the same estimate for  $\tau$  each time the value revealed by the data ( $\hat{\tau} = 0.4194$ ) lies within the domain of the prior distribution. If not, the algorithm gives poor estimates. The reason is that the choice of a rectangular distribution (or more generally of a bounded prior distribution) associates a zero probability with all values outside the domain. If one choose a bounded prior distribution, one has to be absolutely sure that any value outside the domain of the distribution is impossible.

$\tau_1$ in s	$\tau_2$ in s	$\alpha_\tau$	$\beta_\tau$ in s <sup>2</sup>	$\hat{\tau}$ in s	$u(\hat{\tau})$ in s	$I_{95\%}(\tau)$ in s
0.4	0.5	4	$3 \cdot 10^{-5}$	0.4194	0.0004	[0.4187, 0.4201]
0.3	0.4	4	$3 \cdot 10^{-5}$	0.4194	0.0004	[0.4187, 0.4201]
0.4	0.5	4	$3 \cdot 10^{-2}$	0.4194	0.0004	[0.4186, 0.4202]
0.4	0.5	2	$3 \cdot 10^{-5}$	0.4194	0.0004	[0.4185, 0.4200]

Table 11: Sensitivity results with respect to different choice of parameters for the Gaussian prior with hyperparameters.

$\tau_{\inf}$ in s	$\tau_{\sup}$ in s	$\hat{\tau}$ in s	$u(\hat{\tau})$ in s	$I_{95\%}(\tau)$ in s
0.44	0.46	0.4400	0.0001	[0.4400, 0.4401]
0.4	0.5	0.4194	0.0004	[0.4187, 0.4202]
0.25	0.35	0.3500	0.0001	[0.3498, 0.3500]
0.2	0.8	0.4194	0.0004	[0.4186, 0.4201]

Table 12: Sensitivity results with respect to the choice of a rectangular prior distribution with bounds  $\tau_{\inf}$  and  $\tau_{\sup}$ .



## 6 Analysis of immunoassay data – a statistical calibration problem

This case study presents the analysis of a bioanalytical test. The example illustrates a statistical calibration problem, in which the posterior distribution determined in the calibration stage is used as (part of) the prior distribution for the subsequent prediction. This case study also exemplifies how prior distributions can be derived from historical measurements. In contrast to the previous examples the variances of the observations are not constant, and we present their inference from the data. Furthermore, the use of Gibbs sampling and its implementation in BUGS are illustrated.

### 6.1 Statistical modelling

An immunoassay is a type of bioanalytical test that exploits the highly specific binding between antibodies and antigens [15]. Immunoassays are able to measure very small concentrations of a substance in solutions and have an immense range of applications, for example, to detect the presence of an infection, of hormones or drugs. Figure 19 shows calibration data for an ELISA (Enzyme-Linked Immunosorbent Assay) where fluorescence intensity has been measured for known values of concentration. For more details about this problem we refer to [39].

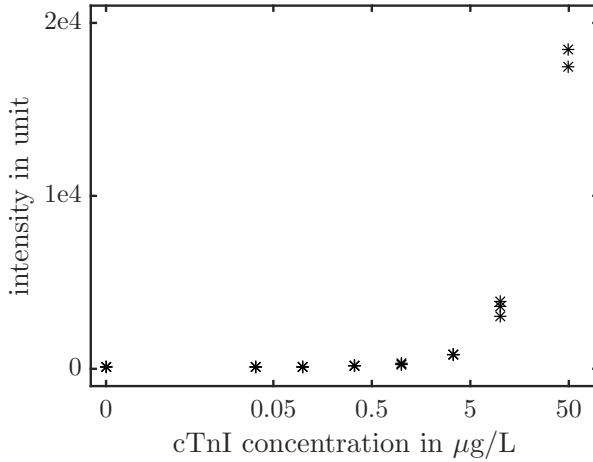


Figure 19: Example data for immunoassay analyses. Displayed are the calibration measurements of an ELISA performed at PTB within the international comparability study CCQM P58.1 [39]. The study assessed the capability of measuring concentrations of human cardiac troponin I (cTnI). The statistical model is given in (57).

The statistical model is

$$y_i = f_{\boldsymbol{\theta}}(x_i) + \epsilon_i, \quad \epsilon_i \sim N(0, ax_i + c), \quad i = 1, \dots, n, \quad (57)$$

where

$$f_{\boldsymbol{\theta}}(x) = \theta_1 + \frac{\theta_2 - \theta_1}{1 + \left(\frac{x}{\theta_3}\right)^{\theta_4}}, \quad (58)$$

and  $\boldsymbol{\delta} = (a, c)^\top$  are additional variance parameters. The  $\epsilon_i$  in (57) are assumed to be independent. The likelihood function is thus given by

$$l(\boldsymbol{\theta}, a, c; \mathbf{y}) = \prod_{i=1}^n \frac{1}{(2\pi(ax_i + c))^{1/2}} e^{-\frac{1}{2(ax_i + c)}(y_i - f_{\boldsymbol{\theta}}(x_i))^2}. \quad (59)$$

The analysis of immunoassays is one of statistical calibration. In the regression task the parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\delta}$  are inferred from measurements  $y_1, \dots, y_n$  of fluorescence intensity at

known concentrations  $x_1, \dots, x_n$ . In the prediction task an unknown concentration  $\tilde{x}$  is then inferred from measurements  $\tilde{y}_1, \dots, \tilde{y}_m$  taken for that concentration. To this end, Bayes' theorem is applied a second time in the form

$$\pi(\tilde{x}, \boldsymbol{\theta}, a, c | \tilde{\mathbf{y}}) \propto \pi(\tilde{x}, \boldsymbol{\theta}, a, c) \times l(\tilde{x}, \boldsymbol{\theta}, a, c; \tilde{\mathbf{y}}). \quad (60)$$

Here,

$$\pi(\tilde{x}, \boldsymbol{\theta}, a, c) = \pi(\tilde{x}) \times \pi(\boldsymbol{\theta}, a, c) \quad (61)$$

is the prior distribution, and involves the posterior distribution  $\pi(\boldsymbol{\theta}, a, c) \equiv \pi(\boldsymbol{\theta}, a, c | \mathbf{y})$  provided as the solution to the regression task, and

$$l(\tilde{x}, \boldsymbol{\theta}, a, c; \tilde{\mathbf{y}}) = \prod_{i=1}^m \frac{1}{(2\pi(a\tilde{x} + c))^{1/2}} e^{-\frac{1}{2(a\tilde{x} + c)}(\tilde{y}_i - f_{\boldsymbol{\theta}}(\tilde{x}))^2} \quad (62)$$

is the likelihood function, and involves the new observations of fluorescence intensity. In contrast to the flow meter calibration problem the regression function is non-linear, and the statistical model is also heteroscedastic, i.e., the variances of the observations are not constant. The sampling distribution, however, is Gaussian as before.

## 6.2 Prior distributions

This example illustrates how historical data can be used to determine an informative prior distribution for a regression problem. The analysis of immunoassays also includes a (subsequent) prediction problem, and we show how to construct the corresponding prior by utilizing the posterior obtained as the solution to the regression problem. The statistical model for this application is given by expressions (57) and (58).

An additional property of the (explicit) priors derived here is that we expect them to be universally applicable in similar immunoassay analyses. The reasons are that noninformative priors are chosen for those parameters that are instrument dependent, and the specification of the informative priors is based on an assumption made about the design of the experiment that an experienced user will likely satisfy.

**Choice of family of priors.** The prior distribution for the regression problem is chosen as

$$\pi(\boldsymbol{\theta}, a, c) = \pi(\theta_1)\pi(\theta_2)\pi(\theta_3)\pi(\theta_4)\pi(a)\pi(c), \quad (63)$$

where the priors for the individual parameters are described below.

**Prior for  $a$  and  $c$ .** We choose the constant priors

$$\pi(a) \propto 1 \ (0 \leq a < \infty), \quad \pi(c) \propto 1 \ (0 \leq c < \infty). \quad (64)$$

These prior distributions are generally applicable for the analysis of immunoassays (provided that the statistical model (57) is appropriate), independent of employed instruments, capabilities, etc. We note that in the immunoassay regression problem the response variable for each concentration is typically measured repeatedly and at different concentrations. Hence, the inference of the two variance parameters should be well-behaved under prior ignorance, and the particular form of the noninformative prior is not expected to be influential.

**Prior for  $\boldsymbol{\theta}$ .** The two parameters  $\theta_1$  and  $\theta_2$  that enter linearly in the regression function (58) are also expected to depend on the employed instrument or the considered substance, and we utilize flat improper priors

$$\theta_1 \sim U(0, \infty), \quad \theta_2 \sim U(-\infty, \infty). \quad (65)$$

Since  $\theta_1$  and  $\theta_2$  enter linearly, their inference under prior ignorance is expected to be well-behaved, provided that the immunoassay design and the priors for  $\theta_3$  and  $\theta_4$  determine the corresponding linear relationship well. Note also that the particular choice of the improper prior (65) for  $\theta_1$  and  $\theta_2$  may be justified by the fact that these unknowns act as location parameters in the likelihood (59).

The two parameters  $\theta_3$  and  $\theta_4$ , finally, enter nonlinearly in the regression function (58), and we use historical data to derive informative priors for these parameters. The historical data stem from the international comparison study CCQM-P58, cf. [37]. Figure 20 shows the posterior distributions for the (scaled)  $\theta_3$  and for  $\theta_4$  gained in a Bayesian analysis [37] of the data from this study.

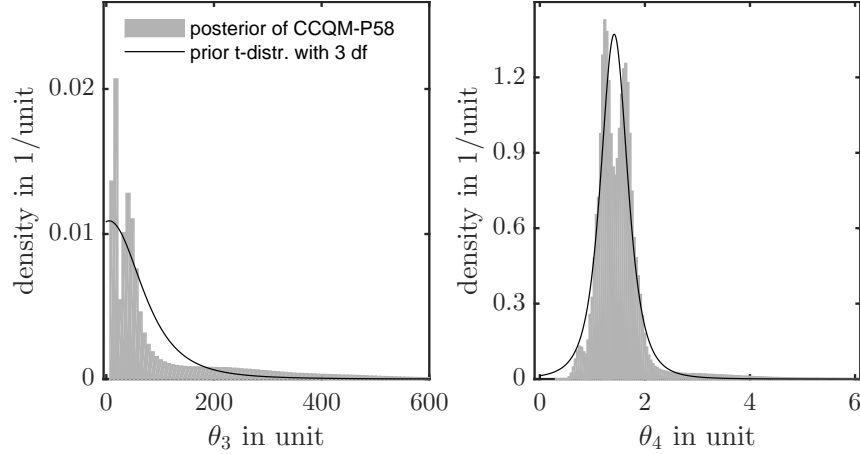


Figure 20: Visualization of the prior distributions (66) and (67) (in black) for  $\theta_3$  (left) and  $\theta_4$  (right) for immunoassay analyses under Model (57). These priors are derived from the posterior distributions of the international comparison study CCQM-P58 [37] with  $x^{\max} = 50$  unit (displayed in grey).

The scaling of  $\theta_3$  was done (for each single regression problem) by dividing the parameter by the largest concentration  $x^{\max} = \max_i x_i$ . The rationale for this scaling is that  $\theta_3$  determines the point of inflection of the regression curve (58), and for properly designed measurements the selected concentrations will fairly cover the whole shape of the regression curve, without adding many extra points on the upper asymptote. In other words, assuming the immunoassay to have been designed properly, we can expect a certain range for  $\theta_3$  in terms of the maximum concentration used. Hence, the ratio of  $\theta_3/x^{\max}$  should be transferable. In order to reach an explicit expression for the prior, a truncated, scaled and shifted  $t$ -distribution with 3 degrees of freedom was fitted to the posterior distribution, c.f. Figure 20 and Table 13. The resulting  $t$ -distribution covers the range of the posterior distribution, and due to the chosen low degree of freedom it allows at the same time for large deviations. The resulting prior for  $\theta_3$  is explicitly given by

$$\theta_3/x^{\max} \sim t_3(0.1, 1.6)U(0, \infty), \quad (66)$$

where  $x^{\max}$  refers to the maximum concentration used in the immunoassay to be analysed.

$\theta_1$ in unit	$\theta_2$ in unit	$\theta_3/x^{\max}$	$\theta_4$	$a$ in $\frac{\text{L} \cdot \text{unit}^2}{\mu\text{g}}$	$c$ in $\text{unit}^2$
$U(0, \infty)$	$U(-\infty, \infty)$	$t_3(0.1, 1.6)U(0, \infty)$	$t_3(1.4, 0.073)U(0, \infty)$	$U(0, \infty)$	$U(0, \infty)$

Table 13: Distributions used in the prior (63) for the regression and additional parameters for immunoassay analyses under model (57).

The parameter  $\theta_4$  is expected to be transferable as long as the immunoassay leads to a regression curve with a reasonable slope in the region of its linear increase, which is usually

ensured by a careful design. Figure 20 shows the posterior distribution for  $\theta_4$  obtained from the CCQM-P58 data. Again, a truncated, scaled and shifted  $t$ -distribution with 3 degrees of freedom was fitted to this posterior distribution, resulting in the prior

$$\theta_4 \sim t_3(1.4, 0.073)U(0, \infty). \quad (67)$$

For further details on the choice of these priors we refer to [39]. Figure 21 illustrates the prior knowledge about the regression curve.

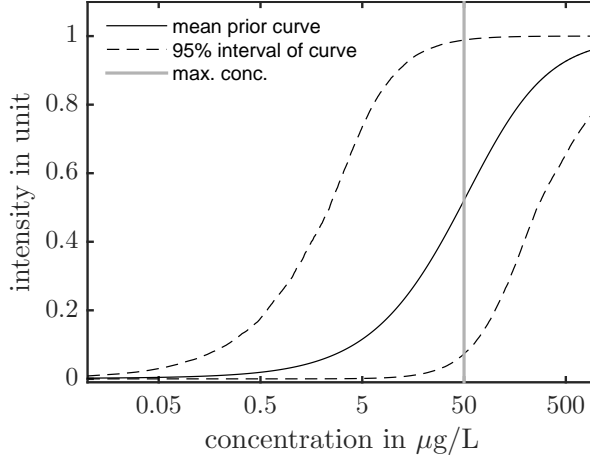


Figure 21: Visualization of the prior distributions (66) and (67) for immunoassay analyses under model (57): Mean regression curve (solid line) and pointwise 95 % intervals (dashed lines) plotted for a maximum concentration of  $50 \mu\text{g L}^{-1}$  and for fixed values of the asymptotes  $\beta_1 = 1$  and  $\beta_2 = 0$ . However, the priors hold for any scaled set of concentrations and any intensity measurement.

**Choice of prior for subsequent prediction.** For the prediction problem the prior

$$\pi(\tilde{x}, \boldsymbol{\theta}, a, c) \propto 1 \times \pi(\boldsymbol{\theta}, a, c | \mathbf{y}) \quad (68)$$

is chosen, i.e., we combine a constant noninformative prior for the unknown concentration  $\tilde{x}$  with the posterior distribution from the regression problem. The prior for the prediction task is thus automatically determined once the regression problem has been solved.

### 6.3 Numerical methods

This example illustrates the flexibility of Gibbs sampling and its powerful BUGS implementation, and we provide explicit source code. A regression problem and a prediction task are considered, where a sample from the posterior of the regression problem is reused for the subsequent prediction. The supplied implementation in WinBUGS treats the regression and the prediction problem simultaneously. The underlying algorithm mainly follows the general description of the Gibbs sampler given in Section 2.3, and corresponding WinBUGS source is given in Appendix A.4.

This WinBUGS source is applied to the immunoassay calibration data in Figure 19. To improve numerical stability, the measured intensities are scaled beforehand by their order of magnitude (i.e.,  $y_i$  is replaced by  $y_i/e^{\lceil \log(y^{\max}) \rceil}$  where  $y^{\max} = \max_i(y_i)$  and  $\lceil z \rceil$  is the smallest integer greater than or equal to  $z$ ). Applying the priors (63) – (68), we obtain the results shown in Table 14 and Figure 22. Table 14 also includes the subsequent fluorescence measurements made at the unknown concentration  $x$ , together with the derived concentration estimate, and Figure 23 shows the corresponding posterior distribution. Figure 22 also shows a residual plot, which indicates no inconsistency of the data and the fitted model. The convergence assessment illustrated in Figure 24 suggests that the Markov chains appear to have reached convergence.

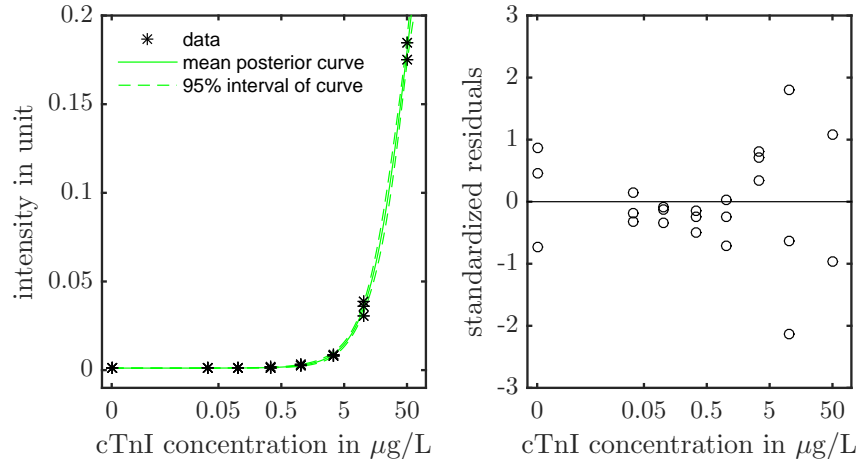


Figure 22: Results for the immunoassay analysis obtained by fitting model (57) and priors (63) – (68) to the data given in Figure 19. Left: Estimated regression function and pointwise 95 % credible intervals. The stars show the data. Right: Expected value of the standardized residuals, i.e., expectation of  $(y_i - f_{\theta}(x_i)) / \sqrt{ax_i + c}$  with respect to the parameters  $\theta$  and  $c$ .

	$\theta_1$ in unit	$\theta_2 \times 10^{-3}$ in unit	$\theta_3$ in $\frac{\mu\text{g}}{\text{L}}$	$\theta_4$
$\hat{\theta}$	0.404	1.14	59.0	1.39
$I_{0.95}(\theta)$	[0.279, 0.710]	[1.05, 1.24]	[34.8, 119]	[1.23, 1.57]

	$a \times 10^{-7}$ in $\frac{\text{L} \cdot \text{unit}^2}{\mu\text{g}}$	$c \times 10^{-9}$ in $\text{unit}^2$
$\hat{\delta}$	4.99	10.2
$I_{0.95}(\delta)$	[2.25, 10.7]	[1.37, 42.0]

	Sample 1	Sample 2
$\tilde{\mathbf{y}}_i$	(280, 305, 334)	(807, 677, 1078)
$\hat{x}_i$	1.20 $\frac{\mu\text{g}}{\text{L}}$	3.26 $\frac{\mu\text{g}}{\text{L}}$
$I_{0.95}(x_i)$	[0.768, 1.72] $\frac{\mu\text{g}}{\text{L}}$	[2.68, 3.89] $\frac{\mu\text{g}}{\text{L}}$

Table 14: Results for the immunoassay analysis obtained by fitting model (57) and priors (63) – (68) to the data given in Figure 19. The MCMC is run with 10 chains and 160,000 iterations, where the first half of the chain is discarded and only every tenth sample is retained. Upper/middle part: Posterior means and 95 % (probabilistically symmetric) credible intervals for the regression and additional parameters. Lower part: Posterior means and 95 % (probabilistically symmetric) credible intervals for the predicted concentration of two additional measurements  $\tilde{\mathbf{y}}$ .

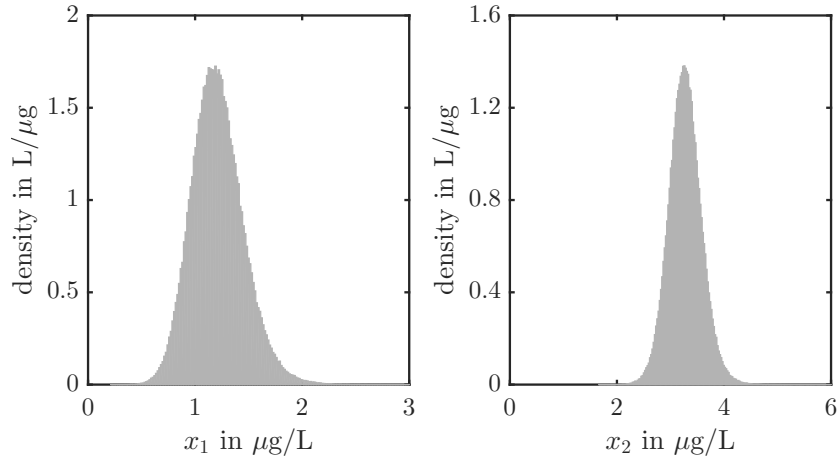


Figure 23: Posterior distributions of the concentrations  $x_1$  (left) and  $x_2$  (right) for the immunoassay analysis, predicted from the fluorescence measurements  $\hat{\mathbf{y}}_1$  and  $\hat{\mathbf{y}}_2$ , respectively, given in Table 14 (bottom) by applying Model (57) and the regression function in Figure 22.

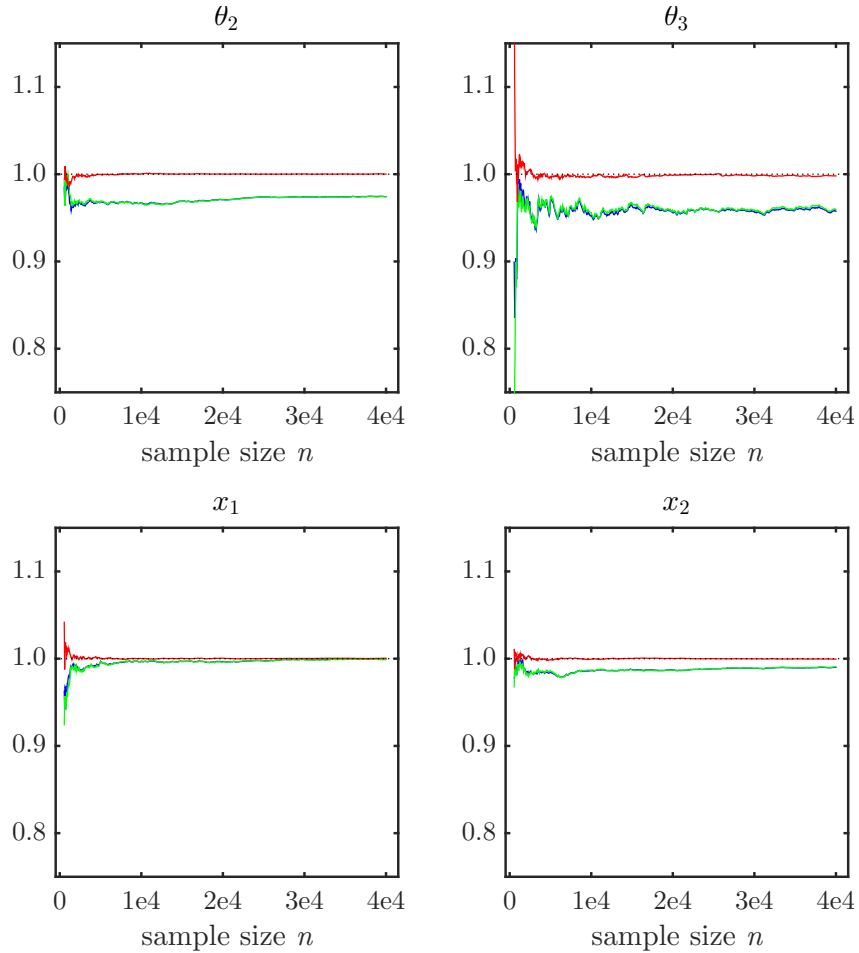


Figure 24: Gelman-Rubin convergence statistic, as modified by [12] for the immunoassay analysis. Displayed are an overall scale measure (i.e., length of the 80 % interval, in green), the corresponding within chain scale measure (in blue, both normalised to a maximum of 1) and their ratio (in red) for the unknown concentrations  $x_1$  and  $x_2$  (bottom row) and for the calibration parameters  $\theta_2$  and  $\theta_3$  (top). After thinning, these measures were calculated for the first  $50 \cdot k$  iterations ( $k = 1, 2, \dots$ ) after the burn in.

## 6.4 Sensitivity analysis

In this example we illustrate a sensitivity analysis with respect to both the employed priors and the statistical model. Since the prior for the prediction problem is taken as the posterior obtained in the regression task, together with a flat noninformative prior for the unknown concentration, the variation of the prior for the regression task naturally carries over to the prediction problem.

We consider the following variations of the priors in (63) for the regression problem. The priors for  $\theta_1, \theta_2, a, c$  are constant over their support and are not be altered. The two informative priors for  $\theta_3$  and  $\theta_4$  are truncated  $t$ -distributions, and they are varried by rescaling the variance parameter of those  $t$ -distributions, i.e., we consider

$$\theta_3/x^{max} \sim t_3(0.1, 1.6k^2) \text{U}(0, \infty), \quad (69)$$

where  $x^{max}$  refers to the maximum concentration used in the immunoassay to be analysed, and

$$\theta_4 \sim t_3(1.4, 0.073k^2) \text{U}(0, \infty), \quad (70)$$

where  $k \geq 1$  controls the reduction in the *a priori* information used. For  $k \rightarrow \infty$ , these priors become noninformative. Table 15 and Figure 25 show the results for the predicted concentration under these priors and when using the observations for the unknown concentration (cf. Table 14), which are similar to those obtained by the original prior ( $k = 1$ ) but imply slightly larger uncertainty.

Prior	$\hat{x}_1/\mu\text{g L}^{-1}$	$I_{0.95}(x_1)/\mu\text{g L}^{-1}$	$\hat{x}_2/\mu\text{g L}^{-1}$	$I_{0.95}(x_2)/\mu\text{g L}^{-1}$
$k = 1$	1.20	[0.768, 1.72]	3.26	[2.68, 3.89]
$k = 2$	1.17	[0.723, 1.71]	3.23	[2.62, 3.88]
$k = 3$	1.16	[0.698, 1.70]	3.21	[2.58, 3.87]

Table 15: Results of the sensitivity analysis for immunoassays. Posterior means and 95 % (probabilistically symmetric) credible intervals for the regression parameters and predicted concentration, obtained by fitting model (57) and priors (69) and (70) for  $k = 1, 2, 3$  as well as priors (63) – (65) and (68) to the data given in Figure 19.

Next we consider variations in the statistical model used for the analysis, where we concentrate on a variation of the error model. Specifically, instead of the assumption made in (57) we assume in the analysis the homoscedastic model

$$y_i = f_{\theta}(x_i) + \epsilon_i, \quad \epsilon_i \stackrel{iid}{\sim} \text{N}(0, \sigma^2), \quad i = 1, \dots, n. \quad (71)$$

Because measurement errors are increasing with increasing concentration, this model does not fit the data in Figure 19, as can be observed in the residual plot in Figure 25.

We conclude this case study by noting that a (partly) informative prior is needed in order to enable an analysis at all. That is, without utilizing prior knowledge the problem turns out to be non-identifiable. For additional reading about this aspect and further results we refer to [39].

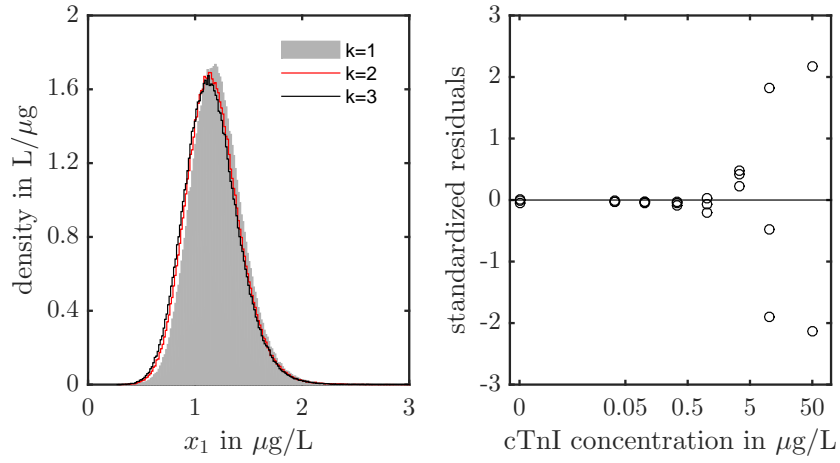


Figure 25: Results of the sensitivity analysis for immunoassays. Left: Posterior distributions of the concentration  $x_1$  obtained by fitting model (57) and priors (69) and (70) for  $k = 2, 3$  (lines) and  $k = 1$  (in grey) as well as priors (63) – (65) and (68) to the data given in Figure 19. Right: The misfit of model (71) when fitted to the data given in Figure 19 with the priors (63) – (68) is illustrated by the expected value of the standardized residuals, i.e., expectation of  $(y_i - f_{\theta}(x_i)) / \sigma$  with respect to the parameters  $\theta$  and  $\sigma$ .



## 7 Analysis of digital PCR data – a parametric inverse problem

This case study represents the analysis of digital PCR data. The example constitutes a parametric inverse-problem rather than a regression task. Here we demonstrate that such problems can be treated by the very same Bayesian methodology as the preceding regression tasks. The example also illustrates the selection of different families of prior distributions to reflect different parameter domains. Furthermore, we present a Metropolis-Hastings algorithm for numerical calculations together with a variable transformation enhancing convergence.

### 7.1 Statistical modelling

Digital PCR (dPCR) [59] is a relatively recent measurement method for the absolute determination of the copy number concentration in a DNA or RNA sample. The method works by partitioning the sample into small volumes with each containing up to a few samples molecules on a chip and amplifying the DNA over a number of cycles. The DNA is measured using a fluorescent marker, and the cycles at which the fluorescence reaches a threshold, called quantification cycles or  $C_q$  can be used to estimate the copy number concentration. Figure 26 shows some dPCR measurements. We refer to [61] for further details about these data.

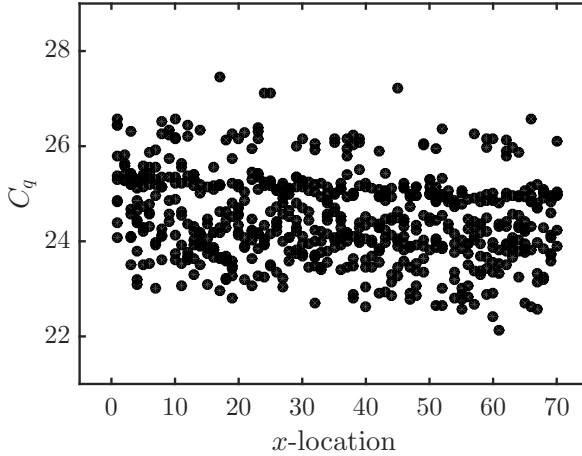


Figure 26: Example data for the digital PCR problem. The plot shows the quantification cycle values  $C_q$ . The number  $n_0$  of partitions for which the threshold is not reached is  $n_0 = 135$ , and the number of low outliers  $n_1$ , high outliers and unreliable  $C_q$  values  $n_2$  and remaining  $C_q$  values  $n_3$  are  $n_1 = 0, n_2 = 4, n_3 = 631$ . The statistical model is given in (72).

The likelihood function is given by

$$l(\mu, \nu, E_0, E, A, b_x, b_y; \mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{n}) \approx p(0, 0 | \mu, \nu)^{n_0} [1 - p(0, 0 | \mu, \nu)]^{n_1} p(1, 0 | \mu, \nu)^{n_2} \times \prod_{j=1}^{n_3} \delta^{-1} \sum_{i=1}^{m2^{c_0}} p(i, c_0 | \mu, \nu) \left[ \Phi(h; G, GA(1-E) \{(1+E_0)(1+E)^{c-1} - 1\}) - \Phi(h; G(1+E)^\delta, GA(1-E)(1+E)^\delta \{(1+E_0)(1+E)^{c+\delta-1} - 1\}) \right], \quad (72)$$

where

$$G = Ai(1+E_0)(1+E)^{c-1} \quad \text{and} \quad c = c_j - c_0 - b_x x_j - b_y y_j. \quad (73)$$

The parameters are the mean initial number of molecules per well  $\mu$ , the dispersion parameter of the Conway-Maxwell distribution of the initial number of molecules  $\nu$ , the efficiency for the first amplification step  $E_0$ , the subsequent efficiency  $E$ , the relative fluorescence per molecule  $A$ , and the  $x$ -trend  $b_x$  and  $y$ -trend  $b_y$ .  $\Phi$  denotes the distribution function of a Normal distribution,  $\delta > 0$  is a small constant used to approximate differentiation,  $p(i, c | \mu, \nu)$  is the probability of there being  $i$  molecules after  $c$  cycles,  $h$  is the fluorescence threshold in relative fluorescence units, and  $x$  and  $y$  are the partition column and row, respectively.

The data (cf. Figure 26) includes  $\mathbf{n} = (n_0, n_1, n_2)$ , where  $n_0$  is the number of partitions for which the threshold is not reached (interpreted as no molecule),  $n_1$  is the number of low outliers plus the number of unreliable  $C_q$  values from the edges of the panel (interpreted as at least one molecule),  $n_2$  is the number of high outliers (interpreted as exactly one molecule). The remaining data comprise  $\mathbf{c} = (c_1, \dots, c_{n_3})$  the other  $C_q$  values along with  $\mathbf{x} = (x_1, \dots, x_{n_3})$  and  $\mathbf{y} = (y_1, \dots, y_{n_3})$ , which are the  $x$ - and  $y$ -locations of the associated partitions.

The distribution of  $p(i, c|\mu, \nu)$  for  $c = 0$  is given by the Conway-Maxwell Poisson distribution parametrised in terms of  $\mu$  and  $\nu$ , and truncated to  $i \leq m$  (to limit computation time). Thus,

$$p(i, c|\mu, \nu) = \frac{\text{CMP}(i; \lambda, \nu)}{\sum_{k=0}^m \text{CMP}(k; \lambda, \nu)}, \quad (74)$$

where  $\lambda$  is such that  $\sum_{k=0}^m k \text{CMP}(k; \lambda, \nu) = \mu$ . The probabilities for subsequent cycles up to  $c = c_0$  are based on molecules duplicating each cycle with probability  $E$ . We use  $c_0 = 6$  and  $m = 7$ .

We note that this problem does not fall into the category of a regression problem, but is in the form of a parametric inverse problem. However, in specifying the statistical model, i.e., the likelihood (72) together with (73), a Bayesian inference can be applied in just the same way as for the other case studies.

## 7.2 Prior distributions

We illustrate the selection of suitable families of prior distributions to reflect knowledge about the likely interval of values for each parameter as well as other available prior knowledge. In particular we introduce the Beta distribution as a flexible family of prior distributions for parameters that take values between zero and one. Similarly to the last example also information from previous measurements is utilized. The statistical model is given by expressions (72) and (73).

**Choice of family of priors.** There are six parameters for which we need priors: the mean initial number of molecules per well  $\mu$ , the dispersion parameter of the Conway-Maxwell distribution of the initial number of molecules  $\nu$ , the efficiency  $E$ , the fluorescence per molecule  $A$ , and the  $x$ -trend  $b_x$  and  $y$ -trend  $b_y$ . For the first three of these parameters we propose informative priors. To this end, we select for each of these parameters a suitable family of distributions, and select their parameters in accordance with our prior belief, including that based on information elicited from experts and previous measurements.

**Prior distribution for mean initial number of molecules per well  $\mu$ .** The range of  $\mu$  is  $(0, \infty)$  and so a natural choice of prior distribution is the Gamma distribution. For dPCR to give good results, it is necessary that  $p(0, 0|\mu, \nu)$  is not too close to zero or one. It may be that a preliminary qPCR analysis is used to estimate  $\mu$  so that after dilution the new  $\mu$  satisfies the required property. Even if such explicit prior information is not available, it may be reasonable prior belief that  $p(0, 0|\mu, \nu)$  is not too close to zero or one. Based on this information, we used a prior for  $\mu$  of  $\Gamma(3/2, 2/3)$ , which has a mean of 1 and 95 % of its mass is between 0.07 and 3.1.

**Prior distribution for dispersion parameter of Conway-Maxwell distribution of initial number of molecules  $\nu$ .** The range of  $\nu$  is  $(0, \infty)$  and so a natural choice of prior distribution is again the Gamma distribution. We chose the parameters of the prior to give a mean of one and a moderate variance, which reflects the default expectation that the initial

molecules follow a Poisson distribution. While we have no information about the possible dispersion, it seems unlikely that there is a physical factor that would lead to significant over- or under-dispersion. We chose  $\Gamma(10, 1/10)$  for which there is a 95 % mass between 0.48 and 1.7.

**Prior distribution for efficiency  $E$ .** The range of  $E$  is  $(0, 1)$  and so a natural choice of prior is the Beta distribution. Prior information about the efficiency  $E$  can be provided by preliminary qPCR experiments. However these are crude estimates for the qPCR efficiency. Furthermore, it is not known how the qPCR efficiency compares with the dPCR efficiency. Thus, priors based on such data ought to have high variance to take into account the uncertainties. For the present work we had access to standard curve data that provided estimates of  $E$ , based on the curves' gradients. These estimates of the efficiency are between 0.90 and 0.94. Taking this and the uncertainties mentioned above into account we use a prior of  $\text{Beta}(60, 5)$  which has a mean of 0.92 and has 95 % of its mass between 0.848 and 0.974.

**Prior distribution for first cycle efficiency  $E_0$ .** With no prior indication that the efficiency in cycle 1 is different to the later efficiency, the prior  $\text{Beta}(60, 5)$  is also used for  $E_0$ .

**Prior distribution for fluorescence per molecule  $A$ .** According to expert testimony, backed up by [53], it is extremely difficult to estimate  $A$  and so construct an informative prior. Instead we use the improper prior  $\pi(A) \propto A^{-1}$ .

**Prior distributions for  $x$ -trend  $b_x$  and  $y$ -trend  $b_y$ .** We do not yet understand the process that leads to the trends represented by  $b_x$  and  $b_y$ . Our only information about them comes from the data. Thus we cannot produce informative priors, and use an improper constant prior for both, i.e.,  $\pi(b_x) \propto 1$  and  $\pi(b_y) \propto 1$ .

Table 16 and Figure 27 show the selected priors. We refer to [61] for further information about this elicitation process.

$\mu$	$\nu$	$E$	$E_0$	$A$	$b_x$	$b_y$
$\Gamma(3/2, 2/3)$	$\Gamma(10, 1/10)$	$\text{Beta}(60, 5)$	$\text{Beta}(60, 5)$	$\pi(A) \propto A^{-1}$	$\pi(b_x) \propto 1$	$\pi(b_y) \propto 1$

Table 16: Prior distributions for the parameters of the digital PCR problem under model (72).

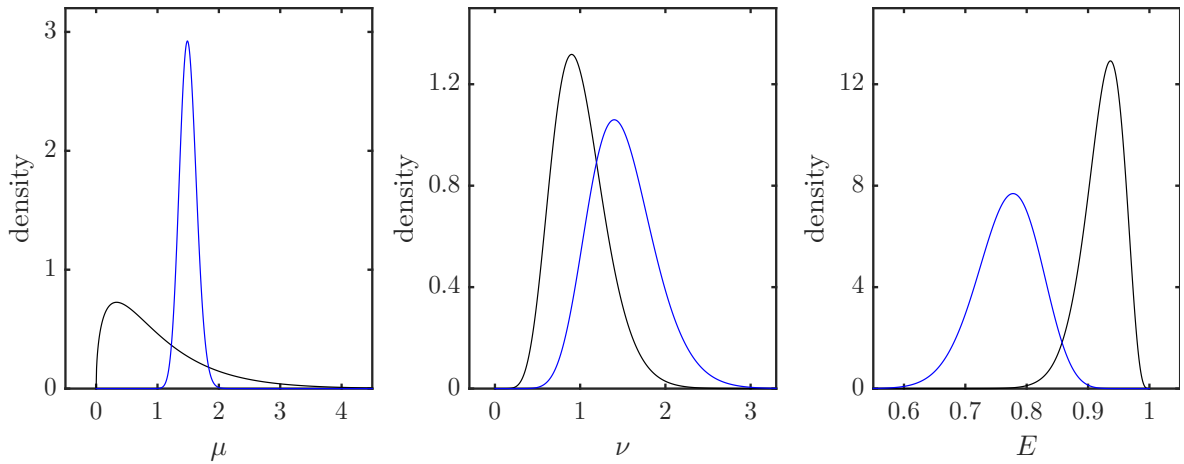


Figure 27: Visualization of the prior distributions (black curves) in Table 16 for the parameters  $\mu$ ,  $\nu$  and  $E$  of the digital PCR problem under model (72). The blue curves show the variation of the prior distributions used in the sensitivity analysis.

### 7.3 Numerical methods

This example illustrates the use of a random-walk Metropolis-Hastings algorithm in connection with a variable transformation to enhance convergence. It also provides a further convergence diagnostic in addition to those considered in the previous case studies. The algorithm is described in Appendix A.5. In using a chain length of  $N = 10^4$  it was applied to the data in Figure 26 yielding the results shown in Figure 28 and Table 17.

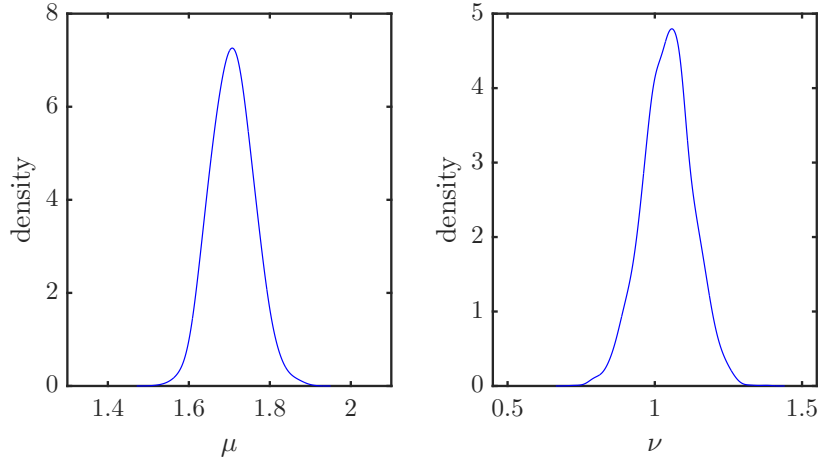


Figure 28: Results obtained by fitting model (72) and the priors in Table 16 to the data given in Figure 26. Displayed are the posterior distributions for the mean copy number  $\mu$  (left) and the dispersion parameter  $\nu$  (right).

	$\mu$	$\nu$	$E$	$E_0$	$A$	$b_x$	$b_y$
$\hat{\theta}$	1.707	1.043	0.930	0.774	$2.83 \cdot 10^{-9}$	-0.0069	-0.0014
std( $\theta$ )	0.051	0.084	0.004	0.020	$1.37 \cdot 10^{-10}$	0.0004	0.0028

Table 17: Results for the digital PCR problem obtained by fitting model (72) and the priors in Table 16 to the data given in Figure 26. The table shows means and standard deviations of the MCMCs for the single parameters.

The transformation used in this study to enhance convergence has 2 parts. The first part reduces the nonlinear relationship between the parameters  $(A, E)$  by transforming  $A$  to  $\log(A)/\log(1 + E)$ . This should improve convergence of the MCMC process. The second part is to transform  $(\mu, E, E_0, \nu)$  so that all the transformed variables take values in the range  $(-\infty, \infty)$ . This makes the optimisation stage more robust. The MCMC process produces a sample in terms of the transformed variables, from which a sample in terms of the original variables can be calculated by reversing the transformation (cf. Appendix A.5). The algorithm uses an estimate of the covariance of the transformed variables at the posterior density mode in the proposal density to achieve good convergence despite the remaining correlation.

Figure 28 illustrates MCMC results for the  $C_q$  data from Figure 26. Plots of MCMC outputs are useful diagnostics. Figure 29 represents a successful MCMC run for which there is good convergence to a unimodal distribution.

The choice to only run a single MCMC chain means that some diagnostics such as Gelman and Rubins diagnostic are inappropriate. Instead the Geweke diagnostic [26] can be used. This is a 2-sample  $t$ -test for a statistical difference between the means from 2 distinct parts of the chain, and uses spectral densities to estimate the variances. A significant difference

indicates that the chain is not converged and that either a longer chain is required, or an initial burn-in section should be removed.

The  $p$ -values using the usual options of the first 10% and last 50% of the chain are shown in Table 18. These suggest that the MCMC has converged for all parameters.

	$\mu$	$\nu$	$E$	$E_0$	$A$	$b_x$	$b_y$
$p$ -value	0.39	0.43	0.79	0.43	0.24	0.79	0.54

Table 18: Convergence analysis in form of  $p$ -values for the Geweke diagnostic [26] for each parameter. The  $p$ -values suggest that convergence has been reached for all parameters.

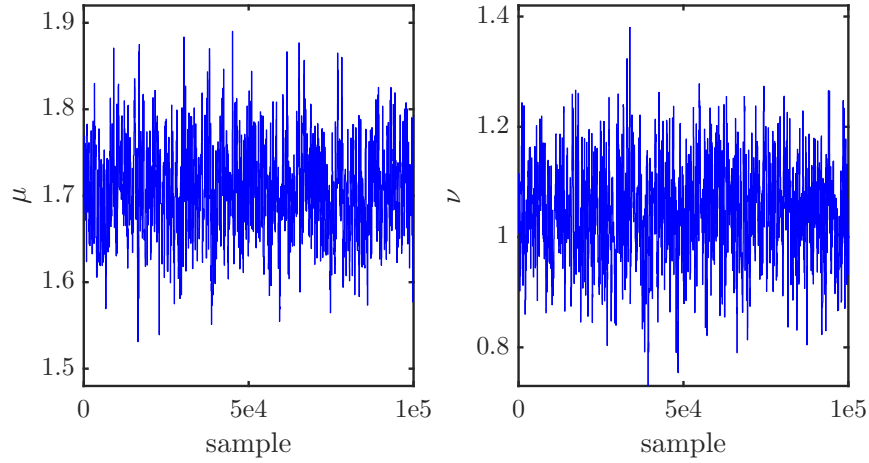


Figure 29: MCMC runs for mean copy number  $\mu$  (left) and the dispersion parameter  $\nu$  (right).

## 7.4 Sensitivity analysis

In this example the sensitivity analysis is performed by varying the priors of individual variables in turn. This enables conclusions to be drawn about the individual impact of the priors for those variables.

The default prior distributions are  $\mu \sim \Gamma(3/2, 2/3)$ ,  $\pi(A) \propto A^{-1}$ ,  $E \sim \text{Beta}(60, 5)$  and  $E_0 \sim \text{Beta}(60, 5)$ ,  $\pi(b_x) \propto 1$ ,  $\pi(b_y) \propto 1$ ,  $\nu \sim \Gamma(10, 1/10)$ . Table 19 shows the means and standard deviations of the univariate posteriors for each parameter, where the prior distribution that has been changed from its default is given in the headings.

The consistent message of these results is that the analysis is robust to changes in the prior distributions as all the parameter estimates are very close. This is not surprising given the standard deviations of the parameters in Table 17.

	NA	$\mu \sim \Gamma(120, 1/80)$	$\nu \sim \Gamma(15, 1/10)$	$E \sim \text{Beta}(50, 15)$
$\mu$	1.707	1.683	1.704	1.724
$\nu$	1.043	1.053	1.049	1.025
$E$	0.930	0.930	0.927	0.930
$E_0$	0.774	0.777	0.776	0.749
$A$	$2.83 \cdot 10^{-9}$	$2.83 \cdot 10^{-9}$	$2.93 \cdot 10^{-9}$	$2.81 \cdot 10^{-9}$
$b_x$	-0.0069	-0.0069	-0.0069	-0.0069
$b_y$	-0.0014	-0.0013	-0.0013	-0.0014

	$E_0 \sim \text{Beta}(50, 15)$	$A \sim \Gamma(10^2, 10^{-11})$	$b_x \sim N(0, 10^{-6})$	$b_y \sim N(0, 10^{-6})$
$\mu$	1.714	1.707	1.705	1.703
$\nu$	1.027	1.046	1.054	1.082
$E$	0.946	0.929	0.930	0.930
$E_0$	0.760	0.776	0.775	0.775
$A$	$2.35 \cdot 10^{-9}$	$2.86 \cdot 10^{-9}$	$2.82 \cdot 10^{-9}$	$2.83 \cdot 10^{-9}$
$b_x$	-0.0069	-0.0059	-0.0069	-0.0069
$b_y$	-0.0024	-0.0025	-0.0002	-0.0013

Table 19: Parameter means under the change of a single univariate prior for the purpose of sensitivity analysis. The column “NA” indicates the default prior, and in each of the subsequent columns the prior of a single parameter is altered as indicated in the heading.

## 8 Acknowledgements

Part of this work has been funded by the European Metrology Research Programme (EMRP) project NEW 04 “Novel mathematical and statistical approaches to uncertainty evaluation”. The EMRP is jointly funded by the EMRP participating countries within EURAMET (European Association of National Metrology Institutes) and the European Union.

## References

- [1] A. Allard, N. Fischer, G. Ebrard, B. Hay, P. M. Harris, L. Wright, D. Rochais, and J. Mattout. A multi-thermogram based bayesian model for the determination of the thermal diffusivity of a material. *submitted to Metrologia*, 2015.
- [2] R. C. Baker. *Flow measurement handbook: Industrial designs, operating principles, performance, and applications*. Cambridge: Cambridge University Press, 1st edition, 2000.
- [3] M. J. Bayarri and M. E. Castellanos. Bayesian Checking of the Second Levels of Hierarchical Models. *Statist. Sci.*, 22:322–343, 2007.
- [4] J. O. Berger. *Statistical decision theory and Bayesian Analysis*. Springer Series in Statistics. New York: Springer, 2nd edition, 1985.
- [5] J. O. Berger. The case for objective Bayesian analysis. *Bayesian Analysis*, 1:385–402, 2006.
- [6] J. O. Berger and J. M. Bernardo. On the development of reference priors. In J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 4, pages 35–60. Oxford: Oxford University Press, 1992.
- [7] J. O. Berger, L. R. Pericchi, J.K. Ghosh, T. Samanta, F. De Santis, J. O. Berger, and L. R. Pericchi. *Objective Bayesian methods for model selection: Introduction and comparison*, volume 38 of *Lecture Notes–Monograph Series*, pages 135–207. Beachwood, OH: Institute of Mathematical Statistics, Beachwood, OH, 2001.
- [8] W. Bich et al. Revision of the ‘guide to the expression of uncertainty in measurement’. *Metrologia*, 49:702–705, 2012.
- [9] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. *Evaluation of measurement data - Guide to the expression of uncertainty in measurement*. Joint Committee for Guides in Metrology, JCGM 100, 2008.
- [10] O. Bodnar and C. Elster. Analytical derivation of the reference prior by sequential maximization of Shannon’s mutual information in the multi-parameter case. *Journal of Statistical Planning and Inference*, 147:106–116, 2014.
- [11] G. E. P. Box and G. C. Tiao. *Bayesian Inferences in statistical analysis*. New York: John Wiley & Sons, 1st edition, 1992.
- [12] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455, 1998.
- [13] P. J. Brown. *Measurement, regression and calibration*, volume 12 of *Oxford Statistical Science Series*. Oxford: Oxford University Press, 1993.
- [14] G. S. Datta and T. J. Sweeting. *Probability matching priors*, volume 25 (Bayesian Thinking: Modeling and Computation) of *Handbook of Statistics*. Amsterdam: Elsevier B.V., 2005.
- [15] C. Davies. *Introduction to immunoassay principles*, in *The Immunoassay Handbook*. Oxford: Elsevier Ltd., 3rd edition, 2005.
- [16] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*, volume 57. CRC press, 1994.



- [17] C. Elster. Bayesian uncertainty analysis compared with the application of the GUM and its supplements. *Metrologia*, 51:S159–S166, 2013.
- [18] EMRP project NEW04 “Novel mathematical and statistical approaches to uncertainty evaluation”, 1 August 2012 - 31 July 2015, <http://www.ptb.de/emrp/new04-home.html> (accessed on 16 February 2015).
- [19] W. A. Fuller. *Measurement error models*, volume 305. John Wiley & Sons, 2009.
- [20] P. H. Garthwaite, J. B. Kadane, and A. O’Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100:680–701, 2005.
- [21] A. Gelman, J. B. Carlin, H. Stern, and D. B. Rubin. *Bayesian data analysis*. Texts in Statistical Science. London: Chapman & Hall/CRC, 2nd edition, 2003.
- [22] A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, pages 1–20, 2013.
- [23] A. Gelman, X. L. Meng, and H. Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6:733–760, 1996.
- [24] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.
- [25] E. I. George and R. E. McCulloch. Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373, 1997.
- [26] J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculating posterior moments. In J.M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *In Bayesian Statistics*, volume 4, pages 169–193. Oxford: Oxford University Press, 1992.
- [27] C. J. Geyer. Practical markov chain monte carlo. *Statistical Science*, 7:473–483, 1992.
- [28] W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in practice*. Interdisciplinary Statistics. London: Chapman & Hall/CRC, 1st edition, 1996.
- [29] M. Goldstein. Subjective Bayesian analysis: Principles and practice. *Bayesian Analysis*, 1:403–420, 2006.
- [30] J. P. Hobert and G. Casella. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91:1461–1473, 1996.
- [31] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14:382–401, 1999.
- [32] MATLAB® 2012 The MathWorks Inc., 2012.
- [33] H. Ishwaran and J. S. Rao. Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, 33:730–773, 2005.
- [34] ISO 7066-2:1988, Assessment of uncertainty in the calibration and use of flow measurement devices – Part 2: Non-linear calibration relationships.
- [35] H. Jeffreys. *Theory of probability*. International series of monographs on physics. Oxford: Clarendon Press, 3rd edition, 1967.

- [36] R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91:1343–1370, 1996.
- [37] K. Klauenberg, B. Ebert, J. Voigt, M. Walzel, J. E. Noble, A. E. Knight, and C. Elster. Bayesian analysis of an international ELISA comparability study. *Clinical Chemistry and Laboratory Medicine*, 49:1459–1468, 2011.
- [38] K. Klauenberg and C. Elster. The multivariate normal mean - sensitivity of its objective Bayesian estimates. *Metrologia*, 49:395–400, 2012.
- [39] K. Klauenberg, M. Walzel, B. Ebert, and C. Elster. Informative prior distributions for ELISA analyses. *Biostatistics*, online first, 2015.
- [40] G. J. P. Kok, A. M. H. van der Veen, P. M. Harris, I. M. Smith, and C. Elster. Bayesian analysis of a flow meter calibration problem. *Metrologia*, *in press*, 2015.
- [41] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger. Mixtures of  $g$  priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103:410–423, 2008.
- [42] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28:3049–3067, 2009.
- [43] A. D. Martin, K. M. Quinn, and J. H. Park. MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*, 42:1–21, 2011.
- [44] A. O’Hagan. Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (Statistician)*, 47:21–35, 1998.
- [45] A. O’Hagan and J. Forster. *Kendalls Advanced Theory of Statistics: Bayesian Inference*, volume 2B. Oxford: Oxford University Press, 2nd edition, 2004.
- [46] W. J. Parker, W. J. Jenkins, C. P. Butler, and G. L. Abbott. Flash method of determining thermal diffusivity, heat capacity and thermal conductivity. *Journal of Applied Physics*, 32:1679–1684, 1961.
- [47] F. Penneccchi et al. Software for calibration problems. *Report*, 2014.
- [48] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008.
- [49] C. P. Robert. *The Bayesian choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. New York: Springer, 2007.
- [50] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. New York: Springer, 2nd edition, 2004.
- [51] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7:110–120, 1997.
- [52] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:319–392, 2009.
- [53] J. M. Ruijter, M. W. Pfaff, S. Zhao, A. N. Spiess, G. Boggy, J. Blom, R.G. Rutledge, D. Sisti, A. Lievens, K. De Preter, S. Derveaux, J. Hellemans, and J. Vandesompele. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. *Methods*, 59:32–46, 2013.

- [54] G. A. F. Seber and C. J. Wild. *Nonlinear regression, 1989*. Wiley, New York, 1989.
- [55] J. V. Stone. *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis*. Sebtel Press, 1st edition, 2013.
- [56] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81:82–86, 1986.
- [57] A. W. Van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistic and Probabilistic Mathematics. Cambridge: Cambridge University Press, 1998.
- [58] A. Vehtari and J. Lampinen. Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation*, 14:2439–2468, 2002.
- [59] B. Vogelstein and K. W. Kinzler. Digital PCR. *Proceedings of the National Academy of Sciences of the United States of America*, 96:9236–9241, 1999.
- [60] J. Wang and N. Zabaras. Hierarchical Bayesian models for inverse problems in heat conduction. *Inverse Problems*, 21:183–206, 2005.
- [61] P. Wilson et al. Modelling quantification cycle data for digital pcr analysis. *To be submitted to BMC Bioinformatics*, 2015.

## A Algorithms and source code

This appendix lists algorithms and explicit software source code for the case studies.

### A.1 MATLAB<sup>®</sup> software for *straight line fit*

MATLAB<sup>®</sup> source code is provided for the Normal linear regression model (cf. Section 3.3). In the main script `Normlinreg_Analyse_Linear` the data as well as the parameter settings for the NIG prior are specified. The corresponding numerical values can be edited to account for different data or different NIG prior distributions. A couple of supporting functions are utilized within the main script:

- The function `normlinreg_define_design_matrix` is used to specify the design matrix  $\mathbf{X}$ . By editing this part of the code the software can easily be adapted to other Normal linear regression problems.
- The function `normlinreg_conj_prior` evaluates the parameters of the NIG posterior as well as the covariance matrix of the model parameters.
- The function `normlinreg_noninf_prior` determines the parameters of the posterior distribution resulting for the noninformative prior.
- A visualization of the marginal posterior distributions is provided by the function `normlinreg_plot_densities`.
- A plot of the resulting model functions including their 95% credible intervals is generated by the function `normlinreg_plot_nig_model`. This function provides in addition a plot of the residuals for the conjugate analysis.
- The function `normlinreg_plot_nig_prior` creates a plot of the NIG prior. This can be helpful during the elicitation of the parameters of the NIG prior.

The plotting routines make use of the functions `gampdf` and `tinv` from the MATLAB<sup>®</sup> Statistics toolbox in order to evaluate the Gamma distribution and to derive quantiles of the t-distribution.

#### Source code

##### Linear analysis

```

1 % -----
2 % NormLinRegAnalyse_Linear.m
3 % -----
4
5 % -----
6 % Data
7 % -----
8 y = [0.11  0.40  0.26  0.45  0.78  0.74  0.70  0.77]';
9
10 % -----
11 % Design matrix for the linear model
12 % -----
13 [X,x,np] = normlinreg_define_design_matrix;
14
15 % -----

```

```

16 % Settings of conjugate prior
17 % -----
18 theta0 = [0.0,1.0]';
19 V0      = 4*eye(np);
20 alp0    = 0.4;
21 bet0    = 0.004;
22
23 % -----
24 % Conjugate NIG analysis
25 % -----
26 [theta,V,alp,bet,cov_theta] = ...
27                               normlinreg_conj_prior(X,y,theta0,V0,alp0,bet0);
28
29 % -----
30 % Inference using non-informative prior
31 % -----
32 [theta_ni,cov_theta_ni,nu_ni,sig2_ni] = normlinreg_noninf_prior(X,y);
33
34 % -----
35 % Output: graphs
36 % -----
37
38 %
39 % Plot NIG Prior
40 %
41 figure(1);
42 normlinreg_plot_nig_prior(x,X,y,theta0,V0,alp0,bet0);
43 %
44 % Plot Model Fit
45 %
46 figure(2);
47 normlinreg_plot_nig_model(x,X,y,theta,V,alp,bet,...
48                           theta_ni,cov_theta_ni,nu_ni)
49 %
50 % Density Plots
51 %
52 figure(3);
53 normlinreg_plot_densities(theta,V,alp,bet,...
54                           theta0,V0,alp0,bet0,theta_ni,...
55                           cov_theta_ni,nu_ni,sig2_ni)
56 %
57 % -----
58 % End.
59 %
60 % No warranty or guarantee applies to this software, and therefore any
61 % users should satisfy themselves that it meets their requirements.
62 % -----

```

### Example of defined design matrix

```

1 function [X,x,np] = normlinreg_define_design_matrix
2 % -----
3 % Example design matrix used for the straight line model
4 % Please edit to define the linear model
5 % -----
6 %
7 % x values
8 x = [0.10  0.21  0.33  0.44  0.56  0.67  0.79  0.90].';
9 %
10 % Design matrix
11 %
12 np = 2; % number of unknown parameters
13 X = zeros(numel(x),np);
14 X(:,1) = ones(size(x));
15 X(:,2) = x;

```

```

16 end
17 %
18 % -----
19 % End.
20 %
21 % No warranty or guarantee applies to this software, and therefore any
22 % users should satisfy themselves that it meets their requirements.
23 % -----

```

---

### Bayesian Normal linear regression using conjugate NIG prior

```

1 % -----
2 function [theta,V,alp,bet,cov_theta] = normlinreg_conj_prior(X,y,...
3                                     theta0,V0,alp0,bet0)
4 % -----
5 % Bayesian normal linear regression using conjugate NIG prior
6 % -----
7
8 ID    = eye(numel(X(1,:)));
9 XTX   = X.'*X;
10 V0i   = V0\ID;
11 V      = (V0i+XTX)\ID;
12 theta = V*(V0i*theta0+X.'*y);
13 alp   = alp0+numel(y)/2;
14 bet    = bet0+(theta0.'*V0i*theta0+y.'*y-theta.'*(V0i+XTX)*theta)/2;
15 cov_theta = (V*bet/alp)*(alp/(alp-1));
16 end
17 %
18 % -----
19 % End.
20 %
21 % No warranty or guarantee applies to this software, and therefore any
22 % users should satisfy themselves that it meets their requirements.
23 % -----

```

---

### Bayesian Normal linear regression using noninformative prior

```

1 function [theta,cov_theta,nu,sig2] = normlinreg_noninf_prior(X,y)
2 % -----
3 % Bayesian normal linear regression using noninformative 1/sig^2 prior
4 % -----
5
6 np    = numel(X(1,:));
7 nu    = numel(y)-np;
8 ID    = eye(np);
9 XTX   = X.'*X;
10 theta = X\y;
11 sig2  = (y-X*theta).'(y-X*theta)/nu;
12 cov_theta = nu*sig2*(XTX\ID)/(nu-2);
13 end
14 %
15 % -----
16 % End.
17 %
18 % No warranty or guarantee applies to this software, and therefore any
19 % users should satisfy themselves that it meets their requirements.
20 % -----

```

---

### Plot of densities

```

1 function normlinreg_plot_densities(theta,V,alp,bet,theta0,V0,alp0,bet0,...
2                                     theta_ni,cov_theta_ni,nu_ni,sig2_ni)
3 % -----
4 % Density plots

```

```

5 % -----
6
7 np = numel(theta);
8 for k = 1:np
9     subplot(1,np+1,k);
10    sx = sqrt(V(k,k)*bet/alp);
11    xx = linspace(theta(k)-4*sx,theta(k)+4*sx,200);
12    plot(xx,tpdf((xx-theta(k))/sx,2*alp)/sx,'-k');
13    hold on
14    sx0 = sqrt(V0(k,k)*bet0/alp0);
15    plot(xx,tpdf((xx-theta0(k))/sx0,2*alp0)/sx0,'-r');
16    sx_ni = sqrt(cov_theta_ni(k,k)*(nu_ni-2)/nu_ni);
17    plot(xx,tpdf((xx-theta_ni(k))/sx_ni,nu_ni)/sx_ni,'-g');
18    hold off; axis tight;
19    xlabel(['\theta_',num2str(k)]); ylabel('density');
20 end
21 %
22 subplot(1,np+1,np+1);
23 sig=linspace(0,10*sqrt(bet0/(alp0+1)),200);
24 invgamma = @(x,a,b)(gampdf(1./x,a,1/b)./(x.^2));
25 plot(sig,2*sig.*invgamma(sig.^2,alp,bet),'-k');
26 hold on;
27 plot(sig,2*sig.*invgamma(sig.^2,nu_ni/2,nu_ni*sig2_ni/2),'-g');
28 plot(sig,2*sig.*invgamma(sig.^2,alp0,bet0),'-r');
29 hold off;
30 legend('NIG','Noninf','Prior');
31 xlabel('\sigma'); ylabel('density'); axis tight;
32 end
33 %
34 % -----
35 % End.
36 %
37 % No warranty or guarantee applies to this software, and therefore any
38 % users should satisfy themselves that it meets their requirements.
39 % -----

```

### Plot of NIG model

```

1 function normlinreg-plot-nig-model(x,X,y,theta,V,alp,bet,theta_ni,...
2                                     cov_theta_ni,nu_ni)
3 % -----
4 % Plot Model fit
5 % -----
6
7 prop = 0.95;
8 post_int = sqrt(bet*dot(X',V*X')'/alp)*tinv(0.5+prop/2,2*alp);
9 %
10 subplot(1,2,1);
11 od = plot(x,y,'ok'); xlabel('x'); ylabel('y');
12 hold on;
13 onig = plot(x,X*theta,'-k');
14 plot(x,X*theta+post_int,'-k');
15 plot(x,X*theta-post_int,'-k');
16 oni = plot(x,X*theta_ni,'-r');
17 post_int_ni = sqrt(dot(X',cov_theta_ni*X')'*...
18                     (nu_ni-2)/nu_ni)*tinv(0.5+prop/2,nu_ni);
19 plot(x,X*theta_ni+post_int_ni,'-r');
20 plot(x,X*theta_ni-post_int_ni,'-r');
21 legend([od,onig,oni],'Data','NIG','Noninf. ');
22 hold off; axis tight;
23 %
24 subplot(1,2,2);
25 invgamma = @(x,a,b)(gampdf(1./x,a,1/b)./(x.^2));
26 sig_hat = quadgk(@(s)(2*s.*s.*invgamma(s.^2,alp,bet)),0,inf);
27 %sig_hat=sqrt(bet/(alp-1));

```

```

28     plot(x,(y-X*theta)/sig_hat , 'ok');
29     hold on;
30     plot(xlim(),[0 0], '-k');
31     hold off; axis tight;
32     xlabel('x'); ylabel('normalized residuals');
33 end
34 %
35 % -----
36 % End.
37 %
38 % No warranty or guarantee applies to this software, and therefore any
39 % users should satisfy themselves that it meets their requirements.
40 % -----

```

---

### Plot of NIG prior

```

1 function normlinreg_plot_nig_prior(x,X,y,theta0,V0,alp0,bet0)
2 % -----
3 % Plot NIG prior
4 % -----
5
6     sig = linspace(0,10*sqrt(bet0/(alp0+1)),200);
7     invgamma = @(x,a,b)(gampdf(1./x,a,1/b)./(x.^2));
8     prop = 0.95;
9     subplot(1,2,1);
10    plot(sig,2*sig.*invgamma(sig.^2,alp0,bet0), '-k');
11    xlabel('\sigma'); ylabel('\pi(\sigma)'); axis tight;
12 %
13    subplot(1,2,2);
14    prior_int=sqrt(bet0*dot(X',V0*X')'/alp0)*tinv(0.5+prop/2,2*alp0);
15    plot(x,y, 'ok'); xlabel('x'); ylabel('y');
16    hold on;
17    plot(x,X*theta0, '-k');
18    plot(x,X*theta0+prior_int, '—k');
19    plot(x,X*theta0-prior_int, '—k');
20    hold off; axis tight;
21 end
22 %
23 % -----
24 % End.
25 %
26 % No warranty or guarantee applies to this software, and therefore any
27 % users should satisfy themselves that it meets their requirements.
28 % -----

```

## A.2 MATLAB<sup>®</sup> software for *flow meter calibration*

MATLAB<sup>®</sup> source code is provided for the flow meter calibration problem (cf. Section 4.3). The software provides results in the form of:

- Estimates and standard uncertainties for the K-factor at three flow rates, viz., the minimum flow rate  $q_{\min}$ , the maximum flow rate  $q_{\max}$  and their midpoint  $q_{\text{mid}}$ , calculated as the average and standard deviation, respectively, of the values  $\{\psi^\top(q)\theta_{(r)}\}$  evaluated at those flow rates;
- A graph showing the prior and posterior distributions for the repeatability standard deviation  $\sigma$ ;
- A graph showing the calibration data, the lower and upper bounds on the values of a calibration curve, the estimated calibration curve  $f_{\hat{\theta}}(q)$  with  $\hat{\theta}$  calculated as the average



of the samples  $\{\theta_{(r)}\}$  from the posterior, and probabilistically-symmetric point-wise credible intervals for the values of the calibration curve corresponding to probability  $P = 0.95$ ;

- A graph showing the standardized residuals  $(k_i - f_{\hat{\theta}}(q_i))/\hat{\sigma}$  with  $\hat{\sigma}$  calculated as the average of the samples  $\{\sigma_{(r)}\}$  from the posterior;
- A graph showing the posterior probability distribution for the K-factor provided by the calibration curve at the minimum flow rate  $q_{\min}$ .

The code includes the main script `Flow_Calibration` and supporting functions

- `flow_design_matrices`,
- `flow_constraint`, and
- `invgamma`.

The code also calls the functions `gamrnd` for drawing random samples from a Gamma distribution, and `mvnrnd` for drawing random vectors from a multivariate Normal distribution, both provided in the MATLAB<sup>®</sup> Statistics toolbox. The user may edit the main script `Flow_Calibration` to modify the form of the calibration curve, in terms of the number and values of the exponents used to define the curve, and the nature of the prior knowledge, in terms of values for  $k_{\text{spec}}$ ,  $\sigma_{0,\text{rel}}$ ,  $\nu_0$  and  $\delta$ . The constraints on the calibration curve are implemented in the function `flow_constraint`, and can also be modified by the user. The data used to provide the results given in Section 4 are read from the files `FlowData_old.txt` and `FlowData_new.txt`, and are described in Appendix B.2.

### Script for flow meter calibration problem

```

1 % -----
2 % Flow_Calibration.m
3 %
4 % Script for applying Bayesian analysis to flow meter calibration problem.
5 % -----
6
7 % -----
8 % Calibration data and prior knowledge.
9 % -----
10 % Data for defining prior knowledge:
11 %   q0 (flow rates) and k0 (measured K-factors).
12 %
13 D = load('FlowData_old.txt');
14 q0 = D(:,1);
15 k0 = D(:,2);
16 %
17 % Data for calibration:
18 %   q (flow rates) and k (measured K-factors).
19 %
20 D = load('FlowData_new.txt');
21 q = D(:,1);
22 k = D(:,2);
23 n = length(q);
24 %
25 % Flow rates for evaluation of results:
26 %   qe (flow rates).
27 %
28 qe = [min(q); (min(q) + max(q))/2; max(q)];

```

```

29 %
30 % Flow rates for evaluation of calibration curves:
31 %     qv (flow rates).
32 %
33     qv = linspace(min(q), max(q), 101)';
34     nv = length(qv);
35 %
36 % Exponents (of q) defining calibration curve.
37 %
38     r = [0, -1, 1, 2, 3];
39     p = length(r);
40 %
41 % Design matrices calculated in terms of normalized flow rates.
42 %
43     [X0, X, Xe, Xv] = flow_design_matrices(r, q0, q, qe, qv);
44 %
45 % Specification of prior knowledge.
46 %
47     kspec = 13.163;           % Manufacturer's specification of K-factor
48     s0rel = 0.025/100;        % Manufacturer's specification of repeatability
49     nu0 = 1;                  % Degree of belief in s0rel
50     theta0 = lscov(X0, k0);   % Parameters of previous calibration curve
51     delta = 0.075/100;        % Relative deviation from previous calibration
52 %
53 % Parameters of prior distribution for sigma^2.
54 %
55     alpha0 = nu0/2;
56     beta0 = (nu0*(s0rel*kspec)^2)/2;
57 %
58 % -----
59 % Calculation of the joint posterior distribution based on Monte Carlo
60 % sampling and rejection.
61 % -----
62 % Ordinary least-squares (OLS) analysis of calibration data.
63 %
64     [theta_OLS, ~, msr, V_OLS] = lscov(X, k);
65 %
66 % Parameters of "unconstrained" posterior NIG(theta1, V1, alpha1, beta1).
67 %
68     theta1 = theta_OLS;
69     V1 = V_OLS/msr;
70     alpha1 = alpha0 + (n - p)/2;
71     beta1 = beta0 + msr*(n - p)/2;
72 %
73 % Samples from posterior distribution.
74 %
75     N = 10^6;
76     sigma2s = zeros(1,N);
77     thetas = zeros(p,N);
78     in = zeros(1,N);
79     for r = 1:N
80 %
81 %         Monte Carlo sampling from "unconstrained" posterior.
82 %
83         sigma2s(r) = 1/gamrnd(alpha1, 1/beta1, 1);
84         thetas(:,r) = mvnrnd(theta1', sigma2s(r)*V1, 1)';
85 %
86 %         Decide whether constraint on values of calibration curve is satisfied.
87 %
88         in(r) = flow_constraint(thetas(:,r), theta0, kspec, delta, Xv);
89 %
90     end
91     sigma2s = sigma2s(logical(in));
92     thetas = thetas(:,logical(in));
93 %

```

```

94 % -----
95 % Output: numerical results and graphs.
96 % -----
97 % Estimates of K-factors and associated standard uncertainties at
98 % flow rates qe.
99 %
100 Ns = length(sigma2s);
101 ks = Xe*thetas;
102 ke = mean(ks,2);
103 uke = std(ks,0,2);
104 fprintf('Number of samples:\n')
105 fprintf('%g \n', Ns)
106 fprintf('Results:\n')
107 fprintf('%9.6f %9.6f %9.6f \n', ke)
108 fprintf('%9.6f %9.6f %9.6f \n', uke)
109 %
110 % Plot prior and posterior distributions for sigma.
111 %
112 sigma = linspace(0, 0.05*kspec/100, 501)';
113 psigma0 = 2*sigma.*invgamma(sigma.^2, alpha0, beta0);
114 psigma1 = 2*sigma.*invgamma(sigma.^2, alpha1, beta1);
115 figure
116 plot(100*sigma/kspec, psigma0, 'k-', 100*sigma/kspec, psigma1, 'b-')
117 xlim([0, 0.05])
118 xlabel('Relative repeatability \sigma/{\it k}_{\rm spec} (%)')
119 ylabel('Probability density \pi(\sigma) (L)')
120 legend('Prior', 'Posterior', 0)
121 %
122 % Plot calibration data, prior knowledge on values of calibration curve,
123 % estimate of calibration curve and point-wise credible intervals.
124 %
125 thetai = mean(thetas,2);
126 prob = 0.95;
127 ilow = floor(Ns*(1-prob)/2);
128 ihgh = ceil(Ns*(1+prob)/2);
129 kl = zeros(nv,1);
130 kh = zeros(nv,1);
131 for i = 1:nv
132     ki = sort(Xv(i,:)*thetas);
133     kl(i) = ki(ilow);
134     kh(i) = ki(ihgh);
135 end
136 figure, hold on
137 plot(q, k, 'kx')
138 plot(qv, Xv*theta0-delta*kspec, 'k—', qv, Xv*theta0+delta*kspec, 'k—')
139 plot(qv, Xv*thetae, 'b-', qv, kl, 'b—', qv, kh, 'b—')
140 plot(qv, Xv*theta0-delta*kspec, 'k—', qv, Xv*theta0+delta*kspec, 'k—')
141 xlabel('Flow rate {\it q} (L min^{-1})')
142 ylabel('K-factor {\it k} (L^{-1})')
143 %
144 % Plot standardized residuals.
145 %
146 sigmae = mean(sqrt(sigma2s));
147 figure, hold on
148 plot(q, (k - X*thetae)/sigmae, 'kx')
149 plot(xlim, [0, 0], 'k—')
150 xlabel('Flow rate {\it q} (L min^{-1})')
151 ylabel('Standardized residual (1)')
152 %
153 % Plot distribution for K-factor at minimum flow rate.
154 %
155 [f, kv] = hist(ks(1,:),75);
156 pkv = f/(Ns*(kv(2)-kv(1)));
157 figure
158 bar(kv, pkv, 1)

```

```

159     h = findobj(gca, 'Type', 'patch');
160     set(h, 'FaceColor', 'w', 'EdgeColor', 'k')
161     xlabel('K-factor {\it k} (L^{-1})')
162     ylabel('Probability density (L)')
163 %
164 % -----
165 % End.
166 %
167 % No warranty or guarantee applies to this software, and therefore any
168 % users should satisfy themselves that it meets their requirements.
169 % -----

```

---

### Evaluate design matrices

```

1 function [X0, X, Xe, Xv] = flow_design_matrices(r, q0, q, qe, qv)
2 % -----
3 % Evaluate design matrices for flow calibration problem.
4 % -----
5 % Maximum flow rate from calibration data for normalization.
6 %
7     qmax = max(q);
8 %
9 % Number of exponents in calibration curve.
10 %
11     p = length(r);
12 %
13 % Evaluate design matrices.
14 %
15     n0 = length(q0);
16     X0 = zeros(n0,p);
17     for i = 1:n0
18         X0(i,:) = (q0(i)/qmax).^r;
19     end
20     n = length(q);
21     X = zeros(n,p);
22     for i = 1:n
23         X(i,:) = (q(i)/qmax).^r;
24     end
25     ne = length(qe);
26     Xe = zeros(ne,p);
27     for i = 1:ne
28         Xe(i,:) = (qe(i)/qmax).^r;
29     end
30     nv = length(qv);
31     Xv = zeros(nv,p);
32     for i = 1:nv
33         Xv(i,:) = (qv(i)/qmax).^r;
34     end
35 %
36 % -----
37 % End.
38 %
39 % No warranty or guarantee applies to this software, and therefore any
40 % users should satisfy themselves that it meets their requirements.
41 % -----

```

---

### Constraint on values of a calibration curve

```

1 function in = flow_constraint(theta, theta0, kspec, delta, Xv)
2 % -----
3 % Decide whether constraint on values of calibration curve is satisfied.
4 % -----
5     d = max(abs(Xv*theta - Xv*theta0));
6     in = (d < delta*kspec);

```

```

7 %
8 % -----
9 % End.
10 %
11 % No warranty or guarantee applies to this software, and therefore any
12 % users should satisfy themselves that it meets their requirements.
13 % -----

```

---

### Evaluate probability density for inverse Gamma distribution

```

1 function p = invgamma(theta, a, b)
2 % -----
3 % Evaluate PDF for IG(a, b).
4 % -----
5 p = (b^a)*(theta.^(-(a+1))).*exp(-b./theta)/gamma(a);
6 %
7 % -----
8 % End.
9 %
10 % No warranty or guarantee applies to this software, and therefore any
11 % users should satisfy themselves that it meets their requirements.
12 % -----

```

### A.3 Metropolis-Hastings algorithm for *inference of thermophysical properties*

A single-site Metropolis-Hastings algorithm is described for the inference of thermophysical properties (cf. Section 5.3). The term single-site refers to the use of a proposal distribution that changes just one of the unknowns. In order to ease notation we let  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top = (\tau, b, \sigma^2, \mu_\tau, \varphi_\tau, \mu_b, \varphi_b)^\top$  denote all unknown parameters (cf. expressions (52), (54) and (55)),  $N$  the length of the Markov chain,  $B$  the proportion of values discarded from the start of the chain ( $0 \leq B \leq 1$ ), and  $S$  a parameter to control the subsampling (or ‘thinning’) of the chain. Furthermore, let  $q_j$  be the proposal distribution for the parameter  $\theta_j$  given by

$$q_j \left( \theta_j | \theta_j^{(i-1)}, \boldsymbol{\theta}_{-j}^{(i)} \right) = \text{N} \left( \theta_j^{(i-1)}, \sigma_{\theta_j}^2 \right), \quad (75)$$

where

$$\boldsymbol{\theta}_{-j}^{(i)} = \left( \theta_1^{(i)}, \dots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \dots, \theta_p^{(i-1)} \right)^\top \quad (76)$$

denotes the set of parameters  $\boldsymbol{\theta}^{(i)}$  with the  $j^{th}$  component removed, and  $\sigma_{\theta_j}$ , the proposal standard deviation, is expressed as a percentage of the proposal mean  $\theta_j^{(i-1)}$ .

The algorithm is as follows [60]:

1. Initialize  $\boldsymbol{\theta}^{(1)}$ ;
2. For  $i = 2 : N$ ,
  - (a) For  $j = 1 : p$ ,
    - i. Sample  $u$  from  $\text{U}(0, 1)$ ;
    - ii. Sample  $\theta_j^{(*)}$  from  $q_j \left( \theta_j | \theta_j^{(i-1)}, \boldsymbol{\theta}_{-j}^{(i)} \right)$ ;
    - iii. If
$$u < \min \left\{ 1, \frac{p(\theta_j^{(*)} | \boldsymbol{\theta}_{-j}^{(i)}, \mathbf{y}) q_j(\theta_j^{(i-1)} | \theta_j^{(*)}, \boldsymbol{\theta}_{-j}^{(i)})}{p(\theta_j^{(i-1)} | \boldsymbol{\theta}_{-j}^{(i)}, \mathbf{y}) q_j(\theta_j^{(*)} | \theta_j^{(i-1)}, \boldsymbol{\theta}_{-j}^{(i)})} \right\}$$

then

$$\theta_j^{(i)} = \theta_j^{(*)}$$

else

$$\theta_j^{(i)} = \theta_j^{(i-1)};$$

3. For  $j = 1 : p$ ,

- (a) Discard the first  $B \times N$  values of the chain in order to leave only those values obtained after the Markov Chain has reached convergence;
- (b) Obtain the final sample by selecting one value from every  $S$  samples in order to limit the correlation between successive samples of the Markov Chain.

The results in Section 5.3 were obtained using the following parameters in the above algorithm and (75):  $\sigma_{\theta_j}^2 = (0.05\theta_j)^2$ ,  $N = 10^4$ ,  $B = 0.2$  and  $S = 5$ . Initial values for the  $\theta_j$  were randomly drawn from the specified prior distribution.

#### A.4 WinBUGS software for *analysis of immunoassay data*

WinBUGS source code is provided for the immunoassay analysis (cf. Section 6.3). WinBUGS is the interactive Windows version of the BUGS program. Except for minor differences, the code supplied below will also run under OpenBUGS, the open source version of the BUGS package which is currently developed further and also runs for non-Windows operating systems. However, observe that the sampling options for the full conditional distributions are different in OpenBUGS [42, Section 4.2.3].

The below source code can be compiled in BUGS, e.g., together with the data listed subsequently. After initialization, the software then draws samples (so-called ‘updates’) for all quantities required. To obtain the numerical results in Section 6.3, 10 chains with 160,000 iterations each were computed. Only every tenth sample of the latter half of the chain was retained for inference in the case study. In particular, the quantities b1, b2, b3, b4 and a, c in the code refer to the calibration parameters, and the quantities X1 and X2 provide the predicted concentration of two additional measurements (cf. Table 14 for all).

##### BUGS source code

```

1  model{
2  ##### PREDICTION OF UNKNOWN CONCENTRATION #####
3
4  #flat non-negative prior on unknown concentration
5  X1~dunif(0,25)
6  X2~dunif(0,25)
7
8  #sample 1: loop over ym1 measurements
9  for (i in 1:ym1){
10   X01[Nasm1[i]]<-X1
11  #measurem. are normally distr. around transformed concentration
12   Ym1[Nasm1[i]] ~dnorm(Y01[Nasm1[i]],epsiP01[Nasm1[i]])
13  #transform concentrations
14   Y01[Nasm1[i]]<-cut.b1+(cut.b2-cut.b1)/
15   (1+pow(X01[Nasm1[i]]/cut.b3,cut.b4))
16  #error model
17   epsiP01[Nasm1[i]]<-1/(cut.c+cut.a*X01[Nasm1[i]])
18  }
19
20  #sample 2: loop over ym2 measurements

```

```

21   for (i in 1:ym12){
22     X02[Nasm2[i]]<-X2
23   #measur. are normally distr. around transformed concentration
24     Ym2[Nasm2[i]]~dnorm(Y02[Nasm2[i]],epsiP02[Nasm2[i]])
25   #transform concentrations
26     Y02[Nasm2[i]]<-cut.b1+(cut.b2-cut.b1)/
27                           (1+pow(X02[Nasm2[i]]/cut.b3,cut.b4))
28   #error model
29     epsiP02[Nasm2[i]]<-1/(cut.c+cut.a*X02[Nasm2[i]])
30   }
31
32   #stop feedback of parameters (between calibration + prediction)
33   cut.c<-cut(c)
34   cut.a<-cut(a)
35   cut.b1<-cut(b1)
36   cut.b2<-cut(b2)
37   cut.b3<-cut(b3)
38   cut.b4<-cut(b4)
39
40   ##### CALIBRATION #####
41
42   for (i in 1:y1){
43   #measur. are normally distr. around transformed concentration
44     Y[Nas[i]]~dnorm(f[i],tau[i])
45   #transform concentrations
46     f[i]<-b1+(b2-b1)/(1+pow((X[Nas[i]]/b3,b4))
47   #error model
48     tau[i]<-1/epsi[i]
49     epsi[i]<-(c+a*(X[Nas[i]]))
50   #monitor standardized residuals
51     WeightResSq[i]<-(f[i]-Y[Nas[i]])/sqrt(epsi[i])
52   }
53
54   #flat non-negative priors on regression parameters theta
55   b1~dunif(0,b1up)
56   b2~dunif(0,YMax)
57   b1up<-100*YMax
58
59   #informative priors on regression parameters theta
60   b3~dt(5.309738953,0.0002478888,3)I(0,5000)
61   b4~dt(1.415119428,13.734710000,3)I(0,10)
62
63   #flat non-negative priors on error parameters delta
64   #(scaled for better sampling)
65   a<-ascale*pow(YMax,2)
66   c<-cscale*pow(YMax,2)
67   ascale~dunif(0,ascmax)
68   cscale~dunif(0,cscmax)
69   #maximum variance in interval [0,YMax] is YMax^2/4
70   ascmax<-1/4/XMax
71   cscmax<-1
72
73   #maximum of intensities and concentrations
74   for (i in 1:y1){
75   #permute intensities and concentrations
76     Yh1lf[i]<-Y[Nas[i]]
77     Xh1lf[i]<-X[Nas[i]]
78   }
79   YMax<-ranked(Yh1lf[],1)
80   XMax<-ranked(Xh1lf[],1)
81 }
82
83 # No warranty or guarantee applies to this software, and therefore any
84 # users should satisfy themselves that it meets their requirements.

```

In addition, WinBUGS requires input data. Below a scaled version of the example data set (cf. Figure 19) is given in WinBUGS format.

```
list(X=c(50, 10, 3.33, 1, 0.333, 0.1, 0.0333, 0, 50, 10, 3.33, 1, 0.333, 0.1,
0.0333, 0, 50, 10, 3.33, 1, 0.333, 0.1, 0.0333, 0), y1=23,
Y=c(0.21049, 0.03894, 0.00802, 0.00260, 0.00156, 0.00123, 0.00118,
0.00110, 0.17479, 0.03579, 0.00847, 0.00310, 0.00152, 0.00128, 0.00113, 0.00119,
0.18438, 0.03066, 0.00790, 0.00279, 0.00166, 0.00122, 0.00120, 0.00107), yml1=3,
ym12=3, Ym1=c(0.00280, 0.00305, 0.00334), Ym2=c(0.00807, 0.00677, 0.01078),
Nas=c(2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
23, 24), Nasm1=c(1, 2, 3), Nasm2=c(1, 2, 3))
```

### A.5 Metropolis-Hastings algorithm for *analysis of digital PCR data*

For the analysis of the data in Figure 26 the posterior is highly peaked, and cannot be calculated in the tails due to the machine precision, and so to ensure efficient calculations we overcome these features by selecting a good starting point, together with an appropriate proposal density.

In order to find the universal maximum during the initial optimization stage, it is necessary to use a number of starting values for  $E$  and  $E_0$ . We use the 9 combinations of  $E$  and  $E_0$  where each equals either 0.5, 0.7 or 0.9.

We estimate  $\mu$  using the standard estimate, which is based on the assumption that the initial molecules are Poisson distributed, and so in addition to  $E$  and  $E_0$  we use the following initial values:

$$\begin{aligned}\mu &= -\log(n_0/n), \\ \nu &= 1, \\ A &= h(1 + E)^{-\bar{c}}, \\ b_x &= \hat{b}_x, \\ b_y &= \hat{b}_y,\end{aligned}\tag{77}$$

where  $\hat{b}_x$   $\hat{b}_y$  are estimates provided by robust linear regression using the R function *lmRob* and  $\bar{c}$  is the mean of  $c$  and  $n = n_0 + n_1 + n_2 + n_3$ .

The variables are further transformed in order to improve performance. The overall transformation  $f(\boldsymbol{\theta}) = \boldsymbol{\varphi}$  is from  $\boldsymbol{\theta} = (\mu, \nu, E, E_0, A, b_x, b_y)^T$  to  $\boldsymbol{\varphi} = (\mu', \nu', E', E'_0, A', b_x, b_y)^T$  given by

$$\begin{aligned}\mu' &= \log(\mu), \\ \nu' &= \log(\nu), \\ E' &= -\log(E^{-1} - 1), \\ E'_0 &= -\log(E_0^{-1} - 1), \\ A' &= \log(A)/\log(1 + E),\end{aligned}\tag{78}$$

with  $b_x$  and  $b_y$  unchanged.

The determinant for this transformation is  $\left| \frac{\partial f}{\partial \boldsymbol{\theta}} \right| = [\mu \nu E (1 - E) E_0 (1 - E_0) \log(1 + E) A]^{-1}$  and the posterior for the transformed parameters  $\boldsymbol{\varphi}$  is calculated as



$\pi(\boldsymbol{\varphi}|\mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{n}) = \pi(f^{-1}(\boldsymbol{\theta})|\mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{n}) / |\frac{\partial f}{\partial \boldsymbol{\theta}}|$ . The mode is found using a Nelder–Mead algorithm starting at these points. The covariance of the posterior at that point,  $\boldsymbol{\Sigma}_{mode}$ , is estimated via the Hessian matrix.

The algorithm for producing the MCMC chain is as follows in which  $N$  is the length of the chain:

1. For each combination of  $E$  and  $E_0$  taking values from 0.5, 0.7 and 0.9, calculate the initial parameter estimates  $\boldsymbol{\theta}_0$  using equations (77);
  - (a) Transform to  $\boldsymbol{\varphi}_0$  using equations (78);
  - (b) Estimate the local maximum  $\boldsymbol{\varphi}_{mode}$  using the Nelder–Mead algorithm starting at  $\boldsymbol{\varphi}_0$ ;
2. From the results of step 1 identify the overall mode  $\boldsymbol{\varphi}_{mode}$ ;
3. Estimate the Hessian matrix  $\mathbf{H}$  at  $\boldsymbol{\varphi}_{mode}$ , and use the matrix to estimate the covariance matrix  $\boldsymbol{\Sigma}_{mode} = \mathbf{H}^{-1}$ . (If  $\mathbf{H}$  is not invertible first reduce the elements on its diagonal so that it is.)
4. Set  $\boldsymbol{\varphi}^{(1)} = \boldsymbol{\varphi}_{mode}$ ;
5. For  $i = 2 : N$ 
  - (a) Sample  $u$  from  $U(0, 1)$ ;
  - (b) Sample  $\boldsymbol{\varphi}^*$  from  $N(\boldsymbol{\varphi}^{(i-1)}, k\boldsymbol{\Sigma}_{mode})$ ;
  - (c) If
 
$$u < \min \left\{ 1, \frac{\pi(\boldsymbol{\varphi}^*|\mathbf{c}, \mathbf{n})}{\pi(\boldsymbol{\varphi}^{(i-1)}|\mathbf{c}, \mathbf{n})} \right\}$$
 then
 
$$\boldsymbol{\varphi}^{(i)} = \boldsymbol{\varphi}^*$$
 else
 
$$\boldsymbol{\varphi}^{(i)} = \boldsymbol{\varphi}^{(i-1)};$$
6. Transform values  $\boldsymbol{\varphi}^{(1)}, \dots, \boldsymbol{\varphi}^{(N)}$  back to  $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$  using the inverse of equations (78);

Steps 3-5 are performed by the R function *MCMCmetrop1R* from the *MCMCpack* package [43], including the selection of  $k$  in step 5b).

## B Data for case studies

This appendix lists data for the case studies.

### B.1 Data for *straight line fit*

Table 20 provides data for the *straight line fit* example.

$x$	0.10	0.21	0.33	0.44	0.56	0.67	0.79	0.90
$y$	0.11	0.40	0.26	0.45	0.78	0.74	0.70	0.77

Table 20: Example data for the Normal linear regression problem.

## B.2 Data for *flow meter calibration*

Tables 21 and 22 contain data for the case study of flow meter calibration comprising, respectively, calibration data used to define the prior knowledge for the flow meter calibration problem and calibration data used as the basis for determining a new calibration curve. Each row of the tables corresponds to measurements made at nominally the same flow rate. The MATLAB<sup>®</sup> script `Flow_Calibration.m` listed in Appendix A.2 reads this data from, respectively, the files `FlowData_old.txt` and `FlowData_new.txt`. It is assumed that within those files the data is organised into two columns, with the first column containing the flow rates  $q_i$  reported in units of  $\text{L min}^{-1}$  and the second column containing the corresponding K-factors  $k_i$  in  $\text{L}^{-1}$ .

$q$	$k$	$q$	$k$	$q$	$k$	$q$	$k$	$q$	$k$
5229.3	13.160631	5220.7	13.159314	5223.9	13.162473	5220.0	13.160894	5224.5	13.161026
4754.7	13.161684	4756.0	13.158261	4755.9	13.158393	4758.7	13.163263	4759.9	13.159973
4327.5	13.157866	4321.6	13.161815	4322.4	13.159314	4326.1	13.161947	4316.9	13.162079
3989.2	13.160499	4025.8	13.159973	4016.1	13.161947	4002.6	13.161420	4021.1	13.160631
3789.5	13.157340	3738.7	13.159709	3761.3	13.160499	3770.2	13.160894	3770.1	13.161552
3332.9	13.159841	3324.6	13.156550	3319.2	13.159709	3329.0	13.156945	3323.8	13.159578
2854.9	13.156287	2873.5	13.154707	2863.9	13.156945	2873.3	13.155497	2864.0	13.156550
2248.3	13.156813	2252.8	13.156813	2239.1	13.156945	2239.3	13.155892	2243.7	13.156945
1823.7	13.159973	1824.3	13.155497	1819.7	13.155892	1828.6	13.157208	1819.7	13.156155
1468.3	13.158393	1472.6	13.160631	1477.3	13.159051	1472.8	13.158393	1468.4	13.157472
1016.9	13.157735	1016.9	13.157603	1012.5	13.155497	1017.1	13.155629	1017.0	13.155892
821.1	13.152338	825.7	13.153128	834.8	13.152338	834.8	13.152206	825.7	13.153128

Table 21: Calibration data used to define the prior knowledge for the flow meter calibration problem. The flow rate  $q$  is reported in units of  $\text{L min}^{-1}$  and the K-factor  $k$  in  $\text{L}^{-1}$ .

$q$	$k$	$q$	$k$	$q$	$k$	$q$	$k$	$q$	$k$
5257.9	13.157208	5248.6	13.157866	5252.3	13.159973	5247.6	13.160236	5252.7	13.158919
4805.5	13.159841	4814.0	13.161552	4810.7	13.158130	4814.1	13.161420	4809.8	13.160499
4473.6	13.157208	4467.5	13.161552	4472.3	13.161157	4477.3	13.159841	4476.5	13.162079
3779.4	13.161026	3774.8	13.161026	3788.6	13.160631	3774.6	13.161684	3784.1	13.160236
3383.3	13.158525	3374.3	13.158393	3387.7	13.159446	3396.8	13.159446	3396.5	13.160367
2818.3	13.157077	2840.7	13.158788	2822.7	13.157735	2823.3	13.154839	2845.4	13.158261
2257.2	13.157998	2266.5	13.156813	2257.0	13.159183	2266.1	13.158919	2266.4	13.157735
1810.4	13.157603	1805.7	13.158261	1805.5	13.159841	1805.5	13.159578		
1395.2	13.159578								
1417.7	13.162210	1417.6	13.162737	1413.2	13.161289	1417.6	13.162737	1418.0	13.159578
1048.4	13.163263	1048.5	13.162079	1048.6	13.160104	1048.4	13.162605	1048.4	13.162737
793.4	13.159314	793.4	13.158788	793.3	13.159841	793.4	13.158261	797.8	13.160499

Table 22: Calibration data for the flow meter calibration problem. The flow rate  $q$  is reported in units of  $\text{L min}^{-1}$  and the K-factor  $k$  in  $\text{L}^{-1}$ .

### B.3 Data for the *analysis of immunoassay data*

Table 23 provides data for the *analysis of immunoassay data* example.

$x$ in $\mu\text{g/L}$	50	10	3.33	1	0.333	0.1	0.0333	0
$y$ in a.u.	21049	3894	802	260	156	123	118	110
	17479	3579	847	310	152	128	113	119
	18438	3066	790	279	166	122	120	107

	Sample 1	Sample 2
$y$ in a.u.	280	807
	305	677
	334	1078

Table 23: Example data for immunoassay analyses. Calibration data (top) and measurements of the two unknown concentrations (bottom).