Supplementary details on Bayesian Calibration of Computer Models

Marc C. Kennedy and Anthony O'Hagan

University of Sheffield

Summary. We provide more detail of the mathematical development involved in the Bayesian calibration methods of Kennedy and O'Hagan (2000), and we present a further example using data from a simulated accident assessment exercise.

1. Introduction

We assume the reader is already familiar with the concepts and notation of Kennedy and O'Hagan (2000). We begin by deriving the posterior distribution of the calibration parameters in Section 2. In Section 3 we discuss the estimation of hyperparameters involved in the model. Section 4 deals with details on calibration, calibrated prediction and calibrated uncertainty analysis. In Section 5 we investigate the sensitivity of our results to some of the model assumptions. Computational issues are discussed in Section 6, and in Section 7 we describe an example.

2. Posterior distribution

Following on from Kennedy and O'Hagan (2000), Section 4.4, we derive the posterior distribution of the parameters $\boldsymbol{\theta}, \boldsymbol{\beta}$ and $\boldsymbol{\phi}$. The full data vector $\boldsymbol{d}$ is normally distributed given $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi})$, and this will yield the likelihood function. To express its mean vector and variance matrix we require some more notation.

We denote the set of points at which the code outputs $\boldsymbol{y}$ are available by $D_1 = \{(\boldsymbol{x}_1^*, \boldsymbol{t}_1), \ldots, (\boldsymbol{x}_N^*, \boldsymbol{t}_N)\}$. Similarly, we denote the set of points for the observations $\boldsymbol{z}$ of the real process by $D_2 = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. Augmenting each of these points by the calibration parameters $\boldsymbol{\theta}$, we define $D_2(\boldsymbol{\theta}) = \{(\boldsymbol{x}_1, \boldsymbol{\theta}), \ldots, (\boldsymbol{x}_n, \boldsymbol{\theta})\}$. If we now let $\boldsymbol{H}_1(D_1)$ denote the matrix with rows $\boldsymbol{h}_1(\boldsymbol{x}_1^*, \boldsymbol{t}_1)^T, \ldots, \boldsymbol{h}_1(\boldsymbol{x}_N^*, \boldsymbol{t}_N)^T$, the expectation of $\boldsymbol{y}$ is $\boldsymbol{H}_1(D_1)\boldsymbol{\beta}_1$. In analogous notation, the expectation of $\boldsymbol{z}$ is $\rho \boldsymbol{H}_1(D_2(\boldsymbol{\theta}))\boldsymbol{\beta}_1 + \boldsymbol{H}_2(D_2)\boldsymbol{\beta}_2$. Hence

$$E(\boldsymbol{d} \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \phi) = \boldsymbol{m}_d(\boldsymbol{\theta}) = \boldsymbol{H}(\boldsymbol{\theta})\boldsymbol{\beta},$$

where

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{H}_1(D_1) & \boldsymbol{0} \\ \rho \boldsymbol{H}_1(D_2(\boldsymbol{\theta})) & \boldsymbol{H}_2(D_2) \end{pmatrix}.$$

To express the variance matrix of $\boldsymbol{d}$, define $\boldsymbol{V}_1(D_1)$ to be the matrix with (j, j') element $c_1((\boldsymbol{x}_j^*, \boldsymbol{t}_j), (\boldsymbol{x}_{j'}^*, \boldsymbol{t}_{j'}))$, so that this is the variance matrix of $\boldsymbol{y}$. Define $\boldsymbol{V}_1(D_2(\boldsymbol{\theta}))$ and $\boldsymbol{V}_2(D_2)$ similarly, and let $\boldsymbol{C}_1(D_1, D_2(\boldsymbol{\theta}))$ be the matrix with (j, i) element $c_1((\boldsymbol{x}_j^*, \boldsymbol{t}_j), (\boldsymbol{x}_i, \boldsymbol{\theta}))$. Then

$$\mathrm{var}(\boldsymbol{d} \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi}) = \boldsymbol{V}_d(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{V}_1(D_1) & \rho \boldsymbol{C}_1(D_1, D_2(\boldsymbol{\theta}))^T \\ \rho \boldsymbol{C}_1(D_1, D_2(\boldsymbol{\theta})) & \lambda \boldsymbol{I}_n + \rho^2 \boldsymbol{V}_1(D_2(\boldsymbol{\theta})) + \boldsymbol{V}_2(D_2) \end{pmatrix}$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

With prior distribution

$$p(\boldsymbol{\theta}, \boldsymbol{\theta}, \boldsymbol{\phi}) \propto p(\boldsymbol{\theta}) p(\boldsymbol{\phi}) \tag{1}$$

we now obtain the full joint posterior distribution

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi} \mid \boldsymbol{d}) \quad \propto \quad & p(\boldsymbol{\theta}) p(\boldsymbol{\phi}) |\boldsymbol{V}_d(\boldsymbol{\theta})|^{-1/2} \\ & \times \exp[-\tfrac{1}{2}\{(\boldsymbol{d} - \boldsymbol{m}_d(\boldsymbol{\theta}))^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1} (\boldsymbol{d} - \boldsymbol{m}_d(\boldsymbol{\theta}))\}]. \end{aligned} \tag{2}$$

Note that we have explicitly shown dependence on $\boldsymbol{\theta}$ but $\boldsymbol{m}_d(\boldsymbol{\theta})$ also depends on $\boldsymbol{\beta}$ and ρ, while $\boldsymbol{V}_d(\boldsymbol{\theta})$ depends on all of $\boldsymbol{\phi}$.

We can complete the square for $\boldsymbol{\beta}$ in the exponent of (2) to find

$$\boldsymbol{\beta} \mid \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{d} \sim N(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{W}(\boldsymbol{\theta})),$$

where

$$\begin{aligned} &\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{W}(\boldsymbol{\theta}) \boldsymbol{H}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1} \boldsymbol{d}, \\ &\boldsymbol{W}(\boldsymbol{\theta}) = (\boldsymbol{H}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1} \boldsymbol{H}(\boldsymbol{\theta}))^{-1}, \end{aligned}$$

and both depend on $\boldsymbol{\phi}$ as well as $\boldsymbol{\theta}$. Then integrating $\boldsymbol{\beta}$ out from (2) yields

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \boldsymbol{d}) \quad \propto \quad & p(\boldsymbol{\theta}) p(\boldsymbol{\phi}) |\boldsymbol{V}_d(\boldsymbol{\theta})|^{-1/2} |\boldsymbol{W}(\boldsymbol{\theta})|^{1/2} \\ & \times \exp[-\tfrac{1}{2}\{(\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1} (\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}))\}]. \end{aligned} \tag{3}$$

3. Estimating hyperparameters

A fully Bayesian analysis would now integrate out the hyperparameters $\boldsymbol{\phi}$ as well to leave the posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{d})$ of the calibration parameters. However, it is clear that (3) is a highly intractable function of $\boldsymbol{\phi}$. Even with the most parsimonious parametrisation of $c_1((\cdot,\cdot),(\cdot,\cdot))$ and $c_2(\cdot,\cdot)$, to integrate over $\boldsymbol{\phi}$ numerically would entail at least a six dimensional quadrature. Since much of the methodology that we develop herein may be rather computationally intensive even conditional on fixed values of $\boldsymbol{\phi}$, the full Bayesian analysis will not typically be practical. It is important to note also that it will generally not be possible to integrate (3) with respect to $\boldsymbol{\phi}$ if $p(\boldsymbol{\phi})$ is improper. To adopt a fully Bayesian analysis will therefore demand full and careful consideration of prior information regarding the hyperparameters.

We propose instead to derive plausible estimates of the components of $\boldsymbol{\phi}$ and then to act as if these were fixed. Thus, for inference about $\boldsymbol{\theta}$ we will use its conditional posterior given the estimated values of $\boldsymbol{\phi}$.

We propose estimating the hyperparameters in two stages. In the first stage we use just the code output data $\boldsymbol{y}$ to estimate the hyperparameters $\boldsymbol{\psi}_1$ of $c_1((\cdot,\cdot),(\cdot,\cdot))$. There is some information about $\boldsymbol{\psi}_1$ in the observational data $\boldsymbol{z}$, but (a) $\boldsymbol{z}$ depends also on the other hyperparameters and (b) the number n of observations in $\boldsymbol{z}$ will typically be very much smaller than the number N of output values in $\boldsymbol{y}$. Therefore very little is lost by this simplification. In the second stage we use $\boldsymbol{z}$ to estimate ρ, λ and the hyperparameters $\boldsymbol{\psi}_2$ of $c_2(\cdot,\cdot)$, having now fixed $\boldsymbol{\psi}_1$. Details of these two stages are as follows.

In stage 1, we estimate $\boldsymbol{\psi}_1$ by maximising $p(\boldsymbol{\psi}_1 \mid \boldsymbol{y})$. (Alternative estimates might be used. For instance, with weak prior information one might use standard geostatistical variogram estimates—see Cressie, 1991.) We have $\boldsymbol{y} \mid \boldsymbol{\beta}_1, \boldsymbol{\psi}_1 \sim N(\boldsymbol{H}_1(D_1)\boldsymbol{\beta}_1, \boldsymbol{V}_1(D_1))$ and integrate $\boldsymbol{\beta}_1$ from the joint distribution $p(\boldsymbol{\beta}_1, \boldsymbol{\psi}_1 \mid \boldsymbol{y})$ to obtain

$$\begin{aligned} p(\boldsymbol{\psi}_1 \mid \boldsymbol{y}) \quad \propto \quad & |\boldsymbol{V}_1(D_1)|^{-1/2} p(\boldsymbol{\psi}_1) |\boldsymbol{W}_1(D_1)|^{1/2} \\ & \exp\{-\tfrac{1}{2}(\boldsymbol{y} - \boldsymbol{H}_1(D_1)\hat{\boldsymbol{\beta}}_1)^T \boldsymbol{V}_1(D_1)^{-1}(\boldsymbol{y} - \boldsymbol{H}_1(D_1)\hat{\boldsymbol{\beta}}_1)\} \end{aligned} \tag{4}$$

where

$$\begin{aligned} &\hat{\boldsymbol{\beta}}_1 = \boldsymbol{W}_1(D_1)\boldsymbol{H}_1(D_1)^T \boldsymbol{V}_1(D_1)^{-1}\boldsymbol{y}, \\ &\boldsymbol{W}_1(D_1) = (\boldsymbol{H}_1(D_1)^T \boldsymbol{V}_1(D_1)^{-1}\boldsymbol{H}_1(D_1))^{-1}. \end{aligned}$$

Note that $\hat{\boldsymbol{\beta}}_1$ is not the same as the first part of $\hat{\boldsymbol{\beta}}$, since $\boldsymbol{\beta}$ depends on $\boldsymbol{z}$ as well as $\boldsymbol{y}$. If $c_1((\cdot,\cdot),(\cdot,\cdot))$ is parametrised in terms of a variance hyperparameter σ_1^2 via

$$c_1(\cdot,\cdot) = \sigma^2 r(\cdot,\cdot) \tag{5}$$

and if σ_1^2 has a conjugate inverse-gamma prior (or the standard 'noninformative' prior) its mode can be found in (4) analytically. We could even integrate σ_1^2 out before maximising with respect to the other components of $\boldsymbol{\psi}_1$, but in practice this would make negligible difference to subsequent analysis. Because of the dependence of $\boldsymbol{z}$ on $\boldsymbol{\psi}_1$, we still could not integrate σ_1^2 out of (3) analytically.

In stage 2, we wish ideally to estimate ρ, λ and $\boldsymbol{\psi}_2$ by maximising $p(\rho, \lambda, \boldsymbol{\psi}_2 \mid \boldsymbol{d}, \boldsymbol{\psi}_1)$. To obtain this density, we can first write

$$p(\boldsymbol{\beta}_2, \rho, \lambda, \boldsymbol{\psi}_2 \mid \boldsymbol{d}, \boldsymbol{\psi}_1) \propto p(\boldsymbol{\beta}_2, \rho, \lambda, \boldsymbol{\psi}_2) p(\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\phi}), \tag{6}$$

since $\boldsymbol{y}$ is independent of the second stage hyperparameters.

We cannot obtain the distribution of $[\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\phi}]$ analytically, but starting with $[\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\phi}, \boldsymbol{\theta}]$, which is normally distributed, we can obtain expressions for its mean vector and variance matrix as follows. The ith element of the mean vector is

$$\begin{aligned} E(z_i \mid \boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\phi}) \quad &= \quad \int E(z_i \mid \boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\phi}, \boldsymbol{\theta}) p(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= \quad \boldsymbol{h}_2(\boldsymbol{x}_i)^T \boldsymbol{\beta}_2 + \rho \left[\int \boldsymbol{h}_1(\boldsymbol{x}_i, \boldsymbol{\theta})^T p(\boldsymbol{\theta})\, d\boldsymbol{\theta}\right] \hat{\boldsymbol{\beta}}_1 \\ &\qquad + \rho \left[\int \boldsymbol{t}(\boldsymbol{x}_i, \boldsymbol{\theta})^T p(\boldsymbol{\theta})\, d\boldsymbol{\theta}\right] \boldsymbol{V}_1(D_1)^{-1}(\boldsymbol{y} - \boldsymbol{H}_1(D_1)\hat{\boldsymbol{\beta}}_1). \end{aligned} \tag{7}$$

where the jth element of $\boldsymbol{t}(\boldsymbol{x}_i, \boldsymbol{\theta})$ is $c_1((\boldsymbol{x}_i, \boldsymbol{\theta}), (\boldsymbol{x}_j^*, \boldsymbol{t}_j))$, for $j = 1, \ldots, N$.

The covariance matrix is $\boldsymbol{V} = \lambda \boldsymbol{I} + \boldsymbol{V}_2(D_2) + \rho^2 \boldsymbol{C}$, where the (i,j) element of $\boldsymbol{C}$ is

$$\begin{aligned} &\int \operatorname{cov}(\eta(\boldsymbol{x}_i, \boldsymbol{\theta}), \eta(\boldsymbol{x}_j, \boldsymbol{\theta}) \mid \boldsymbol{y}, \boldsymbol{\psi}_1) p(\boldsymbol{\theta})\, d\boldsymbol{\theta} = \int c_1((\boldsymbol{x}_i, \boldsymbol{\theta}), (\boldsymbol{x}_j, \boldsymbol{\theta})) p(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &-tr\left\{\boldsymbol{V}_1(D_1)^{-1} \int \boldsymbol{t}(\boldsymbol{x}_j, \boldsymbol{\theta}) \boldsymbol{t}(\boldsymbol{x}_i, \boldsymbol{\theta})^T p(\boldsymbol{\theta})\, d\boldsymbol{\theta}\right\} \\ &+tr\left\{\boldsymbol{W}_1(D_1) \int \boldsymbol{h}_1(\boldsymbol{x}_j, \boldsymbol{\theta}) \boldsymbol{h}_1(\boldsymbol{x}_i, \boldsymbol{\theta})^T p(\boldsymbol{\theta})\, d\boldsymbol{\theta}\right\} \\ &-tr\left\{\boldsymbol{W}_1(D_1)\boldsymbol{H}_1(D_1)^T \boldsymbol{V}_1(D_1)^{-1} \int \boldsymbol{t}(\boldsymbol{x}_j, \boldsymbol{\theta}) \boldsymbol{h}_1(\boldsymbol{x}_i, \boldsymbol{\theta})^T p(\boldsymbol{\theta})\, d\boldsymbol{\theta}\right\} \end{aligned}$$

$$-tr\left\{\boldsymbol{V}_1(D_1)^{-1}\boldsymbol{H}_1(D_1)\boldsymbol{W}_1(D_1)\int \boldsymbol{h}_1(\boldsymbol{x}_j,\boldsymbol{\theta})\boldsymbol{t}(\boldsymbol{x}_i,\boldsymbol{\theta})^T p(\boldsymbol{\theta})\,d\boldsymbol{\theta}\right\}$$
$$+tr\left\{\boldsymbol{V}_1(D_1)^{-1}\boldsymbol{H}_1(D_1)\boldsymbol{W}_1(D_1)\boldsymbol{H}_1(D_1)^T\boldsymbol{V}_1(D_1)^{-1}\int \boldsymbol{t}(\boldsymbol{x}_j,\boldsymbol{\theta})\boldsymbol{t}(\boldsymbol{x}_i,\boldsymbol{\theta})^T p(\boldsymbol{\theta})\,d\boldsymbol{\theta}\right\}.$$

We then approximate $[\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{\beta}_2, \boldsymbol{\phi}]$ by a normal distribution with the above moments for the purpose of estimating ρ, λ and $\boldsymbol{\psi}_2$. To simplify the notation we write the mean vector as $\boldsymbol{H}_2(D_2)\boldsymbol{\beta}_2 + \rho\hat{\boldsymbol{\eta}}(D_2)$, where $\hat{\boldsymbol{\eta}}(D_2)$ is the vector with elements $E(\eta(\boldsymbol{x}_i,\boldsymbol{\theta}) \mid \boldsymbol{y})$, $i = 1,\ldots,n$. We can now integrate out $\boldsymbol{\beta}_2$ from (6) to obtain the approximation

$$\begin{aligned} p(\rho,\lambda,\boldsymbol{\psi}_2 \mid \boldsymbol{d},\boldsymbol{\psi}_1) \quad &\propto \quad p(\rho,\lambda,\boldsymbol{\psi}_2)|\boldsymbol{V}|^{-1/2}|\boldsymbol{W}_2|^{1/2} \\ &\qquad \times \exp\{-\tfrac{1}{2}(\boldsymbol{z}-\boldsymbol{H}_2(D_2)\hat{\boldsymbol{\beta}}_2-\rho\hat{\boldsymbol{\eta}}(D_2))^T\boldsymbol{V}^{-1}(\boldsymbol{z}-\boldsymbol{H}_2(D_2)\hat{\boldsymbol{\beta}}_2-\rho\hat{\boldsymbol{\eta}}(D_2))\}, \end{aligned}$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}}_2 &= \boldsymbol{W}_2\boldsymbol{H}_2(D_2)\boldsymbol{V}^{-1}(\boldsymbol{z}-\rho\hat{\boldsymbol{\eta}}(D_2)), \\ \boldsymbol{W}_2 &= (\boldsymbol{H}_2(D_2)^T\boldsymbol{V}^{-1}\boldsymbol{H}_2(D_2))^{-1}. \end{aligned}$$

The integrals in these expressions can be evaluated analytically for certain combinations of the functions $\boldsymbol{h}_1(\boldsymbol{x},\boldsymbol{\theta})$, $p(\boldsymbol{\theta})$ and $c_1((\cdot,\boldsymbol{\theta}),(\cdot,\boldsymbol{\theta}))$. In particular, we can derive results as follows when $\boldsymbol{\theta} \sim N(\boldsymbol{m}_\theta, \boldsymbol{V}_\theta)$, $c_1((\boldsymbol{x},\boldsymbol{t}),(\boldsymbol{x}',\boldsymbol{t}')) = \sigma_1^2\exp\{-(\boldsymbol{x}-\boldsymbol{x}')^T\boldsymbol{\Omega}_x(\boldsymbol{x}-\boldsymbol{x}')-(\boldsymbol{t}-\boldsymbol{t}')^T\boldsymbol{\Omega}_t(\boldsymbol{t}-\boldsymbol{t}')\}$ and the elements of $\boldsymbol{h}_1(\boldsymbol{x},\boldsymbol{\theta})$ are such that the expectation of the product of any pair can be found with respect to an arbitrary multivariate normal distribution for $\boldsymbol{\theta}$. The appropriateness of such specifications as practical model choices is discussed in Kennedy and O'Hagan (2000), Section 5.2.

Note that $\boldsymbol{\Omega}_x$ and $\boldsymbol{\Omega}_t$ are arbitrary functions of $\boldsymbol{\psi}_1$: we consider simple specifications of these, and of $c_1((\cdot,\cdot),(\cdot,\cdot))$ generally, in our examples. The general closed form expressions for the integrals are as follows. To calculate (7) we have

$$\int \boldsymbol{h}_1(\boldsymbol{x}_i,\boldsymbol{\theta})^T p(\boldsymbol{\theta})\,d\boldsymbol{\theta} = E_\Theta\{\boldsymbol{h}_1(\boldsymbol{x}_i,\boldsymbol{\theta})\}^T,$$

where E_Θ is expectation with respect $N(\boldsymbol{m}_\theta,\boldsymbol{V}_\theta)$, and the jth element of $\int \boldsymbol{t}(\boldsymbol{x}_i,\boldsymbol{\theta})p(\boldsymbol{\theta})\,d\boldsymbol{\theta}$ is

$$\begin{aligned} \int c_1((\boldsymbol{x}_i,\boldsymbol{\theta}),(\boldsymbol{x}_j{}^*,\boldsymbol{t}_j))p(\boldsymbol{\theta})\,d\boldsymbol{\theta} \quad &= \quad \sigma_1^2|\boldsymbol{I}+2\boldsymbol{V}_\theta\boldsymbol{\Omega}_t|^{-1/2}\exp\{-(\boldsymbol{x}_i-\boldsymbol{x}_j{}^*)^T\boldsymbol{\Omega}_x(\boldsymbol{x}_i-\boldsymbol{x}_j{}^*)\} \\ &\qquad \times\exp\{-(\boldsymbol{m}_\theta-\boldsymbol{t}_j)^T(2\boldsymbol{V}_\theta+\boldsymbol{\Omega}_t^{-1})^{-1}(\boldsymbol{m}_\theta-\boldsymbol{t}_j)\}. \end{aligned}$$

The kth column of $\int \boldsymbol{h}_1(\boldsymbol{x}_j,\boldsymbol{\theta})\boldsymbol{t}(\boldsymbol{x}_i,\boldsymbol{\theta})^T p(\boldsymbol{\theta})\,d\boldsymbol{\theta}$ is

$$\begin{aligned} &\sigma_1^2|\boldsymbol{I}+2\boldsymbol{V}_\theta\boldsymbol{\Omega}_x|^{-1/2}\exp\{-(\boldsymbol{x}_i-\boldsymbol{x}_k{}^*)^T\boldsymbol{\Omega}_x(\boldsymbol{x}_i-\boldsymbol{x}_k{}^*)\} \\ &\times\exp\{-(\boldsymbol{m}_\theta-\boldsymbol{t}_k)^T(2\boldsymbol{V}_\theta+\boldsymbol{\Omega}_t^{-1})^{-1}(\boldsymbol{m}_\theta-\boldsymbol{t}_k)\}E'_\Theta(\boldsymbol{h}_1(\boldsymbol{x}_i,\boldsymbol{\theta})), \end{aligned}$$

where the expectation E'_Θ is with respect to the normal distribution $N((\boldsymbol{V}_\theta{}^{-1}+2\boldsymbol{\Omega}_t)^{-1}(\boldsymbol{V}_\theta{}^{-1}\boldsymbol{m}_\theta+2\boldsymbol{\Omega}_t\boldsymbol{t}_k),(\boldsymbol{V}_\theta{}^{-1}+2\boldsymbol{\Omega}_t)^{-1})$. The (k,l) element of $\int \boldsymbol{t}(\boldsymbol{x}_j,\boldsymbol{\theta})\boldsymbol{t}(\boldsymbol{x}_i,\boldsymbol{\theta})^T p(\boldsymbol{\theta})\,d\boldsymbol{\theta}$ is

$$\begin{aligned} &\sigma_1^4|\boldsymbol{I}+4\boldsymbol{V}_\theta\boldsymbol{\Omega}_t|^{-1/2}\exp\{-(\boldsymbol{x}_j-\boldsymbol{x}_k{}^*)^T\boldsymbol{\Omega}_x(\boldsymbol{x}_j-\boldsymbol{x}_k{}^*)-(\boldsymbol{x}_i-\boldsymbol{x}_l{}^*)^T\boldsymbol{\Omega}_x(\boldsymbol{x}_i-\boldsymbol{x}_l{}^*)\} \\ &\times\exp\left\{-\tfrac{1}{2}(\boldsymbol{t}_k-\boldsymbol{t}_l)^T\boldsymbol{\Omega}_t(\boldsymbol{t}_k-\boldsymbol{t}_l)-\tfrac{1}{2}\left(\boldsymbol{m}_\theta-\frac{\boldsymbol{t}_k+\boldsymbol{t}_l}{2}\right)^T\left(\boldsymbol{V}_\theta+\frac{\boldsymbol{\Omega}_t{}^{-1}}{4}\right)^{-1}\left(\boldsymbol{m}_\theta-\frac{\boldsymbol{t}_k+\boldsymbol{t}_l}{2}\right)\right\} \end{aligned}$$

and finally

$$\int \boldsymbol{h}_1(\boldsymbol{x}_j,\boldsymbol{\theta})\boldsymbol{h}_1(\boldsymbol{x}_i,\boldsymbol{\theta})^T p(\boldsymbol{\theta})\,d\boldsymbol{\theta} = E_\Theta(\boldsymbol{h}_1(\boldsymbol{x}_j,\boldsymbol{\theta})\boldsymbol{h}_1(\boldsymbol{x}_i,\boldsymbol{\theta})^T).$$

4. Calibration

Having estimated the hyperparameters $\boldsymbol{\phi}$ we now condition on these, so that we regard the posterior distribution of the calibration parameters to be

$$p(\boldsymbol{\theta} \mid \boldsymbol{\phi},\boldsymbol{d}) \propto |\boldsymbol{V}_d(\boldsymbol{\theta})|^{-1/2}|\boldsymbol{W}(\boldsymbol{\theta})|^{1/2}\exp[-\tfrac{1}{2}\{\boldsymbol{d}-\boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\}^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}\{\boldsymbol{d}-\boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})\}]p(\boldsymbol{\theta}). \tag{8}$$

We can use this to make inference about $\boldsymbol{\theta}$, although its intractability means that numerical methods must be used.

In practice, we will not generally be interested in inference about $\boldsymbol{\theta}$ as such. The purpose of calibration is to use the calibrated model for predicting the real process. We can think of calibration as a preliminary to addressing the other statistical problems of interpolation, sensitivity analysis and uncertainty analysis. Thus, the problem of predicting the true process $z(\boldsymbol{x})$ at some specified variable inputs $\boldsymbol{x}$ can be seen as interpolating the function $z(\cdot)$, and is dealt with in Section 4.1. Uncertainty analysis is addressed in Section 4.2. We do not explicitly deal with sensitivity analysis in this paper: appropriate techniques are outlined in O'Hagan *et al.* (1999).

4.1. Calibrated prediction

The posterior distribution of $z(\cdot)$ conditional on the estimated hyperparameters $\boldsymbol{\phi}$ and the calibration parameters $\boldsymbol{\theta}$ is a Gaussian process. Its mean function is given by

$$E(z(\boldsymbol{x}) \mid \boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{d}) = \boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta})^T\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{t}(\boldsymbol{x},\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}(\boldsymbol{d}-\boldsymbol{H}(\boldsymbol{\theta})\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})), \tag{9}$$

where

$$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta}) = \begin{pmatrix} \rho\boldsymbol{h}_1(\boldsymbol{x},\boldsymbol{\theta}) \\ \boldsymbol{h}_2(\boldsymbol{x}) \end{pmatrix}$$

and

$$\boldsymbol{t}(\boldsymbol{x},\boldsymbol{\theta}) = \begin{pmatrix} \rho\boldsymbol{V}_1((\boldsymbol{x},\boldsymbol{\theta}),D_1) \\ \rho^2\boldsymbol{V}_1((\boldsymbol{x},\boldsymbol{\theta}),D_2(\boldsymbol{\theta})) + \boldsymbol{V}_2(\boldsymbol{x},D_2) \end{pmatrix}.$$

Its covariance function is given by

$$\begin{aligned} \mathrm{cov}(z(\boldsymbol{x}),z(\boldsymbol{x}') \mid \boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{d}) &= \rho^2 c_1((\boldsymbol{x},\boldsymbol{\theta}),(\boldsymbol{x}',\boldsymbol{\theta})) + c_2(\boldsymbol{x},\boldsymbol{x}') \\ &\quad -\boldsymbol{t}(\boldsymbol{x},\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}\boldsymbol{t}(\boldsymbol{x}',\boldsymbol{\theta}) \\ &\quad +(\boldsymbol{h}(\boldsymbol{x},\boldsymbol{\theta})-\boldsymbol{H}(\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}\boldsymbol{t}(\boldsymbol{x},\boldsymbol{\theta}))^T\boldsymbol{W}(\boldsymbol{\theta})(\boldsymbol{h}(\boldsymbol{x}',\boldsymbol{\theta})-\boldsymbol{H}(\boldsymbol{\theta})^T\boldsymbol{V}_d(\boldsymbol{\theta})^{-1}\boldsymbol{t}(\boldsymbol{x}',\boldsymbol{\theta})). \end{aligned}$$

By combining this distribution with the posterior distribution (8) of $\boldsymbol{\theta}$, we can make inferences about $z(\boldsymbol{x})$, again using numerical computation methods. For instance to estimate $z(\boldsymbol{x})$ we might use its posterior mean $E(z(\boldsymbol{x}) \mid \boldsymbol{\phi},\boldsymbol{d})$ (for the estimated values of $\boldsymbol{\phi}$), obtained by integrating (9) with respect to (8).

4.2. Calibrated uncertainty analysis

Now suppose that we wish to predict the real process in the context where one or more of the variable inputs is subject to parametric variability, as discussed in Kennedy and O'Hagan (2000), Section 2.1. The problem of uncertainty analysis is to study the (extra) uncertainty in model outputs induced by this parametric variability. Although uncertainty analysis for computer codes is typically formulated in this way, i.e. with concern for uncertainty in the code outputs, in the present context the larger challenge is to study uncertainty in the real process $z(\cdot)$.

We therefore consider the random variable $z(\boldsymbol{X})$, where the variable inputs $\boldsymbol{X}$ are now random, having a distribution $G_{\boldsymbol{X}}(\boldsymbol{x})$. (In practice, only a subset of the variable inputs will be subject to parametric variability, so $G_{\boldsymbol{X}}(\cdot)$ will be degenerate in the other dimensions.) The task of uncertainty analysis is now to make inference about the *distribution* of $z(\boldsymbol{X})$. In particular, we wish to make inference about properties of this distribution such as the mean $K = E_{\boldsymbol{X}}\{z(\boldsymbol{X})\} = \int_{\mathcal{X}} z(\boldsymbol{x})\, dG_{\boldsymbol{X}}(\boldsymbol{x})$, the variance $L = \mathrm{var}_{\boldsymbol{X}}\{z(\boldsymbol{X})\} = K_2 - K^2$, where $K_2 = \int_{\mathcal{X}} z(\boldsymbol{x})^2\, dG_{\boldsymbol{X}}(\boldsymbol{x})$, or the value at some point g of the distribution function $F(g) = P_X\{z(\boldsymbol{X}) \leq g\} = \int_{z(\boldsymbol{x})\leq g} dG_{\boldsymbol{X}}(\boldsymbol{x})$. Inference about these or other summaries of the distribution of $z(\boldsymbol{X})$ may be derived from the posterior distribution of $z(\cdot)$.

For instance, the posterior mean of K (given the estimated hyperparameters) is

$$\begin{aligned} E(K \mid \boldsymbol{\phi}, \boldsymbol{d}) &= \int_\Theta \left[\int_{\mathcal{X}} E\{z(\boldsymbol{x}) \mid \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{d}\}\, dG_{\boldsymbol{X}}(\boldsymbol{x}) \right] dG^*(\boldsymbol{\theta}) \\ &= \int_\Theta \left[\bar{\boldsymbol{h}}(\boldsymbol{\theta})^T \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \bar{\boldsymbol{t}}(\boldsymbol{\theta})^T \boldsymbol{V}_d(\boldsymbol{\theta})^{-1} (\boldsymbol{d} - \boldsymbol{H}(\boldsymbol{\theta}) \hat{\boldsymbol{\beta}}(\boldsymbol{\theta})) \right] dG^*(\boldsymbol{\theta}), \qquad (10) \end{aligned}$$

where $G^*(\boldsymbol{\theta})$ is the posterior distribution of $\boldsymbol{\theta}$ so that $dG^*(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \boldsymbol{d})\, d\boldsymbol{\theta}$, $\bar{\boldsymbol{h}}(\boldsymbol{\theta}) = \int_{\mathcal{X}} \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\theta})\, dG_{\boldsymbol{X}}(\boldsymbol{x})$ and $\bar{\boldsymbol{t}}(\boldsymbol{\theta}) = \int_{\mathcal{X}} \boldsymbol{t}(\boldsymbol{x}, \boldsymbol{\theta})\, dG_{\boldsymbol{X}}(\boldsymbol{x})$.

These integrals can be evaluated analytically for suitable choices of $dG_{\boldsymbol{X}}(\cdot)$, $\boldsymbol{h}_1(\cdot,\cdot)$, $\boldsymbol{h}_2(\cdot)$, $c_1((\cdot,\cdot),(\cdot,\cdot))$ and $c_2(\cdot,\cdot)$. In particular, if $G_{\boldsymbol{X}}$ is taken to be $N(\boldsymbol{m}_x, \boldsymbol{V}_x)$ and

$$\begin{aligned} c_1((\boldsymbol{x}, \boldsymbol{t}), (\boldsymbol{x}', \boldsymbol{t}')) &= \sigma_1^2 c_{1x}(\boldsymbol{x}, \boldsymbol{x}') c_{1\theta}(\boldsymbol{t}, \boldsymbol{t}') && (11) \\ &= \sigma_1^2 \exp\{-(\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{\Omega}_x (\boldsymbol{x} - \boldsymbol{x}')\} \exp\{-(\boldsymbol{t} - \boldsymbol{t}')^T \boldsymbol{\Omega}_t (\boldsymbol{t} - \boldsymbol{t}')\}, && (12) \\ c_2(x, x') &= \sigma_2^2 \exp\{-(\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{\Omega}_x^* (\boldsymbol{x} - \boldsymbol{x}')\} && (13) \end{aligned}$$

then $\bar{\boldsymbol{t}}(\boldsymbol{\theta})^T = (\bar{\boldsymbol{t}}_1(\boldsymbol{\theta})^T, \bar{\boldsymbol{t}}_2(\boldsymbol{\theta})^T)$ where $\bar{\boldsymbol{t}}_1(\boldsymbol{\theta})^T$ has ith element

$$\bar{\boldsymbol{t}}_1(\boldsymbol{\theta})_i = \rho\sigma_1^2 |\boldsymbol{I} + 2\boldsymbol{V}_x \boldsymbol{\Omega}_x|^{-1/2} \exp\{-(\boldsymbol{\theta} - \boldsymbol{t}_i)^T \boldsymbol{\Omega}_t (\boldsymbol{\theta} - \boldsymbol{t}_i) - (\boldsymbol{x}_i^* - \boldsymbol{m}_x)^T (2\boldsymbol{V}_x + \boldsymbol{\Omega}_x^{-1})^{-1} (\boldsymbol{x}_i^* - \boldsymbol{m}_x)\} \qquad (14)$$

for $i = 1, \ldots, N$, and $\bar{\boldsymbol{t}}_2(\boldsymbol{\theta})^T$ has ith element

$$\begin{aligned} \bar{\boldsymbol{t}}_2(\boldsymbol{\theta})_i &= \rho^2 \sigma_1^2 |\boldsymbol{I} + 2\boldsymbol{V}_x \boldsymbol{\Omega}_x|^{-1/2} \exp\{-(\boldsymbol{x}_i - \boldsymbol{m}_x)^T (2\boldsymbol{V}_x + \boldsymbol{\Omega}_x^{-1})^{-1} (\boldsymbol{x}_i - \boldsymbol{m}_x)\} \\ &\quad + \sigma_2^2 |\boldsymbol{I} + 2\boldsymbol{V}_x \boldsymbol{\Omega}_x^*|^{-1/2} \exp\{-(\boldsymbol{x}_i - \boldsymbol{m}_x)^T (2\boldsymbol{V}_x + \boldsymbol{\Omega}_x^{*-1})^{-1} (\boldsymbol{x}_i - \boldsymbol{m}_x)\} \qquad (15) \end{aligned}$$

for $i = 1, \ldots, n$. Also

$$\bar{\boldsymbol{h}}(\boldsymbol{\theta}) = \begin{pmatrix} \rho E_X\{\boldsymbol{h}_1(\boldsymbol{x}, \boldsymbol{\theta})\} \\ E_X\{\boldsymbol{h}_2(\boldsymbol{x})\} \end{pmatrix},$$

where the expectations are with respect to the $N(\boldsymbol{m}_x, \boldsymbol{V}_x)$ distribution for $\boldsymbol{X}$.

The integration with respect to $\boldsymbol{\theta}$ in (10) must be performed numerically. By similar expansions, following Haylock and O'Hagan (1996), we can evaluate $\mathrm{var}(K \mid \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{d})$, $E(L \mid$

$\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{d})$ and $\mathrm{var}(L \mid \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{d})$. Numerical integration with respect to (8) then yields $\mathrm{var}(K \mid \boldsymbol{\phi}, \boldsymbol{d})$, $E(L \mid \boldsymbol{\phi}, \boldsymbol{d})$ or $\mathrm{var}(L \mid \boldsymbol{\phi}, \boldsymbol{d})$. The approach of Oakley and O'Hagan (1998) can be used to obtain posterior moments of the distribution function $F(\cdot)$.

It is equally straightforward to work in terms of an uncertainty analysis of the code output $\eta(\boldsymbol{x}, \boldsymbol{\theta})$, with respect to either or both of parametric variability in $\boldsymbol{x}$ and parametric uncertainty (after calibration) in $\boldsymbol{\theta}$.

5. Computation

The main computational issues concern the need for numerical integration with respect to the posterior distribution G^* of $\boldsymbol{\theta}$, and the need to invert the matrix $\boldsymbol{V}_d(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$ value in that numerical integration.

If the code $\eta(\cdot,\cdot)$ is complex and computer-intensive, we will expect the number N of code evaluations available to be relatively small (and we expect n to be smaller still). Then the inversion of the $(N+n) \times (N+n)$ matrix $\boldsymbol{V}_d(\boldsymbol{\theta})$ may not be a serious problem. However, for a simpler code we may expect to be able to make larger numbers of runs to obtain more information about $\eta(\cdot,\cdot)$. Then N is potentially very large. In this case considerable computational savings are achieved by the code design D_1 having a Cartesian product form. Suppose first that D_1 is the Cartesian product of an n_x-point variable input design $\{\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_{n_x}^*\}$ and an n_t-point calibration input design $\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_{n_t}\}$, with points arranged row-wise so that $D_1 = \{(\boldsymbol{x}_1^*, \boldsymbol{t}_1), \ldots, (\boldsymbol{x}_{n_x}^*, \boldsymbol{t}_1), \ldots, (\boldsymbol{x}_1^*, \boldsymbol{t}_{n_t}), \ldots, (\boldsymbol{x}_{n_x}^*, \boldsymbol{t}_{n_t})\}$. Then, using the fact that $c_1((\cdot,\cdot),(\cdot,\cdot))$ has the separable form (11) we find that $\boldsymbol{V}_1(D_1) = \sigma_1^2 \boldsymbol{A}_t \otimes \boldsymbol{A}_x$, where $\boldsymbol{A}_t$ and $\boldsymbol{A}_x$ are correlation matrices for the calibration inputs and variable inputs respectively. Then $\boldsymbol{V}_1(D_1)^{-1} = \sigma_1^{-2} \boldsymbol{A}_t^{-1} \otimes \boldsymbol{A}_x^{-1}$ and $|\boldsymbol{V}_1(D_1)| = \sigma_1^{2n_t n_x} |\boldsymbol{A}_t|^{n_x} |\boldsymbol{A}_x|^{n_t}$. We thereby need only work with matrices of dimensions $n_t \times n_t$ and $n_x \times n_x$, instead of $N \times N$ where $N = n_t n_x$. If the calibration input and/or variable input designs are also Cartesian products in the individual inputs we obtain considerable further simplification. Other Kronecker product forms arise for $\boldsymbol{C}_1(D_1, D_2(\boldsymbol{\theta}))$ and for various derived matrices. These, combined with standard results for partitioned matrices, allow efficient computation even when the number of code evaluations is extremely large. Another device that might be considered for computation with large correlation matrices is the local computation approach; Vecchia (1988). However, it is not clear how that idea could be usefully applied in the more complex framework of calibration.

Turning now to the question of numerical integration with respect to $\boldsymbol{\theta}$, in our examples we use the iterative Gauss-Hermite quadrature method of Naylor and Smith (1982). This approach is realistic because the dimensionality of $\boldsymbol{\theta}$ is relatively low, so that quadrature is feasible, and because the code is relatively simple, so that we can afford to use Cartesian product rules and iteration. With more expensive codes or in somewhat higher dimensional $\boldsymbol{\theta}$ space it becomes important to use more efficient quadrature designs (for references see Evans and Schwartz, 1995). For high-dimensional $\boldsymbol{\theta}$, it may become necessary to use simulation methods of integration: we have not explored this yet.

6. Sensitivity to modelling assumptions in the Tomsk example

An analysis of the Tomsk data is described in Kennedy and O'Hagan (2000), in which we made various modelling choices, particularly in relation to the correlation function. These assumptions will not be appropriate for all applications. We now examine how some alternative plausible modelling assumptions affect inferences in the case of the 25-point Tomsk data.

The original model described in Kennedy and O'Hagan (2000) (Strategy 2) will be referred to as M1. Three alternative models are outlined below. In M2 we relax the assumption that the hyperparameters are fixed, in M3 we use an alternative functional form for the correlation function, and in M4 we consider the isotropic form of the gaussian correlation function.

M2: Integration with respect to the roughness parameters. It was suggested in Section 3 that fixing hyperparameters at the posterior modal values, rather than treating them as uncertain, is an acceptable simplification of the model. It is often the case in models of this kind that inferences (especially posterior variances) are sensitive to the choice of the roughness parameters in the correlation function. In an attempt to take more account of the uncertainty about these parameters, we used a simple numerical method to integrate over ω_1, ω_2 in calculating the posterior predictive means and variances. These represent the roughness parameters in our non-isotropic product of 1-dimensional correlation functions. We assumed a weak but proper prior for ω_1, ω_2, to ensure that the integrals we are approximating are proper. The remaining hyperparameters ρ, λ, and σ^2 were fixed at their original posterior modal values. A 25-point Gauss-Hermite integration rule was used (Naylor and Smith, 1982), based on a product of univariate 5-point Gauss-Hermite rules. Although, as was pointed out in Section 3, the posterior distribution is strictly improper when we use improper priors for these hyperparameters, this computation is intended only to show the likely effect of integrating with respect to the hyperparameters if proper but relatively flat prior distributions were used.

M3: Isotropic Matérn correlation. The exponential form of the correlation function is appropriate if the inadequacy function is analytic, and therefore may not be the best for modelling physical systems. We carried out an analysis identical to the one described above but using the isotropic Matérn correlation function suggested by Handcock and Wallis (1994), with hyperparameters $\nu > 0$ and $\alpha > 0$. This function has the form

$$c(t) = \frac{(\alpha t)^{\nu} \mathcal{K}_{\nu}(\alpha t)}{\Gamma(\nu) 2^{\nu-1}} \tag{16}$$

where Γ is a Gamma function and $\mathcal{K}_\nu$ is a modified Bessel function (Abramowitz and Stegun 1965, pp. 374–379). The form (16) represents a much wider class of correlations than the simpler gaussian form. We estimate the hyperparameters (ν, α) in the same way as before. Using the noninformative prior $p(\phi) \propto \sigma^{-2}$, the posterior mode estimates are $(\hat{\sigma}^2, \hat{\nu}, \hat{\alpha}) = (0.242, 3.875, 0.217)$. The implied correlation function is plotted in Figure 1.

M4: Isotropic gaussian correlation function. The final model variation considered uses the isotropic gaussian correlation function

$$c(\boldsymbol{x}, \boldsymbol{x}') = \exp(-b|\boldsymbol{x} - \boldsymbol{x}'|^2).$$

This is equivalent to assuming that $\omega_1 = \omega_2$ in M1. The estimated values of these parameters under M1 differ by a factor of 20. Under M4 we might therefore expect to see quite different inferences than we obtained with M1 if there is sensitivity to these roughness parameters. This function is plotted in Figure 1 with b estimated by its posterior mode.

We now consider the effects of using each of these models. The absolute errors of prediction based on M4 are shown in Figure 2. Very similar patterns of errors are seen

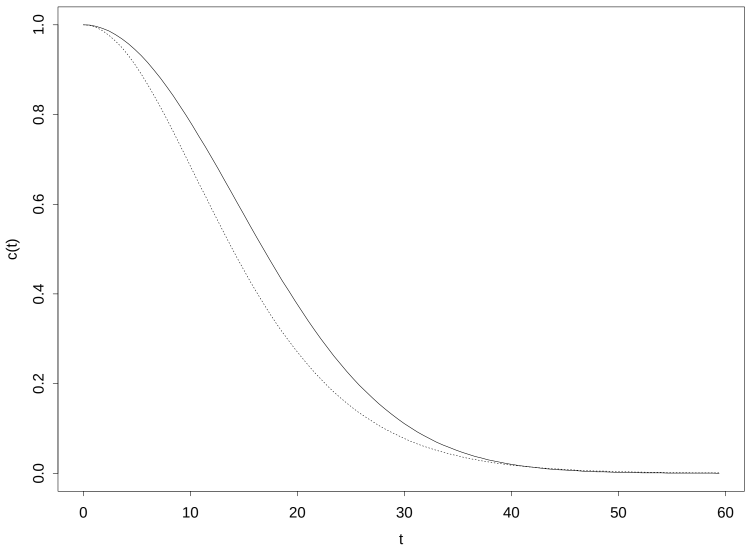

Fig. 1. Matérn correlation function (dashed line) and exponential correlation function (solid line) with hyperparameters estimated from the data

with each of the model variations M1–M4. The errors are greatest in the South-West region close to the source of the release, and also on the north side of the plume. The Q-Q plots of the residuals for each of the model variations are shown in Figure 3. The extreme negative residuals seen in the Q-Q plots, which indicate a bad fit of the model, correspond to an overestimation of deposition close to the source. In this particular application, however, inference about deposition in regions further from the source are much more important. The errors in prediction further North are accompanied by larger variances, and the standardised residuals in this region are approximately $N(0, 1)$.

The RMSE obtained using the modified predictions from M2 is 0.35, which is slightly better than the 0.36 achieved when the ω parameters are assumed to be known. Such a modest improvement does not justify the additional computation required, which is substantial even for a crude 2-dimensional integration.

While taking into account hyperparameter uncertainty appears to give modest improvements in RMSE, as does the assumption of isotropy, the standardised residuals look similar in each case. In particular, for the Tomsk dataset, the difference between the use of gaussian or Matérn class appears to be negligible. This will not always be the case, of course. With additional data points it may be possible to estimate more appropriate hyperparameters for the Matérn correlation function. Also, model inadequacy corrections do appear to be smooth in this problem.

In short, we conclude that (a) any improvement due to integrating with respect to the hyperparameters, as opposed to maximising, is likely to be small, and (b) the effect of using alternative covariance structures is also small. However, we do not propose that any of these models fits perfectly. Figure 2 and the Q-Q plots suggest that the true deposition surface

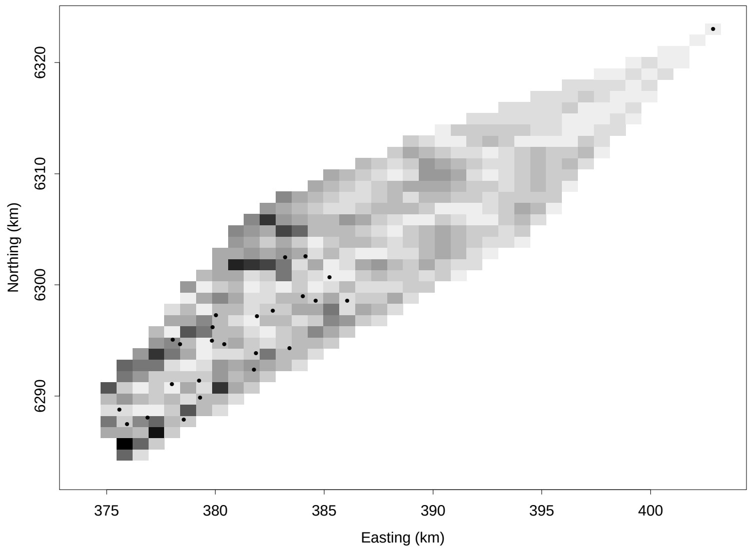

Fig. 2. Absolute errors of prediction based on model M4 (dark regions correspond to the largest errors), and the 25 design points

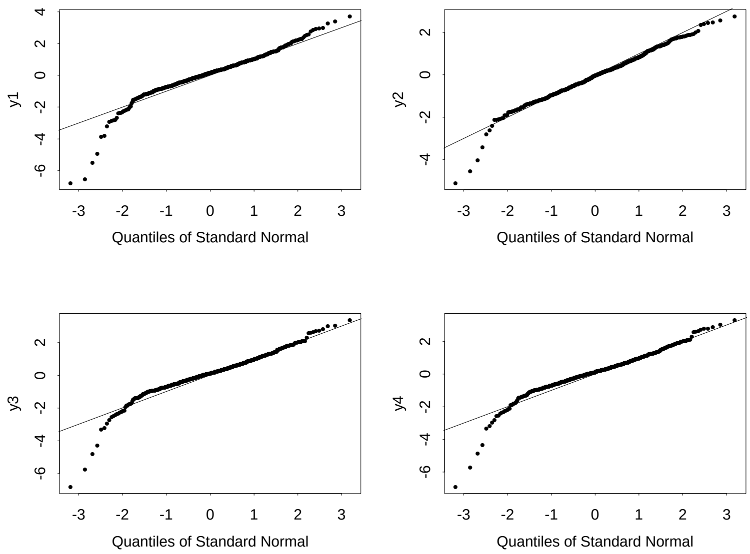

Fig. 3. Normal Q-Q plots of standardised residuals y1–y4 based on four different models M1–M4 and 25 observations

exhibits local features that we are failing to predict. In some sense, all of the models have predictive distributions with tails that are too thin. This is an area for ongoing investigation, but is likely to be application-specific.

7. EXAMPLE: EXERCISE DATA

7.1. Data and model

The data for our second example comes from a recent exercise carried out at the NRPB in which an accidental release was simulated. We should stress that although the data are not real, they are realistic. They were not simply simulated from the plume model, but deliberately incorporated unexpected features designed to test NRPB's procedures thoroughly. Figure 4 shows the 112 log-deposition measurements of I-131 which were supplied to NRPB during the course of the exercise. The contours represent an interpolation of the log-deposition using Splus. The first 20 observations notified to NRPB in the exercise are marked 'D', and these comprise our calibration data. Notice that these are not particularly well placed. Four of these data are the furthest points from the area of greatest radiation deposition, and an area of high deposition in the West is not represented in these data. This is not unusual, and in practice one must use whatever early data are available, but it means that we cannot expect to predict the actual deposition pattern accurately from these data.

The wind direction changed twice while material was still being released from the source. For the code $\eta(\cdot)$, we use a sum of logarithms from 3 Gaussian plume models, each with an unknown source term but known duration and wind direction. The deposition velocity was fixed at 0.001, which is the default value for deposition of iodine particles. The variable inputs $\boldsymbol{x}$ represent the position of a point relative to the source of the release and the wind direction of the first plume. $\boldsymbol{\theta}$ is a vector of three log source terms, one for each of the wind phases. The data comprise $(\boldsymbol{x}_1, z_1), \ldots, (\boldsymbol{x}_{20}, z_{20})$, where $\boldsymbol{x}_i$ is the variable input of the ith measurement and z_i is the corresponding measurement.

In the Tomsk example of Kennedy and O'Hagan (2000), the code was treated as a known function. Although the code used here is also cheap, we now treat the code as an expensive function in order to demonstrate the full analysis in which both the code and the model inadequacy are modelled as Gaussian processes. For the code design we take the coordinates from the first 20 physical data sites and form the cartesian product with a 4-point latin hypercube design in the 3-dimensional $\boldsymbol{\theta}$-space, giving a total of 80 code points. In the absence of real prior information, the prior mean was chosen to minimise the mean squared error between the code and the physical observations. The best-fitting value is $\boldsymbol{\theta}^* = (35.7, 20.8, 21.2)$. Traditional calibration would use the code with the input vector assumed to be fixed at this value. The root mean squared prediction error (RMSE) based on the remaining 92 observations is 3.28. Prior variances of 5 were used to represent weak prior information, so that the prior is $\boldsymbol{\theta} \sim N(\boldsymbol{\theta}^*, 5\boldsymbol{I})$. Although we could be accused of cheating by using the data to choose the prior mean, the effect on the analysis is minimal due to the large prior variances. In a truly real scenario, prior estimates would be elicited from radiation experts before any measurements are taken.

7.2. Results

Using the calibrated prediction method we obtain RMSE=2.45. Interpolation of the physical data alone using a Gaussian process model produces RMSE=2.88. This corresponds to

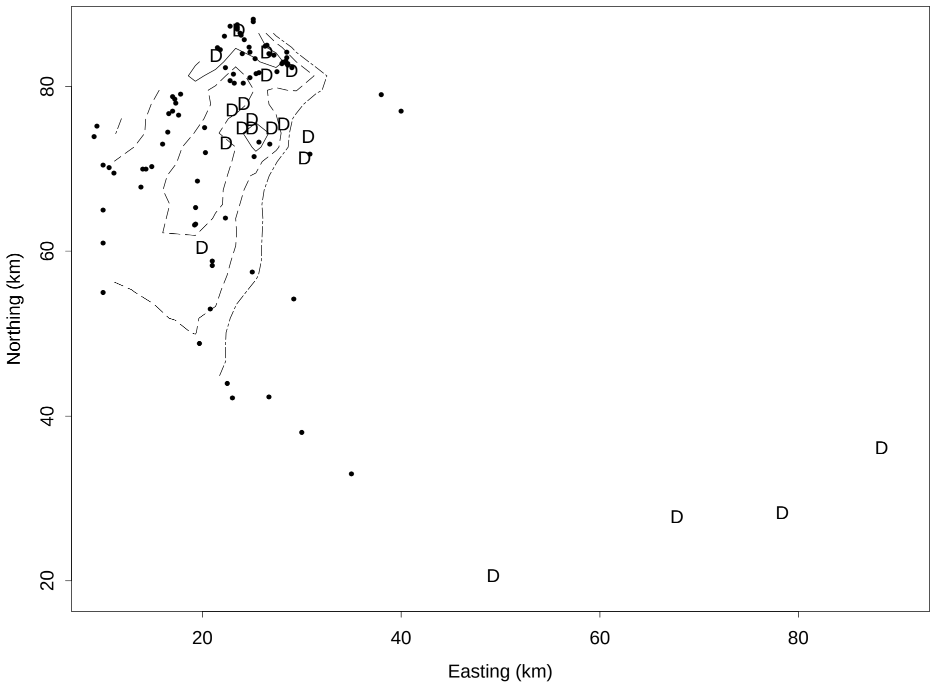

Fig. 4. Measurements of I-131 log-deposition, with contour lines at 12 (solid line), 10 (— — —) and 8 (— - —)

reducing the typical prediction error of concentration from a factor of $\exp(2.88) = 17.8$ to $\exp(2.45) = 11.6$. The errors are very large, due mainly to the poor data quality, but again the improvement is worthwhile. The quantile plots based on these prediction strategies are shown in Figure 5. These plots illustrate how the simple data interpolation underestimates the variance and overestimates deposition generally.

ACKNOWLEDGEMENTS

This research was supported by research grant GR/K54557 from the Engineering and Physical Sciences Research Council, UK, with additional financial contributions from the National Radiological Protection Board and the Environment Agency. We thank Neil Higgins, Tom Charnock and their colleagues at the National Radiological Protection Board for providing us with the data, and for advice on the use of the Gaussian plume model.

REFERENCES

Abramowitz, M. and Stegun, I. (1965) *Handbook of Mathematical Functions*, 9th edition. New York: Dover.

Cressie, N. A. C. (1991) *Statistics for Spatial Data*. New York: Wiley.

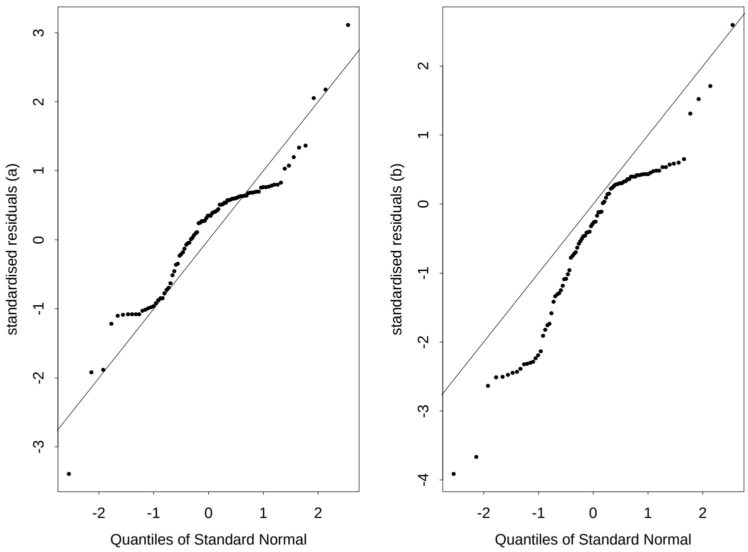

Fig. 5. Quantile-Quantile plots for residuals using (a) Bayesian calibration and model inadequacy correction; (b) simple interpolation of the data

Evans, M. and Schwartz, T. (1995) Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems (with discussion). *Statist. Sci.*, **10**, 254–272.

Handcock, M. S. and Wallis, J. R. (1994) An approach to statistical spatial-temporal modelling of meteorological fields (with discussion) *J. Amer. Statist. Assoc.*, **89**, 368–390.

Kennedy, M. C. and O'Hagan, A. (2000) Bayesian Calibration of Computer models. Submitted to *J. Roy. Stat. Soc.*.

Naylor, J.C. and Smith, A.F.M. (1982) Applications of a method for the efficient computation of posterior distributions, *Appl. Statist.*, **31**, 214–225.

Oakley, J. E. and O'Hagan, A. (1998) Bayesian inference for the uncertainty distribution. Tech. Rep. University of Nottingham, Statistics Section.

O'Hagan, A., Kennedy, M. C. and Oakley, J. E. (1999) Uncertainty analysis and other inference tools for complex computer codes. In: *Bayesian Statistics 6*, (J. M. Bernardo, J. O. Berger, A. P. Dawid and A.F.M. Smith, eds.), 503–524 (with Discussion). Oxford: University Press.

Vecchia, A. V. (1988). Estimation and identification for continuous spatial processes. *J. Roy. Statist. Soc. B*, **50**, 297–312.