

# Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation

Anna Heath,<sup>\*,†</sup> Ioanna Manolopoulou and Gianluca Baio

The Expected Value of Perfect Partial Information (EVPPPI) is a decision-theoretic measure of the ‘cost’ of parametric uncertainty in decision making used principally in health economic decision making. Despite this decision-theoretic grounding, the uptake of EVPPPI calculations in practice has been slow. This is in part due to the prohibitive computational time required to estimate the EVPPPI via Monte Carlo simulations. However, recent developments have demonstrated that the EVPPPI can be estimated by non-parametric regression methods, which have significantly decreased the computation time required to approximate the EVPPPI. Under certain circumstances, high-dimensional Gaussian Process (GP) regression is suggested, but this can still be prohibitively expensive. Applying fast computation methods developed in spatial statistics using Integrated Nested Laplace Approximations (INLA) and projecting from a high-dimensional into a low-dimensional input space allows us to decrease the computation time for fitting these high-dimensional GP, often substantially. We demonstrate that the EVPPPI calculated using our method for GP regression is in line with the standard GP regression method and that despite the apparent methodological complexity of this new method, R functions are available in the package BCEA to implement it simply and efficiently. © 2016 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

**Keywords:** SPDE-INLA; Gaussian Process regression; Health economic evaluation; Value of information

## 1. Introduction

Broadly speaking, the objective of publicly funded health care systems, such as the UK National Health Service, is to maximise health gains across the general population, given finite monetary resources and limited budget. Bodies such as the National Institute for Health and Care Excellence (NICE) in the UK provide guidance on decision-making on the basis of health economic evaluation. This covers a suite of analytic approaches for combining costs and clinical consequences of an intervention, in comparison to alternative options which may already be available, with the aim of aiding decision-making associated with health resources. Much of the recent research has been oriented towards building the health economic evaluation on sound and advanced statistical decision-theoretic foundations, arguably making it a branch of applied statistics [1, 2] and increasingly often under a Bayesian approach [3–6].

In a nutshell, the process involves the identification of suitable measures of clinical benefits (generically termed as ‘effectiveness’) and costs associated with an intervention, which we indicate as  $(e, c)$ . The variable  $c$  usually includes the cost of acquisition and implementation of the health intervention (e.g. a drug), or societal costs such as those related to number of days off work or social care. As for the clinical

Department of Statistical Science, University College London, Department of Statistical Science, University College London, U.K.

\*Correspondence to: Anna Heath, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, U.K.

†E-mail: [anna.heath.14@ucl.ac.uk](mailto:anna.heath.14@ucl.ac.uk)

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

benefits  $e$ , they can be a ‘hard’ measurement (e.g. number of cases averted), but most often are considered in terms of *Quality Adjusted Life Years* (QALYs) [7], combining the quantity and the quality of life provided by a given intervention. Individual level variability in the outcome is expressed in terms of a joint probability distribution  $p(e, c | \theta)$ , indexed by a set of parameters  $\theta$  whose uncertainty is described by a probability distribution  $p(\theta)$ , in a Bayesian context.

According to the precepts of decision theory [8], for each intervention  $t = 0, \dots, T$  ( $t = 0$  being the ‘reference’ alternative, e.g. standard of care and  $t = 1, \dots, T$  the new options being evaluated) the health economic outcomes  $(e, c)$  are combined in order to quantify the overall ‘value’ of the intervention, usually in terms of the monetary net benefit, which is given by the costs associated with the treatments subtracted from rescaled effectiveness (we define this formally in §2). The alternative associated with the highest expected net benefit is deemed as ‘the most cost-effective’, given current evidence — notice that in a Bayesian context, this expectation is taken over the distributions of both the individual level outcomes and population level parameters. From the decision-theoretic point of view, the identification of the overall expected net benefit is all that is needed to reach the optimal decision given the current state of knowledge available to the decision-maker [9, 10].

However, the implementation of a health care intervention is usually associated with risks such as the irreversibility of investments [11]. Moreover, health economic models often involve a relatively large number of parameters, usually estimated using limited information. For these reasons, health technology assessment (HTA) bodies such as National Institute for Health and Care Excellence recommend a thorough investigation of the impact of uncertainty on the decision making process of parametric and model uncertainty, a process known in the health economics literature as *Probabilistic Sensitivity Analysis* (PSA) [12–15].

The analysis of the *value of information* (VoI) [16] is an increasingly popular method to conduct PSA in health economic evaluations [1, 11, 17–22]. The basic idea of VoI analysis is to compare the decision based on current evidence to the one that would be made, had the uncertainty in the parameters been resolved, for example, by observing an (infinitely) large amount of data.

The main advantage of the analysis of the VoI is that it directly addresses the potential implications of current uncertainty, not only in terms of the likelihood of modifying the current decision in light of new and more definitive evidence, but also in terms of the opportunity cost of the incorrect decision. If this cost is low, there is little value in agonising about the decision, even for a low probability of cost-effectiveness, as the implicit penalty is negligible if the decision turns out to be wrong in the face of new evidence. Therefore, the probability of cost-effectiveness alone can dramatically overstate or down-play the decision sensitivity. For this reason, it has been advocated that VoI measures should also be presented when representing decision uncertainty [6, 17, 23, 24].

Despite the useful features of a VoI analysis, its uptake in health economic evaluation has been slow. VoI analysis has been hindered by several different factors, both theoretical and practical. Theoretically, issues include whether the economic model can be trusted and what constitutes a ‘high’ level of decision uncertainty given that VoI measures are unbounded. Practically, VoI measurements can be computationally costly unless restrictive assumptions are made [25]. It is typically easy to calculate numerically (but not analytically) the expected value of learning *all* the model parameters perfectly. This is known as the (overall) Expected Value of Perfect Information (EVPI). However, this quantity has often little practical use, as it will be rarely possible to learn all the model parameters at once in a new study. Thus, the decision-maker is usually interested in the expected VoI about subsets of parameters of interest, sometimes called focal parameters. This will indicate which parameters are driving decision uncertainty, and thus, where future research may add value. This subset analysis is concerned with the computation of the *Expected Value of Perfect Partial Information* (EVPPI) and is usually computationally costly.

In the last few years, research has focused on alternative methods that could be used to speed up the estimation of the EVPPI without compromising its accuracy, so as to increase its applicability in health economic evaluations [26–28]. A review of these methods is given in [29]. A promising development [30] has recently explored the use of non-parametric regression methods, specifically Generalised Additive Models (GAMs) [31] and Gaussian Process (GP) regression [32], to approximate the EVPPI. These methods have been shown to be very effective: GAMs are very flexible and extremely inexpensive in computational terms if the number of parameters of interest  $P$  is relatively small (e.g.  $P < 5$ ). When  $P$  is large, however, GAMs either overfit or simply cannot be used as the number of parameters exceeds the number of data points. GP regression methods may be used as an alternative regression method: on the one hand, they overcome the limitations of GAMs and can be used to estimate the EVPPI for higher dimensions of the subset of parameters of interest, producing a significant improvement over simpler,

but computationally prohibitive methods. On the other hand, however, the computational cost to fit a GP regression model to a realistic problem including a relatively large number of parameters can still be substantial. Moreover, as EVPPI values may need to be calculated for a large number of different parameter sets, this computational time can be multiplied 10-fold or 100-fold.

To overcome this issue, we propose in this paper a faster method to fit GP regression, based on spatial statistics and Bayesian inference using INLA [33]. We translate the estimation of the EVPPI into a ‘spatial’ problem by considering that the simulated net benefit values, for example, the values from the PSA, are ‘observed’ at different points in the parameter space. We can therefore use the available technology for fast Bayesian computation of spatial models [34, 35] to approximate the EVPPI efficiently. Furthermore, as this spatial machinery loses its computational advantages in higher dimensional spaces, we demonstrate that the use of dimension-reduction techniques to project onto a 2-dimensional space is accurate for the examples considered. Thus, we can use this method along with dimensionality reduction to approximate the EVPPI quickly and efficiently in applied cases, irrespective of the complexity of the problem.

The paper is structured as follows: in §2, we present the general framework used for the analysis of the value of information in health economic evaluation and in §2.1, we briefly review the main characteristics of GPs and specifically their application in the computation of the EVPPI. Then, in §3, we present our proposal for a new method to compute the EVPPI; first we briefly review the main features of the spatial statistics literature based on stochastic partial differential equations (described in §3.1), which is used to estimate the correlation function required to fit the GP. Then, in §3.2, we discuss how this can be brought to bear in the modelling and efficient computation of the EVPPI. In §4, we test our method in comparison to existing GP regression models to estimate the EVPPI on a set of health economic examples. We particularly focus on the issues of computational time as well as accuracy of the estimation. Finally, in §5 and §6, we present some technical aspects as well as the main conclusions from our work.

## 2. Value of information analysis in health economics

PSA is usually based on a simulation approach [6, 36, 37]: uncertainty about the relevant parameters  $\theta$  is described by a suitable probability distribution, from which a sample of  $S$  values is obtained, for example, via Monte Carlo (MC) sampling from the prior or Markov Chain MC (MCMC) estimation of the posteriors under a Bayesian framework, or using bootstrap in a frequentist approach. First, for each intervention, the expected value is computed conditionally on each value of the simulated parameters. Assuming the commonly used *monetary net benefit* [38] to quantify the value of the different treatments, this value is estimated by

$$NB_i(\theta_s) = kE [e | \theta_s; t] - E [c | \theta_s; t],$$

where  $\theta_s$  is the  $s$ -th set of simulated values for the parameters vector,  $e$  and  $c$  are the measures of effectiveness and cost, respectively, and  $k$  is the *willingness to pay*, which is used to put the cost and effectiveness measures on the same scale, that is, in terms of the amount of money that the decision maker is willing to pay to increment the benefits by one unit. Notice here that  $(e, c)$  represent the individual level effectiveness and costs for a specific value of  $\theta$  and conditional on any observed data. Therefore, the expectation is taken over the joint distribution of  $(e, c)$  to give the population level net benefit for the specific value of  $\theta$ .

The vector of values  $\mathbf{NB}_t = [NB_t(\theta_1), \dots, NB_t(\theta_S)]'$  is a sample from the distribution of the decisions (randomness being induced by uncertainty in the parameters) and can be analysed to determine the impact of parameter uncertainty on the decision-making process. If the optimal decision, that is, the intervention with the maximum expected net benefit, varies substantially across the simulations, then the decision-making process is sensitive to the uncertainty in the model parameters and more research could be recommended by the health technology assessment bodies.

For example, consider a non life-threatening infectious disease (such as mild influenza) and assume that under current practice ( $t = 0$ ), individuals have a risk of infection  $\pi$ . If they become infected, they are subject to an average duration of the disease of  $\lambda$  days, for each of which they have to follow a course of treatment that costs  $\gamma$  monetary units (say, £). The new treatment, whose cost-effectiveness is being assessed ( $t = 1$ ), has an average implementation cost of £ $\xi$  and it is assumed that the chance of infection is reduced by a factor  $\rho$ . However, if an individual becomes infected then they will still have a period of  $\lambda$  days in which they will require treatment at the cost of £ $\gamma$  per day. Under these (unrealistically!) simple assumptions, we can define  $\theta = (\pi, \lambda, \gamma, \xi, \rho)$ . Assuming further that suitable probability distributions

can be associated with each of the elements of  $\theta$  to describe our current uncertainty, we could reasonably define the net benefits for the two interventions as

$$NB_0(\theta) = k(-\pi\lambda) - \pi\gamma\lambda \quad \text{and} \quad NB_1(\theta) = k(-\pi\rho\lambda) - (\xi + \pi\rho\gamma\lambda),$$

implying that the clinical benefit is given by minus the chance of being infected multiplied by the length of the infection and that cost is given by the sum of the implementation cost of the intervention and the overall cost of treatment if infected.

In general terms, the expected opportunity loss of making a decision based on current evidence instead of on perfect information can be quantified by the *Expected Value of Perfect Information*, defined as

$$EVPI = E_{\theta} \left[ \max_t NB_t(\theta) \right] - \max_t E_{\theta} [NB_t(\theta)], \quad (1)$$

where the expectation is taken with respect to the posterior distribution of  $\theta$ .

Because these expectations are typically not analytically available, they are estimated through simulations. Provided the number  $S$  of simulations used to perform PSA is large enough to characterise the underlying distribution of the decisions, it is straightforward to compute a MC estimate using the simulated values for the net benefits

$$\widehat{EVPI} = \frac{1}{S} \sum_{s=1}^S \max_t NB_t(\theta_s) - \max_t \frac{1}{S} \sum_{s=1}^S NB_t(\theta_s),$$

which usually requires almost no extra computational time, once the PSA samples are available.

In most practical situations, however, the interest is in quantifying the value of reducing uncertainty on a specific subset  $\phi \subset \theta$  of parameters of interest as it may be feasible to conduct a specific clinical trial or literature review in order to potentially reduce the level of current uncertainty in these parameters. For example, for the simple model described earlier, we may be interested in learning the value of investigating the epidemiological parameters  $\phi = (\pi, \lambda, \rho)$  describing the risk and duration of the infection and the effectiveness of the intervention in reducing it, while considering the remaining parameters  $\psi = (\gamma, \xi)$  as nuisance.

This value is known as the Expected Value of *Partial* Perfect Information and in the general setting is defined as

$$EVPPI = E_{\phi} \left[ \max_t E_{\psi|\phi} [NB_t(\phi, \psi)] \right] - \max_t E_{\theta} [NB_t(\phi, \psi)], \quad (2)$$

where  $\theta = (\phi, \psi)$ . The last term in equation (2) is again the maximum expected net benefit under current evidence. The first term is made by two nested elements: the inner part is the maximum expected net benefit that would be obtained if uncertainty in the parameters of interest *only* were resolved. This means that the inner expectation assumes that the value of  $\phi$  is known. Of course, as the ‘true’ value of  $\phi$  is not known, it is necessary to consider the expectation over the current distribution of  $\phi$ .

In simple cases, where the conditional expectation  $E_{\psi|\phi} [NB_t(\phi, \psi)]$  is available analytically as a function of  $\phi$ , then it is possible to calculate the EVPPI using a single sample from  $p(\phi)$ . This can occur in several settings, the most common of which is known as the ‘sum-product’ form for the net benefit. This allows us to calculate the EVPPI based on samples that have already been obtained for PSA as part of a standard health economic analysis.

A more general solution, which involves additional sampling, is to use a nested MC scheme, in which first a sample  $\phi_1, \dots, \phi_{S_{\phi}}$  is obtained from the marginal distribution of  $\phi$  and then, for each  $s = 1, \dots, S_{\phi}$ , a sample  $\psi_1, \dots, \psi_{S_{\psi}}$  from the conditional distribution  $p(\psi | \phi_s)$  is also simulated. This produces a total of  $S_{\phi} \times S_{\psi}$  simulations, where both numbers need to be large in order to reduce the MC error such that the EVPPI estimates are suitably precise; for example, Brennan *et al.* [39] suggest that  $S_{\phi}$  and  $S_{\psi}$  should be in the order of 10 000, although this may need to be higher in complex models. This immense computational burden and the difficulty in deriving analytic results in practical scenarios have been arguably the main reasons for the relatively limited practical use of the EVPPI as a tool for PSA [40, 41].

### 2.1. Gaussian Process regression for efficient computation of the EVPPI

Gaussian Processes are a family of stochastic processes used in statistics and machine learning for non-parametric regression, classification and prediction [32, 42] and can be thought of as an extension of the

multivariate Normal distribution to an infinite vector of observations [32, 43]. Strictly speaking, a GP is an infinite collection of random variables, any subset of which follows a multivariate Gaussian distribution [44]. A GP is entirely defined in terms of its mean and covariance *functions* [45, 46], which calculate the mean vector and covariance matrix for each subset of random variables depending on some input values and a small set of hyperparameters. These inputs determine the specific mean and variance for each random variable in the process. Consequently, GPs can be used for regressing random variables on a set of input values.

To fit a GP for non-parametric regression, the general form of the mean and covariance function is specified by the modeller. In general, the covariance function is taken as a decreasing function of the ‘distance’ between any two input values, that is, points that are ‘closer’ have a higher correlation [32, 47] where ‘distance’ and ‘closeness’ can be measured in different ways depending on the context. These functions typically depend on a set of hyperparameters; for example, the covariance function is often defined in terms of a smoothness parameter that determines the similarity between two points ‘close’ together and a GP marginal variance parameter. Once these general functions are specified, problem-specific values for the hyperparameters need to be determined.

In a Bayesian setting, vague and conjugate priors have been proposed for the hyperparameters allowing for partially analytically tractable posterior distributions [30, 48]. Integration and numerical optimisation can then be used to find maximum a posteriori estimates of the GP parameters. Therefore, GPs are an increasingly popular method of regression because their extreme flexibility typically is obtained at a relatively small computational cost as MCMC methods may be avoided to fit them. However, for large datasets, the cost of fitting a GP is still substantial as numerical optimisation in this setting requires inverting an  $S \times S$  dense matrix, at a computational cost of  $\mathcal{O}(S^3)$ .

The basic idea exploited by Strong *et al.* [30] is to consider the net benefit of each treatment  $t$  computed using the  $s$ -th set of simulated values of the parameters as a noisy observation of the ‘true’ underlying conditional expectation

$$NB_t(\theta_s) = E_{\psi|\phi_s} [NB_t(\theta)] + \varepsilon_s, \tag{3}$$

with  $\varepsilon_s \stackrel{iid}{\sim} \text{Normal}(0, \sigma_\varepsilon^2)$  and assuming conditional independence between the net benefits under the different treatments  $t$ . As the conditional expectation on the right hand side changes as a function of the parameters of interest only, we can equivalently write (3) as

$$NB_t(\theta_s) = g_t(\phi_s) + \varepsilon_s. \tag{4}$$

Once the functions  $g_t(\cdot)$  have been estimated using GP regression methods, the fitted values  $\hat{g}_t(\phi_s)$  can be used to approximate the EVPPi by computing

$$\widehat{\text{EVPPi}} = \frac{1}{S} \sum_{s=1}^S \max_t \hat{g}_t(\phi_s) - \max_t \frac{1}{S} \sum_{s=1}^S \hat{g}_t(\phi_s).$$

Assuming a GP structure for the functions  $g_t(\cdot)$  in a linear regression framework effectively amounts to modelling

$$\begin{pmatrix} g_t(\phi_1) \\ g_t(\phi_2) \\ \vdots \\ g_t(\phi_S) \end{pmatrix} \sim \text{Normal}(\mathbf{H}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{5}$$

where:  $\phi_s$  is the  $s$ -th simulated value for  $\boldsymbol{\phi}$ ;  $\mathbf{H}$  is a design matrix

$$\mathbf{H} = \begin{pmatrix} 1 & \phi_{11} & \cdots & \phi_{1P} \\ 1 & \phi_{21} & \cdots & \phi_{2P} \\ \vdots & & \ddots & \\ 1 & \phi_{S1} & \cdots & \phi_{SP} \end{pmatrix}; \tag{6}$$

$\boldsymbol{\beta}$  is the vector of regression coefficients describing the linear relationship between the parameters of interest  $\boldsymbol{\phi}$  and the conditional expectation of the net benefits; and the covariance matrix  $\boldsymbol{\Sigma}$  is determined

by the covariance function  $\mathcal{C}$ , a matrix operator whose elements  $\mathcal{C}(r, s)$  describe the covariance between any two points  $g_t(\boldsymbol{\phi}_r)$  and  $g_t(\boldsymbol{\phi}_s)$ .

Strong *et al.* [30] use a squared exponential, also known as an exponentiated quadratic, covariance function  $\mathcal{C}_{\text{Exp}}$ , defined by

$$\mathcal{C}_{\text{Exp}}(r, s) = \sigma^2 \exp \left[ - \sum_{p=1}^P \left( \frac{\phi_{rp} - \phi_{sp}}{\delta_p} \right)^2 \right], \quad (7)$$

where  $\phi_{rp}$  and  $\phi_{sp}$  are the  $r$ -th and the  $s$ -th simulated value of the  $p$ -th parameter in  $\boldsymbol{\phi}$ , respectively. For this covariance function,  $\sigma^2$  is the GP marginal variance and  $\delta_p$  defines the smoothness of the relationship between two values that are ‘close together’ in dimension  $p$ . For high values of  $\delta_p$  the correlation between the two conditional expectations with similar values for  $\phi_p$  is small. The  $\delta_p$  values are also treated as hyperparameters to be estimated from the data.

Combining equations (4) and (5), we can directly model the ‘observed’ vector of net benefits as

$$\begin{pmatrix} \text{NB}_t(\theta_1) \\ \text{NB}_t(\theta_2) \\ \vdots \\ \text{NB}_t(\theta_S) \end{pmatrix} \sim \text{Normal}(\mathbf{H}\boldsymbol{\beta}, \mathcal{C}_{\text{Exp}} + \sigma_\varepsilon^2 \mathbf{I}). \quad (8)$$

The model in (8) includes  $2P + 3$  hyperparameters: the  $P + 1$  regression coefficients  $\boldsymbol{\beta}$ , the  $P$  smoothness parameters  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_P)$ , the marginal variance of the GP  $\sigma^2$  and the residual error  $\sigma_\varepsilon^2$ , also known as ‘nugget variance’. In this setting, therefore, the PSA samples for  $\boldsymbol{\phi}$  are the ‘covariates’ used to fit the non-parametric regression model, while the ‘response’ is represented by the net benefits. Given that the estimation of the hyperparameters is the most expensive component of fitting a GP [30], in computational terms the efficiency of this method to estimate the EVPPI depends on the number of parameters of interest.

Strong *et al.* [30] integrate out  $\boldsymbol{\beta}$  and  $\sigma$  analytically and use numerical optimisation to calculate the posterior mode of the other hyperparameters  $\boldsymbol{\delta}$  and  $\sigma_\varepsilon$  analytically. This allows for great flexibility but at a computational cost in the order of  $S^3$ . As PSA is usually based on relatively large number of simulated values from the posterior distributions of  $\boldsymbol{\theta}$  (i.e. in the thousands) this procedure still takes considerable computational effort despite producing a significant improvement over MC methods, especially for larger numbers of parameter of interest.

### 3. Fast computation of the EVPPI using Integrated Nested Laplace Approximation

#### 3.1. Spatial statistics and Stochastic Partial Differential Equations

An interesting application of GP regression is in the field of spatial statistics, where measurements are taken at different points in a spatial domain. For example, these can be the cases of influenza at locations in a geographical area (e.g. a country) or the level of pollution at different monitoring sites. The main assumption in spatial statistics is that points that are ‘closer’ to each other in a geographical sense share more common features and are influenced by common factors than those ‘further away’ [49].

A very popular specification of a spatial model when exact locations are available is based on the Matérn family of covariance functions [50], defined by

$$\mathcal{C}_M(r, s) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} (\kappa \|\boldsymbol{\phi}_r - \boldsymbol{\phi}_s\|)^\nu \text{K}_\nu(\kappa \|\boldsymbol{\phi}_r - \boldsymbol{\phi}_s\|),$$

where  $\boldsymbol{\xi} = (\sigma, \kappa, \nu)$  is a vector of hyperparameters,  $\|\cdot\|$  denotes the Euclidean distance and  $\text{K}_\nu$  is the modified Bessel function of the second kind and order  $\nu$ . The Matérn covariance function is related to the covariance function in (7), which can be obtained when  $\delta_p$  is constant for all  $p = 1, \dots, P$  and  $\nu \rightarrow \infty$  [32]. This implies that the resulting covariance matrix for a specific set of input values is still dense, that is, a large number of the matrix elements are non-zero, which in turn generates a computational cost for matrix inversion in the order of  $S^3$ .

However, Lindgren *et al.* [51] demonstrate that a sparse matrix can be used to approximate a GP with a Matérn covariance function (Matérn GP) by using Stochastic Partial Differential Equations (SPDE), which in general leads to a computation time for matrix inversion in the order of  $S^{\frac{3}{2}}$ . It can be shown

that, in addition to being defined in terms of a relationship with the multivariate Gaussian, a Matérn GP is also exactly equal to the function  $g_t(\boldsymbol{\phi})$  that solves a specific SPDE of the form

$$\tau(\kappa^2 - \Delta)^{\frac{\alpha}{2}} g_t(\boldsymbol{\phi}) = \mathcal{W}(\boldsymbol{\phi}),$$

where  $\mathcal{W}$  is Gaussian noise,  $\Delta$  is the Laplacian operator,  $\alpha = \nu + \frac{P}{2}$  (with  $P = 2$ , in the spatial context) and the marginal variance of the Matérn GP is

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{\frac{P}{2}}} \kappa^{-2\nu} \tau^{-2}.$$

Therefore, finding the solution to this SPDE is exactly equivalent to finding the function  $g_t(\boldsymbol{\phi})$ , which as mentioned in §2.1 is instrumental in estimating the EVPPI.

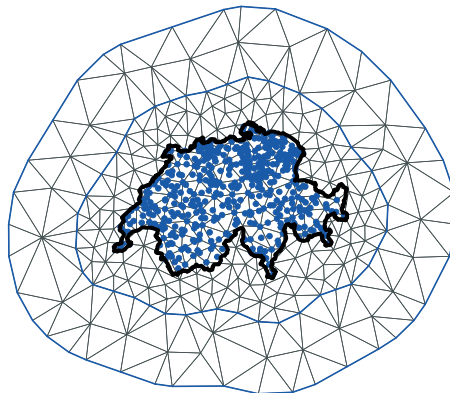
The fundamental implication of this result is that efficient algorithms for solving SPDEs can be used to approximate the Matérn GP. In practice, the SPDE is solved using the finite element method [52]. First, the region over which the SPDE is being solved, that is, the range of  $\boldsymbol{\phi}$ , is split into small areas. In the proper spatial 2-dimensional case, a grid of small triangles is used; an example of this triangulation relating to the amount of rainfall in Switzerland over a pre-specified time horizon is shown in Figure 1. Notice that there are two triangulation boundaries with the border of Switzerland added for clarity. An inner boundary encases (or ‘hugs’) all the data points relatively tightly, while the outer boundary is further from the points. This is because boundary conditions are imposed on the SPDE solver and it is important to avoid these impacting on the estimation of the smooth function  $g_t(\cdot)$ . The value of the Matérn GP is then approximated by simple (linear) functions within each small triangular area. Therefore, the triangles are more tightly packed within the inner boundary to give a good approximation to the Matérn GP, while in the outer region the grid can be rougher as the approximation is not as important.

The Matérn GP is directly approximated at each intersection, where triangles meet (i.e. the vertices). Then, within each triangle it is approximated using linear interpolation by taking a weighted sum of the weights at the three corners of the triangle. Therefore, the vertex weights  $\boldsymbol{\omega}$ , which are estimated from the data (i.e. the net benefit values, in this case), entirely determine the value of the Matérn GP. Lindgren *et al.* [51] demonstrate that these vertex weights can be assumed to have a multivariate Gaussian distribution with a specific precision matrix, conditional on  $\kappa$  and  $\tau$ . This precision matrix is, to a very good degree of approximation, sparse as non-zero entries correspond loosely only with points that ‘share a triangle’.

### 3.2. Computing the EVPPI using SPDE-INLA

Assuming a Matérn covariance function, the model in (8) becomes  $\mathbf{NB}_t \sim \text{Normal}(\mathbf{H}\boldsymbol{\beta}, \mathbf{C}_M + \sigma_\epsilon^2 \mathbf{I})$ , which can be equivalently re-expressed as

$$\mathbf{NB}_t \sim \text{Normal}(\mathbf{H}\boldsymbol{\beta} + F(\boldsymbol{\omega}), \sigma_\epsilon^2 \mathbf{I}), \tag{9}$$



**Figure 1.** An example of the grid approximation used to approximate the Matérn GP in a proper spatial problem. The thick black line represent the border of Switzerland. The blue dots represent the positions where data points have been observed. These data points are used to estimate the value of the Matérn GP throughout the geographical space (i.e. the whole area covered by Switzerland, in this case).

where  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_S)' \sim \text{Normal}(\mathbf{0}, \boldsymbol{Q}^{-1}(\tau, \kappa))$  are the vertex weights and  $\boldsymbol{Q}(\tau, \kappa)$  is the sparse precision matrix determined by the SPDE solution. The function  $F(\cdot)$  maps from the position of the  $\boldsymbol{\omega}$  values to the position of the data points on the grid. As is possible to see in Figure 1, many of the points do not lie exactly on the intersections between the triangles and are therefore calculated as a function of several  $\boldsymbol{\omega}$  values.

Interestingly, the specification in (9) is in fact a *Latent Gaussian Model* [53] as the response of interest  $\mathbf{NB}_i$  is dependent on a latent Gaussian vector  $\boldsymbol{\omega}$  controlled by a small number of hyperparameters ( $\kappa$  and  $\tau$ ). This means that inference can be performed in a very efficient way by using the Integrated Nested Laplace Approximation (INLA) algorithm [53] (cf. Appendix A), programmed in the R package R-INLA [54], which also includes extensions to implement the SPDE method [34, 51, 55–58].

The SPDE-INLA method has been developed and successfully applied in a spatial context [57, 59, 60], where inputs are proper coordinates (i.e. longitude and latitude, hence defined in a 2-dimensional space) which makes the GP approximation extremely fast. However, calculating the EVPPI relies on a set of much higher dimensional inputs. While in theory the SPDE machinery works in higher dimensional spaces, the computational advantages will diminish in these cases.

To fully exploit the computational savings of the SPDE-INLA procedure, we re-express the problem of computing the EVPPI in a ‘spatial’ context. In this case, the simulated parameter vector for  $\boldsymbol{\phi}$  designates a point in the  $P$ -dimensional parameter space. We consider that the net benefit, calculated as a function of  $\boldsymbol{\phi}$ , has been ‘observed’ at this point. We then wish to find a representation of these  $P$ -dimensional points in at most 2-dimensional space, so that we can efficiently estimate the Matérn covariance of this representation using the SPDE-INLA methodology. If no such projection exists, then the computational complexity of fitting a Matérn GP with a high dimensional SPDE means that the standard GP method is more appropriate.

As this projection will be used to predict the net benefit values, it makes sense to use a regression-based dimension reduction method. This class of methods tries to find a *sufficient* reduction, that is, one for which the projection contains the relevant information about the net benefit function. Formally, we can express this condition as  $\mathbf{NB}_i \perp\!\!\!\perp \boldsymbol{\phi} \mid R(\boldsymbol{\phi})$ , where  $R(\cdot)$  is the reduction function from  $P$ , the length of  $\boldsymbol{\phi}$ , to  $d$ , the number of dimensions needed to capture all the information in  $\boldsymbol{\phi}$ . There is a wealth of methods that can be used to estimate this sufficient reduction [61–66]. Specifically, we focus on Principal Fitted Components (PFC) [65, 66] to calculate the EVPPI.

**3.2.1. Principal Fitted Components.** Principal Fitted Components is a model based inverse regression method. This means that in order to find a sufficient reduction, we consider a model for  $\boldsymbol{\phi}$  as a function of  $\mathbf{NB}_i$ . As different models can be specified and lead to different sufficient reductions, the best fit model amongst a set of candidates should be chosen before finding the sufficient reduction.

The general form of these PFC models is based on a linear structure

$$\boldsymbol{\phi} = \boldsymbol{\mu} + \mathbf{Y}f(\mathbf{NB}_i) + \boldsymbol{\epsilon},$$

where  $\boldsymbol{\mu}$  is the intercept;  $\mathbf{Y}$  is a  $P \times d$  matrix to be estimated to determine the sufficient reduction;  $f(\cdot)$  is a vector-valued function of  $\mathbf{NB}_i$ ; and  $\boldsymbol{\epsilon}$  is an error term in the exponential family. The form of the error structure changes the way in which the reduction is calculated and methods have been developed for independent, heteroskedastic and unstructured errors.

In order to use PFC, the function  $f(\cdot)$  and the error structure need to be specified. In our problem, we consider normally distributed errors and set  $f(\mathbf{NB}_i) = [\alpha_1 \mathbf{NB}_i, \alpha_2 \mathbf{NB}_i^2, \dots, \alpha_h \mathbf{NB}_i^h]^T$ , although in general  $f(\cdot)$  can map to any function of  $\mathbf{NB}_i$ . Additionally, the number of dimensions  $d$  needed to capture all the relevant information in  $\boldsymbol{\phi}$  needs to be specified. It is then advisable to select the values of  $d$  and  $h$  (the polynomial degree) associated with the best performing inverse regression specification, for example, in terms of model fitting, as measured by information criteria such as the AIC [66].

The cost of fitting any of these models individually is negligible and thus fitting a number of models in correspondence to a set of chosen  $d$  and  $h$  values adds very little to the computational time required to estimate the EVPPI. In any case, because PFC assumes that the number of dimensions needed to capture the information in  $\boldsymbol{\phi}$  is no larger than the number of dimensions in the function  $f(\cdot)$ , for simple relationships between the net benefit and  $\boldsymbol{\phi}$ , the sufficient regression is low dimensional.

To fully exploit the spatial nature of the SPDE-INLA methodology,  $d$  must be set to two. However, identifying the optimal value for  $d$  allows us to check the validity of the EVPPI approximation. If one



dimension is sufficient to capture all the information in  $\phi$ , there is no harm in using a second component because this will only add information and the reduction will remain sufficient. On the other hand, using two dimensions when the AIC suggests  $d > 2$ , may lead to a loss in information. Consequently, the EVPPI estimate based on a two-dimensional reduction of  $\phi$  may be biased. In light of the large computational savings and the fact that the AIC has a tendency to overestimate  $d$  [66], it may still be worth using the projection to estimate the EVPPI and then perform thorough model checking (e.g. by means of the residual plots) to assess its performance, before resorting to more computationally intensive methods. We return to this point in §4.

From the theoretical point of view, PFC provides a robust method for determining the sufficient reduction [66]. Thus, combining PFC and SPDE-INLA to estimate the EVPPI seems to be a valid strategy. Additionally, due to the flexibility of the INLA algorithm it is possible to cater for more complicated structures in the relationships between the net benefit and the parameters of interest, as we discuss in §5.

We have combined PFC and SPDE-INLA, along with some simple PFC model selection, in the R package BCEA [67, 68] to allow users to integrate standard economic analysis with efficient calculations for the EVPPI in large dimensions. This function relies on the R-INLA and ldr packages [54, 69]. We have also implemented model checking procedures for the non-parametric regression as standard in order to aid practitioners. This potentially improves the use of value of information analysis as a tool for PSA in applied health economic problems.

## 4. Examples

We present two case studies of health economic models and compare the estimates of the EVPPI using the direct GP regression implemented by Strong *et al.* and our SPDE-INLA projection method. For both case studies, random subsets of between 5 and 16 parameters of interest were considered to compare the performance of the GP procedures — notice that this represents the standard range of parameter subsets that would be used practically for EVPPI calculation using GP [30, 70]. For each subset, the EVPPI was calculated using both methods and for a willingness-to-pay threshold of  $k = 20000$  monetary units, say £. The computational time and EVPPI estimate was then recorded for both methods to allow a direct comparison.

### 4.1. Vaccine Study

The first case study (referred to as the ‘Vaccine study’) is a Bayesian health-economic model proposed to analyse the effect of an influenza vaccine on health outcomes and costs. A more detailed description of the example is presented in [6]. The parameters are sampled from their joint posterior distribution using Markov Chain Monte Carlo (MCMC) methods. These sampled parameter values are used to calculate the net benefits and the EVPPI.

Two treatment options are considered, either the vaccine is available to the population ( $t = 1$ ) or not ( $t = 0$ ). If an individual gets influenza, they are treated with anti-viral drugs and will often visit the doctor. Complications may occur, including pneumonia and hospitalisation, in which case there will be some indirect costs such as time off work. The cost of the treatment is the acquisition cost of the drugs, the time in hospital, the doctor’s visits and the cost of the vaccine. The benefit of the treatment is measured in QALYs, to which each adverse effect contributes negatively.

The Vaccine model includes 28 key parameters representing the probability of infection, the reduction in risk due to the vaccine, the occurrence of complications, the monetary costs of the interventions and the QALY loss due to different health states. However, as the model is built as an evidence synthesis, additional sources of uncertainty are present; for instance, the true number of people getting influenza or the true number of people getting side effects are unknown. Considering all the unobserved quantities in the model, the number of parameters increases to 62.

### 4.2. SAVI Study

The second case study is a simple fictional decision tree model with correlated parameters, presented at the Sheffield Accelerated Value of Information web app [71] (hence this example is referred to as the ‘SAVI study’). The model has two treatment options and 19 underlying parameters. A more in-depth model description is presented in [39]. Most importantly, the 19 underlying parameters are assumed to follow a multivariate Gaussian distribution and thus the conditional distribution  $p(\psi | \phi)$  can be computed analytically and the EVPPI can therefore be calculated using MC simulation. The SAVI web app

**Table I.** The computational time required (in seconds) to calculate an EVPPI using both the GP regression method and SPDE-INLA method for increasing numbers of parameters for both case studies.

Number of parameters of interest	Computation time (seconds)			
	Vaccine Example		SAVI Example	
	GP	SPDE-INLA	GP	SPDE-INLA
5	17	9	17	7
6	42	10	14	7
7	45	10	18	7
8	57	11	21	8
9	74	8	26	8
10	86	8	31	9
11	70	7	37	8
12	60	8	47	8
13	84	11	52	7
14	188	8	66	6
15	470	7	70	7
16	121	8	71	7

provides 10 000 PSA samples of all the 19 parameters, along with the simulated costs and benefits of both treatment options. The number of available PSA samples poses a significant challenge for the standard GP regression method. For this reason, only the first 1 000 observations are used for the comparison with our SPDE-INLA method.

#### 4.3. Computational Time

We begin our discussion of the two EVPPI estimation methods by comparing the computational time required to obtain an estimate. The EVPPI estimates were calculated using 1 000 PSA samples for both case studies and both methods. To compare our SPDE-INLA method we used the code available from Strong [72] with a slight modification. This modification changed the numerical optimiser (used to estimate the hyperparameters) to give quicker computation time and more accurate results although in some cases this optimiser can struggle numerically and the slower optimiser must be used.

Additionally, to allow for a fair comparison between the two methods only 500 PSA runs were used to estimate the hyperparameters by numerical optimisation. This is because for each step, an  $S \times S$  dense matrix must be inverted. As this is computationally expensive, it is suggested [30] that the full PSA run is not used to calculate the hyperparameters. Once the hyperparameters have been estimated, all 1 000 PSA samples are used to find the fitted values  $\hat{g}_i(\phi_s)$ , so all the information is utilised. Using all 1 000 observations for the optimisation step can give more accurate results and is sometimes necessary (see for example §4.4).

The computational time for the GP regression increases substantially with the number of parameters of interest (Table I), between 17 and 470 s for the Vaccine case study and 17 and 71 s for the SAVI example. However, interestingly, the computation time does not increase uniformly for GP regression. This is due to the numerical optimisation, as occasionally additional steps are required to reach convergence.

The computation time for our SPDE-INLA method remains constant as the number of parameters increases. The computation time of our SPDE-INLA method is significantly lower than the GP regression method, up to around 70 times faster. Even for 5 parameters of interest, it is between 2 and 2.5 times faster, despite the fact that we are using all the data points to estimate the EVPPI, albeit from a projected input space.

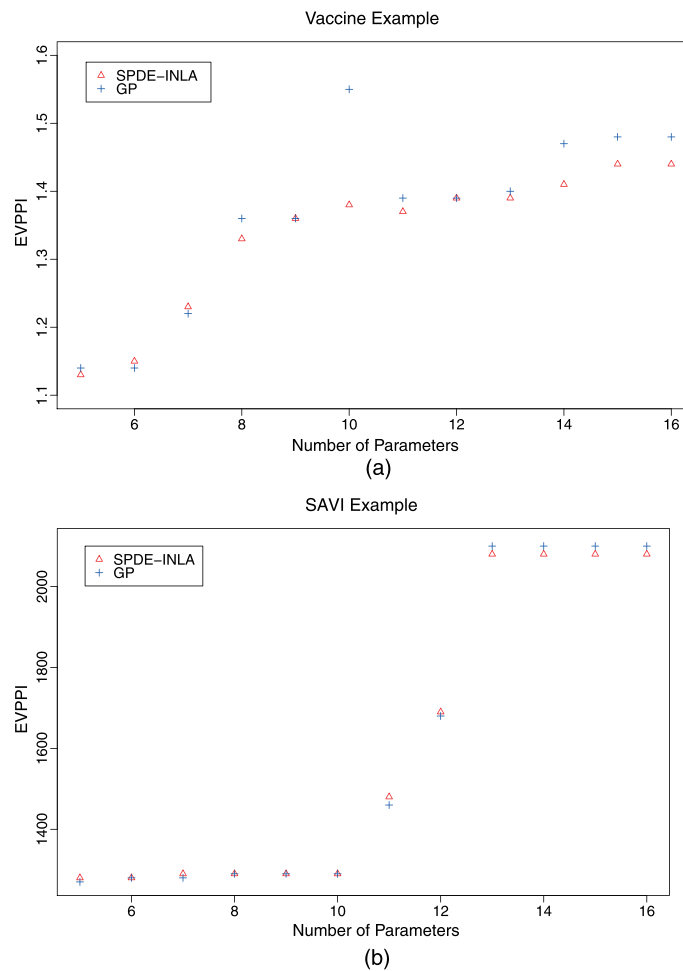
To understand if our method scales to larger PSA datasets, EVPPI estimates using all 10 000 PSA samples from the SAVI case study were also calculated. The computational time required to calculate an EVPPI estimate was between 40 and 80 s with an average time of 56 s. This is significant as the computation time does not increase exponentially using the SPDE-INLA method; the computation time is between than 6 and 10 times slower for a 10-fold increase in number of PSA samples. Crucially, the speed of our SPDE-INLA method depends on the density of the grid approximation. Therefore, its computational effort could be decreased by using a sparser grid, although this would clearly affect the quality of the EVPPI estimate. It would, therefore, be possible to use our method to calculate the EVPPI for larger

PSA data sets. This may be relevant, for instance, in models involving individual level simulations (often referred to as ‘microsimulations’ in the health economic literature) [73], where larger PSA samples are required to fully assess the underlying distributions.

#### 4.4. Accuracy

In general, it is difficult to establish whether an estimate of the EVPPI is accurate because the calculations of the EVPPI are frequently analytically intractable. In fact, for the Vaccine model there is no closed-form expression for the EVPPI, while, given its simplified model structure, for the SAVI example long MC runs can establish the EVPPI up to an inconsequentially small MC error. Thus, it is difficult to determine which method is more accurate when the two approximate EVPPI values diverge, as no baseline comparator is easily available. Nevertheless, there are at least two potential features that we can use to assess the reliability of our estimates.

*4.4.1. Monotonicity with respect to the number of parameters of interest.* It can be easily shown that the EVPPI is a non-decreasing function of the number of parameters of interest (cfr. a proof in Supporting Material). This means that, provided the smaller subsets are entirely contained within the larger subsets, the EVPPI estimates should be non-decreasing. This property provides a possible assessment of the accuracy of the methods: if one method fulfils this property and the other does not, then the former could be more accurate. It is important to note that monotonicity is only a necessary condition for a good EVPPI estimate and not a sufficient one. It is clearly possible to construct a function that gives



**Figure 2.** The EVPPI estimate for the Gaussian Process regression method (GP) and the new method developed in this paper (SPDE) for increasing parameter subset size for the Vaccine (panel a) and the SAVI (panel b) case studies.

a monotone sequence, such as simply giving the number of parameters in the set, but clearly does not estimate the EVPPI at all.

Figure 2 shows the EVPPI estimate for increasing parameter subset sizes for both case studies. The smaller sets of parameters of interest are simply subsets of the larger sets of parameters. For the Vaccine example, shown in panel (a), the standard GP regression method has some difficulty retrieving monotonicity, specifically for the parameter set containing 10 parameters which is clearly overestimated. Our SPDE-INLA method also overestimates the EVPPI for the set containing 10 parameters, but by 0.01 or less than 1%. This gives an indication of the accuracy of our EVPPI estimation method. This overestimation for the standard GP method is due in part to the incorrect estimation of the hyperparameters based on the reduced PSA dataset. If the estimate is obtained using all 1 000 PSA samples then the EVPPI is 1.39, which respects monotonicity but the computation time increases to 256 s compared with 86.

For the SAVI example (panel b) the monotonicity is respected across both methods and the EVPPI values, rounded to three significant figures, are similar. For both examples, the EVPPI values for the smaller parameter subsets are similar across both methods. As the length of  $\phi$  increases the SPDE-INLA method underestimates slightly compared with the standard GP but only by at most 3% of the total EVPPI value. There is evidence that the SPDE-INLA method is accurate while possibly guarding against spurious results that come from estimating the hyperparameters based on a smaller subset of the PSA samples. We note however that some issues with underestimation may occur with larger subsets.

**4.4.2. SPDE-INLA as EVPI approximation.** To investigate whether the SPDE-INLA method underestimates the EVPPI for larger parameter subsets, we compare the results from our method to the overall EVPI, which represents the largest parameter subset available for each example. As mentioned earlier, the overall EVPI (1) can be easily calculated directly from the available PSA data, because as there are no nuisance parameters, we can use a single loop to estimate it. We can then use our method to calculate the overall EVPI, by considering that all the underlying model parameters are of interest. This allows us to compare our method directly with the ‘true’ MC EVPI estimate. The EVPI was calculated for both cases studies and the results are shown in Table II. The computational time required to calculate these estimates is 14 s for the Vaccine example and 8 s for the SAVI.

For both case studies, the SPDE-INLA approximation is correct to two significant figures, with a small discrepancy in the third significant figure. This gives a further indication that the EVPPI estimated using our method is an accurate method for calculating the EVPPI, possibly demonstrating that even for large numbers of parameters of interest the underestimation of the EVPPI is not severe.

For the Vaccine example, there are 62 parameters that contribute to the model uncertainty. We are therefore approximating the Matérn covariance function with a projection from a 62-dimensional to a 2-dimensional space. However, despite the difficulty of preserving the original data structure with this projection, the AIC suggests that this is a sufficient reduction and the EVPI estimate is still very close to the true value.

**Table II.** The EVPI values calculated using the PSA samples directly and our SPDE-INLA method.

Case Study	MC EVPI	SPDE-INLA EVPI
Vaccine	2.52	2.51
SAVI	2100	2080

**Table III.** Comparison of the EVPPI estimation methods, standard GP, GAM regression and the SPDE-INLA method with ‘true’ EVPPI values based on  $10^7$  Monte Carlo simulations.

Parameter Subset	EVPPI estimate (Time to compute in seconds)			
	MC Simulations	GP	GAM	SPDE-INLA
2 Parameters - $\phi_5, \phi_{14}$	248	274 (375)	277 (0.78)	278 (41)
4 Parameters - $\phi_5, \phi_6, \phi_{14}, \phi_{15}$	841	861 (367)	862 (98)	856 (48)
2 Parameters - $\phi_7, \phi_{16}$	536	549 (390)	546 (0.25)	549 (43)

**4.4.3. Analytic Results.** Because the parameters for the SAVI case study are in fact drawn from a known multivariate Normal distribution, the EVPPI can be calculated using a single MC loop. The parameters of interest,  $\phi$ , are sampled from their posterior marginal and then the conditional expectation for each simulated parameter vector is calculated analytically. Strong *et al.* [30] provide these EVPPI values, based on a long MC run, for three different parameters subsets of size 2 and 4, which can be used to test the accuracy of our procedure.

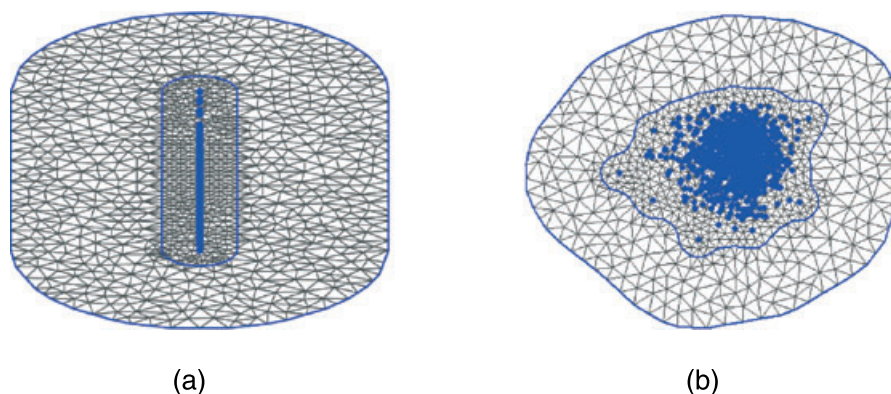
Both GP regression and GAM can be used to calculate the EVPPI using all 10 000 available PSA samples and the willingness-to-pay is fixed at £10 000. Table III gives the single MC loop EVPPI together with all three estimated values, GAM, GP regression and SPDE-INLA. It is clear from this table that our method's performance is in line with the other approximation methods while all three methods overestimate the true EVPPI value in this setting.

The computational time required to calculate these estimates are given in Table III. Clearly, for the 2 parameter setting the GAM regression method is the most appropriate as it takes under a second to calculate the EVPPI estimate. However, for the 4 parameter example, the computational time is lowest for our SPDE-INLA method, which takes half the computational effort of the GAM regression method and around 10% of the standard GP. This demonstrates the computational effort required for the standard GP method using a large number of PSA samples, despite using only 500 PSA samples to find the hyperparameters (see §4.3).

## 5. Technical Considerations

There are several technical aspects relating to the implementation of this method that can affect its estimation performance. These considerations are mostly due to ensuring that the approximations used to estimate the EVPPI are not too rough, whilst retaining the computational advantages of the method.

The most important technical aspect of the SPDE-INLA procedure is the grid approximation used to build up the Finite Element Approximation to the Matérn field. To create an accurate approximation, the triangulation must completely surround the data points with a significant distance from the outermost data point to the final boundary, as there are artificial constraints at the boundary of the triangulation. To reduce the computation time, a tight boundary hugs the data points closely and within this boundary the mesh points are dense to give a good approximation. Outside of this inner boundary, the approximation can be rougher and the triangles are therefore larger. The mesh approximation is most efficient when the 2-dimensions (coming from the projections for EVPPI calculation) are on approximately the same scale. Therefore, the PSA inputs should be rescaled before calculating the projection. This avoids situations such as that shown in Figure 3 (a), where a large number of triangles cover an area with no observations. Rescaling has no effect on the estimated EVPPI value [30] but does significantly decrease that computation time.



**Figure 3.** Two grid approximations for the same data set. The LHS shows the triangulation when the variables are left on their original scale, with the projected data points in blue. Notice that there are a large number of triangles in this case, but a relatively small number that surround the data points. In contrast to this, on the right, where the data points are scaled we note that a much larger number of mesh points cover the data, allowing for a more accurate Matérn field approximation for a fixed computational time.

The triangulation must be dense enough to adequately capture the underlying structure or the EVPPI estimate will be incorrect. However, as the computation time of the method is directly related to the number of vertices, there is a trade-off between accuracy and computational time. Most importantly, larger datasets require denser triangulations to calculate the EVPPI efficiently and typically the number of mesh points should be greater than the number of data points. If some estimation difficulties are observed by a standard model checking procedure, such as residual plots, the estimation can sometimes be improved by making the triangulation more dense.

The inner boundary should completely encase the data points. However, sometimes extreme outliers can be isolated by this boundary and this affects the Matérn field approximation. Therefore, care must be taken to ensure that all points lie within the inner boundary. However, as our 'data',  $\phi$ , are typically from PSA samples (such as MCMC samples), a relatively large number of observations are normally available and thus outliers are rare. An extreme outlier should probably be investigated thoroughly.

In some more complex examples, both estimation methods can struggle and therefore, the small loss in flexibility from our method can slightly worsen its estimation performance. However, this drawback can be overcome by adding additional structure to the linear predictor. This means that the  $\mathbf{H}$  matrix in Equation (6) changes its form to include non-linear functions of the parameters. Mostly importantly, allowing for 2nd or 3rd order interactions between the parameters seems to account for the flexibility lost by using the projections whilst retaining the computational advantages. Adding these extra terms fits easily into the INLA framework and so this can be easily added to the standard EVPPI calculation method.

Finally, we reiterate that model checking should be thoroughly performed to ascertain when this additional flexibility should be used. Specifically, highly structured residuals indicate that the regression curve has not picked up all the relevant relationships present in the data [30, 74]. Therefore, we advocate integrating model checking as a standard part of EVPPI estimation and stress that poorly fitted models result from either a lack of flexibility in the linear predictor or an incorrect PFC model.

Our implementation of the INLA-SPDE method to compute the EVPPI in the R package *BCEA* accounts for all these potential issues, including allowing the user to customise the linear predictor, and thus provides a reliable tool for practitioners. Options are also available to fine tune the model used to find the PFC and to use standard residual plots to check the fit of the non-parametric regression model [67].

## 6. Conclusion

This paper develops a fast method for GP regression in order to reduce the computational effort required to calculate the EVPPI for health economic evaluations. This method is based on a spatial interpretation of GP regression and projections into 2-dimensional space. This in turn allows us to use a fast computation method developed in spatial statistics, based on calculating a sparse precision matrix that approximates a Matérn GP. Finally, this sparse precision matrix allows us to use the INLA methodology for fast Bayesian computation of the hyperparameters for the GP. It also allows us to find fitted values at no additional cost, which are then in turn used to estimate the EVPPI.

Despite the methodological complexity of our GP regression method, the user-friendly R package *R-INLA* can be used to estimate the hyperparameters and find the fitted values. This simplifies the implementation of our method, allowing us to integrate it into a straightforward R function. This GP regression method significantly decreases the computation time required to calculate the EVPPI for larger subsets of parameters of interest as it is at least two times faster than standard GP regression method, taking around 10 s to calculate an EVPPI estimate with 1 000 PSA samples.

There is little loss of accuracy when using our method in the examples we have considered. For larger subsets the EVPPI estimate is slightly underestimated compared with the standard GP methods, but this does not seem to be severe as the EVPI estimates are very close. Additionally, in some examples, our method seemed to be more accurate than the standard GP regression method, due to the breakdown of numerical optimisation. These results are conditional on the dimension reduction being sufficient to capture all the information in  $\phi$ .

There are several important points of further research. Firstly, further work is required to understand the impact of our new method on the bias and standard error of the EVPPI estimate, especially considering the error introduced by using projections. These properties are important and estimation methods have been provided for the standard GP regression method [30]. Secondly, a sparser grid could decrease the computational time required for this method. However, a comprehensive understanding of the impact of the density of the grid on the EVPPI estimate is needed. Ideally, we would determine an optimal

grid density in terms of required computation time and accuracy. Finally, it is important to investigate how successful our method is compared with other fast GP regression methods. Lindgren *et al.* [51] demonstrate that the SPDE framework can be extended to non-stationary fields and thus, this method may provide quick GP regression for non-stationary processes.

Our method has the potential to have an important impact on the practice of health economic evaluation; the analysis of the VoI is well known as a potentially effective tool to determine research priority and the accuracy of decisions made as a result of economic models. Nevertheless, their practical applications has been thwarted by the complexity of the resulting calculations. Our method substantially reduces the computational time and is implemented in an R package, with stand-alone code also available from the authors, which means that practitioners and regulators can use it routinely to assess the impact of uncertainty in models on the decision-making process being investigated.

## Acknowledgements

Anna Heath is funded by the EPSRC. Dr Gianluca Baio is partially funded by a research grant sponsored by Mapi. Additionally, the authors would like to thank Alice Davis for helpful discussion on the SPDE-INLA methodology and two anonymous reviewers for their helpful comments.

## References

1. Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford University Press: Oxford, UK, 2006.
2. Willan A, Briggs A. *The Statistical Analysis of Cost-Effectiveness Data*. John Wiley and Sons: Chichester, UK, 2006.
3. O'Hagan A, Stevens J. A framework for cost-effectiveness analysis from clinical trial data. *Health Economics* 2001; **10**:303–315.
4. O'Hagan A, Stevens J, Montmartin J. Bayesian cost effectiveness analysis from clinical trial data. *Statistics in Medicine* 2001; **20**:733–753.
5. Spiegelhalter D, Abrams K, Myles J. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley and Sons: Chichester, UK, 2004.
6. Baio G. *Bayesian Methods in Health Economics*. CRC Press: Boca Raton, FL, 2012.
7. Loomes G, McKenzie L. The use of QALYs in health care decision making. *Social Science and Medicine* 1989; **28**:299–308.
8. Raiffa H, Schlaifer H. *Applied Statistical Decision Theory*. Harvard University Press: Boston, MA, 1961.
9. Claxton K. The irrelevance of inference: a decision-making approach to stochastic evaluation of health care technologies. *Journal of Health Economics* 1999; **18**:342–364.
10. Claxton K, Lacey L, Walker S. Selecting treatments: a decision theoretic approach. *Journal of the Royal Statistical Society A* 2000; **163**:211–225.
11. Claxton K. Bayesian approaches to the value of information: implications for the regulation of new pharmaceutical. *Health Economics* 1999; **8**:269–274.
12. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, Brazier J, O'Hagan A. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics* 2005; **14**:339–347.
13. EUnetHTA. Methods for health economic evaluations: A guideline based on current practices in Europe - second draft, 2014.
14. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada [3rd Edition], 2006.
15. Department of Health and Ageing. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee: Version 4.3, 2008.
16. Howard R. Information Value Theory. *IEEE Transactions on System Science and Cybernetics* 1966; **SCC-2**(1):22–26.
17. Felli J, Hazen G. Sensitivity analysis and the expected value of perfect information. *Medical Decision Making* 1998; **18**:95–109.
18. Felli J, Hazen G. A Bayesian approach to sensitivity analysis. *Health Economics* 1999; **8**:263–268.
19. Claxton K, Neumann P, Araki S, Weinstein M. Bayesian Value-Of-Information Analysis. *International Journal of Technology Assessment in Health Care* 2001; **17**:38–55.
20. Ades A, Lu G, Claxton K. Expected Value of Sample Information Calculations in Medical Decision Modeling. *Medical Decision Making* 2004; **24**:207–227.
21. Brennan A, Kharroubi S. Efficient Computation of Partial Expected Value of Sample Information Using Bayesian Approximation. *Research Report 560/05*, Department of Probability and Statistics, University of Sheffield UK, 2005.
22. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. *Medical Decision Making* 2006; **26**:480–496.

23. Meltzer D. Addressing uncertainty in medical cost-effectiveness analysis: Implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *Journal of Health Economics* 2001; **20**:109–129.
24. Campbell J, McQueen R, Libby A, Spackman D, Carlson J, Briggs A. Cost-effectiveness Uncertainty Analysis Methods A Comparison of One-way Sensitivity, Analysis of Covariance, and Expected Value of Partial Perfect Information. *Medical Decision Making* 2014; **35**:596–607.
25. Wilson E. A Practical Guide to Value of Information Analysis. *PharmacoEconomics* 2015; **33**(2):105–121.
26. Madan J, Ades A, Price M, Maitland K, Jemutai J, Revill P, Welton N. Strategies for Efficient Computation of the Expected Value of Partial Perfect Information. *Medical Decision Making* 2014:327–342.
27. Strong M, Oakley J. An efficient method for computing single-parameter partial expected value of perfect information. *Medical Decision Making* 2013; **33**(6):755–766.
28. Sadatsafavi M, Bansback N, Zafari Z, Najafzadeh M, Marra C. Need for speed: an efficient algorithm for calculation of single-parameter expected value of partial perfect information. *Value in Health* 2013; **16**(2):438–448.
29. Coyle D, Oakley J. Estimating the expected value of partial perfect information: a review of methods. *The European Journal of Health Economics* 2008; **9**(3):251–259.
30. Strong M, Oakley J, Brennan A. Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample A Nonparametric Regression Approach. *Medical Decision Making* 2014; **34**(3): 311–326.
31. Hastie T, Tibshirani R. *Generalized Additive Models*, Vol. 43. CRC Press: Boca Raton, FL, 1990.
32. Rasmussen C, Williams C. *Gaussian Processes for Machine Learning*. ISBN: 026218253X, 2006.
33. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal the Royal Statistical Society, Series B* 2009; **71**:319–392.
34. Lindgren F. Continuous domain spatial models in R-INLA. *The ISBA Bulletin* 2012; **19**(4):14–20.
35. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology* 2013; **7**:39–55.
36. Baio G, Dawid P. Probabilistic sensitivity analysis in health economics. *Statistical Methods in Medical Research* 2011; **24**:615–634.
37. Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making, 2009.
38. Stinnett A, Mullaly J. Net health benefits a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 1998; **18**(2):S68–S80.
39. Brennan A, Kharroubi S, O'Hagan A, Chilcott J. Calculating Partial Expected Value of Perfect Information via Monte Carlo Sampling Algorithms. *Medical Decision Making* 2007; **27**:448–470.
40. Steuten L, van de Wetering G, Groothuis-Oudshoorn K, Retèl V. A systematic and critical review of the evolving methods and applications of value of information in academia and practice. *Pharmacoeconomics* 2013; **31**(1):25–48.
41. Tuffaha H, Gordon L, Scuffham P. Value of information analysis in healthcare: a review of principles and applications. *Journal of Medical Economics* 2014; **17**:377–383.
42. Davis R. *Gaussian Processes, Encyclopedia of Environmetrics, Section on Stochastic Modeling and Environmental Change*. Wiley: New York, 2001.
43. Ebden M. *Gaussian Processes for Regression: A quick Introduction*, The Website of Robotics Research Group in Department on Engineering Science. University of Oxford: Oxford, 2008.
44. Rasmussen C. Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning*. Springer: Heidelberg, 2004; 63–71.
45. MacKay D. Introduction to Gaussian Processes. *NATO ASI Series F Computer and Systems Sciences* 1998; **168**:133–166.
46. Gibbs M, MacKay DJC. Efficient implementation of gaussian processes. *Technical Report*, Cavendish Laboratory, Cambridge. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.8224>, 1997.
47. O'Hagan A. Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society. Series B (Methodological)* 1978; **40**(1):1–42.
48. O'Hagan A. Bayes–hermite quadrature. *Journal of Statistical Planning and Inference* 1991; **29**(3):245–260.
49. Tobler W. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 1970; **46**:234–240.
50. Cressie N. *Statistics for Spatial Data*. Wiley: New York, 1993.
51. Lindgren F, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011; **73**(4):423–498.
52. Ciarlet P. *The Finite Element Method for Elliptic Problems*. Elsevier: New York, NY, 1978.
53. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (statistical methodology)* 2009; **71**(2):319–392.
54. Rue H, Martino S, Lindgren F, Simpson D, Riebler A, Krainski E. INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using Integrated Nested Laplace Approximation, 2014. R package version 0.0-1404466487.
55. Simpson D, Lindgren F, Rue H. Fast approximate inference with INLA: the past, the present and the future 2011. arXiv preprint arXiv:1105.2982.
56. Lindgren F, Rue H. Bayesian spatial and spatiotemporal modelling with R-INLA. *Journal of Statistical Software* 2013; **63**:1–19.
57. Simpson D, Lindgren F, Rue H. Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics* 2012; **1**:16–29.
58. Krainski E, Rue H. A short introduction on how to fit a SPDE model with INLA, 2013. (Available from: <https://inla.googlecode.com/hg/tutorials/spde/inla-spde-howto/inla-spde-howto.pdf>) [Accessed on 3 May 2016].



59. Cameletti M, Ignaccolo R, Bande S. Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics* 2011; **22**:985–996.
60. Cameletti M, Lindgren F, Simpson D, Rue H. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis* 2013; **97**(2):109–131.
61. Li K. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* 1991; **86**(414):316–327.
62. Cook D, Ni L. Sufficient dimension reduction via inverse regression. *Journal of the American Statistical Association* 2005; **100**:410–428.
63. Cook RD, Weisberg S. Discussion on “sliced inverse regression for dimension reduction” by k. c. li. *Journal of the American Statistical Association* 1991; **86**(414):316–327.
64. Li B, Wang S.i. On directional regression for dimension reduction. *Journal of the American Statistical Association* 2007; **102**(479):997–1008.
65. Cook RD. Fisher lecture: Dimension reduction in regression. *Statistical Science* 2007:1–26.
66. Cook RD, Forzani L. Principal fitted components for dimension reduction in regression. *Statistical Science* 2008; **23**(4):485–501.
67. Baio G, Berardi A, Heath A. *Bayesian Cost Effectiveness Analysis with the R package BCEA*. Springer, forthcoming: New York, 2016.
68. Baio G. BCEA: an R package to run Bayesian cost-effectiveness analysis, 2012. (Available from: <http://www.statistica.it/gianluca/BCEA/BCEA.pdf>) [Accessed on 3 May 2016].
69. Adraghi K, Raim A. ldr: An R software package for likelihood-based sufficient dimension reduction. *Journal of Statistical Software* 2014; **61**(3):1–21.
70. Dong H, Coyle D, Buxton M. Value of information analysis for a new technology: Computer-assisted total knee replacement. *International Journal of Technology Assessment in Health Care* 2007; **23**:337–342.
71. Strong M, Breeze P, Thomas C, Brennan A. SAVI - Sheffield Accelerated Value of Information, Release version 1.013 (2014-12-11), The University of Sheffield, 2014.
72. Strong M. Partial EVPPI Functions, 2012. (Available from: [http://www.shef.ac.uk/polopoly\\_fs/1.305039!/file/R\\_functions.txt](http://www.shef.ac.uk/polopoly_fs/1.305039!/file/R_functions.txt)) [Accessed on 3 May 2016].
73. Weinstein M. Recent developments in decision-analytic modelling for economic evaluation. *Pharmacoeconomics* 2006; **24**:1043–1053.
74. Strong M, Oakley J, Brennan A, Breeze P. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample a fast nonparametric regression-based method. *Medical Decision Making* 2015:570–583.
75. Martins T, Simpson D, Lindgren F, Rue H. Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis* 2013; **67**:68–83.
76. Rue H, Held L. *Gaussian Markov Random Fields: Theory and Applications*. CRC Press: Boca Raton, FL, 2005.

## Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.