SUPPLEMENT

This supplement provides complete proofs for theoretical results, extended numerics and full details to reproduce the experiments presented in the paper.

APPENDIX A: PROOF OF THEORETICAL RESULTS

PROOF OF FACT 1. For a prior $\mathcal{N}(m, c)$ and data $\{(\boldsymbol{x}_i, f_i)\}_{i=1}^n$, standard conjugacy results for GPs lead to the posterior g_n being a GP $\mathcal{N}(m_n, c_n)$, with mean $m_n(\boldsymbol{x}) = m(\boldsymbol{x}) + \boldsymbol{c}(\boldsymbol{x}, X)\boldsymbol{C}^{-1}(\boldsymbol{f} - \boldsymbol{m})$ and covariance $c_n(\boldsymbol{x}, \boldsymbol{x}') = c(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{c}(\boldsymbol{x}, X)\boldsymbol{C}^{-1}\boldsymbol{c}(X, \boldsymbol{x}')$, see Chap. 2 of Rasmussen and Williams (2006). Then repeated application of Fubini's theorem produces

$$\begin{aligned}
\mathbb{E}[\Pi[g_n]] &= \int_\Omega \int_{\mathcal{X}} g_n(\boldsymbol{x}, \omega)\pi(\mathrm{d}\boldsymbol{x})\mathbb{P}(\mathrm{d}\omega) = \int_{\mathcal{X}} m_n(\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x}) \\
\mathbb{V}[\Pi[g_n]] &= \int_\Omega \left[\int_{\mathcal{X}} g_n(\boldsymbol{x}, \omega)\pi(\mathrm{d}\boldsymbol{x}) - \int_{\mathcal{X}} m_n(\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x})\right]^2 \mathbb{P}(\mathrm{d}\omega) \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}}\int_\Omega [g(\boldsymbol{x}, \omega) - m_n(\boldsymbol{x})][g(\boldsymbol{x}', \omega) - m_n(\boldsymbol{x}')]\mathbb{P}(\mathrm{d}\omega)\pi(\mathrm{d}\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x}') \\
&= \int_{\mathcal{X}}\int_{\mathcal{X}} c_n(\boldsymbol{x}, \boldsymbol{x}')\pi(\mathrm{d}\boldsymbol{x})\pi(\mathrm{d}\boldsymbol{x}').
\end{aligned}$$

The proof is completed by substituting the expressions for m_n and c_n into these two equations. (The result in the main text additionally sets $m \equiv 0$.) □

PROOF OF FACT 1. From Eqn. 5 in the main text $\|\hat{\Pi} - \Pi\|_{\mathcal{H}^*} \leq \|\mu(\hat{\pi}) - \mu(\pi)\|_{\mathcal{H}}$. For the converse inequality, consider the specific integrand $f = \mu(\hat{\pi}) - \mu(\pi)$. Then, from the supremum definition of the dual norm, $\|\hat{\Pi} - \Pi\|_{\mathcal{H}^*} \geq |\hat{\Pi}[f] - \Pi[f]|/\|f\|_{\mathcal{H}}$. Now we use the reproducing property:

$$\begin{aligned}
\frac{|\hat{\Pi}[f] - \Pi[f]|}{\|f\|_{\mathcal{H}}} &= \frac{|\langle f, \mu(\hat{\pi}) - \mu(\pi)\rangle_{\mathcal{H}}|}{\|f\|_{\mathcal{H}}} \\
&= \frac{\|\mu(\hat{\pi}) - \mu(\pi)\|_{\mathcal{H}}^2}{\|\mu(\hat{\pi}) - \mu(\pi)\|_{\mathcal{H}}} = \|\mu(\hat{\pi}) - \mu(\pi)\|_{\mathcal{H}}.
\end{aligned}$$

This completes the proof. □

PROOF OF FACT 2. Combining Fact 1 with direct calculation gives that

$$\begin{aligned}
\|\hat{\Pi} - \Pi\|_{\mathcal{H}^*}^2 &= \|\mu(\hat{\pi}) - \mu(\pi)\|_{\mathcal{H}}^2 \\
&= \sum_{i,j=1}^n w_i w_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) - 2\sum_{i=1}^n w_i \int k(\boldsymbol{x}, \boldsymbol{x}_i)\, \mathrm{d}\pi(\boldsymbol{x}) + \iint k(\boldsymbol{x}, \boldsymbol{x}')\, \mathrm{d}\pi(\boldsymbol{x})\mathrm{d}\pi(\boldsymbol{x}') \\
&= \boldsymbol{w}^\top \boldsymbol{K}\boldsymbol{w} - 2\boldsymbol{w}^\top \Pi[\boldsymbol{k}(X, \cdot)] + \Pi\Pi[k(\cdot, \cdot)]
\end{aligned}$$

as required. □

The following lemma shows that probabilistic integrators provide a point estimate that is *at least as good* as their non-probabilistic counterparts:

LEMMA 1 (Bayesian re-weighting). *Let $f \in \mathcal{H}$. Consider the cubature rule $\hat{\Pi}[f] = \sum_{i=1}^n w_i f(\boldsymbol{x}_i)$ and the corresponding BC rule $\hat{\Pi}_{BC}[f] = \sum_{i=1}^n w_i^{\mathrm{BC}} f(\boldsymbol{x}_i)$. Then $\|\hat{\Pi}_{\mathrm{BC}} - \Pi\|_{\mathcal{H}^*} \leq \|\hat{\Pi} - \Pi\|_{\mathcal{H}^*}$.*

PROOF. This is immediate from Fact 2, which shows that the BC weights w_i^{BC} are an optimal choice for the space $\mathcal{H}$. □

The convergence of $\hat{\Pi}_{\mathrm{BC}}$ is controlled by quality of the approximation m_n:

LEMMA 2 (Regression bound). *Let $f \in \mathcal{H}$ and fix states $\{\boldsymbol{x}_i\}_{i=1}^n \in \mathcal{X}$. Then we have $|\Pi[f] - \hat{\Pi}_{\mathrm{BC}}[f]| \leq \|f - m_n\|_2$.*

PROOF. This is an application of Jensen's inequality: $|\Pi[f] - \hat{\Pi}_{\mathrm{BC}}[f]|^2 = (\int f - m_n \mathrm{d}\pi)^2 \leq \int (f - m_n)^2 \, \mathrm{d}\pi = \|f - m_n\|_2^2$, as required. □

Note that this regression bound is not sharp in general (Ritter, 2000, Prop. II.4) and, as a consequence, Thm. 1 below is not quite optimal.

Lemmas 1 and 2 refer to the point estimators provided by BC. However, we aim to quantify the change in probability mass as the number of samples increases:

LEMMA 3 (BC contraction). *Assume $f \in \mathcal{H}$. Suppose that $\|\hat{\Pi}_{\mathrm{BC}} - \Pi\|_{\mathcal{H}^*} \leq \gamma_n$ where $\gamma_n \to 0$ as $n \to \infty$. Define $I_\delta = [\Pi[f] - \delta, \Pi[f] + \delta]$ to be an interval of radius $\delta > 0$ centred on the true value of the integral. Then $\mathbb{P}\{\Pi[g_n] \notin I_\delta\}$ vanishes at the rate $O(\exp(-(\delta^2/2)\gamma_n^{-2}))$.*

PROOF. Assume without loss of generality that $\delta < \infty$. The posterior distribution $\mathrm{o}\Pi[g_n]$ is Gaussian with mean m_n and variance v_n. Since $v_n = \|\hat{\Pi}_{\mathrm{BC}} - \Pi\|_{\mathcal{H}^*}^2$ we have $v_n \leq \gamma_n^2$. Now the posterior probability mass on I_δ^c is given by $\int_{I_\delta^c} \phi(r|m_n, v_n)\mathrm{d}r$, where $\phi(r|m_n, v_n)$ is the p.d.f. of the $\mathcal{N}(m_n, v_n)$ distribution. From the definition of δ we get the upper bound

$$\begin{aligned}
\mathbb{P}\{\Pi[g_n] \notin I_\delta\} \quad &\leq \quad \int_{-\infty}^{\Pi[f]-\delta} \phi(r|m_n, v_n)\mathrm{d}r + \int_{\Pi[f]+\delta}^{\infty} \phi(r|m_n, v_n)\mathrm{d}r \\
&= \quad 1 + \Phi\Big(\underbrace{\frac{\Pi[f] - m_n}{\sqrt{v_n}}}_{(*)} - \frac{\delta}{\sqrt{v_n}} \Big) - \Phi\Big(\underbrace{\frac{\Pi[f] - m_n}{\sqrt{v_n}}}_{(*)} + \frac{\delta}{\sqrt{v_n}} \Big).
\end{aligned}$$

From the definition of the WCE we have that the terms $(*)$ are bounded by $\|f\|_{\mathcal{H}} < \infty$, so that asymptotically as $\gamma_n \to 0$ we have

$$\begin{aligned}
\mathbb{P}\{\Pi[g_n] \notin I_\delta\} \quad &\lesssim \quad 1 + \Phi\big(-\delta/\sqrt{v_n}\big) - \Phi\big(\delta/\sqrt{v_n}\big) \\
&\lesssim \quad 1 + \Phi\big(-\delta/\gamma_n\big) - \Phi\big(\delta/\gamma_n\big) \\
&\lesssim \quad \mathrm{erfc}\big(\delta/\sqrt{2}\gamma_n\big).
\end{aligned}$$

The result follows from the fact that $\mathrm{erfc}(x) \lesssim \exp(-x^2/2)$ for x sufficiently small. □

This result demonstrates that the posterior distribution is well-behaved; probability mass concentrates in a neighbourhood I_δ of $\Pi[f]$. Hence, if our prior is

well calibrated (see Sec. 4.1), the posterior provides uncertainty quantification over the solution of the integral as a result of performing a finite number n of integrand evaluations.

Define the *fill distance* of the set $X = \{\boldsymbol{x}_i\}_{i=1}^n$ as

$$h_X = \sup_{\boldsymbol{x}\in\mathcal{X}} \min_{i=1,\dots,n} \|\boldsymbol{x} - \boldsymbol{x}_i\|_2.$$

As $n \to \infty$ the scaling of the fill distance is described by the following special case of Lemma 2, Oates et al. (2016a):

LEMMA 4. *Let $v : [0,\infty) \to [0,\infty)$ be continuous, monotone increasing, and satisfy $v(0) = 0$ and $\lim_{x\downarrow 0} v(x)\exp(x^{-3d}) = \infty$. Suppose further $\mathcal{X} = [0,1]^d$, π is bounded away from zero on $\mathcal{X}$, and $X = \{\boldsymbol{x}_i\}_{i=1}^n$ are samples from an uniformly ergodic Markov chain targeting π. Then we have $\mathbb{E}_X[v(h_X)] = O\big(v(n^{-1/d+\epsilon})\big)$ where $\epsilon > 0$ can be arbitrarily small.*

PROOF OF THM. 1. Initially consider fixed states $X = \{\boldsymbol{x}_i\}_{i=1}^n$ (i.e. fixing the random seed) and $\mathcal{H} = \mathcal{H}_\alpha$. From a standard result in functional approximation due to Wu and Schaback (1993), see also Wendland (2005, Thm. 11.13), there exists $C > 0$ and $h_0 > 0$ such that, for all $\boldsymbol{x} \in \mathcal{X}$ and $h_X < h_0$, $|f(\boldsymbol{x}) - m_n(\boldsymbol{x})| \leq C h_X^\alpha \|f\|_{\mathcal{H}}$. (For other kernels, alternative bounds are well-known; Wendland, 2005, Table 11.1). We augment X with a finite number of states $Y = \{\boldsymbol{y}_i\}_{i=1}^m$ to ensure that $h_{X\cup Y} < h_0$ always holds. Then from the regression bound (Lemma 2),

$$\begin{aligned}\big|\hat{\Pi}_{\mathrm{B(MC)MC}}[f] - \Pi[f]\big| &\leq \|f - m_n\|_2 = \left(\int (f(\boldsymbol{x}) - m_n(\boldsymbol{x}))^2 \,\mathrm{d}\pi(\boldsymbol{x})\right)^{1/2} \\ &\leq \left(\int (C h_{X\cup Y}^\alpha \|f\|_{\mathcal{H}})^2 \,\mathrm{d}\pi(\boldsymbol{x})\right)^{1/2} = C h_{X\cup Y}^\alpha \|f\|_{\mathcal{H}}.\end{aligned}$$

It follows that $\|\hat{\Pi}_{\mathrm{B(MC)MC}} - \Pi\|_{\mathcal{H}_\alpha^*} \leq C h_{X\cup Y}^\alpha$. Now, taking an expectation $\mathbb{E}_X$ over the sample path $X = \{\boldsymbol{x}_i\}_{i=1}^n$ of the Markov chain (or over the i.i.d. realisation), we have that

$$\mathbb{E}_X\|\hat{\Pi}_{\mathrm{B(MC)MC}} - \Pi\|_{\mathcal{H}_\alpha^*} \leq C\mathbb{E}_X h_{X\cup Y}^\alpha \leq C\mathbb{E}_X h_X^\alpha. \tag{1}$$

From Lemma 4 above, we have a scaling relationship such that, for $h_{X\cup Y} < h_0$, we have $\mathbb{E}_X h_X^\alpha = O(n^{-\alpha/d+\epsilon})$ for $\epsilon > 0$ arbitrarily small. From Markov's inequality, convergence in mean implies convergence in probability and thus, using Eqn. 1, we have $\|\hat{\Pi}_{\mathrm{B(MC)MC}} - \Pi\|_{\mathcal{H}_\alpha^*} = O_P(n^{-\alpha/d+\epsilon})$. This completes the proof for $\mathcal{H} = \mathcal{H}_\alpha$. More generally, if $\mathcal{H}$ is norm-equivalent to $\mathcal{H}_\alpha$ then the result follows from the fact that $\|\hat{\Pi}_{\mathrm{B(MC)MC}} - \Pi\|_{\mathcal{H}^*} \leq \lambda\|\hat{\Pi}_{\mathrm{B(MC)MC}} - \Pi\|_{\mathcal{H}_\alpha^*}$ for some $\lambda > 0$. □

Note that the fill distance was central to the proof of Thm. 1. In fact, the above argument implies that any set of n random points for which the fill distance is asymptotically minimised provides the same rate of contraction for the posterior. Indeed, a deterministic point set with low fill distance could equally be used and our proof demonstrates that the resulting BC point estimator would obtain the optimal $O(n^{-\alpha/d})$ rate for the worst case error, up to logarithmic terms, among all deterministic estimators on $\mathcal{H}_\alpha$.

PROOF OF THM. 2. From Theorem 15.21 of Dick and Pillichshammer (2010), which assumes $\alpha \geq 2$, $\alpha \in \mathbb{N}$, the QMC rule $\hat{\Pi}_{\text{QMC}}$ based on a higher-order digital $(t, \alpha, 1, \alpha m \times m, d)$ net over $\mathbb{Z}_b$ for some prime b satisfies $\|\hat{\Pi}_{\text{BQMC}} - \Pi\|_{\mathcal{H}^*} \leq C_{d,\alpha}(\log n)^{d\alpha} n^{-\alpha} = O(n^{-\alpha+\epsilon})$ for $\mathcal{S}_\alpha$ the Sobolev space of dominating mixed smoothness order α, where $C_{d,\alpha} > 0$ is a constant that depends only on d and α (but not on n). The result follows immediately from norm equivalence and Lemma 1. The contraction rate follows from Lemma 3. □

PROOF OF PROP. 2. Denote by $\mathbb{P}_{n,\lambda}$ the posterior distribution on the integral conditional on a value of λ. Following Prop. 1, this is a Gaussian distribution with mean and variance given by:

$$\begin{aligned}\mathbb{E}_\lambda[\Pi[g_n]] &= \Pi[c_0(\cdot, X)]\boldsymbol{C}_0^{-1}\boldsymbol{f} \\ \mathbb{V}_\lambda[\Pi[g_n]] &= \lambda\{\Pi\Pi[c_0(\cdot,\cdot)] - \Pi[c_0(\cdot, X)]\boldsymbol{C}_0^{-1}\Pi[c_0(X,\cdot)]\}\end{aligned}$$

Furthermore, the posterior on the amplitude parameter satisfies

$$\begin{aligned}p(\lambda|\boldsymbol{f}) &\propto p(\boldsymbol{f}|\lambda)p(\lambda) \\ &= \frac{1}{(2\pi)^{n/2}\lambda^{\frac{n}{2}+1}|\boldsymbol{C}_0|^{\frac{1}{2}}} \exp\left(-\frac{1}{2\lambda}\boldsymbol{f}^\top \boldsymbol{C}_0^{-1}\boldsymbol{f}\right)\end{aligned}$$

which corresponds to an inverse-gamma distribution with parameters $\alpha = \frac{n}{2}$ and $\beta = \frac{1}{2}\boldsymbol{f}^\top \boldsymbol{C}_0^{-1}\boldsymbol{f}$. We therefore have that $(\Pi[g_n], \lambda)$ is distributed as normal-inverse-gamma and the marginal distribution for $\Pi[g_n]$ is a Student-t distribution, as claimed. □

APPENDIX B: KERNEL MEANS

In this section we propose *approximate Bayesian cubature*, ${}_a\hat{\Pi}_{\text{BC}}$, where the weights ${}_a\boldsymbol{w}_{\text{BC}} = \boldsymbol{K}^{-1}{}_a\Pi[\boldsymbol{k}(X,\cdot)]$ are an approximation to the optimal BC weights based on an approximation ${}_a\Pi[\boldsymbol{k}(X,\cdot)]$ of the kernel mean (see also Prop. 1 in Sommariva and Vianello, 2006). The following lemma demonstrates that we can bound the contribution of this error and inflate our posterior to reflect the additional uncertainty due to the approximation, so that uncertainty quantification is still provided.

LEMMA 5 (Approximate kernel mean). *Consider an approximation ${}_a\pi$ to π of the form ${}_a\pi = \sum_{j=1}^m {}_a w_j \delta_{{}_a\boldsymbol{x}_j}$. Then BC can be performed analytically with respect to ${}_a\pi$; denote this estimator by ${}_a\hat{\Pi}_{BC}$. Moreover, $\|{}_a\hat{\Pi}_{BC} - \Pi\|^2_{\mathcal{H}^*} \leq \|\hat{\Pi}_{BC} - \Pi\|^2_{\mathcal{H}^*} + \sqrt{n}\|{}_a\Pi - \Pi\|^2_{\mathcal{H}^*}$.*

PROOF. Define $\boldsymbol{z} = \Pi[\boldsymbol{k}(X,\cdot)]$ and ${}_a\boldsymbol{z} = {}_a\Pi[\boldsymbol{k}(X,\cdot)]$. Let $\boldsymbol{\epsilon} = {}_a\boldsymbol{z} - \boldsymbol{z}$, write ${}_a\hat{\Pi}_{\text{BC}} = \sum_{i=1}^n {}_a w_i^{\text{BC}} \delta_{\boldsymbol{x}_i}$ and consider

$$\begin{aligned}\|{}_a\hat{\Pi}_{\text{BC}} - \Pi\|^2_{\mathcal{H}^*} &= \|\mu({}_a\hat{\pi}_{\text{BC}}) - \mu(\pi)\|^2_{\mathcal{H}} \\ &= \left\langle \sum_{i=1}^n {}_a w_i^{\text{BC}} k(\cdot, \boldsymbol{x}_i) - \int k(\cdot,\boldsymbol{x})\mathrm{d}\pi(\boldsymbol{x}), \sum_{i=1}^n {}_a w_i^{\text{BC}} k(\cdot,\boldsymbol{x}_i) - \int k(\cdot,\boldsymbol{x})\mathrm{d}\pi(\boldsymbol{x}) \right\rangle_{\mathcal{H}} \\ &= {}_a\boldsymbol{w}_{\text{BC}}^\top \boldsymbol{K}\,{}_a\boldsymbol{w}_{\text{BC}} - 2{}_a\boldsymbol{w}_{\text{BC}}^\top \boldsymbol{z} + \Pi[\mu(\pi)] \\ &= (\boldsymbol{K}^{-1}{}_a\boldsymbol{z})^\top \boldsymbol{K}(\boldsymbol{K}^{-1}{}_a\boldsymbol{z}) - 2(\boldsymbol{K}^{-1}{}_a\boldsymbol{z})^\top \boldsymbol{z} + \Pi[\mu(\pi)] \\ &= (\boldsymbol{z}+\boldsymbol{\epsilon})^\top \boldsymbol{K}^{-1}(\boldsymbol{z}+\boldsymbol{\epsilon}) - 2(\boldsymbol{z}+\boldsymbol{\epsilon})^\top \boldsymbol{K}^{-1}\boldsymbol{z} + \Pi[\mu(\pi)] \\ &= \|\hat{\Pi}_{\text{BC}} - \Pi\|^2_{\mathcal{H}^*} + \boldsymbol{\epsilon}^\top \boldsymbol{K}^{-1}\boldsymbol{\epsilon}.\end{aligned}$$

Use $\otimes$ to denote the tensor product of RKHS. Now, since $\epsilon_i = {}_a z_i - z_i = \mu({}_a\hat{\pi})(\boldsymbol{x}_i) - \mu(\pi)(\boldsymbol{x}_i) = \langle \mu({}_a\hat{\pi}) - \mu(\pi), k(\cdot, \boldsymbol{x}_i)\rangle_{\mathcal{H}}$, we have:

$$\begin{aligned}
\boldsymbol{\epsilon}^\top \boldsymbol{K}^{-1}\boldsymbol{\epsilon} &= \sum_{i,i'} [\boldsymbol{K}^{-1}]_{i,i'} \big\langle \mu({}_a\hat{\pi}) - \mu(\pi), k(\cdot, \boldsymbol{x}_i)\big\rangle_{\mathcal{H}} \big\langle \mu({}_a\hat{\pi}) - \mu(\pi), k(\cdot, \boldsymbol{x}_{i'})\big\rangle_{\mathcal{H}} \\
&= \Big\langle \big(\mu({}_a\hat{\pi}) - \mu(\pi)\big) \otimes \big(\mu({}_a\hat{\pi}) - \mu(\pi)\big), \sum_{i,i'} [\boldsymbol{K}^{-1}]_{i,i'} k(\cdot, \boldsymbol{x}_i) \otimes k(\cdot, \boldsymbol{x}_{i'})\Big\rangle_{\mathcal{H}\otimes\mathcal{H}} \\
&\leq \|\mu({}_a\hat{\pi}) - \mu(\pi)\|_{\mathcal{H}}^2 \Big\| \sum_{i,i'} [\boldsymbol{K}^{-1}]_{i,i'} k(\cdot, \boldsymbol{x}_i) \otimes k(\cdot, \boldsymbol{x}_{i'})\Big\|_{\mathcal{H}\otimes\mathcal{H}}.
\end{aligned}$$

From Fact 1 we have $\|\mu({}_a\hat{\pi}) - \mu(\pi)\|_{\mathcal{H}} = \|{}_a\hat{\Pi} - \Pi\|_{\mathcal{H}}$ so it remains to show that the second term is equal to $\sqrt{n}$. Indeed,

$$\begin{aligned}
&\Big\| \sum_{i,i'} [\boldsymbol{K}^{-1}]_{i,i'} k(\cdot, \boldsymbol{x}_i) \otimes k(\cdot, \boldsymbol{x}_{i'})\Big\|_{\mathcal{H}}^2 \\
&\quad = \sum_{i,i',l,l'} [\boldsymbol{K}^{-1}]_{i,i'} [\boldsymbol{K}^{-1}]_{l,l'} \big\langle k(\cdot, \boldsymbol{x}_i) \otimes k(\cdot, \boldsymbol{x}_{i'}), k(\cdot, \boldsymbol{x}_l) \otimes k(\cdot, \boldsymbol{x}_{l'})\big\rangle_{\mathcal{H}} \\
&\quad = \sum_{i,i',l,l'} [\boldsymbol{K}^{-1}]_{i,i'} [\boldsymbol{K}^{-1}]_{l,l'} [\boldsymbol{K}]_{il} [\boldsymbol{K}]_{i',l'} \;=\; \mathrm{tr}[\boldsymbol{K}\boldsymbol{K}^{-1}\boldsymbol{K}\boldsymbol{K}^{-1}] \;=\; n.
\end{aligned}$$

This completes the proof. □

Under this method, the posterior variance $\mathbb{V}[\Pi[{}_a g_n]] := \|{}_a\hat{\Pi}_{\mathrm{BC}} - \Pi\|_{\mathcal{H}^*}^2$ cannot be computed in closed-form, but computable upper-bounds can be obtained and these can then be used to propagate numerical uncertainty through the remainder of our statistical task. The idea here is to make use of the triangle inequality:

$$\|{}_a\hat{\Pi}_{\mathrm{BC}} - \Pi\|_{\mathcal{H}^*} \;\leq\; \|{}_a\hat{\Pi}_{\mathrm{BC}} - {}_a\Pi\|_{\mathcal{H}^*} \;+\; \|{}_a\Pi - \Pi\|_{\mathcal{H}^*}. \tag{2}$$

The first term on the RHS is now available analytically; from Fact 1 its square is ${}_a\Pi_a\Pi[k(\cdot,\cdot)] - {}_a\Pi[\boldsymbol{k}(\cdot, X)]\boldsymbol{K}^{-1}{}_a\Pi[\boldsymbol{k}(X,\cdot)]$. For the second term, explicit upper bounds exist in the case where states ${}_a\boldsymbol{x}_i$ are independent random samples from π. For instance, from (Song, 2008, Thm. 27) we have, for a radial kernel k, uniform ${}_a w_j = m^{-1}$ and independent ${}_a\boldsymbol{x}_i \sim \pi$,

$$\|{}_a\Pi - \Pi\|_{\mathcal{H}^*} \;\leq\; \frac{2}{\sqrt{m}} \sup_{\boldsymbol{x}\in\mathcal{X}} \sqrt{k(\boldsymbol{x},\boldsymbol{x})} + \sqrt{\frac{\log(2/\delta)}{2m}} \tag{3}$$

with probability at least $1-\delta$. (For dependent ${}_a\boldsymbol{x}_j$, the m in Eqn. 3 can be replaced with an estimate for the effective sample size.) Write $C_{n,\gamma,\delta}$ for a $100(1-\gamma)\%$ credible interval for $\Pi[f]$ defined by the conservative upper bound described in Eqns. 2 and 3. Then we conclude that $C_{n,\gamma,\delta}$ is $100(1-\gamma)\%$ credible interval with probability at least $1-\delta$.

Note that, even though the credible region has been inflated, it still contracts to the truth, since the first term on the RHS in Lemma 5 can be bounded by the sum of $\|{}_a\hat{\Pi}_{\mathrm{BC}} - \Pi\|_{\mathcal{H}^*}$ and $\|{}_a\Pi - \Pi\|_{\mathcal{H}^*}$, both of which vanish as $n, m \to \infty$. The resulting (conservative) posterior ${}_a g_n$ can be viewed as a updating of beliefs based on an approximation to the likelihood function; the statistical foundations of such an approach are made clear in the recent work of Bissiri et al. (2016).

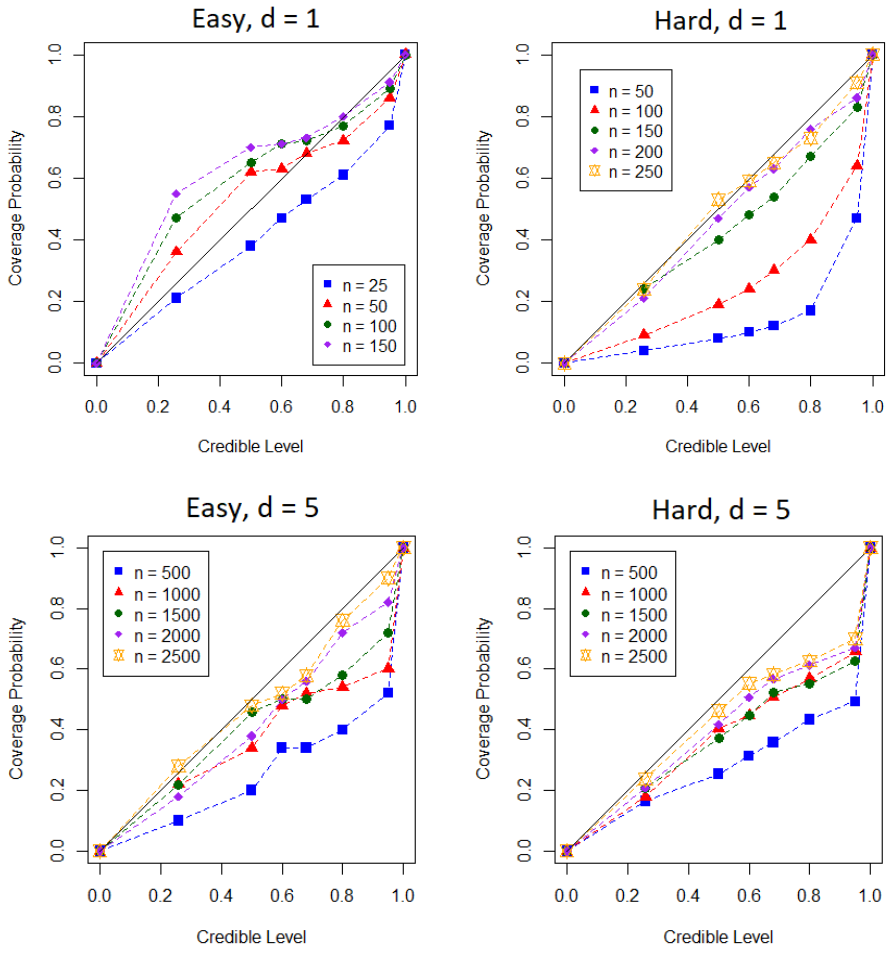

FIGURE 1. *Evaluation of uncertainty quantification provided by EB for both* $\boldsymbol{\sigma}$ *and* λ*. Results are shown for* $d = 1$ *(top) and* $d = 5$ *(bottom). Coverage frequencies* $C_{n,\gamma}$ *(computed from* 100 *(top) or* 50 *(bottom) realisations) were compared against notional* $100(1-\gamma)\%$ *Bayesian credible regions for varying level* γ*. Left: "Easy" test function* f_1*. Right: "Hard" test function* f_2*.*

APPENDIX C: ADDITIONAL NUMERICS

This section presents additional numerical results concerning the calibration of uncertainty for multiple parameters and in higher dimensions.

Calibration in $d = 1$*:* In Fig. 1 (top row) we study the quantification of uncertainty provided by EB in the same setup as in the main text, but optimising over both length-scale parameter σ_1 and magnitude parameter λ. For both "easy" and "hard" test functions, we notice that EB led to over-confident inferences in the "low n" regime, but attains approximately correct frequentist coverage for larger n. Note also that the hyperparameters do not seem to converge to a fixed value as n increases (see Fig. 2).

Calibration in $d = 5$*:* The experiments of Sec. 5.1, based on BMC, were repeated in dimension $d = 5$. Results are shown in Fig. 1 (bottom row). Clearly more integrand evaluations are required for EB to attain a good frequentist coverage of the credible intervals, due to the curse of dimension. However, the frequentist coverage was reasonable for large n in this task.

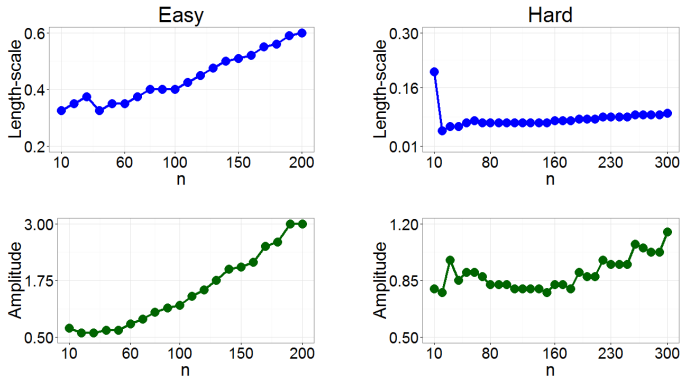

FIGURE 2. *Length scale parameter σ (top) and amplitude parameter λ (bottom) parameters were estimated by the empirical Bayes method as the number n of samples was varied. The "easy" (left) and "hard" (right) test functions considered in the main text were used, here in dimension $d = 1$.*

Calibration for varying prior smoothness: The experiments of Sec. 5.1, based on BMC, were repeated in dimension $d = 1$ for a Matérn kernel with smoothness $\alpha = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$. Results are shown in Fig. 3. There was a clear interaction between the smoothness α of the prior and the number n of samples needed for the length scale σ to be properly estimated and, hence, for the posterior to be well-calibrated.

Empirical convergence assessment: The convergence of BQMC was studied based on higher-order digital nets. The theoretical rates provided in Sec. 3.2.2 for this method are $O(n^{-\alpha+\epsilon})$ for any $\alpha > 1/2$. Fig. 4 gives the results obtained for $d = 1$ (left) and $d = 5$ (right). In the one dimensional case, the $O(n^{-\alpha+\epsilon})$ theoretical convergence rate is attained by the method in all cases $p = \alpha + 1/2 \in \{3/2, 5/2, 7/2\}$ considered. However, in the $d = 5$ case, the rates are not observed for the number n of evaluations considered. This helps us demonstrate the important point that (in addition to numerical conditioning) the rates we provide are asymptotic, and may require large values of n before being observed.

APPENDIX D: SUPPLEMENTAL INFORMATION FOR CASE STUDIES

D.1 Case Study #1

MCMC: In this paper we used the manifold Metropolis-adjusted Langevin algorithm (Girolami and Calderhead, 2011) in combination with population MCMC. Population MCMC shares information across temperatures during sampling, yet previous work has not leveraged evaluation of the log-likelihood f from one sub-chain t_i to inform estimates derived from other sub-chains $t_{i'}$, $i' \neq i$. In contrast, this occurs naturally in the probabilistic integration framework, as described in the main text.

Here MCMC was used to generate a small number, $n = 200$, of samples on a per-model basis, in order to simulate a scenario where numerical error in computation of marginal likelihood will be non-negligible. A temperature ladder with $m = 10$ rungs was employed, for the same reason, according to the recommendation of Calderhead and Girolami (2009). No convergence issues were experienced; the same MCMC set-up has previously been successfully used in Oates et al.

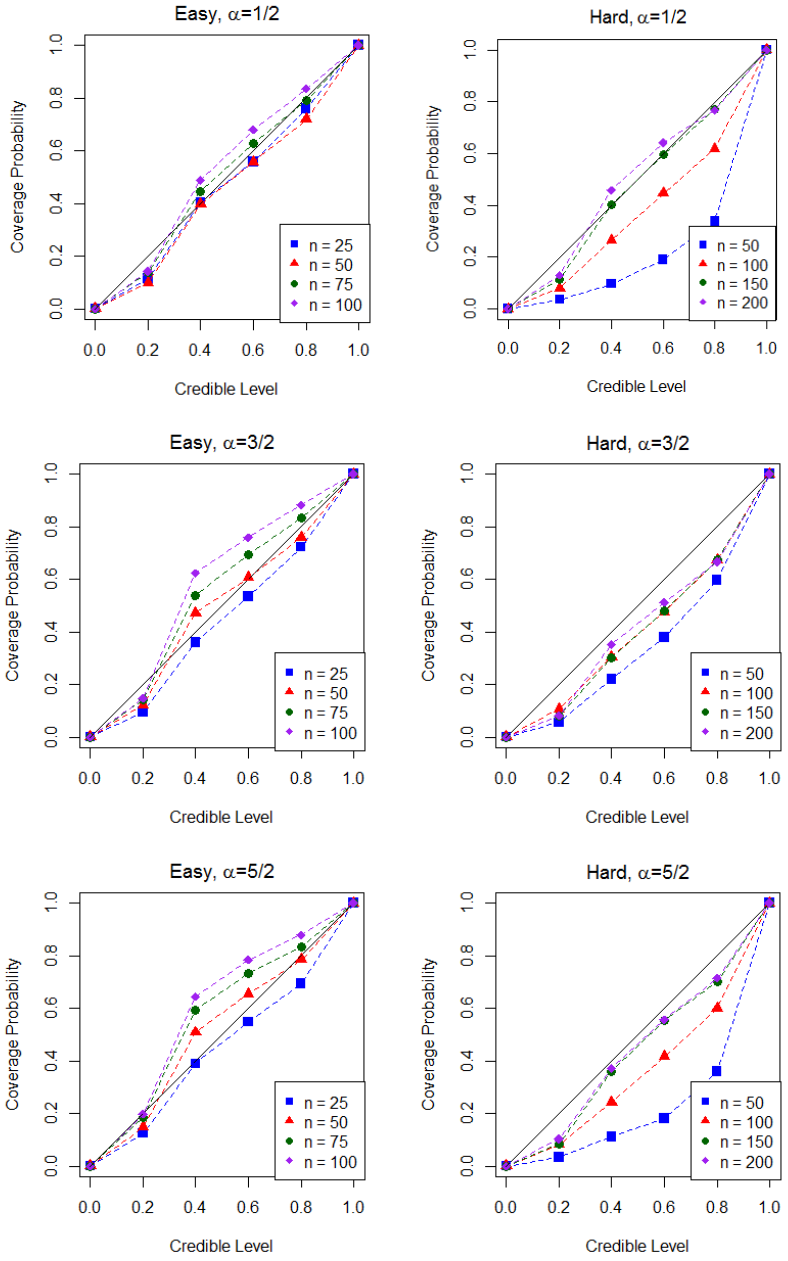

FIGURE 3. *Evaluation of uncertainty quantification being provided. Here* λ *was marginalised whilst* σ *was estimated via EB. Results are shown for* $\alpha = \frac{1}{2}$ *(top),* $\alpha = \frac{3}{2}$ *(middle) and* $\alpha = \frac{5}{2}$ *(bottom). Coverage frequencies* $C_{n,\gamma}$ *were compared against notional* $100(1-\gamma)\%$ *Bayesian credible regions for varying level* γ*. Left: "Easy" test function* f_1*. Right: "Hard" test function* f_2*.*

(2016b).

Prior elicitation: Here we motivate a prior for the unknown function g based on the work of Calderhead and Girolami (2009), who advocated the use of a power-law schedule $t_i = (\frac{i-1}{m-1})^5$, $i = 1, \ldots, m$, based on an extensive empirical comparison of possible schedules. A "good" temperature schedule approximately satisfies the criterion $|g(t_i)(t_{i+1} - t_i)| \approx m^{-1}$, on the basis that this allocates

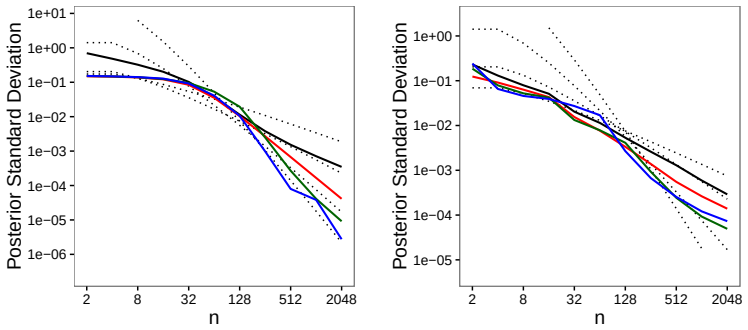

FIGURE 4. *Empirical investigation of BQMC in $d = 1$ (left) and $d = 5$ (right) dimensions and a Sobolev space of mixed dominating smoothness $\mathcal{S}_\alpha$. The results are obtained using tensor product Matérn kernels of smoothness $\alpha = 3/2$ (red), $\alpha = 5/2$ (green) and $\alpha = 7/2$ (blue). Dotted lines represent the theoretical convergence rates established for each kernel. The black line represents standard QMC. Kernel parameters were fixed to $(\sigma_i, \lambda) = (0.005, 1)$ (left) and $(\sigma_i, \lambda) = (1, 0.5)$ (right).*

equal area to the portions of the curve g that lie between t_i and t_{i+1}, controlling bias for the trapezium rule. Substituting $t_i = (\frac{i-1}{m-1})^5$ into this optimality criterion produces $|g(t_i)|((i+1)^5 - i^5) \approx m^4$. Now, letting $i = \theta m$, we obtain $|g(\theta^5)|(5\theta^4 m^4 + o(m^4)) \approx m^4$. Formally treating θ as continuous and taking the $m \to \infty$ limit produces $|g(\theta^5)| \approx 0.2\theta^{-4}$ and so $|g(t)| \approx 0.2t^{-4/5}$. From this we conclude that the transformed function $h(t) = 5t^{4/5}g(t)$ is approximately stationary and can reasonably be assigned a stationary GP prior. However, in an importance sampling transformation we require that $\pi(t)$ has support over $[0, 1]$. For this reason we took $\pi(t) = 1.306/(0.01 + 5t^{4/5})$ in our experiment.

Variance computation: The covariance matrix $\boldsymbol{\Sigma}$ cannot be obtained in closed-form due to intractability of the kernel mean $\Pi_{t_i}[k_f(\cdot, \boldsymbol{\theta})]$. We therefore explored an approximation ${}_a\boldsymbol{\Sigma}$ such that plugging in ${}_a\boldsymbol{\Sigma}$ in place of $\boldsymbol{\Sigma}$ provides an approximation to the posterior variance $\mathbb{V}[\log p(\boldsymbol{y})]$ for the log-marginal likelihood. This took the form

$$ {}_a\Sigma_{i,j} \quad := \quad {}_a\Pi_{t_i\,a}\Pi_{t_j}[k_f(\cdot,\cdot)] - {}_a\Pi_{t_i}[\boldsymbol{k}_f(\cdot, X)]\boldsymbol{K}_f^{-1}{}_a\Pi_{t_j}[\boldsymbol{k}_f(X,\cdot)] $$

where an empirical distribution ${}_a\pi = \frac{1}{100}\sum_{i=1}^{100}\delta_{\boldsymbol{x}_i}$ was employed based on the first $m = 100$ samples, while the remaining samples $X = \{\boldsymbol{x}_i\}_{i=101}^{200}$ were reserved for the kernel computation. This heuristic approach becomes exact as $m \to \infty$, in the sense that ${}_a\Sigma_{i,j} \to \Sigma_{i,j}$, but under-estimates covariance at finite m.

Kernel choice: In experiments below, both k_f and k_h were taken to be Gaussian covariance functions; for example: $k_f(\boldsymbol{x}, \boldsymbol{x}') = \lambda_f \exp\big(-\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2/2\sigma_f^2\big)$ parametrised by λ_f and σ_f. This choice was made to capture smoothness of both integrands f and h involved. For this application we found that, while the σ parameters were possible to learn from data using EB, the λ parameters required a large number of data to pin down. Therefore, for these experiments we fixed $\lambda_f = 0.1 \times \text{mean}(f_{i,j})$ and $\lambda_h = 0.01 \times \text{mean}(h_i)$. In both cases the remaining kernel parameters σ were selected using EB.

Data generation: As a test-bed that captures the salient properties of model selection discussed in the main text, we considered variable selection for logistic

regression:

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{\beta}) &= \prod_{i=1}^{N} p_i(\boldsymbol{\beta})^{y_i}[1-p_i(\boldsymbol{\beta})]^{1-y_i} \\ \text{logit}(p_i(\boldsymbol{\beta})) &= \gamma_1\beta_1 x_{i,1} + \ldots \gamma_d\beta_d x_{i,d}, \quad \gamma_1,\ldots,\gamma_d \in \{0,1\} \end{aligned}$$

where the model $\mathcal{M}_k$ specifies the active variables via the binary vector $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_d)$. A model prior $p(\boldsymbol{\gamma}) \propto d^{-\|\boldsymbol{\gamma}\|_1}$ was employed. Given a model $\mathcal{M}_k$, the active parameters β_j were endowed with independent priors $\beta_j \sim \mathcal{N}(0,\tau^{-1})$, where here $\tau = 0.01$.

A single dataset of size $N = 200$ were generated from model $\mathcal{M}_1$ with parameter $\boldsymbol{\beta} = (1,0,\ldots,0)$; as such the problem is under-determined (there are in principle $2^{10} = 1024$ different models) and the true model is not well-identified. The selected model is thus sensitive to numerical error in the computation of marginal likelihood. In practice we limited the model space to consider only models with $\sum\gamma_i \leq 2$; this speeds up the computation and, in this particular case, only rules out models that have much lower posterior probability than the actual MAP model. There were thus 56 models being compared.

D.2 Case Study #2

Background on the model: The Teal South model is a PDE computer model for an oil reservoir. The model studied is on an 11×11 grid with 5 layers. It has 9 parameters representing physical quantities of interest. These include horizontal permeabilities for each of the 5 layers, the vertical to horizontal permeability ratio, aquifer strength, rock compressibility and porosity. For our experiments, we used an emulator of the likelihood model documented in Lan et al. (2016) in order to speed up MCMC; however this might be undesirable in general due to the additional uncertainty associated with the approximation in the results obtained.

Kernel choice: The numerical results in Sec. 5.3 were obtained using a Matérn $\alpha = 3/2$ kernel given by $k(r) = \lambda^2\big(1+\sqrt{3}r/\sigma\big)\exp\big(-\sqrt{3}r/\sigma\big)$ where $r = \|\boldsymbol{x}-\boldsymbol{y}\|_2$, which corresponds to the Sobolev space $\mathcal{H}_{3/2}$. We note that $f \in \mathcal{H}_{3/2}$ is satisfied. We used EB over the length-scale parameter σ, but fixed the amplitude parameter to $\lambda = 1$.

Variance computation: Due to intractability of the posterior distribution, the kernel mean $\mu(\pi)$ is unavailable in closed form. To overcome this, the methodology in Supplement B was employed to obtain an empirical estimate of the kernel mean (half of the MCMC samples were used with BC weights to approximate the integral and the other half with MC weights to approximate the kernel mean). Eqn. 2 was used to upper bound the intractable BC posterior variance. For the upper bound to hold, states ${}_a\boldsymbol{x}_j$ must be independent samples from π, whereas here they were obtained using MCMC and were therefore not independent. In order to ensure that MCMC samples were "as independent as possible" we employed sophisticated MCMC methodology developed by Lan et al. (2016). Nevertheless, we emphasise that there is a gap between theory and practice here that we hope to fill in future research. For the results in this paper we fixed $\delta = 0.05$ in Eqn. 3, so that $C_{n,\gamma} = C_{n,\gamma,0.05}$ is essentially a $95(1-\gamma)\%$ credible interval. A formal investigation into the theoretical properties of the uncertainty quantification studied by these methods is not provided in this paper.

D.3 Case Study #3

Kernel choice: The (canonical) *weighted* Sobolev space $\mathcal{S}_{\alpha,\boldsymbol{\gamma}}$ is defined by taking each of the component spaces $\mathcal{H}_u$ to be Sobolev spaces of dominating mixed smoothness $\mathcal{S}_\alpha$. i.e. the space $\mathcal{H}_u$ is norm-equivalent to a tensor product of $|u|$ one-dimensional Sobolev spaces, each with smoothness parameter α. Constructed in this way, $\mathcal{S}_{\alpha,\boldsymbol{\gamma}}$ is an RKHS with kernel

$$k_{\alpha,\boldsymbol{\gamma}}(\boldsymbol{x},\boldsymbol{x}') = \sum_{u\subseteq\mathcal{I}} \gamma_u \prod_{i\in u} \left(\sum_{k=1}^{\alpha} \frac{B_k(x_i)B_k(x_i')}{(k!)^2} - (-1)^\alpha \frac{B_{2\alpha}(|x_i - x_i'|)}{(2\alpha)!} \right),$$

where the B_k are Bernoulli polynomials.

Theoretical results: In finite dimensions $d < \infty$, we can construct a higher-order digital net that attains optimal QMC rates for weighted Sobolev spaces:

THEOREM 1. *Let $\mathcal{H}$ be an RKHS that is norm-equivalent to $\mathcal{S}_{\alpha,\boldsymbol{\gamma}}$. Then BQMC based on a digital $(t,\alpha,1,\alpha m\times m,d)$-net over $\mathbb{Z}_b$ attains the optimal rate $\|\hat{\Pi}_{\mathrm{BQMC}} - \Pi\|_{\mathcal{H}^*} = O(n^{-\alpha+\epsilon})$ for any $\epsilon > 0$, where $n = b^m$.*

PROOF. This follows by combining Thm. 15.21 of Dick and Pillichshammer (2010) with Lemma 1. □

The QMC rules in Theorem 1 do not explicitly take into account the values of the weights $\boldsymbol{\gamma}$. An algorithm that tailors QMC states to specific weights $\boldsymbol{\gamma}$ is known as the *component by component* (CBC) algorithm; further details can be found in (Kuo, 2003). In principle the CBC algorithm can lead to improved rate constants in high dimensions, because effort is not wasted in directions where f varies little, but the computational overheads are also greater. We did not consider CBC algorithms for BQMC in this paper.

Note that the weighted Hilbert space framework allows us to bound the WCE *independently of dimension* providing that $\sum_{u\in\mathcal{I}}\gamma_u < \infty$ (Sloan and Woźniakowski, 1998). This justifies the use of "high-dimensional" in this context. Further details are provided in Sec. 4.1 of Dick et al. (2013).

D.4 Case Study #4

Kernel choice: The function spaces that we consider are Sobolev spaces $\mathcal{H}_\alpha(\mathbb{S}^d)$ for $\alpha > d/2$, obtained using the reproducing kernel $k(\boldsymbol{x},\boldsymbol{x}') = \sum_{l=0}^{\infty}\lambda_l P_l^{(d)}(\boldsymbol{x}^\top\boldsymbol{x}')$, $\boldsymbol{x},\boldsymbol{x}'\in\mathbb{S}^d$, where $\lambda_l \asymp (1+l)^{-2\alpha}$ and $P_l^{(d)}$ are normalised Gegenbauer polynomials (Brauchart et al., 2014). A particularly simple expression for the kernel in $d = 2$ and Sobolev space $\alpha = 3/2$ can be obtained by taking $\lambda_0 = 4/3$ along with $\lambda_l = -\lambda_0\times(-1/2)_l/(3/2)_l$ where $(a)_l = a(a+1)\dots(x+l-1) = \Gamma(a+l)/\Gamma(a)$ is the Pochhammer symbol. Specifically, these choices produce $k(\boldsymbol{x},\boldsymbol{x}') = 8/3 - \|\boldsymbol{x}-\boldsymbol{x}'\|_2$, $\boldsymbol{x},\boldsymbol{x}'\in\mathbb{S}^2$. This kernel is associated with a tractable kernel mean $\mu(\pi)(\boldsymbol{x}) = \int_{\mathbb{S}^2} k(\boldsymbol{x},\boldsymbol{x}')\mathrm{d}\pi(\boldsymbol{x}') = 4/3$ and hence the initial error is also available $\Pi[\mu(\pi)] = \int_{\mathbb{S}^2}\mu(\pi)(\boldsymbol{x})\mathrm{d}\pi(\boldsymbol{x}') = 4/3$.

Theoretical results: The states $\{\boldsymbol{x}_i\}_{i=1}^n$ could be generated with MC. In that case, analogous results to those obtained in Sec. 3.2.1 can be obtained. Specifically, from Thm. 7 of Brauchart et al. (2014) and Bayesian re-weighting (Lemma 1), classical MC leads to slow convergence $\|\hat{\Pi}_{\mathrm{MC}} - \Pi\|_{\mathcal{H}^*} = O_P(n^{-1/2})$. The

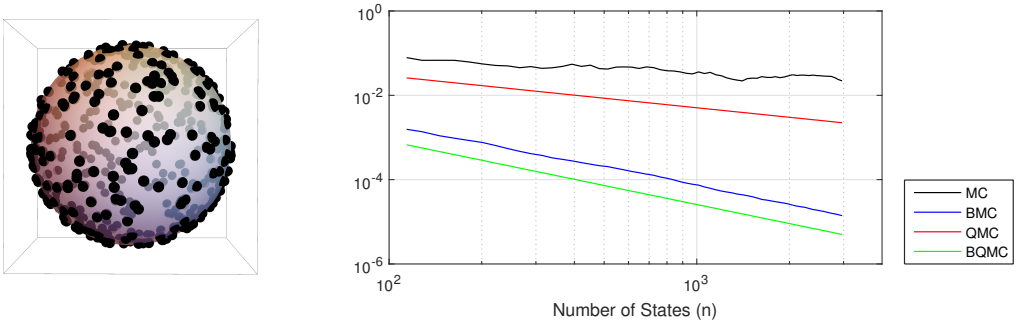

FIGURE 5. *Application to global illumination integrals in computer graphics. Left: A spherical t-design over* $\mathbb{S}^2$*. Right The WCE, or worst-case-error, for Monte Carlo (MC), Bayesian MC (BMC), Quasi MC (QMC) and Bayesian QMC (BQMC).*

regression bound argument (Lemma 2) together with a functional approximation result in Le Gia et al. (2012, Thm. 3.2), gives a faster rate for BMC of $\|\hat{\Pi}_{\mathrm{BMC}} - \Pi\|_{\mathcal{H}^*} = O_P(n^{-3/4})$ in dimension $d = 2$.

Rather than focus on MC methods, we present results based on spherical QMC point sets. We briefly introduce the concept of a *spherical t-design* (Bondarenko et al., 2013) which is define as a set $\{\boldsymbol{x}_i\}_{i=1}^n \subset \mathbb{S}^d$ satisfying $\int_{\mathbb{S}^d} f \mathrm{d}\pi = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i)$ for all polynomials $f : \mathbb{S}^d \to \mathbb{R}$ of degree at most t. (i.e. f is the restriction to $\mathbb{S}^d$ of a polynomial in the usual Euclidean sense $\mathbb{R}^{d+1} \to \mathbb{R}$).

THEOREM 2. *For all* $d \geq 2$ *there exists* C_d *such that for all* $n \geq C_d t^d$ *there exists a spherical t-design on* $\mathbb{S}^d$ *with n states. Moreover, for* $\alpha = 3/2$ *and* $d = 2$*, the use of a spherical t-designs leads to a rate* $\|\hat{\Pi}_{\mathrm{BQMC}} - \Pi\|_{\mathcal{H}^*} = O(n^{-3/4})$.

PROOF. This property of spherical t-designs follows from combining Hesse and Sloan (2005); Bondarenko et al. (2013) and Lemma 1. □

The rate in Thm. 2 is best-possible for a deterministic method in $\mathcal{H}_{3/2}(\mathbb{S}^2)$ (Brauchart et al., 2014). Although explicit spherical t-designs are not currently known in closed-form, approximately optimal point sets have been computed[1] numerically to high accuracy. Additional theoretical results on point estimates can be found in Fuselier et al. (2014). In particular they consider the conditioning of the associated linear systems that must be solved to obtain BC weights.

Numerical results: In Fig. 5, the value of the WCE is plotted[2] for each of the four methods considered (MC, QMC, BMC, BQMC) as the number of states increases. Both BMC and BQMC appear to attain the same rate for $\mathcal{H}_{3/2}(\mathbb{S}^2)$, although BQMC provides a constant factor improvement over BMC. Note that $O(n^{-3/4})$ was shown by Brauchart et al. (2014) to be best-possible for a deterministic method in the space $\mathcal{H}_{3/2}(\mathbb{S}^2)$.

REFERENCES

Bissiri, P., Holmes, C. and Walker, S. (2016). A general framework for updating belief distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(5):1103–1130.

[1] our experiments were based on such point sets provided by R. Womersley on his website `http://web.maths.unsw.edu.au/~rsw/Sphere/EffSphDes/sf.html` [Accessed 24 Nov. 2015].

[2] the environment map used in this example is freely available at: `http://www.hdrlabs.com/sibl/archive.html` [Accessed 23 May 2017].

Brauchart, J., Saff, E., Sloan, I. H. and Womersley, R. (2014). QMC designs: Optimal order quasi Monte Carlo integration schemes on the sphere. *Math. Comp.*, 83:2821–2851.

Fuselier, E., Hangelbroek, T., Narcowich, F. J., Ward, J. D. and Wright, G. B. (2014). Kernel based quadrature on spheres and other homogeneous spaces. *Numer. Math.*, 127(1):57–92.

Hesse, K. and Sloan, I. A. (2005). Worst-case errors in a Sobolev space setting for cubature over the sphere S2. *Bull. Aust. Math. Soc.*, 71(1):81–105.

Kuo, F. Y. (2003). Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces. *J. Complexity*, 19(3):301–320.

Le Gia, Q. T., Sloan, I. H. and Wendland, H. (2012). Multiscale approximation for functions in arbitrary Sobolev spaces by scaled radial basis functions on the unit sphere. *Appl. Comput. Harmon. Anal.*, 32:401–412.

Sloan, I. H. and Woźniakowski, H. (1998). When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *J. Complexity*, 14(1):1–33.

Sommariva, A. and Vianello, M. (2006). Numerical cubature on scattered data by radial basis functions. *Computing*, 76(3-4):295–310.

Wu, Z. and Schaback, R. (1993). Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.*, 13(1):13–27.