Bayesian Emulation of Complex Computer Models with Structured Partial Discontinuities

Jonathan Owen

Supervisors: Dr Ian Vernon & Prof Michael Goldstein

Durham University

iCASE PhD sponsored by Emerson

Collaborator: Prof Jonathan Carter, Coventry University

jonathan.owen@durham.ac.uk

April 28, 2021
Table of Contents

1 Introduction and Motivation

2 Bayesian Emulation of Functions with Structured Discontinuities

3 TNO OLYMPUS Well Placement Optimisation Challenge

4 Conclusion
Introduction

Aim

Develop **Bayesian emulation methodology** which can handle computer simulators (or functions) with **structured (partial) discontinuities**, for use within **decision support tasks**.
Introduction

Aim

Develop **Bayesian emulation methodology** which can handle computer simulators (or functions) with **structured (partial) discontinuities**, for use within **decision support tasks**.

Limitations to use of computer models include:

- Long simulator evaluation times.
- Complex structure of the models and underlying stochasticity.
- High-dimensional input and output spaces.
- Many sources of uncertainty.
Introduction

Aim

Develop **Bayesian emulation methodology** which can handle computer simulators (or functions) with structured (partial) discontinuities, for use within decision support tasks.

Limitations to use of computer models include:

- Long simulator evaluation times.
- Complex structure of the models and underlying stochasticity.
- High-dimensional input and output spaces.
- Many sources of uncertainty.

Structured Discontinuities

- Simulators may possess a set of structured discontinuities.
- These are poorly handled by standard emulation methodology, e.g. Gaussian Processes [Kennedy and O’Hagan, 2001].
Motivation

Well Placement Optimisation in the petroleum industry

Maximise the Net Present Value (NPV) as a function of well location which possesses **structured discontinuities** with respect to the **partial fault boundaries**.
Well Placement Optimisation in the petroleum industry

Maximise the Net Present Value (NPV) as a function of well location which possesses **structured discontinuities** with respect to the **partial fault boundaries**.

TNO OLYMPUS Field Development Optimisation Challenge [ISAPP, 2018]

Aim is to encourage research and development of technology for optimising production systems under uncertainty.
Motivation

Well Placement Optimisation in the petroleum industry

Maximise the Net Present Value (NPV) as a function of well location which possesses structured discontinuities with respect to the partial fault boundaries.

TNO OLYMPUS Field Development Optimisation Challenge [ISAPP, 2018]

Aim is to encourage research and development of technology for optimising production systems under uncertainty.

- Optimisation under uncertainty uses ensembles of models.
- Current industry approaches use computationally expensive optimisation algorithms [Tanaka et al., 2018].
- Decision support incorporating emulators has the ability to fully explore the decision space.
TNO OLYMPUS Reservoir Model [TNO, 2017]

- OLYMPUS is a fictitious oil reservoir model.
- Geological uncertainty is represented by an ensemble of 50 model realisations.
- Simulations are computationally expensive.
Plot of mean oil content over 50 geologies

J coordinate

I coordinate

Transformed TNO OLYMPUS Map
A Bayesian emulator is a stochastic belief specification for a deterministic function that provides a fast and efficient statistical approximation, yielding predictions for as yet unevaluated parameter settings and a corresponding statement of the uncertainty.
Bayesian Emulation [Vernon et al., 2010]

Bayesian Emulation

A Bayesian emulator is a stochastic belief specification for a deterministic function that provides a fast and efficient statistical approximation, yielding predictions for as yet unevaluated parameter settings and a corresponding statement of the uncertainty.

Emulator Equation

\[ f(x) = \sum_j \beta_j g_j(x_A) + u(x_A) + w(x) \]
Bayesian Emulation [Vernon et al., 2010]

**Bayesian Emulation**

A Bayesian emulator is a stochastic belief specification for a deterministic function that provides a fast and efficient statistical approximation, yielding predictions for as yet unevaluated parameter settings and a corresponding statement of the uncertainty.

**Emulator Equation**

\[ f(x) = \sum_j \beta_j g_j(x_A) + u(x_A) + w(x) \]

The **squared exponential covariance function** for \( x \in \mathbb{R}^p \):

\[
\text{Cov}[u(x_A), u(x'_A)] = \sigma_u^2 \exp \left\{ - (x_A - x'_A)^T \Sigma_p^{-1} (x_A - x'_A) \right\}
\]
Bayesian Emulation [Vernon et al., 2010]

Bayesian Emulation

A Bayesian emulator is a stochastic belief specification for a deterministic function that provides a fast and efficient statistical approximation, yielding predictions for as yet unevaluated parameter settings and a corresponding statement of the uncertainty.

Emulator Equation

\[
f(x) = \sum_j \beta_j g_j(x_A) + u(x_A) + w(x)
\]

The squared exponential covariance function for \( x \in \mathbb{R}^p \):

\[
\text{Cov}[u(x_A), u(x'_A)] = \sigma_u^2 \exp \left\{ -\frac{\|x_A - x'_A\|^2}{\theta^2} \right\}
\]
Challenges emulating functions with discontinuities

- Diagnostics will be severely violated in the vicinity of the discontinuity.
- Emulator parameter estimation may produce strange results leading to global issues.

Solution: Embed the input parameter space in higher dimensions.

Define the embedding function $v(x)$ with the form chosen to characterise the discontinuities, e.g. torn along the discontinuities.

Embed $x \in \mathbb{R}^p$ in higher dimensional space $v = v(x) \in \mathbb{R}^q$, where $q > p$.

Emulation proceeds but with input $v(x)$ and variance matrix, $\Sigma^D$.

The squared exponential covariance function is modified to:

$$\text{Cov}[u(v), u(v')] = \sigma^2 \exp\{-\frac{(v - v')^T \Sigma^{-1} q (v - v')}{2}\}$$
Challenges emulating functions with discontinuities

- Diagnostics will be severely violated in the vicinity of the discontinuity.
- Emulator parameter estimation may produce strange results leading to global issues.

Solution: Embed the input parameter space in higher dimensions

- Define the embedding function $v(x)$ with the form chosen to characterise the discontinuities, e.g. torn along the discontinuities.
- Embed $x \in \mathcal{X} \subset \mathbb{R}^p$ in higher dimensional space $v = v(x) \in \mathcal{V} \subset \mathbb{R}^q$, where $q > p$.
- Emulation proceeds but with input $v(x)$ and variance matrix, $\Sigma_{qD}$. 

The squared exponential covariance function is modified to:

$$\text{Cov}[u(v(x)), u(v'(x))] = \sigma^2 \exp\left\{-\frac{(v - v')^T \Sigma_{qD}^{-1} (v - v')}{2}\right\}$$
Challenges emulating functions with discontinuities

- Diagnostics will be severely violated in the vicinity of the discontinuity.
- Emulator parameter estimation may produce strange results leading to global issues.

Solution: Embed the input parameter space in higher dimensions

- Define the embedding function $\mathbf{v}(\mathbf{x})$ with the form chosen to characterise the discontinuities, e.g. torn along the discontinuities.
- Embed $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$ in higher dimensional space $\mathbf{v} = \mathbf{v}(\mathbf{x}) \in \mathcal{V} \subset \mathbb{R}^q$, where $q > p$.
- Emulation proceeds but with input $\mathbf{v}(\mathbf{x})$ and variance matrix, $\Sigma_{qD}$.

The squared exponential covariance function is modified to:

$$\text{Cov}[u(\mathbf{v}), u(\mathbf{v}')] = \sigma_u^2 \exp \left\{ - (\mathbf{v} - \mathbf{v}')^T \Sigma_{qD}^{-1} (\mathbf{v} - \mathbf{v}') \right\}$$
Let \((x, y) \in [0, 2] \times [0, 2]\).

\[
f(x, y) = 0.4 \sin(5x) + 0.4 \cos(5y) \\
+ 1.2(x - 1)^2 \mathbb{1}_{x>1} \mathbb{1}_{y>1.25} \\
- 0.6(x - 0.6)^2 \mathbb{1}_{x>0.6} \mathbb{1}_{y<0.75}
\]

Define the embedding function for \(z = v(x, y)\) to be:

\[
v(x, y) = \begin{cases} 
0 & \text{if } y \geq 1.25 \\
0.6 \left(x - (0.6 + 0.8(y - 0.75))\right)^2 \cdot \mathbb{1}_{x>0.6+0.8(y-0.75)} & \text{if } 0.75 \leq y < 1.25 \\
-0.6(x - 0.6)^2 \cdot \mathbb{1}_{x>0.6} & \text{if } y < 0.75
\end{cases}
\]

\[
\Sigma_{3D} = \begin{pmatrix} \theta^2 & 0 & 0 \\
0 & \theta^2 & 0 \\
0 & 0 & \theta^2 \end{pmatrix}
\]
2D Example – True Function
2D Example – Embedding Surface
2D Example – Emulator Expectation
2D Example – Emulator Variance

Middle section curved in x1x2 and uncorrected
Embedding a 2D input parameter space in 3D

- Let \( x = (x, y) \in \mathcal{X} \subset \mathbb{R}^2 \) and \( \Sigma_{2D} = \begin{pmatrix} \theta^2 & 0 \\ 0 & \theta^2 \end{pmatrix} \) for all \( x \in \mathcal{X} \).
- Define the embedding \( v = v(x) = \begin{pmatrix} x \\ y \\ v(x,y) \end{pmatrix} \in \mathcal{V} \subset \mathbb{R}^3 \).
- Consider a point \( x_0 = (x_0, y_0) \in \mathcal{X} \).
- The partial derivatives are: \( v_x = \frac{\partial v}{\partial x} \bigg|_{x=x_0} \) and \( v_y = \frac{\partial v}{\partial y} \bigg|_{x=x_0} \), and can be used to construct the tangent plane at \( v(x_0, y_0) \).
Embedding a 2D input parameter space in 3D

- Let \( \mathbf{x} = (x, y) \in \mathcal{X} \subset \mathbb{R}^2 \) and \( \Sigma_{2D} = \begin{pmatrix} \theta^2 & 0 \\ 0 & \theta^2 \end{pmatrix} \) for all \( \mathbf{x} \in \mathcal{X} \).
- Define the embedding \( \mathbf{v} = \mathbf{v}(\mathbf{x}) = \begin{pmatrix} x \\ y \\ v(x,y) \end{pmatrix} \in \mathcal{V} \subset \mathbb{R}^3 \).
- Consider a point \( \mathbf{x}_0 = (x_0, y_0) \in \mathcal{X} \).
- The partial derivatives are: \( v_x = \frac{\partial v}{\partial x}\bigg|_{\mathbf{x}=\mathbf{x}_0} \) and \( v_y = \frac{\partial v}{\partial y}\bigg|_{\mathbf{x}=\mathbf{x}_0} \), and can be used to construct the tangent plane at \( \mathbf{v}(x_0, y_0) \).

Aim

Construct \( \Sigma_{3D} \) which induces \( \Sigma_{2D} \), using the tangent plane to approximate \( \mathbf{v}(x, y) \).
Embedding a 2D input parameter space in 3D

- Let \( x = (x, y) \in \mathcal{X} \subset \mathbb{R}^2 \) and \( \Sigma_{2D} = \begin{pmatrix} \theta^2 & 0 \\ 0 & \theta^2 \end{pmatrix} \) for all \( x \in \mathcal{X} \).
- Define the embedding \( v = v(x) = \begin{pmatrix} x \\ y \\ v(x,y) \end{pmatrix} \in \mathcal{V} \subset \mathbb{R}^3 \).
- Consider a point \( x_0 = (x_0, y_0) \in \mathcal{X} \).
- The partial derivatives are: \( v_x = \frac{\partial v}{\partial x} \bigg|_{x=x_0} \) and \( v_y = \frac{\partial v}{\partial y} \bigg|_{x=x_0} \), and can be used to construct the tangent plane at \( v(x_0, y_0) \).

**Aim**

Construct \( \Sigma_{3D} \) which induces \( \Sigma_{2D} \), using the tangent plane to approximate \( v(x, y) \).

Work in the orthonormal \( w \)-basis defined by the tangent plane:

\[
\begin{align*}
\mathbf{w}_1 & \propto v_x \mathbf{e}_x + v_y \mathbf{e}_y + (v_x^2 + v_y^2) \mathbf{e}_z \\
\mathbf{w}_2 & \propto -v_y \mathbf{e}_x + v_x \mathbf{e}_y \\
\mathbf{w}_3 & \propto -v_x \mathbf{e}_x - v_y \mathbf{e}_y + \mathbf{e}_z
\end{align*}
\]
Embedding a 2D input parameter space in 3D

**Proposed form of $\Sigma_{3D}$ by SVD:**

$$
\Sigma_{3D} = \alpha_1^2 w_1 w_1^T + \alpha_2^2 w_2 w_2^T + \alpha_3^2 w_3 w_3^T
$$
Embedding a 2D input parameter space in 3D

**Proposed form of $\Sigma_{3D}$ by SVD:**

\[
\Sigma_{3D} = \alpha_1^2 w_1 w_1^T + \alpha_2^2 w_2 w_2^T + \alpha_3^2 w_3 w_3^T
\]

Let $u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ be in the tangent plane of $v(x)$ at $x_0$ where

\[
u(x) = A x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ v_x & v_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.\]

Recentre at $x_0 = 0$ and $u_0 = v(x_0) = 0$.

Mahalanobis distance preservation requires: $x^T \Sigma_{2D}^{-1} x = u^T \Sigma_{3D}^{-1} u$. 
Embedding a 2D input parameter space in 3D

Proposed form of $\Sigma_{3D}$ by SVD:

$$\Sigma_{3D} = \alpha_1^2 w_1 w_1^T + \alpha_2^2 w_2 w_2^T + \alpha_3^2 w_3 w_3^T$$

Let $u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ be in the tangent plane of $v(x)$ at $x_0$ where

$u = Ax = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. Recentre at $x_0 = 0$ and $u_0 = v(x_0) = 0$.

Mahalanobis distance preservation requires: $x^T \Sigma_{2D}^{-1} x = u^T \Sigma_{3D}^{-1} u$.

$$\implies \Sigma_{2D}^{-1} = A^T \Sigma_{3D}^{-1} A$$

$$\implies \begin{pmatrix} \frac{1}{\theta^2} & 0 \\ 0 & \frac{1}{\theta^2} \end{pmatrix} = \frac{1 + r^2}{\alpha_1^2 r^2} \begin{pmatrix} v_x^2 & v_x v_y \\ v_x v_y & v_y^2 \end{pmatrix} + \frac{1}{\alpha_2^2 r^2} \begin{pmatrix} v_y^2 & -v_x v_y \\ -v_x v_y & v_x^2 \end{pmatrix}$$

where $r^2 = v_x^2 + v_y^2$
Embedding a 2D input parameter space in 3D

**Proposed form of $\Sigma_{3D}$ by SVD:**

$$
\Sigma_{3D} = \alpha_1^2 w_1 w_1^T + \alpha_2^2 w_2 w_2^T + \alpha_3^2 w_3 w_3^T
$$

Let $u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ be in the tangent plane of $v(x)$ at $x_0$ where $u = Ax = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ v_x & v_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. Recentre at $x_0 = 0$ and $u_0 = v(x_0) = 0$. Mahalanobis distance preservation requires: $x^T \Sigma_{2D}^{-1} x = u^T \Sigma_{3D}^{-1} u$.

$$
\Sigma_{3D} = \begin{pmatrix}
\theta^2 + \frac{\alpha_3^2 v_x^2}{1 + r^2} & \frac{\alpha_3^2 v_x v_y}{1 + r^2} & v_x \left( \theta^2 - \frac{\alpha_3^2}{1 + r^2} \right) \\
\frac{\alpha_3^2 v_x v_y}{1 + r^2} & \theta^2 + \frac{\alpha_3^2 v_y^2}{1 + r^2} & v_y \left( \theta^2 - \frac{\alpha_3^2}{1 + r^2} \right) \\
v_x \left( \theta^2 - \frac{\alpha_3^2}{1 + r^2} \right) & v_y \left( \theta^2 - \frac{\alpha_3^2}{1 + r^2} \right) & \theta^2 r^2 + \frac{\alpha_3^2}{1 + r^2}
\end{pmatrix}
$$
It is clear that the form of $\Sigma_{3D}$ is **not constant over the 3D input space** for general embedding functions, $v(x, y)$.

We need to construct **non-stationary emulators** since $\Sigma_{3D}(x)$.

Non-stationary emulators of [Dunlop et al., 2018] are employed to accommodate for this.
Non-Stationary Emulators

- It is clear that the form of $\Sigma_{3D}$ is **not constant** over the 3D input space for general embedding functions, $v(x, y)$.
- We need to construct **non-stationary emulators** since $\Sigma_{3D}(x)$.
- Non-stationary emulators of [Dunlop et al., 2018] are employed to accommodate for this.

**Quadratic Form, where $v = v(x)$ and $\Sigma(x) = \Sigma_{3D}(x)$**

$$Q(v, v') = (v - v')^T \left( \frac{\Sigma(x) + \Sigma(x')}{2} \right) (v - v')$$

**Non-Stationary Squared Exponential Covariance Function**

$$\text{Cov}[u(v), u(v')] = \sigma_u^2 \frac{2^{\frac{3}{2}} |\Sigma(x)|^{\frac{1}{4}} |\Sigma(x')|^{\frac{1}{4}}}{|\Sigma(x) + \Sigma(x')|^{\frac{1}{2}}} \exp \left\{ -Q(v, v') \right\}$$
2D Emulator Variance NOT using embedding correction
2D Emulator Variance using embedding correction

Middle section curved in x1x2 and corrected using tangent plane method.
2D Emulator Expectation NOT using embedding correction
2D Emulator Expectation using embedding correction
2D Emulator Expectation using embedding correction
2D Example – True Function
Well Placement Optimisation

Aim
Design a well placement strategy including each well location, trajectory and type, as well as the number of wells and drilling sequence, for the TNO OLYMPUS oil reservoir model which maximises the expected NPV over the field lifetime, under the uncertainty captured by the 50 geological realisations.
Aim

Design a well placement strategy including each well location, trajectory and type, as well as the number of wells and drilling sequence, for the TNO OLYMPUS oil reservoir model which maximises the expected NPV over the field lifetime, under the uncertainty captured by the 50 geological realisations.

Objective Function

\[
\mathbb{E}[\text{NPV}](\mathbf{x}) \approx \text{NPV}(\mathbf{x})
\]

\[
\text{NPV}_j(\mathbf{x}) = \sum_{i=1}^{N_t} \frac{R_j(\mathbf{x}, t_i)}{(1 + d)^{\frac{t_i}{\tau}}}
\]

\[
R_j(\mathbf{x}, t_i) = Q_{j,op}(\mathbf{x}, t_i) \cdot r_{op} - Q_{j,wp}(\mathbf{x}, t_i) \cdot r_{wp} - Q_{j,wi}(\mathbf{x}, t_i) \cdot r_{wi} - P_j(\mathbf{x}, t_i) - D_j(\mathbf{x}, t_i)
\]
Transformed TNO OLYMPUS Map

Plot of mean oil content over 50 geologies

J. Owen (Durham University)
Bayesian Emulation with Structured Discontinuities
April 28, 2021 24 / 40
OLYMPUS – Embedding Surface
OLYMPUS – Embedding Surface

Plot of Embedding Surface

I coordinate

J coordinate
OLYMPUS – Embedding Surface

Plot of Embedding Surface

I coordinate

J coordinate

-2
-1
0
1
2

40 60 80 100

60
70
80
90
100
110
120
130
Covariance Matrix for Vertical Slices through the Location Parameter Space

Plot of Embedding Surface

Induced Covariance of vertical strip at x = 42

J. Owen (Durham University) Bayesian Emulation with Structured Discontinuities April 28, 2021 29 / 40
Covariance Matrix for Vertical Slices through the Location Parameter Space
Covariance Matrix for Vertical Slices through the Location Parameter Space
Covariance Matrix for Vertical Slices through the Location Parameter Space

Induced Covariance of vertical strip at x = 116

J. Owen (Durham University)
Oil reservoir engineering insight suggests the first well should be a **producer**.

An initial 41 point design is constructed by sequentially selecting points which **minimises the average emulator variance** over the input parameter space.

This is based on an emulator for functions with structured discontinuities.

An additional 6 points are added in 3 pairs to **explore the existence and magnitude of the (partial) discontinuities**.

**Ghost runs with zero NPV** are manually added in the water region surrounding the field to pin down the oil-water boundary or the location parameter space.
NPV Emulator Expectation for Well 1 Wave 1

![Plot of Emulator Expectation Well 1 Wave 1](image)
NPV Emulator Expectation plus 2 SD for Well 1 Wave 1
NPV Emulator Standard Deviation for Well 1 Wave 1
NPV Emulator Expectation for Well 1 Wave 2
The proposed emulation methodology is able to handle general forms of structured partial discontinuities whilst still maintaining a flexible choice of emulator form.

These emulators have been successfully demonstrated to help choose well locations for the TNO OLYMPUS Challenge.

Emulators which can handle structured partial discontinuities can be incorporated within our sequential decision support strategy for the full well placement optimisation problem.

An efficient strategy is to simultaneously choose locations for groups of wells. For example, 3 wells represents a 6D problem requiring an embedding of the location parameter space in 9D.

The embedded emulator methodology is fully generalisable and transferable to many scientific and industrial applications possessing such structured discontinuities.


TNO (2017). Olympus oil reservoir model input decks. OLYMPUS Oil Reservoir Model Input Decks and documentation designed for use in the TNO OLYMPUS field development optimisation challenge.