Emulation for Multiphysics Simulations

Susan E. Minkoff¹, Tamara Dolski¹, Elaine Spiller², and Matt McChesney,³

¹Department of Mathematical Sciences, The University of Texas at Dallas
²Department of Mathematical and Statistical Sciences, Marquette University
³Battelle Memorial Institute, Columbus, Ohio

February 27, 2024

Acknowledgements: Supported through NSF grants # DMS-2053858 and # DMS-2053872 and the UT Dallas 3D+4D Seismic FWI research consortium.
Outline

1. Description of the seismic inverse problem and why surrogate models are useful.

2. Discussion of coupled multiphysics application problems including hydraulic fracturing.

3. Gaussian process emulators and what can go wrong when emulating coupled physics.


5. Example problem results comparing emulators
   - Trigonometric toy problem
   - Terzaghi consolidation problem

6. Emulation results for hydraulic fracturing.
Deterministic inversion gives a single model of the desired parameter. UQ methods result in distributions of velocity fields.

Problem: MCMC can take many models (100’s of thousands or millions) to converge.

Often 40-90% of samples are rejected!

Recall Bayes’ Rules:

\[ P(c|\text{data}) \propto P(\text{data}|c)P(c) \]

Likelihood function:

\[ P(d|\theta) = \exp \left( -\frac{\| F(\theta) - d \|^2}{2\sigma^2\| d \|^2} \right) \]

\[ G. K. \ Stuart, \ S. \ E. \ Minkoff, \ and \ F. \ Pereira, \ “A \ Two-Stage \ Markov \ Chain \ Monte \ Carlo \ Method \ for \ Seismic \ Inversion \ and \ Uncertainty \ Quantification,” \ Geophysics, \ Vol. \ 84, \ No. \ 6, \ pp. \ R1015-R1032, \ 2019. \]
Gaussian Stochastic Process Emulators (GASP) which once trained act as interpolators for other data points without requiring runs of full physical simulators.

GASP emulators provide credible intervals and means at non-design points.²

Applications include modeling leakage for CO₂ sequestration, hydraulic fracturing simulation, and coupled time lapse simulation studies.

Standard Gaussian Process

Predictive mean and variance:

\[
\hat{y}(x^*) = m(x^*) + r^T(x^*)\hat{R}^{-1}(y^D - m(X^D)) \\
\hat{s}^2(x^*) = \sigma^2 \left(1 - r(x^*)^T\hat{R}^{-1}r(x^*) + \frac{(1 - 1^T\hat{R}^{-1}r(x^*))^2}{1^T\hat{R}^{-1}1}\right)
\]

\(\hat{y}(x^*)\) = predictive mean  \(m(\cdot)\) = trend function
\(\{x^D, y^D\}\) = training data  \(\sigma^2\) = estimated variance
\(\hat{s}(x^*)\) = the variance  \(r(x^*) = (c(x^*, x_1^D), \ldots, c(x^*, x_n^D))^T\)
\(R\) = correlation matrix  \(x^*\) = untested input

Linked Emulators vs. Composite Emulators

Mathematically we have $g(f(x), z)$ with the following inputs and outputs:

![Diagram showing linked and composite emulators](image)

**Figure 1:** Pink lines are true functions. Dark green lines are emulator means. Green shaded region gives confidence interval. Circles are design points.

---

Parallel Partial Emulator:
- Predict vector-valued functions
  \[ \hat{y}(x^*) = m(x^*) + r^T R^{-1} (Y^D - m(X^D)) \].

Linked Emulator:
- Predict coupled scalar-valued functions
  \( g(f(x), z) \)

Parallel Partial Linked Emulator:
- Predict vector-valued coupled functions
  \( g(f(x), z) \)

Predictive Mean and Variance:
\[ \hat{y}(z^*, w^*) = m(W^*, z^*) + l^T R^{-1}(g^D - m(z^D, w^D)) \]
\[ \hat{s}^2(z^*, w^*) = V_1 + \sigma^2 V_2 \]

Where \( l_i = \prod_{j=1}^{d} \xi_{ij} \prod_{j=1}^{p} r_{ij}(z_j^D, z_{ij}^*) \), and \( V_1 \) and \( V_2 \) give a closed form expression for \( \text{Var}(\mu(W, z)) \) and \( \mathbb{E}[\sigma^2(W, z)] \) respectively with \( \mu \) and \( \sigma \) being mean and variance of \( \tilde{g} \).

---

Composite Trigonometric Function Example

\[ f(x) = \langle \sin(c_1 x_1) + c_2 x_2^2, \sin(c_3 x_1 \cos(c_4 \pi x_2)) \rangle \]

\[ g(w, z) = \langle \cos(c_5 z_2) \sin(c_6 w_1) + \sin(c_7 z_1) \cos(c_8 w_2), \rangle \]

\[ \eta_1(x, z) = g_1(w, z) \]

\[ \cos(c_9 z_2) \sin(c_{10} w_1) + \sin(c_{11} z_1) \cos(c_{12} w_2) \]

\[ \eta_2(x, z) = g_2(w, z) \]

**Figure 3:** Emulation results at nominal values \([c_1, \ldots, c_{12}] = [5, 1, 3, 1, 3, 3, 3, 4, 2, 4, 2]\): (a) \(\eta_1\) with PPLE; (b) \(\eta_2\) with PPLE; (c) \(\eta_1\) with PPCE; (d) \(\eta_2\) with PPCE. Red squares are true values. Black dots are predicted values from GP mean. Bars are 95% credible intervals.

---

Flow equation:

\[ \rho_0 \frac{\partial \phi p}{\partial t}(x, t) = \nabla \left( \frac{k}{\mu c} \nabla p \right) \]

- \( \rho_0 \) = initial fluid density
- \( \phi \) = porosity
- \( k \) = permeability
- \( \mu \) = fluid viscosity
- \( c \) = fluid compressibility
- \( p \) = pressure

Mechanical Deformation equation:

\[-(\lambda + \hat{\mu}) \frac{d^2 u}{dx^2} = f\]

- \( \lambda \) and \( \hat{\mu} \) = Lamé constants
- \( u \) = displacement
- \( f \) = the external load

---


Coupling: Terzaghi Consolidation Problem

**Coupling in the Simulation**

- **Fluid Flow**
- **Change in Pressure**
- **Porosity**
- **Mechanics**

- Change in pressure: \( \frac{dp}{dx} \)
- Porosity: \( \phi(x, t) = 1 - \frac{1 - \phi_0}{e^{\varepsilon}} \), \( \varepsilon = \frac{du}{dx} \)
- \( \varepsilon \) = strain
- \( \phi_0 \) = porosity

**Coupling in the Emulator**

- **\( \vec{f}_1 \)**
- **\( \vec{f}_2 \)**
- ... 
- **\( \vec{f}_d \)**
- **\( W_1 \)**
- **\( W_2 \)**
- ... 
- **\( W_d \)**
- **\( \bar{g} \)**
- **\( Y \)**

- **x**: \([\nu, E, k, \phi_0]\)
- **W**: \(\phi\)
- **z**: \([k, \phi_0]\)
- **f**: mechanics equation
- **g**: fluid flow equation
- **Y**: Pressure

---

Constant permeability case. Time $t = 0.3$ s

(a) Simulated pressure values along the column of mud used for testing the emulators; (b) RMSE over the column for PPCE (blue solid line), PPLE PCA (long dashed black line), and PPLE gKDR (short dashed red line); (c) $L_{CI}$ over the column for each method with line styles corresponding to those shown in b.

<table>
<thead>
<tr>
<th></th>
<th>PPLE PCA</th>
<th>PPLE gKDR</th>
<th>PPCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing (s)</td>
<td>2</td>
<td>377</td>
<td>2</td>
</tr>
<tr>
<td>RMSE (psi)</td>
<td>4.1e3</td>
<td>3.9e3</td>
<td>1.4e4</td>
</tr>
<tr>
<td>$L_{CI}$ (psi)</td>
<td>1.7e4</td>
<td>6.5e3</td>
<td>2.1e4</td>
</tr>
<tr>
<td>Coverage (%)</td>
<td>97</td>
<td>93</td>
<td>92</td>
</tr>
</tbody>
</table>
(a) Simulated pressure values along the column of mud used for testing the emulators; (b) the root mean squared error (RMSE) plots for CE laGP (solid blue line), LE laGP (dash-dotted orange line), CESC (long dashed black line), and LESC (short dashed purple line); (c) the average 95% credible interval ($L_{CI}$) for these methods.

<table>
<thead>
<tr>
<th></th>
<th>LESC</th>
<th>CESC</th>
<th>LE laGP</th>
<th>CE laGP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing (s)</td>
<td>143</td>
<td>69</td>
<td>342</td>
<td>135</td>
</tr>
<tr>
<td>RMSE (psi)</td>
<td>3.6e4</td>
<td>1.7e5</td>
<td>5.7e4</td>
<td>2.0e5</td>
</tr>
<tr>
<td>$L_{CI}$ (psi)</td>
<td>3.9e4</td>
<td>4.7e5</td>
<td>2.7e4</td>
<td>7.9e5</td>
</tr>
<tr>
<td>Coverage (%)</td>
<td>93</td>
<td>99</td>
<td>80</td>
<td>94</td>
</tr>
</tbody>
</table>
Hydraulic Fracturing (CFRAC) Simulation Setup

<table>
<thead>
<tr>
<th>Quantity</th>
<th>max(QoI)</th>
<th>G</th>
<th>$\phi_{edil}^\circ$</th>
<th>$P_0$</th>
<th>$q_{max}$</th>
<th>$\sigma_{xx}$</th>
<th>$\sigma_{yy}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta V_{ME}$ ($m^3$)</td>
<td>0.039</td>
<td>13572</td>
<td>1.07</td>
<td>39.7</td>
<td>70.9</td>
<td>57.3</td>
<td>109.7</td>
</tr>
<tr>
<td>$\frac{\Delta V_{ME}}{\Delta V_{total}}$ (%)</td>
<td>0.5</td>
<td>13178</td>
<td>1.1</td>
<td>44.5</td>
<td>71.7</td>
<td>60.3</td>
<td>108.9</td>
</tr>
<tr>
<td>$\Delta D_{ME}$ ($m$)</td>
<td>0.22</td>
<td>13932</td>
<td>0.86</td>
<td>34.5</td>
<td>75</td>
<td>63.9</td>
<td>112.1</td>
</tr>
<tr>
<td>$\frac{\Delta D_{ME}}{\Delta D_{total}}$ (%)</td>
<td>2.2</td>
<td>12463</td>
<td>1.08</td>
<td>34.8</td>
<td>72.2</td>
<td>64.5</td>
<td>99.6</td>
</tr>
<tr>
<td>$\sum M_0$ ($N \cdot m$)</td>
<td>1.8e13</td>
<td>13094</td>
<td>1.1</td>
<td>42.1</td>
<td>74.8</td>
<td>56.8</td>
<td>110.2</td>
</tr>
<tr>
<td>Num. of Events</td>
<td>164</td>
<td>13932</td>
<td>0.86</td>
<td>34.5</td>
<td>75</td>
<td>63.9</td>
<td>112.1</td>
</tr>
</tbody>
</table>

**Table 1:** The maximum of each quantity of interest from the CFRAC design and the input parameters at these maximum values.

- Running 1 CFRAC simulation takes 3 – 12 hours.
- Evaluating the emulator for CFRAC at the $>> 1000$ untested inputs takes $<< 10$ minutes with training!
Figure 4: Output of all six emulators. Black marks are the coordinates that correspond to the maximum values of the following: cumulative volume change during seismicity (triangles), cumulative displacement change during seismicity (pentagrams), cumulative volume change during seismicity over total volume change (hexagrams), cumulative displacement change during seismicity over total displacement change (squares), cumulative moment (circles), and the number of events (diamonds).
1. The parallel partial linked emulator accounts for both coupling in the computer model and spatial or temporal output for the quantity of interest.

2. Parallel partial linked emulators outperform other emulation strategies based on the performance metrics of the root mean squared error, average length of the 95% credible interval and coverage.

3. Emulation is a powerful tool to predict occurrence of volume creation.

4. Volume creation does not necessarily occur with the largest number of microseismic events; in fact, fractures might open and slide simultaneously even with smaller number of events is occurring.