Bayes Linear Methods for Multiscale Emulation of the Gullfaks Oil Reservoir

Jonathan Cumming

June 30, 2008
Oil Reservoirs

- A oil reservoir is an underground region of porous rock which contains and oil and/or gas.
- The hydrocarbons are trapped above by a layer of impermeable rock and below by a body of water, thus creating the reservoir.
- The oil and gas are pumped out of the reservoir (for £), and sometimes fluids is pumped into the reservoir (to boost production).
- The purpose of the simulator is to model the flows and distributions of contents of the reservoir over time.
Each cell in the reservoir has a collection of associated parameters:

- **Permeability** – dictates how easily fluids flow through the reservoir. Directionally dependent.
- **Porosity** – how much oil/water/gas can be contained within a fixed volume of rock

There are also several other parameters:

- **Fault transmissibility** – how easily fluids can flow through geological fault lines
- **Aquifer features** – size, permeability and porosity of the aquifer
- **Saturation properties** – intrinsic properties of the oil and its behaviour when mixed with water and gas

Since there are a huge number of these cells in the reservoir it is common to use a scalar multiplier over the whole reservoir (or subregions) to adjust values.
The model outputs comprise the behaviour of the various wells and injectors in the reservoir.

Output is a time series on the following variables for each well:

- **Pressures**  Bottom-hole pressure, Tubing head pressure
- **Production/Injection rates and totals**  for each of oil, water and gas.
- **Fluid ratios**  Water cut, Gas-oil ratio

The resolution of the time series can be varied from months to years.

With a large number of wells, daily output, or a long operating period there will be a *lot* of output data.
Gullfaks

- Gullfaks is a Norwegian hydrocarbon reservoir located in the North Sea
- The model is based on a grid of size $38 \times 87 \times 25$ and contains 43 production and 13 injection wells
- The model simulates 10 years of production taking 1.5–3 hours per simulation
- Inputs: Field multipliers for porosity ($\phi$), permeabilities ($k_x, k_z$), critical saturation ($crw$), and aquifer properties ($A_p, A_h$)
- Outputs: Focus on oil production rate for a 3-year period
Map

Jonathan Cumming
Bayes Linear Methods for Multiscale Emulation of the Gullfaks Oil Reservoir
Jonathan Cumming
Bayes Linear Methods for Multiscale Emulation of the Gullfaks Oil Reservoir
Preliminaries

- Obtain a coarse simulator $f^c(x)$ by coarsening the vertical gridding by a factor of 10
- 1000 runs of $f^c(x)$ in a LHC over the input parameters
- Focus on a 3-year window in the outputs, and the 10 production wells active in that period
- Consider the output quantity ‘oil production rate’ at each well
- Take 4-month averages over the time series
- Screen the wells (PV methods) – 4 wells capture 87% of the variation of the collection
Graphical Model – Coarse Emulation

\[ f^c(x_i) \rightarrow f_{\text{suff}} \]
Coarse Emulation

- $n$ is large so we fit emulators to each output individually and using information from the model runs
  \[
  f_i^c(x) = g_i(x[i])^T \beta_i^c + u_i^c(x[i]) + v_i^c(x), \tag{1}
  \]

- Determine the active variables $x[i]$ and basis functions $g_i(\cdot)$ by model selection
- Specify $E(\beta_i^c) = \hat{\beta}_i^c$, $\text{Var}(\beta_i^c) = \hat{\Sigma}_i^c$ where $\hat{\beta}_i^c$ and $\hat{\Sigma}_i^c$ are OLS estimates
- $u_i^c(x[i])$ is a residual process with a Gaussian correlation function, $v_i^c(x)$ is a nugget process with zero correlation length
- $w_i^c = u_i^c(x[i]) + v_i^c(x)$ and $\text{Var}(w_i^c)$ is typically small
- Obtain hyperparameter values by variogram methods
Graphical Model – Multiscale Emulation

\[ f^a(x_i) \rightarrow f^a_{\text{suff}} \]

\[ f^c(x_i) \rightarrow f^c_{\text{suff}} \]
Multiscale Emulation

- We consider the accurate emulator to have the form
  \[ f_i^a(x) = g_i(x_{[i]}) \beta_i^a + \beta_{wi}^a w_i^c(x) + w_i^a(x), \]  
  \( (2) \)

- We have the same \( g_i(\cdot) \) and \( x_{[i]} \) as with \( f^c \)
- Link (1) and (2) via the coefficients
  \[ \beta_{ij}^a = \rho_{ij} \beta_{ij}^c + \gamma_{ij} \]  
  \( (3) \)

- Since we will have less information on \( f^a \) we correlate the accurate emulators by correlating the \( \beta_{ij}^a \) over time
- Now perform 20 model runs and adjust \( \beta^a \) by new values
Model Runs

Jonathan Cumming
Bayes Linear Methods for Multiscale Emulation of the Gullfaks Oil Reservoir
## Emulation Summaries

<table>
<thead>
<tr>
<th>Well</th>
<th>Time</th>
<th>$x[i]$</th>
<th>No. Model Terms</th>
<th>Coarse Simulator $R^2$</th>
<th>Accurate Simulator $\tilde{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2</td>
<td>4</td>
<td>$\phi, crw, A_p$</td>
<td>9</td>
<td>0.886</td>
<td>0.951</td>
</tr>
<tr>
<td>B2</td>
<td>8</td>
<td>$\phi, crw, A_p$</td>
<td>7</td>
<td>0.959</td>
<td>0.958</td>
</tr>
<tr>
<td>B2</td>
<td>12</td>
<td>$\phi, crw, A_p$</td>
<td>10</td>
<td>0.978</td>
<td>0.995</td>
</tr>
<tr>
<td>B2</td>
<td>16</td>
<td>$\phi, crw, k_z$</td>
<td>7</td>
<td>0.970</td>
<td>0.995</td>
</tr>
<tr>
<td>B2</td>
<td>20</td>
<td>$\phi, crw, k_x$</td>
<td>11</td>
<td>0.967</td>
<td>0.986</td>
</tr>
<tr>
<td>B2</td>
<td>24</td>
<td>$\phi, crw, k_x$</td>
<td>10</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>B2</td>
<td>28</td>
<td>$\phi, crw, k_x$</td>
<td>10</td>
<td>0.975</td>
<td>0.981</td>
</tr>
<tr>
<td>B2</td>
<td>32</td>
<td>$\phi, crw, k_x$</td>
<td>11</td>
<td>0.980</td>
<td>0.951</td>
</tr>
<tr>
<td>B2</td>
<td>36</td>
<td>$\phi, crw, k_x$</td>
<td>11</td>
<td>0.983</td>
<td>0.967</td>
</tr>
</tbody>
</table>
Graphical Model – History Matching/Calibration

Jonathan Cumming
Bayes Linear Methods for Multiscale Emulation of the Gullfaks Oil Reservoir
History Matching

- We want to reduce the size of the input space $X$ by eliminating any input points whose simulator evaluations are unlikely to reproduce the behaviour of the physical system.
- Use the implausibility measure

$$I_{(i)}(x) = \frac{|E(f^a_i(x)) - z_i|^2}{\text{Var}(f^a_i(x) - z_i)},$$

- Or the multivariate form

$$I(x) = \frac{(E(f^a(x)) - z)^T \text{Var}(f^a(x) - z)^{-1}(E(f^a(x)) - z)}{q},$$

- Evaluate $I(x)$ over a grid in the input parameters and identify and exclude regions of poor match quality.
- Obtain a reduced space $X^*$ of potential matches.
Implausibility Plots

Jonathan Cumming
Bayes Linear Methods for Multiscale Emulation of the Gullfaks Oil Reservoir
Refocusing

- Make the restriction $X^* = \{x : l(x) \leq 4\} \simeq \{x : \phi < 0.79\}$
- Now consider final 4 time points in original data, plus an additional point 1 year beyond the end of the previous series to be forecast
- Since reducing the space many of the old model runs are no longer valid, so supplement with additional evaluations
- $262+100$ coarse runs, $6+20$ accurate runs
- Re-fit the coarse and fine emulators, using the old emulator structure as a starting point
- Allow for new AVs and basis functions
Graphical Model – Forecasting

Jonathan Cumming
Bayes Linear Methods for Multiscale Emulation of the Gullfaks Oil Reservoir
We now forecast the “future” system value $y_P$ given the observed historical data $z_H$ (using our collection of emulators)

$$E_{z_H}(y_P) = E(y_P) + \text{Cov}(y_P, z_H)\text{Var}(z_H)^{-1}(z_H - E(z_H))$$

$$\text{Var}_{z_H}(y_P) = \text{Var}(y_P) - \text{Cov}(y_P, z_H)\text{Var}(z_H)^{-1}\text{Cov}(z_H, y_P).$$

Plug in $z_H = y_H + e_H$ and $y = f^a(x^*) + \epsilon$

Requires specification of model discrepancy $\epsilon$ and some prior specification for $x^*$

$$E_{z_H}(y_P) = \mu^*_P + (\Sigma^*_P + \Sigma^e_P)(\Sigma^*_H + \Sigma^e_H + \Sigma^e_H)^{-1}(z_H - \mu^*_H)$$

$$\text{Var}_{z_H}(y_P) = (\Sigma^*_P + \Sigma^e_P) - (\Sigma^*_P + \Sigma^e_P)(\Sigma^*_H + \Sigma^e_H + \Sigma^e_H)^{-1} (\Sigma^*_H + \Sigma^e_H).$$
Forecasting Results

Jonathan Cumming
Bayes Linear Methods for Multiscale Emulation of the Gullfaks Oil Reservoir
In summation

- Consider a more complex version of the simulator – regional multipliers for more inputs, more wells
- Incorporate spatial correlation between emulators of different wells
- More sophisticated assessments of discrepancy accounting for spatio-temporal features
- More than two simulator scales