Monte Carlo filtering and regional sensitivity analysis (RSA)

M. Ratto, IPSC-JRC, European Commission

DYNARE Workshop, CEF2011, San Francisco

http://eemc.jrc.ec.europa.eu
Background

• environmental sciences, early 80’s;

• complex numerical and analytical models, based on first principles, conservation laws, ... ;

• ill-defined parameters, competing model structures (different constitutive equations, different types of process considered, spatial/temporal resolution, ... )

• need to establish magnitude and sources of prediction uncertainty;

• Monte Carlo simulation analyses.
MC filtering and RSA

One of the earliest landmark application of MC simulation: eutrophication in the Peel Inlet, SW Australia.

Regional Sensitivity Analysis developed and employed.


MC filtering and RSA

Two tasks for RSA:

• qualitative definition of the system behaviour

  [a set of constraints: thresholds, ceilings, time bounds based on available information on the system];

• binary classification of model outputs based on the specified behaviour definition.

  [qualifies a simulation as \textit{behaviour} \( (B) \) if the model output lies within constraints, \textit{non-behaviour} \( (B^c) \) otherwise]
MC filtering and RSA

Define a range for \( k \) input factors \( x_i \), \( 1 \leq i \leq k \), reflecting uncertainty in parameters and model constituent hypotheses.

Each Monte Carlo simulation is associated to a vector of values of the input factors.

Classifying simulations as either \( B \) or \( \overline{B} \), a set of binary elements is defined allowing to distinguish two sub-sets for each \( x_i \):

\[
(x_i \mid B) \quad (x_i \mid \overline{B})
\]
MC filtering and RSA

http://eemc.jrc.ec.europa.eu
MC filtering and RSA

The Kolmogorov-Smirnov two-sample test (two-sided version) is performed for each factor \textit{independently}

\[
H_0 : f_m(x_i | B) = f_n(x_i | \bar{B}) \\
H_1 : f_m(x_i | B) \neq f_n(x_i | \bar{B})
\]

Test statistic:

\[
d_{m,n}(x_i) = \sup_{y} \left\| F_m(x_i | B) - F_n(x_i | \bar{B}) \right\|
\]

where

\( F \) are marginal cumulative probability functions, \( f \) are probability density functions

http://eemc.jrc.ec.europa.eu
MC filtering and RSA

At what significance level $\alpha$ does the computed value of $d_{m,n}$ determine the rejection of $H_0$?

$\alpha$ is the probability to reject $H_0$ when it is true (i.e., to recognise a factor as important when it is not).

The importance of each parameter is inversely related to this significance level.
MC filtering and RSA

Derive the critical level $D_\alpha$ at which the computed value of $d_{m,n}$ determines the rejection of $H_o$ (the smaller $\alpha$, the higher $D_\alpha$).

If $d_{m,n} > D_\alpha$, then $H_o$ is rejected at significance level $\alpha$. 
MC filtering and RSA

\begin{align*}
\text{Prior} & \quad F_{m}(X_{i}|B) \\
\phantom{\text{Prior}} & \quad F_{n}(X_{i}|\overline{B})
\end{align*}

\[ d_{m,n} \]
RSA: comment

RSA has many ‘global’ properties:

The whole range of value of input factors is considered, all factors are varied at the same time.

CAVEAT:

RSA classification analyses univariate marginal distributions.

Lacks for interaction structure.
Mapping stability: example (1)

Phillips curve:

GDP \( y \), inflation \( i \), output gap \( c \)

\[
\begin{align*}
y_t &= p_t + c_t \\
p_t &= p_{t-1} + a_{p,t} \\
c_t &= 2A \cos(2\pi / \tau)c_{t-1} - A^2 c_{t-2} + a_{c,t} \\
i_t &= \omega \cdot i_{t-1} + (1 - \omega)E_t i_{t+1} + t_y c_t + a_{i,t}
\end{align*}
\]
Mapping stability: example (1)

Stability is assured if

$$\omega > 0.5$$
Mapping stability: example (1)
Mapping stability: example (2)

Phillips curve, extended version:
GDP $y$, inflation $i$, output gap $c$

\[ y_t = p_t + c_t \]
\[ p_t = p_{t-1} + a_{p,t} \]
\[ c_t = 2A \cos \left( \frac{2\pi}{\tau} \right) c_{t-1} - A^2 c_{t-2} + a_{c,t} \]
\[ i_t = \omega_b \cdot i_{t-1} + \omega_f E_t i_{t+1} + t_y c_t + a_{i,t} \]
Mapping stability: example (2)

Stability is assured if

$$\omega_b + \omega_f < 1$$
Mapping stability: example (2)
Mapping stability: example (2)

Bivariate analysis

\[ cc = -0.50539 \]

\[ \omega_f \]

\[ \omega_b \]
Mapping stability: example (3)

- As the complexity of the model structure and its parameterization increases, it is not trivial to know a priori the set of model coefficients assuring stability;
Mapping stability : example (3)

• The Lubik Schorfheide (2005) model

• 12 parameters;
Lubik Schorfheide model (2005)

\[
y_t = \frac{E_t y_{t+1} - \left[\tau + \alpha(2 - \alpha)(1 - \tau)\right](R_t - E_t \pi_{t+1})}{\tau + \alpha(2 - \alpha)(1 - \tau)} - \alpha E_t \Delta q_{t+1} - \alpha(2 - \alpha) \frac{1 - \tau}{\tau} \Delta y_{t+1} - E_t \bar{z}_{t+1}
\]

\[
\pi_t = \beta E_t \pi_{t+1} + \alpha \beta E_t \Delta q_{t+1} - \alpha \Delta q_t + \frac{k}{\tau + \alpha(2 - \alpha)(1 - \tau)}(y_t - \bar{y}_t)
\]

\[
\pi_t = \Delta e_t + (1 - \alpha) \Delta q_t + \pi^*_t
\]

\[
R_t = \rho_R R_{t-1} + (1 - \rho_R)(\psi_1 \pi_t + \psi_2(y_t - \bar{y}_t) + \psi_3 \Delta e_t) + e_{R,t}
\]

\[
\Delta q_t = \rho_q \Delta q_{t-1} + e_{q,t}
\]

\[
y_s = \rho_y^* y_{t-1} + e_{y^*,t}
\]

\[
\pi^*_t = \rho_{\pi^*} \pi^*_{t-1} + e_{\pi^*,t}
\]

\[
\bar{z}_t = \rho_z \bar{z}_{t-1} + e_{z,t}
\]
# Lubik Schorfheide model (2005)

<table>
<thead>
<tr>
<th>Name</th>
<th>Density</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi_1$</td>
<td>Gamma</td>
<td>1.5</td>
<td>0.5</td>
</tr>
<tr>
<td>$\psi_2$</td>
<td>Gamma</td>
<td>0.25</td>
<td>0.125</td>
</tr>
<tr>
<td>$\psi_3$</td>
<td>Gamma</td>
<td>0.25</td>
<td>0.125</td>
</tr>
<tr>
<td>$\rho_R$</td>
<td>Beta</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Beta</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>$r$</td>
<td>Gamma</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Gamma</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Gamma</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>$\rho_q$</td>
<td>Gamma</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>$\rho_z$</td>
<td>Gamma</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>$\rho_y^*$</td>
<td>Gamma</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>$\rho_{\pi}^*$</td>
<td>Gamma</td>
<td>0.7</td>
<td>0.15</td>
</tr>
<tr>
<td>$\sigma_R$</td>
<td>InvGamma</td>
<td>1.2533</td>
<td>0.6551</td>
</tr>
<tr>
<td>$\sigma_q$</td>
<td>InvGamma</td>
<td>2.5066</td>
<td>1.3103</td>
</tr>
<tr>
<td>$\sigma_z$</td>
<td>InvGamma</td>
<td>1.2533</td>
<td>0.6551</td>
</tr>
<tr>
<td>$\sigma_y^*$</td>
<td>InvGamma</td>
<td>1.2533</td>
<td>0.6551</td>
</tr>
<tr>
<td>$\sigma_{\pi}^*$</td>
<td>InvGamma</td>
<td>1.88</td>
<td>0.9827</td>
</tr>
</tbody>
</table>
Lubik Schorfheide model (2005)

http://eemc.jrc.ec.europa.eu
## 3.1 Sampling options

<table>
<thead>
<tr>
<th>Option Name</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nsam</td>
<td>2048</td>
<td>Size of MC sample</td>
</tr>
</tbody>
</table>
| ilptau            | 1       | \(1 = \text{use } L_{P_r} \text{ quasi-Monte Carlo} \)
                  |         | \(0 = \text{use LHS Monte Carlo} \)                                       |
| pprior            | 1       | \(1 = \text{sample from prior distributions} \)
                  |         | \(0 = \text{sample from multivariate normal} \) N(\hat{\theta}, \Sigma) \), \( \hat{\theta} \) is posterior mode
                  |         | \( \Sigma = H^{-1}, H \) is Hessian at the mode                           |
| prior_range       | 1       | \(1 = \text{sample uniformly from prior ranges} \)
                  |         | \(0 = \text{sample from prior distributions: this requires MATLAB Statistics Toolbox} \) |
| morris            | 0       | \(0 = \text{no Morris sampling for screening} \)
                  |         | \(1 = \text{Morris sampling for screening} \)                              |
| morris_nliv       | 6       | number of levels in Morris design                                           |
| morris_ntra       | 20      | number of trajectories in Morris design                                     |
| ppost             | 0       | \(0 = \text{don’t use Metropolis posterior sample} \)
                  |         | \(1 = \text{use Metropolis posterior sample: this overrides any other sampling option!} \)
## 3.2 Stability mapping

<table>
<thead>
<tr>
<th>option name</th>
<th>default</th>
<th>description</th>
</tr>
</thead>
</table>
| stab         | 1       | 1 = perform stability mapping  
|              |         | 0 = no stability mapping is performed |
| load_stab    | 0       | 0 = generate a new sample  
|              |         | 1 = load a previously created sample |
| alpha2_stab  | 0.4     | critical value for correlations $\rho$ in filtered samples:  
|              |         | plot couples of parameters with  
|              |         | $|\rho| > \text{alpha2}_\text{stab}$ |
| ksstat       | 0.1     | critical value for Smirnov statistics $d$:  
|              |         | plot parameters with $d > \text{ksstat}$ |
Mapping the fit

How do parameters adjust to fit the multivariate set of observed covariates?

Are there trade-off's?
Mapping the fit

- Sample structural coefficients prior distributions (or ranges);
  OR
- Use Metropolis posterior sample;

Compute RMSE’s of the 1-step ahead model predictions for each of the N observed series.
Mapping the fit

N filtering rules: defining as $B$ the samples corresponding to the best 10% RMSE’s, -$B$ otherwise.

For each parameter we get $N$ ‘behavioural’ sets:

$f_j(X_i/B); j = 1, ..., N$
Mapping the fit

Trade off occurs when:
1) one structural parameter drives significantly the fit of more than one observed series;

AND

2) $f_j(X_i|B) \neq f_k(X_i|B)$
Mapping the fit: prior

http://eemc.jrc.ec.europa.eu
Mapping the fit: prior

http://eemc.jrc.ec.europa.eu
Mapping the fit: posterior

\[
\sigma_R
\]

\[
\sigma_q
\]

\[
\sigma_z
\]

\[
\sigma_{y^*}
\]

\[
\sigma_{\pi^*}
\]

\[
\psi_1
\]
Mapping the fit: posterior

- \( \psi_2 \)
- \( \psi_3 \)
- \( \rho_R \)
- \( \alpha \)
- \( r \)
- \( \kappa \)
Mapping the fit: posterior
Mapping the fit: posterior

\[ R = \ldots + (1 - \rho_R)(\psi_1(\Delta e_t + (1 - \alpha)\Delta q_t + \pi_t^*) + \ldots + \psi_3 \Delta e_t) + \ldots \]

\[ = \ldots + (1 - \rho_R)(\psi_1((1 - \alpha)\Delta q_t + \pi_t^*) + (\psi_1 + \psi_3)\Delta e_t) + \ldots \]
### 3.4 Mapping the fit

<table>
<thead>
<tr>
<th>option name</th>
<th>default</th>
<th>description</th>
</tr>
</thead>
</table>
| rmse        | 0       | 0 = no RMSE analysis  
             |         | 1 = do RMSE analysis |
| load_rmse   | 0       | 0 = make a new RMSE analysis  
             |         | 1 = load previous RMSE analysis |
| lik_only    | 0       | 0 = compute RMSE’s for all observed series  
             |         | 1 = compute only likelihood and posterior |
| var_rmse    | varobs  | list of observed series to be considered |
| pfilt_rmse  | 0.1     | filtering threshold for RMSE’s: default it to filter the best 10% for each observed series |
| istart_rmse | 1       | start computing RMSE’s from istart_rmse: use 2 to avoid big initial error |
| alpha_rmse  | 0.002   | critical value for Smirnov statistics $d$: plot parameters with $d > alpha_rmse$ |
| alpha2_rmse | 1       | critical value for correlation $\rho$  
             |         | plot couples of parameters with $|\rho| > alpha2_rmse$ |
| glue        | 0       | prepare for GLUE graphical interface |

http://eemc.jrc.ec.europa.eu
RSA: problems

Correlation and interaction structures of the $B$ subset [$\Rightarrow$ also Beck’s review, 1987]:

- Smirnov test is a sufficient test only if $H_0$ is rejected (i.e. the factor is IMPORTANT);

- any covariance structure induced by the classification is not detected by the univariate $d_{m,n}$ statistic.
  
  e.g. factors combined as products or quotients may compensate

- bivariate correlation analysis is not revealing, either.
Example 1. $Y = X_1 \times X_2$, $X_1, X_2 \sim U[-0.5, 0.5]$

Behavioural runs: $Y > 0$.

Scatter plot of the B subset in the $X_1, X_2$ plane

Smirnov test fails

Correlation analysis would help (e.g. PCA)
RSA problems: example (1)

Example 1. Cumulative distributions of $X_1$
(Behavioural runs: $Y > 0$.)
RSA problems: example (1)

In this case, a correlation analysis would be helpful.

the correlation coefficient of \((X_1, X_2)\) for the B set is \(\rho=0.75\).

This suggests performing a Principal Component Analysis on the B set, obtaining the two components (the eigenvectors of the correlation matrix):
RSA problems: example (1)

PC1 = (0.7079, 0.7063); accounting for the 87.5% of the variation of the B set.

PC2 = (-0.7063, 0.7079); accounting for the 12.5% of the variation of the B set.

The direction of the principal component PC1 (associated with the highest eigenvalue) indicates the privileged orientation for acceptable runs.
RSA problems: example (1)
RSA problems: example (1)

http://eemc.jrc.ec.europa.eu
RSA problems: example (1)

Now the level of significance for rejecting the null hypothesis when it is true is very small ($\alpha<0.1\%$), implying a very strong relevance of the linear combinations of the two input factors, defined by the principal component analysis.
RSA problems: example (2)

Example 2. \( Y = X_1^2 + X_2^2, \ X_1, X_2 \sim U[-0.5, 0.5] \)

Behavioural runs: \([0.2 < Y < 0.25]\)

Scatter plot of the B subset in the \( X_1, X_2 \) plane

Smirnov test fails
Correlation fails as well (sample \( \rho \approx -0.04 \))
RSA problems (ctd.)

To address the RSA limitations and to better understand the impact of uncertainty and interaction in the high-dimensional parameter spaces of models, Spear et al. (1994) developed the computer intensive tree-structured density estimation technique (TSDE).

Interesting applications of TSDE in environmental sciences can be found in Spear (1997), Grieb et al. (1999) and Osidelle and Beck (2001).