Application of the control variate technique to the estimation of total sensitivity indices

S. KUCHERENKO
B. DELPUECH
Imperial College London (UK)
s.kucherenko@imperial.ac.uk

B. IOOSS
Electricité de France (France)

S. TARANTOLA
Joint Research Centre of the European Commission, Ispra (Italy)

Outline

ANOVA decomposition and Sobol' Sensitivity Indices.

Improved direct formula for evaluation of Sobol' Main Effect Sensitivity Indices with small values.

Monte Carlo estimates. Variance reduction techniques.

Improved direct formula for evaluation of Sobol' Total Effect Sensitivity Indices using control variates.

Results.

ANOVA decomposition and Sensitivity Indices

Consider a model

x is a vector of input variables

f(x) is square integrable

$$Y = f(x)$$
$$x = (x_1, x_2, \ldots, x_k) \in H^n$$
$$0 \le x_i \le 1$$

ANOVA decomposition is unique if variables are independent

$$Y = f(x) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,k}(x_1, x_2, \ldots, x_k),$$
$$\int_0^1 f_{i_1 \ldots i_s}(x_{i_1}, , \ldots, x_{i_s}) dx_{i_k} = 0, \ \forall k, \ 1 \le k \le s, \rightarrow \int_0^1 f_{i_1 \ldots i_s} f_{i_1 \ldots i_l} dx_{i_k} dx_{i_l} = 0, \ \forall i_k \ne i_l$$

Variance decomposition:

$$D^2 = \sum_i D_i^2 + \sum_{i,j} D_{ij}^2 + \ldots D_{1,2,\ldots,n}^2$$

Sobol' SI:

$$1 = \sum_{i=1}^{k} S_i + \sum_{i<j} S_{ij} + \sum_{i<j<l} S_{ijl} + \ldots + S_{1,2,\ldots,k}$$

Evaluation of Sobol' Sensitivity Indices

Consider two subsets of the total input vector: $x = \left(y, z \right)$

Direct formulas for evaluation of Sobol' Sensitivity indices

$$S_y = \frac{1}{D^2} [\int_0^1 f(y, z')^2 \, dydzdz' - f_0^2,$$

$$S_y^{tot} = \frac{1}{2D^2} \int_0^1 [f(y,z) - f(y',z)]^2 \, dydzdz',$$

$$D^2 = \int_0^1 f^2(y, z) \, dydz - f_0^2$$

Evaluation of sensitivity indices is reduced to high-dimensional integration using MC/QMC methods

Evaluation of Sobol' Main Effect Sensitivity Indices with small values

$\underline{x} = (y, z), \;\; x' = (y', z')$

using values $f(y,z), \;\; f(y,z'), \;\; f(y',z)$

$$S_y = \frac{1}{D^2}\int_0^1 f(y,z) f(y,z')\, dydzdz' - f_0^2$$

for small indices $S_y << 1$

$$\int_0^1 f(y,z) f(y,z')\, dydzdz' \approx f_0^2$$

$\rightarrow$ loss of accuracy

Improved formula for evaluation of Sobol' Main Effect Sensitivity Indices

Notice that $f_0^2 = \int_0^1 f(y,z)dydz \int_0^1 f(y',z')\, dy'dz'$

using values $f(y,z),\ \ f(y',z),\ \ f(y',z')$

$$S_y = \frac{1}{D^2}\int_0^1 f(y,z) f(y',z)\, dydy'dz - f_0^2 \rightarrow$$

$$S_y = \frac{1}{D^2}[\int_0^1 f(y',z')[f(y',z) - f(y,z)]\, dydy'dzdz'$$

- gives much more accurate results (Kucherenko, Mauntz, 2002)

Additional advantage (Saltelli 2002):

Requires $N(n{+}2)$ model evalution rather than

$N(2n{+}1)$ for original Sobol' formulas.

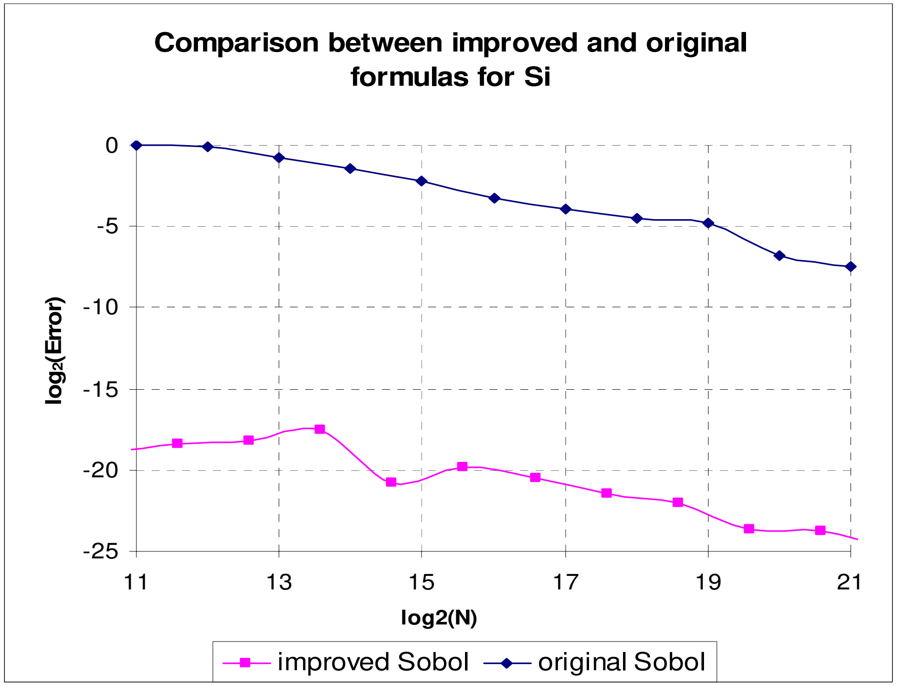
Improved formulas for small indices: S. Kucherenko: 2002, 2007:

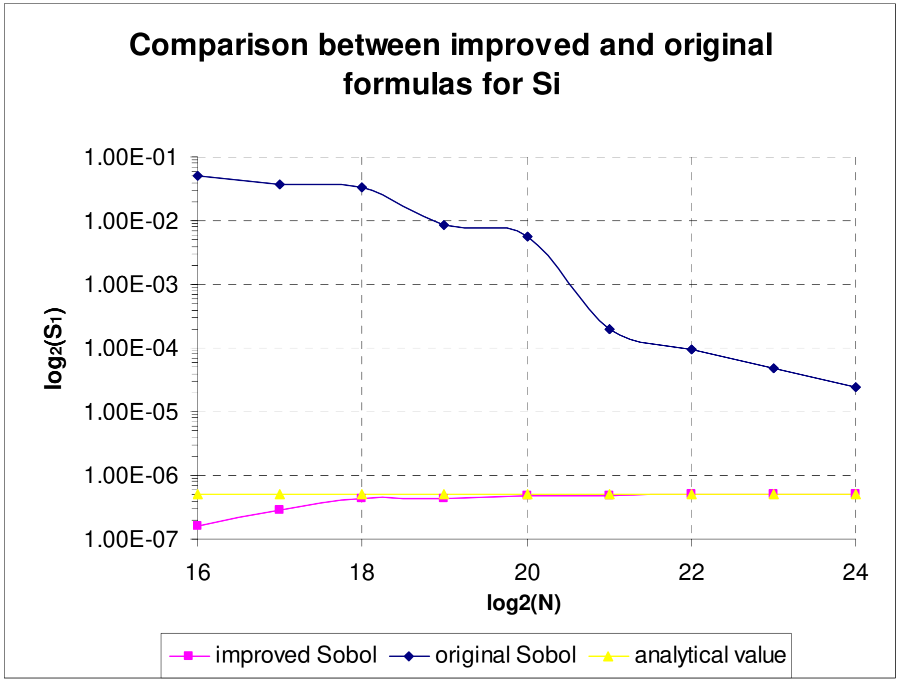
Further improvements: Sobol' and Mishetskaya 2007, A. Owen 2012

Improved formula for Sobol' Main Effect Sensitivity Indices. Test

$$f(x)=\sum_{i=1}^{n} i x_i\,,\quad S_i = S^T = \frac{6}{n(n+1)(2n+1)}$$

$$n=180,\quad S_1 = 5.1\ 10^{-7}$$





Monte Carlo integration

$$I[f] = \int_{H^n} f(\vec{x})d\vec{x}$$

Monte Carlo :

$$I[f] = E[f(\vec{x})]$$

Crude Monte Carlo Estimate : $I_N[f] = \frac{1}{N}\sum_{i=1}^{N} f(\vec{z}_i)$

$\{\vec{z}_i\}$ − is a sequence of random (MC) or

quasi-random (QMC) points in H^n

Error: $\varepsilon = \left| I[f] - I_N[f] \right|$

$$\varepsilon_N^{MC} = (E(\varepsilon^2))^{1/2} = \frac{\sigma(f)}{N^{1/2}}$$

Convergence does not depent on dimensionality

but it is slow

How to improve MC ?

Slow convergence: $\varepsilon_N = \dfrac{\sigma\ (f)}{N^{1/2}}$

To improve MC convergence:

Decrease $\sigma\ (f)$ by applying variance reduction techniques:

- antithetic variables;
- control variates;
- stratified sampling;
- importance sampling.

Monte Carlo Integration. Variance reduction. Control Variate Method

Define a new function: $\bar{f}(x) = f(x) + C(g(x) - \mu_g)$

The control $g(x)$ is sampled along with $f(x)$

$\mu_g = \int g(x)dx$, so that $E[\bar{f}(x)] = E[f(x)]$

Variance Decomposition

$$Var[\bar{f}] = Var[f] + 2C \cdot Cov(f, g) + C^2 Var[g]$$

Where $Cov(f, g)$ is the covarience between $f(x)$ and $g(x)$

The optimal control parameter C is obtained by minimizing the variance:

$$C = -\frac{Cov(f, g)}{Var(g)}$$

Reduction of variance: $Var[\bar{f}] = Var[f]\left[1 - \frac{Cov(f, g)}{Var[f]Var[g]}\right]$

Evaluation of Sobol' Total Sensitivity Indices

Recall Sobol-Jansen formula for one input variable:

$$x = (x_j, x_{\sim j})$$

$$S_j^{tot} = \frac{1}{2D^2} \int_0^1 [f(x) - f(x_j', x_{\sim j})]^2 \, dydzdz'$$

Ideally we need to find the control function for $[f(x) - f(x_j', x_{\sim j})]$

But it is not possible in a general case

Evaluation of Sobol' Total Sensitivity Indices using control variates approach

ANOVA decomposition:

$$f(x) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \ldots + f_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n)$$

Use the first order terms as "control variates" $g(x) = f_0 + \sum_{i=1}^{n} f_i(x_i)$

Theorem: $$S_j^{tot} = \frac{1}{2D^2} \int_0^1 [f(x) - f_j(x_j) - [f(x') - f_j(x_j')]]^2 \, dx dx_j' + S_j$$

It is presented for the case of a single variable x_j. The same approach can be easily generalised for the case of a set of variables.

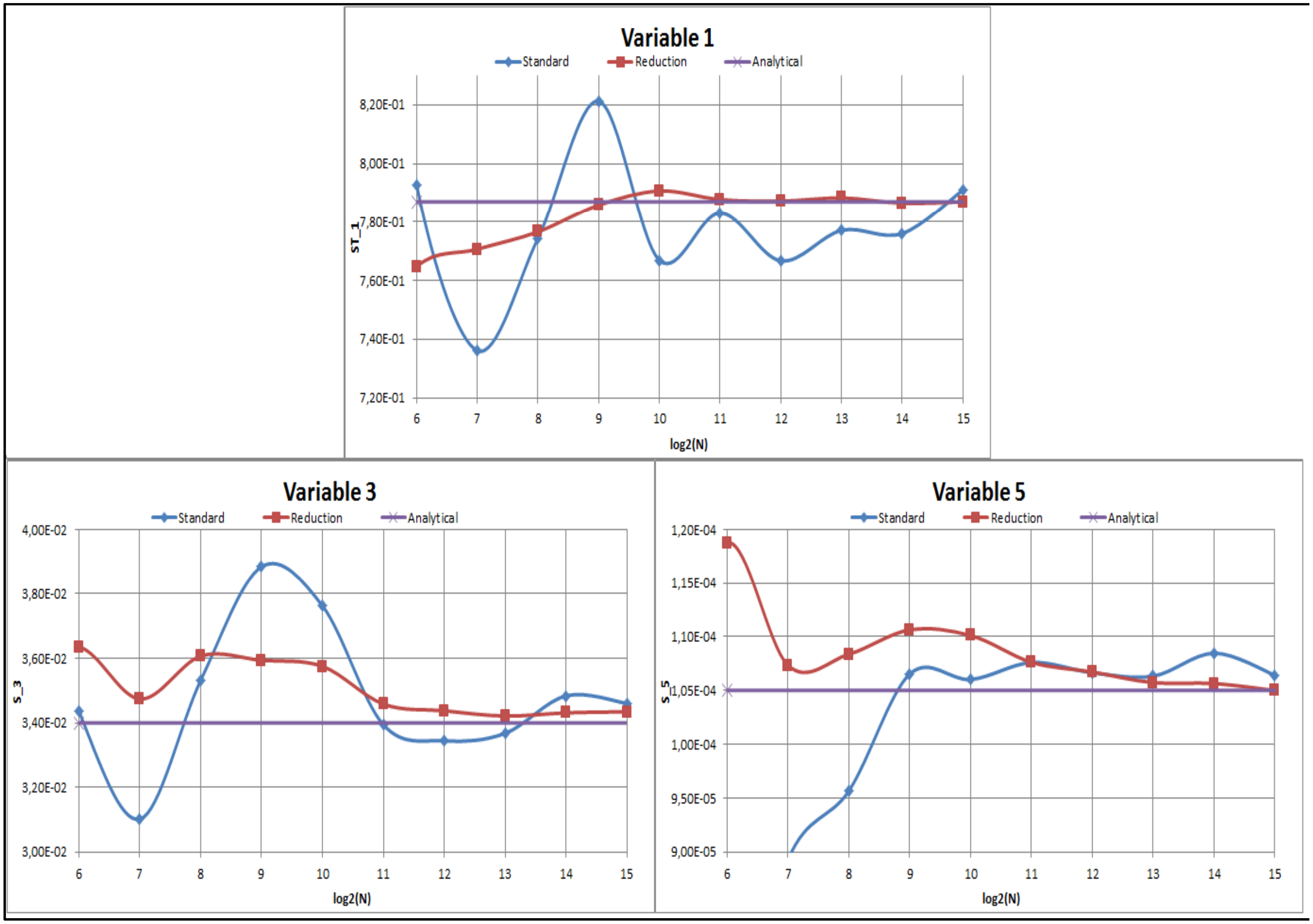
The efficiency of this method can be increased further by adding higher order terms to control variate $g(x) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j)$

Can result in significant improvement in the efficiency of MC estimates, provided the first order terms $f_j(x_j)$ and corresponding S_j are known. They can be found for A) explicitly given analytical function; B) functions approximated by metamodels.

Improved formula for Sobol' Total Sensitivity Indices, Test: G-function

$$f=\prod_{i=1}^{n} g_i(x_i),\quad g_i(x_i)=\frac{|4x_i-2|+a_i}{1+a_i},$$

$$a_i=\{0,\ 1,\ 4.5,\ 9,\ 99,\ 99,\ 99,\ 99\}$$

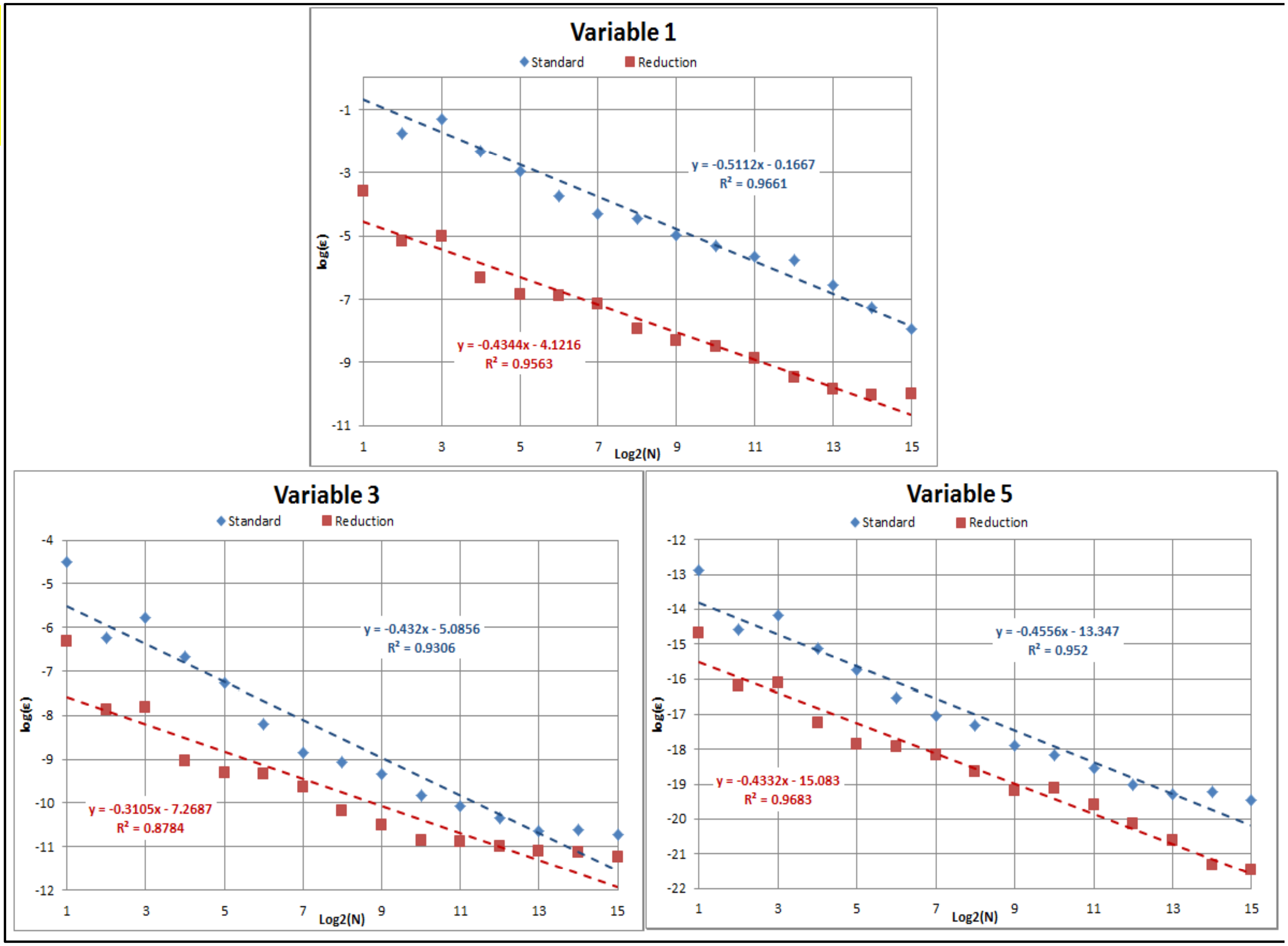


Comparison of convergences numerically computed total SI to the analytical values. Standard – blue, Reduced variance- red.

Improved formula for Sobol' Total Sensitivity Indices, Test: G-function

$$f=\prod_{i=1}^{n} g_i(x_i),\ \ g_i(x_i)=\frac{|4x_i-2|+a_i}{1+a_i},$$

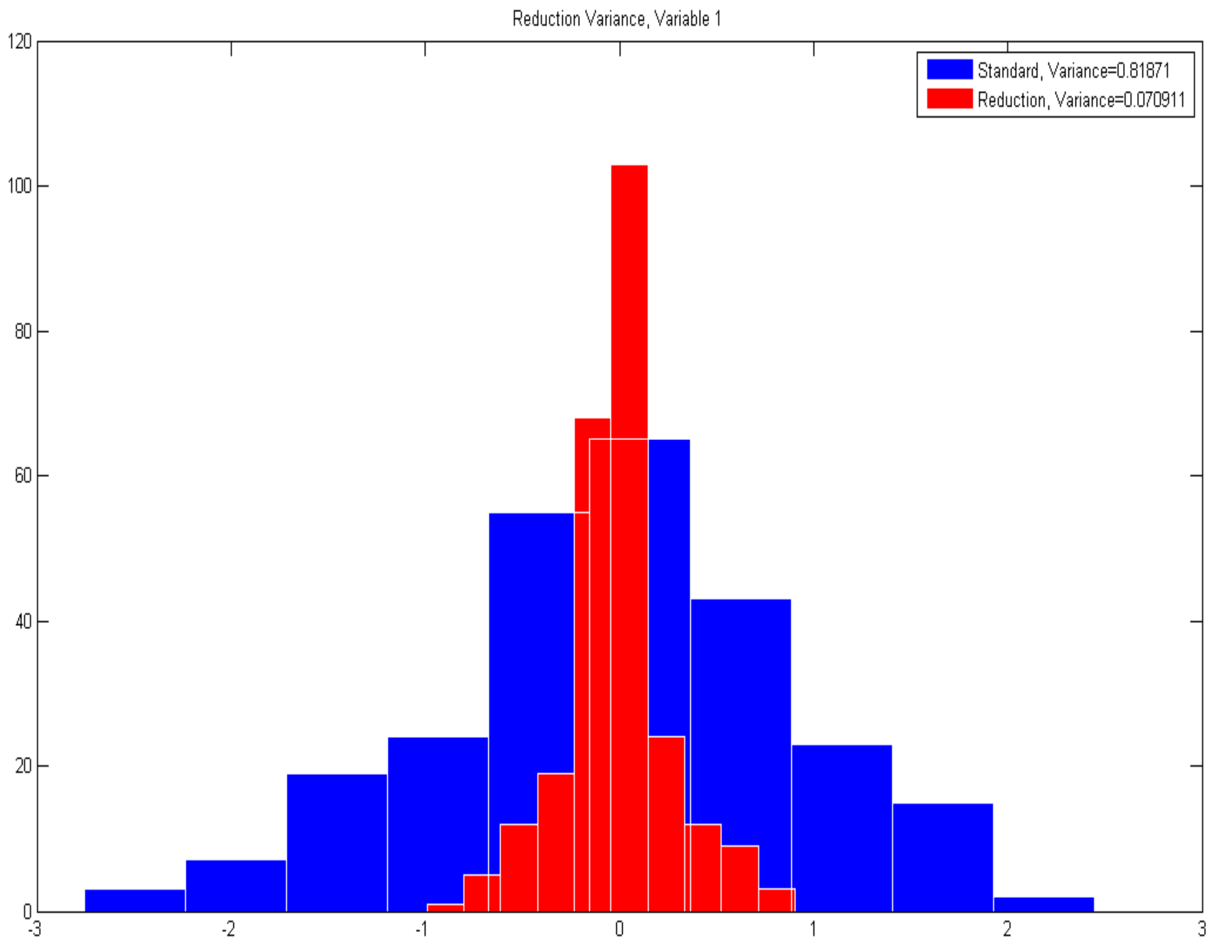
$$a_i=\{0,\ 1,\ 4.5,\ 9,\ 99,\ 99,\ 99,\ 99\}$$



Root Mean Square Error vs. Number of sampled points

Improved formula for Sobol' Total Sensitivity Indices, Test: G-function

$$f = \prod_{i=1}^{n} g_i(x_i),\ \ g_i(x_i) = \frac{|4x_i - 2| + a_i}{1 + a_i},$$
$$a_i = \{0,\ 1,\ 4.5,\ 9,\ 99,\ 99,\ 99,\ 99\}$$



Comparison of Histograms:

$$[f(x) - f(x')]$$

$$[f(x) - f_j(x_j) - [f(x') - f_j(x_j^{'})]]$$

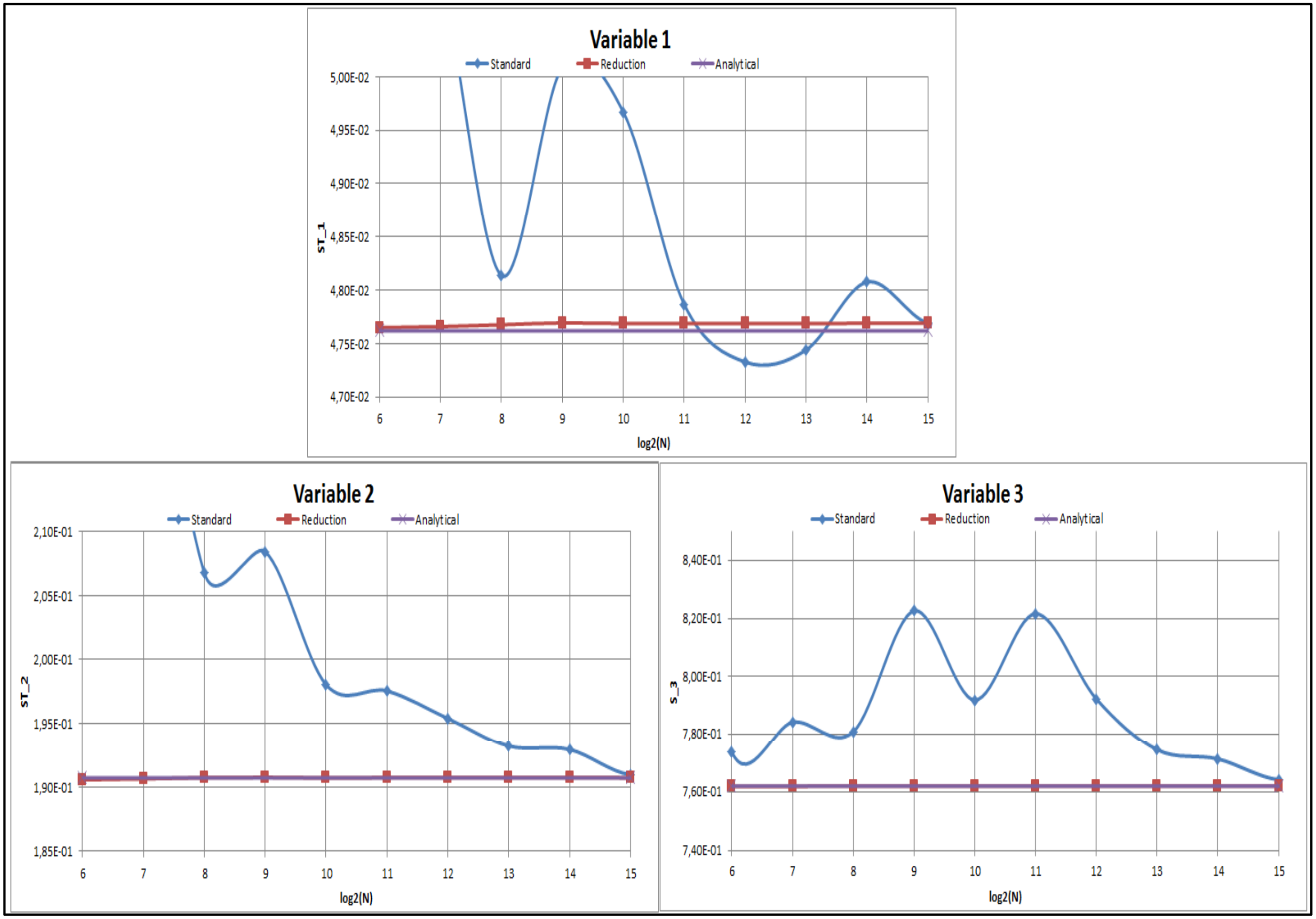
Dramatic variance reduction !

Improved formula for Sobol' Total SI,
Test: Modified G-function

$$f = \prod_{i=1}^{n} g_i(x_i),$$

$$g_i(x_i) = \frac{|4x_i - 2| + 2 + 3a_i}{1 + a_i},$$

$$a_i = \{19,\ 9,\ 4\}$$



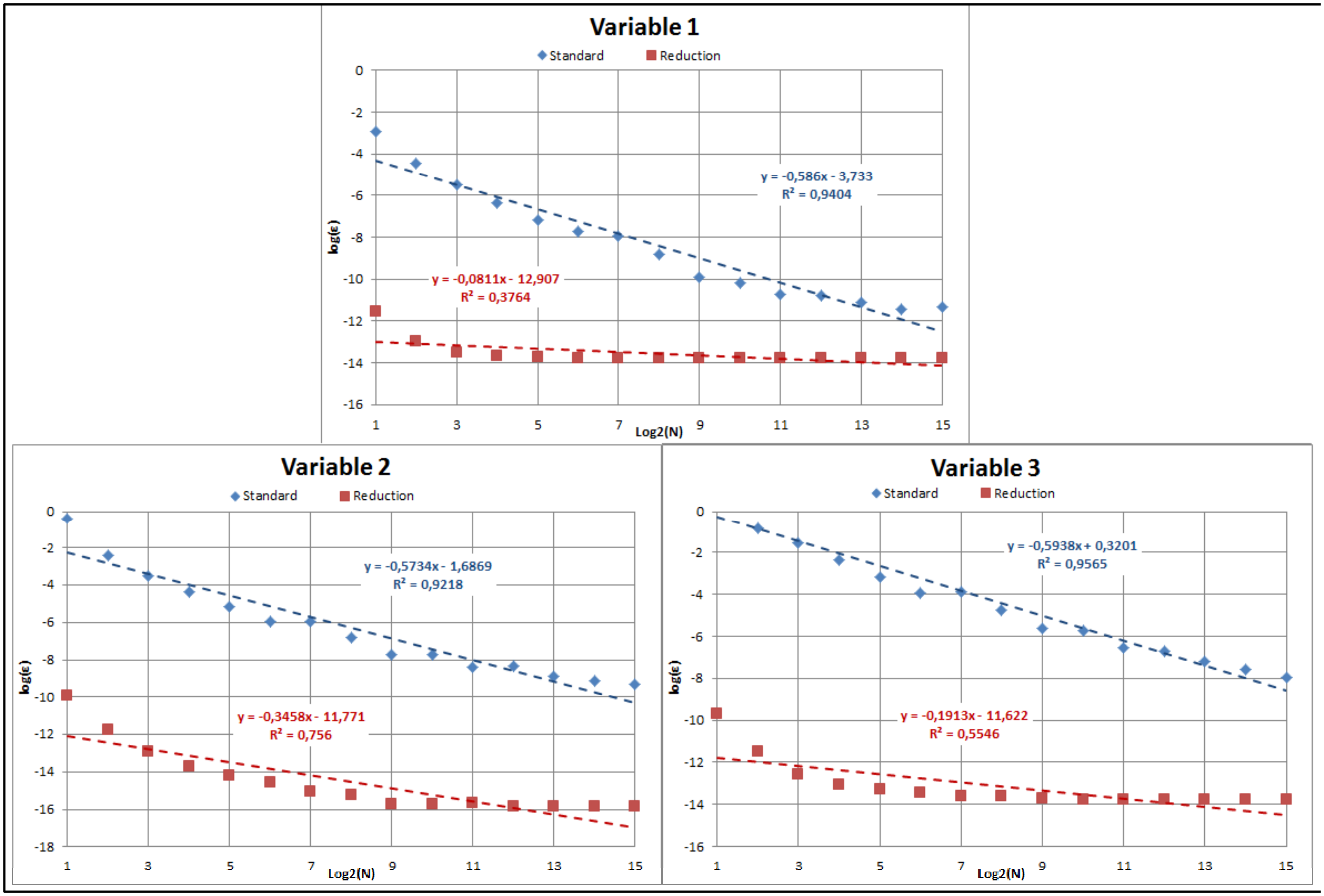
Comparison of convergences numerically computed total SI to the analytical values. Standard – blue, Reduced variance- red.

Improved formula for Sobol' Total Sensitivity Indices, Test: Modified G-function

$$f = \prod_{i=1}^{n} g_i(x_i),$$

$$g_i(x_i) = \frac{|4x_i - 2| + 2 + 3a_i}{1 + a_i},$$

$$a_i = \{19,\ 9,\ 4\}$$



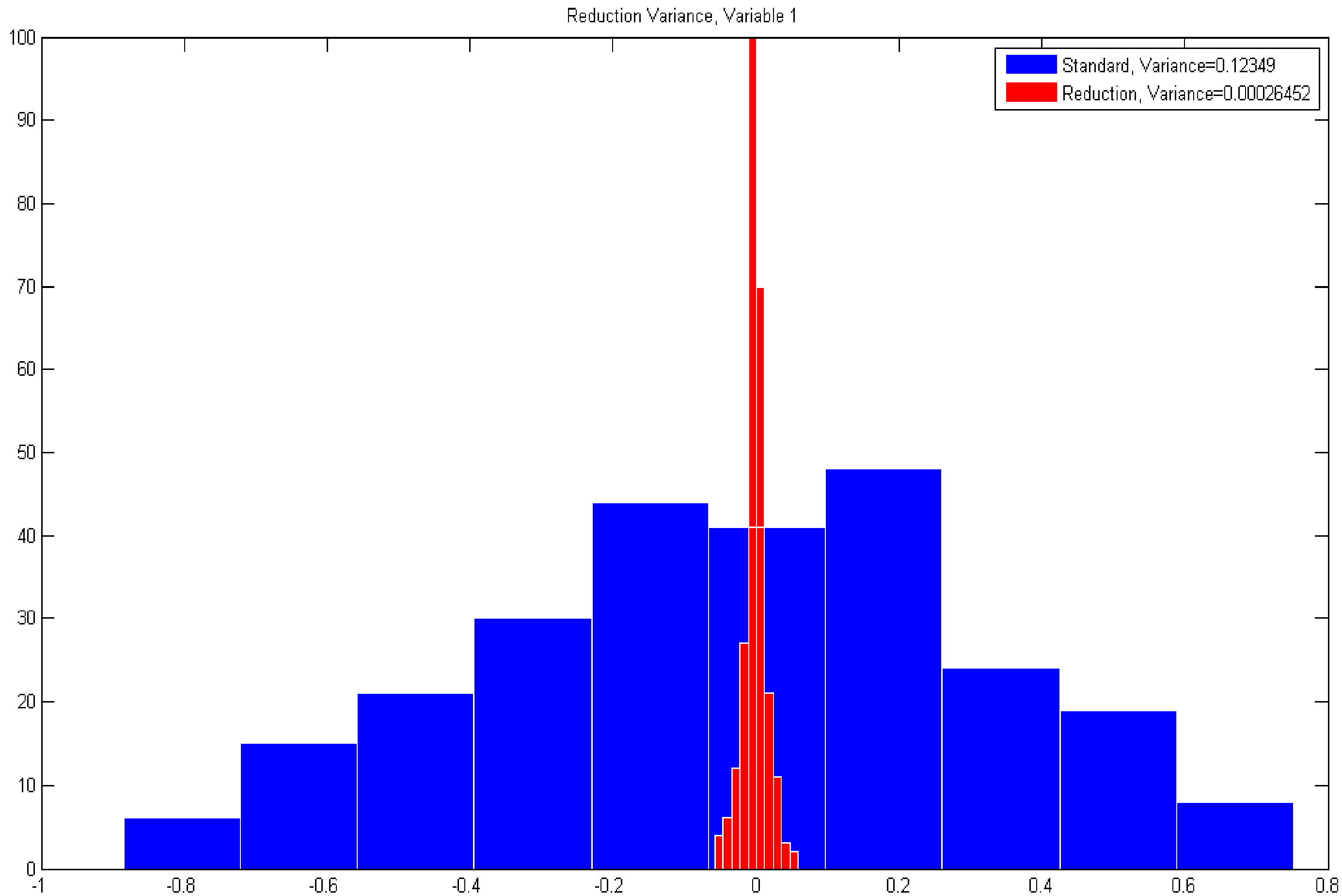
Root Mean Square Error vs. Number of sampled points

Improved formula for Sobol' Total Sensitivity Indices, Test: Modified G-function

$$f = \prod_{i=1}^{n} g_i(x_i),$$

$$g_i(x_i) = \frac{|4x_i - 2| + 2 + 3a_i}{1 + a_i},$$

$$a_i = \{19,\ 9,\ 4\}$$



Comparison of Histograms:

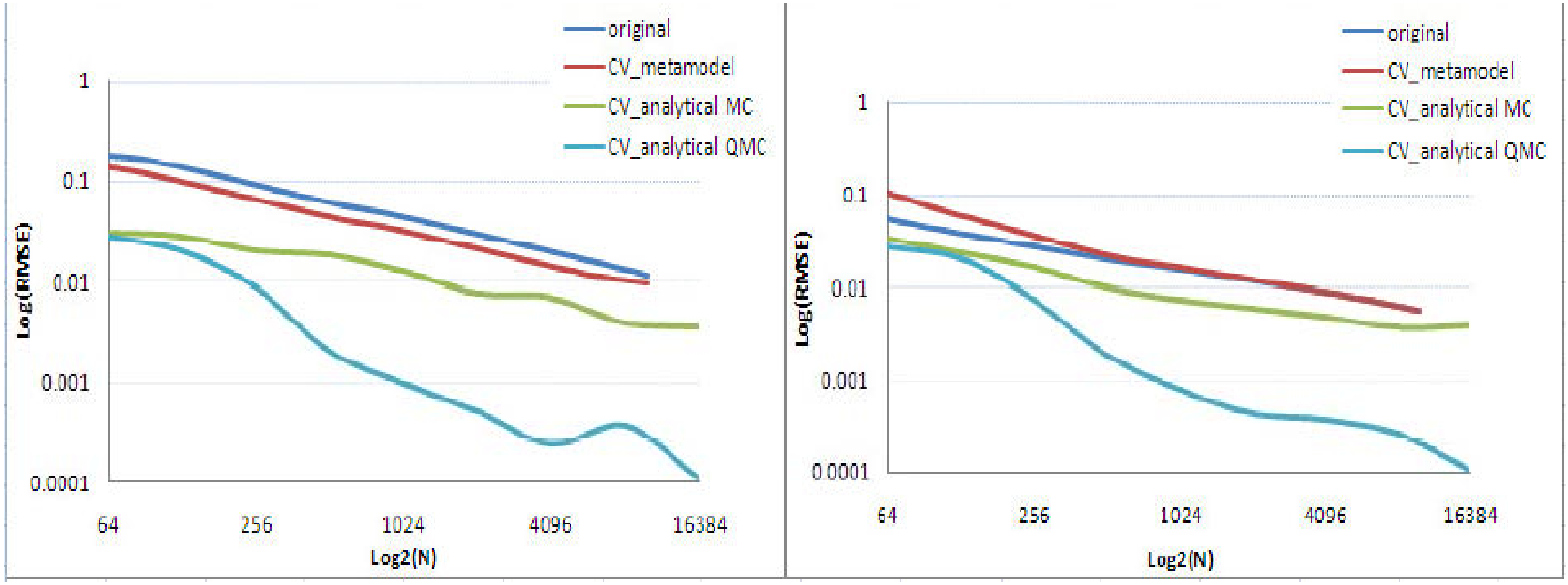
$$[f(x) - f(x')]$$

$$[f(x) - f_j(x_j) - [f(x') - f_j(x_j')]]$$

Dramatic variance reduction !

Improved formula for Sobol' Total Sensitivity Indices, Test: Ishigami function

$$f(x_1,x_2,x_3)=\sin(x_1)+7\sin^2(x_2)+0.1x_3^4\sin(x_1), -\pi\le x_i\le\pi,\ i=1,2,3$$



Variable 1 Variable 2

Root Mean Square Error vs. N.

original Sobol-Jansen formula (dark blue line);

improved formulas: based on metamodel (red line), analytical results for $f_j(x_j)$ and S_j using MC (green line), analytical results for $f_j(x_j)$ and S_j using QMC (light blue line).

Thank you for your attention !

Acknowledgments

Financial support: EPSRC Grant

Publications:

Sobol, I.M., Tarantola S., Gatelli D., Kucherenko S., Mauntz W. Estimating the Approximation Error when fixing Unessential Factors in Global Sensitivity Analysis, *Reliability Engineering & System Safety*, 92(7): 957-960, 2007.

Kucherenko S., Feil B., Shah N., Mauntz W. The identification of model effective dimensions using global sensitivity analysis *Reliability Engineering & System Safety* 96 (2011) 440–449

Sobol' indices for models with dependent variables:

S. Kucherenko, S. Tarantola, P. Annoni. Estimation of global sensitivity indices for models with dependent variables, *Computer Physics Communications, V. 183 (2012) 937–946*