



HAL
open science

Model-Based Process Optimization in the Presence of Parameter Uncertainty

Antti Solonen, Heikki Haario

► **To cite this version:**

Antti Solonen, Heikki Haario. Model-Based Process Optimization in the Presence of Parameter Uncertainty. Engineering Optimization, 2011, pp.1. 10.1080/0305215X.2011.617817 . hal-00762320

HAL Id: hal-00762320

<https://hal.science/hal-00762320>

Submitted on 7 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Model-Based Process Optimization in the Presence of Parameter Uncertainty

Journal:	<i>Engineering Optimization</i>
Manuscript ID:	GENO-2011-0084.R3
Manuscript Type:	Original Article
Date Submitted by the Author:	29-Jul-2011
Complete List of Authors:	Solonen, Antti; Lappeenranta University of Technology Haario, Heikki; Lappeenranta University of Technology
Keywords:	Process Optimization, Bayesian Estimation, MCMC, Stochastic Optimization
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
geno_mcmc_R3.tex	

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Engineering Optimization

Vol. 00, No. 00, October 2011, 1–24

Model-Based Process Optimization in the Presence of Parameter Uncertainty

Antti Solonen^{a,b*} and Heikki Haario^a

^a*Lappeenranta University of Technology, Lappeenranta, Finland*

^b*Finnish Meteorological Institute, Helsinki, Finland*

(Received 00 Month 20xx; final version received 00 Month 20xx)

In model-based process optimization one uses a mathematical model to optimize a certain criterion, for example the product yield of a chemical process. Models often contain parameters that have to be estimated from data. Typically, a point estimate (e.g. the least squares estimate) is used to fix the model for the optimization stage. However, parameter estimates are uncertain due to incomplete and noisy data. In this paper, it is shown how parameter uncertainty can be taken into account in process optimization. To quantify the uncertainty, Markov Chain Monte Carlo (MCMC) sampling, an emerging standard approach in Bayesian estimation, is used. In the Bayesian approach, the solution to the parameter estimation problem is given as a distribution, and the optimization criteria are functions of that distribution. The formulation and implementation of the optimization is studied, and numerical examples are used to show that parameter uncertainty can have a large effect in optimization results.

Keywords: Process Optimization; Bayesian Estimation; MCMC; Stochastic Optimization

*Corresponding author. Email: solonen@lut.fi

1. Introduction

In mathematical modeling, one often has unknown parameters in the model that need to be estimated from measured data. After fitting the model to the measurements, the model can be used to study the phenomenon of interest. One common use for the fitted model is process optimization, where the goal is to find operating conditions that optimize a certain criterion, e.g. maximize the yield of a desired product in a chemical reaction system. The problem in the optimization is that parameter estimates are often uncertain, since they are estimated from incomplete and noisy measurements. Traditionally, one chooses a specific point estimate for the parameter, e.g., the least squares estimate, and ignores the uncertainty in the parameters. In this paper, incorporating parameter uncertainty into process optimization is studied. To quantify the parameter uncertainty, Bayesian parameter estimation and MCMC sampling is used (see section 2 for a short introduction and the appendix for more details).

Let us define more clearly what is meant by model-based process optimization. Nonlinear models of form $y = f(x, \theta) + \epsilon$ are considered, where x are the controllable variables and θ the unknown parameters. In parameter estimation, the task is to estimate θ from measurements y . Measurement error is denoted by ϵ . In process optimization, one wishes to optimize some criterion $c(x, \theta)$ with respect to the control variables x . For example, in chemical reaction engineering, x could be the temperature of a reaction, θ might represent the reaction rate constants and $c(x, \theta)$ could be the yield of a product. A common way to optimize processes is to use response surface methods, where a sequence of experiments is used to create empirical regression models, that are used to find the optimal process, see e.g. (Myers et al. 2009). In many cases, however, a mechanistic model $f(x, \theta)$ is available, that enables more comprehensive design optimization, often with much smaller amount of experimental work. This approach is called *model-based process optimization*.

In process optimization literature, a few approaches for handling parameter uncertainty have been proposed. In these approaches – see e.g. (Ma and Braatz 2003, Rooney and Biegler 2003) and the review paper of Lee and Chen (2009) – a fixed parametric form for the parameter uncertainty (e.g. Gaussian) is used. This is in accordance with the standard nonlinear parameter estimation procedures in engineering: usually a Gaussian approximation of parameter uncertainty is used to calculate confidence intervals for parameter estimates, based on linearization around a point estimate. The approach presented in this paper provides a fully nonlinear extension to the existing methods, that utilizes the output from MCMC parameter estimation.

Recently, Bayesian model fitting methods have opened a way for proper statistical analysis of parameter estimation for nonlinear models. For examples of Bayesian parameter estimation and MCMC for model fitting in engineering applications (chemical reaction engineering) see e.g. (Vahteristo et al. 2008, Kuosa et al. 2009). In the Bayesian approach, one can avoid linear approximations in parameter uncertainty analysis. In this paper, it is shown how these results can further be used as input for process optimization. The approach for solving the resulting stochastic optimization problem is based on the same random sampling methods as those used in parameter estimation.

The paper is organized as follows. Section 2 begins by shortly recalling how Bayesian parameter estimation with MCMC works. Then, the process optimization approach together with some general purpose optimization criteria are presented that can be used when the parameters are given as a distribution instead of a point estimate. Section 3 contains remarks and discussion and in Section 4 two synthetic examples of the approach are given. Section 5 concludes the paper. Since MCMC is not a standard tool in engi-

neering, a brief introduction and a simple parameter estimation example is given in the appendix.

2. Parameter Uncertainty in Process Optimization

When unknown parameters are estimated from data, it is important to obtain information about the uncertainty of the estimates. Uncertainties can be quantified using statistical methods. Classical statistical analysis, that gives the optimal parameter values and error estimates for them, is approximate (based on linearization of the model) and may sometimes be quite misleading. Moreover, the question of the reliability of the model predictions is left open, i.e., how is the uncertainty in model parameters reflected to the model response. Both of these problems may be properly treated by Markov chain Monte Carlo (MCMC) methods. In MCMC, the estimation of model parameters and predictions are performed according to the Bayesian paradigm. All uncertainties in the data as well as the modeling results are treated as random variables that have statistical distributions. Instead of a single fit to the data, 'all' parameterizations of the model that, statistically, fit the data 'equally well' are determined. A distribution of the unknown parameters is generated using available prior information (e.g. previous studies) and statistical knowledge of the observation noise. Computationally, the distribution can be generated using the Markov chain Monte Carlo (MCMC) sampling approach.

MCMC gives the solution to the parameter estimation problem as a set of samples from the distribution of the parameters, instead of a point estimate. See the appendix for a brief introduction to MCMC and for a demonstration of the differences between MCMC and classical parameter estimation. In the rest of the paper, it is assumed that MCMC parameter estimation can be done, and samples from the parameter distribution are available. In this Section, it is shown how the uncertainty presented in the form of samples can be incorporated into model-based optimization problems.

In more detail, Bayesian estimation considers θ as a random variable that has probability density $p(\theta|\mathbf{y})$, from which one can produce samples using e.g. MCMC (see the appendix for more details). Following this interpretation, also the optimization criterion $c(\mathbf{x}, \theta)$ is a distribution that can obtain a range of possible values at any point \mathbf{x} . Thus, we are dealing with an optimization problem with a stochastic target function. There are numerous approaches for these kind of optimization problems, see e.g. (Shapiro *et al.* 2009) for a methodological introduction. In this paper, a novel approach is presented, based on the same MCMC methods that are used for the parameter estimation.

Instead of optimizing $c(\mathbf{x}, \theta)$ for a specific fixed θ , the Bayesian approach allows one to optimize a function of the distribution of $c(\mathbf{x}, \theta)$. However, it is not obvious how this statistical knowledge of the modeling uncertainty should be taken into account. Should one, for example, maximize the expected value of the cost function? Or should one employ a more conservative 'worst-case' approach where the smallest predicted product yield is maximized, taking into account the parameter distribution? Moreover, one has to consider the algorithmic details of the optimization, to avoid excessive CPU times that often plague Monte Carlo type calculations.

Below, some 'distribution-based' optimization criteria are introduced, and two ways to evaluate them are presented. First, the obvious (but computationally costly) Monte Carlo approach is discussed, where the distribution of $c(\mathbf{x}, \theta)$ is directly simulated with different parameter samples given by MCMC parameter estimation. Then, an efficient

way to compute similar criteria using MCMC integration is presented.

2.1. Direct Monte Carlo Sampling

An obvious candidate criterion for optimization is the expectation: one averages $c(\mathbf{x}, \theta)$ over θ and optimizes the mean. This reduces the risk of obtaining \mathbf{x} that produces a good value for $c(\mathbf{x}, \theta)$ only locally for a specific value for θ . The mean criterion is defined as

$$C(\mathbf{x}) = E_{p(\theta|\mathbf{y})}[c(\mathbf{x}, \theta)] = \int c(\mathbf{x}, \theta)p(\theta|\mathbf{y})d\theta. \tag{1}$$

This integral can be approximated using the existing MCMC samples from $p(\theta|\mathbf{y})$. The most obvious way is to pick a (large) number of samples $(\theta_1, \dots, \theta_N)$ from the MCMC output and use a direct Monte Carlo approximation:

$$C(\mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^N c(\mathbf{x}, \theta_i). \tag{2}$$

However, using expectation alone as an optimization criterion does not take into account the variability in the criterion. For instance, it might be useful to look for solutions that both give good criterion values on average and have small variances. A requirement of small variance can be added to the optimization criterion by e.g. penalizing large standard deviations:

$$C(\mathbf{x}) = E_{p(\theta|\mathbf{y})}[c(\mathbf{x}, \theta)] - \alpha \text{Std}_{p(\theta|\mathbf{y})}[c(\mathbf{x}, \theta)]. \tag{3}$$

Tuning parameter α defines the weight given for the variability in the criterion: decreasing α gives more weight to the expectation. Note that this criterion is close to the *mean-risk models* discussed, e.g., by Shapiro *et al.* (2009). This *robust mean* criterion in equation 3 can also be computed easily using the direct Monte Carlo approximation using empirical mean and standard deviation formulas.

As the third option, the 'worst case criterion' is mentioned, where the worst possible value of $c(\mathbf{x}, \theta)$ is maximized:

$$C(\mathbf{x}) = \min_{\theta} c(\mathbf{x}, \theta). \tag{4}$$

In this case, direct Monte Carlo approximation means calculating $c(\mathbf{x}, \theta)$ with different possible values for θ and finding the minimum from the calculated samples.

The direct Monte Carlo approximation of the above criteria is simple to implement and include in different optimization routines. However, the approach is computationally challenging, since we might need a large number of points θ_i to evaluate $C(\mathbf{x})$ just once. This can be a problem, if the optimization criterion (simulation model) is computationally expensive to evaluate. In the following, another way to approximate $C(\mathbf{x})$ using MCMC sampling and to optimize $C(\mathbf{x})$ using a simulated annealing approach are presented, following the optimal experimental design method of Müller *et al.* (2004).

2.2. MCMC Sampling and Simulated Annealing

The idea, originally presented by Müller *et al.* (2004) in a design of experiments context, is to use MCMC sampling for approximating $C(\mathbf{x})$. When the expectation is maximized (see equation 1), this can be done by running MCMC with target density

$$\pi(\mathbf{x}, \theta) = c(\mathbf{x}, \theta)p(\theta|\mathbf{y}), \quad (5)$$

assuming that $c(\mathbf{x}, \theta)$ is non-negative and bounded for all \mathbf{x} and θ . The above target density admits $C(\mathbf{x})$ as its marginal. Thus, one can use MCMC to sample from $\pi(\mathbf{x}, \theta)$ and read $C(\mathbf{x})$ from the resulting chain. Sampling from $\pi(\mathbf{x}, \theta)$ can be done using e.g. the Metropolis algorithm by proposing a new \mathbf{x} , picking a sample from $p(\theta|\mathbf{y})$ from an existing MCMC chain, evaluating $c(\mathbf{x}, \theta)$ and accepting or rejecting the proposed \mathbf{x} according to the Metropolis acceptance rule (see the appendix). The MCMC sampler is run in the joint space of \mathbf{x} and θ , where parameters θ are proposed directly from the parameter posterior distribution and controls \mathbf{x} from a separate proposal distribution as in the Metropolis algorithm.

Thus, in addition to using MCMC to produce samples from $p(\theta|\mathbf{y})$, MCMC sampling is used to explore the 'x-space' and approximate $C(\mathbf{x})$ as well. This might seem contradictory, since we are dealing with an optimization problem instead of a sampling problem. However, using MCMC can add value to optimization: $C(\mathbf{x})$ might not have a unique optimum and some variables in \mathbf{x} might not even have a significant effect on $C(\mathbf{x})$. This kind of sensitivity information can be obtained by MCMC.

The mean criterion surface $C(\mathbf{x})$ might be flat so that it is hard to distinguish the optimal \mathbf{x} from the obtained samples. As noted by Müller *et al.* (2004) in an optimal design context, a more peaked surface can be obtained by sampling from an augmented target

$$\pi(\mathbf{x}, \theta_1, \dots, \theta_J) = \prod_{i=1}^J c(\mathbf{x}, \theta_i)p(\theta_i|\mathbf{y}) \quad (6)$$

that has the marginal distribution $C(\mathbf{x})^J$. In this case, one has to pick $(\theta_1, \dots, \theta_J)$ at each MCMC iteration randomly from the existing parameter chain. Increasing J concentrates the samples more tightly around the regions of high $C(\mathbf{x})$. Increasing J along the sampling process is analogous to *simulated annealing*, which is a popular stochastic optimization method. Thus, MCMC sampling can be turned into optimization by increasing J .

The robust mean criterion in equation (3) cannot be expressed as a convenient integral like the expectation, and the MCMC sampling approach used with expectation in equation (5) cannot be directly applied. However, note that the expression $\text{Std}_{p(\theta|\mathbf{y})}[c(\mathbf{x}, \theta)]$ essentially gives weight to deviations of the cost function, $|c(\mathbf{x}, \theta) - c(\mathbf{x}, \theta')|$. Thus, a similar effect of penalizing large deviations in $c(\mathbf{x}, \theta)$ can be obtained by defining the criterion as

$$C(\mathbf{x}) = \mathbb{E}_{p(\theta|\mathbf{y})}[c(\mathbf{x}, \theta)] - \alpha \mathbb{E}_{p(\theta, \theta'|\mathbf{y})}[|c(\mathbf{x}, \theta) - c(\mathbf{x}, \theta')|] \quad (7)$$

where $p(\theta, \theta'|\mathbf{y}) \propto p(\theta|\mathbf{y})p(\theta'|\mathbf{y})$ is the joint distribution of two independent and identically distributed parameter vectors. It is straightforward to write this as an integral,

which can be approximated by MCMC sampling from target

$$\pi(\mathbf{x}, \theta, \theta') = (c(\mathbf{x}, \theta) - \alpha|c(\mathbf{x}, \theta) - c(\mathbf{x}, \theta')|) p(\theta|\mathbf{y})p(\theta'|\mathbf{y}). \tag{8}$$

Thus, at each MCMC iteration, one picks two independent random samples θ and θ' from the parameter distribution and evaluates the expression $c(\mathbf{x}, \theta) - \alpha|c(\mathbf{x}, \theta) - c(\mathbf{x}, \theta')|$. Naturally, the annealed target (see equation 6 for expectation) can be used with the robust mean criterion as well. This can be done simply by picking $(\theta_1, \dots, \theta_J)$ and $(\theta'_1, \dots, \theta'_J)$ from the parameter distribution and using MCMC to sample from the augmented target

$$\pi(\mathbf{x}, \theta_1, \dots, \theta_J, \theta'_1, \dots, \theta'_J) = \prod_{i=1}^J \pi(\mathbf{x}, \theta_i, \theta'_i). \tag{9}$$

The presented MCMC sampling and simulated annealing approaches cannot be implemented for all target function formulations. The requirement is that the criterion can be written as an integral (expectation) over the parameter distribution. For example, the worst-case criterion in equation (4) cannot be written in such a form, and one has to use the direct Monte Carlo approximation.

3. Remarks and Discussion

- Lee and Chen (2009) conduct a comparative study of different methods to quantify uncertainties in different performance measures (model outputs) caused by uncertain model inputs (parameters in our case). They consider direct numerical integration methods, based on quadrature formulas, and more sophisticated methods like the polynomial chaos expansion (PCE) for approximating the distribution of the performance measures. In the comparisons, they use certain fixed forms for the parameter distributions. The starting point of this study is different: we estimate the true, analytically intractable parameter distribution by MCMC from measured data, and study how the output of the (increasingly popular) MCMC estimation can be used as input in optimization tasks. For process optimization after MCMC parameter estimation, Monte Carlo type of approaches are natural, since the parameter distribution is only available as samples, not in closed form. The posterior distribution often does not follow any known parametric form, and approximating the posterior with a certain distribution might give in biased parameter estimation and optimization results.
- Another source of uncertainty in process optimization is the variation in 'external' operating conditions $\tilde{\mathbf{x}}$ that one cannot control (related to e.g. weather). It would be desirable to control the system so that it works well for a range of possible values for $\tilde{\mathbf{x}}$. While this paper concentrates on handling parameter uncertainty, note that it might be useful consider also $\tilde{\mathbf{x}}$ as a random variable, with some given density $p(\tilde{\mathbf{x}})$. The uncertainty in $\tilde{\mathbf{x}}$ can be taken into account by integrating over $\tilde{\mathbf{x}}$. For example, for the mean criterion one can write

$$C(\mathbf{x}) = \int \int c(\mathbf{x}, \theta, \tilde{\mathbf{x}})p(\theta|\mathbf{y})p(\tilde{\mathbf{x}})d\theta d\tilde{\mathbf{x}}.$$

The integration over $\tilde{\mathbf{x}}$ can be done using direct Monte Carlo integration or MCMC

1
2
3
4
5 integration, in a similar way as with model parameters θ . In MCMC integration,
6 the uncertainty in $\tilde{\mathbf{x}}$ can be incorporated simply by picking $\tilde{\mathbf{x}}$ at each MCMC step
7 from the known distribution $p(\tilde{\mathbf{x}})$. The distribution $p(\tilde{\mathbf{x}})$ can be of some parametric
8 form or it can be the empirical distribution (density estimate) obtained from existing
9 measurements.

- 10
11
- 12 • When using direct Monte Carlo approximation in evaluating the criteria, one has
13 to decide which samples θ_i are used for evaluating $C(\mathbf{x})$. In practice, one could
14 always pick the same (large) subset of the MCMC chain for the evaluation of $C(\mathbf{x})$,
15 which would result in a deterministic optimization problem that could be solved
16 with many nonlinear optimization methods. However, the results calculated in
17 this way would be dependent on the chosen parameter subset. Another option is
18 to pick the parameter points randomly from the MCMC chain whenever $C(\mathbf{x})$ is
19 evaluated. This leads to a stochastic optimization problem: $C(\mathbf{x})$ evaluated twice
20 with the same \mathbf{x} gives two different values. In this case, one has to use an opti-
21 mization method that can handle noise in the target function. The amount of noise
22 can be controlled by the number of samples chosen for the Monte Carlo approximation.
23
 - 24 • In this paper, the parameter α that defines the weight given for the variance
25 in $c(\mathbf{x}, \theta)$ in the optimization is chosen by hand. A useful approach might be to
26 consider the problem as a multi-objective optimization problem with two objec-
27 tives: large expectation and small deviations. The 'optimal compromises' could
28 be read from the resulting Pareto-optimal front. The challenge in applying multi-
29 objective optimization techniques is the stochastic nature of the optimization problem.
30
 - 31 • The downside of the presented approach is, as in many Monte Carlo methods, the
32 large number of model evaluations required. In the simulated annealing approach,
33 the amount of computation per step increases as J is increased. However, parallel
34 computing is directly applicable here: one can evaluate each term in the product in
35 equation (6) independently.
36
 - 37 • If parameter estimation is done in a Bayesian way, using for example MCMC, it is
38 natural to ask how the parameter estimation output can be used in other common
39 statistical analyses, such as simulation, experimental design and model-based opti-
40 mization. Simulation is straightforward: one can simply solve the model with different
41 parameter values given by MCMC estimation and study the uncertainty in model sim-
42 ulations. Recently, also optimal experimental design methods that can utilize MCMC
43 output have been proposed, see e.g. (Müller *et al.* 2004, Solonen *et al.* 2011). Thus,
44 these common statistical analyses for mathematical models can now be performed in
45 a unified Bayesian way.
46
47
48
49
50
51

52 4. Examples

53
54 In this section, process optimization using MCMC output is demonstrated with two
55 synthetic examples. Data is simulated with assumed true parameter values, based on
56 which MCMC parameter estimation is performed. Based on the resulting MCMC chain,
57 a process criterion is optimized, both by fixing θ to its MAP-estimate (least squares
58 estimate) and by taking the possible parameter values given by MCMC into account in
59
60

the proposed way. In the examples, we use the robust mean criterion.

4.1. Chemical Reaction

First, a simple chemical reaction $A \rightarrow B \rightarrow C$ is considered. Compound A transforms into B with rate $k_1(T)$ and B into C with rate $k_2(T)$. The reaction rates depend on temperature T . The model is written as an ODE system as

$$\begin{aligned} \frac{dA}{dt} &= -k_1(T)A \\ \frac{dB}{dt} &= k_1(T)A - k_2(T)B \\ \frac{dC}{dt} &= k_2(T)B. \end{aligned}$$

The temperature dependency of the reaction rates is defined using the Arrhenius equation

$$k_i(T) = a_i \exp\left(-\frac{E_i}{R} \left(\frac{1}{T} - \frac{1}{T_0}\right)\right), \tag{10}$$

where a_i are the reaction rates at the reference temperature T_0 , E_i are the activation energies of the reactions, and R is the gas constant. In parameter estimation, the goal is to estimate $\theta = (a_1, a_2, E_1, E_2)$ by measuring A and B .

In process optimization, the task is to calculate the collection time and the temperature that maximize the concentration of the intermediate product B . Thus, the control variables are $\mathbf{x} = (t, T)$ and the criterion function is defined as

$$c(\mathbf{x}, \theta) = B(t, T, \theta) \tag{11}$$

where $B(t, T, \theta)$ denotes the component B of the solution of the above ODE at time t with temperature T and parameter values θ .

4.1.1. Parameter Estimation

Synthetic data is generated, assuming that the true parameter values are $\theta_{true} = (0.005, 0.001, 4E4, 3.8E4)$. The values are chosen by hand, roughly corresponding to 'typical' Arrhenius parameter values in chemical kinetics. Gaussian noise is added to the true model response. Instead of using the model directly, noise is added using $\sqrt{y_i} = \sqrt{f_i} + \epsilon_i$. This transformation produces positive measurements, where measurement error increases as the model response increases, see e.g. (Malve *et al.* 2006). As a reference temperature, $T_0 = 323.15K$ is used. Data is simulated at two temperatures, $T = 313.15K$ and $T = 343.15K$.

The same error structure that was used in data generation is used as the likelihood in MCMC sampling. Thus, the likelihood is formulated as

$$p(\mathbf{y}|\theta) \propto \exp\left(-\frac{1}{2\sigma^2} \|\sqrt{\mathbf{y}} - \sqrt{f(\mathbf{x}, \theta)}\|^2\right). \tag{12}$$

A priori, positivity constraints are set for the parameters. In data generation, $\sigma = 0.5$ is used. For MCMC sampling, an effective adaptive MCMC method called Delayed Rejec-

tion Adaptive Metropolis (DRAM) of Haario *et al.* (2006) is employed, see the appendix for some discussion of adaptive MCMC. The DRAM algorithm is run for 10000 iterations.

Simulated data, model fit and parameter posterior $p(\theta|y)$ obtained by MCMC are plotted in figure 1. From the fit one can see that all parameters can be estimated from data with rather good accuracy. However, even this small uncertainty has an effect in the process optimization, as seen in the next section.

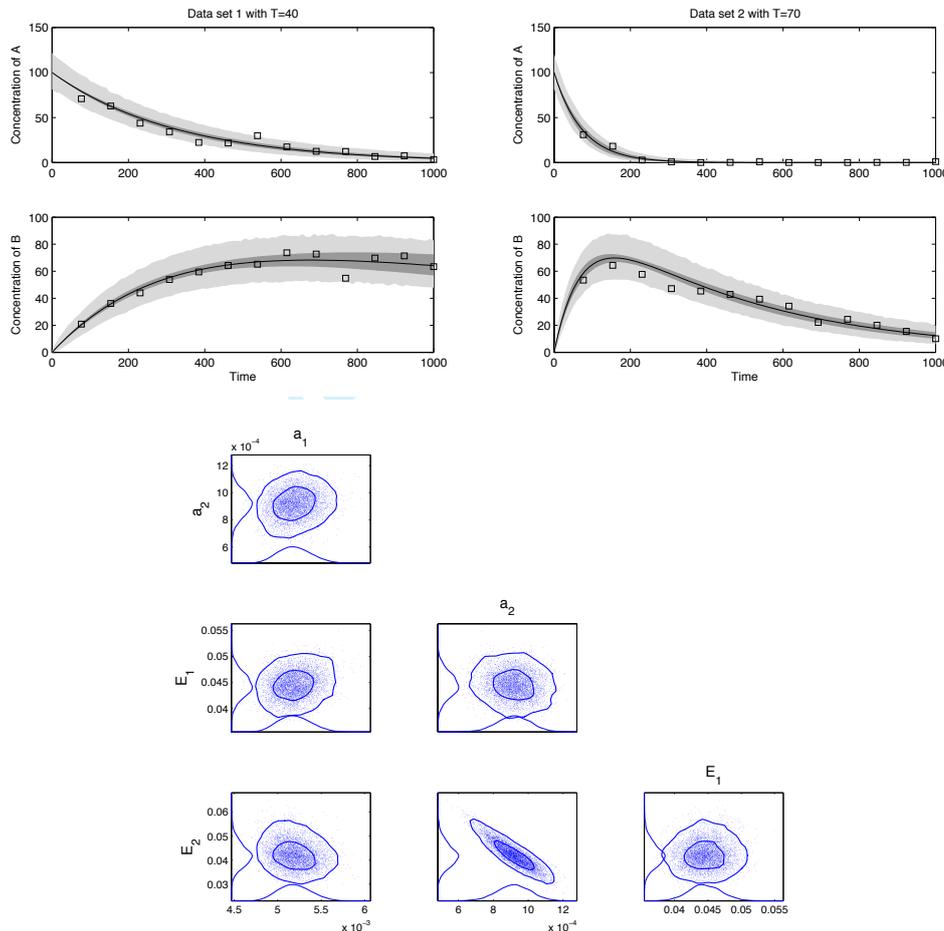


Figure 1. Top: simulated data and predictive distributions calculated from MCMC. Bottom: pairwise marginals from the parameter posterior distribution with 1d and 2d density estimates.

4.1.2. Optimal Process

Since there are only two control variables in this case and a simple model (the ODE system can be solved analytically), it is possible to demonstrate different optimization criteria by calculating direct Monte Carlo approximations on a 2D-grid of different values for the two control variables. Here, the optimal processes given by a point estimate (least squares estimate) and by the robust mean criterion in equation 8 are compared. The term \mathbf{x}_α^* denotes the optimal process condition given by the robust mean criterion with weight α given for the deviations, and $\mathbf{x}_{\text{map}}^*$ denotes the optimal condition given by using only the MAP estimate for θ .

For evaluating the robust mean criterion (equation 3), 1000 samples from the parameter distribution are chosen. The different optimization criteria in a 100x100 grid that has evenly based points in intervals $t \in [5, 100]$ and $T \in [60, 160]$ are evaluated. In figure 2, the robust mean criterion with different values for α and the criterion value with the MAP estimate are compared. The results differ significantly for varying values for α . The optimum with the MAP estimate is at $\mathbf{x}_{\text{map}}^* = (10.8, 143.8)$. The robust mean criterion gives $\mathbf{x}_0^* = (16.5, 128.7)$, $\mathbf{x}_{0.1}^* = (24.2, 117.6)$, $\mathbf{x}_{0.5}^* = (98.1, 81.2)$ and $\mathbf{x}_1^* = (100, 80.2)$. With larger values for α , the optimum does not change much from \mathbf{x}_1^* .

Using the MAP estimate only in the optimization suggests that in order to get maximal product yield, the temperature should be set high and the product should be collected soon after the process starts. Optimization with the robust mean criterion tells another story: it is better to put the temperature to a lower value and to wait for a bit longer before collecting the product. This difference is caused by the fact that, based on the parameter estimation, we do not know accurately how the system behaves, especially at high temperatures where we have no measurements. To illustrate this, the distributions of model responses with $\mathbf{x}_{\text{map}}^*$ and \mathbf{x}_1^* are compared in figure 4. The distribution of the product yields with different values for α are given in figure 3.

Note that taking the parameters into account as a distribution can yield different solutions for the process optimization problem, even if the parameter uncertainty is relatively small (as in this toy example). In real situations, one often cannot estimate parameters as well as in this example. The more uncertainty there is in parameter estimates, the more important it is to consider the optimized quantity as a distribution instead of a point estimate.

4.1.3. Comparison to the Classical Approach

From the MCMC output, one can study the sensitivity of the process optimum found with different methods, see e.g. figure 3, but the analysis does not tell how one method compares to another in real life. In synthetic cases, however, the 'true' behavior of the model is known, and one can compare how the MCMC approach performs compared to the classical approach, where a fixed point estimate is used and the parameter uncertainty is neglected.

Here, such an experiment is performed as follows. First, synthetic data is generated, as in the previous sections, by adding noise to the model simulated with 'true' parameter values. Then, parameters are estimated using both least squares and MCMC, and the optimization is performed using the obtained estimates. Thus, in the least squares approach the obtained point estimate is fixed at the optimization stage, and in the MCMC approach the parameter uncertainty is taken into account as in the previous section. Finally, the obtained optimization results are plugged in to the true model (model with assumed true parameters) to see how large product yields were actually obtained with the two methods. The above procedure is repeated a number of times (here 400) to get statistics for the performance comparison.

The results of the comparison are illustrated in figure 5. For the 400 test cases, the proposed approach that takes parameter uncertainty into account gives robust results, whereas using the point estimate only can lead to poor product yield. For instance, the product yield was below 60 in 33/400 cases when the point estimate was used, but only in 1/400 cases when the MCMC approach was applied. Note that it is possible that the least squares approach gives better results for individual cases, e.g. if the obtained point estimate happens to be very close to the 'true' value. In the long run, however, it pays off to take the uncertainty into account.

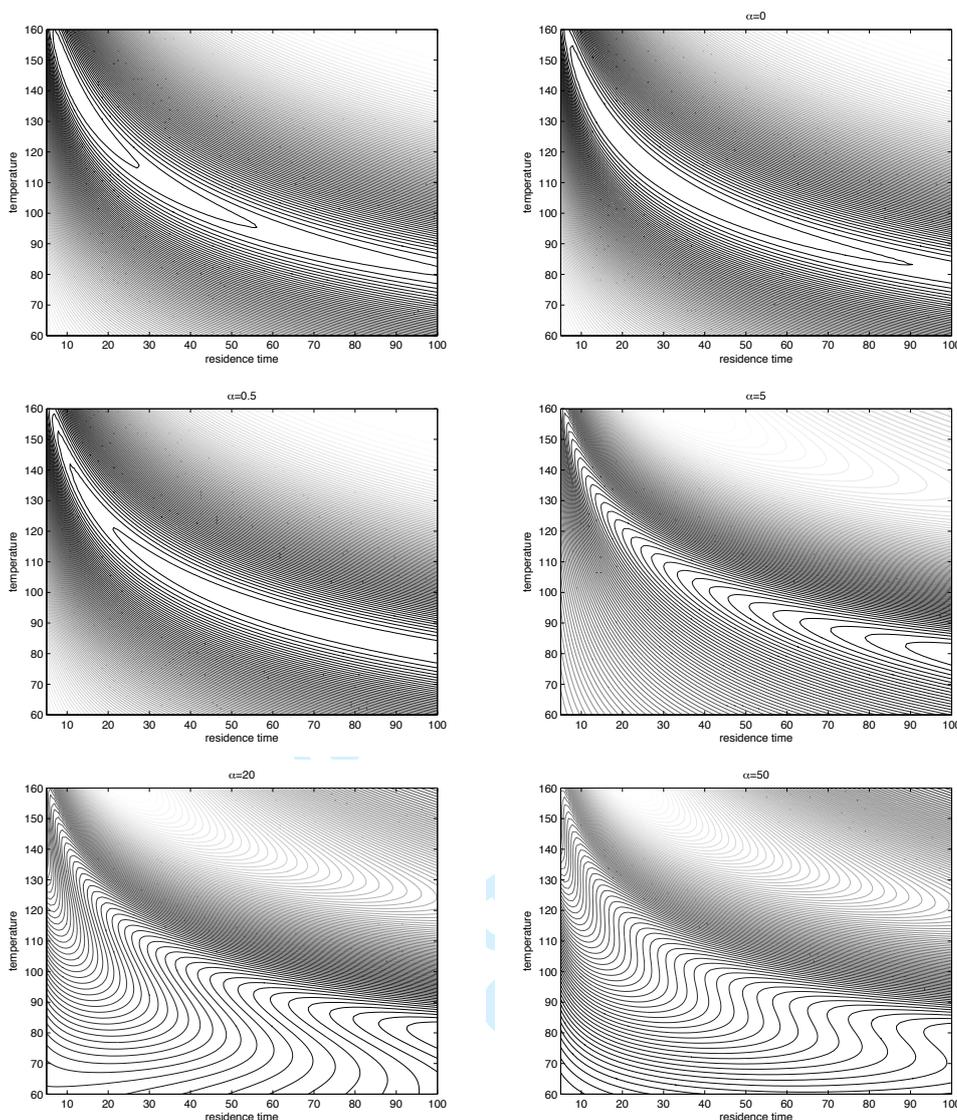


Figure 2. Criterion surfaces with the MAP estimate (top left) and with the robust mean criterion with different values for α . Temperatures are in Celcius.

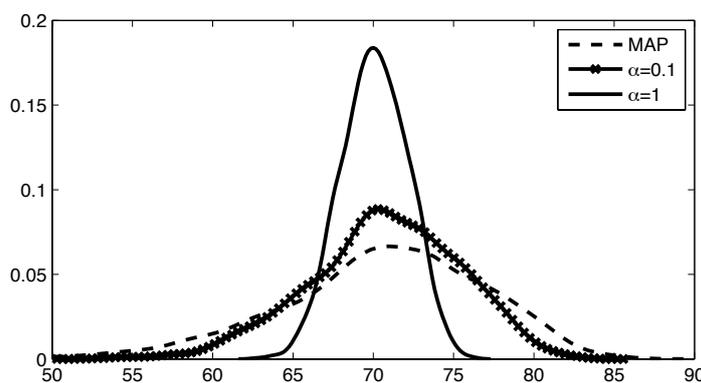


Figure 3. Distributions of product yield at MAP-optimal conditions and with the robust mean criterion with different values for α .

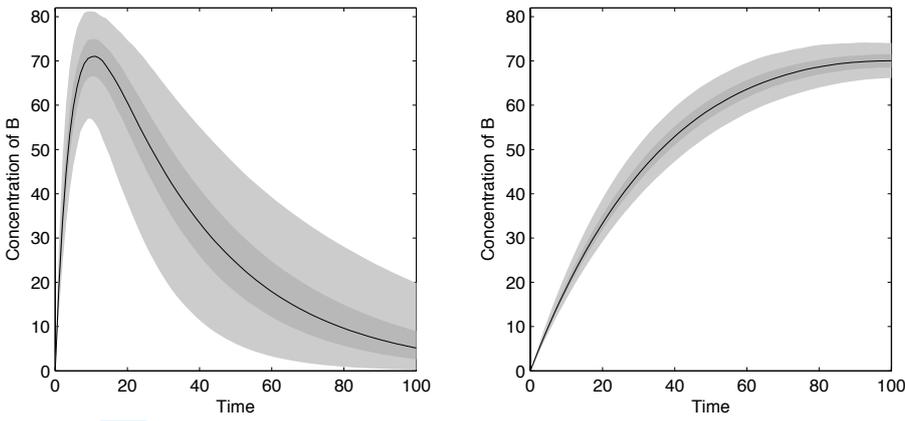


Figure 4. Distribution of model response at the MAP-optimal conditions (left) and at the optimal temperature calculated with the robust mean criterion (right). Two confidence envelopes, 50% (dark gray) and 99% (light gray) are plotted.

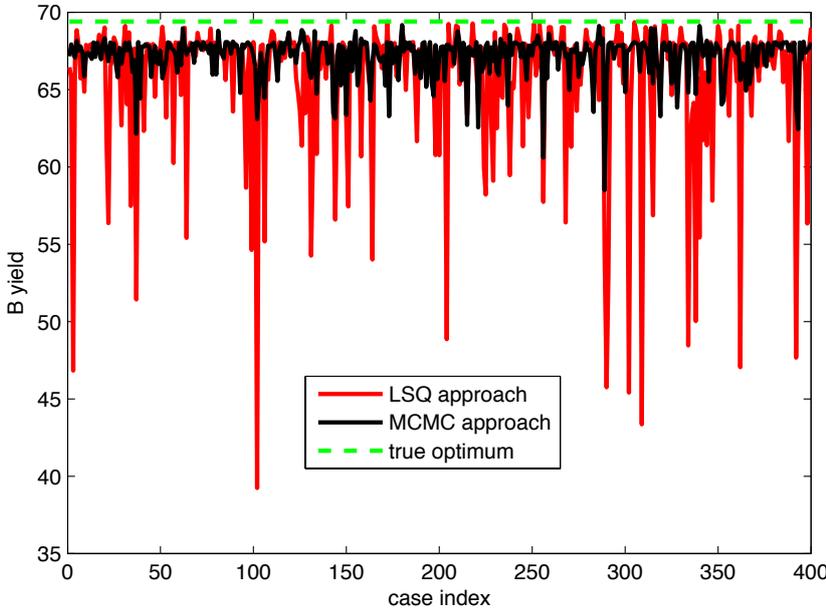
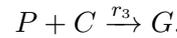
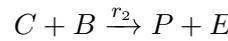


Figure 5. The optimal product yield (dashed line) and the actual product yield obtained using the proposed approach and the classical way (solid lines).

4.2. Williams-Otto Process

In this example, we consider the well-known Williams-Otto plant, introduced by Williams and Otto (1963), which is a commonly used benchmark problem in process optimization. The process consists of an ideal stirred tank reactor, in which three elementary chemical

reactions take place:



The raw materials A and B are fed into the reactor with rates F_A and F_B . The reaction mixture is directed into a cooler with rate F_R , after which the residue G is totally separated in a decanter. The remaining mixture is fed into a distillation column, where the desired product P is separated. The recovery of P is not complete and some P is left in the bottom of the distillation column. A fraction of the bottom product is recycled back to the reactor (flow F_T) and the rest is removed from the process. The process is given as a flowchart in figure 6.

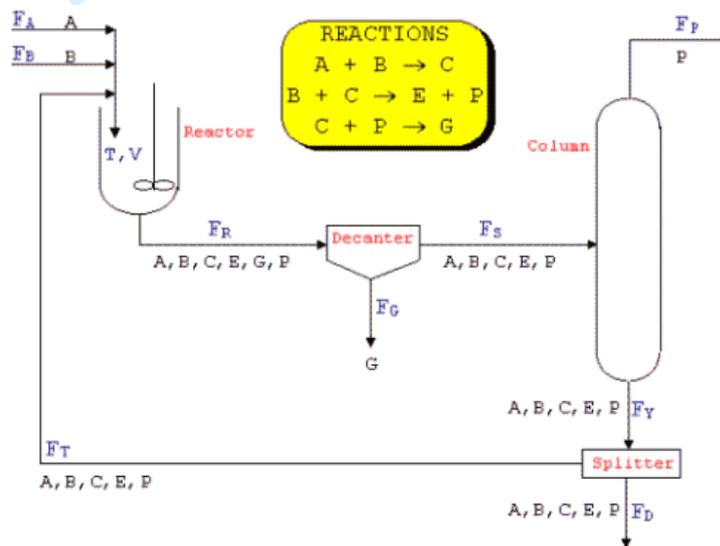


Figure 6. The Williams-Otto plant as a flowchart. Figure provided by prof. Ugur Akman.

The reaction rates are given by

$$r_1 = k_1 F_{RA} F_{RB} V \rho / F_R^2$$

$$r_2 = k_2 F_{RB} F_{RC} V \rho / F_R^2$$

$$r_3 = k_3 F_{RC} F_{RP} V \rho / F_R^2$$

where F_{RA} , F_{RB} , F_{RC} and F_{RP} are the flow rates in the reactor. Reactor volume and density of the reaction mixture are denoted by V and ρ respectively. The reaction rate coefficients follow the Arrhenius equation:

$$k_i(T) = q_i \exp\left(-b_i \left(\frac{1}{T} - \frac{1}{T_0}\right)\right) \quad (13)$$

where T is the reactor temperature and T_0 the reference temperature. The kinetic parameters $\theta = (\mathbf{q}, \mathbf{b})$ are estimated from measurements (details below).

Finally, the model can be written as a steady state model using the flow rates. Here, the model is written separately for different parts of the plant. First, for the reactor the model reads as

$$\begin{aligned}
 F_A + F_{TA} - F_{RA} - r_1 &= 0 \\
 F_B + F_{TB} - F_{RB} - r_1 - r_2 &= 0 \\
 F_{TC} - F_{RC} + 2r_1 - 2r_2 - r_3 &= 0 \\
 F_{TE} - F_{RE} + 2r_2 &= 0 \\
 F_{TG} - F_{RG} + 1.5r_3 &= 0 \\
 F_{TP} - F_{RP} + r_2 - 0.5r_3 &= 0.
 \end{aligned}$$

The equations in the decanter are trivial. In the distillation column, some product is left in the bottom of the column, since P forms an azeotrope with E . In the Williams-Otto plant, the amount of P retained in the column is taken to be 10% of the flow rate of E . Thus, one can write

$$\begin{aligned}
 F_P &= F_{SP} - 0.1F_{SE} \\
 F_{YP} &= 0.1F_{SE}.
 \end{aligned}$$

In the splitter, a fraction α of the total flow is removed from the process and the rest is recycled back to the reactor:

$$\begin{aligned}
 F_{Ti} &= (1 - \alpha)F_{Yi} \\
 F_{Di} &= \alpha F_{Yi},
 \end{aligned}$$

where $i = (A, B, C, P, E)$. For the flow rates, the model is a nonlinear system of algebraic equations and can be solved with numerical methods.

4.2.1. Parameter Estimation

To estimate the kinetic parameters $\theta = (\mathbf{q}, \mathbf{b})$, the chemical reaction is studied separately by running an experiment in a separate batch reactor with no flows. The kinetic system is now written as an ODE system:

$$\begin{aligned}
 \frac{dA}{dt} &= -k_1AB \\
 \frac{dB}{dt} &= -k_1AB - k_2CB \\
 \frac{dC}{dt} &= 2k_1AB - 2k_2CB - k_3CP \\
 \frac{dP}{dt} &= k_2CB - 0.5k_3CP \\
 \frac{dE}{dt} &= 2k_2CB \\
 \frac{dG}{dt} &= 1.5k_3PC.
 \end{aligned}$$

Assume that we can measure the concentration of components A and B at different temperatures. Synthetic measurements are generated by adding normally distributed random noise to model responses with assumed 'true' parameter values given in table 1. To get the model response, the ODE system is solved numerically, using initial values $A(0) = B(0) = 1$ and $C(0) = P(0) = E(0) = G(0) = 0$. Three batches of data are simulated in temperatures 320K, 340K and 360K, and 350K is used as the reference temperature T_0 . The reactor is run for 0.1 hours in each case and the concentrations of the compounds are collected in 20 evenly spaced times. An MCMC chain of length 10000 is created from the parameter distribution using the DRAM algorithm of Haario *et al.* (2006) and the appendix for DRAM details. Figures 7 and 8 show the distribution of the parameters and the model fit. Note that again the situation is very idealized: in real problems, the data and parameter identifiability are often worse. As shown in the next section, even this seemingly harmless parameter uncertainty has a clear impact on process optimization results.

Table 1. True parameter values for the Williams-Otto plant, obtained from (Kajaluoto 1984).

Parameter	True value
q_1	31.9223
b_1	6666.67
q_2	118.57622
b_2	8.333e+03
q_3	157.1899
b_3	1.111e+04

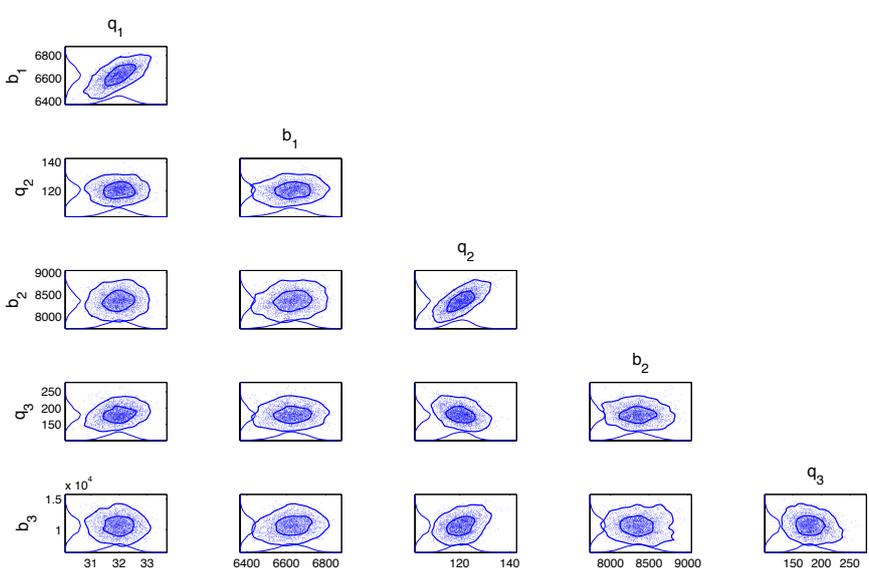


Figure 7. Posterior distribution of the kinetic parameters.

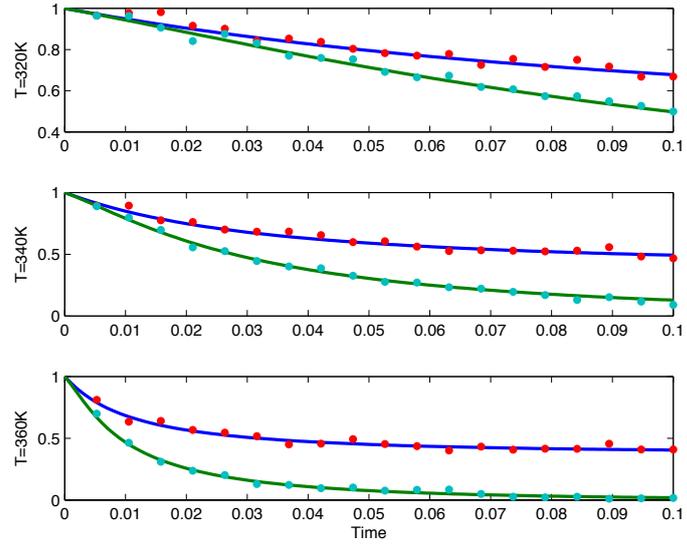


Figure 8. Simulated data and model fit for measured components *A* and *B*.

4.2.2. Process Optimization

Given a specific parameter value, the steady-state model can be used, and the system can be solved in terms of flow rates. The control variables are the input flow rates of the raw materials, the reactor temperature and the purge fraction: $\mathbf{x} = (F_A, F_B, T, \alpha)$.

As a cost function, the "return of investment", a similar cost function that was used by Kajaluoto (1984), is used. In practice, raw materials and disposal of unwanted products cost money and profit is obtained from the desired product *P*. The cost function is written as

$$c(\mathbf{x}, \theta) = (84A_{in} - 201.96\dot{m}_D - 336\dot{m}_G + 1995.52\dot{m}_P - 2.22\dot{m}_R)/(6V\rho). \quad (14)$$

The robust mean criterion and MCMC integration for process optimization (equation 8) are employed. The estimation is run with $\alpha = 0.5$, $\alpha = 1$ and $\alpha = 3$. The annealed target with $J = 40$ is used to obtain enough contrast in the surface $C(\mathbf{x})$. The robust mean criterion is given in figure 9 as an example for $\alpha = 1$.

When using the MAP estimate $\hat{\theta}$ only (traditional approach), one obtains optimal conditions $\mathbf{x}_{map}^* = [6489, 15589, 379.3, 0.1002]$ that gives objective function value $c(\mathbf{x}_{map}^*, \hat{\theta}) = 144.3$. The results agree well with the ones obtained by Kajaluoto (1984). The optimal conditions \mathbf{x}_α^* given by the robust mean criterion with different values for α are $\mathbf{x}_{0.5}^* = [6415, 12390, 370.5, 0.081]$, $\mathbf{x}_1^* = [6016, 13760, 367.6, 0.081]$ and $\mathbf{x}_3^* = [4395, 8034, 358.6, 0.0556]$. The distributions of the cost functions with the robust mean criterion and the MAP-optimal point are illustrated in figure 10.

It turns out that the distribution of the objective function value, given the uncertainty in the reaction kinetics, is rather wide, indicating that the objective function is sensitive to the values of the kinetic parameters (see figure 10). With the results given by the robust mean approach, the risk of obtaining poor (even negative) return of investment is minimized. However, putting more weight to small deviations comes at a cost: the expected value decreases slightly as the deviations are emphasized.

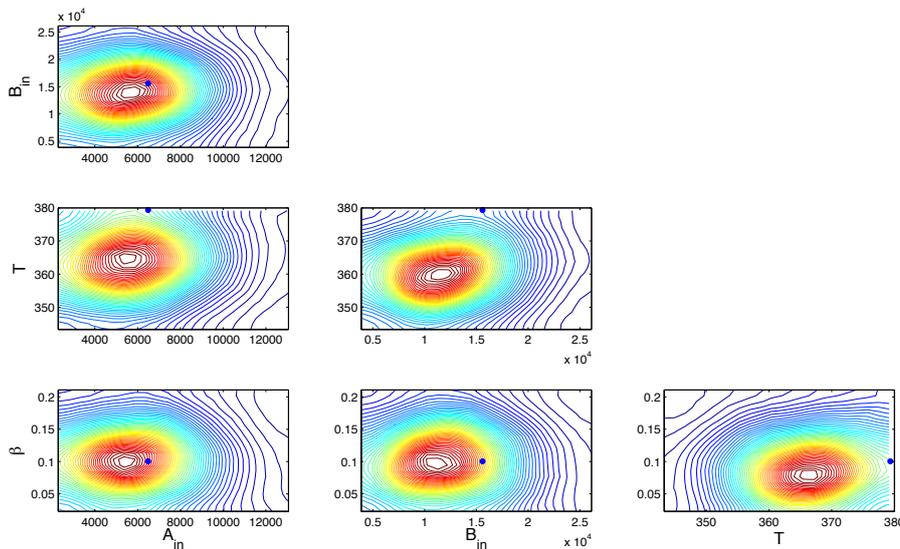


Figure 9. Robust mean criterion surface with $\alpha = 1$ and the MAP-optimal conditions (dot).

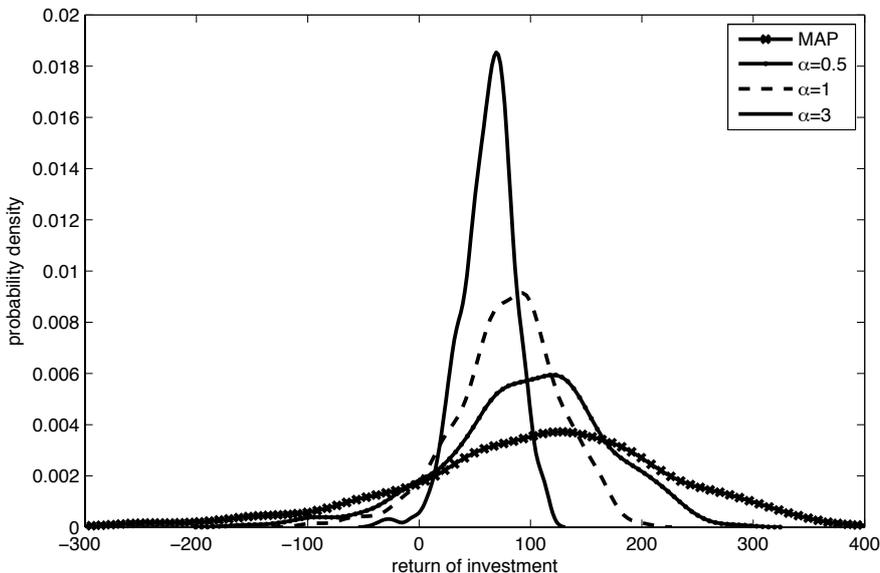


Figure 10. Distributions of product yield at MAP-optimal conditions and optimal conditions calculated with the robust mean criterion with different values for α .

5. Conclusions

In this paper, a way to incorporate parameter uncertainty into process optimization is presented. The approach utilizes the output from MCMC methods that are more and more routinely used in statistical model fitting problems. The same MCMC methods can also be used in calculating the optimization criteria effectively, instead of direct Monte Carlo approximations. The approach is tested in two simple chemistry examples. Even in

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

these simple cases, with small parameter uncertainties, the difference to the traditional approach is clearly visible. With the presented approach it is possible to reduce the risk of obtaining poor process optimization results caused by incomplete knowledge of the process.

Acknowledgements

We wish to thank Prof. Ugur Akman for providing the flowchart figure of the Williams-Otto process. Mr. Kalle Riihimäki is acknowledged for suggesting the Williams-Otto experiment.

This work was funded by the Finnish Center of Excellence in Inverse Problems Research consortium (Academy of Finland).



Appendix

In this appendix, the basic concepts and algorithms of Bayesian parameter estimation and MCMC are briefly summarized. In addition, a simple example that demonstrates the difference of the MCMC approach to classical nonlinear parameter estimation procedures is given.

MCMC basics

In the Bayesian approach the unknown parameter vector is interpreted as a random variable. The aim of the analysis is to find its distribution. Before experimental data is available the parameter θ has a *prior distribution* $p(\theta)$. The observations \mathbf{y} update the distribution $p(\theta)$ to the *posterior distribution* by the *Bayes formula*

$$\pi(\theta) = \frac{p(\mathbf{y}|\theta)p(\theta)}{\int p(\mathbf{y}|\theta)p(\theta) d\theta}. \quad (15)$$

Here $p(\mathbf{y}|\theta)$ is the *likelihood function* that gives the likelihood of data \mathbf{y} for given parameter value θ . The *posterior distribution* $\pi(\theta) = p(\theta|\mathbf{y})$ gives the probability distribution of parameter values, given the measured data \mathbf{y} . The integral $\int p(\mathbf{y}|\theta)p(\theta) d\theta$ is needed as the normalizing constant, to ensure that π indeed is a probability measure, with total measure equal to one, $\int_{\theta} \pi(\theta) d\theta = 1$.

In the usual settings the parameter vector θ and data \mathbf{y} are connected by a model $\mathbf{y} = f(\mathbf{x}, \theta) + \epsilon$, where the experimental error $\epsilon \sim N(0, \sigma^2 \mathbf{I})$, i.e., in all experiments the measurement noise is Normally distributed, independent and with standard deviation of size σ . It is not difficult to see that then

$$p(\mathbf{y}|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{i=1}^n (y_i - f(x_i, \theta))^2 / 2\sigma^2\right).$$

So, one arrives at the familiar least squares function: maximizing the likelihood function turns out to be equivalent to minimizing the residual sum of squares.

In principle, the Bayes formula solves the estimation problem in a fully probabilistic sense: one finds the peak, the maximum a posteriori (MAP) point, of the parameter distribution. Then one determines a required portion of the probability mass (typically some 95% or 99% of the mass) around it. However, the problems of how to define the a priori distribution and how to calculate the integral of the normalizing constant are faced. Often, only a 'flat' prior is used, that is, a uniform distribution that only defines physically possible lower and upper bounds for each parameter. However, the integration of the normalizing constant often is a formidable task, even for only moderately high number of parameters in a nonlinear model. So, a direct application of the Bayes formula is intractable for all but trivial nonlinear cases. The MCMC methods provide a tool to handle this problem. They generate a sequence of parameter values $\theta_1, \theta_2, \dots, \theta_N$, whose empirical distribution approximates the true posterior distribution for large enough sample size N .

The trick here is that the distribution from which to sample is not known, but it is still possible to generate samples from it. Instead of sampling from the true distribution, one samples from an artificial *proposal* distribution. Combining the sampling with a simple accept/reject procedure, the posterior can be correctly approximated. The simplest

MCMC method is the *Metropolis algorithm* introduced by Metropolis *et al.* (1953):

- (1) Initialize by choosing a starting point θ_1 .
- (2) Choose a new candidate $\hat{\theta}$ from a suitable proposal distribution $q(\cdot|\theta_n)$ that may depend on the previous point of the chain.
- (3) **Accept** the candidate with probability

$$\alpha(\theta_n, \hat{\theta}) = \min \left(1, \frac{\pi(\hat{\theta})}{\pi(\theta_n)} \right).$$

If rejected, repeat the previous point in the chain. Go back to item 2.

So, points with $\pi(\hat{\theta}) > \pi(\theta_n)$, i.e., steps 'uphill', are always accepted. But also points with $\pi(\hat{\theta}) < \pi(\theta_n)$, i.e., steps 'downhill', may be accepted, with probability that is given by the *ratio* of the π values. In practice, this is done by generating a uniformly distributed random number $u \in [0, 1]$ and accepting $\hat{\theta}$ if $u \leq \pi(\hat{\theta})/\pi(\theta_i)$. Note that only the ratios of π at consecutive points are needed, so the main difficulty is omitted: the calculation of the normalizing constant is not needed since the constant cancels out!

However, the choice of the proposal distribution may still pose a problem. It should be chosen so that the 'sizes' of the proposal q and target distributions suitably match. This may be difficult to achieve. An unsuitable proposal leads to inefficient sampling, typically due to

- the proposal being too large. Then the new candidates mostly miss the essential region of π ; they are chosen at points where $\pi \simeq 0$ and only rarely accepted.
- the proposal being too small. The new candidates mostly are accepted, but from a small neighborhood of the previous point. So, the chain moves only slowly, and may not cover the target π in finite number of steps.

For simple cases, the proposal might be relatively easy to find by some hand-tuning. However, the 'size' of the proposal distribution is not a sufficient specification. In higher dimensions, especially, the shape and orientation of the proposal are crucial. The most typical proposal is a multi-dimensional Gaussian (Normal) distribution. In the *random walk* version, the center point of the Gaussian proposal is chosen to be the current point of the chain. The task then is to find a covariance matrix that produces efficient sampling.

Several efficient adaptive methods have been recently proposed, see, e.g., the Adaptive Metropolis (AM) and the Delayed Rejection Adaptive Metropolis (DRAM) algorithms (Haario *et al.* 2001, 2006). In adaptive MCMC, one uses the sample history to automatically tune the proposal distribution 'on-line' as the sampling proceeds. In AM, the empirical covariance from the samples obtained so far is used that as the covariance of a Gaussian proposal. The simplicity of AM adaptation allows for its use in more advanced adaptation schemes. The DRAM algorithm combines the delayed rejection (DR) method by Mira (2001) with AM. This DRAM method has been shown to be efficient in many applications, see e.g. (Villagran *et al.* 2008, Smith and Marshall 2008). In DR, when a proposed candidate point in a Metropolis chain is rejected, a second stage move is proposed around the current point. For example, one can use downscaled versions of the proposals given by AM adaptation as second stage proposals in DR. This is especially helpful to get the sampler moving (to get accepted points) in the beginning of the MCMC run.

In this paper, the DRAM method was used for all sampling tasks. The DRAM calculations were performed by using the Matlab adaptive MCMC Toolbox, **that may**

be downloaded from the Internet, see (Laine 2008).

Example: MCMC vs. Classical Methods

In this section, the difference between classical nonlinear regression analysis and Bayesian model fitting is illustrated using a simple toy example. The model $\mathbf{y} = \theta_1(1 - \exp(-\theta_2\mathbf{x}))$, used to model e.g. the biological oxygen demand, is fitted to data $\mathbf{x} = (1, 3, 5, 7, 9)$, $\mathbf{y} = (0.076, 0.258, 0.369, 0.492, 0.559)$. First, the parameters are estimated using the classical least squares approach, and the covariance of the parameters is approximated using a linearization around the least squares estimate $\hat{\theta}$. In practice the covariance is estimated as $\text{Cov}(\hat{\theta}) \approx \sigma^2(\mathbf{J}^T\mathbf{J})^{-1}$, where \mathbf{J} is the Jacobian of the parameters evaluated at $\hat{\theta}$ and σ^2 is the measurement error variance (assumed here to be i.i.d. Gaussian). The measurement error variance was estimated from the residuals, giving $\sigma = 0.014$.

Then, the DRAM sampler is run for 10000 steps to get samples from the posterior distribution. In figure 11, the parameter distributions obtained with the classical linearization-based way and with the MCMC approach are presented. One can see how the classical approach can go wrong, if the parameter distribution is not Gaussian. In the figure, the predictive distribution calculated from the MCMC output is also given, illustrating that extending the uncertainty analysis to predictions (and other functions of the parameters) can be easily achieved simply by simulating the model with different parameter values given by MCMC.

Review Only

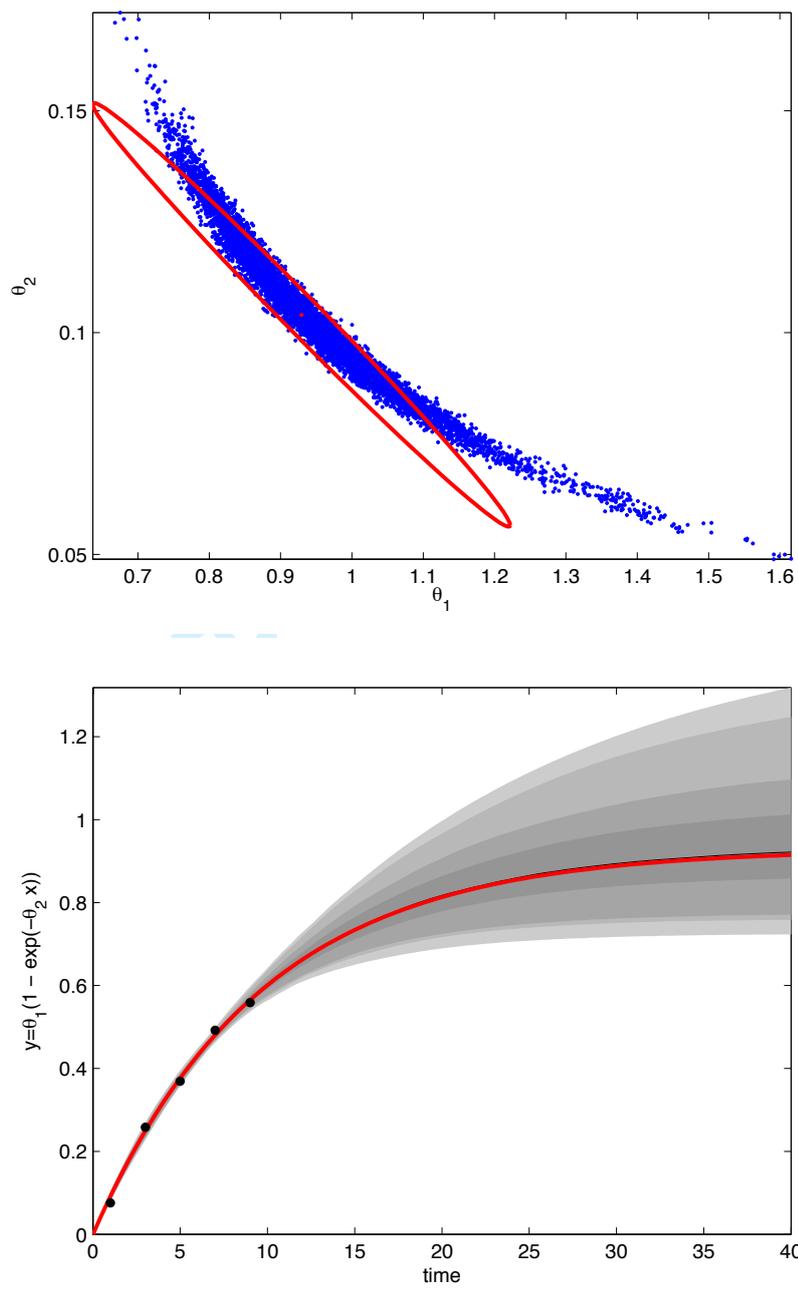


Figure 11. Top: parameter posterior with MCMC (blue) and linearization around the LSQ estimate (red). Bottom: predictive distributions with MCMC and the single prediction produced by the LSQ estimate. Gray colors correspond to 50%, 80%, 95% and 99% confidence envelopes.

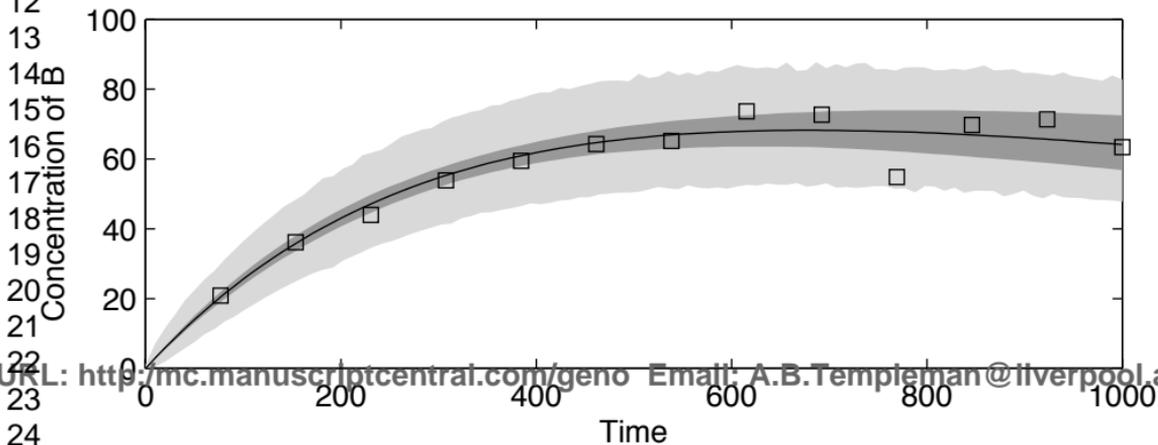
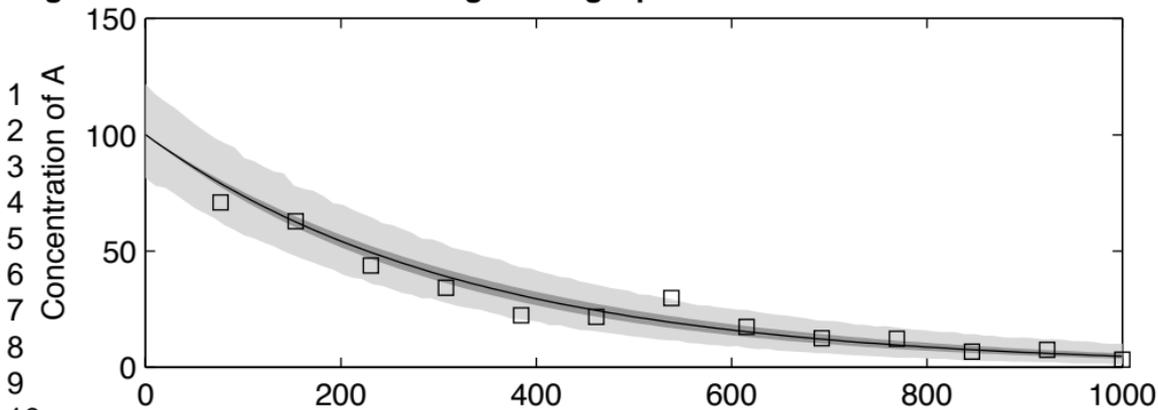
References

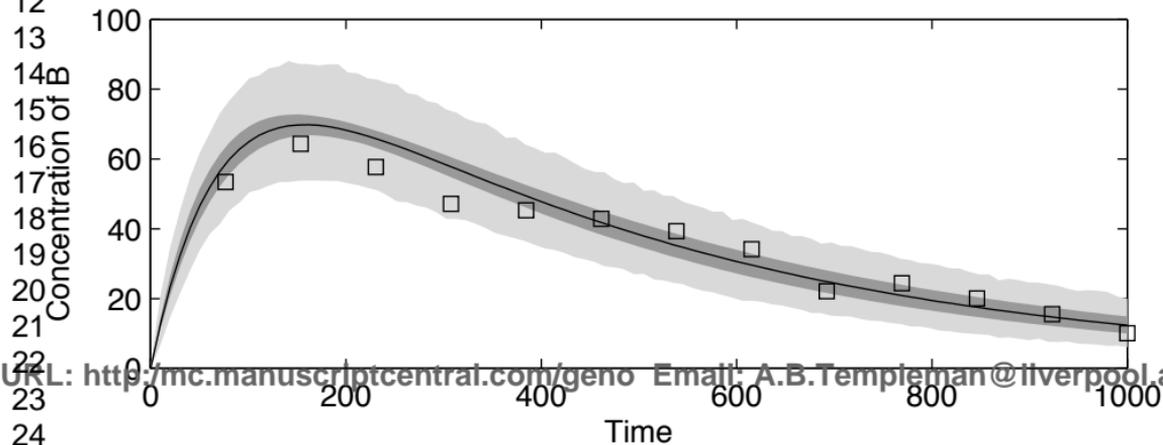
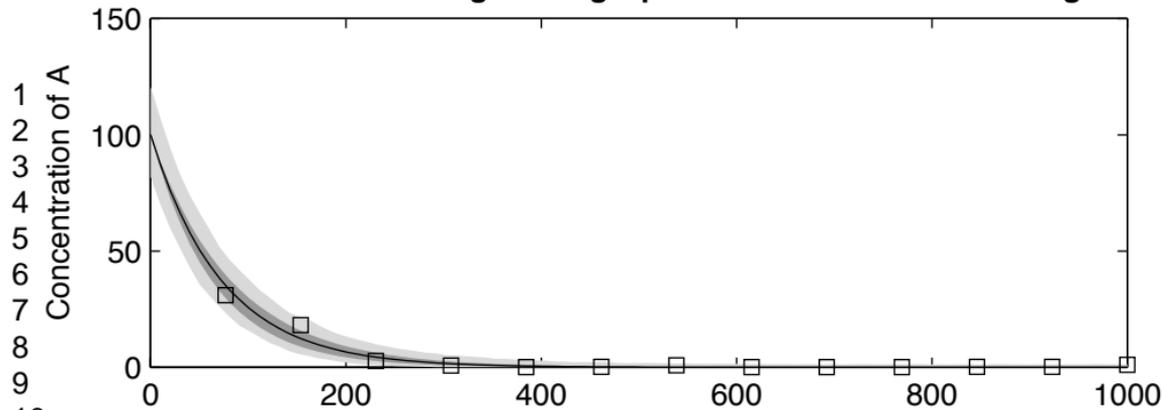
- 1
2
3
4
5
6
7 Haario H., Saksman, E. and Tamminen J., 2001. An adaptive Metropolis algorithm.
8 *Bernoulli*, 7 (2), 223–242.
- 9 Haario H., Laine M., Mira A. and Saksman E., 2006. DRAM: Efficient adaptive MCMC.
10 *Statistics and Computing*, 16 (4), 339–354.
- 11 Kaijaluoto S., 1984. *Process optimization by flowsheet simulation*. PhD thesis. Espoo,
12 Finland: Technical Research Centre of Finland.
- 13 Kuosa M., Laari A., Solonen A., Haario H. and Kallas J., 2009. Multicomponent reaction
14 kinetics for the ozonation of p-nitrophenol and its decomposition products under
15 acidic conditions at constant pH. *Chemical Engineering Science*, 64 (10), 2332–2342.
- 16 Laine M., 2008. Adaptive MCMC Methods with Applications in Environmental and
17 Geophysical Models. *Finnish Meteorological Institute Contributions*, 69. Available
18 from: <http://www.helsinki.fi/~mjllaine/mcmc/> (Accessed June 2011).
- 19 Lee S.H., Chen W., 2009. A comparative study of uncertainty propagation methods for
20 black-box-type problems. *Structural and Multidisciplinary Optimization*, 27, 239–
21 253.
- 22
23 Ma D.L. and Braatz R.D., 2003. Robust identification and control of batch processes.
24 *Computers and Chemical Engineering*, 27, 1175–1184.
- 25 Malve O., Laine M., Haario H., Kirkkala T. and Sarvala J., 2006. Bayesian modelling of
26 algae mass occurrences - using adaptive MCMC methods with a lake water quality
27 model. *Environmental Modelling & Software*, 22 (7), 966–977.
- 28 Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E., 1953.
29 Equations of State Calculations by Fast Computing Machines. *Journal of Chemical*
30 *Physics*, 21 (6), 1087–1092.
- 31 Mira A., 2001. On Metropolis-Hastings algorithms with delayed rejection. *Metron*, LIX
32 (34), 231–241.
- 33 Müller P., Sanso B. and De Iorio M., 2004. Optimal Bayesian Design by Inhomoge-
34 neous Markov Chain Simulation. *Journal of the American Statistical Association*,
35 99 (467), 788–798.
- 36 Myers R.H., Montgomery D.C., Anderson-Cook C.M., 2009. *Response Surface Method-*
37 *ology: Process and Product Optimization Using Designed Experiments*. John Wiley
38 and Sons.
- 39 Rooney W.C. and Biegler L.T., 2003. Optimal Process Design with Model Parameter
40 Uncertainty and Process Variability. *AIChE Journal*, 49 (2), 438–449.
- 41 Shapiro A., Dentcheva D. and Ruszczyński A., 2009. *Lectures on Stochastic Programming,*
42 *Modeling and Theory*. MPS-SIAM series on optimization; 9.
- 43 Smith T. J. and Marshall L. A., 2008. Bayesian methods in hydrologic modeling: A study
44 of recent advancements in Markov chain Monte Carlo techniques. *Water Resources*
45 *Research*, 44 (W00B05), 1–9.
- 46 Solonen A., Haario H. and Laine M., 2011. Simulation-Based Optimal Design Using a
47 Response Variance Criterion. To appear in *Journal of Computational and Graphical*
48 *Statistics*.
- 49 Vahteristo K., Laari A., Solonen A. and Haario H., 2008. Estimation of kinetic parameters
50 in neopentyl glycol esterification with propionic acid. *Chemical Engineering Science*,
51 63 (3) 587–598.
- 52 Villagran A., Huerta G., Jackson C. S. and Sen M. K., 2008. Computational methods
53 for parameter estimation in climate models. *Bayesian Analysis*, 3(3), 1–27.
- 54 Williams T.J. and Otto R. E., 1963. A Generalized Chemical Processing Model for the
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

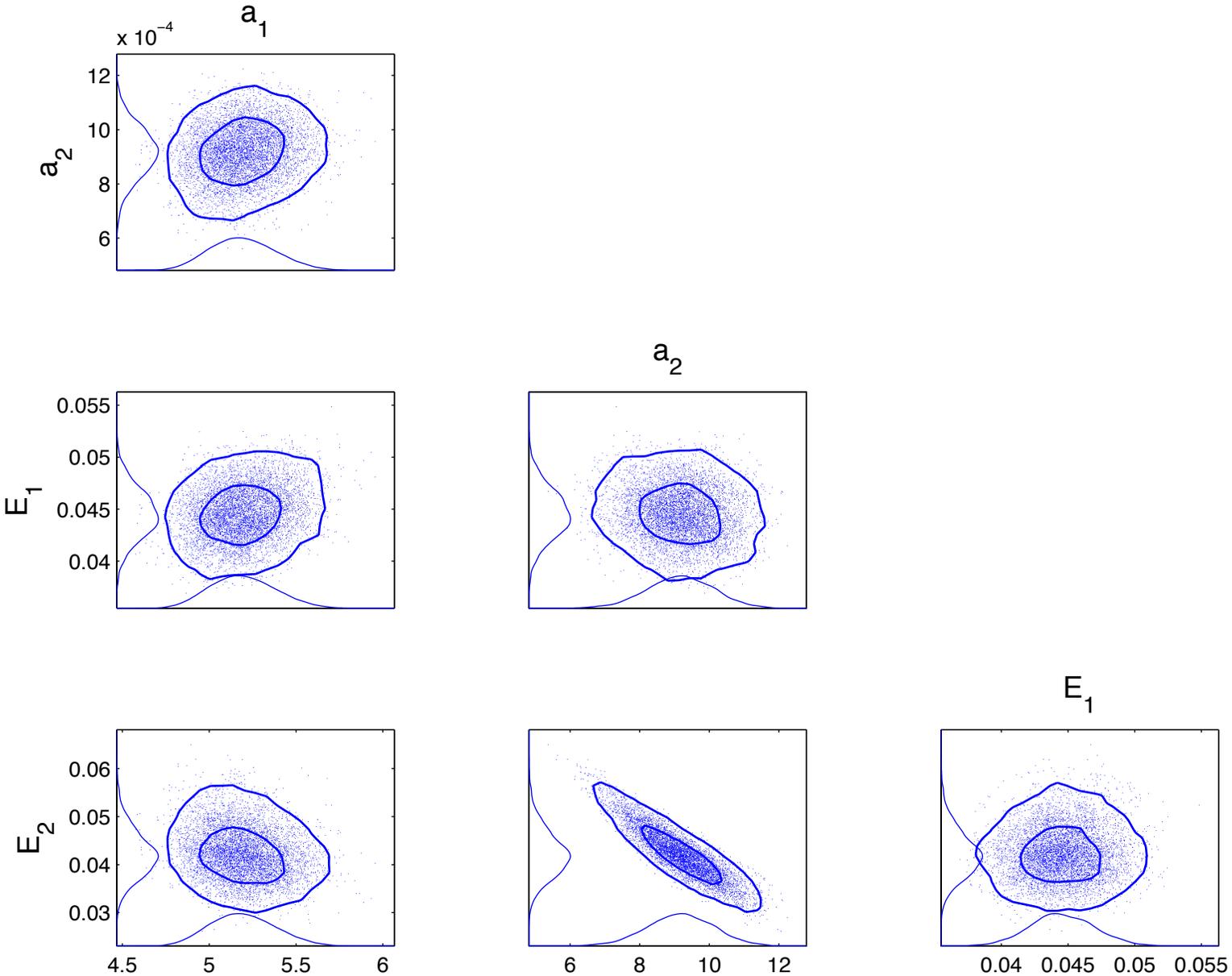
Investigation of Computer Control. *AIEE Transactions*, 79, 458.

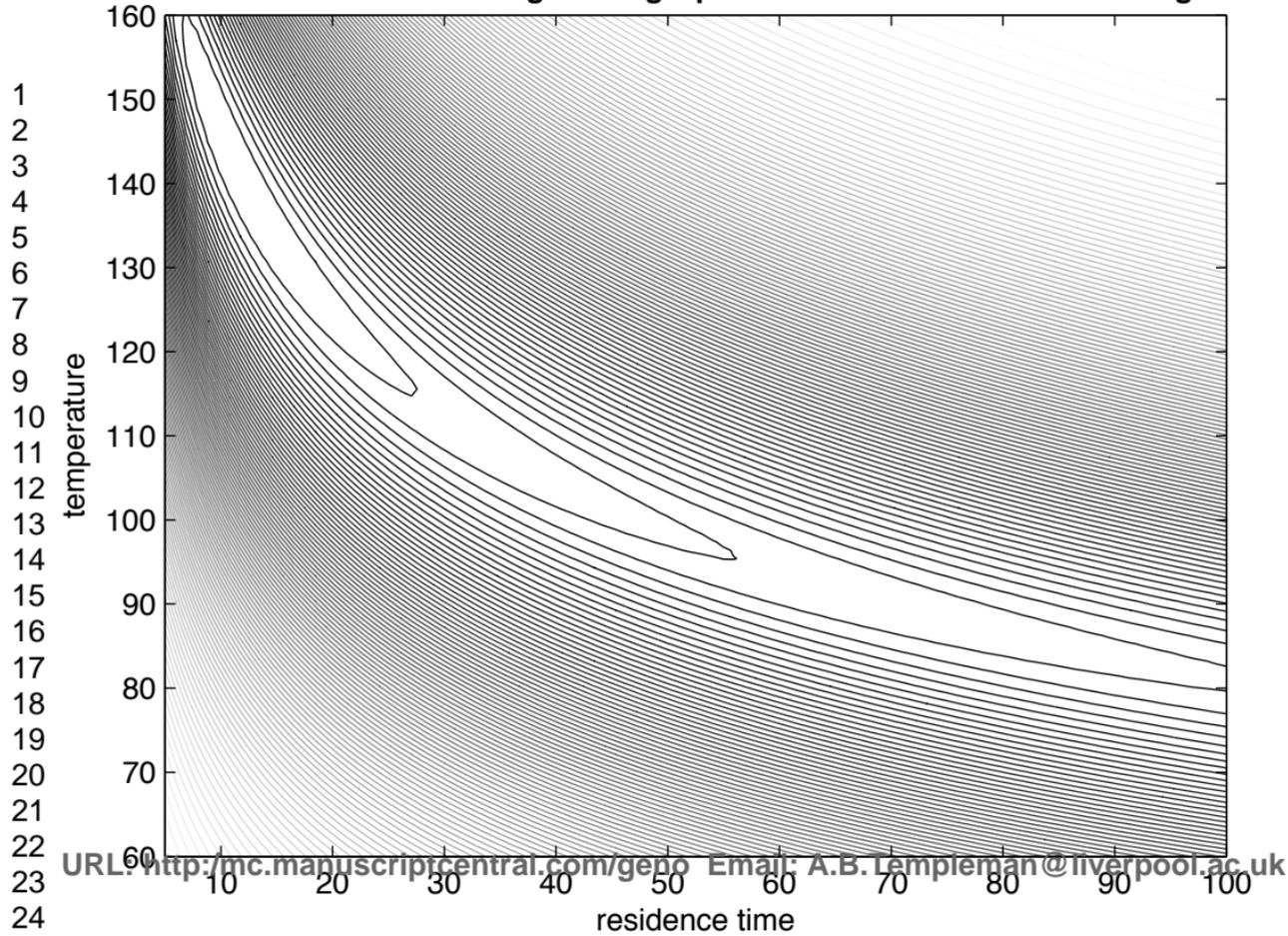
For Peer Review Only



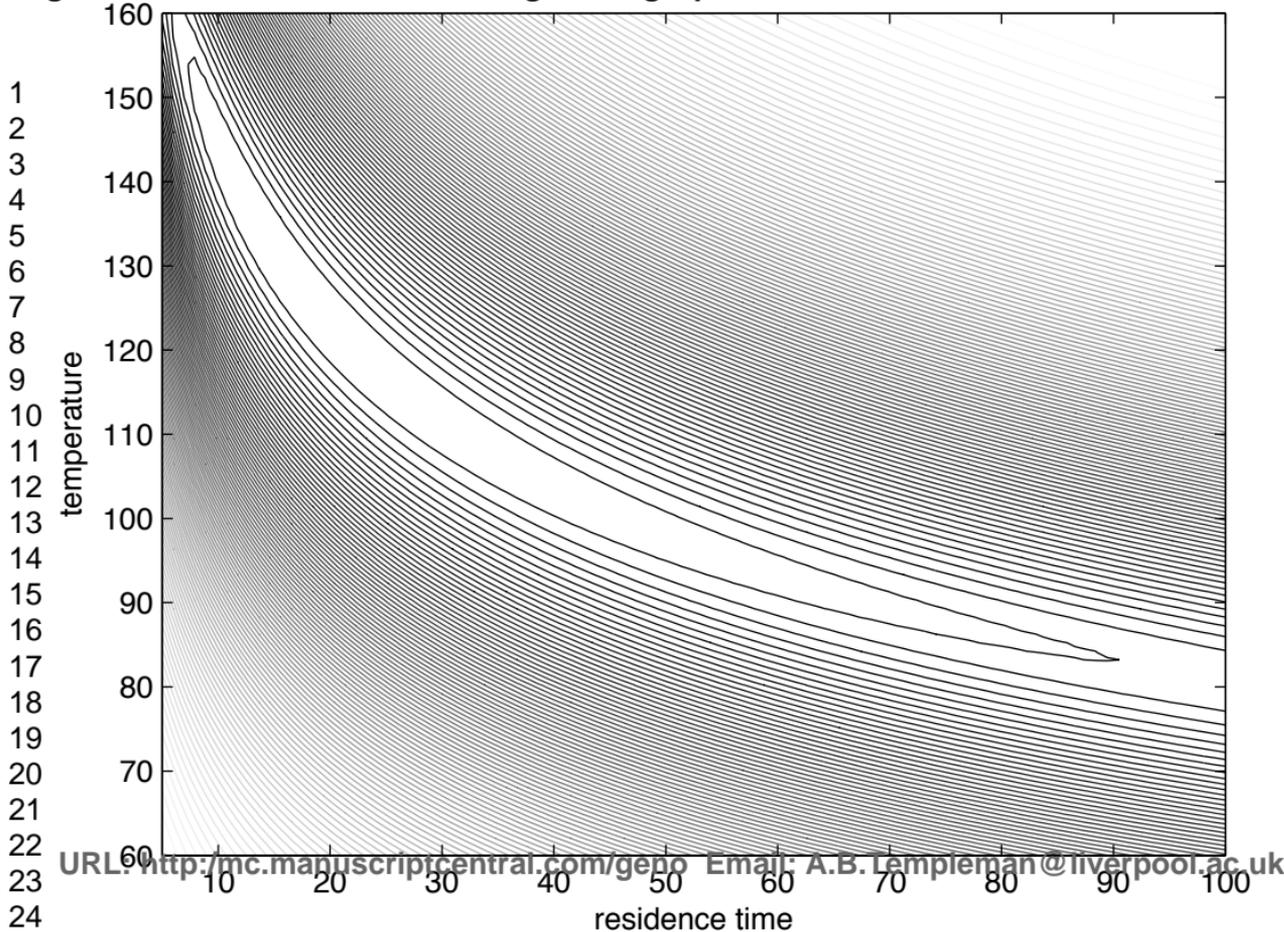


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

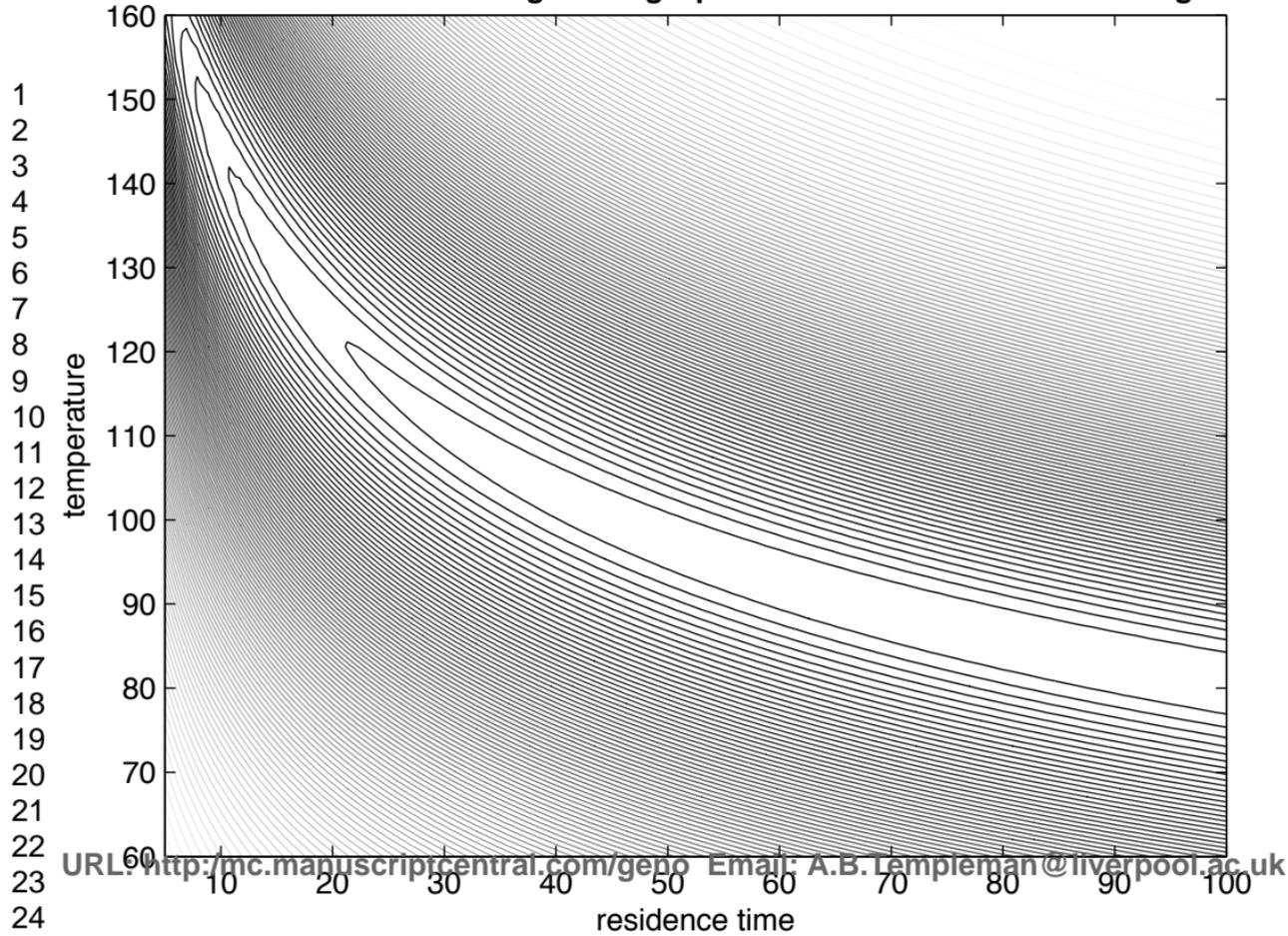




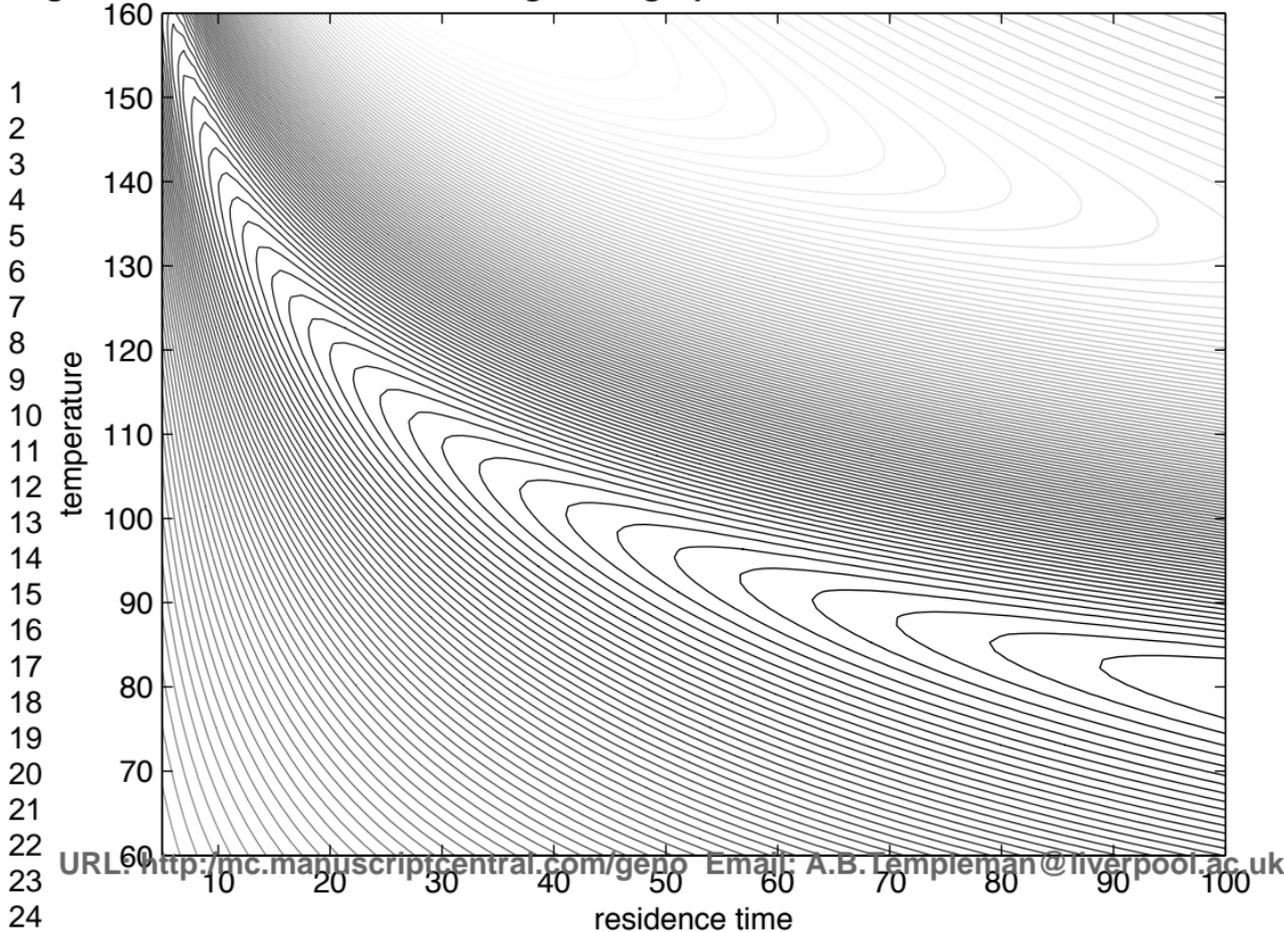
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

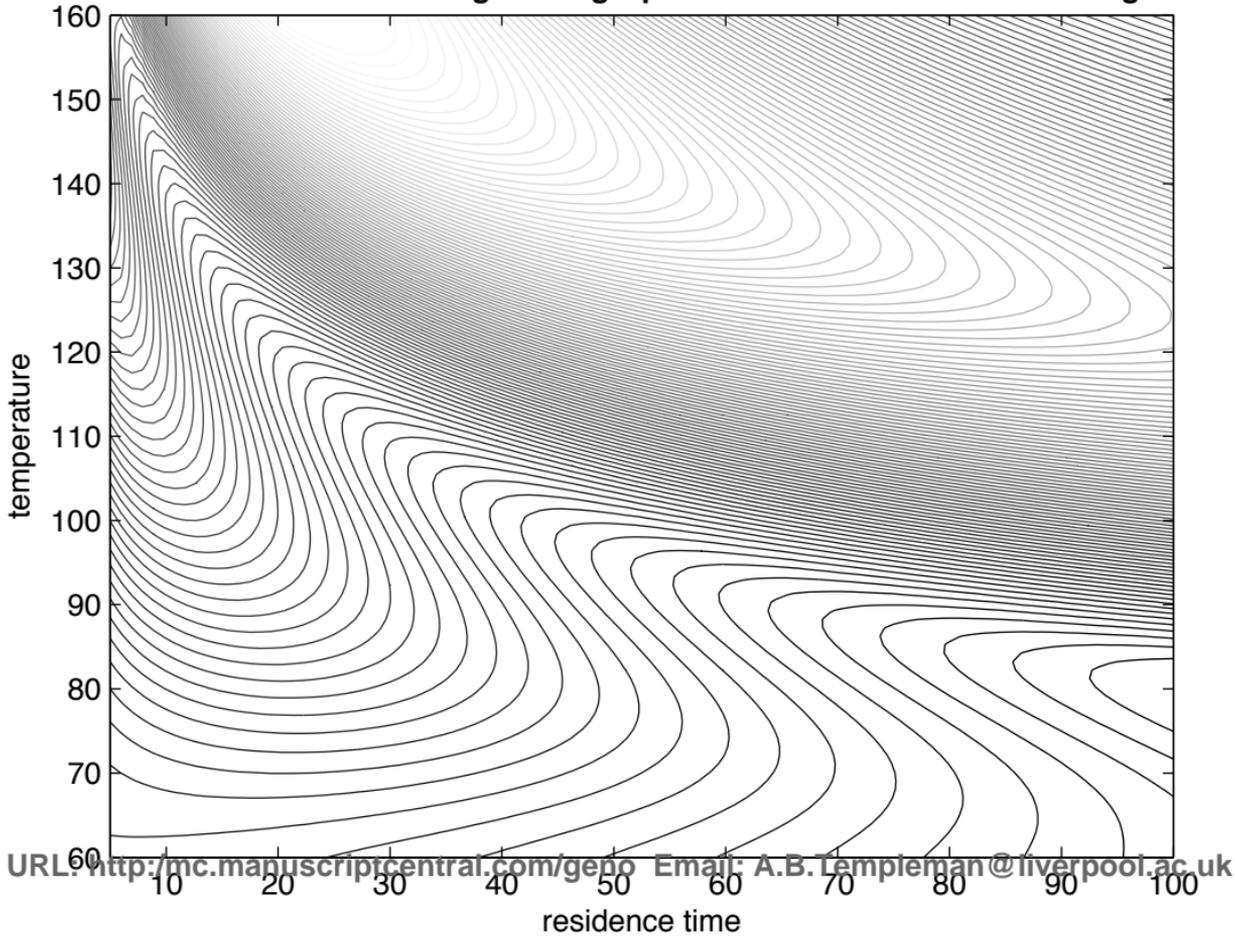


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

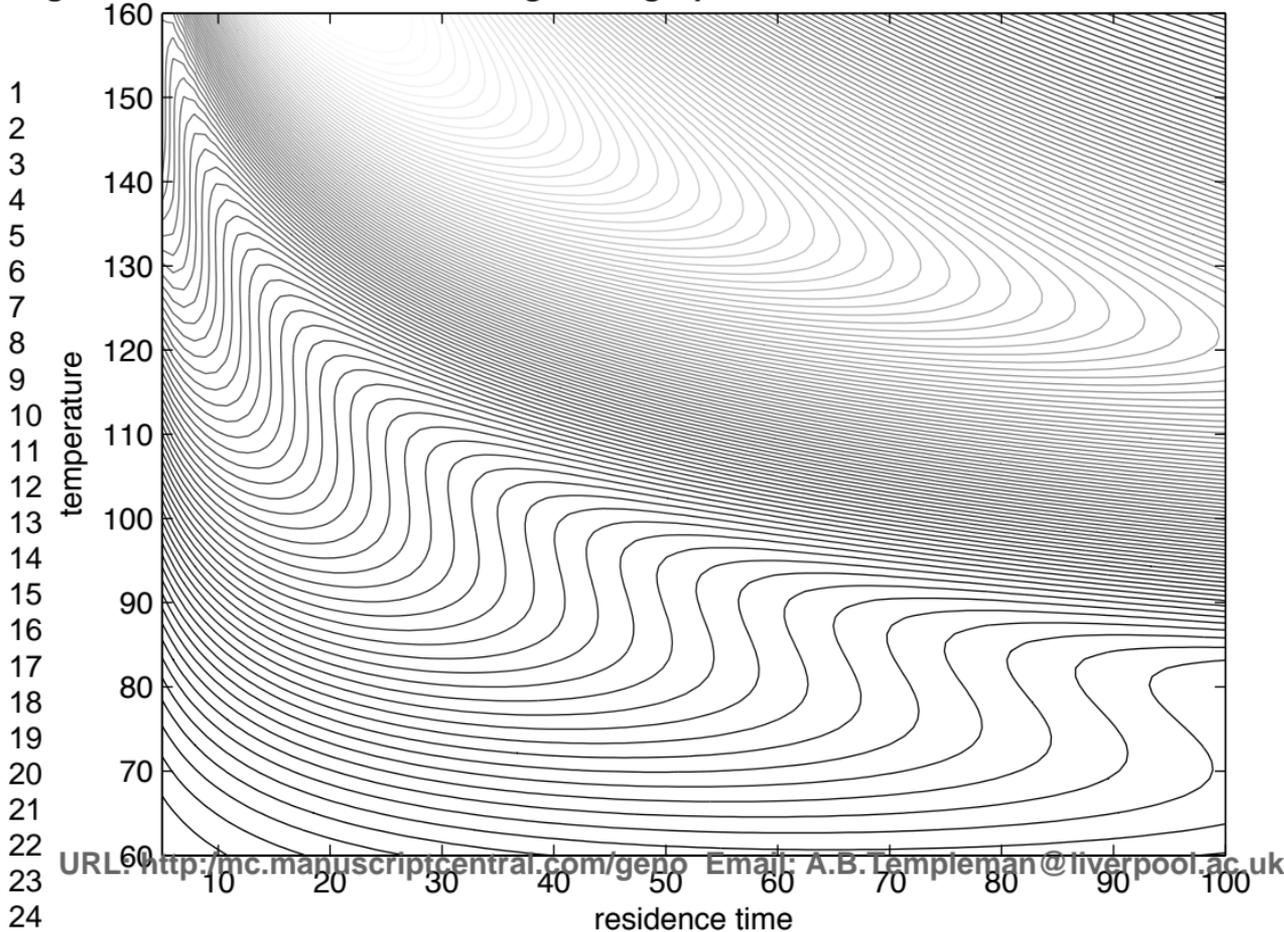


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

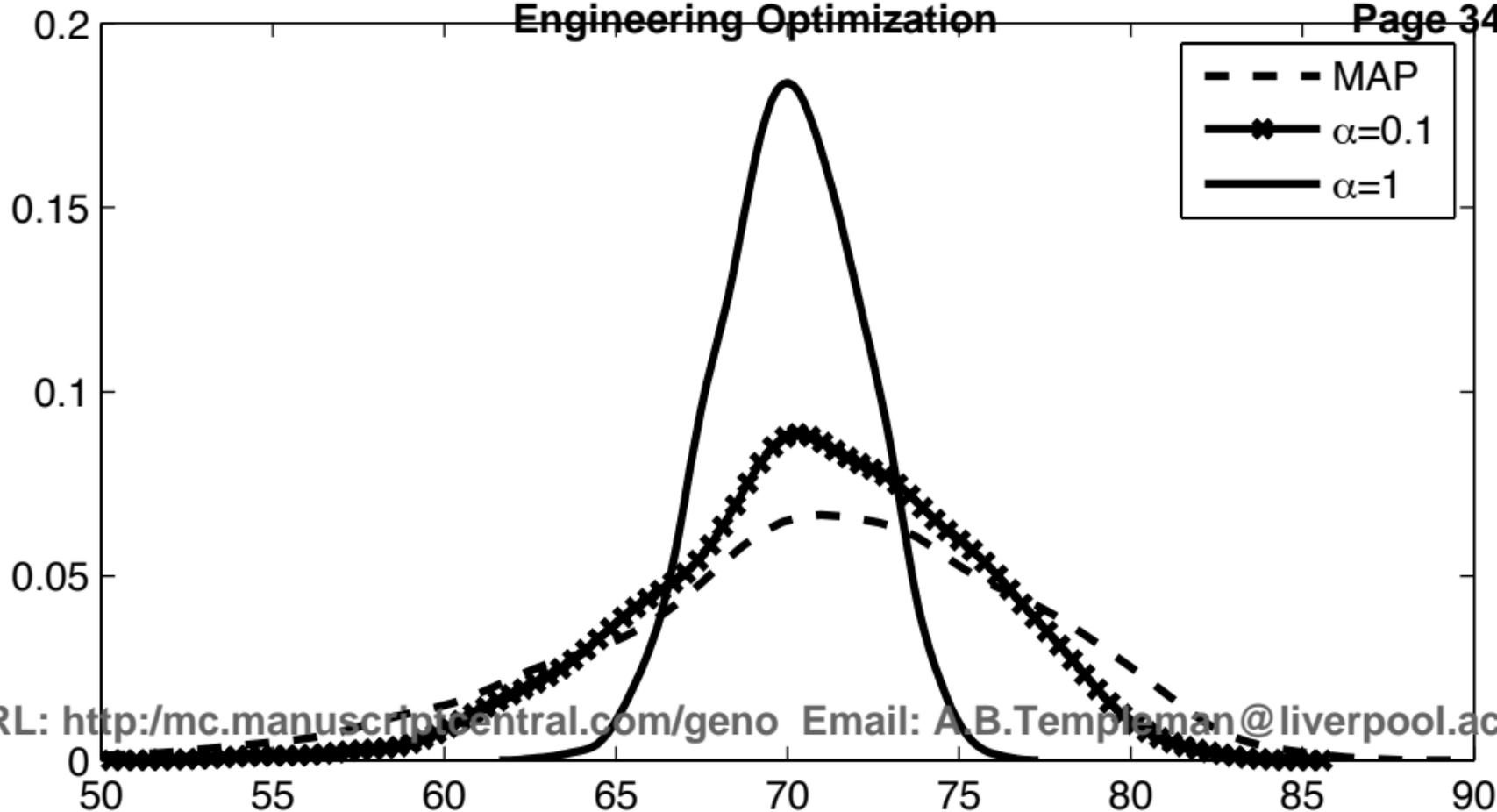




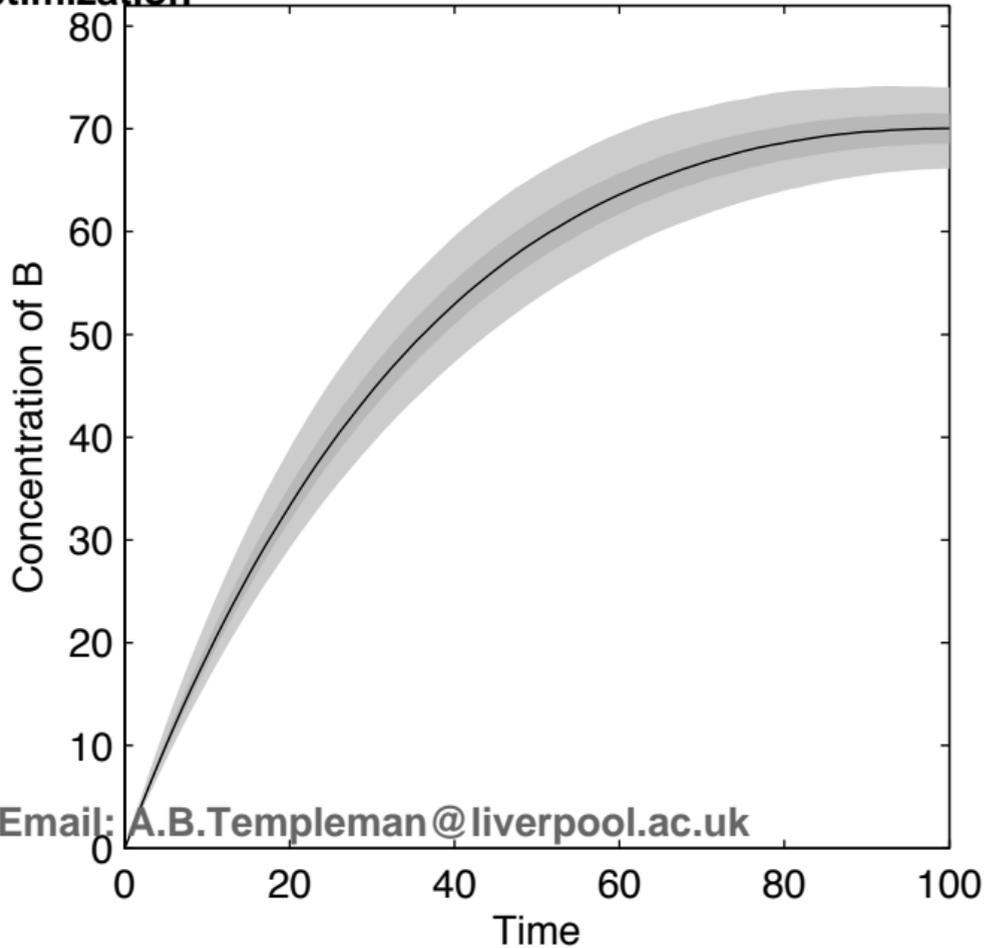
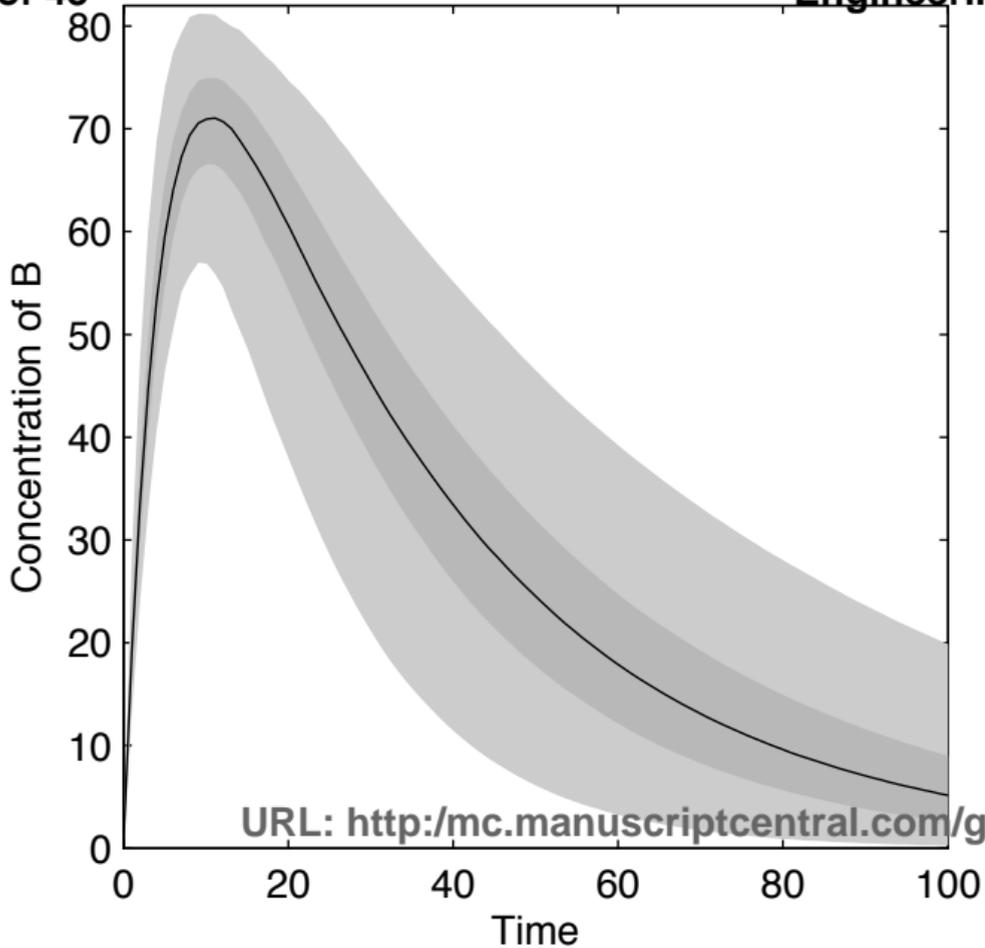
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25



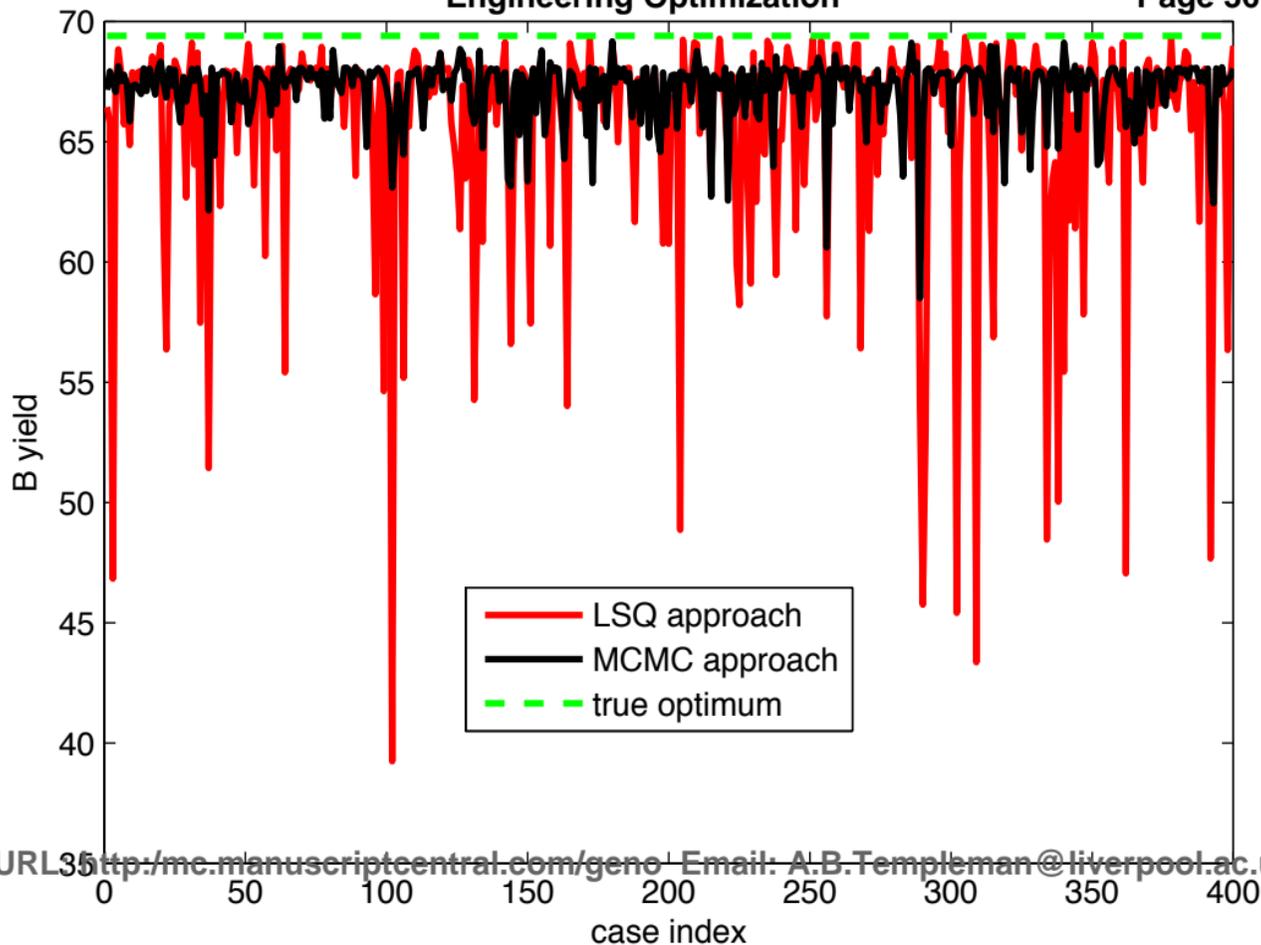
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

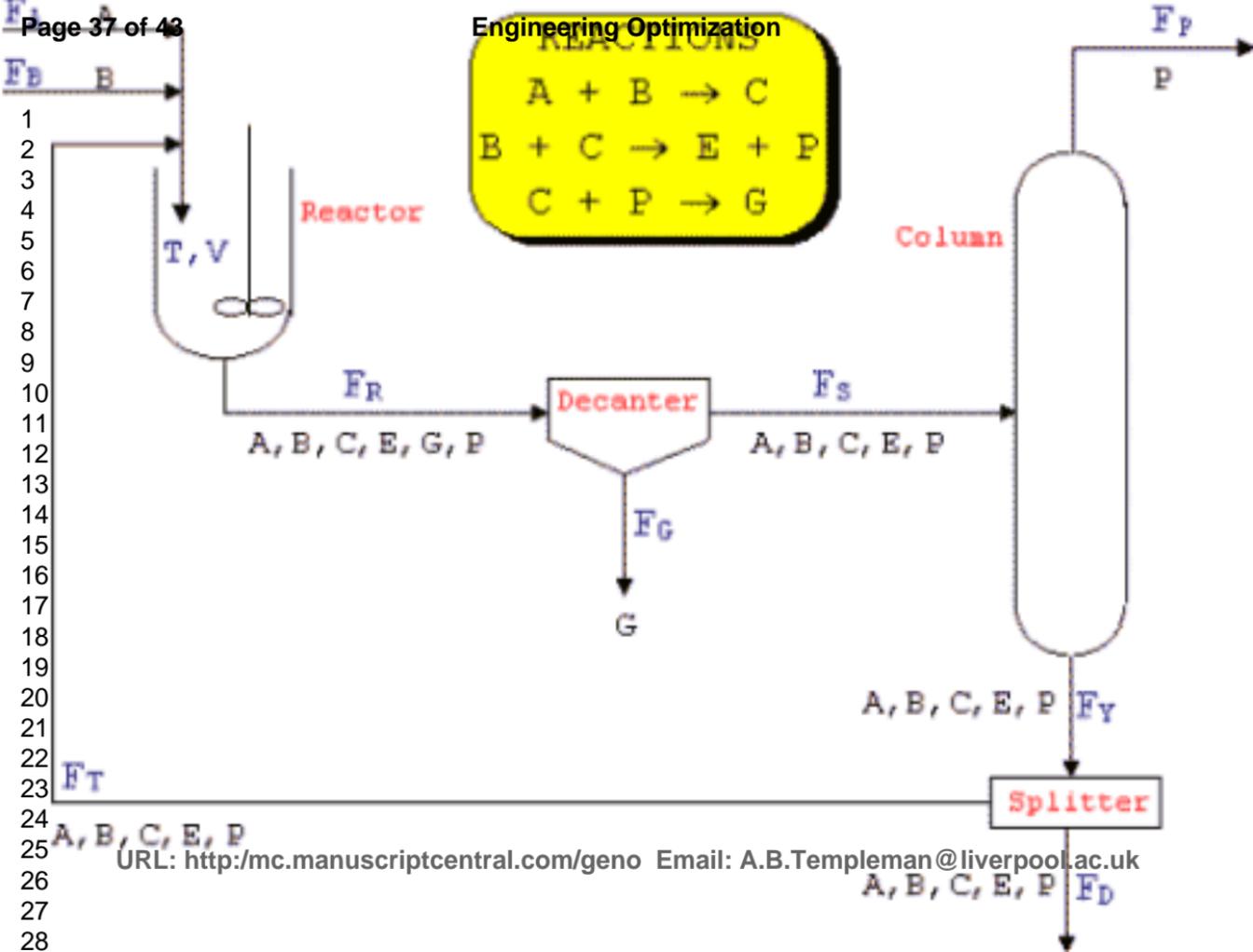
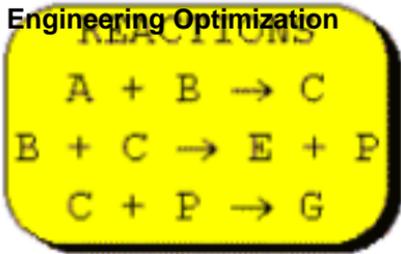
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

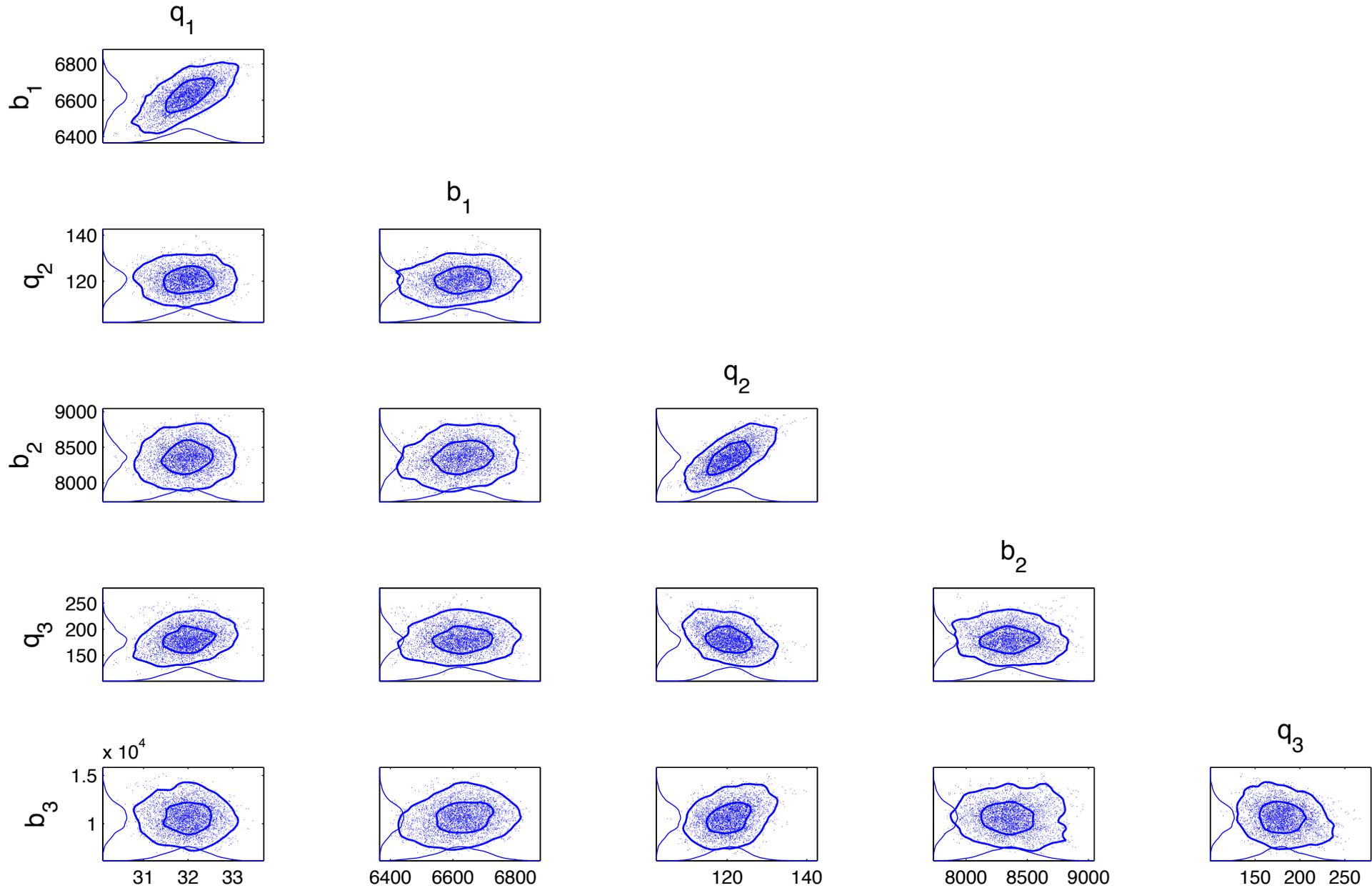


URL: <http://mc.manuscriptcentral.com/geno> Email: A.B.Templeman@liverpool.ac.uk



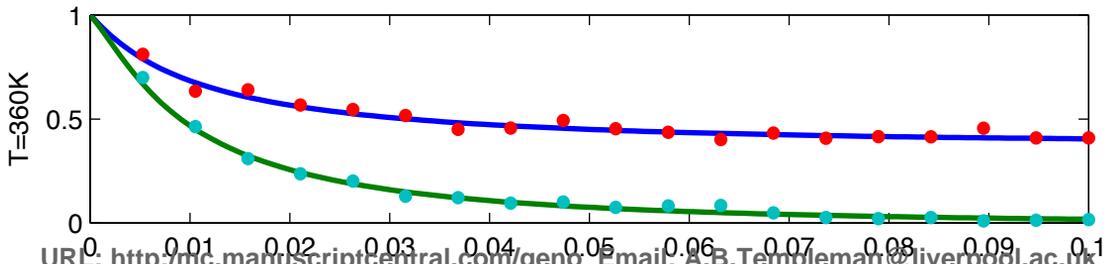
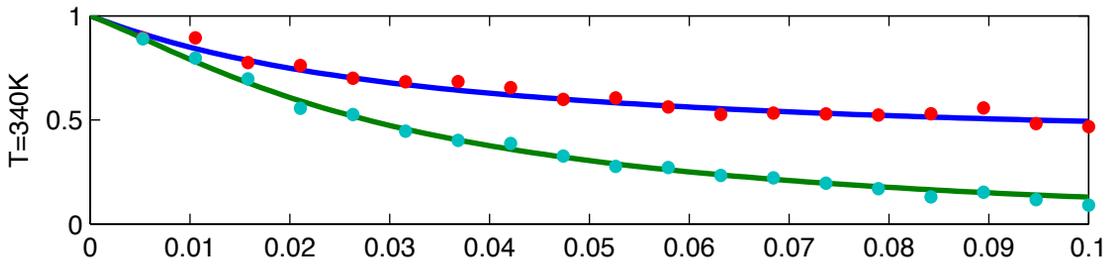
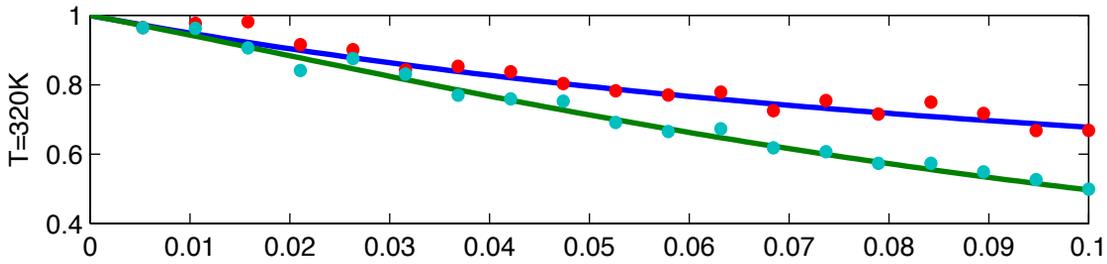


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

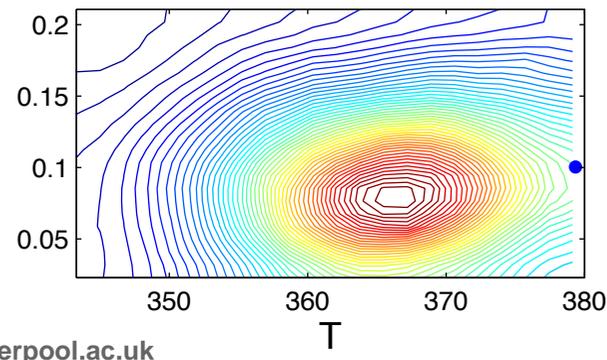
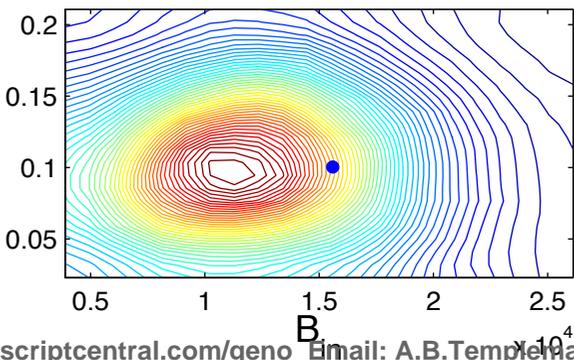
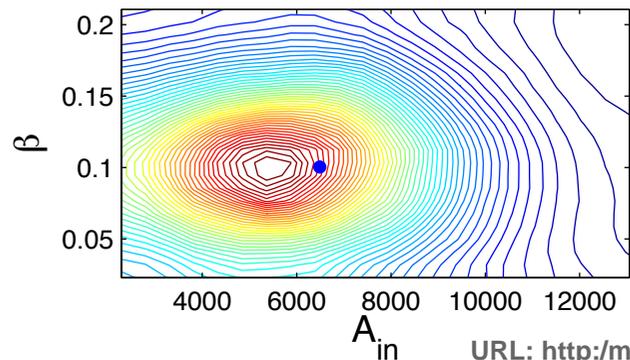
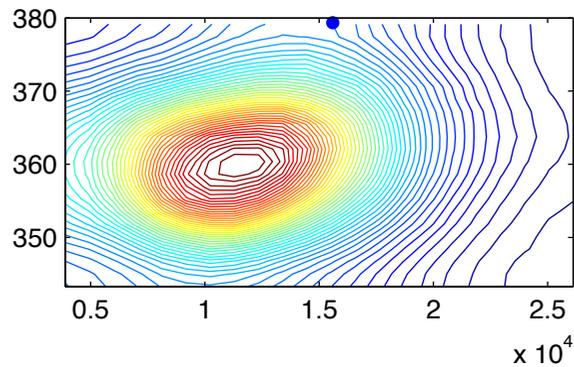
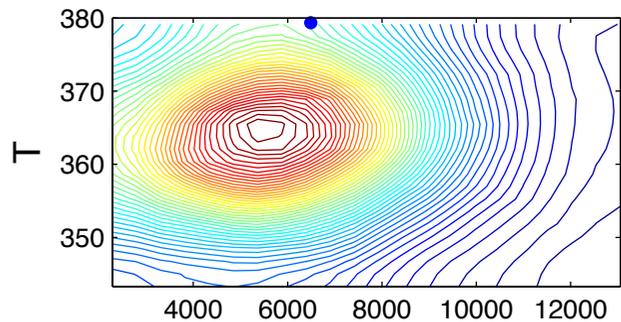
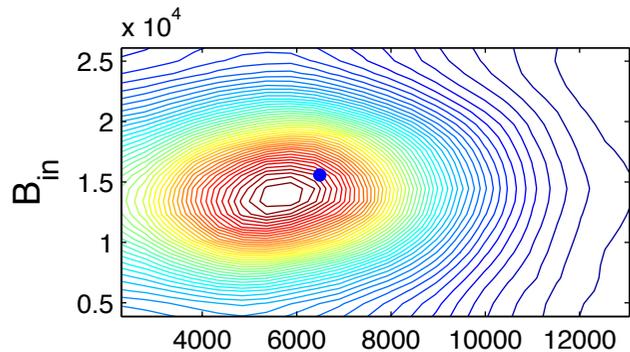


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

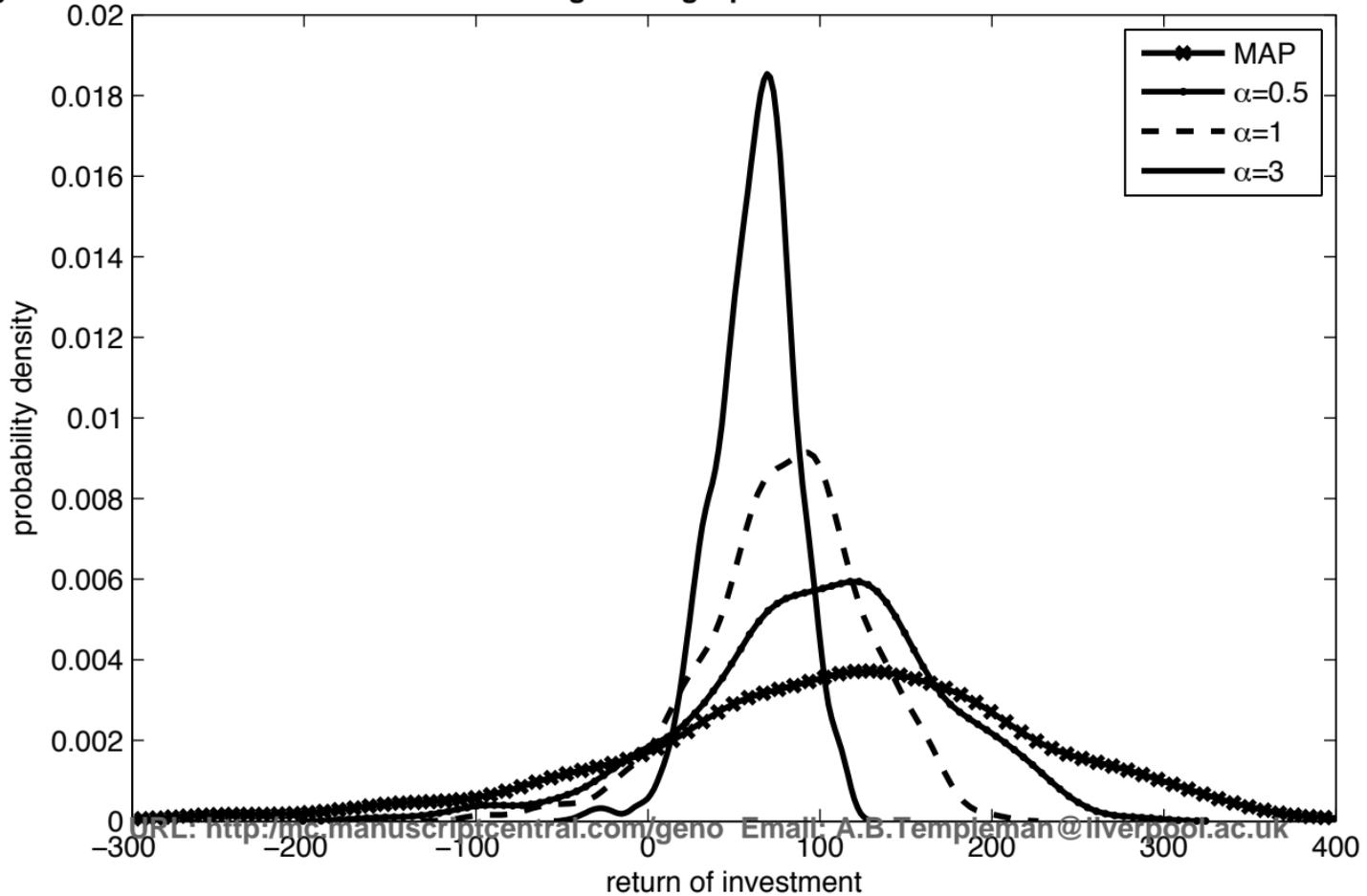
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29



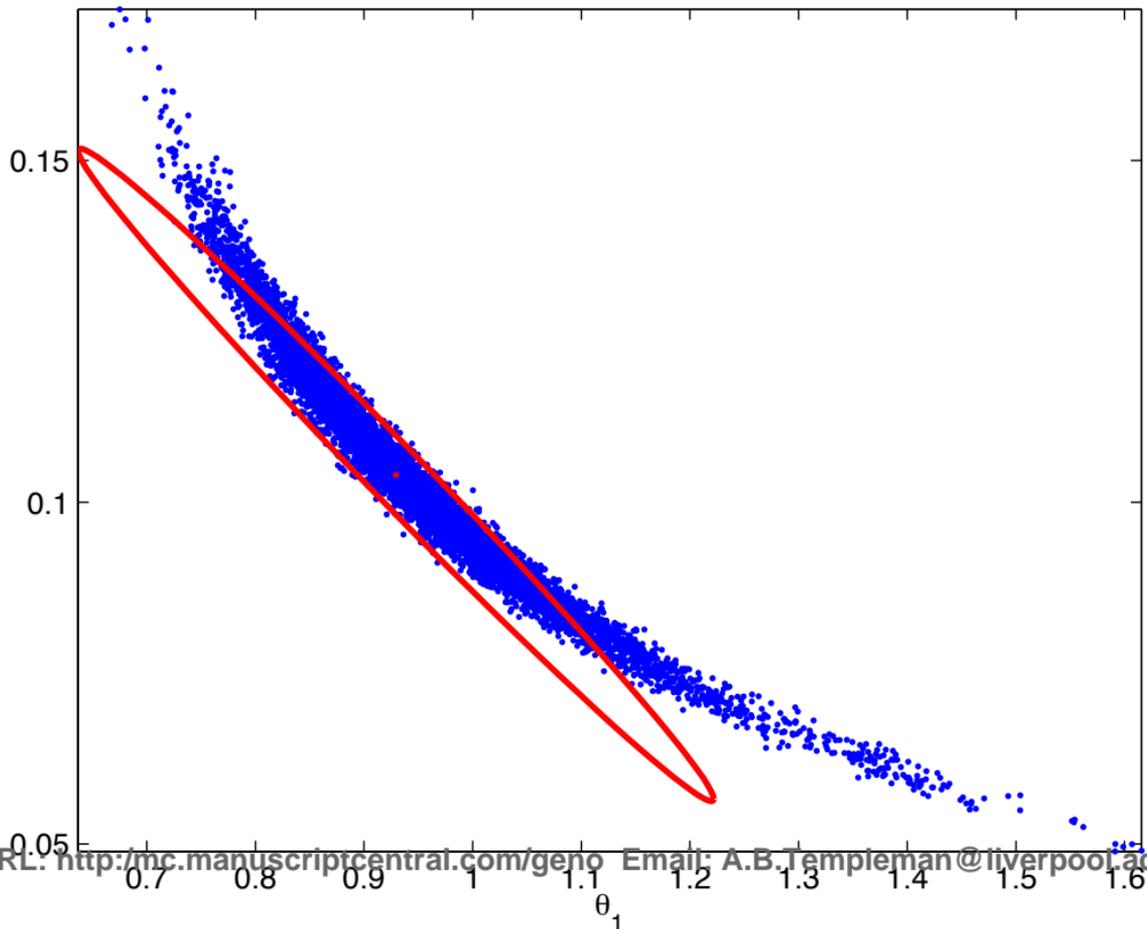
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

