 DOI: 10.1002/masy.201000046

Diagnostic Checks and Measures of Information in the Bayesian Design of Experiments with Complex Polymerizations

Afsaneh Nabifar,[1] *Neil T. McManus,*[1] *Eduardo Vivaldo-Lima,*[2] *Alexander Penlidis**[1]

Summary: A case study is presented from the bimolecular nitroxide-mediated radical polymerization of styrene. A Bayesian approach is implemented in order to design two optimal sequences of 2-trials each, and compared with a standard fractional factorial design. The improved effectiveness of the Bayesian design is demonstrated through a discussion of diagnostic criteria (on the quality of prior knowledge and the significance of estimated effects) and enhanced information content measures.

Keywords: Bayesian design of experiments; kinetics (polym.); nitroxide-mediated radical polymerization; polymerization (general); polystyrene (PS)

Introduction

A Bayesian approach to the design of experiments has the same objectives as standard experimental (full or fractional factorial) designs but with significant practical benefits over standard design methods. The ability of the Bayesian approach to incorporate prior process knowledge (which is available in most of the cases but usually discarded) into the design (prior knowledge coming from a variety of sources) is a distinct advantage. This and other advantages of the Bayesian methodology (improvements with respect to information content retrieved from process data, relative ease in changing factor levels in the middle of experimentation, flexibility with factor ranges and overall ''cost''-effectiveness (time and effort/resources) with respect to the number of experiments) were recently highlighted in Nabifar et al.,[1] with examples from complex polymerization scenarios drawn specifically from the nitroxide-mediated radical polymerization (NMRP) of styrene, under both bimolecular and unimolecular initiating options. Even more recently, Nabifar et al.,[2] in a companion paper, expanded on and reinforced the advantages of the Bayesian technique by looking at additional polymerization case studies and demonstrating the attractiveness of the technique, which can suggest a set of future process experiments in an optimal, sequential and iterative fashion. Exploiting the Bayesian design in complex polymerizations (and in other processes, since the technique is perfectly general), could hopefully lead to optimal performance in fewer trials, thus saving time and money.

Table 1 gives a compendium of experimental design issues that can effectively be handled by the Bayesian design approach (and hence pose typical limitations for standard experimental designs). Most (if not all) of these issues along with the more superior performance of the Bayesian approach in handling them have been demonstrated via the case studies of Table 2. An ''identifier letter'' in Table 1

[1] Institute for Polymer Research (IPR), Department of Chemical Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
Fax: (519) 746-4979;
E-mail: penlidis@cape.uwaterloo.ca
[2] Facultad de Química, Departamento de Ingeniería Química, Universidad Nacional Autónoma de México, CP 04510 México D.F., México

Table 1.
Overview of issues handled by the Bayesian design approach.

Issue	Identifier
Flexible wrt number of trials that can be designed	A
Changing factor level/range in the middle of experimentation	B
Accommodating extra trial(s) mid-way through experimentation	C
Sequential nature (n-trials vs. sequences of fewer trials)	D
Process constraints (and impractical treatment combinations)	E
Situations with missing observation(s)	F
Increase of information content	G
Possible detection of nonlinearities	H
Incorporation of prior knowledge	I
Flexible wrt source of prior knowledge (screening experiments vs. models and/or combinations)	J
Flexible wrt quality of prior knowledge (informative vs. non-informative priors)	K
Factors with several (or combination of) levels	L
Dropping/adding factors	M
Iterative fashion	N
Single vs. multi-response scenarios	O

Table 2.
Case studies/examples with Bayesian design implementations on complex polymerization scenarios.

Case study	Process details	Responses	Identifier
Design 2 experiments (Full factorial experiment (exp) = 8)	Bimolecular NMRP Three factors (T, [I], [N]) Prior knowledge: mechanistic model	Single: batch time	A, I, K, N
Design 4 exp (two sequences of 2-trials each vs. 4-trials; comparison with fractional factorial) (Full factorial exp = 8; half fraction of full factorial = 4)	Bimolecular NMRP Three factors (T, [I], [N]) Prior knowledge: mechanistic model	Single: batch time	A, D, G, I, K, N
Design 4 experiments (expanded T range in the middle of experimentation) (Full factorial exp = 8)	Bimolecular NMRP Three factors (T, [I], [N]) Prior knowledge: mechanistic model	Single: batch time	A, B, D, H, I, L, N
Design 3 experiments (accommodating extra trial to gain new information) (Full factorial exp = 8)	Unimolecular NMRP Three factors (T, [I], $\overline{M}_n(I)$) Prior knowledge: empirical model	Single: number-average molecular weight	A,C, D, G, I, J, K, N
Design 5-trial experiment vs. two sequences of 3- and 2- trials each (Full factorial exp = 8)	Bimolecular NMRP Three factors (T, [I], [N]) Prior knowledge: mechanistic model	Single: weight-average molecular weight	A, D, G, I, N
Two sequences of 2-trials each (design experiments that are nearly optimal for all responses) (Full factorial exp = 8)	Unimolecular NMRP Three factors (T, [I], $\overline{M}_n(I)$) Prior knowledge: reduction of mechanistic model from bimolecular to unimolecular	Two: batch time and weight-average molecular weight	A, D, G, I, N, O
Three sequences of 4-, 4- and 10-trials each (Full factorial exp = 128)	Emulsion terpolymerization Seven factors (T, [I], [monomer], [CTA], [impurity], [emulsifier], I type) Prior knowledge: mechanistic model	Four: batch time, terpolymer composition, weight-average molecular weight, total # of particles	A, B, C, D, E, F, I, L, M, N, O
Two sequences of 4-trials each (Full factorial exp = 64)	Suspension polymerization Six factors ([CTA], [stabilizer], speed of agitation, dispersed phase hold-up, two factors from impeller geometry Prior knowledge: mechanistic model	Two: mean particle size, coefficient of variation of the particle size distribution	A, E, F, I, N, O

cross-references the entries (design issues) of Table 1 with corresponding case studies in Table 2, where the issues have been addressed. A few explanations/remarks for the entries of Table 2: (a) T, [I] and [N] stand for polymerization temperature, initiator concentration (I represents initiator, in general) and controller (for example, TEMPO) concentration, respectively (b) $\overline{M}_n$ (I) represents the average molecular weight of the unimolecular initiator (c) CTA is chain transfer agent (d) Comparisons of the number of Bayesian trials are made (see first column of Table 2) with either full factorial trials (2^k, k = number of experimental factors considered in the design) or fractional factorial trials (usually, half fraction of a full factorial design); observe how much fewer the Bayesian runs are with respect to both full and fractional factorial trials (e) The column of responses (third column of Table 2) gives a feel about the many different measured variables that can be employed.

Since details about the Bayesian analysis steps have been discussed in Nabifar et al.,[1,2] they are not repeated herein for the sake of brevity. Instead, a summary of the different steps for implementing the Bayesian design of experiments is given in a point form in Table 3, along with the pertinent equations used. One should note that the trials suggested by the Bayesian design are optimal with respect to the criterion of Eq. 6 of Table 3. In addition, one can easily appreciate the sequential and

Table 3.
Summary steps for implementing the Bayesian design of experiments.

1. Select the design factors and their levels; select response(s).
2. Choose a model for process response(s), as a function of factors (variables) under study and related parameters:

$$\underline{y} = f(\underline{X}, \underline{\theta}) + \underline{\varepsilon} \quad (1)$$

2.1 For simplicity, a linear regression model can be used:

$$\underline{y} = \underline{X}\underline{\theta} + \underline{\varepsilon} \quad (2)$$

where $\underline{y}$ is an $n \times 1$ vector of observations with y_i representing the i^{th} observation. $\underline{X}$ is an $n \times p$ design matrix, with its elements having the values of $+1$ and -1, corresponding to the high and low levels of the design factors, respectively. A row of the $\underline{X}$ matrix corresponds to an experimental trial (treatment). n is the number of trials and p is the number of parameters or effects. $\underline{\theta}$ is the $p \times 1$ vector of true, yet unknown, values of parameters. Finally, $\underline{\varepsilon}$ is the error vector which is assumed to be normally distributed with mean $\underline{0}$ and variance matrix $\underline{I}\sigma^2$ ($\underline{I}$ is the identity matrix). σ^2 is the variance of the response(s).
3. Assume a distribution for the prior knowledge about the parameters ($\underline{\theta}$).
3.1 For simplicity, the prior knowledge held about $\underline{\theta}$ before the experimentation is assumed to be following a multivariate normal distribution:

$$\underline{\theta} : N[\underline{\alpha}; \underline{U}] \quad (3)$$

$\underline{\alpha}$ is a $p \times 1$ known vector of the parameter means and $\underline{U}$ is a positive definite $p \times p$ known matrix of the variances (covariances, if they exist) of the parameter means.
4. Apply Bayes' theorem to obtain the posterior distribution.
4.1 Here, the posterior distribution is also normally distributed with mean, $\underline{\hat{\theta}}$, and updated variance/covariance matrix $\underline{U}$:

$$\underline{\hat{\theta}} = \left[\underline{U}^{-1} + \left(1/\sigma^2\right)\underline{X}'\underline{X}\right]^{-1}\left[\underline{U}^{-1}\underline{\alpha} + \left(1/\sigma^2\right)\underline{X}'\underline{y}\right] \quad (4)$$

$$\underline{U} = \left[\underline{U}^{-1} + \left(1/\sigma^2\right)\underline{X}'\underline{X}\right]^{-1} \quad (5)$$

5. Choose design criteria to select the "best" set of experiments.
5.1 Here, a design based on D-optimality has been applied, by choosing the best set of trials of a full factorial experiment to minimize the variance/covariance matrix (Eq. 5), which is equivalent to maximizing determinant H:

$$H = \left|\underline{I} + \left(1/\sigma^2\right)\underline{X}\underline{U}\underline{X}'\right| \quad (6)$$

6. Run and analyze the experiments (which have been selected using Eq. 6)
7. Update vector of parameter means (θ) and variance/covariance matrix (U) using Eq. 4 and Eq. 5.
8. Given the new $\underline{U}$, use Eq. 6 to select the next sequence of trials.
9. Analyze the experiments and update $\underline{\theta}$ and $\underline{U}$; repeat steps 4 to 9.
10. Stop the experimentation once the values of parameters of interest are known with accepted degree of certainty. Update the vector of parameters, after the analysis of the final sequence of experiments.

iterative nature of the technique by scrutinizing the steps (1-10) of Table 3.

Two typical questions that often arise in Bayesian design implementations have to do with how effectively one can make statements about the quality of prior knowledge and the significance of the estimated effects (from the designed experiments), and about the gain in information content. These two important questions were not addressed in detail in earlier Bayesian design implementations,[1–4] and hence they are the topic of the present contribution. In other words, the following questions, intimately related to the Bayesian experimental design technique, will be addressed in what follows: (1) What statistical diagnostic criteria can one use in order to shed light on the quality of prior knowledge (see step 3, and implicitly step 2, of Table 3) and the significance of estimated effects (see step 7 of Table 3)? (2) What measures of information (content) are available and what different aspects of the design procedure can they emphasize (see steps 4 and 5 of Table 3)? Addressing questions (1) and (2) above not only clarifies the design steps further but also could only make one more confident in the effectiveness and practicality of the Bayesian design of experiments procedure.

Results and Discussion

The Bayesian design of experiments is now implemented to the bimolecular nitroxide-mediated radical polymerization of styrene, as an example of a complex polymerization process. However, the Bayesian technique is perfectly general and can therefore be potentially applied to other complex polymerization systems as well as other chemical engineering processes with uncertain models and/or parameters.

In a typical bimolecular NMRP, a nitroxide radical, like 2,2,6,6-tetramethyl-1-piperidinyloxy (TEMPO), is present in addition to a free radical initiator, like benzoyl peroxide (BPO) or azobisisobutyronitrile (AIBN). The nitroxide radical reacts with the carbon-centered free radical of the growing chain in a reversible process and dramatically lowers the concentration of the active (free) radicals. This, coupled with the inability of the nitroxide radical to initiate new chain growth, leads to controlled polymerization. We are not offering a literature (or background) review on NMRP, as the emphasis here is to demonstrate the benefits of specific aspects of the Bayesian design of experiments and not the kinetics of NMRP. For the interested reader, kinetic and modeling information closely related to the case study presented herein could be found in references [5-7]. The case study that follows implements the Bayesian design steps of Table 3, with added emphasis on the interpretation of diagnostic criteria and information measures.

The first step is to select the design factors with their levels, and the responses. Three factors are chosen and shown (with their initial levels) in Table 4. The selection of factors and their low and high levels were based on a detailed and critical analysis of the literature, combined with some of our previous experience about the process.[6,7] Reaction (batch) time (in hrs) to reach 60% conversion is chosen as the single response in this case study. It was felt (based on process information and experience) that due to various sources of experimental error, the time for a batch to reach 60% conversion would vary by ± 1 hr about 95% of the time.

Casting the prior knowledge into the vector of parameter means ($\underline{\alpha}$) and the variance/covariance matrix ($\underline{U}$) is the next important step (see Table 3). To generate the initial values for the parameter means,

Table 4.
Selected factors and their levels (T = temperature, [I] = initiator concentration, [N] = nitroxide concentration).

Level	T (°C)	[I] (M)	[N] (M)
low	120	0.0305	0.0324
high	130	0.036	0.0396

Table 5.
Elements of **prior** $\underline{\alpha}$ and $\underline{U}$.

Parameter	α_i	U_{ii}
Mean	7.795	11.9497
T	−3.189	0.25
[I]	−0.295	0.25
[N]	0.250	6.25
T × [I]	0.167	0.0068
T × [N]	−0.178	1.5625
[I]×[N]	−0.239	0.5625
T × [I] × [N]	0.144	0.0156

Table 6.
Four possible 2-trial experiments for the **first** sequence.

Experiment	T	[I]	[N]	[N]/[I]	Batch time response
1	−1	−1	−1	1.06	9.63
	−1	−1	1	1.30	13.26
2	−1	1	−1	0.9	–
	−1	1	1	1.1	–
3	1	−1	−1	1.06	–
	1	−1	1	1.30	–
4	1	1	−1	0.9	–
	1	1	1	1.1	–

a 2^3 standard factorial design is used and the corresponding batch times are obtained from a general mechanistic model developed for NMRP of styrene.[5,7,8] This model, based on the general reaction mechanism, consists of mole balances for the main recipe ingredients and detailed population balances for the molecular weight part, which eventually leads to the development of 22 ordinary differential equations. The vector of parameter means ($\underline{\alpha}$), shown in the second column of Table 5, is obtained via conducting a linear regression on the results (see Eq. 2 in Table 3). In order to determine the variances of the parameters, each parameter is examined separately and based on our knowledge of the process, a reasonable estimate of the variance is made. This stage is where most of the brainstorming takes place. In cases where the parameter (or the effect) is considered to be well known, a smaller interval of uncertainty is given, i.e., a smaller fluctuation about the mean is tolerated. Since in the bimolecular NMRP of styrene nitroxide concentration plays an important role, it is genuinely desired to find out more information about the model term involving this factor. Hence, higher initial variances are allocated to the parameters related to nitroxide concentration ([N]) and its interactions. The diagonal elements of the prior variance/covariance matrix (U) are shown in the third column of Table 5; the off-diagonal elements are initially all set to zero.

The next step is selection of the ''best'' experiments using the search algorithm. As an example, it was decided to run a total of four experiments in two sequences of 2-trials each (the Bayesian technique allows one to design any number of trials per sequence, and any number of sequences). Following the metric for designing experiments, i.e., maximizing the determinant H (Eq. 6 of Table 3), four 2-trial experiments were found that corresponded to a maximized H (H = 460.77), as shown in Table 6. Now let's make a few practical remarks, in order to demonstrate the efficiency of the design procedure and the importance of the interplay between Bayesian steps and ''process'' sense. If this case study was examined blindly, somebody could have picked one of the four experiments of Table 6 completely randomly. However, combining knowledge about the bimolecular NMRP of styrene with the suggested results from the Bayesian design, we can now make a more judicious choice of the appropriate experiment, which is optimal from both a statistical and process sense. For instance, from previous experience with the bimolecular NMRP of styrene we know that the ratio of concentrations of nitroxide to initiator ([N]/[I]) is important. Experiments 1 and 3 (and experiments 2 and 4) in Table 6 lead to trials with the same nitroxide to initiator ratio at two different temperatures. It is decided to choose the lower temperature (120 °C), as it is the most common temperature used in bimolecular NMRP. Hence, running experiments 1 and 2 would be preferable. Between experiments 1 and 2, experiment 1 would appear even more preferable, as in this experiment

Table 7.
Posterior variance/covariance matrix U after the **first** sequence of experiments.

0.9286	0.2306	0.2306	0	−0.0063	0	0	0
0.2306	0.2452	−0.0048	0	0.0001	0	0	0
0.2306	−0.0048	0.2452	0	0.0001	0	0	0
0	0	0	1.8563	0	1.0984	0.3954	−0.0110
−0.0063	0.0001	0.0001	0	0.0068	0	0	0
0	0	0	1.0984	0	1.2879	−0.0989	0.0027
0	0	0	0.3954	0	−0.0989	0.5269	0.0010
0	0	0	−0.0110	0	0.0027	0.0010	0.0156

a wider range of [N]/[I] is examined (1.06–1.3 vs. 0.9–1.1). Hence, experiment 1 is conducted for the first sequence of 2 trials and the corresponding batch time responses are shown in Table 6. Here we demonstrated the importance of a parallel correct scientific/practical decision, not a statistical one, which might not be apparent, if not for the added emphasis of the Bayesian methodology on prior information.

For the next step, Eq. 5 in Table 3 is employed to calculate the posterior variance/covariance matrix (shown in Table 7; one can now observe the presence of nonzero off-diagonal elements, i.e., covariance elements, in the updated U matrix). The updated variance/covariance matrix can now be used back in Eq. 6 (of Table 3) to design (in an iterative sequential fashion) the next sequence that maximizes the determinant H. Only one 2-trial experiment corresponds to the highest value of H (H = 78.73) this time. The corresponding levels for the factors are shown in Table 8,

Table 8.
The **second** sequence of 2 trials.

T	[I]	[N]	[N]/[I]	Batch time response
1	1	−1	0.9	3
1	1	1	1.1	5

along with the batch time responses obtained from conducting experiments in the lab. Scrutinizing again Tables 6 and 8, one can see that the Bayesian design is choosing the second sequence at a higher temperature level (130 °C) and different ratios of [N]/[I]. Therefore, in a very succinct way, the total of 4 runs (two sequences of 2-trials) cover both temperatures and all four different [N]/[I] ratios.

The updated variance/covariance matrix is shown in Table 9. Comparing Table 7 with Table 9, one can see that after the second sequence of experiments, the uncertainties about the parameters (diagonal elements of the U matrix) have decreased. Also, most of the covariances (off-diagonal elements of the U matrix; measures of the strength of the correlation between parameters) appearing in the first sequence (Table 7) are zero in the second sequence (Table 9). Eq. 4 of Table 3 can now be employed to calculate the posterior $\underline{\theta}$ after each sequence of 2-trials (results are shown in the 3rd and 4th column of Table 10; the superscript in $\theta_i^{()}$ denotes the sequence of trials).

A series of statistical diagnostic tests can now be carried out in addition to the steps of Table 3. These diagnostic tests serve to quantify the relative importance of the

Table 9.
Posterior variance/covariance matrix U after the **second** sequence of experiments.

0.2514	0	0	0	−0.0067	0	0	0
0	0.1667	−0.0833	0	0	0	0	0
0	−0.0833	0.1667	0	0	0	0	0
0	0	0	0.2548	0	0	0	−0.0150
−0.0067	0	0	0	0.0068	0	0	0
0	0	0	0	0	0.5345	−0.3701	0
0	0	0	0	0	−0.3701	0.4293	0
0	0	0	−0.0150	0	0	0	0.0156

Table 10.
Summary results of diagnostic tests.

Parameter	Prior α_i	$\theta_i^{(1)}$	$\theta_i^{(2)}$	Test 1	Test 2 (after 2nd seq.)	Test 3 (after 2nd seq.)
Mean	7.795	7.795	7.561	**2.2549**	**15.0793**	−0.4667
T	−3.189	−3.189	−3.269	**−6.3779**	**−8.0065**	−0.1952
[I]	−0.295	−0.295	−0.374	**−0.5890**	−0.9166	−0.1952
[N]	0.250	0.953	1.220	0.1000	**2.4167**	1.9215
T × [I]	0.167	0.167	0.167	**2.0211**	**2.0200**	−0.0016
T × [N]	−0.178	−0.354	−0.170	−0.1422	−0.2331	0.0100
[I] × [N]	−0.239	−0.302	−0.236	−0.3185	−0.3605	0.0040
T × [I] × [N]	0.144	0.146	0.147	1.1551	1.1759	0.0194

parameters (i.e., factor effects) and their interactions, as well as the quality of prior knowledge (in other words, the adequacy of the model used to generate the ''prior knowledge'', as the case might be). Results from these statistical diagnostic tests for our example are shown in Table 10. The 2nd column presents the initial values of the parameter means ($\underline{\alpha}$); the 3rd and 4th columns contain the updated estimates of the parameter means after the first and second sequence, respectively. Test 1 (5th column of Table 10) is defined as the ratio of the prior mean (2nd column of Table 10) to the prior standard deviation of the mean $[\alpha_i/(U_{ii})^{1/2}]$, where (U_{ii}) is shown in the third column of Table 5. This test checks the null hypothesis that $\alpha_i = 0$ purely in the opinion of the ''expert'' (the person who assigned the values for the prior effects and variances). It is essentially a measure of the uncertainty of the ''expert''. A value greater than 2 or less than -2 is considered significant (this is equivalent to a 95.44% confidence interval). Test 2 is a measure of the actual significance of an effect. For instance, the test 2 value after the second sequence of experiments (6th column of Table 10) is equal to the second updated estimate of the effect, $\theta_i^{(2)}$, divided by the square root of the corresponding diagonal element of the last posterior variance/covariance matrix (see Table 9; $[\theta_i^{(2)}/(U_{post(ii)})^{1/2}]$). Once again, a value greater than 2 or less than -2 implies significance. Finally, test 3 is a measure of the quality of the expert's opinion. For instance, test 3 after the second sequence of experimentation (7th column of Table 10) is equal to $(\theta_i^{(2)} - \alpha_i)$ divided by the square root of the diagonal element of the last posterior variance/covariance matrix (see again Table 9; $[(\theta_i^{(2)} - \alpha_i)/(U_{post(ii)})^{1/2}]$). A significant value greater than 2 or less than -2 implies that portion(s) of the mechanistic model related to the response(s) and the effects in question may need further refinement.

A careful analysis of the summary results of Table 10 can lead to several interesting remarks. Based on the results of test 1, the influential factors on the batch time response, purely in the opinion of the ''expert'', are temperature (T) and the interaction between temperature and initiator concentration (T × [I]). The results of test 2 reveal that the actual significant parameters, based on the new observations, are T, [N], and T × [I]. Initiator concentration ([I]) is not affecting the batch time response (within the considered operating conditions). This result explains why in kinetic studies of bimolecular NMRP, most of the time the concentration of initiator is kept constant while the concentration of nitroxide is varied to study the effect of [N]/[I].[6,9–11] Finally, test 3 shows that the results are in agreement with the expert's opinion. However, the value of test 3 for [N] is arguably very close to 2 and thus indicates that the expert's opinion about the nitroxide concentration was not overly accurate. The result of test 3 indicates that the Bayesian design has indeed ''spotted'' this and identified that the expert's opinion was not valid about [N], yet another of the implicit advantages of Bayesian design.

If a standard fractional factorial design was chosen to be carried out in this example, a half fraction of a full factorial design ($\frac{1}{2}2^3$) with resolution III would have been chosen. This means that all of the main factors would be confounded with 2 factor interactions. An important difference between the Bayesian and standard experimental design is that employing prior information usually prevents the complete confounding which is present in the results of the corresponding standard design experiments.

Another flexible trait of the Bayesian approach is that it allows one to easily run an experiment with a non-standard number of trials (for example, two sequences of 2-trials each, as in this example). This can only result in a more efficient use of resources, while providing quite acceptable parameter estimates.

In order to ''visualize'' and demonstrate the improvement achieved by the Bayesian methodology (compared to a standard fractional factorial design), various possible measures of information (content) are discussed next. The Fisher information is a way of measuring the amount of information that an experiment carries about an unknown parameter θ. The Fisher information and the estimated variance are reciprocal, thus minimizing the variance corresponds to maximizing the information. In the case of multi-parameter scenarios, the means of the parameters form a vector and their variances are included in a variance/covariance matrix. The inverse of the variance/covariance matrix is called the ''information matrix''. Since the Fisher information is now in the form of a matrix, this matrix could be compressed using a real-valued summary statistic (like the determinant or trace of the matrix). Being real-valued functions, these ''information criteria'' can now be maximized. In the Bayesian design of experiments, the information matrix is the reciprocal of the posterior variance/covariance matrix (Eq. 5 of Table 3). The summary statistic used in our approach is the determinant. Hence, Fisher information $(\mathrm{FI}) = |\underline{U}^{-1} + (1/\sigma^2)\,\underline{X}'\,\underline{X}|$. Scrutinizing the determinant H (Eq. 6 of Table 3), one can see that our experimental criterion to arrive at optimal experiments is indeed a function of the Fisher information. For our example case study presented here, FI after two sequences of 2-trials each is 8.42×10^7. The Fisher information would be 6.65×10^7 in the case of a half fraction of a full factorial design. Therefore, the Bayesian design is an improvement with respect to information content.

Another advantage of the Bayesian design is that it moves in the direction of minimizing the variance of the parameters. For example, Figure 1 illustrates the evolution of the variances for the nitroxide concentration effect (corresponding to element U_{44} of the variance/covariance matrix) throughout the experimentation. As one can see, this value has dropped from 6.25 in the prior (Table 5) to 1.85 (Table 7) to 0.25 after the second sequence (Table 9), thus demonstrating that the Bayesian design in its sequential nature minimizes the uncertainty in the parameter values (translating uncertainty into variance of the effects and the related parameters).

Another indication that the Bayesian approach is an improvement over the corresponding standard fractional factorial design can be visualized through the estimated response from the regression model (Y_{model}), where $Y_{\mathrm{model}} = \underline{X\theta}$. Figure 2

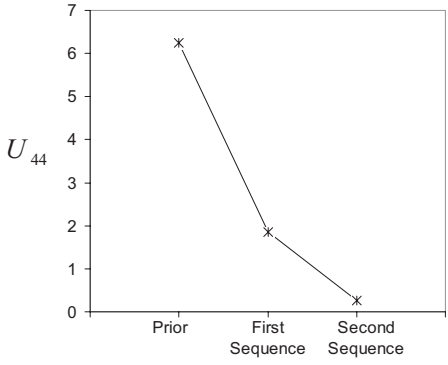

Figure 1.
Evolution of variance (U_{44}) through the course of experimentation.

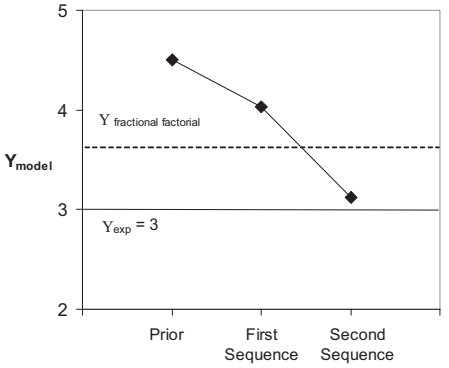

Figure 2.
Evolution of Y_{model} through the course of experimentation.

illustrates an example of the evolution of the estimated response for the conditions of the first trial of the second sequence (the 1st row of Table 8) through the course of experimentation and compares the corresponding estimated response with the actually observed value (Y_{exp}; solid line in Figure 2). The estimated response from the fractional factorial experiment is also shown for comparison (dashed line in Figure 2). As one can see, Y_{model} after the second sequence of the Bayesian experiments is very close to the actually observed Y_{exp}, while the corresponding Y_{model} calculated for the fractional factorial experiment is farther away from the actually observed value. Once more, in the Bayesian steps, Y_{model} moves in the direction of minimizing its difference with Y_{exp}.

Finally, Figure 3 illustrates yet another measure of information content, as expressed by the 95% joint confidence region (JCR). It can be seen in Figures 3a, 3b and 3c that the areas of the joint

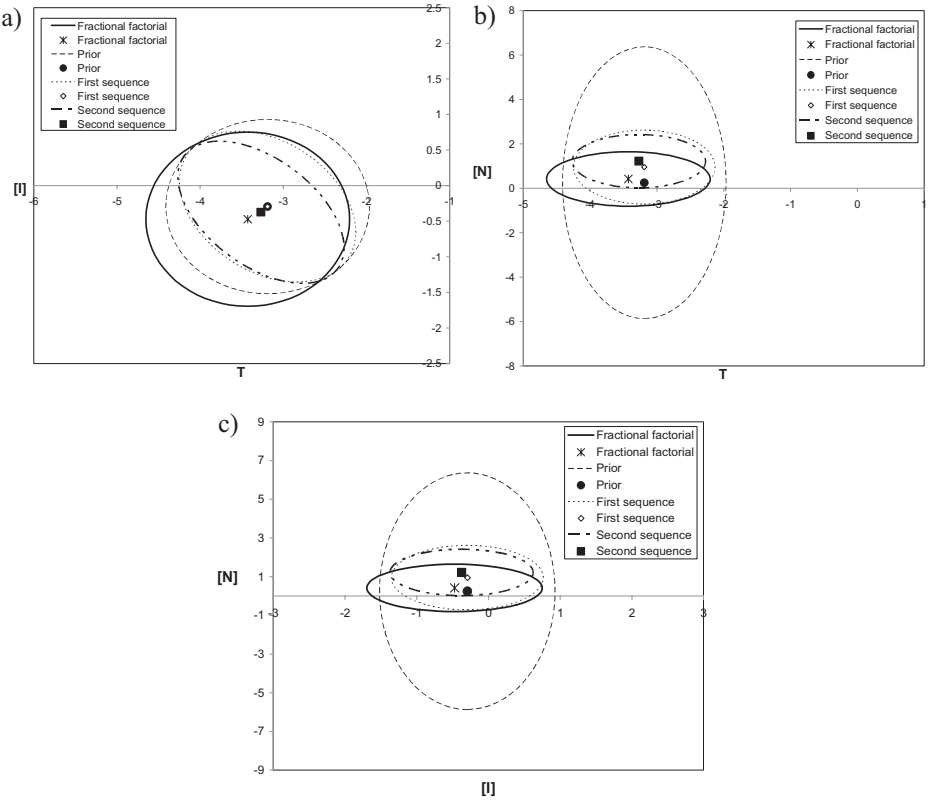

Figure 3.
95% joint confidence regions (JCRs) for parameters, a) T-[I], b) T-[N], and c) [I]-[N].

confidence regions are significantly reduced from prior to the first sequence, and then from the first sequence to the second sequence of experimentation. The reduction indicates that the precision in parameter estimates has increased. These figures also contain the JCRs for the corresponding standard fractional factorial design and they all demonstrate that the novel Bayesian approach is indeed significantly more effective than the standard fractional factorial (much smaller JCRs for the Bayesian designed experiments).

The joint confidence region for the T-[I] parameters is slightly tilted after the second sequence (see Figure 3a), indicating that the T and [I] related parameters are moderately correlated. Going back to the results of Table 7 and Table 9, ranslating covariances into the corresponding correlation values (Covariance (i, j)/[Variance (i) $\times$ Variance (j)]$^{0.5}$), one would see that most of the correlation coefficients are considerably smaller in absolute value than 1. In theory, the magnitude of the correlation values is the factor that would affect the orientation of the final joint confidence regions of the parameters. For instance, in Table 9, U_{32} shows the covariance between [I] and T. U_{32} being non-zero will give rise to a joint confidence region that is tilted. The negative covariance ($U_{32} = -0.0833$) implies that initiator concentration, [I], tends to decrease as temperature, T, increases (see Figure 3a). The JCRs in Figures 3b and 3c are completely horizontal, representing completely uncorrelated parameters (which again confirms the results of Table 7 and Table 9).

Conclusion

Two important questions, intimately related to the Bayesian design of experiments, were addressed via a complex polymerization example. The first had to do with statistical diagnostic criteria that one can use in order to make statements about the quality of prior knowledge and the significance of estimated factor effects. These statistical diagnostic checks can ''spot'' and identify situations when the expert's opinion (prior knowledge) is not valid about certain parameters (hence, factor effects). The second question was concerned with enhanced information measures (obtained from the Bayesian procedure versus a standard design of experiments), as another illustration of the effectiveness and practicality of the approach. It was shown that the Fisher information (FI) matrix is maximized in the Bayesian approach, which corresponds to minimizing the variances and reducing the 95% joint confidence regions (JCR), hence improving the precision of the parameter estimates.

Acknowledgements: The authors wish to acknowledge financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Canada Research Chair (CRC) program. Also, thanks go to UWW/OMNOVA Solutions, USA, for a special scholarship to A. Nabifar for 2009/ 2010.

[1] A. Nabifar, N. T. McManus, E. Vivaldo-Lima, P. M. Reilly, A. Penlidis, *Macromol. React. Eng.* **2010**, *4*, 387.
[2] A. Nabifar, N. T. McManus, E. Vivaldo-Lima, A. Penlidis, *Chem. Eng. Tech.* **2010**, *33*, to appear in issue 11.
[3] M. A. Dube, A. Penlidis, P. M. Reilly, *J. Polym. Sci. Part A-Polym. Chem.* **1996**, *34*, 811.
[4] E. Vivaldo-Lima, A. Penlidis, P. E. Wood, A. E. Hamielec, *J. Appl. Polym. Sci.* **2006**, *102*, 5577.
[5] J. Bonilla, E. Saldivar, A. Flores-Tlacuahuac, E. Vivaldo-Lima, R. Pfaendner, F. Tiscareno-Lechuga, *Polym. React. Eng. J.* **2002**, *10*, 227.
[6] A. Nabifar, N. T. McManus, E. Vivaldo-Lima, L. M. F. Lona, A. Penlidis, *Can. J. Chem. Eng.* **2008**, *86*, 879.
[7] M. Roa-Luna, A. Nabifar, M. P. Diaz-Barber, N. T. McManus, E. Vivaldo-Lima, L. M. F. Lona, A. Penlidis, *J. Macromol. Sci. Part A: Pure Appl. Chem.* **2007**, *44*, 337.
[8] B. J. Belincanta-Ximenes, P. V. R. Mesa, L. M. F. Lona, E. Vivaldo-Lima, N. T. McManus, A. Penlidis, *Macromol. Theory Simul.* **2007**, *16*, 194.
[9] R. P. N. Veregin, P. G. Odell, L. M. Michalak, M. K. Georges, *Macromol.* **1996**, *29*, 3346–3352.
[10] R. P. N. Veregin, P. G. Odell, L. M. Michalak, M. K. Georges, *Macromol.* **1996**, *29*, 2746–2754.
[11] M. Zhang, W. H. Ray, *J. Appl. Polym. Sci.* **2002**, *86*, 1630–1662.

