Organic Process Research & Development **2001**, *5*, 659–664

Optimization of an Imidate Hydrolysis Reaction: Use of DOE and Mathematical Modeling to Deliver Yield and Productivity Improvements

Dermot J. Gavin* and Carlos A. Mojica

Process Development Laboratories, Pfizer Global Manufacturing, Eastern Point Road, Groton, Connecticut 06340, U.S.A.

Abstract:

Experimental design and mathematical modeling techniques were used in the optimization of a Pinner quench process. The detailed effects of water, agitation, and temperature on both product yield and quality were elucidated. Significant improvements in both yield (5%) and throughput (>50%) were attained. This work demonstrates how the use of experimental design techniques in chemical process development adds value by allowing efficient optimization of a given synthetic step. The progress of the project from initial lab-scale investigations to full plant-scale implementation is described.

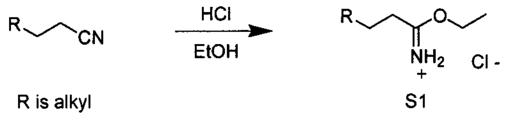
The material in question was a Pfizer pharmaceutical product. It was made by a multistep route from commercial starting materials. The high production volumes required necessitated constant improvements in yield and cycle time. Two key synthetic steps are described below.

The first step consisted of the Pinner reaction[1] of the primary cyano group with HCl gas in anhydrous ethanol, giving the imidate ester product (S1) (Scheme 1).

The second step was the hydrolysis of the imidate ethyl ester to the corresponding keto-ester (S2) under aqueous conditions at 35 °C. The batch was added to an equal weight of water and then stirred until reaction completion was achieved (Scheme 2). The reaction mixture was bi-phasic. The reactions progress can be monitored chromatographically (GC) and typically took approximately 20 h. Once the reaction was over, the aqueous layer was decanted and the organic layer washed with weak caustic solutions and then water, to remove impurities.

Both the gassing and quench reactions were prone to formation of a known impurity X; typically this material made up the remainder of the mass balance (ex GC) in low-yielding reactions. This material was believed to be formed via a thermal degradation pathway from imidate. The S2 product, the Pinner starting material and impurity X were visible on GC. This hydrolysis reaction was known to be sensitive to temperature and to the amount of water used for quenching. Since the reaction time was relatively long, it was presumed that mass transfer between the two layers was a key factor in the reaction rate. In all S2 experiments ~20% of yield is lost as material that is not visible on GC. The evidence suggests that a sequence of hydrolysis and polymerization reactions during the highly acidic quench account for this loss. The caustic washes are effective in removing these impurities.

Scheme 1. **Pinner reaction**



Scheme 2. **Pinner hydrolysis**

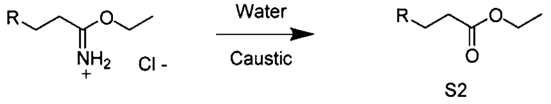


Table 1. **Range data for each factor and initial plant conditions**

factor	name	units type	−1 level	+1 level	currently
A	temp	deg °C	35.00	65.00	35
B	water vol	kg water/kg batch	0.50	1.50	1.0
C	agitation	rpm	150.00	600.00	~350 eq.

Table 2. **Design experiments and results**

std order	temp	water	agitation	S2 in-situ yield (8 h)	S2 isolated yield (wt)
1	35	0.5	150	36.2	37.2
2	65	0.5	150	18.2	20.5
3	35	1.5	150	50.2	47.7
4	65	1.5	150	57.5	60.3
5	35	0.5	600	32.2	31.8
6	65	0.5	600	29.1	33.4
7	35	1.5	600	56.6	62.1
8	65	1.5	600	58.6	61.4
9	50	1	375	52.5	55.3
10	50	1	375	54.5	57.4
11	50	1	375	53.2	57.4
12	50	1	375	52.0	57.9
calc. CP					42.3

A designed experiment (DOE)[2] study which evaluated the effect of three factors over broad ranges was felt to be the most efficient way of understanding and then optimizing this key step (S2). The factors chosen were temperature, quench water volume, and agitation rate. As well as providing a

---

* To whom correspondence should be addressed. Telephone: (860)-4418867. Fax: (860)-4416369. E-mail: dermot.gavin@pfizer.com.

(1) (a) Pinner, A.; Klein, F. *Ber.* **1877**, *10*, 1889. (b) For reviews see: Roger, R.; Neilson, D. *Chem. Rev.* **1961**, *61*, 179.

(2) (a) Box, G.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; Wiley: New York, 1978. (b) Montgomery, D. C. *Design and Analysis of Experiments*; Wiley: New York, 1997. (c) Schmidt, S. R.; Launsby, R. G. *Understanding Industrial Designed Experiments*; Air Academy Press: New York, 1978. (d) Carlson, R. *Design and Optimization in Organic Synthesis*; Elsevier: New York, 2000.

10.1021/op010225+ CCC: $20.00

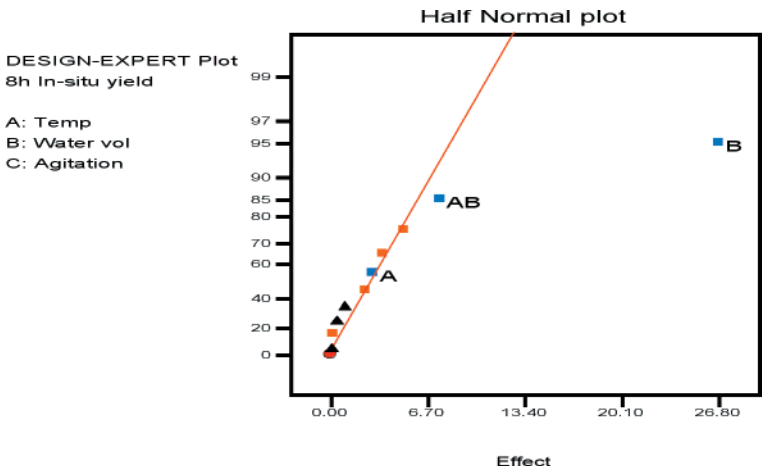


Figure 1. **Half-normal probability plot showing that water amount (B), reaction temperature (A), and their interaction effect (AB) were significant effects.**

mathematical model for desired outputs (yield, etc.) based on variation of inputs (agitation, etc.), the DOE was intended to show the overall data trend and thereby indicate the best areas for further development work.

It was decided that a two-level factorial would be the most efficient design type for this experimental program. In a two-level factorial with three factors and eight experiments, all effects due to individual factors would be identified, as well as effects due to interactions between any factor pair (Table 1). Third-order interactions (ABC), although rare, would also be identified. The current value for agitation is a guess due to difficulties in estimating the lab agitation needed to accurately mimic plant equipment. The approach taken was to find an agitation rate that just disturbed the surface of the two-phase mixture and use this as the low level; on the lab scale this was 150 rpm. The high level was correspondingly taken as the agitation rate that led to extremely efficient mixing (mixture appeared as one layer); this was 600 rpm. In general, a decision was made to evaluate the effect of large ranges so that the highest-level view of the data was obtained.

The experiments carried out and the results obtained are shown in Table 2. For clarity only the 8 h in-situ yields and the final isolated S2 yields are shown. The close agreement between center-point results indicated good reproducibility of data from experimental block-to-block and gave confidence that no external factors were at work. The experiments are laid out in standard order for clarity but were obviously carried out in random or run order.

Our experimental strategy involved sampling at 2 h intervals and calculating the in-situ yield at each stage of each experiment. The first sample was taken 2 h after addition to the water of the HCl-gassed organic stream, the fourth sample was then taken after 8 h of reaction. GC was used for reaction monitoring with a sample of known purity used as reference material. In addition to the design experiments, four midpoint experiments were also run. The data from any mathematical model generated will predict the S2 values at the midpoints, and thus the actual values can be used as a check on the model. The experiments were planned so that one center-point experiment was carried out in each block of runs. This allowed any variation in uncontrolled variables to be detected, for example variation in S1 input quality, variation in operator performance, and so forth. The actual center-point experiments are highlighted in bold in the table as is the predicted center-point from the resultant DOE model.

In the statistical analysis[3] shown in Figure 1, the amount of water used (B) was most important, the interaction effect between temperature and water charge (AB) was next, and the temperature of the reaction mixture (A) was fifth. Factor A was picked despite being fifth in overall size of effect due to the fact that AB was significant (principle of hierarchy).[2] Interestingly, the agitation rate appeared to be unimportant within the experimental region probed, despite the fact that a lab agitation rate of 150 rpm was very slight. This presumably indicated that the solubility of imidate in water was very high and therefore that mass transfer was not reaction rate-limiting. The water experiments above showed the negative impact of this factor at low levels.

Use of A, B, and AB in a multiple regression model led to a high correlation coefficient (R^2, of 81%) indicating the presence of a strong relationship between these factors and the S2 in-situ yield. Analysis of variance[4] testing (ANOVA) shows that the effects are significant and that the probability of these effects being due to random variation is low. A mathematical model is displayed[5] in Figure 2.

The maximum in-situ S2 yield at 8 h corresponded to high water volume and high temperature, with water volume having a much larger effect as evidenced by the respective graph slope on each axis. The corresponding yield equation is:

$$\text{yield} = 45.74 - 0.60T + 1.51V + 0.51TV$$

where T is temperature and V is water volume. Analysis of

---

[3] Lenth, R. V. Quick and easy analysis of unreplicated factorials. *Technometrics* **1989**, *31*, 469.

[4] Draper, N. R.; Smith, H. *Applied Regression Analysis*; Wiley-Interscience: New York, 1998.

[5] We had the opportunity to try several commercially available software packages for experimental design and analysis. In this case both *Design Expert*, version 5, by Stat-Ease and *Minitab*, version 13.30, from Minitab Inc. were used.

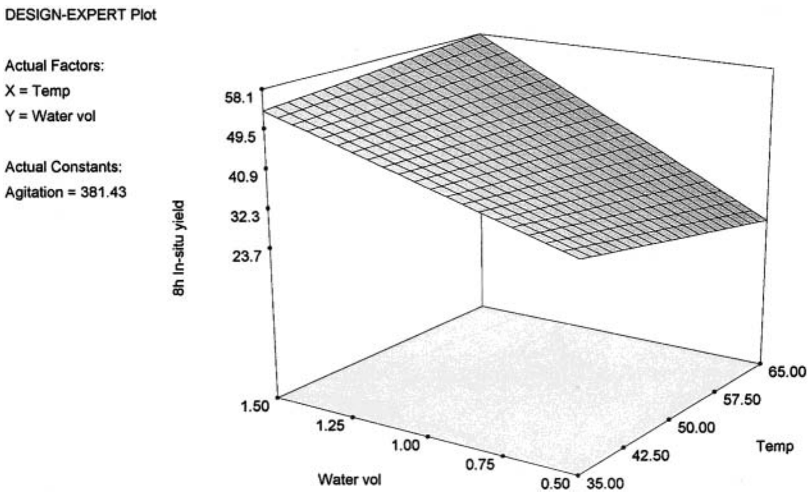


Figure 2. **Linear surface plot showing the effect of water volume and temperature on in-situ yield.**

Table 3. **In-situ data from 2, 4, 6, 8 h samples and isolated S2 yield data**

	in-situ yield				
std order	2 h	4 h	6 h	8 h	isolated S2 yield
1	24.0	30.6	34.3	36.2	37.2
2	41.6	29.4	22.1	18.2	20.5
3	56.2	53.6	51.6	50.2	47.7
4	66.0	62.5	59.6	57.5	60.3
5	18.6	25.5	29.5	32.2	31.8
6	48.4	40.1	34.0	29.1	33.4
7	56.3	57.3	57.2	56.6	62.1
8	53.6	58.6	58.3	58.6	61.4
9	55.7	55.0	53.5	52.5	55.3
10	57.9	56.8	55.3	54.5	57.4
11	55.2	55.2	53.5	53.2	57.4
12	67.6	52	52.0	52.0	57.9

the isolated S2 yield gave similar results as did analysis of the 2, 4, and 6 h samples from each run, see Table 3.

From a statistical perspective it was noteworthy that analysis of area % data gave a much different picture of the reaction. Collinearity[2] introduced by use of normalized chromatographic area data is a likely cause of this difference. Use of data from an external standard approach is more reliable, however.

The large curvature effects seen in all models evaluated underscores the complexity of the system and the need to utilize response surface methods to get a more definitive model of the data. The presence of lurking third-order interactions (ABC or triple-order interactions) is also indicative of this inherent complexity. The predicted yield value, from the model, for the center point experiments (∼42%) was significantly lower than that obtained from the corresponding runs (∼56%). This indicates that a relatively small variation of factors close to the center points might not lead to significant yield losses. This was an important result, since "common knowledge" indicated that operation at higher concentrations gave significant amounts of impurity X.

In addition to the formal DOE analysis, simple inspection of all results, (Table 3), revealed the following:

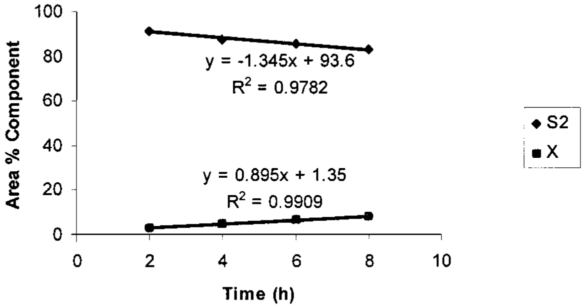


Figure 3. **Relationship between impurity X and S2 (area %).**

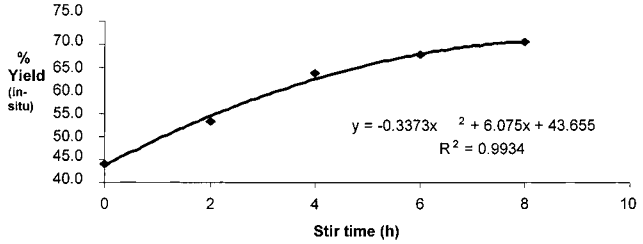


Figure 4. **In-situ yield data from plant samples.**

(i) For many runs the maximum in-situ yield was obtained at 2 h with the values falling off slightly after that time. This indicates that the reaction may be much faster than expected under control conditions also (see below).

(ii) In-situ yields are good indicators of the isolated yield indicating low workup losses.

(iii) Experiments at low water (0.5 kg water per kg batch) uniformly gave low S2 yields and high impurity X (e.g., 25% by GC area for standard order no. 6). Data analysis across all lab runs indicated that impurity X was being formed from S2 rather than from S1 as had been previously thought (Figure 3).

The complexity of the system shown by the lab DOE analysis indicated that direct correspondence between lab and plant might be unlikely, although the same trends would be expected to apply. To find out if typical plant runs ($T = 35$ °C, concentration = 1.0 kg water/kg batch, agitation ≈ 350 rpm) were complete earlier than 20 h a typical batch was sampled at 0, 2, 4, 6, and 8 h and assayed against an external standard for % in-situ yield, (Figure 4).

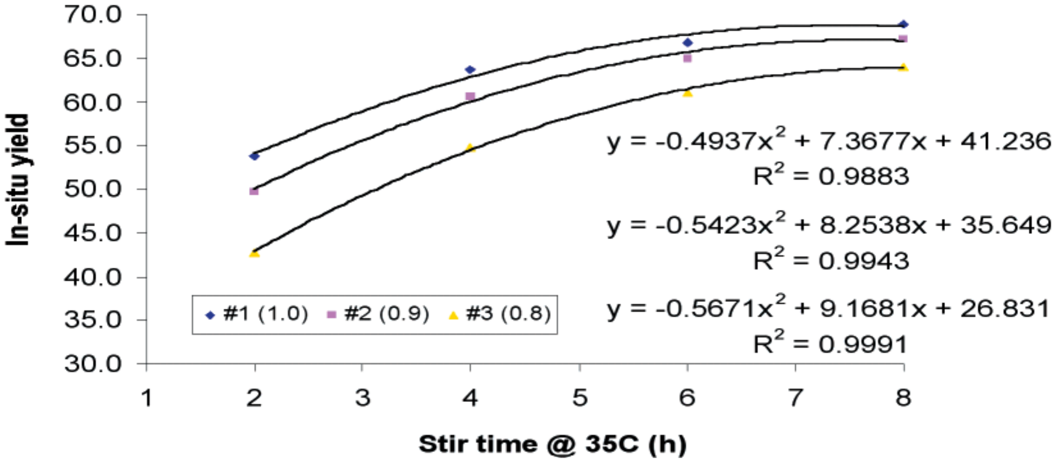


Figure 5. **Variation of in-situ yield with time at 35 °C (three concentrations).**

The polynomial equation fitted to this yield profile indicated that the yield would plateau at 10 h but then decay after that. This result was important for two reasons: (i) the maximum in-situ yield (70%) was significantly greater than then current values and (ii) a ∼50% decrease in cycle time for this portion of the reaction could be possible. To verify this result a typical plant batch (after 16 h at 35 °C) was assayed, and the in-situ yield was 65.8%. A ∼5% yield increase was therefore realizable from decreasing the quench stir time.

Examination of the DOE data also revealed that although low water levels (0.5 kg per kg batch) were very harmful for yield, the model predicted much lower yields for the center-point experiments than was actually obtained. This indicated that some decrease in water level might be possible without any major yield penalty. This potential for increased productivity was important and warranted further investigation. To this end plant samples were quenched with varying amounts of water (1.0 (current water loading), 0.9, and 0.8 kg water/kg batch) and the in-situ yields monitored at 2, 4, 6, and 8 h stir times (Figure 5).

The data showed that a 10% concentration increase was possible without a large change in in-situ yield (69% at 1.0 vs 67% at 0.9 kg water/kg batch), a 20% concentration increase had a more significant yield penalty however (64% at 0.8 kg water/kg batch). Quality results (assay and impurities) were comparable to those seen in the original process. It was noteworthy that the in-situ yield at 8 h was 69% for the experiment where 1.0 kg of water/kg batch was used. This was further evidence that shorter quench stir times should provide a yield benefit in the plant. The in-situ yield at any value of stir time can obviously be calculated from the appropriate predictive equation. The corresponding isolated yield data is shown in Table 4.

Table 4. **Variation of isolated yield with concentration (35 °C)**

exp	water amt	isol. yield S2
1	1	70.9
2	0.9	68.5
3	0.8	64.4

Another three experiments that were run at 22 °C showed that the yield convergence seen in the above plot did not occur. This implied that the yields obtained at the three concentrations might be closer at even higher temperatures, thereby facilitating a further concentration increase at little yield penalty. Towards this end a concentration study was undertaken at 45 °C. The pilots were sampled at 2, 4, 6, and 8 h with in-situ yield calculation as before. The data obtained is shown in Figure 6.

Although the yields from the three concentrations were indeed closer together, they were (a) lower on average than the corresponding 35 °C experiments and (b) had higher impurity X levels (∼3% vs the 0.8% max seen at 35 °C). This process would be more sensitive to holding also due to the yield decline seen from 5 h on. The color of the isolated liquor was noticably darker from the 45 °C experiments also. In light of these data it seemed advisable to focus on process improvements that could be achieved at 35 °C.

The earlier DOE data provided a mathematical model for yield and quality based on variation of temperature and so forth but was generally carried out under conditions that differed substantially from the then current quench process. This was deliberate, as the aim was to understand how the factors impacted yield and quality over large input ranges. The large curvature effects seen in the earlier DOE data underscored the need to utilize response surface methods to get a definitive model of the reaction. The data described immediately above was closer to current control values and so was more suitable for use in developing a model to explain the effect of factor variation under current conditions. The factors used were temperature and water amount since agitation was previously found to be unimportant over the range studied (Table 5).

Although this was not a typical response surface design, the coverage of the experimental space was easily sufficient to allow estimation of an accurate model (Figure 7). The points in red indicate the values of water loading (concentration) and temperature in the appropriate experiment, while the yellow lines indicate contours of constant yield response from the predicted model. The choice of temperature and water ranges studied reflected process and regulatory limits.

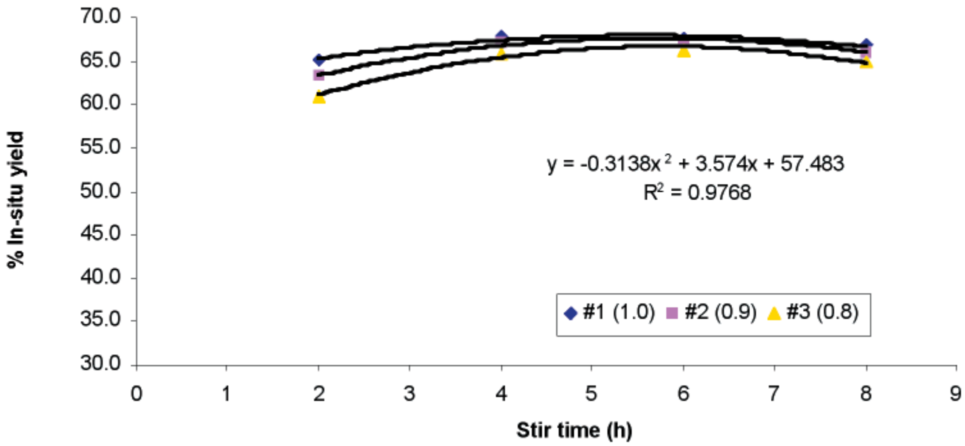


Figure 6. **Variation of in-situ yield with time at 45 °C (three concentrations).**

Table 5. **Design experiments and results (in-situ yields at 8 h)**

A: water amt (L/kg batch)	B: temp (°C)	in-situ yield (%)
0.7	22	33.7
0.8	22	37.8
0.8	35	64.0
0.8	45	64.9
0.9	22	40.3
0.9	35	67.2
0.9	45	66.1
1.0	35	68.9
1.0	45	66.8
1.0	35	68.6
1.1	35	69.3
1.2	35	68.8

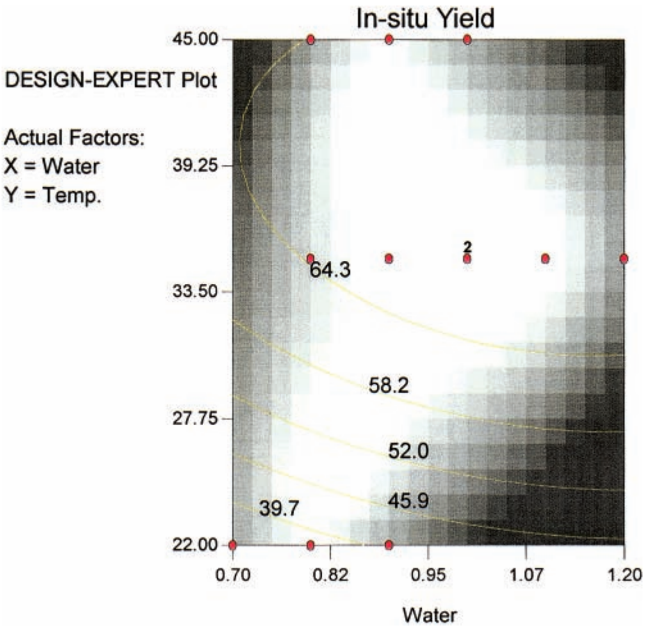


Figure 7. **Coverage of design space by the experiment set.**

Multiple regression analysis was used to determine the relationship between these factors and the S2 in-situ yield (Table 6). A variety of models were evaluated, but a model that incorporated curvature for both factors gave the best results. The yield quoted is on an in-situ basis after 8 h, as before this agreed well with isolated yields.

Table 6. **Response surface regression: 8 h yield versus water amount and temperature**[a]

term	coefficient	SE coefficient	*T*	*P*
estimated regression coefficients for 8 h yield				
constant	−145.1	6.6190	−21.919	0.000
water amount	116.0	12.5406	9.251	0.000
temperature	7.8	0.2046	38.196	0.000
water amount × water amount	−41.7	6.8832	−6.064	0.001
temperature × temperature	−0.1	0.0025	−36.514	0.000
water amount × temperature	−0.6	0.2069	−3.031	0.023

$S = 0.4456$ $R^2 = 99.9\%$ $R^2_{adj} = 99.9\%$

source	DF	Seq SS	Adj SS	Adj MS	*F*	*P*
analysis of variance for 8 h yield						
regression	5	2062.75	2062.749	412.550	2E+03	0.000
linear	2	1738.06	302.779	151.390	762.41	0.000
square	2	322.86	265.877	132.938	669.49	0.000
interaction	1	1.82	1.824	1.824	9.18	0.023
residual error	6	1.19	1.191	0.199		
lack-of-fit	5	1.15	1.146	0.229	5.10	0.324
pure error	1	0.05	0.045	0.045		
total	11	2063.94				

observation	8 h yield	fit	SE fit	residual	st resid[b]
unusual observations for 8 h yield					
3	64.000	64.710	0.304	−0.710	−2.18R

[a] The analysis was done using uncoded units. [b] R denotes an observation with a large standardized residual.

Use of A, B, A^2, B^2, and AB in the multiple regression model led to a high correlation coefficient (R^2, of 99.9%) indicating the presence of a strong relationship between these factors and the S2 in-situ yield. Analysis of variance testing (ANOVA) showed that the effects were significant and that the probability of these effects being due to random variation was extremely low. The model is displayed in Figure 8). The portions of the graph that fall to baseline indicate regions where experimental data was not available. Interrogation of the model revealed that the yield optimum was 71% at 40 °C and 1.0 kg water/kg batch. As well as showing the process optimum this model provides a detailed "map" of the

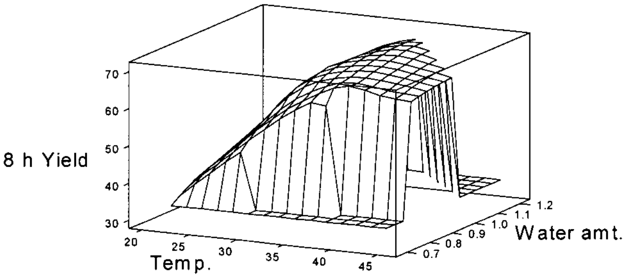


Figure 8. **3-D model of the Pinner quench reaction data.**

experimental space which has been very useful in evaluating the effects of process changes on scale. The model is capable of predicting yield results at various conditions to within ±0.5% (residuals analysis).

Conclusions

In summary the use of DOE and mathematical modeling techniques allied to rigorous experimental investigations led to an optimal quench process. A yield improvement of 5% and a cycle-time improvement of >50% were gained for this step while the quality results remained unchanged. The possibility of increasing throughput by utilizing concentration increases was demonstrated. Implementation of these changes on plant is proceeding without issue.

Acknowledgment

We thank Dr. Matt Priest for his thorough review of the manuscript.

Received for review August 20, 2001.

OP010225+