

Sensitivity Analysis of SAGD using Experimental Design and a Proxy

Vahid Dehdari and Clayton V. Deutsch

Oil Injecting hot steam into the reservoir is the most efficient method for producing heavy oil in the Alberta. For simulating this process, and predicting amount of produced and injected steam, STARS software which is very expensive can be used. Simulation running time is another concern for petroleum engineers. Just for simulating this mechanism in a well pair, running time is about 10 hours. As a result, running time of a full model is about one week. For this reason, in the department of mining and petroleum engineering, a proxy for fast and accurate simulating of this process has been developed. Running time using this proxy is less than 1 minute which is very fast. But correlation coefficient between simulator and proxy results should be reasonably high, otherwise finding an inaccurate result is not acceptable. Usually correlation coefficient of 0.9 or higher is great. This proxy can be used for realization ranking application. After Geostatistical modeling of realizations, they should be ranked based on the highest oil production and lowest steam injection for transferring uncertainty of reservoir to the simulation models. There are four different parameters in the proxy which are uncertain. All other parameters are fixed. These parameters are number of discretized cells in the horizontal direction (n), steam pressure (P), steam temperature (T) and Enthalpy multiplier (H). Two levels have been considered for each factor. 100 reservoir realizations have been generated for this sensitivity analysis. Then these realizations have been tested with simulator. After that, they have been tested by proxy using different level of parameters. For each case, realizations ranked based on their NPV values. Then ranking indices of proxy plotted against ranking results of simulator for finding correlation coefficient. Sensitivity of correlation coefficient by changing different factors measured for finding significant factors and regression models.

Introduction

Steam-assisted gravity drainage (SAGD) is an efficient method for producing heavy oil by steam-flooding. In this method, two horizontal wells should be drilled parallel to each other. Bottom well is production well and it should be drilled close to the bottom of reservoir. Top well is injection well and it should be drilled 5-10 meters above the production well. Length of each well is around 1 kilometer. Steam from top well should be injected to the reservoir to increase temperature of bitumen between wells and also above injector. This mechanism decrease bitumen viscosity and bitumen can be produced by gravity forces. Also process can be began by injecting steam into both of wells for 1-3 months in order to heated bitumen enough to start flowing to the lower wells, then injecting steam to the producers should be stopped and it should be injected only to the injector. During SAGD, cone shaped steam chamber moves to upwards during rising period and then will be moved horizontally during spreading period. Fig. 1 shows development of steam chamber by injecting steam.

In this method, amount of oil production depends highly on the efficiency of steam injection. Usually in these reservoirs there are shale barriers in different layers. If steam reaches to the shale layer, it cannot move upward any more. Also if there is a shale barrier between injector and producer, oil cannot be produced through producers. For this reason, location of producer and injector is very important in this process. Robust optimization should be done by considering uncertainty of static properties in the reservoir. Finding reservoir simulation model and trying different well trajectories for finding optimal value of oil production is not efficient. Due to the large computation time, it is almost impossible to find optimal location in a reasonable time. Even by considering uncertainty in the reservoir model, running time would be much more than using a single model. Finding a proxy model for fast modeling of SAGD process is a good option. In this paper, we used Butler theory for modeling SAGD process. Using this method, running time can be decreased significantly. Because running time of proxy is much less than flow simulation, considering uncertainty is possible and optimization can be done in a reasonable short time.

Data collection and implementation of the experiment

In this project, we want to compare SAGD (steam assisted gravity drainage) simulation results of commercial simulator with results of proxy (program has been written by FORTRAN language which is much faster than commercial simulator, but results are not as accurate as simulator).

Steam-assisted gravity drainage (SAGD) is an efficient method for producing heavy oil by steam-flooding. In this method, two horizontal wells should be drilled parallel to each other. Bottom well is production well and it should be drilled close to the bottom of reservoir. Top well is injection well and it should be drilled 5-10 meters above the production well. Steam from top well should be injected to the reservoir to increase temperature of bitumen between wells and also above injector. This mechanism decrease bitumen viscosity and bitumen can be produced by gravity forces. Fig. 1 shows development of steam chamber by injecting steam.

In order to compare results, different Geostatistical realizations should be generated. In this project, we generated 100 reservoir realizations completely randomly using Geostatistical methods. Then we run each realization with simulator and calculate oil and steam production based on the results of simulator. As a result, 100 NPV (net present value) can be obtained by considering both of oil and steam productions from these 100 realizations. Then, correlation coefficient between 100 NPV from simulator and 100 NPV from proxy can be obtained as response. But some parameters in proxy (number of cells, pressure, temperature and enthalpy) are uncertain and can be considered as factors with different levels. By changing these parameters, NPV values for proxy can be changed. As a result, correlation coefficient between simulator and proxy ranking indices will change (but simulator results will not change, because nothing is uncertain in the simulator runs). Generating realizations and assigning different realizations to different replicates were completely random. Figs 2-3 show input parameter file (just part of it) and output result of proxy for one of the replicates.

After finding proxy and simulator results, realizations can be ranked based on the highest NPV (highest oil production and lowest steam injection). Then ranking indices can be plotted against each other for finding correlation coefficient. High correlation coefficient is desirable for ranking application. Fig. 4, shows result of one of the runs, but all other results are similar to each other and all of data are around a 45 degree line passed through the origin for ensuring high correlation coefficient.

This experiment is said to be four-way layout. The objective of proposed experiment is to define main parameters of the proxy, select the most influential ones and obtain their appropriate values for desired behavior of the process.

From described experiment assembly, number of grids, steam pressure, steam temperature and enthalpy multiplier are treated as crucial setting parameters, i.e. four quantitative factors with two levels each are examined in the experiment. The factors are summarized in the Table 1. The response is correlation coefficient between simulator and proxy results.

Analysis of variance

The experiment is analyzed using a Four-way layout design with two levels on each factor. The four-way layout model considers the influence of the main factor effects, the two factor interaction effects and the three factor interaction effect.

$$y_{ijk} = \eta + \alpha_i + \beta_j + \gamma_k + \delta_t + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\delta)_{it} + (\beta\gamma)_{jk} + (\beta\delta)_{jt} + (\gamma\delta)_{kt} + (\alpha\beta\gamma\delta)_{ijkl} + \varepsilon_{ijkl}$$

$$i = 1,2 \quad j = 1,2 \quad k = 1,2 \quad t = 1,2$$

2^k factorial design

In this project, analysis using a 2^k factorial design is carried out. In this case, there are 4 factors and each of them has 2 levels. Different runs showed that effect of steam temperature on the correlation coefficient is negligible. We can prove it using analysis of variance and also 2^k factorial design. Table 2 shows response for different factor-level combinations using a 24 factorial design. In this case there are 4 factors each at 2 levels.

Also Fig.5 shows different factors and their levels. All factors are quantitative factors.

In order to derive mathematical model for correlation coefficient, the experiment is conducted with deemed the most influential 4 different factors each at 2 levels. The factors are number of cells (N), steam pressure (P), steam temperature and enthalpy multiplier (H).

Since each factor has 2 levels and all factor combinations have corresponding measurements (no aliases of main effects), 2^k full factorial design analysis is applied to the experiment. Recall that we are dealing with the nominal-the-best problem with target value y_t of correlation coefficient 0.85. Also we can treat this problem as a higher the better by maximizing location and minimizing dispersion. In other words there is need of minimization of quadratic loss function $L(y, y_t)$. So the two-step procedure is used to meet the target value with minimum dispersion.

$$E[L(y, y_t)] = c.VAR(y) + c.(E[y] - y_t)^2 \quad (1)$$

where the constant c is determined by any considerations.

Analysis for 2^k full factorial design

In this section experimental design analysis is constructed. Plots of main effects and factor interactions are drawn, and some preliminary results are obtained.

To get extended ANOVA table the data from Table 3 is used.

Based on the above table, factorial and interaction effects can be calculated easily. Also for finding significant effects, half normal plot can be used. In below, you can see box plot for different levels of each factor. In can also shows the main effect of each factor. It seems that steam temperature did not change significantly and by changing its level, it has constant response.

In Fig. 6, the bounds of each box are 25% and 75% quantiles, the whiskers are the extremes, the line in the box is the median median. As you can see, in factor A, changing level has significant effect on changing the median. Probably this factor is a significant factor. Also Variation within box plots of two levels in factor D is not significant. As a results, probably this factor is not significant, but for better understanding of significant effects, half normal plot should be shown. Before showing half normal plots, in Fig.7 you can see interaction plots.

It seems that except the interaction between AB and AC, other interactions are not significant and these interaction effects do not have significant effect on the model. In these cases, both of lines are parallel to each other and changing one factor does not have significant effect on changing other factor. But as we talked before, we can find all significant factors or interaction effects, based on statistic tests or half normal plot. For this reason, in the Fig. 8, you can see half normal plot for location effect. In half normal plot, the estimated effect of an unimportant factor will typically be on or close to a near-zero line, while the estimated effect of an important factor will typically be displaced well off the line.

From this plot, we can find that factor A (number of cells), factor B (steam pressure), factor C (enthalpy) and also interaction effects AB and AC are significant.

For making sure these effects are significant, ANOVA table can be found to see which factors can reject null hypothesis. Table 4 shows ANOVA table for regression model with above significant factors:

As you can see all factors A, B, C, AB and AC are significant with $\alpha=0.05$ because their probability is very small and null hypothesis has been rejected In null hypothesis, first we assume all factors are insignificant, then probability for this assumption should be computed and if it is a small value, we can say that null hypothesis has been rejected and factor is significant.

Assumption for the regression model that we used is based on the normality of residuals. For these reason, residual plot should be shown to check this assumption. We checked normality assumption before finding the regression models, because if they are not normal, we should transform them for making them normal. But in this case we showed results after doing analysis. Fig.9 shows plots of residual vs. different levels of each significant factor.

If the model is correct and if the assumptions are satisfied, the residuals should be structureless. As you can see in this case, residuals are structureless. The plot also shows that there is more variability in the high level of factor A (number of cells), but for the factors B (steam pressure) and factor C (enthalpy) variability is more in the lower level.

Also Fig.10 shows normal QQ plot for checking normality assumption.

As you can see mid-data are close to 45 degree line and also distribution is S-shape and it means that normality assumptions are correct.

After finding significant factors for both location and dispersion, we can find model for each of them by using regression model. Final equation in terms of coded factors are:

$$\hat{y} = 0.865 - 0.04x_{A_i} - 0.01x_{B_i} + 0.018x_{C_i} + 0.005x_{A_iB_i} + 0.017x_{A_iC_i}$$

Response surface methodology

Factor D is not included into the response surface methodology, since it has been identified as insignificant for response location. For considering nonlinear effect and plotting response surfaces, 3 levels of factors A, B and C has been considered. This is 3^k design and k is number of factors which is 3. The central composite design is used for analysis of the data, because it is one method of response surface methodology. Fig. 11 shows response surface of each 2 factors for finding the optimal level of each factor.

Conclusions

- The number of cells, steam pressure and enthalpy were the most significant of all the factors.
- Significant interaction effects are between the number of cells and pressure and also the number of cells and enthalpy.
- Maximum correlation coefficient can be found by selecting number of cells and steam pressure at lower level, and enthalpy at higher level. Using this combination, correlation coefficient of about 0.9 can be obtained.
- The model assumptions are validated by checking the residuals.
- From residual plots we can find that there is more variability in + level of number of cells and - level of pressure and enthalpy.

References

- Wu, C.F.J, Hamada, M., Experiments: Planning, Analysis, and Parameter Design Optimization, Wiley, 2000, pages 96 - 204.
- Douglas C. Montgomery, Design and Analysis of Experiments, Fifth Edition, John Wiley & Sons, Inc.

Table 1: Parameter settings

Treatment effect	Level 1	Level 2
Number of discretized cells (n)	7	15
Steam pressure (kpa)	3000	4000
Steam temperature (c)	150	200
Enthalpy multiplier (H)	0.97	1.05
Response		
Correlation coefficient between simulator and proxy results		

Table 2: Response for different factor-level combinations

n	Pressure (P)	Enthalpy (H)	Temperature (T)	Correlation coefficient
5	3000	0.97	150	0.894
			200	0.894
		1.05	150	0.894
			200	0.893
	4000	0.97	150	0.877
			200	0.877
		1.05	150	0.88
			200	0.878
15	3000	0.97	150	0.832
			200	0.828
		1.05	150	0.866
			200	0.865
	4000	0.97	150	0.823
			200	0.825
		1.05	150	0.861
			200	0.859

Table 3: Design matrix and response data with sample variance

Run	Factor levels				Responses (z)
	N	P	H	T	Mean y
1	-	-	-	-	0.894
2	-	-	-	+	0.894
3	-	-	+	-	0.894
4	-	-	+	+	0.893
5	-	+	-	-	0.877
6	-	+	-	+	0.877
7	-	+	+	-	0.88
8	-	+	+	+	0.878
9	+	-	-	-	0.832
10	+	-	-	+	0.828
11	+	-	+	-	0.866
12	+	-	+	+	0.865

13	+	+	-	-	0.823
14	+	+	-	+	0.825
15	+	+	+	-	0.861
16	+	+	+	+	0.859

Table 4: ANOVA table for significant location effects

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.865375	0.0003423	2527.922	< 2e-16
A	-0.0205	0.0003423	-59.884	6.72E-12
B	-0.005375	0.0003423	-15.701	2.70E-07
C	0.009125	0.0003423	26.656	4.22E-09
A:B	0.0025	0.0003423	7.303	8.36E-05
A:C	0.00875	0.0003423	25.56	5.88E-09
B:C	0.000375	0.0003423	1.095	0.305
A:B:C	-0.00025	0.0003423	-0.73	0.486

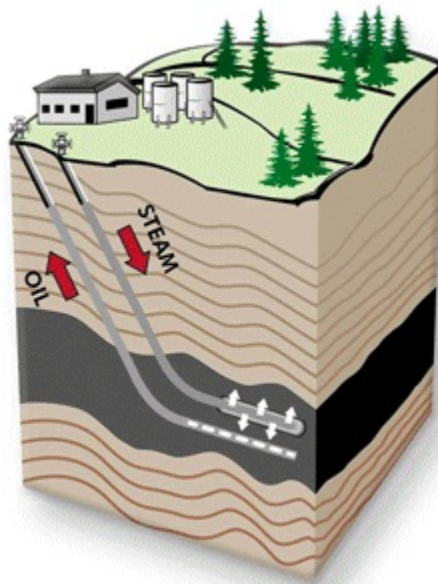


Figure 1: Development of steam chamber by injecting steam in SAGD process (source: www.treehugger.com)

```

Parameter file for FIT-FAST-RUN
*****

**
** The program searches for the keywords; the parameters are specified for
** a single wellpair and will be used for fitting the model to simulation
** results using simulated annealing
** The parameters should be specified on the same line as the keyword or
** immediately following the keyword as in the example parameter file.
**
** Distributions are entered by a distribution type and some parameters:
** 0 for constant (cte)
** 1 for uniform (xmin,xmax)
** 2 for triangular (xmin,xmod,xmax)
** 3 for normal (xmean,xst)
** 4 for lognormal (xmean,xst)
**
*****
** Information common to all data sets
**
*****
PRJNAME '3D real model'
FILEPRE $1-$2-$3-$4 ** Prefix used for output files
NSETS 1 ** Number of data sets used in the fitting process
SEED 929929 ** Randa number seed

** Time specifications for results and criteria to stop the runs
**
TIMEFREQ 30.0 ** Frequency for reporting results, days
ENDTIME 5400.0 ** Final time for producing the wells, days
MINTIME 1000.0 ** Minimum time for producing the wells, days
WCUTEND 1.0 ** Water cut criteria for stopping production, fraction

**
** Adjustment factors used in the fitting process
**
AFR 0 1.0 2.5 ** Adjusting factor for rising period
AFS 0 1.0 2.5 ** Adjusting factor for spreading period
PERMF 0 1.0 1.5 ** Adjusting factor for averaging permeability
AFCSOR 0 $3 2.5 ** Adjusting factor for csor
AFPRESS 0 1.0 2.5 ** Adjusting factor for steam chamber pressure
ANFLG 0 ** Flag, if 1 then AFR = AFS otherwise (0) they are different
    
```

Figure 2: Proxy input parameter files

Realization #	OOIP (RC)	OOIP (SC)	RF	Qo m3	Qw m3	NPV
1	376943.844	327941.156	0.522	171279.484	196547.000	75812376.000
2	360152.562	313332.719	0.472	147808.609	168498.828	65479372.000
3	385221.000	335142.281	0.594	198933.656	252894.719	86822088.000
4	368968.125	321002.281	0.635	203950.984	214633.938	91243768.000
5	392792.562	341729.531	0.665	227278.625	226956.391	102291464.000
6	376334.062	327410.656	0.533	174643.172	181652.828	78238936.000
7	359894.781	313056.250	0.524	164180.047	183551.406	72912456.000
8	376666.219	327699.625	0.658	215690.344	199173.391	97886472.000
9	344634.906	299832.375	0.404	121035.094	190953.422	50969868.000
10	365492.938	317978.875	0.519	165155.781	159506.062	74602592.000
11	373418.625	324874.188	0.302	98031.789	144651.844	41783308.000
12	397212.812	345575.125	0.520	179543.953	184574.953	80543232.000
13	392146.062	341167.062	0.710	242208.766	219272.938	110140704.000
14	372474.219	324052.562	0.416	134835.656	167494.500	59043108.000
15	376364.062	327436.750	0.587	192142.953	198351.969	86153864.000
16	331454.094	288365.062	0.388	111985.789	145719.219	48706924.000
17	408304.031	355224.531	0.633	224718.984	220817.938	101318600.000
18	356456.344	310117.000	0.518	160735.453	143381.078	73198688.000
19	369426.281	321400.844	0.613	196892.344	214437.797	87724280.000
20	354251.469	308198.781	0.582	179479.000	182637.094	80607632.000
21	350376.781	304827.781	0.415	126402.688	160746.578	55164012.000
22	419803.438	365229.000	0.693	252931.703	220352.109	115448192.000
23	372949.688	324466.219	0.593	192423.250	178006.891	87311272.000
24	375154.031	326384.000	0.571	186266.391	212642.641	82501072.000
25	341261.500	296897.500	0.435	129133.852	173137.750	55910028.000

Proxy Version: 2.000 Finished

Figure 3: Output results of proxy

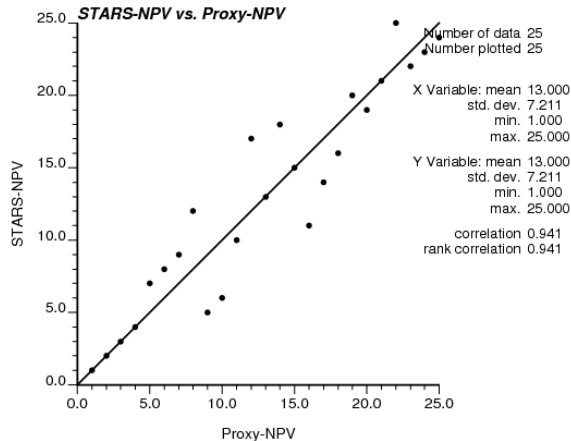


Figure 4: Rank correlation between simulator and proxy result

	n	P	T	H
-	7	3000	150	0.97
+	15	4000	200	1.05

Figure 5: Data configuration for 24 designs

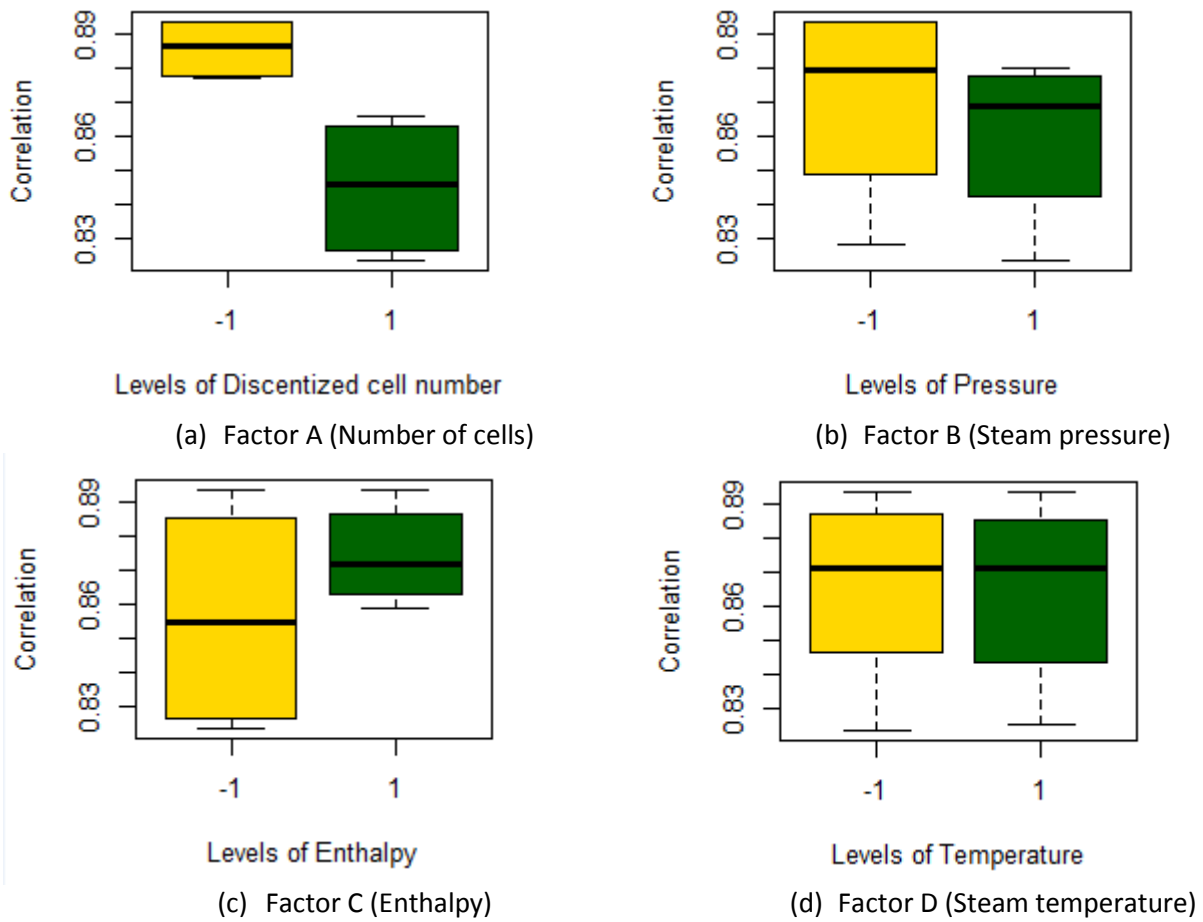
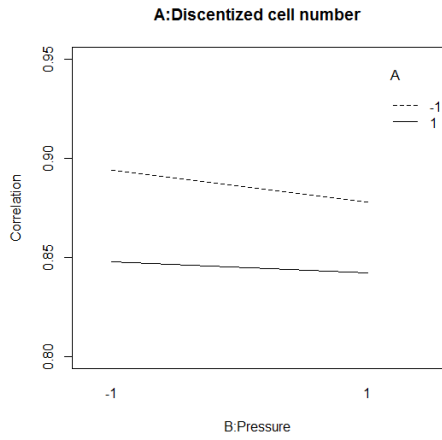
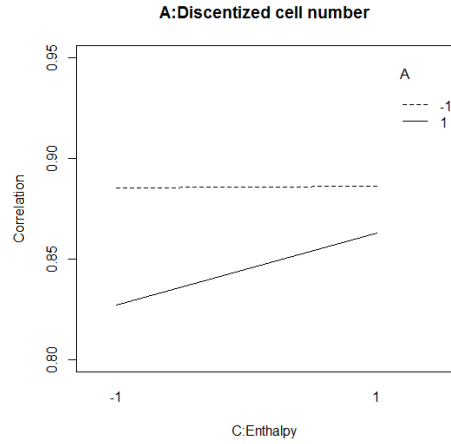


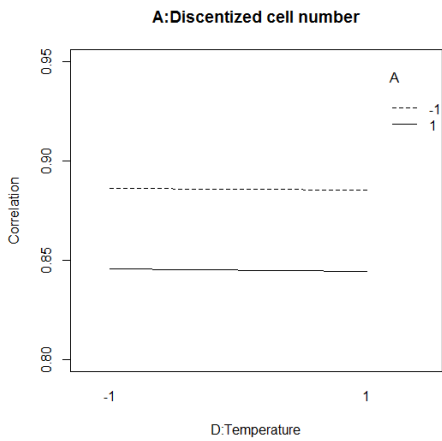
Figure 6: Box plot for different levels of each factor



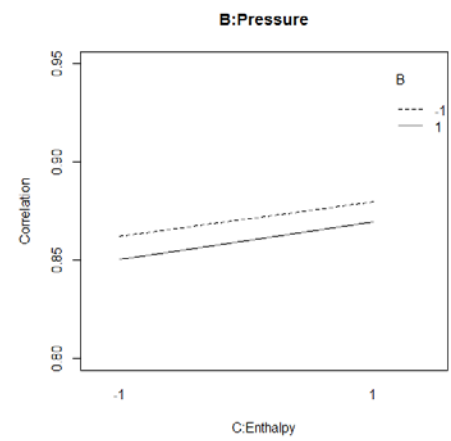
(a) Interaction between n and p



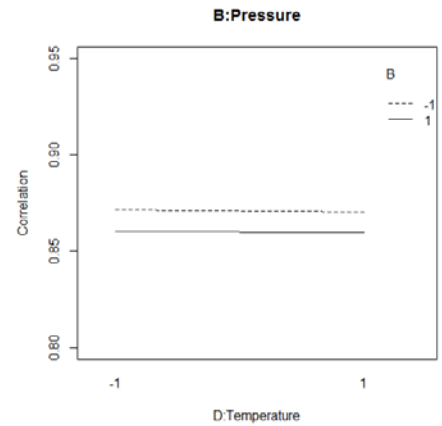
(b) Interaction between n and H



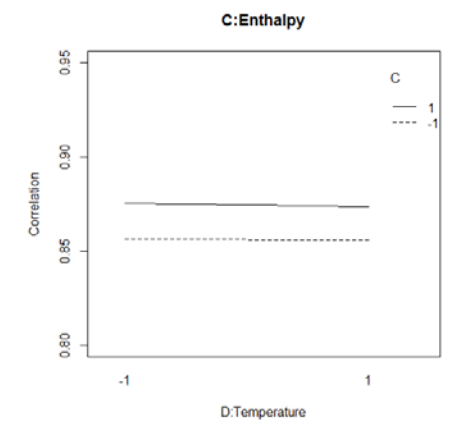
(c) Interaction between n and T



(d) Interaction between P and H



(e) Interaction between P and T



(f) Interaction between H and T

Figure 7: Interaction effects between different factors

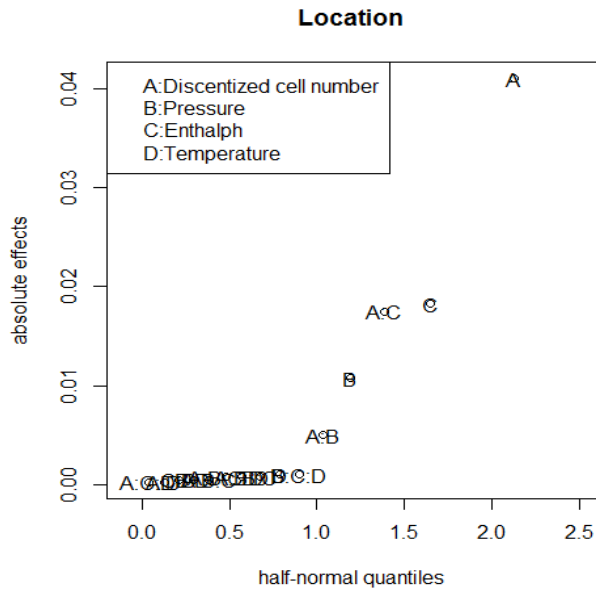


Figure 8: Half normal plot for location

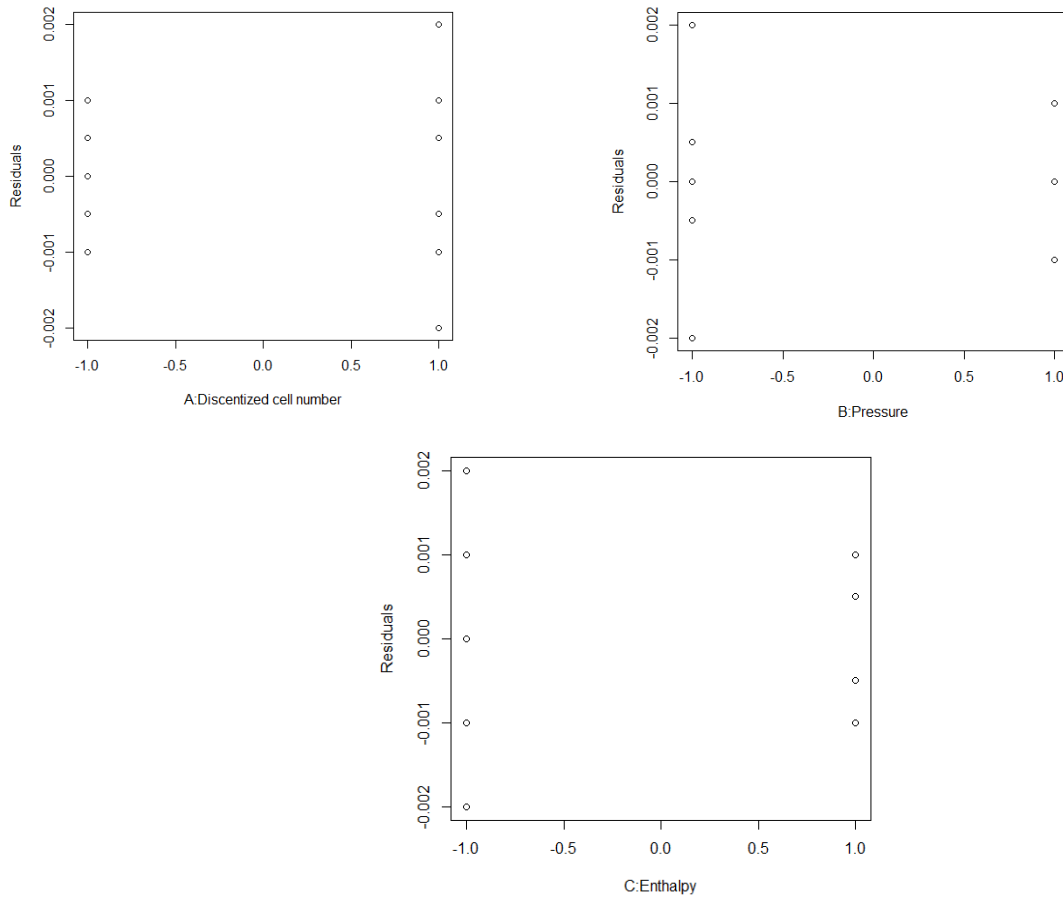


Figure 9: Residual vs. different levels of each significant factor

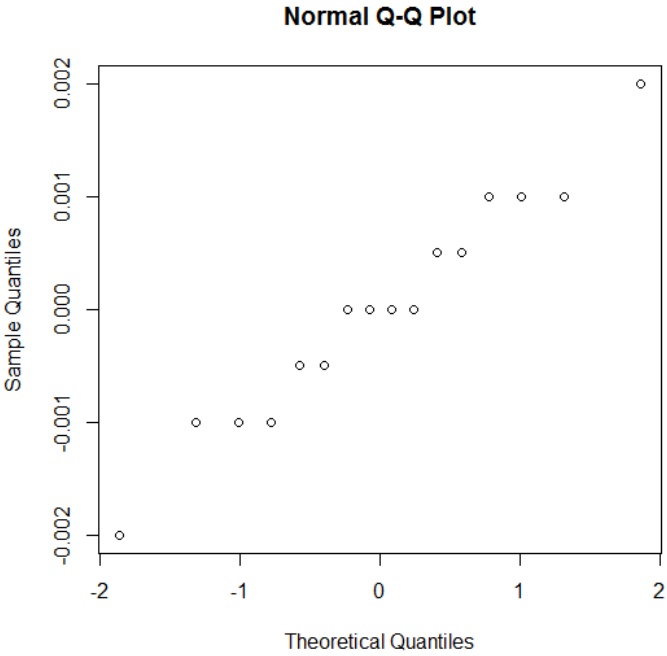
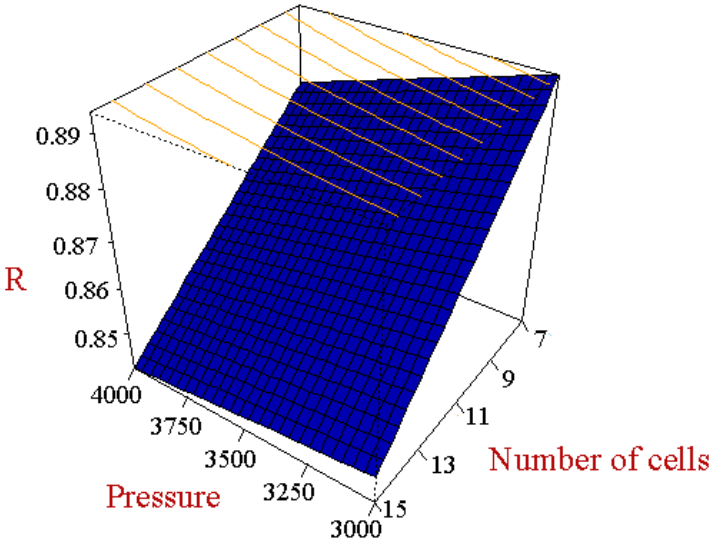
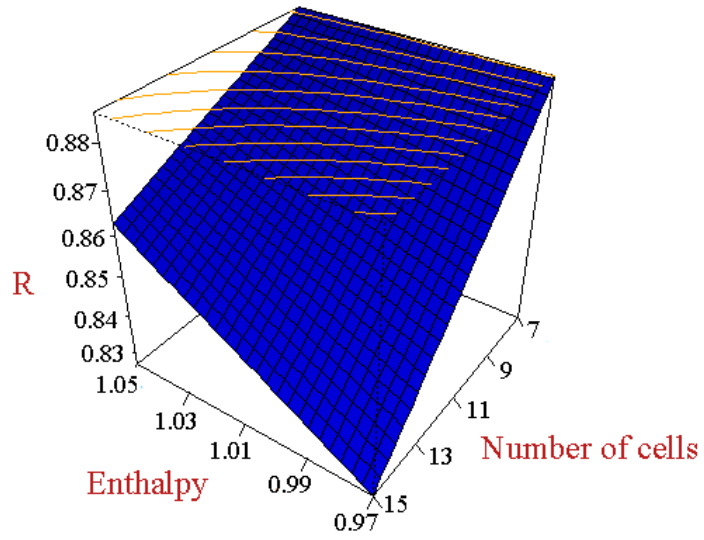


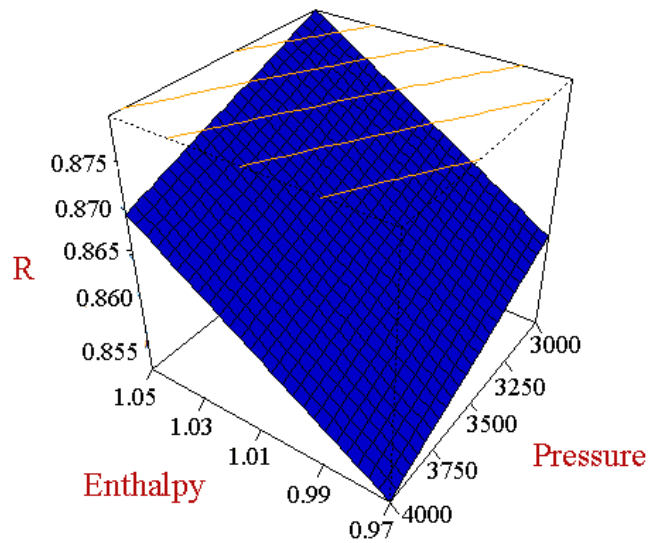
Figure 10: QQ plot for checking normality assumptions



(a) Number of cells vs. Pressure



(b) Number of cells vs. Enthalpy



(c) Pressure vs. Enthalpy

Figure 11: Response surface of different factors