The Quantification of Uncertainties in Production Prediction Using Integrated Statistical and Neural Network Approaches: An Iranian Gas Field Case Study

A. Abdollahzadeh[i], M. Hosseini, Gh. Zargar[ii]

ABSTRACT

Uncertainty in production prediction has been subject to numerous investigations. Geological and reservoir engineering data comprise a huge number of data entries to the simulation models. Thus, uncertainty of these data can largely affect the reliability of the simulation model. Due to these reasons, it is worthy to present the desired quantity with a probability distribution instead of a single sharp value.

For the case-study, numbers of parameters which are believed to contribute largely the uncertainty of Field Gas Production Total are recognized. A sensitivity analysis was done to find the most significant initial parameters. Screening experiments are designed in order to recognize the main factors and the significant interactions of factors that we need to certainly include in the response function. Later, experiments of response surface are designed objective to model the response surface function of Field Gas Production Total. This has been done based on applying two methods, Response Surface Methodology and Artificial Neural Networks. The probability distribution of Total Field Gas Production was then plotted using Monte Carlo simulation.

KEYWORDS

Reservoir, Simulation, Uncertainty, Gas, Sensitivity

1. INTRODUCTION

Based on the field owner reports which are marked as confidential, the filed is a gas field with 20 km length and 6 to 10 km width. It was discovered in 1988. Gas in place and recoverable gas were estimated to be about 1.47 and 1.19 TSCF, respectively. The condensate gas ratio in this field is about 9.77 BBL/MMCF. This field is known to be a dry gas reservoir.

The reservoir, i.e., Asmari formation, is a fractured carbonate type and the production takes place through fractures. The main production mechanism is believed to be gas expansion drive. It is not certain whether it is a combined process of aquifer and gas expansion drive.

Depth of fluid contact was petrophysically estimated to be at 5400 ft. Average total thickness for zone 2 of the Asmari formation is 170.6 ft. The maximum Net-to-Gross (*NTG*) value for zone 2 is 0.72. This uncertainty in *NTG* values is crucial through its effect on Initial Gas in Place. Maximum effective porosity for zone 2 is 17.9%. Uncertainty of this parameter is also crucial through its effect on Initial Gas in Place. Water saturation of zone 2 attains an average value of 38%.

Based on the field reports, no rock compressibility test was done in this field. This value was estimated utilizing two methods of Hall and Knaap. There was neither routine nor special core analysis done. Corey method was utilized to calculate the relative permeabilities for this field by means of a plot of Permeability vs. Porosity of a well of the field. The vertical permeability was considered to be half times the horizontal permeability. This multiple value is not certain.

There is a requirement to investigate the reliability of fracture parameters; permeability, porosity, size of matrix blocks, and fractures shape factor.

These described sources of uncertainties, along with several other un-described ones, are led to the results of simulation model highly uncertain to rely on, which brings necessity of performing an uncertainty study for production prediction in this field. The desired parameter is Total Field Gas Production (*FGPT*).

2. SENSITIVITY ANALYSIS

A sensitivity study was done to determine the main

i Corresponding Author, Asaad Abdollahzadeh is with the Center for Exploration and Production Studies and Research (CEPSR), Research Institute of Petroleum Industry (RIPI), Tehran, Iran (e-mail:abdollahzadeha@ripi.ir).

ii Ghasem Zargar is with the Department of Petroleum Engineering, Petroleum University of Technology, Tehran, Iran (e-mail: zargar@put.ac.ir).

uncertain factors among the following uncertainties choices: aquifer, net to gross ratio, porosity, rock compressibility, permeability in X direction, ratio of vertical to horizontal permeability, and fracture parameters (permeability, porosity, size of matrix block and sigma factor). The standard approach is to vary one parameter at a time, keeping all other parameters at the base-case value [2].

The results show aquifer, net to gross ratio, porosity, permeability, fracture permeability, and fracture porosity as the main uncertain parameters. For each parameter, based on the field reports, sufficient number of uncertainty levels (two or three) and their probability of occurrence are specified (Table 2).

A. Aquifer

There has been no report for water production of wells in this field and it is therefore believed that the aquifer has little effects on gas production. The main production mechanism is believed to be gas expansion drive. However, there is no firm reason to neglect the effect of aquifer on gas production. This issue has been studied under two different cases: either the reservoir is connected to the aquifer (which is the base case) or either the aquifer is not connected to the reservoir (and therefore it does not play a role in production). This factor is noted by "*Aquifer*" in the response functions.

B. Net to Gross Ratio

The amount of production depends strongly on the Initial Gas in Place; therefore the uncertainty of net to gross ratio affects the production prediction. To investigate the effect of this parameter, other than the base case, two lower and upper cases have been considered for this factor. "*NTG*" notation in response functions refer to this parameter.

C. Matrix Porosity

Porosity has been known as the second most uncertain parameter which contributes to uncertainty in production prediction by affecting the amount of Initial Gas in Place. This factor is also a three-level factor and is noted as "*PHI*" in the response functions.

D. Matrix Permeability

In this field, matrix permeability is considered to be isotropic in XY-plane. The sensitivity of production to permeability in x-direction, k_x, is investigated. Its downside, base and upside values are specified. k_y is set equal to k_x and k_z is equal to $0.5k_x$. "*KX*" represents this factor in response functions.

E. Fracture Parameters

In a fracture reservoir, the production takes place through the fractures; the fracture parameters are therefore quite noticeable to be considered as an uncertain parameter. The model showed to be sensitive to fracture permeability and porosity. These parameters are also considered as three-level factors. "*KF*" and "*PHI_FRAC*" express fracture permeability and fracture porosity in response functions, respectively.

TABLE 1: PARAMETERS ABBREVIATIONS.

Abbreviation	Description
FGPT	Total field gas production
NTG	Net to gross
AQU	Aquifer
PHI	Matrix porosity
KX	Matrix permeability in X direction
KF	Fracture permeability
PHI_FRAC	Fracture porosity
RSM	Response Surface Methodlogy
ANN	Artificial Neural Network
Sigma	Fracture shape factor

TABLE 2: MAIN UNCERTAIN PARAMETERS ALONG WITH THEIR PROBABILITY OF OCCURRENCE

Main Uncertainties	Cases	Probability of Occurrence	*Aquifer*	*NTG*	*PHI*	*KF*	*KX*	*PHI_FRAC*
Aquifer	Base-Case	0.6	Y	0	0	0	0	0
	NO_AQU	0.4	N	0	0	0	0	0
NTG	NTG_LOW	0.2	Y	-1	0	0	0	0
	Base-Case	0.6	Y	0	0	0	0	0
	NTG_UP	0.2	Y	1	0	0	0	0
PHI	IGP_PHI_LOW	0.2	Y	0	-1	0	0	0
	Base-Case	0.6	Y	0	0	0	0	0
	IGP_PHI_UP	0.2	Y	0	1	0	0	0
KF	KF_LOW	0.2	Y	0	0	-1	0	0
	Base-Case	0.6	Y	0	0	0	0	0
	KF_UP	0.2	Y	0	0	1	0	0
KX	KX_LOW	0.2	Y	0	0	0	-1	0
	Base-Case	0.6	Y	0	0	0	0	0
	KX_UP	0.2	Y	0	0	0	1	0
PHI_FRAC	PHI_FRAC_LOW	0.2	Y	0	0	0	0	-1
	Base-Case	0.6	Y	0	0	0	0	0
	PHI_FRAC_UP	0.2	Y	0	0	0	0	1

3. Experimental Design

Experimental Design is a well-known technique to maximize the information obtained from a set of experiments [3]. In this work, the experimental design was applied for two purposes; screening and modeling.

A. Screening Design

We refer to a design as a screening design if its primary purpose is to identify significant main effects, rather than interaction effects, the latter being assumed an order of magnitude less important [8]. For this purpose, a resolution VI fractional factorial experiment with 32 runs was designed. This response function was fit to data with *R2 adjusted* of 0.995:

$$FGPT = b_0 + b_1{*}NTG + b_2{*}PHI + b_3{*}KF + b_4{*}KX + b_5{*}PHI\_FRAC + b_6{*}Aquifer + b_7{*}NTG{*}PHI + b_8{*}NTG{*}KF + b_9{*}PHI{*}KF + b_{10}{*}PHI{*}Aquifer + b_{11}{*}NTG{*}Aquifer \quad (1)$$

Table 3 shows the coefficients (b_0 to b_{11}) and their significance for the above function. According to this table, *KX* is the least significant factor followed by *PHI_FRAC*. Response surface functions will be fitted based on this design, i.e., it has to include all the factors and considerable interactions, i.e., *NTG*PHI*, *NTG*KF*, *PHI*KF*, *PHI*Aquifer* and *NTG*Aquifer*, in response surface models.

TABLE 3: TABLE OF COEFFICIENTS AND THEIR SIGNIFICANCE FOR THE RESOLUTION VI FRACTIONAL FACTORIAL DESIGN.

Coefficients	Value of Coefficients	*P-value*
b_0	1038.4	2.49467E-36
b_1	209.66	1.82972E-22
b_2	209.34	1.8847E-22
b_3	95.66	9.00421E-16
b_4	4.656	0.281
b_5	11.66	0.01171
b_6	61.09	4.27496E-12
b_7	38.22	1.52091E-08
b_8	23.66	1.64651E-05
b_9	23.59	1.70225E-05
b_{10}	-17.47	0.000487
b_{11}	-16.66	0.000766

B. Modeling Designs

A response surface expresses a response variable as an empirical function of one or more quantitative factors. A general form of this type of response function is

$y = f(x_1, x_2, \ldots, x_k)$

Where y is the response and x_1, x_2, …, x_k are quantitative levels of the factors of interest [10].

Four response surface functions based on four different designs were modeled; Full Factorial, Box-Behnken, Central Composite and D-optimal designs.

I) Full Factorial Design

A Full Factorial design is a design with all possible high/low combinations of all the input factors. For this design $3^5{\times}2$ runs is needed [8]. This design will generate the population sample. Following model with *R^2 adjusted* of *0.995* is fit to the model:

$$FGPT = 1135.3 + 219.11{*}PHI + 218.94{*}NTG + 95.58{*}KF + 69.09{*}Aquifer - 72.98{*}KF{*}KF + 39.25{*}NTG{*}PHI + 23.40{*}PHI{*}KF + 22.44{*}NTG{*}KF - 15.74{*}PHI{*}Aquifer - 14.77{*}NTG{*}Aquifer + 12.56{*}PHI\_FRAC + 10.81{*}KF{*}Aquifer - 11.99{*}NTG{*}NTG - 10.58{*}PHI{*}PHI + 4.917{*}KX{*}Aquifer + 3.620{*}KX \quad (2)$$

It is expected that this model to be the best representative of the population sample and therefore it has been selected as the basis for comparing following models to.

II) Box-Behnken Design

Box-Behnken design is formed by combining 2k factorials with incomplete block designs [10]. This design needs 46×2 number of runs [10]. A model with *R2 adjusted* of 0.997 fits the data with this response function:

$$FGPT = 1131.8 + 232.22{*}NTG + 232.13{*}PHI + 75.90{*}Aquifer + 95.81{*}KF - 71.77{*}KF{*}KF + 42.13{*}NTG{*}PHI + 13.00{*}KF{*}Aquifer + 12.59{*}PHI\_FRAC - 12.06{*}PHI{*}Aquifer + 24.13{*}PHI{*}KF - 11.59{*}NTG{*}Aquifer + 23.00{*}NTG{*}KF - 6.312{*}NTG{*}NTG - 6.104{*}PHI{*}PHI + 3.625{*}KX{*}Aquifer + 3.500{*}KX \quad (3)$$

Based on this response surface function, utilizing Monte Carlo simulation for 1000 samples, probability distribution of *FGPT* is sketched and compared to that of Full Factorial as the population space (Figure 1).

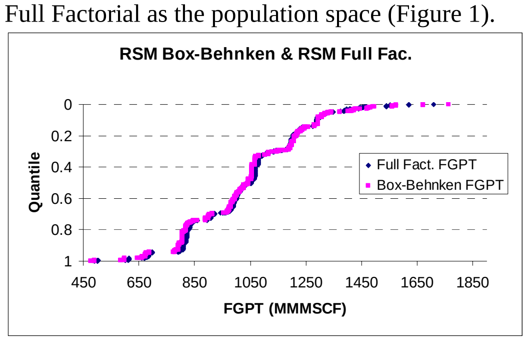

Figure 1: Comparison of RSM Full Factorial *FGPT* with RSM Box-Behnken *FGPT*.

III) Central Composite Design

A Box-Wilson Central Composite Design contains an imbedded Factorial or Fractional Factorial design with center points that is augmented with a group of 'star points' that allow us to estimate the curvature [8]. A total number of 2× (25+2×5+8) runs is needed for this design [10]. The following response function is fitting the data with *R2 adjusted* of 0.995:

$$\mathrm{FGPT} = 1130.8 + 211.44*\mathrm{NTG} + 211.09*\mathrm{PHI} + 95.74*\mathrm{KF} + 68.12*\mathrm{Aquifer} - 91.28*\mathrm{KF}*\mathrm{KF} + 38.05*\mathrm{NTG}*\mathrm{PHI} + 23.77*\mathrm{PHI}*\mathrm{KF} + 23.73*\mathrm{NTG}*\mathrm{KF} - 16.97*\mathrm{PHI}*\mathrm{Aquifer} - 16.38*\mathrm{NTG}*\mathrm{Aquifer} + 12.00*\mathrm{PHI\_FRAC} + 6.941*\mathrm{KF}*\mathrm{Aquifer} + 5.088*\mathrm{KX}*\mathrm{Aquifer} + 4.441*\mathrm{KX} \tag{4}$$

Based on this response surface function, Monte Carlo simulation was utilized to plot the *FGPT* probability distribution (Figure 2).

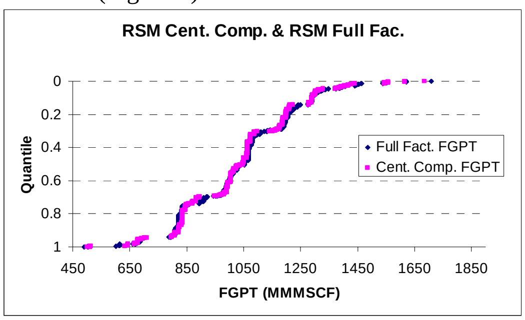

Figure 2: Comparison of RSM Full Factorial *FGPT* with RSM Central Composite *FGPT*.

IV) D-optimal Design

D-optimal design is one form of design provided by a computer algorithm. The optimality criterion used in generating D-optimal deigns is maximizing |XTX|, the determinant of the information matrix XTX. This design need 67 runs; least number of runs is the main advantage of this design [10].

The response surface function obtained from the D-optimal design is:

$$\mathrm{FGPT} = 1131.3 + 221.24*\mathrm{NTG} + 221.21*\mathrm{PHI} + 104.00*\mathrm{KF} + 4.320*\mathrm{KX} + 14.45*\mathrm{PHI\_FRAC} + 72.85*\mathrm{Aquifer} - 69.31*\mathrm{KF}*\mathrm{KF} + 43.05*\mathrm{NTG}*\mathrm{PHI} + 30.70*\mathrm{NTG}*\mathrm{KF} + 18.25*\mathrm{KF}*\mathrm{Aquifer} + 20.45*\mathrm{PHI}*\mathrm{KF} - 14.51*\mathrm{NTG}*\mathrm{Aquifer} - 16.95*\mathrm{PHI}*\mathrm{Aquifer} - 15.82*\mathrm{NTG}*\mathrm{NTG}. \tag{5}$$

The *R2 adjusted* for this response surface function is 0.988. Monte Carlo simulation was used to plot *FGPT* probability distribution based on this function. Figure 3 illustrates comparison between *FGPT* obtained by Full Factorial vs. *FGPT* obtained by D-optimal.

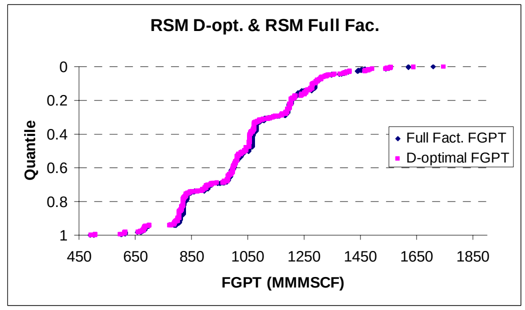

Figure 3: Comparison of RSM Full Factorial *FGPT* with RSM D-optimal *FGPT*.

4. NEURAL NETWORK

One can think of Artificial Neural Network (ANN) as an alternative to the Response Surface Methodology [7]. ANN is no magic math; the most commonly used artificial neural network are nothing more than non-linear regressions and discriminate analysis models that can be implemented with standard statistical softwares [7]. A dataset containing the variable input parameters and the corresponding output is used to train the ANN and test with a number of additional simulation runs in a further step [7]. According to each response surface design, an ANN was trained with three hidden nodes (Figure 4), and tested utilizing K-fold crossvalidation method [5]. Then For each model, the *FGPT* probability distribution was sketched and compared with the population space (Figure 5, Figure 6, Figure 7 and Figure 8).

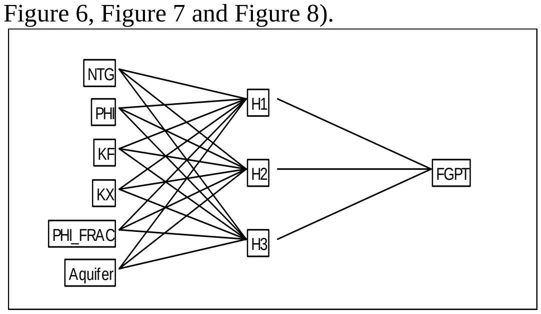

Figure 4: The Model JMP6 Utilizes to build the Neural Networks.

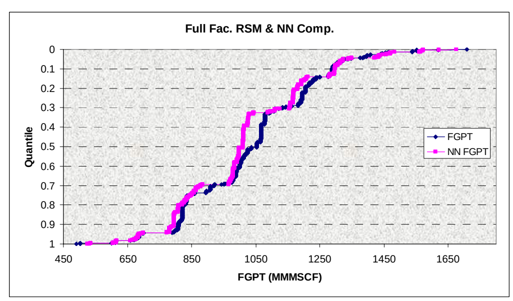

Figure 5: Comparison of RSM and ANN drawn probability distribution of the *FGPT*; basis: Full Factorial Design.

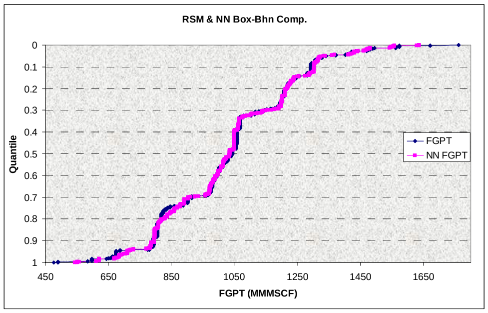

Figure 6: Comparison of RSM and ANN drawn probability distribution of the *FGPT*; basis: Box-Behnken Design.

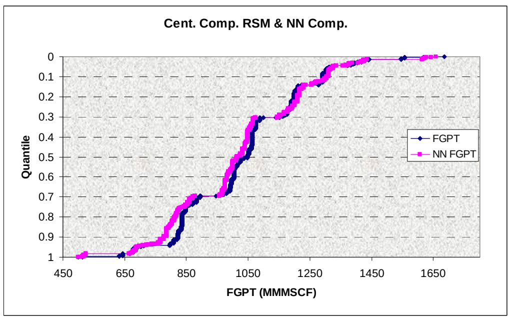

Figure 7: Comparison of RSM and ANN drawn probability distribution of the *FGPT*; basis: Central Composite Design.

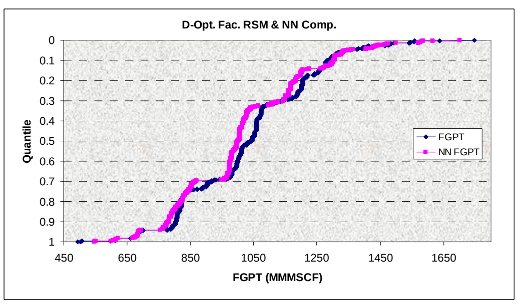

Figure 8: Comparison of RSM and ANN drawn probability distribution of the *FGPT*; basis: D-optimal Design.

5. Discussion

Comparing *FGPT* probability distribution sketched based on RSM-derived response surface function with the population sample, for each of the Box-Behnken, Central Composite and D-optimal design, illustrates close correspondence with the *FGPT* probability distribution sketched based on full Factorial design (Figure 1, Figure 2 and Figure 3;Table 4). However, the probability distribution sketched based on D-optimal design better estimates the population sample than the other two. This design also possesses the advantage of requiring least number of runs.

Comparison of *FGPT* probability distribution sketched from both Neural Network and Response Surface Methodology (Figure 5, Figure 6, Figure 7 and Figure 8) does not show worthy results. Only in case of Box-Behnken design an excellent correspondence between *FGPT* probability distributions from both RSM and ANN is observed. Not observing this in case of Full Factorial can be justified as having the ANN model overfit. Full Factorial design requires 35×2 number of runs [8]. This introduces a large number of data with large variety to the ANN model and, consequently, contaminates the ability of model to predict for non-introduced data.

For case of D-optimal design, justification is that the ANN model is underfit due to the fact that model is not trained with adequate number of data, 67 in this case. Meanwhile, it is expected that Central Composite design illustrates better correspondence between its RSM- and ANN-derived probability distribution than the other two; and Figure 7 provides good expectation.

To model the ANN based on each of Box-Behnken, Central Composite and D-optimal designs, the Box-Behnken model better represents the population space of *FGPT* (Figure 9, Figure 10 and Figure 11 and Table 5). As expressed above, this can be justified in regards to adequate number of data introduced to the ANN model.

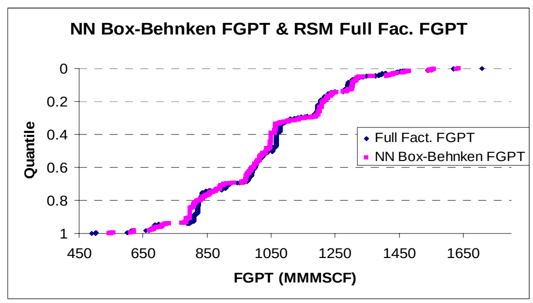

Figure 9: Comparison of RSM Full Factorial *FGPT* with NN Box-Behnken *FGPT*.

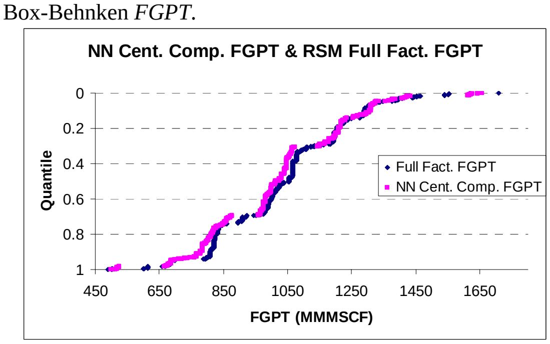

Figure 10: Comparison of RSM Full Factorial *FGPT* with NN Central Composite *FGPT*.

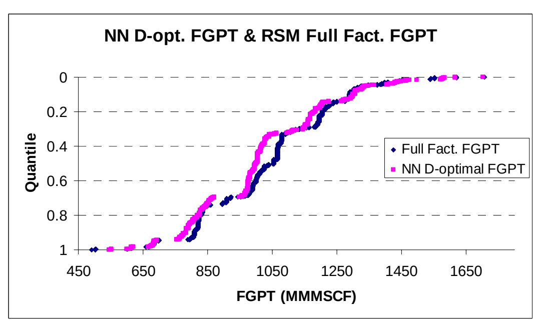

Figure 11: Comparison of RSM Full Factorial *FGPT* with NN D-optimal *FGPT*.

TABLE 4: NUMERIC COMPARISON OF RSM BOX-BEHNKEN, CENTRAL COMPOSITE AND D-OPTIMAL WITH RSM FULL FACTORIAL *FGPT*.

	RSM Box-Behnken	**RSM Cent. Comp.**	**RSM D-opt.**	**RSM Full Fact.**
Mean	1032.66587	1033.472663	1033.46226	1035.868392
P_{10}	1293.523	1290.74	1282.7	1289.227
P_{20}	1214.825	1198.92	1204.15	1204.39
P_{30}	1145.469	1150.401	1138.49	1133.233
P_{40}	1056.025	1062.68	1060.44	1067.507
P_{50}	1043.31	1048.199	1039.68	1053.65
P_{60}	994.436	1003.86	998.14	999.2
P_{70}	913.91	894.601	917.84	921.02
P_{80}	817.951	835.267	820.74	825.933
P_{90}	793.188	822.62	802.56	809.247

TABLE 5: NUMERIC COMPARISON OF NN BOX-BEHNKEN, NN CENTRAL COMPOSITE AND NN D-OPTIMAL WITH RSM FULL FACTORIAL *FGPT*.

	NN Box-Behnken	**NN Cent. Comp.**	**NN D-opt.**	**RSM Full Fact.**
Mean	1034.754208	1019.927307	1014.65898	1035.868392
P_{10}	1304.447427	1304.604939	1300.74843	1289.227
P_{20}	1216.425817	1209.217144	1184.14175	1204.39
P_{30}	1155.766981	1151.499749	1144.62911	1133.233
P_{40}	1049.426803	1048.264999	1019.31429	1067.507
P_{50}	1034.999081	1012.809431	999.290226	1053.65
P_{60}	996.0394781	982.3751629	976.233124	999.2
P_{70}	904.1340492	871.9452471	865.652627	921.02
P_{80}	826.084712	813.6972471	821.316628	825.933
P_{90}	796.5525969	777.0703089	783.785759	809.247

6. CONCLUSIONS

Numerous uncertain parameters are contributing to the simulation model of this gas field. More investigation in order to study the influence of these uncertainties in prediction of *FGPT* is needed. Based on this investigation, the following results are acquired:

1) Sensitivity analysis recognized Net to Gross Ratio, Porosity, Fracture Permeability, Fracture Porosity, Permeability (k_x) and Aquifer as the main uncertain factors.

2) A resolution VI Fractional Factorial screening design introduced *NTG*PHI*, *NTG*KF*, *PHI*KF*, *PHI*Aquifer*, and *NTG*Aquifer* as considerable interactions and that we have to account for in response surface models.

3) We recommend D-optimal design as representing the best estimation for population space utilizing Response Surface Methodology. This design also possesses the advantage of requiring least number of runs comparing to other designs.

4) A design giving the most reliable probability distribution utilizing Response Surface Methodology will not necessarily result in the most reliable probability distribution utilizing ANN. An ANN model trained by Box-Behnken design best estimates the population space.

7. ACKNOWLEDGEMENT

The authors would like to acknowledge the support and contribuations of the Research Institute of Petroleum Industry (RIPI) at Tehran, Petroleum University of Technology (PUT) and University of Calgary (UC).

8. REFERENCES

[1] Abdollahzadeh, A.: "Koma Field's Full-Field Model Update and History Matching"; IFP-school, France; Final MS Thesis in Reservoir Geoscience and Engineering; 2002.

[2] Chewaroungroaj, J.; Varela, Omar J., SPE; and Lake, Larry W. SPE, The University of Texas at Austin: "An Evaluation of Procedures to Estimate Uncertainty in Hydrocarbon Recovery Predictions", paper SPE 59449 presented at the 2000 SPE Asia Pacific Conference on Integrated Modelling for Asset Management held in Yokohama, Japan, 25-26 April 2000.

[3] Cox, D.R.; Reid, N.: "The Theory of the Design of Experiments", Chapman & Hall/CRC; 1 edition; June 6, 2000.

[4] Damsleth E., SPE; Hage, Asmund, Norsk Hydro A/S; Volden, Rolf, Norwegian Computing Center: "Maximum Information at Minimum Cost: A North Sea Field Development Study with an Experimental Design", paper SPE 23139; 1992.

[5] Dejean, J. P., IFP; and Blanc, G., SPE, IFP: "Managing Uncertainties on Production Predictions Using Integrated Statistical Methods", paper SPE 56696 presented at the 1999 SPE Annual Technical Conference and Exhibition held in Houston, Texas, 3-6 October 1999.

[6] JMP6 Software Manual"; SAS Co; 2005.

[7] Lechner, J. P., SPE; OMV AG, and Zangl., G., SPE, Schlumberger Information Solutions: "Treating Uncertainties in Reservoir Performance Prediction with Neural Networks", paper SPE presented at the SPE Europec/EAGE Annual Conference held in Madrid, Spain, 13-16 June 2005.

[8] Mason, Robert L.; Gunst, Richard F.; Hess, James L.: "Statistical Design and Analysis of Experiments with Applications to Engineering and Science"; Second Edition, A JOHN WILEY & SONS PUBLICATION; 2003.

[9] NIST/SEMATECH "e-Handbook of Statistical Methods", http://www.itl.nist.gov/div898/handbook/, 2006.

[10] "Smith, O.J.; Hendry, D.J.; and Crowther, A.R., BP Exploration: "The Quantification and Management of Uncertainty in Reserves", paper SPE 26056 presented at the Wostorn Regional Meeting held in Anchorago, Alaska, U.S.A., 26-28 May 1993.

[11] "Statistica Software Manual"; Statsoft Co; 2003.