THE PETROLEUM SOCIETY PAPER NO. 97-45

Data Integration and Analysis for Optimal Field Development

L. Fedenczuk
Gambit Consulting Ltd.

K. Hoffmann
Gambit Consulting Ltd.

THIS PAPER IS TO BE PRESENTED AT THE 46TH ANNUAL TECHNICAL MEETING OF THE PETROLEUM SOCIETY OF CIM IN BANFF, ALBERTA, CANADA, JUNE 8-11, 1997. DISCUSSION OF THIS PAPER IS INVITED AND MAY BE PRESENTED AT THE MEETING IF FILED IN WRITING WITH THE TECHNICAL PROGRAM CHAIRMAN PRIOR TO THE CONCLUSION OF THE MEETING. THIS PAPER AND ANY DISCUSSION FILED WILL BE CONSIDERED FOR PUBLICATION IN CIM JOURNALS. PUBLICATION RIGHTS ARE RESERVED. THIS IS A PRE-PRINT SUBJECT TO CORRECTION.

ABSTRACT

This paper presents an integrated approach to data interpretation during pool or field development. The methodology is based on the integration of data at different levels of resolution, multivariate statistical analysis, and advanced computer graphics.

Statistical techniques compress and organise large amounts of data into a small set of information. They aid in the identification of the most significant factors, any valid relationships, and patterns hidden in geological databases. Statistical processing estimates the degree of uncertainty surrounding exploration events, while visualization techniques provide a presentation and interpretation tool. Special attention has been given to the application of scientific visualization for presenting and analysing multidimensional data sets.

The paper explains how to convert multidimensional geological data sets into a single parameter defined as the production probability. This parameter is visualized together with the three-dimensional properties of the formation. In addition, the paper presents a method for overlaying different multivariate data sets representing non-overlapping sparse matrices.

The techniques presented in this paper improve the testing of geological hypotheses and lead to advances in the understanding of "cause and effect" relationships between formation properties, field activities, and well performance.

INTRODUCTION

The methods and techniques described in this paper are useful in identifying any valid relationships and patterns hidden in exploration and development databases. In general, geological data is multidimensional, often noisy, nonreproducible and worst of all, the majority of geological samples are mixtures of simpler components. Statistical and numerical methods can help to unmix these samples, find the original compositions of sample data, find variance and patterns, and define and test geological hypotheses [1]. These techniques help to summarise the multivariate data and relate it to geological events. In addition, they help to identify the most important factors, which are then used to develop the predictive models. These models can be quantified, and used to direct exploration and development efforts in areas with the highest possible potential. Computer graphics and specifically, scientific visualization, improve the interpretation process of the data or results of the statistical analysis [2,3,4].

Statistical analysis and scientific visualization were applied to estimate the production capacity or production probability based on available geological, petrophysical, and engineering data in a naturally fractured reservoir. Results of the analysis were used to provide answers and build hypotheses in the exploration process. Statistical models allowed for the selection of regions of interest or specific sites (wells) with a higher potential than the surrounding locations.

The following is a list of steps in the analysis:

- development of the integrated data base
- univariate analysis with the necessary data checks
- estimation of pairwise correlations
- estimation of correlations between groups (canonical correlations)
- development of mathematical models to predict production rates
- development of models which discriminated between good and poor performers
- visualization of data and models.

The next section presents the most important steps of the analysis.

DATA ANALYSIS

Data Base

The project data base contained geological, petrophysical, DST, and completion parameters. The geological and petrophysical sets contained tops, net pays, length of perforations, and petrophysical parameters representing average estimates for four zones. Reservoir pressure and initial production came from DST tests. The petrophysical parameters (e.g. resistivity, neutron porosity) were estimated using different averaging methods. Arithmetic, geometric, and harmonic means were calculated. Data transformations and the selection of variables into the final models were guided by frequency distributions and diagnostics from the multivariate procedures [5].

Parameter Selection

Stepwise regression, backward regression, and Rsquare regression procedures were used to identify the independent variables which significantly influence the outcome of the dependent variable. The stepwise discriminant procedure selected the best subset of parameters for predicting well membership into a poor/good classification.

All possible subsets of variables were tested by adding and deleting variables one at a time until a reasonable stopping point was reached. The diagnostics were used to determine whether the models adequately represented the data upon which they were based. The diagnostics indicated when more terms should be added to the model, when variables should be transformed, and how to identify outliers [6,7].

Numerical Models

The optimal set of parameters was employed to build valid statistical models which supported geological interpretations. Regression techniques were used for quantitative variables (continuous scale). The dependent variable was associated with the oil/gas production rate. In addition to the quantitative variables, predictive models incorporated qualitative variables, such as the perforation presence in the specific formation. A good regression model with six variables was developed which predicted an average oil production (r^2=0.8). The length of the perforation intervals in each of the formations had a considerable effect on the predictions. Observations with major deviations between predicted and observed values were attributed to completion problems or erroneous production data. The impact of completion techniques has been observed in other studies [8].

The major thrust of the study was to develop discriminant models. Thus, a function was developed that differentiated between good and poor wells. The discriminant function was based on the well quality and represented a linear combination of variables which supported the best discrimination between the two classes of wells in the data set. These classes (*good* and *poor* producers) were based on the initial production and the selected cut-off value. The second group (*poor*) included abandoned wells. The discriminant analysis acted in this study as a tool to overlay multidimensional information. The discriminant function was built on a base of the selected set of variables from the stepwise discriminant analysis. The verification of the function was done on a test set. The discriminant analysis was performed with two data sets: a learning and a testing set [6]. These sets were created by random selection of the observations from the initial data set. The misclassification error in the testing set was around 25%.

The final discriminant function was used to classify each of the observations into one of the groups. The final result of such analysis was the posterior probability of membership in each group. This probability was defined as the production probability, and was colour-coded on the surface of the three-dimensional structure (see Figure 1). Details of the probability visualization can be found in the next section.

Furthermore, wells which had a higher production potential than reported (e.g. abandoned wells that were reclassified into producers) became candidates for workovers. Those wells were probably damaged during the completion process.

Visualization

Advanced visualization techniques were used to interpret and integrate results from the discriminant analysis. In the visualization process, a probability (p), a scalar function of three variables in a form $p=p(x,y,z)$, can be regarded as a function which maps the fourth dimension (p) into three dimensions. Volume data can be displayed with the use of a three-dimensional analogue of the two-dimensional contour plots and pseudo colour maps in the form of isosurfaces [9], 3D points [10], 3D self-emitting voxels [11], translucent voxels [12], colour-coded density data [4], or opaque voxels [13].

Data processing involved a sequence of operations which converted the probability set into computer images. Initially, the production probability was transformed into a form suitable for visualization operations. Operations such as gridding and interpolation were applied because real data was provided on a sparse grid. Furthermore, each 3D point acquired a colour that was associated with the probability, opacity, surface properties (reflection), and volume properties (refraction). Optionally, this stage can lead to a series of realisations in the experimental space, with additional time dependence (if animation is required).

Next, the rendering process was performed on polygonized surfaces or volume cells produced from a model. During this process, the data model and mathematical description of the environment were processed by computer to determine object position, viewing direction, visibility, colour, light source, reflection, refraction, and light scattering [14]. The final images were visualized on the graphics workstation.

Our visualization of the geological structure, together with the original data and the results of the statistical analysis, is related to surface and volume mapping in three-dimensional space [9,15].

Examples of the surface mapping are presented in Figure 1 and Figure 2. Both images show the same geological structure visualized from two different observation points. The variable of interest (production probability) is mapped as colour on the surface of the uppermost formation. Formations are represented only by the surfaces of their tops, and these surfaces are represented by polygons formed on the grid pattern. In volume mapping (Figure 3), the volume occupied between two subsequent tops is modelled as an element of constant properties (colour, opacity, etc.). Each volume is mapped into a screen space by its polygonal faces.

Shading and hidden surface computations can be done using a variety of rendering engines. In both cases, rendering was performed with an A-buffer renderer [16]. Phong shading was applied in order to obtain better visual effects without the artefacts associated with discontinuities along the polygons and volume elements. Phong shading exaggerates the solid appearance of the cells, while effectively blending neighbouring cells through the normal interpolation [14].

Overlaying Probability

Boolean operations (union, complement, intersection, and difference) can be applied to posterior probabilities. If applied recursively, these operations allow us to construct expressions to add and extract regions of interest from the statistical solutions, and enhance their graphical representations. Furthermore, boolean operations can overlay different responses recorded at two different grid patterns [8].

Below, we define these operations for two sets, P and Q, which represent posterior probabilities. We assume that the probabilities are spatially defined as $p(x,y,z)$ and *q(x,y,z)* and that they are derived from two discriminant functions. The operation formulas are:

intersection

$$P \bigcap Q = p(x,y,z) \bullet q(x,y,z)$$

complement

$$\overline{P} = 1.0 - p(x,y,z)$$

difference

$$P - Q = P \bigcap \overline{Q}$$
$$= p(x,y,z) - p(x,y,z) \bullet q(x,y,z)$$

union

$$P \bigcup Q = \overline{\overline{P} \bigcap \overline{Q}}$$
$$= p(x,y,z) + q(x,y,z) - p(x,y,z) \bullet q(x,y,z)$$

The above operations do not form a complete boolean algebra set due to the fact that $P \bigcap P \neq P$ and $P \bigcap \overline{P} \neq \mathrm{O}$.

Interpretation

Colour-mapped three-dimensional structures can greatly enhance the visual interpretation of geological parameters and the results of statistical analysis. The images presented here show differences in the probability of success along with the geological structure. Spatial structures of the production probability correspond to overlays of several contour maps on the 3D light table.

The volume representation in Figure 3 shows not only the production probability along with the structure itself, but also the netpay of the formations. A review of the probability and the netpay of the images explained why the netpay was negatively correlated to well performance. It proved that the structurally low areas, and the presence of fractures, had a significant effect on well performance. At the time of the study, the data set did not include the fracture orientation or density.

One needs to apply thresholds or cut-off values of probability as criteria for the interpretation and selection of areas having higher chances of success. For example, an 80% or 90% probability of the well being a producer can serve as a criterion. A choice of a high probability value as a cut-off will cause some of the potentially producing areas not to be included in the area of interest. On the opposite side, a cut-off value below the 0.8 level will increase the risk of nonproducers. However, in the second case, one will not overlook too many potential producers.

Special attention should be given to two areas of interest which are based on two probability tails. The first tail corresponds to the highest values of the production probability, and the second tail corresponds to a very low production probability. Areas corresponding to these tails can be reviewed by interdisciplinary teams to understand the differences and build new hypotheses. Special attention in the areas of high production probability should be given to all abandoned wells.

This approach can be the most effective in reservoirs where formation damage is suspected, or where wells may have been abandoned due to the misinterpretation of the DST tests. On the other hand, the area outside of these two tails often represents the opportunity area, where advanced technology can improve the exploration odds.

CONCLUSIONS

In this study, different data sets were integrated and analysed with multivariate statistical methods. Geological structure and the production probability were visualized in the form of three-dimensional surfaces or volume cells.

The colour schema provided a way to present the results of the statistical analysis, test the correctness of the geological hypotheses, validate the analysis processes, and enhance the interpretation of results.

The methodology can help to focus activity in areas with the most chance of success, areas requiring optimisation, areas where infill wells should be drilled, or areas where significant numbers of wells do not behave as expected.

This methodology is recommended for a quick analysis of reservoirs or fields being considered for optimisation, acquisition, or development.

In addition, these techniques help to link members of interdisciplinary teams by integrating all available data sets and selecting the most important factors without any bias.

REFERENCES

1. R. Ehrlich and W.E. Full, Sorting Out Geology-Unmixing Mixtures, in Use and Abuse of Statistical Methods in Earth Sciences, Edited by W. B. Size, Oxford University Press, New York, 1987, pp. 33-46.

2. E.T. Connolly, L.L. Fedenczuk, Chemical Water Analysis - Methods of Formation Recognition, Use and Quality Control by Computer Manipulation and Mapping, Canadian Well Logging Society Journal, Vol. 15, No. 1, December 1986, pp. 21-80.

3. L.L. Fedenczuk, M. Bercov, TETRAPLOT - Four Vertices Are Better Than Three, SAS Users Group International Seventeenth Annual Conference, Honolulu, Hawaii, April 1992, pp. 534-538.

4. G.M. Nielson, T.A. Foley, B. Hamann, and D. Lane, Visualizing and Modeling Scattered Multivariate Data, IEEE Computer Graphics and Applications, Vol. 11, No.3, pp. 47-55, May 1991.

5. John C. Davis, Statistics and Data Analysis in Geology, John Wiley & Sons, New York, 1986, ISBN 0-471-08079-9.

6. SAS/STAT User's Guide, Version 6, SAS Institute Inc. Cary, NC, USA, ISBN 1-55544-376-1.

7. Mark L. Berenson, David M. Levine, Basic Business Statistics, Prentice-Hall, 1983, ISBN 0-13-057620-4.

8. R. Forth, B. Slevinsky, D. Lee, L.Fedenczuk, Application of Statistical Analysis to Optimize Reservoir Performance, Journal of Canadian Petroleum Technology, 1996, Vol. 35, No. 8, pp. 36-42.

9. G. Wyvill, C. McPheeters, B. Wyvill, Data Structures for soft Objects. The Visual Computer, Vol. 2, No. 4, 1986.

10. H.E. Cline, W.E. Lorensen, S. Ludke, C.R. Crawford, and B.C. Teeter, Two Algorithms for the Three-Dimensional Reconstruction of Tomograms, Medical Physics, Vol. 15, No.3, May/June 1988, pp. 320-327.

11. P. Sabella, A Rendering Algorithm for Visualizing 3D Scalar Fields, Computer Graphics, Vol. 22, No. 4, 1988.

12. M. Levoy, Efficient Ray Tracing of Volume Data, ACM Transactions on Graphics, Vol. 9, No. 3, July 1990, pp. 245-261.

13. L.S. Chen, G.T. Herman, R.A. Reynolds, J.K. Udupa, Surface Shading in the Cuberille Environment, IEEE Computer Graphics and Applications, 5(12), 1985.

14. D.F. Rogers and R.A. Earnshaw, Computer Graphics Techniques Theory and Practice, Springer-Verlag, 1990

15. L. Fedenczuk and B. Wyvill, Visualization of Connectivity in Octree Based Models, The Journal of Visualization and Computer Animation, Vol. 6, No.3, 1995 pp. 173-189

16. James D. Foley and Andries van Dam and Steven Feiner and John Hughes, Computer Graphics Priniciples and Practice, Addison-Wesley, 1990

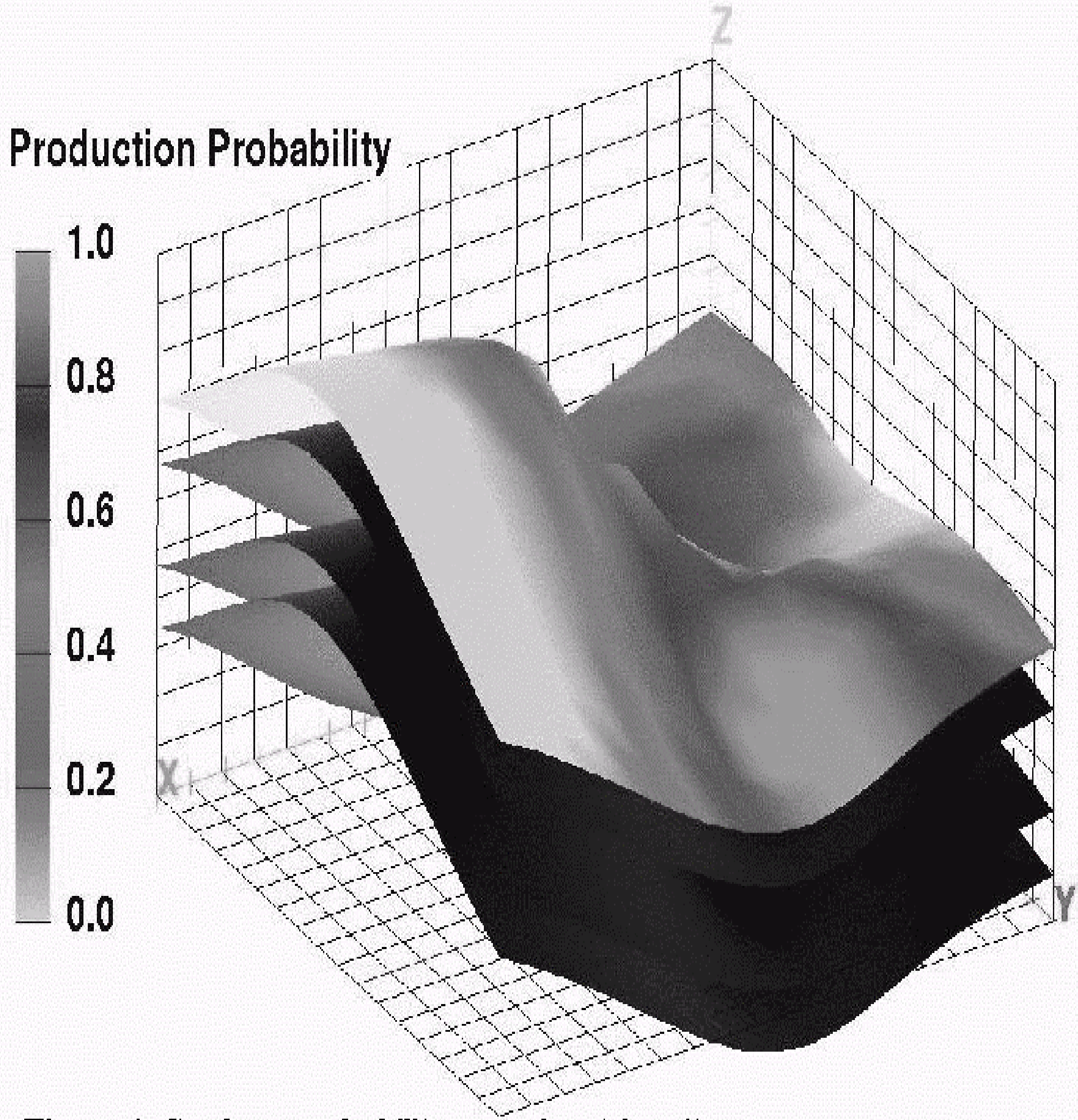

Figure 1. Surface probability mapping (view 1).

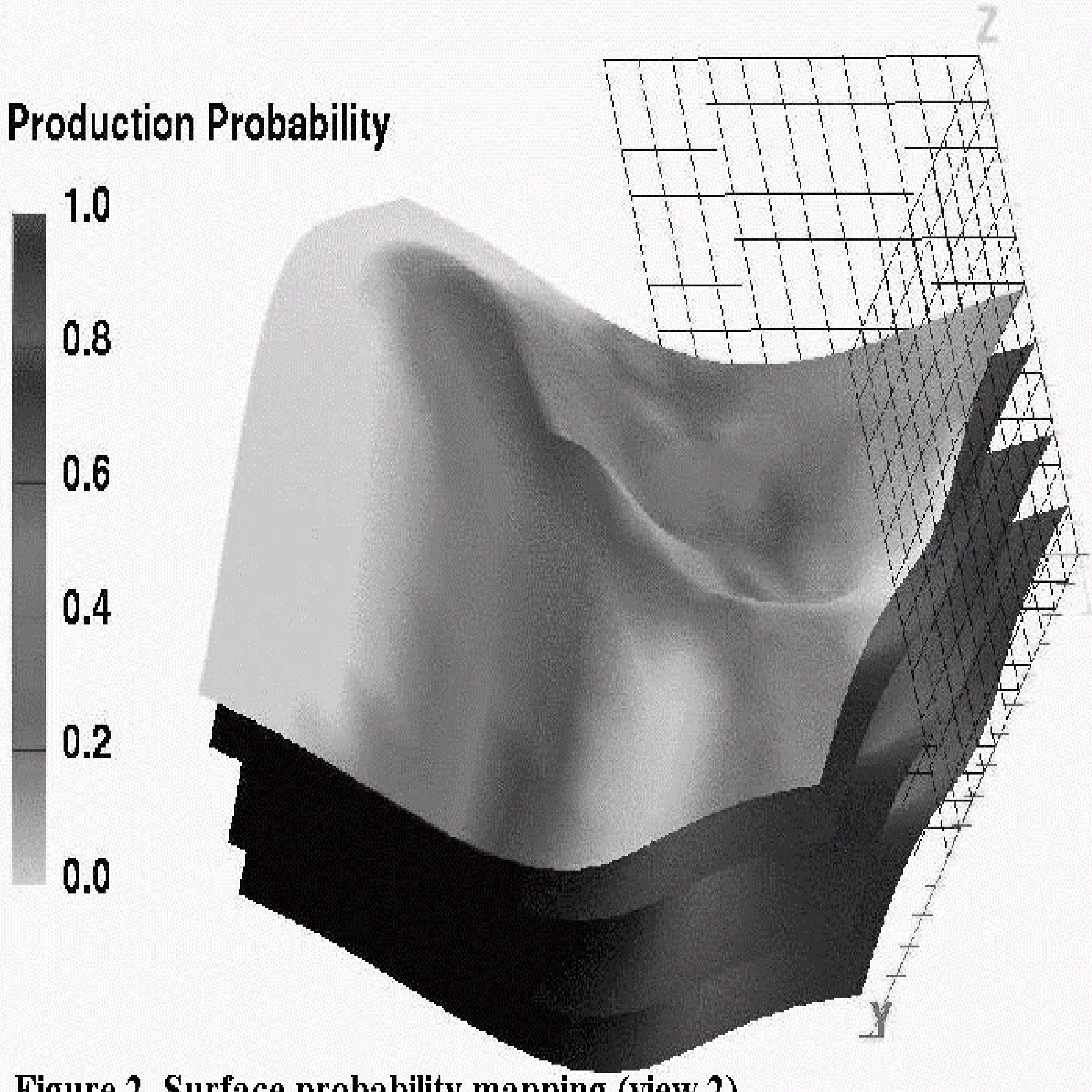

Figure 2. Surface probability mapping (view 2).

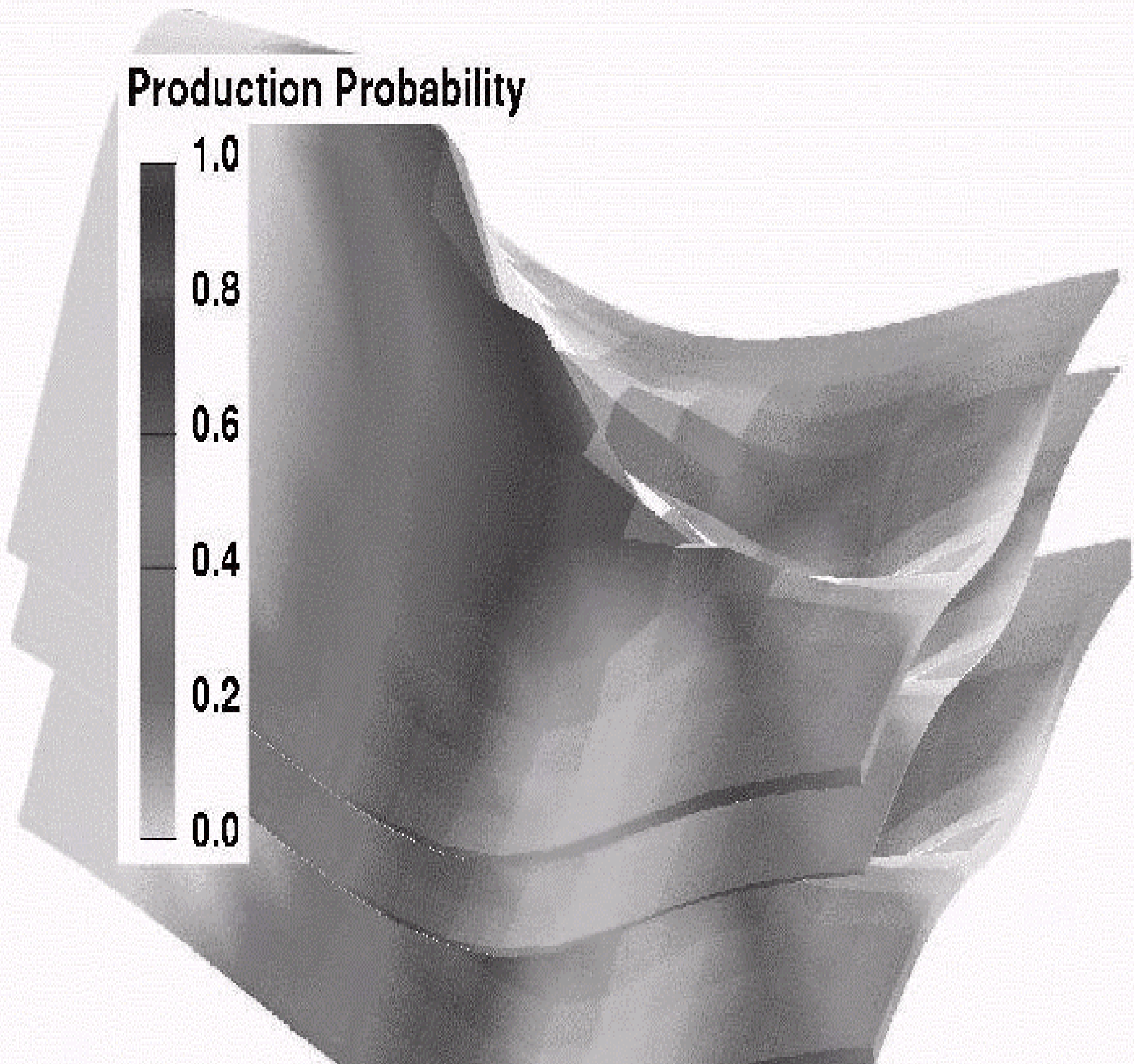

Figure 3. Volume probability mapping.