7th International Conference & Exposition on Petroleum Geophysics

"HYDERABAD 2008"

P - 200

Markov Bayes Simulation for Structural Uncertainty Estimation

Samik Sil*, Sanjay Srinivasan and Mrinal K Sen.

University of Texas at Austin, samiksil@gmail.com

Summary

Reservoir models are built using disparate datasets each of which may be prone to experimental and interpretational errors and therefore a resulting reservoir model is generally associated with uncertainties. One of the primary sources of uncertainties lies in the structure (or reservoir architecture) estimation from seismic data. Geostatistics can be used to integrate seismic data with well data for the purpose of structural uncertainty estimation. In this paper we present a case study from the Gulf of Mexico, where structural uncertainty associated with a seismic horizon is modeled using Markov-Bayes stochastic simulation. For this simulation, seismic data is used as "soft" or secondary data while well log derived marker depths are used as hard data. Simulation results show uncertainty distributions with smaller variance in the vicinity of the wells. However, in regions away from the wells, the interpreter-picked horizon appears to fall outside the error bounds predicted by our stochastic algorithm. Lack of well control, existence of faults, improper choice of seismic processing parameters (error in time migrated images) and interpreters' bias are some of the plausible causes of this disparity.

Introduction

Seismic data play a major role in the development of reservoir models since these are the only data that image the entire reservoir albeit at a low resolution. Final higher resolution reservoir models are built by integrating data of all types. Integrated data can be physical parameters (e.g. porosity, permeability, fluid saturation etc.) mainly obtained from log and well production information and seismic attributes (e.g. Vp/Vs, reflectivity, AVO gradient, two way time etc.) obtained from seismic data. Uncertainty from each data source gets propagated resulting in a reservoir model whose architecture and other properties may be uncertain. Multiple reservoir models may fit the available data equally well. This is not meant to say that the subsurface geology is stochastic but that our knowledge of subsurface derived from disparate data types are not certain. Assessment of uncertainty associated with a reservoir model can help avoid unnecessary 'surprises' during production of the reservoir and can be used for better reservoir management.

A major contribution of geostatistics in reservoir modeling is integration of data from various data sources. A recent trend in geostatistical modeling involves reservoir characterization integrating seismic and well log data (Xu et. Al., 1992, Eidsvik et al., 2001, Mukherjee et al, 2001,

Larsen et al, 2006). Reservoir models generated by integrating data using geostatistical tools help estimate uncertainty in well performance prediction and reserve estimation.

In seismic exploration, seismic data is generally exhaustive albeit at low resolution. On the other hand, well data are limited, but their accuracy is better than any other sources of information. Geostatistics tends to use well data as the primary source of information (hard data) and seismic as the secondary source (soft data). The goal is to obtain maps of the primary variable considering both hard and soft data. The mapping can be done using: 1) Interpolation algorithm and 2) Stochastic imaging or simulation (Xu et al, 1992). Spatial interpolation algorithm (e.g. Kriging) yields a unique result and provides a local measure of uncertainty. On the other hand, stochastic simulation provides multiple possible results, each honoring the same set of data constraints. Fluctuations in realizations provide a visual and quantifiable measure of uncertainty (Xu et al., 1992). Because of the need to get accurate global representation of uncertainty, the recent trend has been to adopt stochastic simulations for reservoir modeling.

Here we report on the determination of uncertainty associated with a reservoir horizon using the Markov Bayes stochastic simulation using 3-D seismic data and well control from the Gulf of Mexico. We performed our analysis over a horizon named A. The original seismic horizon A (soft data for this study) was picked from 3-D seismic data by the interpreters and may possibly be biased. Fifty eight wells are available in our study area, which are used as hard data for our study. Even though the number of wells is considerably high, their geometric locations are well separated. In this scenario we performed multiple simulations (or realizations) of the A horizon integrating both soft and hard data using Markov Bayes simulations. Localized nature of the hard data locations affects the simulation results. Hence in regions away from wells, the constraint placed on stochastic simulation is only due to the seismic. This paper describes the entire process of simulation of the horizon and uncertainty estimations and the limitations associated with the procedure.

Theory of Markov Bayes Stochastic simulation

In order to incorporate the effect of soft data (seismic data for our case) in spatial interpolation, indicator co-Kriging (IK) can be performed (Deutsch and Journel, 1998). IK can generate posterior conditional distribution (ccdf) considering soft data and other prior information.

Different prior information can come from (i) local hard indicator data $i(\mathbf{u}^a;z_1)$ originating from local hard data $z_1(\mathbf{u}^a)$, (ii) local soft indicator data $y(\mathbf{u}^{a'};z_2)$ originating from ancillary information providing prior probabilities about the value $z_2(\mathbf{u}^{a'})$ (ii) $F(z_1)$ global prior information common to all location **u** within the study area A, where (a) $Z_1(u)$ is the primary variable of interest (in our case reservoir horizon in time units) at a particular location **u** over the field of interest A (horizon A for our case), (b) $z_1(\mathbf{u}^a)$ are sparse hard data at the location $\mathbf{u}^a$, a =1,2, ….n (n=number of hard data, for our case number of well data), (c) $z_2(u^{a'})$ are the secondary data points at the location $\mathbf{u}^{a'}$, a'=1,2…n' (n'=number of secondary data, for our case seismic data). Then the indicator Kriging or Bayesian updating of the local prior cdf $F_{Z1}(z_1)$ into a posterior cdf $F_{Z1}(z_1|z_1(\mathbf{u}^a),y(\mathbf{u}^{a'}))$ can be done using the following formula:

$$\text{Prob}\{z(u) <= z \mid (n + n')\} = \lambda_0(u)F(z) + \sum_{a=1}^{n} \lambda_a(u; zl)i(u^a; z1) + \sum_{a'=1}^{n'} \upsilon_{a'}(u; z2)y(u^{a'}; z2), \quad (1)$$

where λa and υa' are the weights associated with the primary and secondary indicator data. These weights are obtained by solving the indicator kriging system:

$$\sum_{b=1}^{n} \lambda_b(u;z_1)C_I(h_{ab}; z_1) + \sum_{a'=1}^{n'} \upsilon_{a'}(u;z_2)C_{IY}(h_{aa'}; z_1, z_2) = C_I(h_{ao};z_1) \quad ,a=1,2,...,n$$

$$\sum_{a=1}^{n} \lambda_a(u;z_1)C_{IY}(h_{aa'};z_1,z_2) + \sum_{b'=1}^{n'} \upsilon_{b'}(u;z_2)C_Y(h_{a'b'};z_2) = C_{IY}(h_{a'o}; z_1, z_2), \; a' = 1,2,.., n'. \quad (2)$$

There are (n + n') equations for each set of thresholds

(z1,z2). The lag hab refers to the data pair at ua and ub. To perform the updating, we require indicator covariances of

hard dataCI (h; z1) , covariance of soft dataCY(h;z2) ,

and cross covariance between hard and soft indicator dataCIY (h; z1, z 2) (Kelkar and Perez, 2002).

If enough data (hard and soft) are available then it is possible to infer the covariances. However in reality the amounts of data (especially hard data) are always limited. To overcome this difficulty a Markov's hypothesis is used to simplify the problem. According to this hypothesis, hard information i.e. $z(u^a)$ screens any collocated soft information ($z(u^{a'})$ at the places $u^a=u^{a'}$). Under such an assumption, the required covariances can be derived as (Deutsch and Journel, 1998):

$$C_{IY}(\mathbf{h}; z) = B(z)C_I(\mathbf{h}; z)$$
$$C_Y(\mathbf{h};z) = B^2(z)C_I(\mathbf{h};z) \quad h>0. \quad (3)$$

Thus the soft autocovariance model for h>0 and cross covariance model are deduced by simple linear rescaling of the hard covariance model. Hard covariance model can be generated from the semivariogram of hard data. Then using the value of the Beez, B(z) one can determine the soft covariance model and cross covariance model. The value of B(z) can be ±1. The B(z) regulate the influence of the soft data on the estimation results. When B(z)=1, soft indicator data corresponding to the threshold z will be treated as hard data, and when B=0, soft indicator data will be ignored. Under the Markov assumption, the indicator kriging equation (1) and the corresponding system (2) simplifies since only the soft data coincident at the simulation location u is used in the interpolation.

Equations (1) and (2) together with the auxiliary conditions (3) for the auto and cross-covariances amount to a Bayesian updating of the prior distribution FZ1(z1) to the posterior conditional probability distribution employing a Markov hypothesis that facilitates the inference of the requisite auto and cross covariances.

Interpolation algorithms such as indicator co-kriging suffer from the drawback that covariance reproduction between estimated nodes does not reproduce the specified covariance. In addition, kriging yields a deterministic (unique) map that is locally accurate, but does not provide an assessment of global uncertainty. The Markov-Bayes estimation procedure can therefore be extended to a sequential stochastic simulation in which the simulation nodes are visited along a random path and at each node the Bayesian updating procedure is carried out including the simulated value at the previously visited node. Several realizations can be generated by changing the random path and the random draw. These realizations reproduce (in a ergodic sense), the specified covariance model for the hard data as well as the calibration relationship between the hard and soft information.

In general, the process of simulation can be divided into the following steps:

- The primary variable is converted into indicators defined at different cutoffs.
- Indicator variograms are calculated from all the available primary data for each cutoff to calculate primary indicator covariance.
- The secondary variable is discretized in different classes guided by the histogram of the secondary data.
- A calibration scattergram of the primary values versus the secondary values is plotted.
- For each class of the secondary variable, the scattergram values are used to calculate the probability distributions.
- The scattergram also used to calculate the coefficients, B(z). These coefficients are defined as:

$$B(z) = m_{sec}(Z > z) - m_{sec}(Z = z)$$

m_{sec} is the mean of the secondary data calculated on the basis of the scattergram. The coefficients are therefore a measure of how effective the secondary data is in discriminating between different outcomes of the primary variable.

- These coefficients and the primary variable covariances are used to calculate the covariances and cross covariances for the secondary variables approximated by the Markov hypothesis (using B values obtained from cross-plot).
- Once all the covariances are available, the posterior cdf's are calculated using the Bayesian updating formula.

Data Analysis

We apply Markov-Bayes simulation to a 3D seismic dataset from the Gulf of Mexico. The reservoir has three horizons of interest; we perform the structural uncertainty analysis on the horizon 'A'. The horizon is extracted from 3-D seismic data using a typical interpretation technique, which serves as soft data for our analysis. The entire horizon is divided into three fields (X, Y and Z) based on the geographic location. The horizon is cut by fifty eight wells (Figure 1) which serve as our hard data. Even though the number of wells is relatively large their distribution is not uniform over the entire 'A' horizon.

To prepare the scattergram and to determine B values, we generated a cross plot between well marker depths (primary data) and collocated seismic two way time data (secondary variable). The cross-plot is shown in Figure 2. A code BICALIB from GSLIB (Deutsch and Journel, 1998) is used to determine the value of B(z) (Eqn. 2) for three thresholds corresponding to X, Y and Z zones of the horizon A. Moderate B values (0.68, 0.49 and 0.54) are obtained for the three thresholds implying *moderate influence of the soft data on the final simulations.*

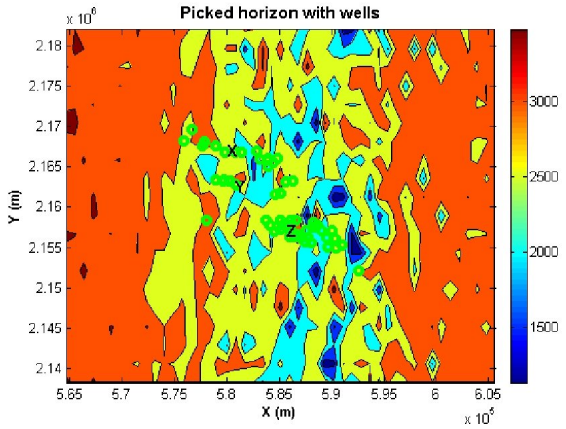

Figure 1: 2-D view of the picked TWT (mSec) of horizon A (soft data). Well locations (hard data) are plotted with green circles. The regions X, Y and Z shown in the figure where different threshold values are chosen.

For the final simulation, indicator covariances are determined from the variograms of the hard data corresponding to the three thresholds (Figure 3). The

B(z) values are then used to generate soft indicator covariances and cross-covariances.

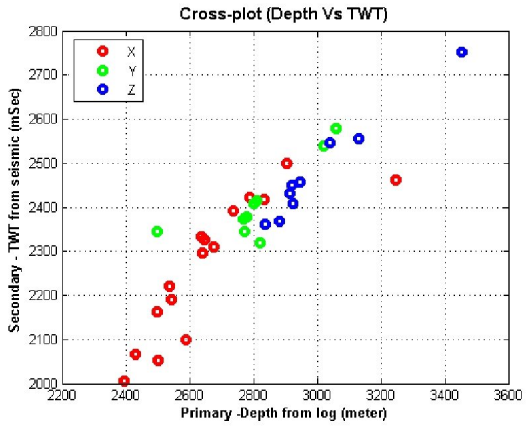

Figure 2: The cross plot between primary and secondary variables. Thresholds are determined from the mean value of hard and soft data for each zone (X, Y and Z). These data are used to generate B values for simulation.

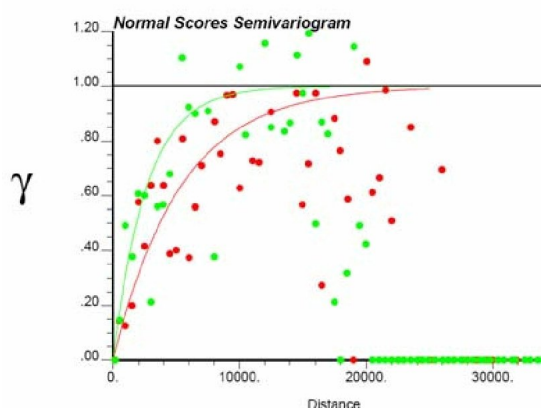

Figure 3: Indicator semivariogram corresponding to the threshold for X. Variograms showed anisotropic behavior. Modeled variograms (red line- major axis of anisotropy and green lines- minor axis of anisotropy) along with the lag distances (h) are used to determine indicator covariances.

Those covariances are then used to determine the required weights for Equation 1. Preparation of hard and soft indicator data were done using the threshold values (mean values of well depths and corresponding mean two way time values for X, Y and Z region of the horizon). A code SISIM from GSLIB (Deutsch and Journel, 1998) is used to generate twenty simulation results. All the simulated results along with the interpreted horizon are plotted in Figure 4. A simple interpretation of the simulated results is that the "true" horizon can be anywhere within the cloud of the simulated results. The standard deviation of all the simulated results can be considered as the uncertainty in the reservoir horizon at a particular location.

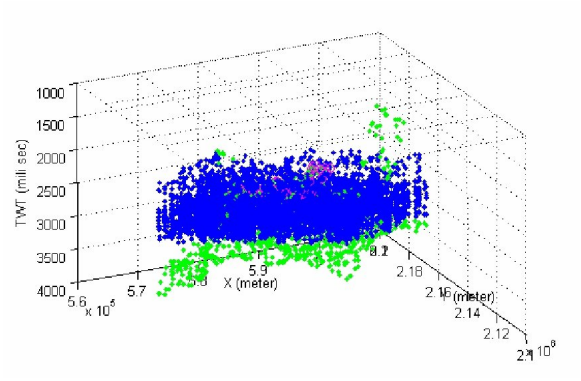

Figure 4: All the simulated horizons (blue dots) are plotted along with the picked horizon from seismic (green).

Discussion

Markov Bayes stochastic simulation integrating soft seismic data and hard well data generated several realizations of the A horizon. A mean of all those realizations along with its standard deviation can be treated as the uncertainty associated with the model (Figure 5). However, the results obtained from our study raised some questions. We can see from figure 5 that the simulated model even with its uncertainty limit has no similarity to the picked horizon at the edges. Thus we need to determine how much confidence we can put on these simulated results. In Figure 6, we plot the absolute value of differences between the mean simulated models and the picked horizon from the seismic. From Figure 6 it is clear that near the well locations, the difference between the mean model and the picked horizon is small. Thus the structural uncertainty assessed from simulation is definitely valid for a localized zone (black ellipse) which is well constrained by the log data. If we believe in the knowledge of the seismic data interpreter then we definitely cannot rely on the simulation results far away from the hard data location. Thus even though simulated results theoretically yield better estimation of the horizon, due to lack of random distribution pattern of hard data, the results may not represent the true picture for the entire horizon. On the other hand, the central part of the A horizon is interpreted with the presence of large displacement faults. Those interpretations may be questionable. Thus within the elliptical zone of figure 6, which is well constrained by the log data, we observe sparse high magnitude residuals. These residuals are indication of interpreted faults which are not at all verified by the log data. Simulation results reasonably questioned those interpretations. We, therefore, conclude

that based on the available data settings, simulation performed an excellent job of horizon uncertainty estimation.

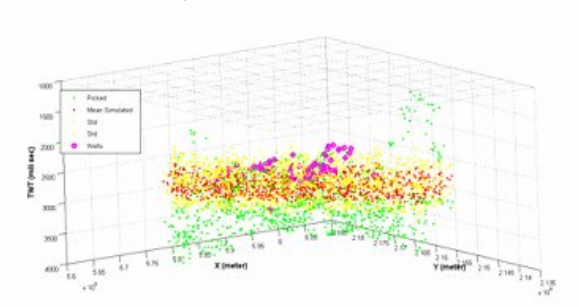

Figure 5: Mean of simulated horizons (red dot) along with its upper and lower uncertainty ranges (yellow dots). Magenta dots are well locations. Note that the picked horizon (green dots) lies outside the uncertainty ranges predicted by our simulation.

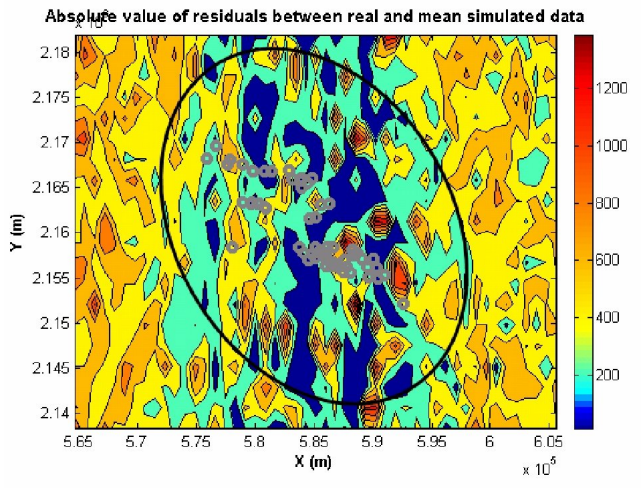

Figure 6: Absolute value of residuals between the mean simulated models and picked A horizon. The difference is small inside an elliptical zone (blue patches within the ellipse). Wells (gray circles) are located almost along the major axis of the ellipse. The simulated results may be questionable far from the center of the A horizon due to localized distribution of the hard data and certain places inside the ellipse which are not controlled by hard data.

Conclusion

Prediction of uncertainty associated with a seismic horizon is addressed using Markov Bayes simulation process. The simulation shows excellent results at and around the hard data locations. Far from the hard data locations, drastic differences between simulation and picked result are observed. Even though the number of hard data is large in this study, its lack of randomness in spatial distribution poses a problem in simulation results. Therefore the uncertainty estimated away from the hard data locations is not good enough to describe the uncertainty of the picked A horizon. At the center of the A zone, where all the wells are located, we still observe localized mismatches between the mean simulated results and picked horizon. These mismatches are likely due to interpreted faults from the seismic data. These faults are not constrained by well data and their presence is questionable. In this scenario, the results obtained from the simulation may be a better estimate of the horizon.

References

Deutsch, C., and Journel, A., 1998, GSLIB: Geostatistical Software Library and User's Guide, Oxford University press, 1-369

Eidsvik, J., Avseth. *P.*, More, H., Mukerji, T., and Mavko., G., 2004, Stochastic reservoir characterization using prestack seismic data; Geophysics, 69, 978-993 Larsen, A., Ulvmoen, M., Omre, and Buland, A., 2006, Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model, Geophysics, 71, R69- R78

Mukherji, T., AVseth, P., Mavko, G., Takahashi,I., Gonzalez, E.,2001, Statistical rock physics: Combining rock physics, information theory, and geostatistics to reduce uncertainty in seismic reservoir characterization, The Leading Edge, March, 313-319

Kelkar, M., and Perez, D., 2002, Applied Geostatistics for Reservoir Characterization, SPG,1-263

Xu, W., Tran, T., Srivastava, R., and Journel, A., 1992, Integrating Seismic Data in Reservoir Modelling : The Collocated Cokriging Alternative, SPE Extended Abstract, 833-842

Acknowledgment

We are thankful to other members of this project from the Petroleum Engineering department of University of Texas at Austin. This work is a part of the reservoir model uncertainty estimation project led by the petroleum engineering department of UT Austin.