GEOLOGICAL SURVEY OF CANADA
OPEN FILE 6253

# A non-parametric discovery process model - A least squares approach

G. Chen, Z. Chen and P. Tzeng

2010

Natural Resources Canada    Ressources naturelles Canada

**GEOLOGICAL SURVEY OF CANADA**
**OPEN FILE 6253**

# A non-parametric discovery process model - A least squares approach

**G. Chen, Z. Chen and P. Tzeng**

**2010**

©Her Majesty the Queen in Right of Canada 20xx

This publication is available from the Geological Survey of Canada Bookstore
(http://gsc.nrcan.gc.ca/bookstore_e.php).
It can also be downloaded free of charge from GeoPub
(http://geopub.rncan.gc.ca/).

G. Chen, Z. Chen and P. Tzeng, 2010: A non-parametric discovery process model - A least squares approach, Open File 6253, 12 p

Open files are products that have not gone through the GSC formal publication process.

# Table of Contents

## Abstract

This open file discusses an algorithm that uses a least square approach for a non-parametric discovery process model. The mathematical derivation of the model is presented, and it is followed by two application examples from different play types and sample sizes to illustrate the use of the proposed method in petroleum resource estimation. A computer code of the algorithm in "R" language and the datasets in the application examples are also included in this open file.

## 1 Introduction

The legacy product of petroleum resource assessment, PETRINES, was developed in the 1980s by the Geological Survey of Canada (GSC) (Lee and Wang, 1985). Demands for improvement, additional functionality and a better user interface are high from the Canadian provinces and territories as well as the petroleum industry. The Geological Survey of Canada has undertaken and continues to conduct methodology studies on petroleum resource assessment under the former Secure Canadian Energy Supply Program and the current GEM Energy Program to meet these growing demands. This proposed algorithm is one of the outputs from the above mentioned programs.

Logan (2005) developed a truncated lognormal discovery process model that uses a least squares solution instead of the maximum likelihood solution of the lognormal discovery process model developed by Lee and Wang (1985). The least squares approach has the advantage of graphically displaying the goodness of fit between the model predicted sequence and the discovery sequence, which enables a direct comparison of the modelled and real discovery sequences in a conventional way and provides a graphic visualization that helps the resource assessor to judge the output estimations of the petroleum resource potential in an area under study. A disadvantage of the least squares solution is the strong dependence on the initial values of input for the unknown parameters under estimation as exhibited in Logan's Excel program. A search for all possible solutions in a constrained parameter space would be more desirable.

This open file describes the derivation of the least squares solution of a non-parametric discovery process model. Two application examples of different play types with a contrast in discovery sequence sizes are given to illustrate the use of the proposed least squares approach of the non-parametric discovery process model in petroleum resource assessment. A discussion of advantages and drawbacks of the method follows. Computer code of the algorithm in 'R' language, as well as the application datasets are provided in the Appendices.

## 2 Methodology

Consider a play with (unknown) $N$ fields made up from (known) $K$ classes where class $j$ has (unknown) $N_j$ fields, each with volume $v_j$, $j$ = 1, ... , $K$. When $n$ discoveries $y_1$, $y_2$,⋯, $y_n$ are made, we observe $n_j$ fields of volume $v_j$ , $j$ = 1, .. . ,$K$ , where $n_1 + n_2 + ... + n_K = n$. The $n$ discoveries can be treated as the result of sampling from the play without replacement and under the assumption that a discovery of a field with volume $v_j$ is proportional to a weight $w_j$, say, $w_j = v_j^{\beta}$, $j$ = 1, ... , $K$ , where $\beta$ is known as the discoverability coefficient.

A graphical way to represent the discovery process is to plot the cumulative discovered volume against the order of discovery, namely, we plot $S_i = y_1 + . .. + y_i$ against $i$ , where $i$ = 1, 2, .. . ,$n$ . For the non-parametric approach, we let

$$\hat{N}_j = \frac{n_j}{1 - \exp\{-w_j \lambda\}}, \qquad \lambda \neq 0, \qquad j = 1,...,K, \qquad (1)$$

denote the Horvitz-Thompson estimator of the unknown $N_j$. When $i$ discoveries are made, we let $m_j(i)$ denote the mean number of class $j$ fields remaining to be discovered, $j$ = 1, ... $K$. By definition, $m_j(0) = N_j$, but we do not know $N_j$, so we use $m_j(0) = \hat{N}_j$ instead. Then through iteration, we use the method in Ninpong *et al* (1992) to find $m_j(i)$ as

$$m_j(i) = \max\{m_j(i-1) - q_j(i-1), 0\}, \qquad i = 1,...,n, \qquad j = 1,...,K, \qquad (2)$$

with

$$q_j(i-1) = \frac{w_j m_j(i-1)}{w_1 m_1(i-1) + w_2 m_2(i-1) + ... + w_K m_K(i-1)},$$

where the maximum is taken to guarantee that $m_j(i) \geq 0$. The above is based on a difference equation version of the Arps-Robert differential equation, and we can therefore produce an estimate of the mean cumulative discovered volume when $i$ discoveries are made as

$$t_i = \sum_{j=1}^{K} (\hat{N}_j - m_j(i)) v_j, \qquad i = 1,...,n. \qquad (3)$$

For a given discovery series the estimate $t_i$ depends on two parameters β (through $w_i$) and λ (through $\hat{N}_j$). To determine these two parameters, one can apply the method of least squares by minimizing the following target function

$$L(\beta, \lambda) = \sum_{i=1}^{n} (s_i - t_i)^2 \qquad (4)$$

over the parameter space: $-\infty < \beta < +\infty$ and $\lambda \neq 0$.

To apply the above to a real discovery series $y_1, .. . , y_n$, one needs to first create $K$ class intervals $I_1, I_2, ... I_K$ to cover all of the observed $y_i$'s. Let $I_1 = [a_0, a_1)$, $I_2=[a_1, a_2),\ldots,$ $I_K=[a_{k-1}, a_k)$. Then $n_j$ is the number of $y_i$'s that fall into the interval $I_j$, $j = 1 ... K$. Let $v_j$ be the middle point of the interval $I_j$, namely $v_j = (a_{j-1}+ a_j)/2$ and replace the observed $y_i$ with $v_j$ if $y_i$ falls into class $j$ to have an adjusted discovery series.

## 3 Application examples

Two application examples are discussed in this open file, a large dataset example from the Niagaran pinnacle reef play of the Michigan Basin, and a small dataset example from the Mesozoic clastic structural petroleum play in the western Sverdrup Basin, to illustrate the applicability of the proposed method for different play types that differ in size and lithology of the sedimentary succession. The discovery sequence dataset of the Niagaran pinnacle reef play comes from Geological Survey of Canada database collected by Lee and Gill (1999) and the dataset for the western Sverdrup basin is from Chen and Osadetz (2009).

### Niagaran Pinnacle Reef Play

The Niagaran play is a well established oil and gas play in the Middle Silurian pinnacle reef of Michigan Basin (Figure 1) with a total oil and gas reserve of 1029.4 million of barrels oil equivalent (Mboe). The dataset consists of 538 discoveries of oil, gas and oil

& gas fields, ranging in sizes from $30 \times 10^3$ to $21 \times 10^6$ barrels of oil equivalent (boe) discovered from 1968-1982 (Figure 2). Details of the geology and petroleum exploration history are referred to Gill (1994a and b). This dataset has been used by Lee and Gill (1999) for comparing the estimates of petroleum resource potentials from different discovery process models that were available at that time. This provides an ideal dataset for testing the proposed least squares algorithm of the non-parametric discovery process model described in the previous section.

There are two unknown parameters, $\lambda$ and $\beta$, to be estimated. To understand the behaviors of the target function and investigate the locations of possible local minimums, a grid search for $\lambda$ ranging from 1 to 3.1 with an increment of 0.02 and $\beta$ ranging from 0.1 to 1.0 with an increment of 0.01 leads to the contour plot or map of the target function in Figure 3. One can see from Figure 3 that the minimum of the target function is located around $\lambda = 2.4$ and $\beta = 0.58$. Figure 4a shows the fit between the discoveries and the model estimates when the target function is minimised. A comparison of the discovered field size distribution with the predicted field size distribution is in Figure 4b, from which it is clear that the remaining undiscovered fields predicted from this approach are relatively small and there are not many fields to be discovered.

As compared to the petroleum resource estimates from different methods discussed by Lee and Gill (1999), the least-squares non-parametric approach gives conservative estimates both in resource potential and number of fields yet to be discovered (Figure 5). The conservative nature of the estimations could be a result of the lack of spatial representatives of discoveries in the offshore area of the Great Lakes (Figure 6); only two discoveries were made in the offshore areas. If the offshore represents one third of the pinnacle reef area, the resource potential of the entire pinnacle reef play could be in the order of 1500 Mboe by areal analog. The lack of spatial representatives is a common, but complicated problem for discovery process models (Kaufman et al., 1988) and will be discussed elsewhere.

### Mesozoic Clastic Structural Play of Western Sverdrup Basin

The oil and gas found in this play are accumulated in anticline structures of the Mesozoic sandstone succession in western Sverdrup basin, Canadian Arctic region. The geology, exploration history and estimated resource potentials are referred to in Chen et al. (2000 and 2009). Figure 7 shows the location of the play and exploration status and Figure 8 is the stratigraphic distribution of the discovered oil and gas in the play. A grid search over $\lambda$ ranging from 0.01 to 3.0 in increments of 0.05 and $\beta$ ranging from 0.1 to 1.5 in increments of 0.01 gives the mapped target function surface in Figure 9. From the target function value map and geological information, it appears that $\lambda$=0.03 and $\beta$=0.82 represents a local minimum that gives the best results honoring the observed discoveries (Figure 10a) and being consistent with known geological constraints. A comparison of the predicted and observed distributions of the Mezosoic clastic structural play of the western Sverdrup Basin is shown in Figure 10b, which suggests that most of the remaining undiscovered field sizes are intermediate and small. Both the estimated play resource and the number of fields by this proposed approach are close to the estimates from the Geoanchored method and a calibrated volumetric approach by Chen and Osadetz (2009).

## 4 Discussion

The least squares approach has the graphic capability of displaying the goodness of fit of the modelled discovery sequence against the observations and provides a graphic visualization that could help resource assessors to judge the output estimations of the petroleum resource potential. This graphic capacity is particularly useful for those assessors with a weak statistical background because it helps them to visualize the model data fitting and parameter determination.

The least squares approach of the non-parametric discovery model in equation (1) finds a solution by minimizing the difference between the observations and modeled values. There could be many local minimums in the parameter space. The proposed non-parametric algorithm converges to the same value if the initial values for ($\lambda$ and $\beta$) are close enough to a local minimum in parameter space. If the initial values give $\hat{N}_j$ that are close to $n_j$, the non-parametric method will basically fit the observed discovery series exactly. Therefore, to obtain a geologically reasonable solution, one should start with a broad initial parameter search space in order to capture multiple possible solutions first, then eliminate contradictory solutions against available geological and exploration information. Because there are only two unknown parameters for this non-parametric approach, one alternative is to calculate target function values over a grid of the parameter space to map the local minimums. A pair of $\lambda$ and $\beta$ from a local minimum that fits both data and geological constraints could provide the "best" solution. Like all other available discovery process models, additional geological and exploration information and data for the study area are always important for constraining the resulting estimates.

## 5 Acknowledgements

The authors thank the internal reviewer Dr. Dale Issler, Geological Survey of Canada, for his helpful comments and suggestion.

## 6 References

Chen, Z. and Osadetz, K. G., 2009. Improving reservoir volumetric estimations in petroleum resource assessment using discovery process models; Petroleum Science, v. 6 (2), p. 105-118.

Chen, Z., Osadetz, K., Embry. A, Gao, H., and Hannigan, P., 2000. Petroleum potential in western Sverdrup Basin, Canadian Arctic Archipelago; Bulletin of Canadian Petroleum Geology, v. 48 (4), p. 323-338.

Gill, D., 1994a. Niagaran reefs of northern Michigan, part I: exploration portrait; Journal of Petroleum Geology, v. 17 (1), p. 99-110.

Gill, D., 1994b. Niagaran reefs of northern Michigan, part I: resource appraisal; Journal of Petroleum Geology, v. 17 (2), p. 213-242.

Kaufman, G. M., Crovelli, R .A., Chow, S., Grace, J. D., Sinding-Larsen, R., Sollie, B. H., and Wang, P. C. C., 1988. Oil and gas modeling and forecasting, in Chung, C. F., Fabbri and Sinding-larsen, R. eds., Quantitative Analysis of Mineral and Energy Resources, D. Reidel Publication Company, p. 695-706.

Lee, P. J. and Gill, D., 1999. Comparison of Discovery Process Methods for Estimating undiscovered Resources; Bulletin of Canadian Petroleum Geology, *v*.47(1), p. 19- 30.

Lee, P. J. and Wang, P. C. C., 1985. Prediction of oil and gas sizes when discovery record is available; Mathematical Geology, v.17 (2),  p. 95-113.

Logan, K. G., 2005. The Truncated Discovery Process Model; Natural Resources Research, *v*.14 (3), p. 265- 281.

Ninpong, R., Power, *II*. and Fuller J. D., 1992. A Rapid Approximation for Predicting the Hydrocarbon Discovery Rate: Part 1 Assessing the Accuracy; Natural Resource Modeling, *v*.14 (3), p. 285- 303.

## Figure Captions

Figure 1. Distribution of principal pinnacle Middle Silurian reef zone in the Michigan Basin, USA (figure from Gill, 1994a).

Figure 2. Discovery sequence (2a) and field size distribution (2b) of the discovered oil and gas fields in the Niagaran pinnacle reef play.

Figure 3. Map of calculated target function value as a function of the two unknown parameters, outlining possible parameter space for the least square solution. The color bar indicates the logarithmically converted target function value.

Figure 4. A least square solution to the discovery dataset, showing the predicted total play resource and number of fields in the Niagaran reef play (4a) and a comparison of observed and predicted field size distributions (4b), Niagaran pinnacle reef play.

Figure 5. Comparison of the estimated mean values of play resource and number of fields from this study with estimates from other methods. See Lee and Gill (1999) for the abbreviations and methods.

Figure 6. Spatial distribution of discovered oil and gas fields and general geological setting showing the play definition of the Niagaran reef play in the Michigan Basin.

Figure 7. Map shows the location of the western Sverdrup basin and petroleum exploration status of the Mesozoic clastic structural play.

Figure 8. Diagrammatic representation of the Mesozoic clastic structural play (left) and stratigraphic distributions of discovered oil and gas resources of the western Sverdrup Basin.

Figure 9. Map of calculated target function value as a function of two unknown parameters, showing the likely parameter space for the least square solution, the western Sverdrup Basin clastic structural play in the Canadian Arctic.

Figure 10. A least squares fit of the model to the observations in cumulative volume of discoveries (million boe), showing both resource potential and number of remaining fields (10a) and a comparison of discovered and predicted field size distributions (10b) of the Mesozoic clastic structural play in the western Sverdrup Basin.

## Appendix Captions

Appendix A. Computer code in 'R' for the least squares approach of non-parametric discovery process model.

Appendix B. Niagaran play dataset.

Appendix C. Western Sverdrup Basin Mesozoic structural play dataset.