PETROVIETNAM
SCIENCE, TECHNOLOGY & INNOVATION PETROVIETNAM

IMPROVING FDP DECISIONS UNDER SPARSE DATA: AN ADAPTIVE METROPOLIS APPROACH FOR NET-PAY QUANTILES

Pham Xuan Phuc, Le Tran Minh Tri, Tran Vu Tung, Nguyen Hong Kien, Nguyen Ngoc Tuan Anh, Pham Tien Trung
Phu Quoc Petroleum Operating Company (Phu Quoc POC)
Email: tungtv@phuquocpoc.vn
https://doi.org/10.47800/PVSI.2025.06-04

Summary

Early field-development decisions often rely on very small samples of reservoir property measurements, where common spreadsheet workflows (typically lognormal fits) become fragile and analyst-dependent. This paper presents the Metropolis-Adaptive Distribution Range-Constrained (M-ADRC) method, a practical Metropolis-Hastings (MH) approach with adaptive constraints for generating representative synthetic populations from low number of observations. The workflow combines (i) domain-informed parameter bounds, (ii) adaptive proposal tuning, (iii) automatic switching between Pearson Type III and lognormal distribution families, and (iv) adaptive iteration policy for samples with extreme bias.

Using a Southeast Asia gas field netpay dataset, the representative population generated by M-ADRC is benchmarked against the traditional lognormal spreadsheet approach across diverse net pay samples. Results show that M-ADRC achieves significantly lower Wasserstein Distance and smaller Swanson's mean error, reaching the "excellent" Wasserstein threshold with roughly eight samples, only half the sample requirement of the spreadsheet approach.

The workflow is lightweight, fully open-source, and requires no specialized hardware. It includes default settings, diagnostic checks, and sensitivity results, enabling reservoir engineers to apply M-ADRC without specialized tuning expertise. M-ADRC yields more reliable small-sample quantiles, reducing the risk of biased inputs in FDP scenarios and enhancing decision confidence. Workflow limitations such as restricted distribution types, resilience to extreme sampling bias, and the lack of geospatial properties will be addressed in future work.

Key words: Reservoir characterization, metropolis-hastings, uncertainty quantification, field development plan.

1. Introduction

Field development plant (FDP) is a strategy outlining the characteristics of an oil and gas field obtained through the exploration phase, and recommending the optimal procedure to extract the hydrocarbon safely, economically, while complying with all regulations and operational constraints. A central cornerstone of any FDP is reserve estimation, which is carried out through reservoir characterization [1]. One popular estimation method during the early stage of development is volumetric [2], where the reservoir is simplified as a porous box. The original oil and gas in place (OOGIP) is then calculated by multiplying the volume of the box (area times thickness) with porosity, hydrocarbon saturation, and dividing by formation volume factor [3]:

$$OOIP = \frac{Ah\phi S_o}{B_o}$$

$$OGIP = \frac{Ah\phi S_g}{B_g}$$

Therefore, quantifying the uncertainty in formation parameters such as porosity, water saturation, and net pay is pivotal for effective employment of volumetric methods [4, 5]. However, the complex nature of field exploration makes direct sampling difficult. To overcome limited sample data, reservoir engineers and geoscientists often rely on geostatistical methods, ranging from simple Kriging to sequential Gaussian simulation [6]. Research on these reservoir properties has been significant, with several papers proposing variations of methods and discussing the importance of these properties [7, 8].

Date of receipt: 22/9/2025.
Date of review and editing: 22/9 - 13/11/2025.
Date of approval: 13/11/2025.

Regarding the challenge of modeling the true distribution of geophysical properties from a small number of samples, a wide variety of statistical methods have been proposed [9]. GhojehBeyglou [10] compared Kriging, Sequential Gaussian Simulation (SGS), and Gaussian Random Function Simulation (GRFS) in determining porosity distribution and concluded that Kriging excels at giving single-point prediction, while SGS and GRFS are more compatible in capturing the variability. For 1D distribution, Aleardi [11] investigated the best statistical models for predicting the multi-dimensional distribution of porosity and litho-fluid facies based on a histogram.

In addition to the aforementioned statistical methods, a popular approach that has demonstrated considerable effectiveness is the application of Monte Carlo simulation [7], a repeated sampling method that has seen extensive use in the oil and gas industry and significantly refined over time through advances in computer engineering. However, traditional Monte Carlo implementations, such as those in Crystal Ball, require large datasets to produce reliable results, posing challenges in data-scarce environments like offshore fields. Therefore, the Metropolis-Hastings (MH) algorithm [12] is introduced as a Markov Chain Monte Carlo (MCMC) technique used to sample from complex probability distributions, applicable for low to very low numbers of samples. This method generates a sequence of correlated samples by proposing candidate points and accepting them with a probability that ensures convergence to the target distribution. MH has been researched thoroughly in petroleum engineering and geoscience, including reservoir modeling [13], well parameters estimation [14], and, most notably, history matching [15 - 17].

This study introduces a practical Metropolis-Hastings-based algorithm called Metropolis-Adaptive Distribution Range-Constrained (M-ADRC) to sample reservoir parameters by incorporating statistical data of the population to create an accurate representative synthetic population. The algorithm is packaged and deployed in a web environment for internal use. A case study using net pay data from Southeast Asian gas fields further validates the M-ADRC approach and demonstrates its superiority over the traditional approach.

2. Methodology

2.1. Distribution

The proposed algorithm currently deals with lognormal and Pearson III distributions.

Lognormal distribution

The lognormal distribution is a continuous probability distribution in which the logarithm of the variable is normally distributed. It is characterized by a positive skew (right-tailed). There are two parameters controlling a lognormal distribution (Figure 1):

- μ: The mean of the natural logarithm of the distribution.

- σ: The standard deviation of the natural logarithm of the distribution.

In petroleum engineering, this distribution type is frequently applied to model reservoir properties such as net pay, permeability, reserves, as these parameters often arise from multiplicative geological processes and exhibit right-skewed distributions [18, 19].

Pearson III distribution

The Pearson Type III distribution is a three-parameter probability distribution that represents a shifted and scaled form of the gamma distribution [20]. It is defined by:

- Skew (α) - controls the skewness. Negative means left-skewed and positive means right-skewed. Zero means normal distribution.

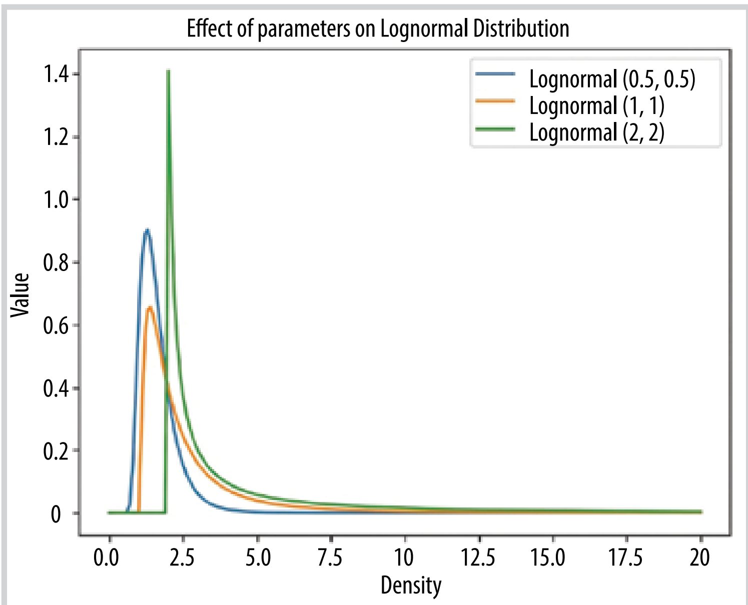

Figure 1. *Effect of parameters on lognormal distribution.*

- Location (μ) - shifts the curve left or right.
- Scale (β) - stretches or compresses the distribution.

Due to its flexibility in representing both positively and negatively skewed data (Figure 2), it has been widely applied in hydrology, sedimentology, and petroleum reservoir characterization. However, its complexity also poses challenges in modeling compared to simpler distribution types such as lognormal.

Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is a robust MCMC method designed to sample complex probability distributions where direct sampling is infeasible. It constructs a Markov chain that converges to the target distribution (P(x)), enabling the generation of representative samples. Assuming a sample set of x_0, x_1,… x_n from a presumably lognormal distribution of shape μ and scale σ, the algorithm's steps are as follows:

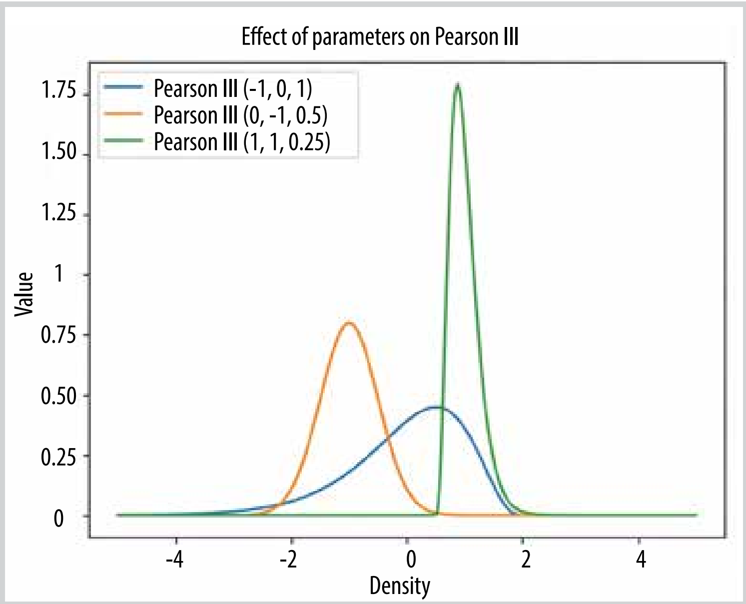

***Figure 2.** Effect of parameters on Pearson III distribution.*

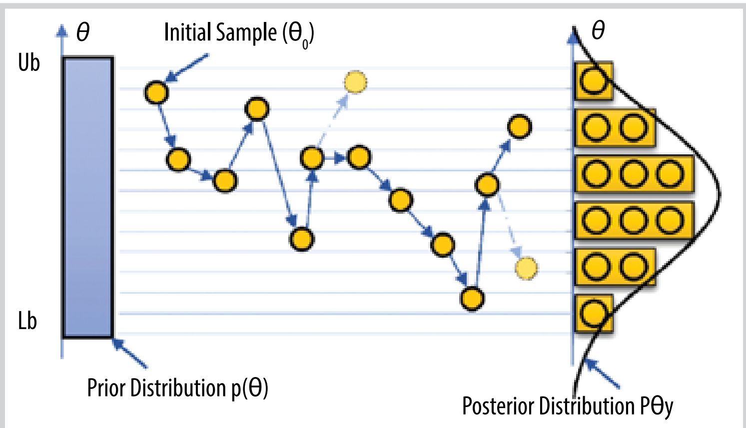

***Figure 3.** Graphical representation of Metropolis-Hasting algorithm [21].*

Initialization: Start at iteration i = 0, select an initial value (μ_0 and l_0 with l = log (σ)) from the parameter space, typically based on domain knowledge or the mean/median value of the samples.

Proposal: Generate new candidate values μ_i and l_i from values in the previous iteration by adding a proposal term ∈ randomly selected from a symmetric, usually normal, distribution:

$$\mu_i = \mu_{i-1} + \epsilon_\mu$$
$$l_i = l_{i-1} + \epsilon_l$$
$$\epsilon \in N(0, \sigma)$$

Acceptance probability: The acceptance probability is calculated as:

$$\alpha = min\left(1, \frac{p(\mu_i, l_i)}{p(\mu_{i-1}, l_{i-1})}\right)$$

Acceptance/rejection: Draw a random number u from a uniform distribution. If α is larger than or equal to u then accept μ_i and l_i into the Markov chain. If not, then reject them and add the old values $\mu_{(i-1)}$ and $l_{(i-1)}$ instead.

Iteration: Repeat steps 2 - 4 for N number of iterations to generate Markov chains of μ and l. After a burn-in period, typically 10% of N, the chains converge and samples are collected.

Optimal parameters selection: From the chains, multiple methods can be employed to select the best set of parameters. For this study, simple median selection is sufficient.

$$\mu_i = \mu_{i-1} + \epsilon_\mu$$
$$l_i = l_{i-1} + \epsilon_l$$

For other types of distributions with a different number of parameters, the same concept is applicable. Figure 3 represents the graphical illustration of Metropolis-Hastings: generate a posterior distribution from a prior uniform distribution.

M-ADRC modified algorithm:

From the Metropolis-Hasting fundamentals, many functionalities are incorporated into M-ADRC algorithm:

Range restrictions

One of the key requirements for the algorithm to perform efficiently is a range limit of parameters (skewness, location, scale, mean of population), which can be obtained from domain experts' analysis and analogy data from nearby fields. The range should be wide to avoid accidentally excluding the true parameter values, also not so broad that the search space becomes inefficient and defeats the purpose of parameters range restriction. When the algorithm proposes an out-of-bound sample, it will be rejected. This improves the stability and efficiency of the search process.

Adaptive proposal

Acceptance rate is a key attribute in evaluating the reliability of Metropolis-Hasting process. Too low (< 1%) or too high (> 60%) rate implies poor sampling procedure. Acceptance rate can be adjusted through the proposal width of each parameter. To avoid manually tuning the width, a piecewise function is integrated into the algorithm. During burn-in period, the function divides this phase into 10 subsections and monitors the acceptance rate in each. If the acceptance rate is out of bound (1 - 60%), the function will multiply the rate by an adjustment factor depending on the magnitude of the acceptance rate. The adjustment factors are user-inputs. Future work may explore auto-adaptive algorithms for proposal width, such as works proposed by Rosenthal [22].

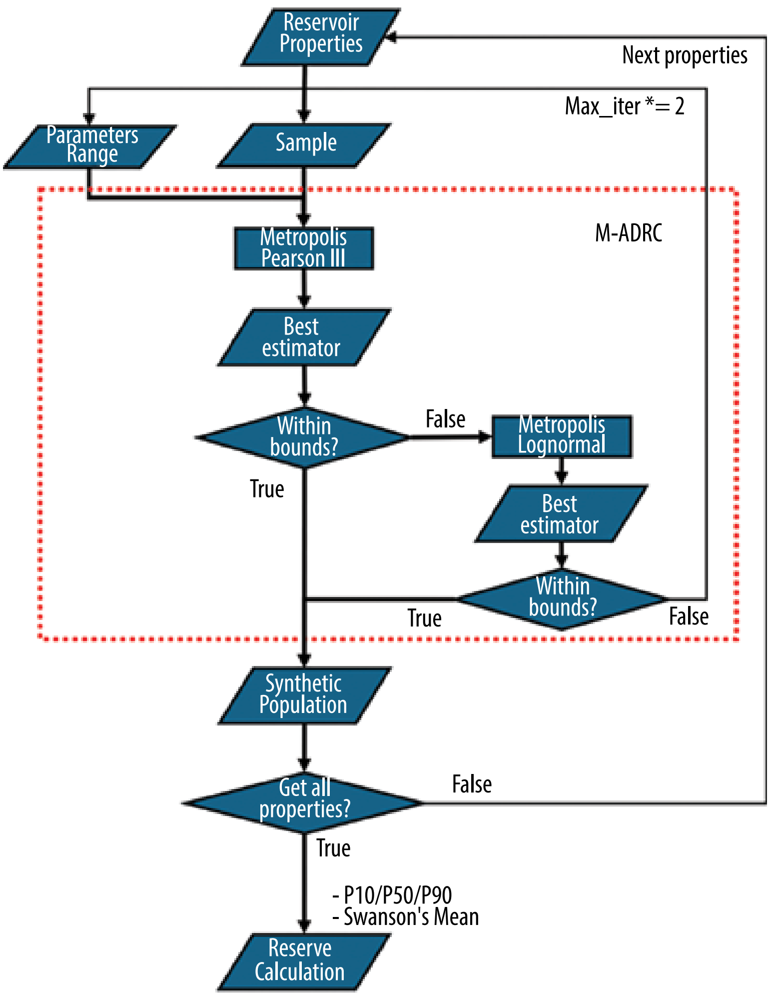

***Figure 4.** M-ARDC workflow incorporated into reserve calculation.*

Distribution switch

If the original Pearson III fit fails to produce a correct final solution (out-of-bounds), which can happen in extreme cases of sampling bias, the algorithm will switch to a lognormal distribution and restart the entire process. Lognormal distributions are much more robust than Pearson III [23] and almost guarantee to produce a "reasonable" solution. A "risk factor" to quantify the risks involved in using these samples for modeling distribution will be introduced in future work.

Adaptive iterations

In worst-case scenario where both the Pearson III and lognormal distributions fail, the algorithm pulls a last-ditch effort by restarting the entire process from the beginning, with double the number of iterations. Our analysis shows that the main reason for failure is due to zero acceptance rate. This, however, can be remedied by increasing number of iterations so that the algorithm has time to stabilize during burn-in. This is, nevertheless, a quick fix that does not address the underlying sampling bias problem. This issue will be revisited in our future work.

Workflow

The M-ADRC workflow is visualized in Figure 4. Detailed procedures are as follows:

First, the algorithm will attempt to fit a Pearson III distribution on the sample and collect skew, location, scale values as initial guess.

The modified Metropolis algorithm is then executed to obtain distribution of these parameters, ensuring that all parameters must be within bounds. The best-guess value for each parameter is chosen as the median.

If the best-guess values do not meet requirements, the algorithm will switch to lognormal distribution and rerun the process.

If lognormal distribution also fails, the algorithm will restart from the beginning with double number of iterations.

The reserve estimation is incorporated into the M-ADRC workflow. For each property such as net pay, porosity, saturation, the algorithm will generate a synthetic population accordingly. Once all results have been acquired, it is then the decision of the petroleum engineers on how to proceed with reserve calculation: take the P10/P50/P90 or Swanson's mean of each property and multiply them in classical volumetric formula, or perform Monte Carlo on the joint synthetic population to get a synthetic distribution of the reserves.

2.2. Lognormal spreadsheet-based approach (LSA)

Many petroleum operating companies assume lognormal distribution for majority of reservoir properties like net pay, porosity, reserve. In order to generate distribution, they fit a lognormal curve to the samples, get μ and σ, generate the Percentage Point Function (PPF) and use it for Monte Carlo simulation. This process, while simple and reliable in many cases, can be highly misleading under some circumstances. The number of samples must be sufficient to generate a representative distribution. Commercial software may require a minimum of twelve (12) samples to work. In many field development projects, the number of core samples can be as low as four (4), which significantly decreases the reliability of the synthetic distribution. In addition, sometimes the sample can be of poor quality: too far apart or too clustered, left-skewed. A practice employed by our engineers is to duplicate the samples so that the number of samples exceeds the threshold of commercial software. This is, however, an unproven practice with no scientific justification.

3. Case study setup

The case study focuses on net pay data from nearly 2,000 wells in gas fields across Southeast Asia, characterized by a lognormal

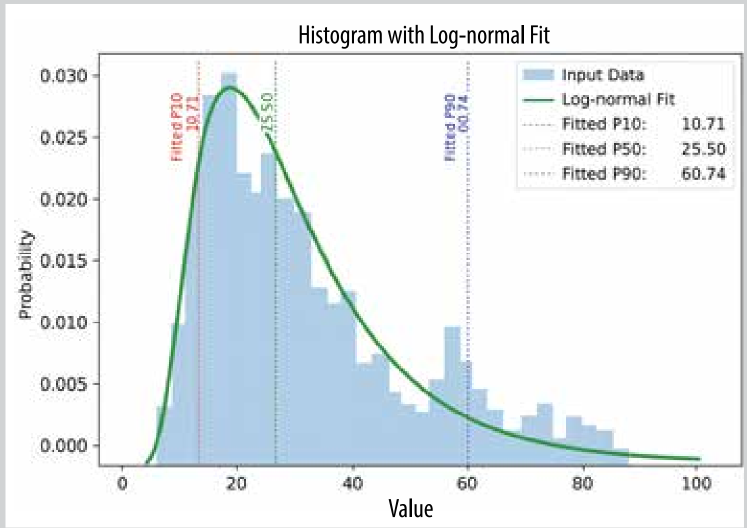

Figure 5. *Lognormal fitted curve.*

distribution. A smaller subset of 500 wells with similar geological and operational attributes to our operating fields has been selected and will be used for this case study. The net pay data for this population can be characterized by either lognormal, Johnson's SU, or Pearson III distribution, as shown in Figure 6. Currently the algorithm focuses on Pearson III and lognormal distributions. Future work may extend further to include more comprehensive distribution types.

All simulations are implemented in Python and executed on standard office machine, with no specialized software or hardware.

For the parameters range as input for M-ADRC (Table 1), this case study uses very wide range, much wider than the actual values as boundaries:

The simulation parameters are listed in Table 2.

The flowchart for validation process is presented in Figure 7. Detailed explanation is as follows:

Sample selection: A number of random samples were drawn from the 500-well dataset.

Simulation: The M-ADRC algorithm and traditional lognormal-based spreadsheet approach (LSA) were executed to generate the percentage point function (PPF).

Validation: The PPF generated from the proposed algorithm and spreadsheet method were validated against actual population using two metrics:

Wasserstein distance (WD): Also known as Earth Mover's distance, measures distance between two probability distributions. It is widely used in uncertainty quantification (Scheidt and Caers, 2009). The common form is the 1-Wasserstein distance:

$$W_1(P,Q) = \int_0^1 \left|F_P^{-1}(q) - F_Q^{-1}(q)\right| dq$$

Table 1. Distribution parameters used by M-ADRC

Distribution parameters	Range	Actual population value
α (Pearson III)	0 - 40	1.2
μ (Pearson III)	0 - 2,000	138
β (Pearson III)	0 - 2,000	89
Population mean	20 - 300 (95th percentile)	138
μ (Lognormal)	3 - 6.4	4.9
σ (Lognormal)	0-2	0.7

Where q is the quantile and F^{-1} (q) is the quantile function. Two identical distributions will have a distance that equals zero. The unit of Wasserstein distance is the same as the unit of the distribution. For this study, Wasserstein distance between the PPF of the synthetic and actual population will be compared. A "decent" WD score depends on the magnitude of the distributions. Figure 8 shows that for this dataset, a WD score of 25 or less can be considered excellent, as the synthetic population accurately represents actual population.

Swanson's mean (SM): Is an empirical approximation of mean for lognormal or moderately right-skewed distribution based on P10, P50, and P90. The most common formula in oil and gas [24] is:

$$Swanson's\ Mean = 0.3P_{10} + 0.4P_{50} + 0.3P_{90}$$

For this study, the relative difference between SM of synthetic and actual population will be recorded. It is calculated as:

$$\%\,Relative\ Difference = \left|\frac{Synthetic - Actual}{Actual}\right| \times 100$$

Iteration: The process was repeated 1,000 times to assess robustness, with results recorded for statistical analysis.

Repeat: Repeat the entire process but with a higher number of samples. This study will demonstrate an analysis for number of samples from 4 to 12.

4. Results and discussion

4.1. Wasserstein distance

Figure 9 shows the median Wasserstein distance between actual PPF of the true

Table 2. Simulation parameters

Simulation parameters	Value
Proposal α	0.1
Proposal μ (Pearson III)	0.5
Proposal β	0.5
Proposal μ (Lognormal)	0.4
Proposal σ	0.4
n_iteration	10,000
Burn-in	1,000

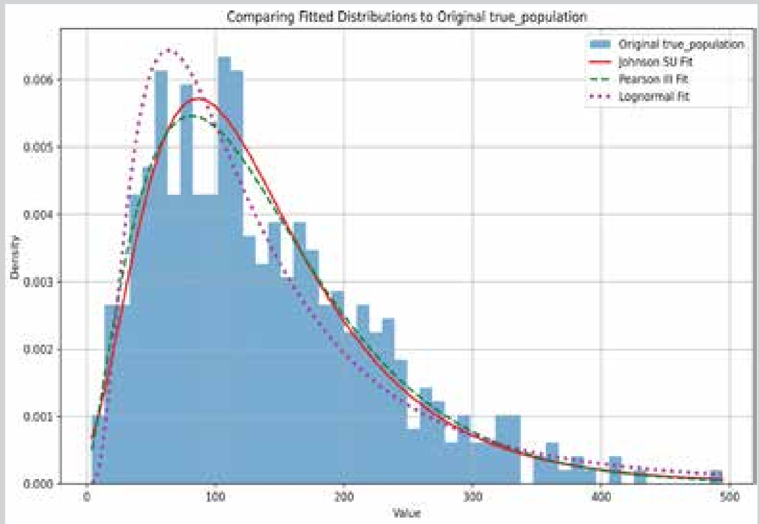

Figure 6. Compare distribution types to data.

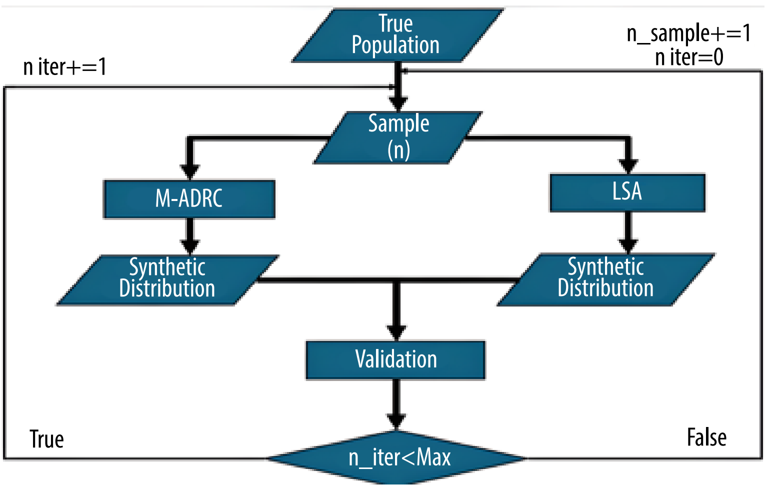

Figure 7. Workflow for the case study.

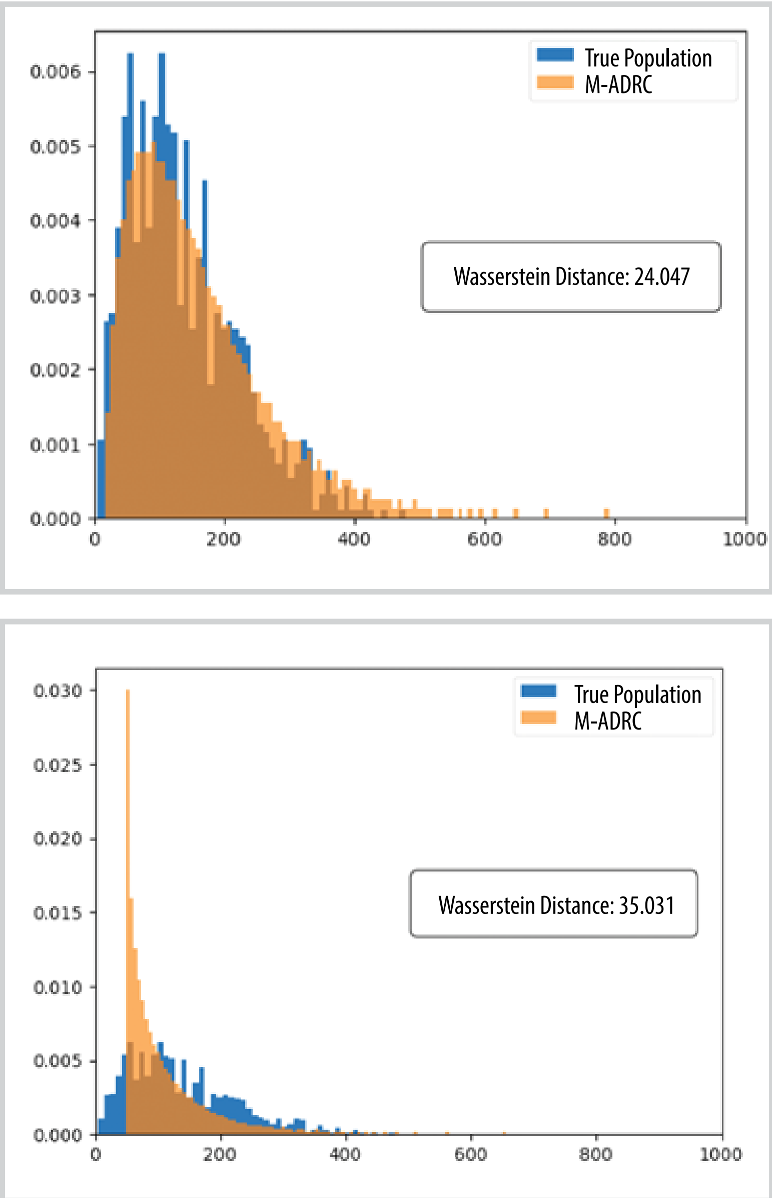

Figure 8. *Example of Wasserstein distance between synthetic and actual population.*

population and synthetic PPF generated from M-ADRC and LSA with increasing number of samples. Each data point represents the median WD value for 1,000 simulation cases with different samples. M-ADRC demonstrates superiority over conventional LSA method. If we consider a Wasserstein distance of 25 as the threshold for excellent representation, then M-ADRC reaches this threshold at around 8 samples, while LSA needs 16 - 18 samples.

Figure 10 further demonstrates the percentage of cases in each run where M-ADRC outputs lower Wasserstein distance value than LSA. To prove its effectiveness, M-ADRC needs to demonstrate improvement in comparison to LSA for more than 50% of cases. The results show that M-ADRC exceeds this threshold by a wide margin, starting at nearly 70% and increasing further as the number of samples grows.

4.2. Swanson's mean

Similar to Wasserstein distance, M-ADRC also shows significantly improved performance in terms of Swanson's mean difference compared to LSA. Figure 11 plots the median relative difference of Swanson's mean between the actual population and synthetic distribution generated from M-ADRC and LSA. Each data point is the median relative difference for 1,000 simulation cases. Although M-ADRC continues to outperform LSA, the performance gap narrows as sample size increases. This is expected because estimating a distribution's mean via Swanson's mean is far less demanding than modeling the full distribution using the Wasserstein distance.

Figure 12 plots the percentage of cases where M-ADRC outperforms LSA in Swanson's mean relative difference metric for different number of samples. The gap between two methods remains relatively constant rather than widening as seen in Figure 10. Nevertheless, M-ADRC maintains at least 60% ratio across all runs.

Results of the case study clearly demonstrate the vast improvement of M-ADRC compared to conventional lognormal spreadsheet approach (LSA) in representing the true population based on a small number of samples. By integrating an advanced MCMC algorithm with adaptive tuning, range restrictions, and distribution switch, M-ADRC is established as a state-of-the-art algorithm, capable of providing reliable synthetic distribution.

4.3. Limitations

The approach requires domain knowledge to limit the range of distribution parameters, especially the mean value. At current stage of development, the algorithm still lacks capability to effectively handle extreme sampling bias. Poor sample selections significantly reduce the

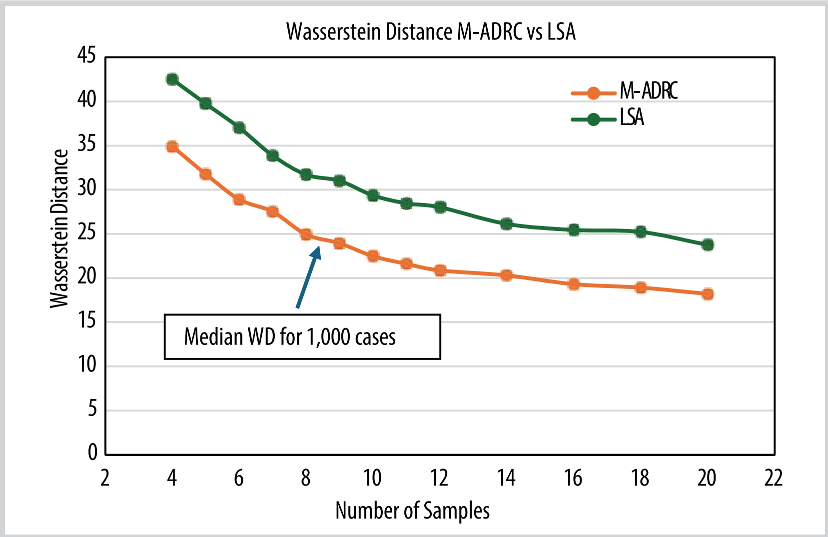

Figure 9. *Median Wasserstein distance between synthetic and actual population: M-ADRC vs LSA.*

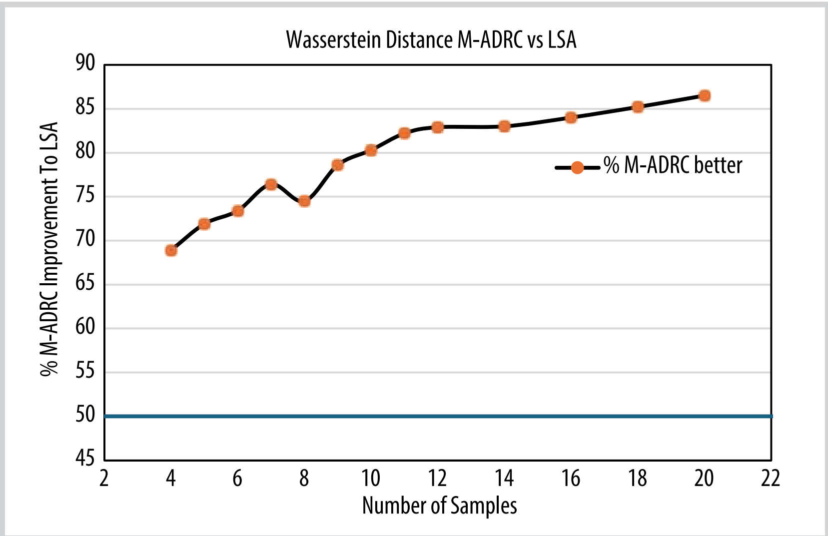

Figure 10. *Percentage of cases where M-ADRC yields improved results to LSA - Wasserstein distance.*

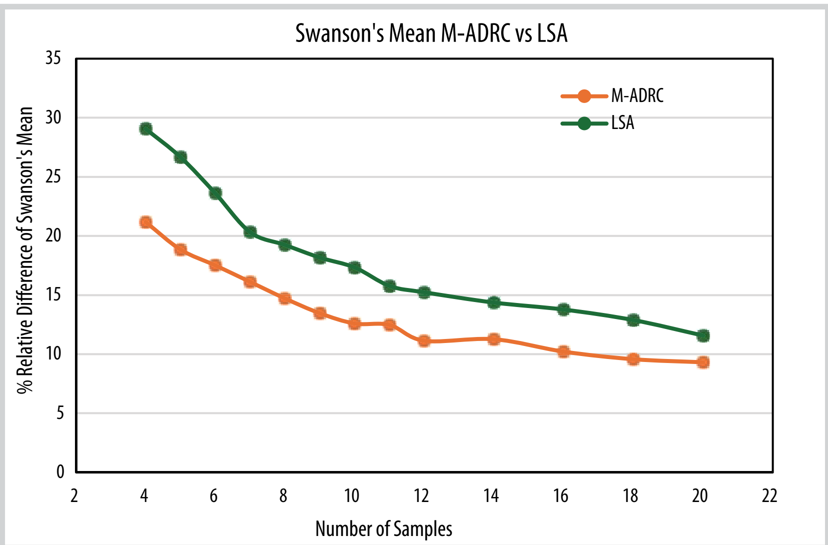

Figure 11. *Median % relative difference of Swanson's mean between synthetic and actual population: M-ADRC vs LSA.*

performance of the algorithm. Figure 13 illustrates that when input samples differ substantially from the true population distribution, the resulting synthetic population will poorly represent the actual population. Furthermore, the workflow does not take into account geospatial properties.

- Web applications

The algorithm is packaged into a web-based solution consisting of two applications (Figure 14): Metropolis sampling to create a synthetic population, and best fit to fit distribution parameters to the synthetic or actual population.

The web application offers several advantages:

+ Accuracy: Robust sampling with only six samples, reducing data requirements.

+ Scalability: Applicable to other parameters (e.g., porosity, water saturation) with similar distributions.

+ Usability: Intuitive interface accessible to non-expert users, with automated parameter estimation.

- Metropolis sampling

The web application is designed to streamline the application of the Metropolis algorithm for reservoir engineers and geophysicists, even those without extensive computational expertise. The workflow includes:

Data Input: Users upload a .csv file containing formation parameter data (e.g., net pay, porosity, water saturation). The file must include a header row and numeric columns, with missing values handled via imputation or exclusion.

Parameter Selection: Users select a numeric column (e.g., “Net_Pay”) and a distribution function (e.g., lognormal, Gaussian) from a dropdown menu.

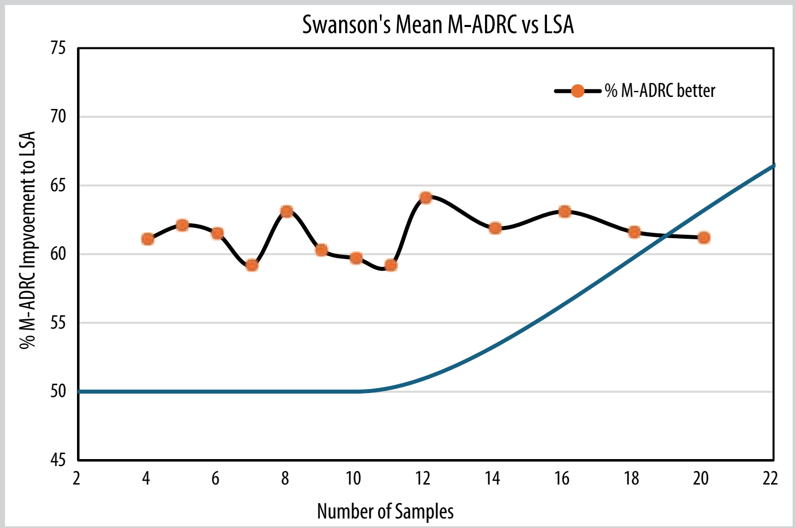

***Figure 12.** Percentage of cases where M-ADRC yields improved results to LSA - Swanson's mean.*

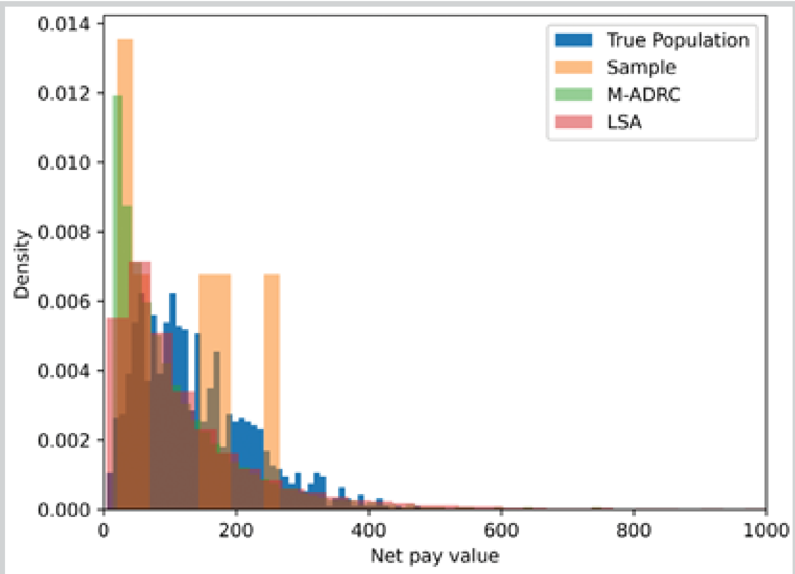

***Figure 13.** Effect of poor sampling on synthetic population generation.*

The application estimates distribution parameters (e.g., μ, σ for log-normal) using maximum likelihood estimation.

Sampling Process: A "Run Sampling" button triggers the metropolis algorithm. The application runs N iterations with a burn-in of 10% of N, generating a synthetic population of 90%N samples.

Output and Validation: Results are displayed as a histogram of the synthetic population alongside the input data distribution. Users can download the synthetic data as a .csv file for further analysis (e.g., in reservoir simulation software).

The application is built using Streamlit framework with a Python backend, ensuring scalability and ease of deployment (Figure 15).

- Best fit application

The best fit application is designed to take a population as input, fit a distribution type (normal, lognormal, triangular) to the population, generate fitted parameters, and compare the fitted distribution to the actual distribution. Figure 16 demonstrates the workflow of the application:

Data Input: Users upload a .csv file containing the population.

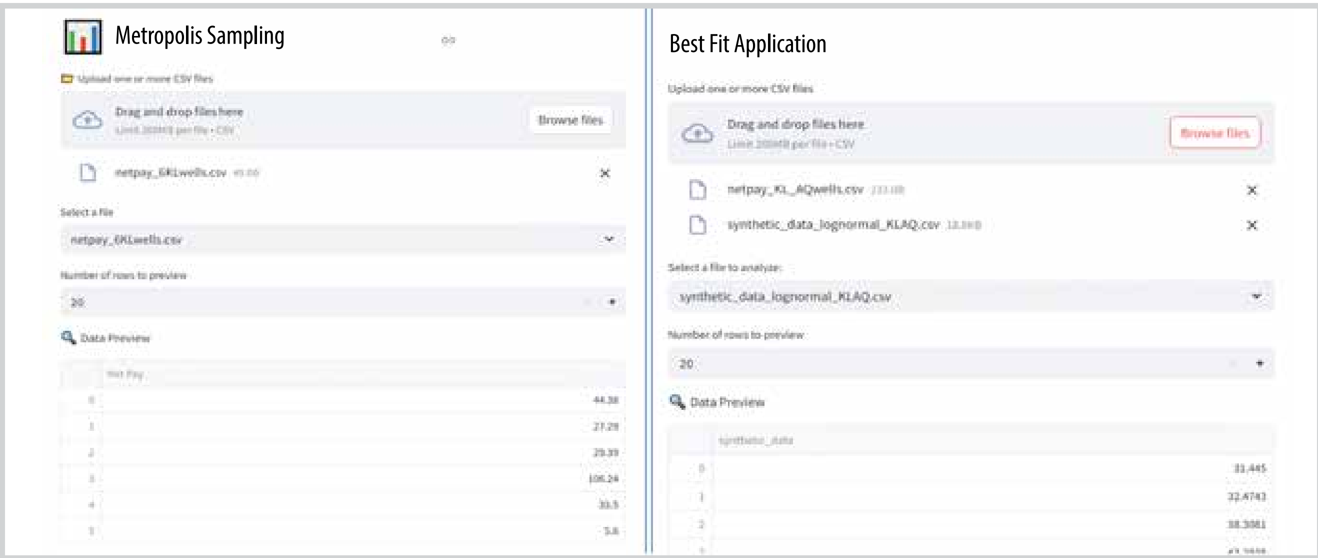

***Figure 14.** The solution consists of two applications: Metropolis Sampling and Best Fit.*

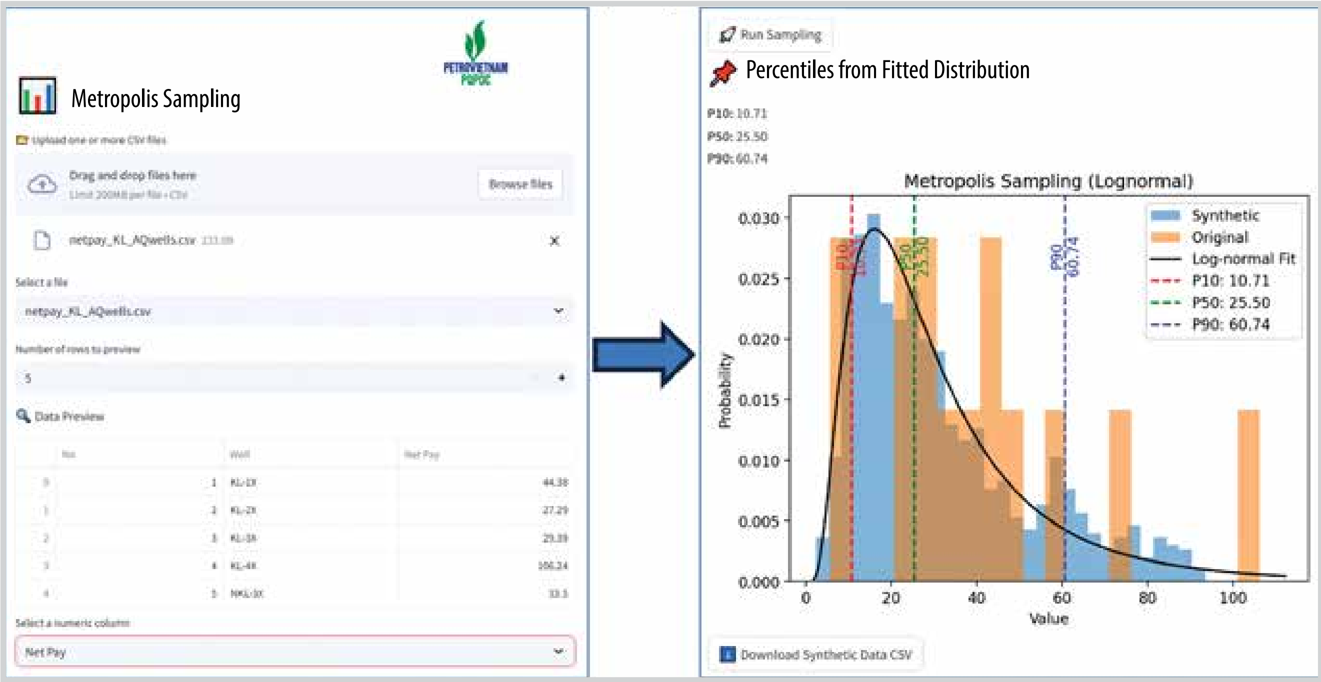

Figure 15. Metropolis Sampling output.

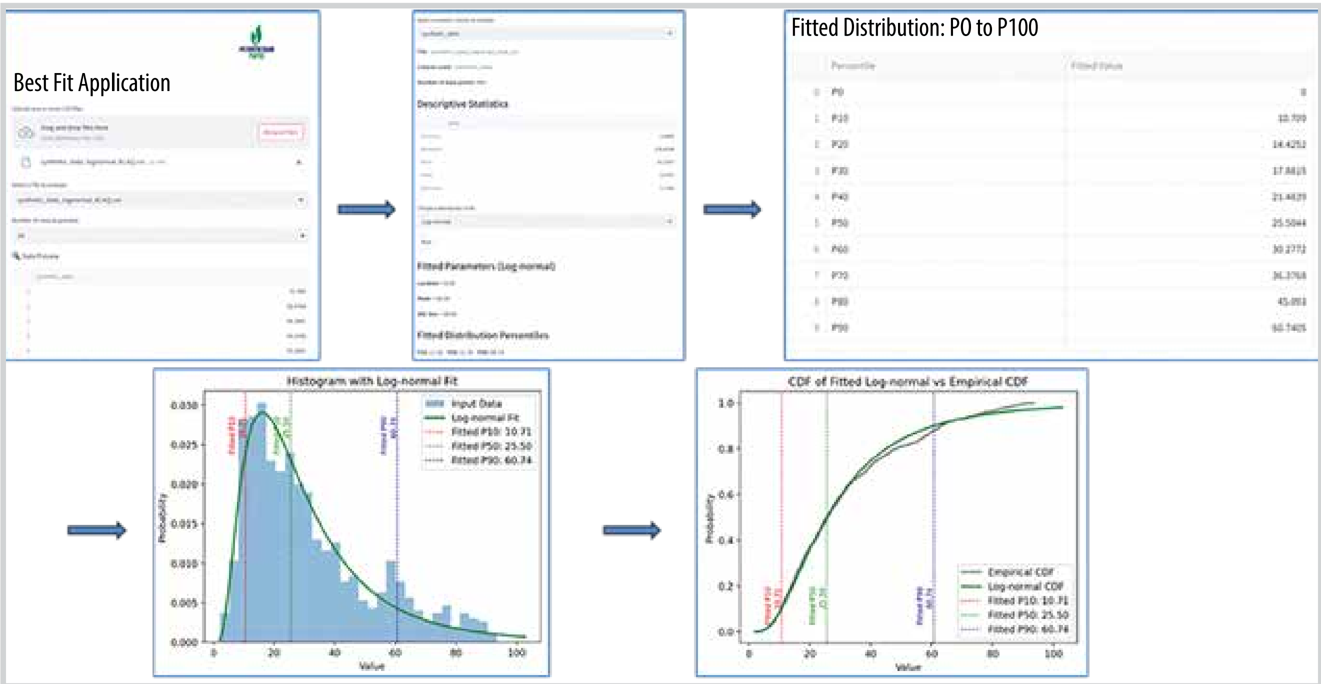

Figure 16. Best Fit Application workflow.

Parameter Selection: Users select a numeric column (e.g., "Net_Pay") and a distribution function (e.g., log-normal, normal, triangular) from a dropdown menu. The application displays descriptive statistical information and estimates distribution parameters.

Fit Process: A "Find best fit" button triggers the curve fitting process. The application finds the best fit parameters for the input population and the chosen distribution type.

Output and Validation: Results are displayed as a histogram of the input population alongside the fitted distribution. Cumulative density function (CDF) plot of the population and the fitted distribution is also presented. Users can download the output as a .csv file for further analysis.

5. Conclusions and recommendations

This study presents M-ADRC, an adaptive Metropolis-Hastings workflow with range constraints, adaptive tuning, and automatic distribution selection designed to deliver representative synthetic population from datasets with few samples. The workflow is tailored for lognormal and Pearson III distribution, which are commonly used to model reservoir properties such as porosity, net pay, saturation. Range restriction of the population relies on opinion of domain experts or analogy from similar areas. There are multiple failsafe mechanism designed to prevent infeasible results within the workflow.

Applied to net pay dataset from Southeast Asian gas fields, the M-ADRC consistently outperforms traditional lognormal spreadsheet-based approach in term of Wasserstein distance and Swanson's mean for all sample size. The workflow is lightweight, reproducible, and runs on standard desktop hardware. Key limitations include its reliance on reasonable parameter bounds, resilience to extreme bias in sampling, and restriction to two distribution families, in addition to negligence of geospatial properties. Future extensions could incorporate broader distribution families (e.g., Johnson SU), censored/error models, hierarchical pooling across fields, spatial correlation, and advanced samplers.

References

[1] Taiwo Oludele Osunrinde, Iloka Chiamaka, and Yawale Ahmad, "Field developmental plan analysis: a case study of 'x' reservoir", *Journal of Petroleum Exploration and Production Technology*, Volume 9, pp. 2185 - 2203, 2019. DOI: 10.1007/s13202-019-0622-0.

[2] C. Karacaer and M. Onur, "Analytical probabilistic reserve estimation by volumetric method and aggregation of resources", *SPE Hydrocarbon Economics and Evaluation Symposium, Calgary, Alberta, Canada*, *24 - 25 September 2012*. DOI: 10.2118/162875-MS.

[3] B.C. Craft, M.F. Hawkins, and Ronald E. Terry, *Applied petroleum reservoir engineering, second edition*. Prentice Hall PTR, 1991.

[4] Paul F. Worthington, "Net pay: What is it? What does it do? How do we quantify it? How do we use it?", *SPE Reservoir Evaluation & Engineering*, Volume 13, Issue 5, pp. 812 - 822, 2010. DOI: 10.2118/123561-PA.

[5] Tian Yang, Yingchang Cao, Yanzhong Wang, Keyu Liu, Chen He, and Shaomin Zhang, "Determining permeability cut-off values for net pay study of a low-permeability clastic reservoir: A case study of the Dongying Sag, eastern China", *Journal of Petroleum Science and Engineering*, Volume 178, pp. 262 - 271, 2019. DOI: 10.1016/j.petrol.2019.03.030.

[6] Annan Boah Evans, Abraham Aidoo Borsah, Chukwugozie Jekwu Ejeh, Asamoah Emmanuel, and Daniel Ocran, "Mapping of porosity, permeability and thickness distribution: Application of geostatistical modeling for the Jubilee oilfield in Ghana", *Scientific & Academic Publishing*, Volume 9, Issue 2, pp. 27 - 49, 2019. DOI: 10.5923/j.geo.20190902.01.

[7] Biswajit Thander, Anirbid Sircar, and G.P. Karmakar, "Hydrocarbon resource estimation: Application of Monte Carlo simulation", *IJLTEMAS*, Volume III, Issue IV, pp. 30 - 47, 2014.

[8] P.F. Worthington, "An electrical analog facility for hydrocarbon reservoirs", *SPE Annual Technical Conference and Exhibition, Dallas, Texas, 9 - 12 October 2005*. DOI: 10.2118/96718-MS.

[9] Mohammad Bakhtiyari, Jafar Qajar, Amir Torghabeh, and Ali Dehghan Abnavi, "Rock typing and uncertainty assessment in geological and petrophysical properties by integrating electrofacies, hydraulic flow units, and geostatistical techniques in the Kangan gas field, Zagros basin", *Acta Geophysica*, Volume 72, pp. 2323 - 2347, 2023. DOI: 10.1007/s11600-023-01214-1.

[10] GhojehBeyglou and MohammadHossein, "Geostatistical modeling of porosity and evaluating the local and global distribution", *Journal of Petroleum Exploration and Production Technology*, Volume 11, pp. 4227 - 4241, 2021. DOI: 11. 10.1007/s13202-021-01308-w.

[11] Mattia Aleardi, "Analysis of different statistical models in probabilistic joint estimation of porosity and litho-fluid facies from acoustic impedance values", *Geosciences*, Volume 8, Issue 11, 2018. DOI: 10.3390/geosciences8110388.

[12] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller, "Equation of state calculations by fast computing machines", *The Journal of Chemical Physics*, Volume 21, Issue 6, pp. 1087 - 1092, 1953. DOI:10.1063/1.1699114.

[13] T. Cui, C. Fox, and M.J. O'Sullivan, "Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm", *Water Resources Research*, Volume 47, Issue 10, 2011. DOI: 10.1029/2010WR010352.

[14] Zhe Ban, Ali Ghaderi, Nima Janatian, and Carlos

F. Pfeiffer, “Parameter estimation for a gas lifting oil well model using Bayes' rule and the Metropolis-Hastings algorithm”, *Modeling, Identification and Control*, Volume 43, Issue 2, pp. 39 - 53, 2022. DOI: 10.4173/mic.2022.2.1.

[15] Silpakorn Dachanuwattana, Wei Yu, and Kamy Sepehrnoori, “An efficient MCMC history matching workflow using fit-for-purpose proxies applied in unconventional oil reservoirs”, *Journal of Petroleum Science and Engineering*, Volume 176, pp. 381 - 395, 2019. DOI: 10.1016/j.petrol.2019.01.070.

[16] Sufia Khatoon, Jyoti Phirani, and Supreet Singh Bahga, “Accelerated Bayesian inference-based history matching of petroleum reservoirs using polynomial chaos expansions”, *Inverse Problems in Science and Engineering*, Volume 29, Issue 13, pp. 3086 - 3116, 2021. DOI: 10.1080/17415977.2021.1973455.

[17] Zhen Zhang, Xupeng He, Marwah AlSinan, Hyung Kwak, and Hoteit Hussein, "Robust method for reservoir simulation history matching using bayesian inversion and long-short-term memory network-based Proxy", *SPE Journal*, Volume 28, pp. 983 - 1007. DOI: 10.2118/203976-PA.

[18] Eckhard Limpert, Werner A. Stahel, and Markus Abbt, “Log-normal distributions across the sciences: Keys and clues”, *BioScience*, Volume 51, Issue 5, pp. 341 - 352, 2001. DOI: 10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.

[19] A.G. Journel, “The lognormal approach to predicting local distributions of selective mining unit grades”, *Mathematical Geology*, Volume 12, Issue 4, pp. 285 - 303, 1980. DOI: 10.1007/BF01029417.

[20] William W.S. Chen, and Samuel Kotz, “The riemannian structure of the three-parameter gamma distribution”, *Applied Mathematics*, Volume 4, Issue 3, pp. 514 - 522, 2013. DOI: 10.4236/am.2013.43077.

[21] Abdalrhman Abrahim Milad, Ibrahim Adwan, Sayf A. Majeed, Zubair Ahmed Memon, Munder Bilema, Hend Ali Omar, Maher G.M. Abdolrasol, Aliyu Usman, and Nur Izzi Md Yusoff, “Development of a hybrid machine learning model for asphalt pavement temperature prediction”, *IEEE Access*, Volume 9, pp. 158041 - 158056, 2021. DOI: 10.1109/ACCESS.2021.3129979.

[22] Jeffrey S. Rosenthal, “Optimal proposal distributions and adaptive MCMC”, Handbook of Markov Chain Monte Carlo, 2011.

[23] Babak Naghavi, James F. Cruise, and Kishore Arora, “Comparative evaluation of three estimators of the log Pearson type 3 distribution”, *Transportation Research Record 1279*, pp. 103 - 112.

[24] A. Hurst, G.C. Brown, and R. Swanson, “Swanson's 30-40-30 rule”, *AAPG Bulletin*, Volume 84, Issue 12, pp. 1883 - 1891, 2000. DOI: 10.1306/8626C70D-173B-11D7-8645000102C1865D.

[25] Caoxiong Li, Mian Lin, Lili Ji, and Wenbin Jiang, “Multiphase flow in tight sandstone: An improved application for 3D intermingled fractal model”, *Journal of Petroleum Science and Engineering*, Volume 177, pp. 403 - 414, 2019. DOI: 10.1016/j.petrol.2019.02.030.

[26] Sarah Lee, “Monte carlo simulation in petroleum engineering: A comprehensive guide to uncertainty analysis and risk assessment”, 2025. [Online]. Available: www.numberanalytics.com/blog/monte-carlo-simulation-petroleum-engineering.

[27] Li Li, Jie Yu, Tao Huang, Lina Tang, Dan Wei, Mingyu Li, and Xin Nie, “Integrated net pay cut-off evaluation workflow for tight sandstone reservoirs: A case study of the linxing gas field, ordos basin”, *Frontiers in Earth Science*, Volume 13, 2025. DOI: 10.3389/feart.2025.1488079.

[28] Kevin P. Murphy, *Machine learning: A probabilistic perspective*. The MIT Press Cambridge, Massachusetts London, England, 2012.

[29] Oracle, Crystal Ball User’s Guide, 11.1.2.4.850, 2017[Online]. Available: https://docs.oracle.com/cd/E57185_01/CYBUG/CYBUG.pdf.

[30] Mehdi Qassamipour, Elnaz Khodapanah, and Seyyed Alireza Tabatabaei-Nezhad, “A comprehensive method for determining net pay in exploration/development wells”, *Journal of Petroleum Science and Engineering*, Volume 196, 2021, DOI: 10.1016/j.petrol.2020.107849.

[31] Céline Scheidt and Jef Caers, "Uncertainty quantification in reservoir performance using distances and Kernel methods - Application to a West Africa deepwater turbidite reservoir", *SPE Journal*, Volume 14, Issue 4, pp. 680 - 692, 2009. DOI: 10.2118/118740-PA.