

Similarity measure of sedimentary successions and its application in inverse stratigraphic modeling

Taizhong Duan¹

Received: 28 October 2016 / Published online: 24 July 2017
© The Author(s) 2017. This article is an open access publication

Abstract This paper presents a unique and formal method of quantifying the similarity or distance between sedimentary facies successions from measured sections in outcrop or drilled wells and demonstrates its first application in inverse stratigraphic modeling. A sedimentary facies succession is represented with a string of symbols, or facies codes in its natural vertical order, in which each symbol brings with it one attribute such as thickness for the facies. These strings are called attributed strings. A similarity measure is defined between the attributed strings based on a syntactic pattern-recognition technique. A dynamic programming algorithm is used to calculate the similarity. Inverse stratigraphic modeling aims to generate quantitative 3D facies models based on forward stratigraphic modeling that honors observed datasets. One of the key techniques in inverse stratigraphic modeling is how to quantify the similarity or distance between simulated and observed sedimentary facies successions at data locations in order for the forward model to condition the simulation results to the observed dataset such as measured sections or drilled wells. This quantification technique comparing sedimentary successions is demonstrated in the form of a cost function based on the defined distance in our inverse stratigraphic modeling implemented with forward modeling optimization.

Keywords Similarity quantification · Sedimentary succession · Inverse stratigraphic modeling · Global optimization · Syntactic approach

1 Introduction

Quantitative study of sedimentary successions unavoidably involves the formal description of discrete or symbolic properties such as facies, rock texture, or structure, and until recent our ability has been still very limited on how to quantify such type properties, for instance, the difference or similarity between sedimentary facies successions from measured sections in outcrop, or drilled sections. More traditional ways to do such comparison are almost exclusively either qualitative or graphic such as a simple description “the two facies successions look very similar,” or a plot representation usually used as shown in Fig. 1. On the other hand, economically efficient recovery of natural resources such as oil and gas demands better and formal quantification of geological models such as geocellular modeling in reservoir characterization.

In hydrocarbon reservoir modeling, geostatistical methods currently dominate, to large extent due to their data conditioning capacity, that is, the model can honor the observed dataset easily. In contrast, forward stratigraphic modeling has yet to be accepted as the major modeling technique in reservoir facies modeling as it seems it should have. Forward stratigraphic modeling is geological process based and is more relevant to petroleum reservoirs (Bosence and Waltham 1990; Granjeon and Joseph 1999; Griffiths et al. 2001), and this method has been developed since the 1960s (Harbaugh and Bonham-Carter 1970), compared to the geostatistics also developed since the 1960s (Matheron 1962, 1989). One of the main reasons

✉ Taizhong Duan
duantz.syky@sinopec.com

¹ Petroleum Exploration and Production Research Institute, Sinopec, 31 Xueyuan Road, Haidian District, Beijing 100083, China

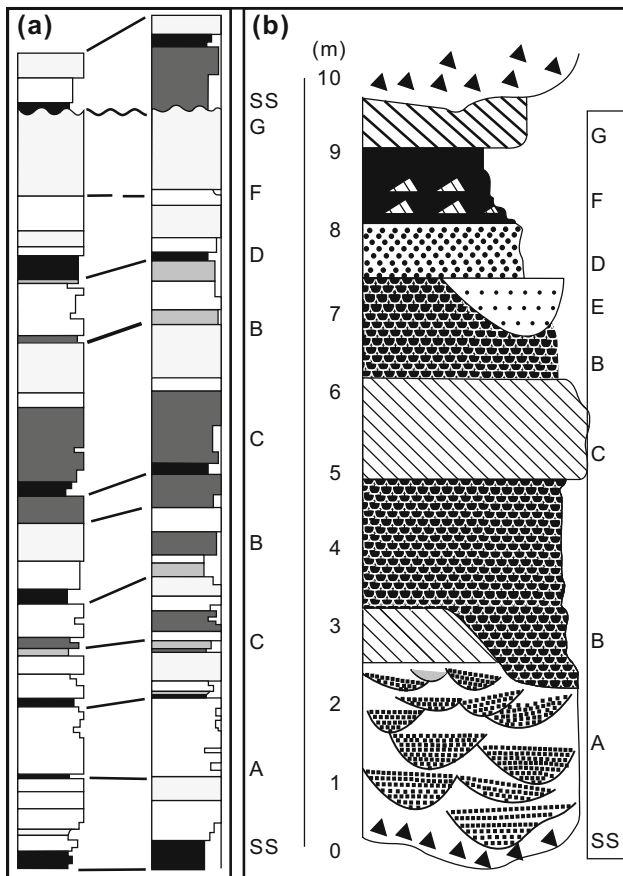


Fig. 1 Typical graphic representation of sedimentary facies succession. *Patterns and gray scale* represent different facies: **a** two “very similar” measured facies sections (modified from Kerans et al. 1994); **b** a plot channel model (modified from Cant and Walker 1976), with ordered facies codes to form a string presentation of the channel facies: SSABCBEDFG (right) or SSACBCBDFGSS (left)

why this technique has been delayed to dominate in petroleum reservoir modeling is its inability to implement data conditioning. Since late 1990s, similar techniques but under different names were proposed to overcome the inability and initiated a new research front in computational stratigraphy and sedimentology, such as inverse stratigraphic modeling (ISM) (Griffiths et al. 1996; Lessenger and Cross 1996; Cross and Lessenger 1999; Duan et al. 2001a; Imhof and Sharma 2006; Charvin et al. 2009; Griffiths 2009; Charvin et al. 2011), adaptive modeling (Duan et al. 1998), modeling optimization (Bornholdt and Westphal 1998; Wijns et al. 2003, 2004), or model calibration (Falivene et al. 2014). However, the progress of these techniques, all of which will be called ISM afterward for simplicity, has been limited, and one of the major hurdles is still the data conditioning.

Therefore, in current version of ISM as shown in Fig. 2, among others, a critical technique needed to enhance the procedure is how to quantify the similarity or distance between simulated results and the observed dataset. That is,

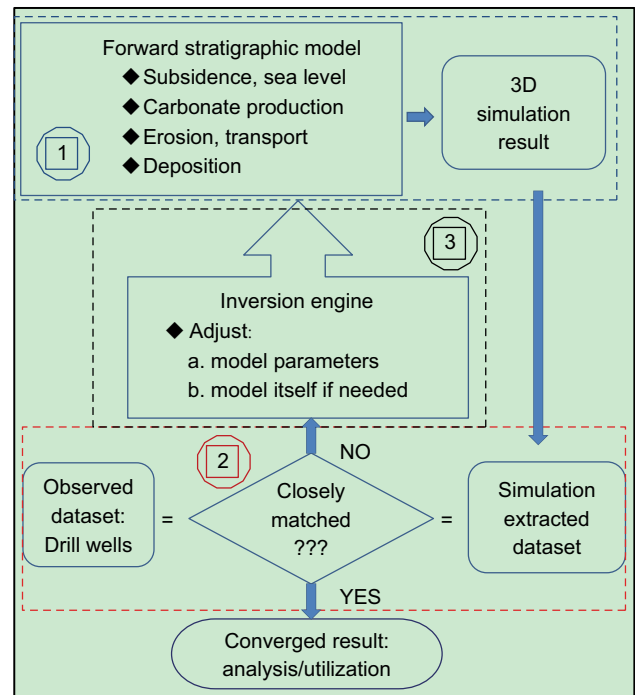


Fig. 2 ISM core techniques and workflow: 1 forward modeling generates 3D simulations; 2 comparing technique quantifies the matching between the simulated and observed datasets; 3 inversion engine updates new simulations for a better match

a proper distance measure can speed up inversion and make it more robust and can lead to better practical application of geological process-based modeling in general. This paper presents a unique and formal method of quantifying the distance between sedimentary facies successions based on a discrete, or symbolic computing technique, combined with other numerical techniques. The formal definition of the distance measure between sedimentary facies successions will be presented first, and then, its application as the cost function in ISM will be demonstrated.

2 Definition of sedimentary facies successions distance

2.1 Formal representation of sedimentary facies successions

In order to quantify the distance or similarity between sedimentary facies successions, it is essential to define what a sedimentary facies succession is and what is the distance between the sedimentary facies successions or the sedimentary facies successions distance (SFSD) formally and quantitatively. A gene-typing technique was proposed for correlation of petrophysically derived numerical lithologies between boreholes (Bakke and Griffiths 1989; Griffiths and Bakke 1990). A syntactic methodology

developed in pattern recognition (Fu 1982) was first proposed to formally describe the language of sedimentary rocks (Griffiths 1990), due to its ability in characterizing the naturally discrete or symbolic feature of sedimentary facies. The more detailed syntactic approach to the analysis of sedimentary successions for reservoir characterization can be found in Duan, Griffiths, and Johnsen (Duan et al. 1999, 2001a), among others. In the syntactic approach, a sedimentary facies succession (one-dimensional vertical stratigraphic section) is represented in a string of facies symbols, each of which contains one or more attributes, i.e., a string with attributes in symbolic computation language (Duan et al. 2001a). For instance, the bottom to top facies type of a typical channel sedimentary facies succession can be coded by symbols as shown in Fig. 1b, whereas the code of each facies can be associated with a number or an attribute representing the thickness of each facies in the parentheses as in the following attributed string of codes:

SS(0.5) A(3.0) C(0.5) B(2.5) C(1.5) B(1.3) E(0.5) D(0.4)
F(1.5) G(0.3) SS(0.1)

Besides the normal facies, the erosional surfaces are treated as a special facies. SS(0.1) and SS(0.5) mean two erosional surfaces with 0.1 and 0.5 million years time gap, respectively, whereas A(3.0) and C(0.5) mean 3.0-m trough-cross-bedded sandstone facies and 0.5-m tabular-cross-bedded sandstone facies, respectively. Of course, each code may have more than one attribute, either numerical or symbolic, such as grain size, color, fossils, and mineral composition. It should be noticed that facies coding and attributing can be compensated with each other. For instance, if you already account for grain size in facies coding such as conglomerate, coarse sandstone, fine sandstone, siltstone, and mudstone, you may not need to repeat the same information by adding a grain size attribute for the above facies code.

The similar approach was first proposed for stratigraphic correlation between drilling wells or outcrop sections, which suffered from being incapable of handling the facies change problem between sections, and has been almost forgotten in the geological community. The so-called facies change dilemma in automatic strata correlation is that two sections of strata with the same facies can be defined equivalent, and two sections of strata with totally different facies also need to be defined equivalent if facies change occurs between, which is not mathematically sound. However, in stratigraphic inversion, similarity quantification between simulated and observed successions naturally avoids the facies change problem, two successions (simulated and observed) to be compared are fundamentally the same. As we understand, the application of this technique to the stratigraphic inverse modeling is like the similar technique's application to the comparison of

DNA or RNA in biology. Therefore, we believe this technique has great potential in improving stratigraphic inverse modeling.

2.2 Definition of distance between sedimentary facies successions

To define a distance or similarity measure between sedimentary facies successions, it is important to understand what characters are essential in distinguishing different sedimentary facies successions and what would be fundamental requirements for a distance definition mathematically.

From the point of view of sedimentology and stratigraphy, comparison of sedimentary facies successions should account for following aspects: (1) facies types and their division in sections, including special facies such as erosional surfaces; (2) the thickness of each identified and divided facies (maybe repeated) in sections; and (3) the vertical order or sequence of the coded facies in each sedimentary facies successions. When it is said that two sedimentary facies successions are equivalent, it means all three of the above-mentioned characters should be the same. For instance, following are two code strings X and Y from neighboring sedimentary successions:

X: SS(0.5) A(3.0) C(0.5) B(2.5) C(1.5) B(1.3) E(0.5)
D(0.4) SS(0.15)
Y: SS(0.5) A(3.0) C(0.5) B(2.5) C(1.5) B(1.3) E(0.5)
D(0.4) F(1.5) G(0.3) SS(0.1)

They are not equal, because firstly the string-X lacks "F(1.5) G(0.3)," implying that part of the channel top deposits may be eroded away. Secondly, its top SS attribute 0.15 is different from the string-Y's 0.1, implying that the time gap represented by the SS of the string-X is 0.05 million years longer than the string-Y's.

For the following two strings:

U: B(2.0) D(2.0) F(2.0) and
V: F(2.0) D(2.0) B(2.0)

They are not equal, because the vertical order of the facies code is different, though three facies types and their thickness are the same. Actually, string-U may represent upper channel deposits, while string-V may represent crevasse splay deposits.

A distance measure of sedimentary facies successions was first proposed in a syntactic approach (Duan et al. 2001a) which accounts for all three aspects of the above-mentioned characters in sedimentary succession comparison. It is based on a series of syntactic distance measures such as the Levenshtein distance (Levenshtein 1966), generalized Levenshtein distance (Fu 1986), and distance between attributed strings (Fu 1986). For details of the definition, refer to Duan et al. (2001b) from Definition 1

through to Definition 5 and related concepts. However, for continuation and readability, the main definition is reproduced as follows.

Definition 1 Let x and y be two attributed strings,

$$x = a_1 a_2 \dots a_n$$

$$y = b_1 b_2 \dots b_m$$

Corresponding attributes of x and y are denoted as:

$$x' = (a_1^1 a_1^2 \dots a_1^k) (a_2^1 a_2^2 \dots a_2^k) \dots (a_n^1 a_n^2 \dots a_n^k)$$

$$y' = (b_1^1 b_1^2 \dots b_1^k) (b_2^1 b_2^2 \dots b_2^k) \dots (b_m^1 b_m^2 \dots b_m^k)$$

It is assumed that each terminal symbol has k attributes. The distance between x and y is defined as:

$$d^{AS}(x, y) = \alpha d^{GL}(x, y) + \beta d^A(x, y) \text{ (or } \beta d^{A'}(x, y))$$

where α and β are two positive weights; $d^{GL}(x, y)$ is the generalized Levenshtein distance between x and y ; and $d^A(x, y)$ is the attribute distance between x and y after optimal alignment with only insertion (when substitution is accepted, a variant $d^{A'}(x, y)$ is obtained) is carried out to make one string equal another syntactically. Let the transferred string be:

$$x_t = c'_1 c'_2 \dots c'_k (y_t)$$

$$= c''_1 c''_2 \dots c''_k \{ \max\{|x|, |y|\} \leq k \leq (|x| + |y|) \}$$

Then

$$d^A(x, y) = \sum \lambda_j d(A(c'_j), A(c''_j)) \quad j = 1, 2, \dots, k$$

where $A(c'_j)$ and $A(c''_j)$ are attribute vectors of c'_j and c''_j , respectively; $d(A1, A2)$ can be any p -norm distance ($p = 1$ used in our case); λ_j are weighting coefficients defined as:

1. $\lambda_j = \lambda_C(c'_j, c''_j) = 1 \ c'_j = c''_j$ (continuation)
2. $\lambda_j = \lambda_I(c'_j, c''_j) \geq 1 \ c'_j \neq c''_j$ (insertion)
3. $\lambda_j = \lambda_S(c'_j, c''_j) \geq 1 \ c'_j \neq c''_j$ (substitution)

Note that when insertion occurs, the inserted symbol's attributes are assigned zero. If substitution is allowed and occurred, attributes remain the same in related symbols.

As Duan et al. (2001b) pointed out, the distance measure defined must satisfy following mathematical requirements (Kaufman and Rousseeuw 1990):

- D(1): $d^{AS}(x, y) \geq 0$
- D(2): $d^{AS}(x, x) = 0$
- D(3): $d^{AS}(x, y) = d^{AS}(y, x)$
- D(4): $d^{AS}(x, z) \leq d^{AS}(x, y) + d^{AS}(y, z)$;

For instance, the following calculation demonstrates the Rule D(4) validation in our SFSD, whereas it is more obvious to validate the rule D(1) to D(3). Let:

$$x: a(9)b(7)c(3)$$

$$y: a(2)b(7)c(3)$$

$$z: b(7)c(3)$$

$$\alpha = 1, \beta = 1$$

$$\omega(i, j) = 0 \ i = j$$

$$\omega(i, j) = 1 \ i \neq j$$

$$\omega(k, 0) = \omega(0, k) = 1; \ i, j, k = \{a, b, c\}$$

$$\lambda_j = \lambda_C(c'_j, c''_j) = 1 \ c'_j = c''_j \text{ continuation}$$

$$\lambda_j = \lambda_I(c'_j, c''_j) \geq 1 \ c'_j \neq c''_j \text{ insertion}$$

$$\lambda_j = \lambda_S(c'_j, c''_j) \geq 1 \ c'_j \neq c''_j \text{ substitution}$$

Then, the minimum-length alignment is

$$x_t [a \ b \ c] \ x'[9 \ 7 \ 3]$$

$$y_t [a \ b \ c] \ y'[2 \ 7 \ 3]$$

$$z_t [a \ b \ c] \ z'[0 \ 7 \ 3]$$

and distances are

$$d^{AS}(x, y) = \alpha d^{GL}(x, y) + \beta d^A(x, y)$$

$$d^{GL}(x, y) = 0; \ d^A(x, y) = |2 - 9| + |7 - 7| + |3 - 3| = 7;$$

$$d^{AS}(x, y) = \alpha \times 0 + \beta \times 7 = 7;$$

$$d^{GL}(x, z) = 1; \ d^A(x, z) = |0 - 9| + |7 - 7| + |3 - 3| = 9;$$

$$d^{AS}(x, z) = \alpha d^{GL}(x, z) + \beta d^A(x, z) = 10;$$

$$d^{GL}(y, z) = 1; \ d^A(y, z) = |0 - 2| + |7 - 7| + |3 - 3| = 2;$$

$$d^{AS}(y, z) = \alpha d^{GL}(y, z) + \beta d^A(y, z) = 3;$$

$$d^{AS}(x, y) + d^{AS}(y, z) = 7 + 3 \geq d^{AS}(x, z) = 10$$

where $x, y,$ and z are strings; a, b, c are terminal symbols or facies codes; $\omega(i, j)$ weights for $d^{AS}(x, y)$, and $\lambda_C(i, j), \lambda_I(i, j), \lambda_S(i, j)$ are weights for $d^A(x, y)$ in symbol continuation, insertion, and substitution operations.

A dynamic programming algorithm (Fu 1986; Duan et al. 2001a) is modified to calculate the SFSD as defined in this section.

3 Application of sedimentary facies successions distance in inverse stratigraphic modeling

Distance measures between sedimentary facies successions can be of significant application in ISM (Duan et al. 1998, 2001a), among others (Duan et al. 2001a). The ISM can be especially beneficial to reservoir evaluation in the early stage of field development when the data quality or amount is not enough for geological reservoir modeling to take advantage of geostatistics-based geomodeling techniques.

Of great importance in ISM is quantifying mismatch between simulated results and the observations as shown in Fig. 2. Our SFSD defined in previous section provides a unique formal way to measure the mismatch properly: (1) it considers both syntactic and attribute distances between

facies successions, i.e., facies type and their thickness; (2) it also naturally considers the vertical order of facies types in the succession, compares the succession as a whole, and does not need further within-succession time calibration between the simulated and observed successions; and (3) it can easily be adapted to accounting for time gaps associated with erosional surfaces as coded into the attribute.

To quantify the mismatch between simulated and observed successions, an absolute time framework needs to be established within the succession, so that facies/thickness formed in the same time interval can be compared each other. For instance, the strata formed between time-1 and time-2 in the simulated succession should be compared with the strata of the same time interval, time-1 and time-2 in the time-calibrated observed succession. The succession simulated by forward modeling very commonly contains time resolution, say, 5000 year (modeling time step), whereas the observed succession contains dated time resolution usually in million years, or at higher resolution hundreds of thousand years. The method published so far to calculate the mismatch of simulated and observed successions is to only compare the thickness of the smallest dated stratigraphic units such as a strata cycle (Cross and Lessenger 1999; Charvin et al. 2009), or the unit thickness maps (Falivene et al. 2014). Of course, in these methods, the higher resolution of time calibration the observed succession has, the more accurate the comparison of the simulated and observed successions is. However, it is almost impossible to time calibrate the observed succession with an order of modeling time step resolution. Therefore, the advantage of our method is very obvious, indicated by the properties (1) to (3) mentioned in the previous section. Moreover, the property (2) implies that the method does not need a higher-resolution chronostratigraphic framework within the observed succession to quantify the mismatch better.

The 3D carbonate forward stratigraphic model used in our ISM is energy and sediment flux based (Duan et al. 2000; Shafie and Madon 2008) and can simulate progradation, aggradation, and retrogradation of a carbonate platform simplified to account for the main factors controlling platform evolution such as basin subsidence, basement flexure, sea-level change, carbonate productivity, sediment transport, erosion, and deposition. The 2D model used is simplified from the 3D model as a tester simulating mainly the subsidence, sea-level change, carbonate productivity, and sedimentation.

3.1 Characterization of the cost function based on sedimentary facies successions distance

The sensitivity analysis of model parameters to the inversion process was run studying the landscape of the SFSD-

based cost function in 3D ISM. Firstly, “an observed dataset” of 5 pseudowell facies sections was generated by extracting from a 3D facies model (reference model) which was simulated with known model parameters by the forward model. Then, a series of forward model runs was set up with all known parameters from the reference model fixed, but one selected parameter each time that did vary systematically across its known value in the reference model. Obviously, the change of the selected parameter to the known value used in the reference model will cause the simulation to be different from the reference model. Thirdly, the cost function values can be calculated against the varying selected parameter by quantifying the difference between the simulations and the reference model with the extracted 5 pseudowell sections based on the SFSD defined in previous section. Finally, the cross-plots between cost function values and the selected parameter were created as shown in Fig. 3.

In total, there were 29 model parameters sensitivity analyzed, with representatives shown in Fig. 3. Twenty-five out of 29 parameters are sensitive to the model inversion, with most showing the typical V-shaped curve (Fig. 3a, P17 curve as an example), and a few U-shaped (Fig. 3a, P18 curve), or L-shaped (Fig. 3b). The V- or U-shaped curves behave very similarly, both with a major minimum at the true parameter value of the reference model, and if multimimums exist, the major one is much more significant than other smaller ones (also shown in Fig. 3c), which makes the convergence of model inversion much easier, and the inverted parameter values more accurate. The L-shaped curves, only a few of them, usually correspond to those model parameters, the value change of which beyond a specific limit no longer makes a contribution to the simulation results, and the true value of which is close to the limit, behaving just like a half-U curve.

The other 4 parameters seemingly insensitive to model inversion can be called as flat-shaped (Fig. 3d, P5 and P6 curves). But in fact in most cases, they are pseudoinsensitive, mainly caused by too large a sampling interval of parameter values in sensitivity analysis calculations. If high-resolution sensitivity is carried out, they would become sensitive to inversion. As shown in Fig. 3d, P5 and P6 curves will become more like P4 if their high-resolution curves are calculated.

The landscape of this cost function can be described as multimodel, stepped, probably noisy, and one-minimum dominated. This type of cost function is complex enough, but can be handled well in inversion with direct search algorithms of global optimization (Ingber and Rosen 1992; Storn and Price 1997). These features of the cost function based on SFSD have made our 3D model inversion possible with reasonably stable results.

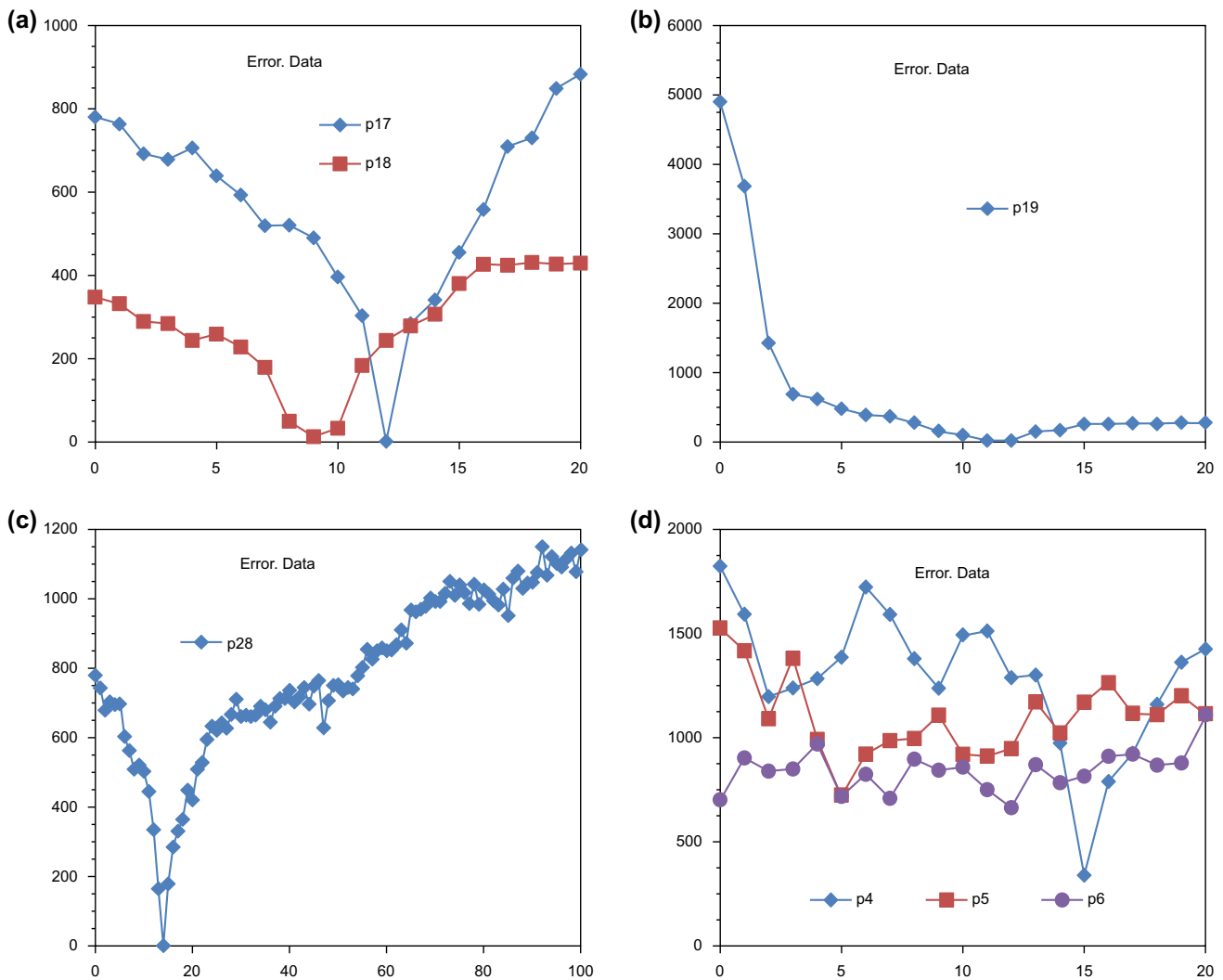


Fig. 3 Plots of cost function (Y-axis) to parameter value (X-axis) in sensitivity analysis. Function value was evaluated with systematic change of each parameter within assigned value range, while all other parameters fixed at their correct values of the reference model

3.2 Convergence behavior of stratigraphic inversion based on sedimentary facies successions distance

ISM is said converged if the cost function values become close to zero in acceptable ranges. A series of 2D and 3D ISM were run using a synthetic dataset, examining the convergence behavior of the inverse process by tackling the evaluations of the SFSD-based cost function during inversion. Both a genetic algorithm and simulated annealing were used as our inversion engine.

Our numerical experiments show that the stratigraphic inversion based on the SFSD cost function behaves very robustly for both 2D and 3D modeling as shown in Fig. 4. The behavior of the parameter inverting process can be grouped into four types. Figure 4a represents a most typical, straightforward parameter inversion in which the true value was inverted smoothly and the convergence was

relatively simple. The inversion process had correctly explored focusing around the true parameter value. Figure 4c represents the most complex convergence process. The inversion had worked on a local minimum for over half of its time and then focused on the other values a while before reaching the true value. Figure 4b represents an example between the two cases. The inversion had entered a local minimum briefly and then jumped out and started to focus on the true parameter value, though finally reached the true value after working on 5 other groups of values. Figure 4d represents a case in which even though the true value was outside of the assigned parameter range for searching, the inversion still can find the best value within the range.

The detailed analysis of convergence behavior, together with sensitivity analysis for each parameter, has helped guide the setup of model and parameters for inversion

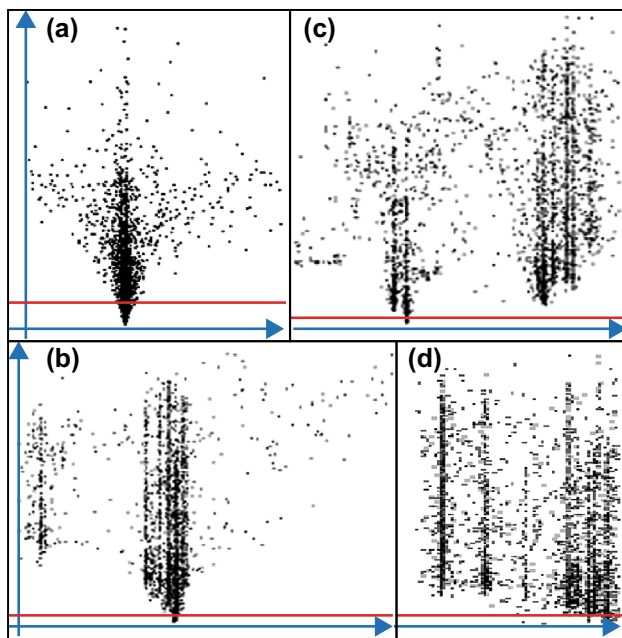


Fig. 4 Cost function value or error (Y-axis) to parameter value (X-axis) obtained in a converged inversion. Each dot represents one function evaluation during the inversion. Horizontal lines indicate the error threshold accepted for uncertainty analysis. Axis scale is normalized between 0 and 1

during our inverse modeling. For instance, those insensitive parameters in a specific time–space scale model will be excluded in inversion; more attention should be put on parameters with complex behavior in determining their searching ranges; and parameter ranges can be increased or decreased according to converging tendency from several short scoping modeling runs, so that the range is wide enough to include possible solutions, but narrow enough to speed up convergence.

3.3 Non-uniqueness of inverted results based on sedimentary facies successions distance

In a physical system such as a depositional system, non-uniqueness implies that more than one reason may cause the same consequence; for instance, an observed unconformity may be caused either by a sea-level fall or by a tectonic lift. For the given observed sedimentary facies successions, can there be more than one set of model parameters found meeting the forward stratigraphic model? A series of runs of 2D and 3D ISM were also used to study the convergence behavior regarding the non-uniqueness issue of inversion, and multiple converged results were robustly achieved. In the 2D example, six vertical sections of the synthetic strata profile (Fig. 5a) were taken as our observed dataset (Fig. 5d), and parameter values related to basin subsidence, sea-level change, and carbonate productivity were inverted. It can be shown that inversions can

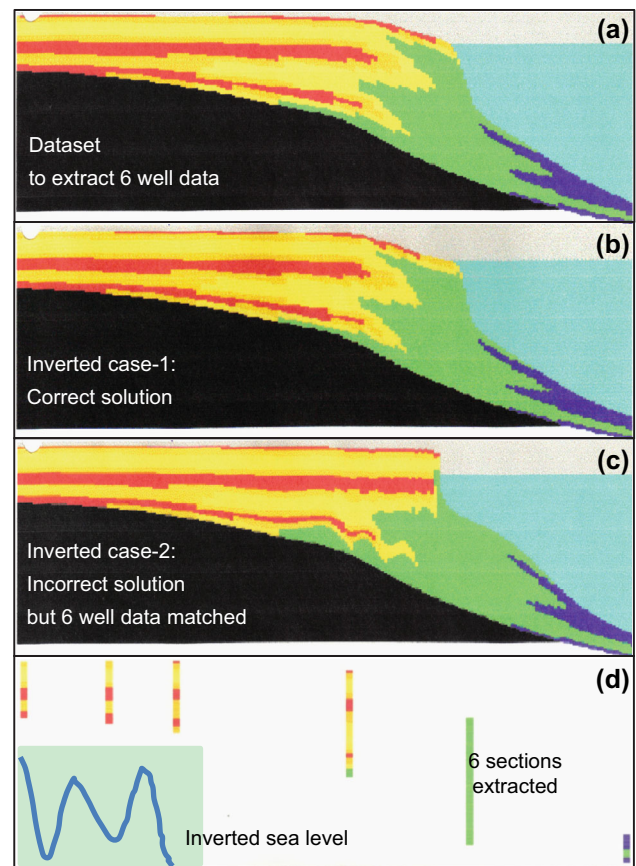


Fig. 5 Comparison of simulated results and observation (synthetic data). **a** Original carbonate dataset; **b**, **c** are two inversion results with inverted production rate shown in Fig. 6a, d, respectively; **d** extracted facies successions of 6 pseudowells as data

recover model parameter values correctly or closely with cost function errors close to zero (errors 3.0 to 8.1; zero is the minimum). There is almost no difference between the inverted strata profile and the original (Fig. 5a, b), or a difference difficult to identify graphically in many cases. For the specific parameter values, the inverted and original ones are also very close such as carbonate productivity (Fig. 6a–c). For 3D model experiments, very similar results are achieved.

However, non-unique solutions of the inversion are found as indicated in Fig. 5c, where the inverted profile is significantly different from the original though the cost function error is close to zero (3.0). The unexpected solution is caused by an unexpected inverted parameter productivity shown in Fig. 6d, which corresponds to a much higher productive rate than the original at specific water depths, causing an unusual sedimentary profile although at 6 observed locations, simulated and original are very close (converged in inversion). In practical study, this type of the non-uniqueness can be recognized easily as an incorrect solution by using additional geological information, such as the gradient of carbonate platform slope from seismic

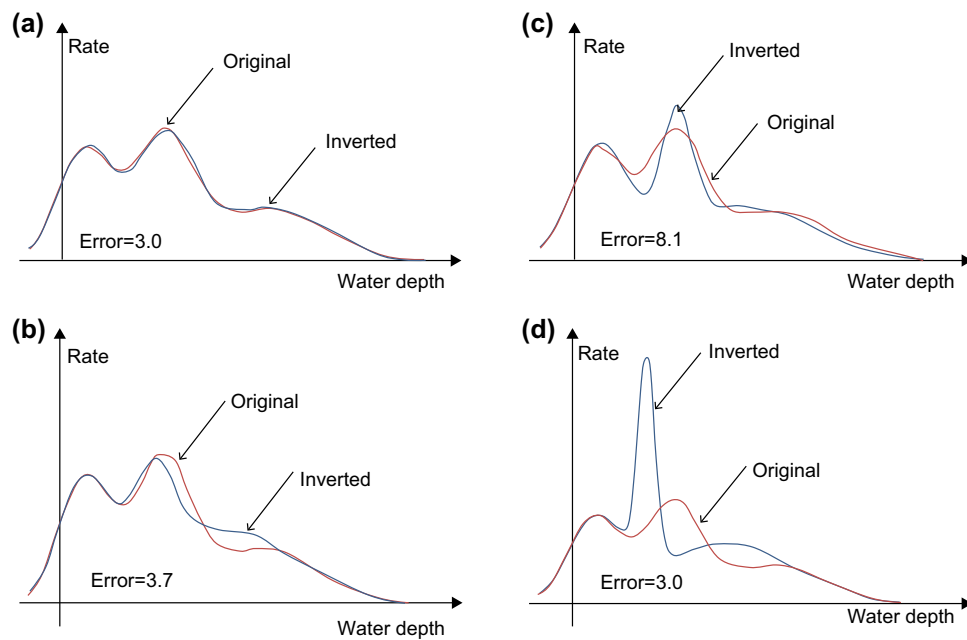


Fig. 6 Inverted carbonate production rate (Y-axis) to water depth (X-axis). **a–d** Curves corresponding to different error levels of inverted rates. *Note d* is associated with the inverted result shown in Fig. 5c. Axis scale is normalized between 0 and 1

data in our example, and it will not confuse our meaningful geological interpretation of the inversion results. However, this may imply that in simpler systems, the non-uniqueness may be a more common issue in inversion.

The non-uniqueness of a model solution means the same solution can be achieved by different sets of parameter values in modeling. Replacing of one set to another set of parameter values to generate the same modeling result is called equivalent-effect parameter substitution. No such similar, obvious non-uniqueness case as seen in our 2D inversions was found in over 10 inverted example realizations in our 3D ISM starting with different initial points of parameter space. The explanation may be that, in a system complex enough as in the 3D stratigraphic modeling, non-uniqueness caused by the equivalent-effect parameter substitution is less probable, compared to a simpler system as in the 2D model case. That is, the probability that equivalent-effect parameter substitution can generate non-uniqueness in a modeling system may be inversely proportional to the complexity of the system.

4 Conclusion

(1) A distance measure of sedimentary facies successions is formally defined based on syntactic presentation of attributed strings and symbolic computation; (2) application of the distance measure used as cost function in ISM is demonstrated with synthetic datasets to be robust in terms of convergence behavior of inversion in both 2D and 3D

modeling; and (3) therefore, the distance measure or other similar ones potentially would be very useful in facies or discrete feature-based inverse geological modeling.

Acknowledgements This research financially was supported by Colorado School of Mines, and are supported by the Science and Technology Ministry of China (2016ZX05033003), China Academy of Sciences (XDA14010204) and Sinopec (G5800-15-ZS-KJB016). The Petroleum Science editors and four anonymous reviewers are thanked for their constructive comments and suggestions. WB Zhang and PQ Lian are thanked for redrawing figures.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bakke S, Griffiths CM. Interactive stratigraphic matching of petrophysically derived numerical lithologies based on gene-typing techniques. In: Collinson J, editor. Correlation in hydrocarbon exploration. London: Norwegian Petroleum Society, Graham & Trotman; 1989. p. 61–76. doi:10.1007/978-94-009-1149-9_7.
- Bornholdt S, Westphal H. Automation of stratigraphic simulations: Quasi-backward modelling using genetic algorithms. In: Masclé A, et al. editor. Cenozoic foreland basins of western Europe: Geological Society (London) Special Publications. 1998;134(1):371–9. doi:10.1144/GSL.SP.1998.134.01.17.
- Bosence D, Waltham D. Computer modeling the internal architecture of carbonate platforms. *Geology*. 1990;18(1):26–30. doi:10.1130/0091-7613(1990)018<0026:CMTIAO>2.3.CO;2.

- Cant DJ, Walker RG. Development of a braided-fluvial facies model for the Devonian Battery Point Sandstone, Quebec. *Can J Earth Sci.* 1976;13(1):102–19. doi:[10.1139/e76-010](https://doi.org/10.1139/e76-010).
- Charvin K, Gallagher K, Hampson GL, et al. A Bayesian approach to inverse modelling of stratigraphy, Part 1: method. *Basin Res.* 2009;21(1):5–25. doi:[10.1111/j.1365-2117.2008.00369.x](https://doi.org/10.1111/j.1365-2117.2008.00369.x).
- Charvin K, Hampson GL, Gallagher K, et al. Characterization of controls on high-resolution stratigraphic architecture in wave-dominated shoreface-shelf parasequences using inverse numerical modeling. *J Sediment Res.* 2011;81(8):562–78. doi:[10.2110/jsr.2011.48](https://doi.org/10.2110/jsr.2011.48).
- Cross TA, Lessenger MA. Construction and application of stratigraphic inverse model. In: Harbaugh JW, et al. editor. *Numerical experiments in stratigraphy: recent advances in stratigraphic and sedimentologic computer simulations*: SEPM. 1999;62(1):69–83.
- Duan T, Cross TA, Lessenger MA. 3-D carbonate stratigraphic model based on energy and sediment flux. In: AAPG annual convention, New Orleans, Louisiana, 2000.
- Duan T, Cross TA, Lessenger MA. Reservoir- and exploration-scale stratigraphic prediction using a 3-D inverse carbonate model. In: AAPG annual convention, Denver, Colorado, 2001a.
- Duan T, Griffiths CM, Cross TA, et al. Adaptive stratigraphic forward modeling: Making forward modeling adapt to conditional data. In: AAPG Annual convention and exhibition, Salt Lake City, Utah, 1998.
- Duan T, Griffiths CM, Johnsen SO. Conditional simulation of 2-D parasequences in shallow marine depositional systems by using attributed controlled grammar. *Comput Geosci.* 1999;25(6):667–81. doi:[10.1016/S0098-3004\(98\)00162-9](https://doi.org/10.1016/S0098-3004(98)00162-9).
- Duan T, Griffiths CM, Johnsen SO. High-frequency sequence stratigraphy using syntactic methods and clustering applied to the upper Limestone Coal Group (Pendleian, E1) of the Kincardine basin, UK. *Math Geol.* 2001b;33(7):825–44. doi:[10.1023/A:1010950814715](https://doi.org/10.1023/A:1010950814715).
- Falivene O, Frascati A, Gesbert S, et al. Automatic calibration of stratigraphic forward models for predicting reservoir presence in exploration. *AAPG Bull.* 2014;98(9):1811–35. doi:[10.1306/02271413028](https://doi.org/10.1306/02271413028).
- Fu KS. Syntactic pattern recognition and applications. Englewood Cliffs: Prentice-Hall; 1982. p. 595–6.
- Fu KS. A step towards unification of syntactic and statistical pattern recognition. *IEEE Trans Pattern Anal Mach Intell v PAMI-8* 1986; p. 398–404.
- Granjeon D, Joseph P. Concepts and applications of a 3-D multiple lithology, diffusive model in stratigraphic modeling. In: Harbaugh JW, et al. editors. *Numerical experiments in stratigraphy: recent advances in stratigraphic and sedimentologic computer simulations*: SEPM Special Publication. 1999;62(1):197–210.
- Griffiths CM. The language of rocks—an example of the use of syntactic analysis in the interpretation of sedimentary environments from wireline logs. In: Hurst A, Lovell MA, Morton AC, editors. *Geological application of wireline logs*: Geological Society of London, Special Publication. 1990;48(1):77–94. doi:[10.1144/GSL.SP.1990.048.01.08](https://doi.org/10.1144/GSL.SP.1990.048.01.08).
- Griffiths CM. What should an ideal objective function for 4D stratigraphic units look like? In: Presented at 18th Modsim, IMACS, World Congress 09 international congress on modelling and simulation, Cairns, Australia. 13–17 July 2009.
- Griffiths CM, Bakke S. Interwell matching using a combination of petrophysically derived numerical lithologies and gene-typing techniques. In: Hurst A, Lovell MA, Morton AC, editors. *Geological applications of wireline logs*. Geological Society of London. 1990;48(1):133–51. doi:[10.1144/GSL.SP.1990.048.01.12](https://doi.org/10.1144/GSL.SP.1990.048.01.12).
- Griffiths CM, Duan T, Mitchell A. How to know when you get it right: a solution to the section comparison problem in forward modelling. In: *Proceedings numerical experiments in sedimentology*. May 1996, University of Kansas.
- Griffiths CM, Dyt C, Paraschivoiu E, et al. SedSim in hydrocarbon exploration. In: Merriam DF, Davis JC, editors. *Geologic modeling and simulation*. New York, Kluwer Academic. 2001. p. 71–97. doi:[10.1007/978-1-4615-1359-9_5](https://doi.org/10.1007/978-1-4615-1359-9_5).
- Harbaugh J, Wand Bonham-Carter G. *Computer simulation in geology*. New York: John Wiley and Sons, 1970. p. 111–71.
- Imhof M, Sharma AK. Quantitative seismostratigraphic inversion of a prograding delta from seismic data. *Mar Pet Geol.* 2006;23(7):735–44. doi:[10.1016/j.marpetgeo.2006.04.004](https://doi.org/10.1016/j.marpetgeo.2006.04.004).
- Ingber L, Rosen B. Genetic algorithms and very fast simulated reannealing: a comparison. *Math Comput Model.* 1992;16(11):87–100. doi:[10.1016/0895-7177\(92\)90108-W](https://doi.org/10.1016/0895-7177(92)90108-W).
- Kaufman L, Rousseeuw PJ. *Finding groups in data: an introduction to cluster analysis*. New York: Wiley; 1990. p. 341–2.
- Kerans C, Lucia FJ, Senger RK. Integrated characterization of carbonate ramp reservoirs using outcrop analogs. *AAPG Bull.* 1994;78(2):181–216.
- Lessenger MA, Cross TA. An inverse stratigraphic simulation model—Is stratigraphic inversion possible? *Energy Explor Exploit.* 1996;14:627–37.
- Levenshtein VI. Binary codes capable of correcting deletions, insertions and reversals. *Cybern Control Theory.* 1966;10(8):707–10.
- Matheron G. *Traité de géostatistique appliquée Tome 1*, Editions Technip, Paris. 1962. p. 333–4.
- Matheron G. *Estimating and choosing*. Berlin: Springer; 1989.
- Shafie KRK, Madon M. A review of stratigraphic simulation techniques and their applications in sequence stratigraphy and basin analysis. *Geol Soc Malays Bull.* 2008;54(1):81–9.
- Storn R, Price K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Global Optim.* 1997;11(4):341–59. doi:[10.1023/A:1008202821328](https://doi.org/10.1023/A:1008202821328).
- Wijns C, Boschetti F, Moresi L. Inverse modelling in geology by interactive evolutionary computation. *J Struct Geol.* 2003;25(10):1615–21. doi:[10.1016/S0191-8141\(03\)00010-5](https://doi.org/10.1016/S0191-8141(03)00010-5).
- Wijns C, Poulet T, Boschetti F, et al. Interactive inverse methodology applied to stratigraphic forward modelling. *Geol Soc Lond Spec Publ.* 2004;239(1):147–56. doi:[10.1144/GSL.SP.2004.239.01.10](https://doi.org/10.1144/GSL.SP.2004.239.01.10).