

Development of an objective function to measure the dissimilarity between data observed in wells and results of *Stratigraphic Forward Models*-SFM

Jessica de Souza Brugognolle¹, Lindaura Maria Steffens², Mathieu Ducros³, Antonio Guilherme François Soares de Marco Grapiglia², Luis Fernando Lamas de Oliveira^{1,2}, Ana Paula Soares², Daniel Fabian Bettu², Paulo César Soares²

¹*Programa de Pós-Graduação em Engenharia Mecânica, Universidade do Estado de Santa Catarina
Paulo Malschitzki Street, 200, Joinville, 89219-710, Santa Catarina, Brazil
jessica.brugognolle@edu.udesc.br, luis.lamas@udesc.br*

²*Departamento de Engenharia de Petróleo, Universidade do Estado de Santa Catarina
Lourival Cesário Avenue, Balneário Camboriú, 88336-275, Santa Catarina, Brazil
lindaura.steffens@udesc.br, antonioguilhermo38@gmail.com, daniel.bettu@udesc.br, ana.soares@udesc.br,
p_soares@terra.com.br*

³*C6+ (Geoscience Consulting and Research)
Rua Marquês de Abrantes, 107, 22230-060, Rio de Janeiro, RJ, Brazil
mathieu.ducros@c6mais.com*

Abstract. In the area of oil production investment decisions and production planning depend on oil production predictions, which are made using reservoir simulation techniques. Simulations reproduce the flows of hydrocarbons in the reservoirs, described through a geological model. Geological models integrate data (well, seismic) and geological interpretations. Stratigraphic Forward Models (SFM) consists in modeling geological processes that considers the principles of conservation of mass, energy and reproduces processes in the study domain (sedimentary basin). SFM provides the facies distribution in the sedimentary basin, resulting from the geological processes. Once the SF modeling has been carried out, it is necessary to evaluate the results obtained and apply a calibration since the results do not honor information from the well data. In order to solve the inversion problem, several authors have developed different methodologies. These approaches provide numerous contributions of extreme relevance, but none of them presents a possible solution for operational application due to an incomplete or inadequate configuration of the objective function (generally not sufficiently geologically representative). Therefore, this work describes a possibility more adequate objective function that clearly and quantitatively translates the degree of similarity between the numerical results and the data is one of the main keys of a successful inversion. The proposed objective function was developed based on previous work, which used an approach originated from automated well correlations. Some modifications of this method were proposed to make the objective function more adapted to SFM calibration. Among these modifications are the fact that the objective function can now compare SFM results whose resolution differ greatly from the resolution of the calibration data.

Keywords: optimization, geological model, reservoir, simulation, objective function.

1 Introduction

Numerical models of quantitative simulation of a geological model were developed based on the interaction of physical functions that represent the production, transport and deposition of sediments in a sedimentary basin. Stratigraphic Forward Models (SFM) reproduces physical changes in the depositional environment during a given time interval, through the interaction of compositional parameters (quantity and type of sediments) and environmental controls (accommodation space, slope, flows and waves, etc). The simulation results from the solution of differential equations that describe the control parameters behavior. Simplified physical rules are used to simulate geological processes at larger scales instead of simulating what happens to each sedimentary grain,

like sediment diffusion laws based on fluid flow and surface slope. Its result is a robust and consistent prediction of 3D stratigraphic sequences, bathymetry and sedimentary grain proportions, among others.

Direct numerical simulation of geological processes, using SFM for instance, is usually accompanied by inverse problems. The resolution of the inverse problem is the process of estimating information about the physical object or system, which we are interested, based on the observed data. In the case of numerical modelling, it corresponds to an estimation of the input parameter values of the simulation, which allows the observed data result to be reproduced.

For SFM the observed measurement values are mainly well data and seismic data. These observed data are also called calibration data, because they guide the calibration (or adjustment) of input parameter values for *forward* numerical modeling.

There are many theoretical ways to perform the inverse process, with varying complexity. However, due to the lack of a more satisfactory method for operational studies, geologists follow a *trial and error method*. This is an iterative process which can be cumbersome.

The main problem of the resolution of the inverse problem in the case of SFM seems to be related to the definition of a consistent objective function Duan *et al.* [2]. The objective function (or cost function) is a measure of the error or distance between the simulation result and the observed data. Until now, attempts to solve the inverse problem in SFM stumbled on the definition of an objective function that would be embed sufficient geological information to provide geologically acceptable simulation results. Duan *et al.* [2] presented a promising quantitative method for calibrating forward stratigraphic models. This approach is derived from techniques of automated well correlations and embed, therefore, geological concepts. However, it relies on some assumptions (e.g. resolution of the observed data) which does not correspond to classical operational situations. Therefore, we revisited the objective function proposed by the authors to make it more adapted to real situations of stratigraphic forward modeling in oil and gas studies.

2 Method to compare observed well data with simulation results

One of the main challenges of SFM calibration is to determine an OF to quantify the similarity between the observed data and the simulation results, being geologically representative. Therefore, this work aims at developing an OF to measure and quantify the similarity between observed and simulated data for sedimentary successions, taking into account some geological knowledge and realistic resolutions of data used in oil and gas studies.

For the purpose of comparison and adjustment measures between objective information and simulations, a formal representation of the existing facies is necessary in order to obtain a quantitative stratigraphic analysis and modeling, in which each layer of facies can be represented by a symbol, in this case, a letter (alphabetic code), forming a syntactic expression or string. Each symbolic code may be accompanied by an attribute representing the thickness or other property of interest. A succession represented by lithofacies symbols is showed in Fig. 1. The sequence, in this case, would be the lithofacies symbol with its thickness (attribute) in parentheses, SS (0.0) D (2.3) C (1) B (1.3) A (0.9) B (1.7) A (2.1) SS (0.0).

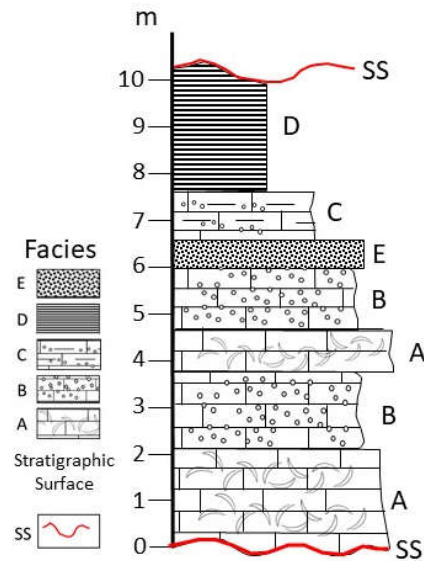


Figure 1 - A succession represented by lithofacies symbols.

2.1 Duan's approach

In their approach, Duan *et al.* [1] define a distance between sedimentary successions to compare their dissimilarity, where the definition of distance is based on attributed syntactic representation. That is, one-dimensional successions are represented by a string of facies or lithofacies symbols sequentially, being that each symbol can also have a vector of attributes that can provide other information on lithofacies such as thickness. Therefore, the distance of any two successions is then defined consisting of its syntactic and attribute subdistances. The total distance $d^{AS}(x, y)$ of the x and y sequences proposed by authors is defined as:

$$d^{AS}(x, y) = \alpha d^F(x, y) + \beta d^A(x', y') \quad (1)$$

being

$d^F(x, y)$: distance between facies of the sequences, also called Levenshtein distance;

α : weight coefficient for the distance between facies;

$d^A(x', y')$: distance between attributes of the sequences;

β : weight coefficient for the distance between attributes;

x', y' : attributes of the corresponding sequences.

Through the term d^F , in equation (1), the well logs (record of rock description along a vertical well) are compared using three possible operations to make the logs equals (substitution, exclusion and insertion of rock records). Thus, the Levenshtein distance, used to compute the dissimilarity between two well logs, is then defined by the minimum number of substitutions, exclusions and insertions needed to transform a well log into another. A generalized version of the Levenshtein distance consider a different weight for each of these operations. The distance between attributes, d^A , is given by the Euclidean distance.

Well data have generally a much higher resolution ($\sim 0,2m$) than SFM results (several meters). Moreover, the total thickness of a geological section, and its corresponding section within the numerical model can be different (and different from a simulation to another as the objective of SFM is to simulate the distribution of sediments within a sedimentary basin). These two elements make difficult the construction of an objective function as there is no obvious geologically meaningful way to perform the comparison of the simulation results with the observed data. The problem related to the difference of resolution is not dealt with by the approach proposed by Duan *et al.* [1].

2.2 Adaptation of the approach to operational studies

In order to correctly deal with operational case studies, we propose a new objective function derived from the work of Duan *et al.* [1] to geologically consider the difference of resolution between the observe data and the simulation results, and account for the difference of nature of the two rock sequences which are compared. In this new approach, the well logs transformations which are allowed to measure the distance between the well data and the simulation results are only insertion of a hiatus (a geological event of $0m$ thickness) and addition, which means the grouping of one or all the following layers of the well. Exclusion and insertion of sediment records, meaningful in the case of well correlation, is not adapted for the comparison of the result of a deterministic simulation and an observed log of well. Insertion of hiatus and grouping of observation in a same unit, within the well log record, still seem interesting to consider as we have no direct information of sedimentation rate. The grouping of several observations can generate a well unit with several facies. In that case the distance is weighted by the relative occurrence of each one of these facies. This approach offers the advantages to consider geological aspects, such as imposing a constant sedimentation rate (normalization of both the simulated an observed records to make them directly comparable) consider geologically meaningful upscaling of the sedimentological record and take into account both facies errors and sedimentation rate errors.

In the new proposal, the distance related to facies will be calculated with the values resulting from the *forward* modeling, provided by an SFM software, for the simulated data and will be used to compare with the well data. For example, if *forward* modeling simulation provides results of deposition bathymetry and sand proportion for the definition of facies model, these results will be used to generate a normalized data matrix for calculating the distances between the simulated and observed facies. In other words, the position of the simulated facies is selected in the data matrix and the closest boundary to the well facies is found. An illustration is shown in Tab.1, with simulated facies and 60% sand proportion and 3m bathymetry and well facies F (highlighted in red). Using these values, it can be calculated using the Euclidean distance. If there are more than one facies in the group, this distance is multiplied by the percentage of each facies in the group, generating a weighting for each facies.

Table 1. Euclidean distance representation between the simulated layer facies (e.g., E) and the well layer facies (e.g., F), considering the closest boundary.

	Sand proportion									
Bathymetry	0.00	0.11	0.22	0.33	0.44	0.56	0.67	0.78	0.89	1.00
0.00	A	A	A	B	B	B	B	B	B	B
0.10	C	C	E	E	E	E	E	E	E	E
0.52	F	F	F	F	G	G	G	G	G	G
1.00	F	F	F	F	F	H	H	H	H	H

Therefore, in this new approach, to calculate the distance between the facies, the term that in the Duan *et al.* [1] function referred as Levenshtein's distance, is formulated mathematically as:

$$d^F(x, y) = \sum_{i=1}^{n_F} (\%facies_i \times d_i^E(x, y)) \tag{2}$$

where

$d^F(x, y)$: distance between facies;

$\%facies_i$: percentage of facies i in the well sequence grouping;

$d_i^E(x, y)$: Euclidian distance of the facies i in the well sequence grouping compared to the facies of the simulated layer;

n_F : total number of facies in the well sequence grouping.

And, considering n attributes for facies definition, the Euclidean distance between the parameters is given by:

$$d_i^E(x, y) = \sqrt{(P1_{facies} - P1_{simulated})^2 + \dots + (Pn_{facies} - Pn_{simulated})^2}$$

In the particular case, considering two attributes (bathymetry and sand proportion, as mentioned in the illustration above):

smallest total distance. Tab. 2 displays part of these groupings for this synthetic example, with numbering entered in the first column. The other columns represent the thickness and facies grouped for each layer of the simulated data. For example, grouping 1 would have a hiatus for the first 3 simulated layers and the last layer would be the grouping of the 200 layers of the well, resulting in a thickness of 40m and facies ratio of 21.5% B; 29.0% E; 24.5% A; 20.5% H; 4.5% C. For the grouping 204, there would be a hiatus for the first simulated layer, a thickness of 0.2m of facies B for the second simulated layer, a thickness of 0.6m of facies B for the third simulated layer, and the rest (196 layers of the well) would form the fourth layer, with a thickness of 39.2m and a ratio of facies 19.9% B; 29.6% E; 25.0% A; 20.9% H; 4.6% C.

Well		
Layer	Thickness	Facies
1	0.2	B
2	0.2	B
...		
42	0.2	B
43	0.2	B
44	0.2	E
45	0.2	E
...		
199	0.2	C
200	0.2	C

Simulated		
Layer	Thickness	Facies
1	9	B
2	5	E
3	10	A
4	5.8	H

Figure 3. Test Case 2 - Data: Well data with 200 layers (left) and simulated data with 4 layers (right).

Table 2. Illustration of the possible groupings of the well data for each simulated layer (horizontal axis) for the synthetic example of well data with 200 layers and simulated data with 4 layers.

Groupings	Well data grouping							
	Simulated layer 1		Simulated layer 2		Simulated layer 3		Simulated layer 4	
	Thickness	Facies	Thickness	Facies	Thickness	Facies	Thickness	Facies
1	Hiatus	-	Hiatus	-	Hiatus	-	40	21.5% B; 29.0% E; 24.5% A; 20.5% H; 4.5% C
2	Hiatus	-	Hiatus	-	0.2	100% B	39.8	21.1% B; 29.1% E; 24.6% A; 20.6% H; 4.5% C
3	Hiatus	-	Hiatus	-	0.4	100% B	39.6	20.7% B; 29.3% E; 24.7% A; 20.7% H; 4.5% C
4	Hiatus	-	Hiatus	-	0.6	100% B	39.4	20.3% B; 29.4% E; 24.9% A; 20.8% H; 4.6% C
...								
201	Hiatus	-	0.2	100% B	Hiatus	-	39.8	21.1% B; 29.1% E; 24.6% A; 20.6% H; 4.5% C
202	Hiatus	-	0.2	100% B	0.2	100% B	39.6	20.7% B; 29.3% E; 24.7% A; 20.7% H; 4.5% C
203	Hiatus	-	0.2	100% B	0.4	100% B	39.4	20.3% B; 29.4% E; 24.9% A; 20.8% H; 4.6% C
204	Hiatus	-	0.2	100% B	0.6	100% B	39.2	19.9% B; 29.6% E; 25.0% A; 20.9% H; 4.6% C
...								
20.301	0.2	100% B	Hiatus	-	Hiatus	-	39.8	21.1% B; 29.1% E; 24.6% A; 20.6% H; 4.5% C
20.302	0.2	100% B	Hiatus	-	0.2	100% B	39.6	20.7% B; 29.3% E; 24.7% A; 20.7% H; 4.5% C
20.303	0.2	100% B	Hiatus	-	0.4	100% B	39.4	20.3% B; 29.4% E; 24.9% A; 20.8% H; 4.6% C
20.304	0.2	100% B	Hiatus	-	0.6	100% B	39.2	19.9% B; 29.6% E; 25.0% A; 20.9% H; 4.6% C
...								
1.373.700	40	21.5% B; 29.0% E; 24.5% A; 20.5% H; 4.5% C	Hiatus	-	Hiatus	-	Hiatus	-

For this example, the layers of the well data were grouped in all possible ways (1,373,701), also considering hiatus. Fig. 4 shows the results of the well data grouping, the simulated data and the lowest objective function value resulting from this grouping. The number of combinations increases factorially with the number of simulated cells and exponentially with the number of well data, which can make the method impractical for operational purposes (cases which can correspond to more than 1E40 combinations). Therefore, it is necessary to look for alternatives ways to reach the value of the objective function.

Well		Simulated	
Thickness	Facies	Thickness	Facies
9.0	95.6% B; 4.4% E	9	B
11.0	100% E	5	E
10.0	2% E; 98% A	10	A
10.0	82% H; 18% C	5.8	H

Total distance = 10.422

Figure 4. *Test Case 2 - Results*: well layers grouped as a result of the proposed method (left), layers resulting from the simulation (right) and OF value of the total distance (bellow).

Thus, alternatives to optimize the performance of the FO calculation need to be investigated, so that the approach proposed here can be computationally executable. A possible alternative to improve this troublesome can be given with the application of optimization algorithms to find the groupings of well data that minimize the value of the OF or applying some algorithms related to graph theory.

4 Conclusions

Published papers on the calibration of SFM show that the measurement of dissimilarity between the simulated and observed model is one of the key points to obtain good results. Therefore, it is necessary to focus the efforts on the proper definition of facies and the objective function, in order to advance in the development of an operational inversion approach.

The results obtained so far are preliminary, as the tests have been carried out using synthetic examples. However, these results are already promising because they have led to a geologically meaningful comparison between the observed data and the well data.

However, many tests still need to be performed. Firstly, to improve make possible the calculation of the OF for realistic cases ($>1E40$ combinations).

From a more geological point of view, it will be necessary to study the weights (α and β) influence, as this can lead to different solutions of the inverse problem, since α takes more into account the difference between the facies and the β gives more relevance to the difference between attributes. Therefore, it is necessary to find the balance that best represents this result from a geological point of view. One way to estimate the best values of the weight could be through the analysis and ranking according to the FO of several simulations obtained by SFM modeling and comparison with classification by geologists.

Acknowledgements. The authors would like to thank FAPESC, for the financial support through the PAP/UDESC/FAPESC program. The students Antonio Guilherme François Soares de Marco Grapiglia and Jessica de Souza Bruggnolle would also like to thank PROIP/UDESC-PIPES for the undergrad scholarship for Science Introduction and PETROBRAS for master scholarship, respectively.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

- [1] DUAN, T.; GRIFFITHS, C. M.; JOHNSON, S. O. 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.*, Ed. 7, 33, p. 825–44, 2001.
- [2] DUAN, T. Similarity measure of sedimentary successions and its application in inverse stratigraphic modeling. *Petroleum Science*, 14, p. 484–492, 2017.