



HAL
open science

Oil production uncertainty assessment by predicting reservoir production curves and confidence intervals from arbitrary proxy responses.

Gaétan Bardy, Pierre Biver, Guillaume Caumon, Philippe Renard

► To cite this version:

Gaétan Bardy, Pierre Biver, Guillaume Caumon, Philippe Renard. Oil production uncertainty assessment by predicting reservoir production curves and confidence intervals from arbitrary proxy responses.. Journal of Petroleum Science and Engineering, 2019, 176, pp.116-125. 10.1016/j.petrol.2019.01.035 . hal-02136465

HAL Id: hal-02136465

<https://hal.univ-lorraine.fr/hal-02136465v1>

Submitted on 24 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



1 **Oil production uncertainty assessment by predicting reservoir production curves and confidence**
2 **intervals from arbitrary proxy responses.**

3 **Gaétan Bardy**^{1,2}(gaetan.bardy@gmail.com), **Pierre Biver**²(pierre.biver@total.com), **Guillaume**
4 **Caumon**¹(guillaume.caumon@univ-lorraine.fr) and **Philippe Renard**³(philippe.renard@unine.ch)

5 1 – GeoRessources, Université de Lorraine-CNRS-CREGU, ENSG, Rue du doyen Marcel Roubault,
6 F-54518 Vandoeuvre-lès-Nancy (France)

7 2 – CSTJF Total S.A. Avenue Larribau 64018 Pau Cedex (France)

8 3 – Centre for Hydrogeology and Geothermics, University of Neuchâtel, 11 rue Emile Argand, 2000
9 Neuchâtel (Switzerland)

10 **Abstract:**

11 Underground fluid flow in hydrocarbon reservoirs (or aquifers) is difficult to predict accurately due to
12 geological and petrophysical uncertainties. To quantify that uncertainty, several spatial statistical
13 methods are often used to generate an ensemble of subsurface models representing and sampling these
14 uncertainties. However, to predict the uncertainties in terms of flow responses, one needs to run a
15 forward flow simulator (often multiphase flow in transient state) on every model of this ensemble and
16 this generally entails intractable computational costs. Approximate solutions (flow proxies) can help
17 addressing this challenge but introduce physical simplifications whose impact on the uncertainty
18 quantification is difficult to characterize. This paper proposes a workflow to assess the dynamic
19 reservoir behavior uncertainties from an input ensemble of realizations sampling geological and
20 geophysical uncertainties. Analytical reservoir production curves are estimated from proxy distances
21 computed between all ensemble members and from a few accurate flow responses computed on a
22 subset of the ensemble. A randomization process accounting for proxy quality and for model selection
23 is used to assess confidence intervals about reservoir production quantile curves. The process can use
24 both static and dynamic proxies and also permits to compare their efficiency. A case study on a real
25 turbiditic reservoir shows the applicability of the method, and highlights the value of even a simple
26 proxy to increase the confidence about future reservoir production.

27 **Keywords:**

28 Uncertainty Quantification, Model Ranking, Geostatistics, Ensemble Methods.

29 **Highlights:**

- 30 • Minimization procedure to determine parameters of an analytical oil production profile
31 • Minimization is guided by the model distances computed on dynamic or static reservoir proxy.
32 • Analytical profiles used with real production profiles to assess uncertainties on oil production

33



1 Introduction

Dynamic reservoir simulations are used to help decision making in the oil industry. Advances in computational methods make complex flow simulation feasible nowadays on detailed deterministic reservoir models honoring subsurface data (e.g., Obi et al, 2014; Wang et al, 2015). However, there is often a practical gap between the number of flow simulations that engineers would need to run to solve a field management problem and the actual number of simulations that can be managed with available computational resources. Indeed, a large number of simulations are needed both to test multiple production scenarios and to account for subsurface uncertainty (see for instance Caers, 2011; Corre et al., 2000; Fetel and Caumon, 2008; Gross, 2012; Manceau et al, 2002; Maschio et al. 2010; Schiozer et al., 2004; Subbey et al., 2004, Zabalza-Mezghani et al., 2004). In this paper, we focus on the specific problem of capturing efficiently the uncertainties on the temporal evolution of field production under a fixed reservoir production scenario. The knowledge of these uncertainties may affect development decisions of green fields (reservoirs not yet in production where no historical data are available), and also provides important information for narrowing the search space during production history matching of brown fields (mature fields with several years of production history). As geological and geophysical interpretations may take many forms, we choose to incorporate this expert interpretive knowledge using a large ensemble of detailed static 3-D reservoir models.

Accurate flow simulation on each of these models being computationally prohibitive, one may need to select only a few models from the initial ensemble to make flow forecasts. However, statistics derived only from a few samples are not robust. This calls for carefully selecting a set of representative models from the full ensemble using some prior ranking of models (Ballin et al., 1992; Deutsch, 1998; Majdi Yazdi and Jensen, 2014; Rahim et al., 2015) and using specific methodologies such as bootstrap to assess confidence intervals around reservoir responses (Scheidt and and Caers, 2010). To-date, several approximate solutions (proxies) have been proposed to help quickly select representative models. Upscaling computes equivalent petrophysical properties at a coarser scale than the initial detailed model (Durlafsky, 2005). Alternatively, simplified physics such as streamline simulation (Thiele et al., 1996) may be used on the fine-scale reservoir models. The computational cost of upscaling and streamline simulation and the relative complexity of these methods have motivated even simpler approximations of the dynamic reservoir response based on connectivity (Alabert and Modot, 1992; Ballin et al., 1992; Deutsch, 1998; Renard and Allard, 2013) or Fast Marching methods (Hovadik and Larue 2011; Xie et al., 2015). To investigate how far it is possible to go with the simplification (and therefore with the reduction of the computing cost) we can even consider using an extremely simple proxy such as the Stock Tank Original Oil In Place (STOOIP). In this case, we will obtain only a scalar value and will not get an approximate recovery curve from this computation, but we will still be able to use it as a proxy in the proposed methodology.

In all these cases, the surrogate or proxy responses cannot or should not be used directly to characterize production and the corresponding uncertainties. Instead, it should somehow be calibrated against accurate flow simulations on a few models. Indeed, inaccuracies in the proxy results may lead to biased reservoir forecasts and a possible underestimation of the associated uncertainties (Doherty and Christensen, 2011; Josset and Lunati, 2013; Josset et al., 2015).

For this reason, combining proxy evaluation on all models and accurate flow simulation on a few models seems to provide a good compromise to obtain representative uncertainty assessment while harnessing computational costs (Doherty and Christensen, 2011; Effendiev et al., 2009; Ginsgourger et al., 2013; Josset and Lunati, 2013; Josset et al., 2015; Maschio and Schiozer, 2014; Scheidt and Caers, 2009; Scheidt et al., 2011). Yet, the choice of a particular proxy can be consequential in practical case



1 studies, and it is difficult to assess which proxy provides the best compromise between the
2 performance and the quality of uncertainty assessment, which may vary from one case study to
3 another.

4 In this work, we build on a class of methods which exploit distances between reservoir models,
5 initially published by Suzuki and Caers (2008). These distances can be defined from many possible
6 proxies, including simple scalar evaluations such as STOOIP, and do not explicitly call for a particular
7 type of parameterization of the static models. For instance, these approaches can be applied to
8 reservoir models of different geometries (Suzuki et al., 2008). The core idea is to define a distance
9 which is correlated to the flow response, then to use this distance to explore the model space
10 efficiently. Based on these concepts, Scheidt and Caers (2009) proposed a general model selection
11 technique which forms the basis of the present paper and of other recent works dealing with reservoir
12 uncertainty assessment (Josset and Lunati, 2013; Josset et al., 2015a; Scheidt and Caers, 2010; Scheidt
13 et al., 2018) and history matching (Ginsbourger et al., 2013; Josset et al., 2015b; Scheidt et al., 2011).
14 Essentially, this class of approaches computes distances from proxies between all possible models,
15 then uses Multi-Dimensional Scaling (MDS) to map models in a feature space where clustering and
16 model selection is performed. Accurate flow simulation is then performed on this small subset to
17 assess reservoir performance uncertainty. To account for limited sample size, Scheidt and Caers
18 (2010) propose a bootstrap method to assess the uncertainties about reservoir responses derived from
19 these small set of samples. Josset and Lunati (2013) extend this methodology to estimate true
20 responses for all the models from multi-scale flow simulation on all the models in order to obtain more
21 robust uncertainty assessments and confidence intervals about tracer breakthrough curves in aquifers.
22 Their method is partly similar to ours, but only considers flow simulation proxies which account for
23 time and space explicitly, as it relies on the comparison of approximate and exact breakthrough curves
24 at each simulation time step. Our method, in contrast, can use the simpler and faster proxies discussed
25 above such as connectivity-based proxies and static proxies such as STOOIP.

26 Our main contribution is to propose a methodology applicable to all classes of reservoir proxies. For
27 this, we present a technique able to generate surrogate analytical production curves for all models of
28 the initial ensemble based on the full set of proxy distances and a few accurate flow simulations on a
29 few reference models. After defining the problem in more formal terms, this paper presents an
30 overview of the method. We then introduce the real reservoir data set used to illustrate the method
31 before explaining in more details the three different steps of the methodology: first, the reconstruction
32 of full physics distances between models according to proxy distances; then the determination of the
33 mathematical curve which best fits the oil production profiles; and last, the determination of the
34 missing dynamic profiles by non-linear optimization. Finally, we present the results and discuss the
35 ability to compute more accurate quantiles on production forecasting for uncertainty evaluation.

36 2 Problem definition

37 We consider a large set \mathcal{S} of N reservoir models \mathcal{M}_i covering the range of possible reservoir
38 geometries and heterogeneities

$$\mathcal{S} = \{\mathcal{M}_i\}_{i=1,2,\dots,N}. \quad (1)$$

39 Each of these models can be generated using alternative interpretations scenarios and Monte-Carlo
40 sampling. This ensemble represents the known structural and parameter uncertainty about the
41 reservoir. Reservoir production forecasts reflecting geological uncertainties should ideally consider all



1 these possible models. However, to keep the computational cost realistic, multiphase flow simulation
 2 can only be performed on a subset \mathcal{S}_1 of m models deemed representative of the ensemble:

$$\begin{aligned} \mathcal{S}_1 &= \{\mathcal{M}_i\}_{i=1,2,\dots,m} \\ \mathcal{S}_2 &= \{\mathcal{M}_i\}_{i=m+1,m+2,\dots,N} \\ \mathcal{S}_1 + \mathcal{S}_2 &= \mathcal{S} \\ m &\ll N \end{aligned} \quad (2)$$

3 Finding these m models is a problem *per se*, which calls for *a priori* ranking of all the models of \mathcal{S}
 4 based on their likely dynamic behavior. In a second step, the analysis of the production scenarios is
 5 usually based on production curve quantiles. The few selected models of \mathcal{S}_1 are used to simulate the
 6 dynamic response and some specific quantiles (Q10, Q50 and Q90) are computed from the production
 7 curves at each time step (e.g., Subbey et al. (2004)).

8 The first limitation with this approach is that only the information from \mathcal{S}_1 is taken into account and all
 9 the remaining models (from \mathcal{S}_2) are discarded. The second limitation is that the quantiles are computed
 10 from very few data and therefore they are very sensitive to the model selection.

11 The goal of our methodology is to address these two challenges by: (1) using the results of flow
 12 simulation proxies on the models of \mathcal{S}_2 ; (2) providing confidence intervals around the quantiles of the
 13 production curves, which reflect both the limited number of model realizations and the quality of the
 14 proxy.

15 3 Overview of the proposed methodology

16 **Fig. 1** illustrates the overall methodology proposed in this paper. The general aim is to minimize the
 17 loss of information and the sensitivity of the quantile estimation. This is attained by estimating
 18 surrogate oil recovery profiles for all the models belonging to \mathcal{S}_2 in order to account for the entire set
 19 of models. This allows to estimate the uncertainties more accurately while keeping the computational
 20 cost relatively low.

21 As illustrated on the top left of **Fig. 1**, we use a proxy for every model belonging to \mathcal{S} following the
 22 same philosophy as Josset and Lunati (2013). Each physically-based simulation proxy provides an
 23 approximate production curve $\varphi_i^P(t)$, where $i \in [1, N]$ refers to a given reservoir model and t is a
 24 pseudo-time step. We use these data to compute a distance between each pair of models. The notion of
 25 distance between geomodels is presented by Suzuki and Caers (2008) to quantify the dissimilarity
 26 between the models. Here, we compute the squared distance between proxy curves as

$$d_{i,j}^P = \sum_{t=0}^{t_f} (\varphi_i^P(t) - \varphi_j^P(t))^2, \quad (3)$$

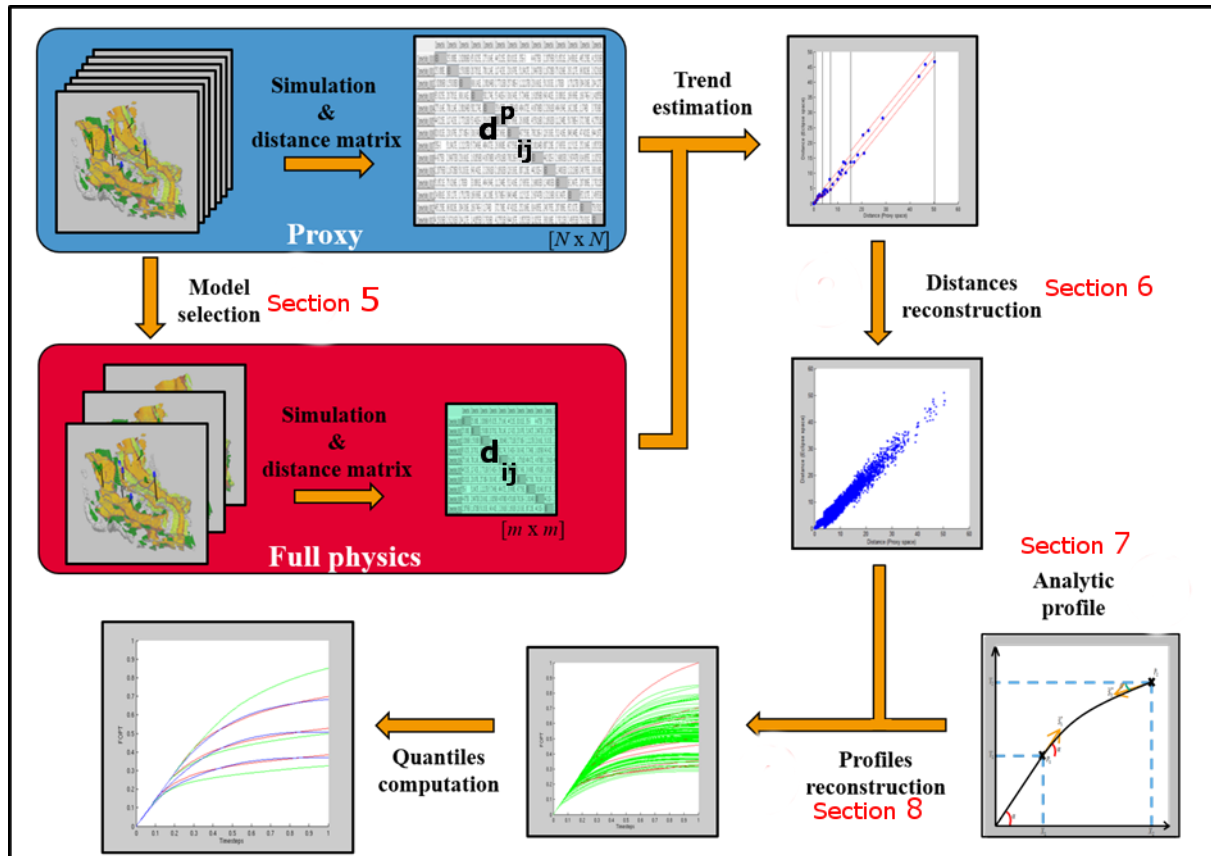
27 where $\varphi_i^P(t)$ represents the proxy curve of the model i at pseudo-time step t , $\varphi_j^P(t)$ the proxy curve of
 28 the model \mathcal{M}_j at pseudo-time step t and t_f denotes the last time step of the simulation. Alternatively,
 29 proxies may consist of simple scalar measures φ_i^P , such as oil in place or pore volumes connected to
 30 producing wells. In these cases, the distance becomes:

$$d_{i,j}^P = |\varphi_i^P - \varphi_j^P|, \quad (4)$$

31 The proxy distances are used to select m representative models (the \mathcal{S}_1 ensemble). For each of the
 32 selected models, the complete and accurate flow simulator is used to compute the dynamic flow
 33 response (red rectangle in **Fig. 1**), following the methodology of Scheidt and Caers (2009). The
 34 distances between the accurate flow responses are computed as well.

1 Then the relations between the two sets of results are analyzed statistically. Using a simple analytical
 2 expression for the production profile (surrogate profile), the statistical analysis between the proxy and
 3 full physics response distances, we estimate the surrogate profiles for all the models belonging to \mathcal{S}_2
 4 (distance reconstruction and profile reconstruction in **Fig. 1**). This allows to generalize the workflow
 5 of Josset and Lunati (2013) in the case of diverse types of proxies.

6 Finally, we estimate uncertainties on oil recovery by computing the production quantiles Q10, Q50
 7 and Q90 using both the accurate responses obtained for the models belonging to \mathcal{S}_1 and the surrogate
 8 profiles for the models belonging to \mathcal{S}_2 (lower left corner in **Fig. 1**). All the steps are described in
 9 detail in the following sections.



10

11 **Fig. 1 – Proposed workflow. A few representative models are selected from proxy simulation (Section 5) and accurate**
 12 **simulations are performed. Using the correlation between proxy distances and accurate distances, all missing dynamic**
 13 **distances are determined (Section 6), then an analytic profile (Section 7) is fitted to determine the missing dynamic**
 14 **profiles (Section 8). Finally, oil recovery quantiles are computed (Section 9).**

15 4 Test case

16 To illustrate the proposed methodology and test both the workflow and its individual steps a real
 17 reservoir case study is used. The reservoir consists of turbiditic channel deposits and is bounded
 18 laterally by pelagic shales of negligible permeability. The reservoir layers have been deformed and
 19 affected by several faults (**Fig. 2**). The field was developed with four producing wells. Three water
 20 injection wells and an aquifer maintain the pressure during reservoir production. For confidentiality
 21 reasons, we cannot disclose reservoir location, dimensions and formation names.

22 In this case, the ensemble \mathcal{S} contains 163 corner-point reservoir models of 793 000 cells ($130 \times$
 23 100×61) describing the static reservoir uncertainties. This number of models corresponds to the



1 subset of the initial 200 generated models for which the full physics flow simulations converged. We
 2 preferred to discard the models for which flow simulation did not converge in order to have a set of
 3 comparable models. Solving the convergence issues would require changing some numerical
 4 parameters and it could lead to differences between the model responses that would not be related to
 5 the geology. The grid geometry was generated from seismic interpretation and well tops. To reflect
 6 structural uncertainties, the reference structural interpretation was perturbed within bounds that reflect
 7 seismic time-to-depth conversion uncertainties and uncertainties in picking the reservoir layers on the
 8 migrated seismic image (Abrahamsen, 1993; Thore et al., 2002). As a result, all grids in the set of
 9 realizations have distinct geometries. For the petrophysical modeling, we first simulated lithologies
 10 using a multiple point simulation technique, then we used sequential Gaussian simulation (SGS)
 11 within each facies for net-to-gross, effective porosity and co-located simulation for the permeability
 12 field. In the SGS, we used a proportion cube from a seismic attribute. For the simulations, we honored
 13 the target statistics inferred from the wells and spatial variability models (variograms) obtained from
 14 available data (wells) and analogs. Finally, the oil-water contact was sampled within a range of 20
 15 meters and corresponding initial water saturations were simulated using again sequential Gaussian
 16 simulation (**Fig. 2**). The aim is to forecast field production for 20 years.

17 Proxy simulation was performed on all models. In another study (Bardy et al., 2014), we investigated
 18 the quality of several proxies on the same ensemble of models. In this paper, we do not use an actual
 19 flow proxy, but simply consider the Stock Tank Original Oil In Place (STOOIP) as an indicator of
 20 reservoir production performance. STOOIP characterizes every model by an integer value using the
 21 following equation:

$$STOOIP = \sum_{i=1}^n GRV(i) * NTG(i) * \Phi(i) * \frac{(1-S_{water})}{Bo}, \quad (5)$$

22 where: n is the number of cells in the model, GRV (Gross Rock Volume) is the effective part of rock
 23 in the cell, NTG (Net To Gross) corresponds to the proportion of reservoir rock in the cell, S_{water} is
 24 the water saturation in the cell and Bo the oil volume factor (volume ratio between reservoir and
 25 surface conditions). Although STOOIP is not a direct reservoir production proxy, it can be taken as
 26 such by our distance-based method. As expected theoretically and shown by Bardy et al. (2014), this
 27 proxy is less accurate than upscaling to provide a good approximation of model responses, but we
 28 chose it because it is very simple and quick to compute. As a result, it is often used in the oil industry
 29 to rank static reservoir models, even though the relationship between static and dynamic reservoir
 30 attributes can be complex. This choice was also made to evaluate to what extent our method can
 31 estimate production curves from a scalar static proxy. As this proxy only provides a single value per
 32 model, the distance was computed using Eq. (4).

33 The flow simulation forecast over 20 years were performed on all the stochastic models using the
 34 finite volume based software ECLIPSE® from Schlumberger®, configured with three fluid phases.

35 Note that another test case for the methodology is presented in Bardy (2015).

36

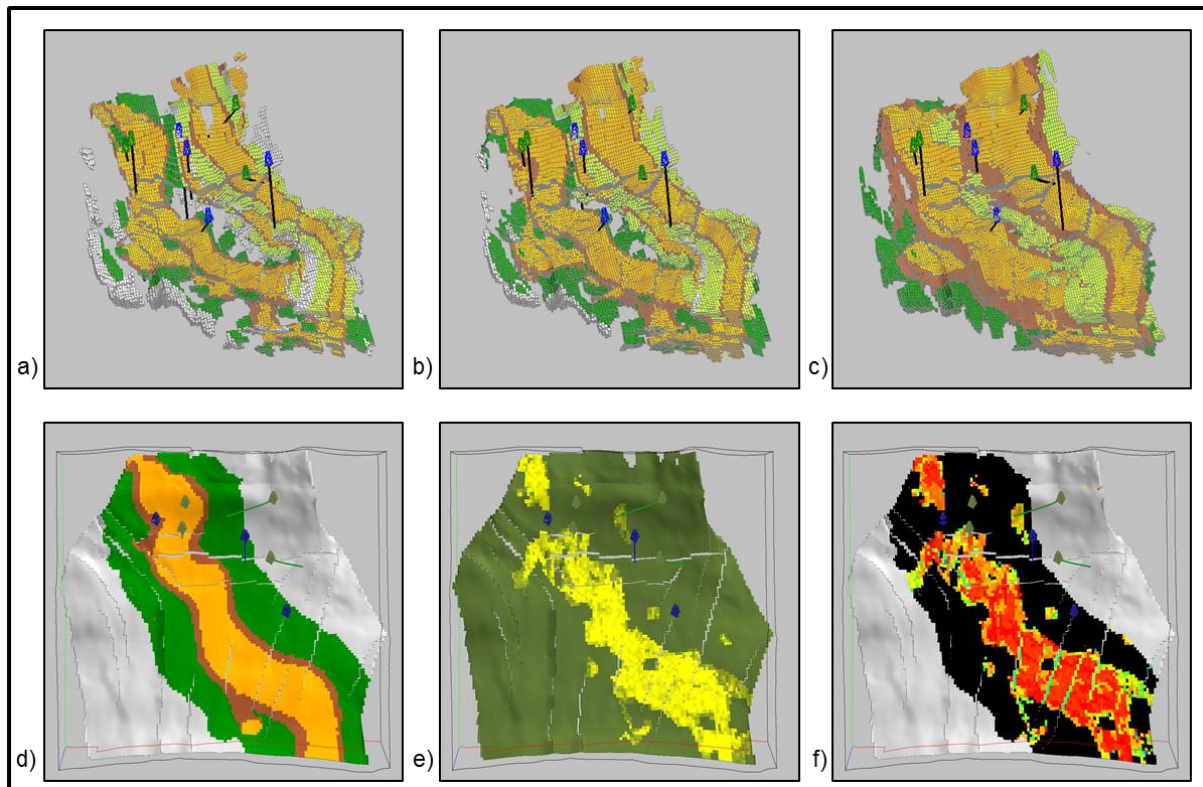


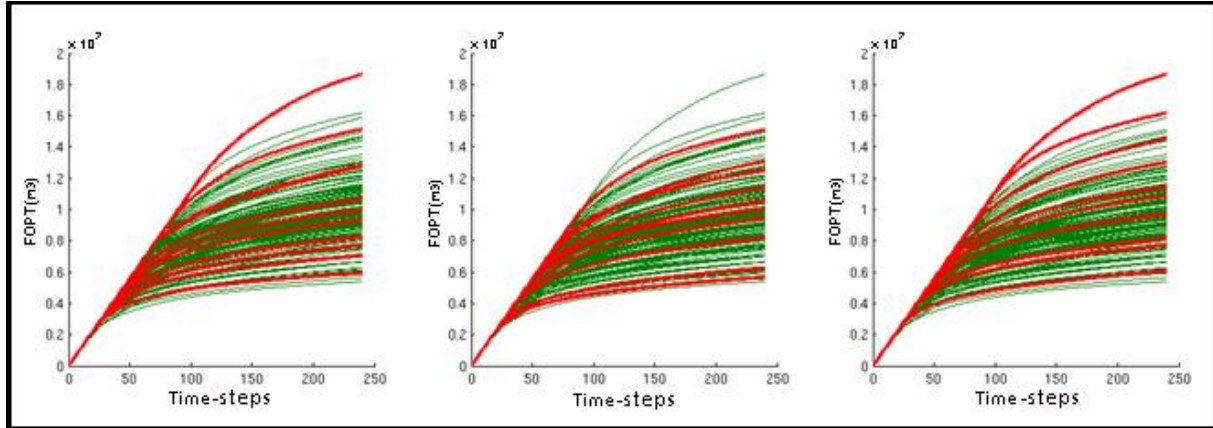
Fig. 2 – Our case study is made of several realizations (a-c: different geometry and facies distribution) in 3D view. Each realization is made of several petrophysical properties simulated successively (d: facies, e: porosity, f: permeability). Injector (blue) and producer (green) wells trajectory are also visible.

5 Model selection

The selection of the subset \mathcal{S}_1 must be as representative of \mathcal{S} as possible to keep the diversity into the responses and to allow an accurate uncertainty assessment on production. To achieve that, we use the Distance Kernel Method (DKM) proposed and described in detail by Scheidt and Caers (2007, 2009) as well as Josset and Lunati (2013). Here we provide only a brief summary of this technique. The interested reader is referred to the original papers for more details. The distance matrix between each pair of model can be computed by Eq. (3) or Eq. (4) depending on the retained proxy. Then the Multi-Dimensional Scaling algorithm (MDS) takes the proxy distance matrix as input and allows to plot the models in an abstract space of low dimensionality relatively to the number N of models. In this space, a set of representative models can finally be selected with a K-medoid algorithm (which is a clustering algorithm applied after a kernel transformation, see Kaufman and Rousseeuw, 1987). This clustering algorithm finds m models (medoids) which partition the model space so that the sum of distances between each model and the closest medoid is minimal. The main advantage of this method as compared to a classical K-means is that it is returning the medoids (models closest to the center of each cluster) which can be used directly, instead of the centers of each cluster which do not correspond to an existing model. The choice of m is important because it should neither be too high (which will be too time consuming to simulate full physics) nor too small (which will not be sufficiently representative). The selected models belonging to \mathcal{S}_1 on which flow simulations will be performed correspond to these medoids.

In our case study, we chose to use ten models ($m = 10$) as it is sufficient to correctly calibrate the method while minimizing the computational burden. As k-means, the k-medoid method solves a global optimization problem iteratively; as a result, different sets of models may be obtained

1 depending on the starting points. Although this can be a problem for some unsupervised clustering
 2 applications, this feature is actually interesting in our case, as it allows to introduce some randomness
 3 in the model selection process. Some examples shown in **Fig 3** illustrate, in the data space, the
 4 distribution of the selected models and the stochasticity of the technique. We will further exploit this
 5 stochasticity in Sect. 9 to analyze the stability of the confidence intervals around production curves.



6
 7 **Fig. 3 – Examples of models selection. The red curves represent the 10 selected models while the green curves**
 8 **represent the other models. The vertical axis represents the cumulative oil production or Field Oil Production Total**
 9 **(FOPT)**

10 6 Missing Distance reconstruction

11 In the previous step, the distances between the proxy responses of the models was used to select the
 12 models belonging to the subset \mathcal{S}_1 . The detailed and accurate flow simulation is performed only for the
 13 models belonging to \mathcal{S}_1 , and results in an oil recovery profile $\varphi_j(t)$ with $t \in [0; t_f]$ and $j \in [0, m]$. In
 14 most approaches using model selection, one would estimate the overall uncertainty on oil recovery by
 15 computing directly the quantiles from this limited set of profiles.

16 Here, we propose to go a step further and use all the information that was obtained by computing the
 17 proxies for all the models to improve the overall uncertainty quantification. The idea is that the
 18 distances between the proxy responses can be correlated with the distances between the real flow
 19 responses. Using that correlation and using in addition a simple expression for the shape of the
 20 profiles, it may be possible to reconstruct the variability of the unknown profiles reasonably well
 21 without solving the accurate flow problem and then use them to improve the uncertainty estimate. To
 22 do so, the first step consists in estimating the unknown distances between the true profiles from the
 23 distances between the proxy responses.

24 More precisely, from the accurate data, it is possible to compute the “true” squared distance ($d_{i,j}$)
 25 between model responses:

$$d_{i,j} = \sum_{t=0}^{t_f} (\varphi_i(t) - \varphi_j(t))^2. \quad (6)$$

26 For each pair (i, j) of models belonging to \mathcal{S}_1 , the distances between the proxy curves $d_{i,j}^p$ and the
 27 distances between the full physics curves $d_{i,j}$ are available.

28 **Fig. 4a** presents the cross-plot between full flow simulation distances and proxy distances obtained in
 29 this case; note that m different selected models provide $\frac{m*(m-1)}{2}$ distinct distances.

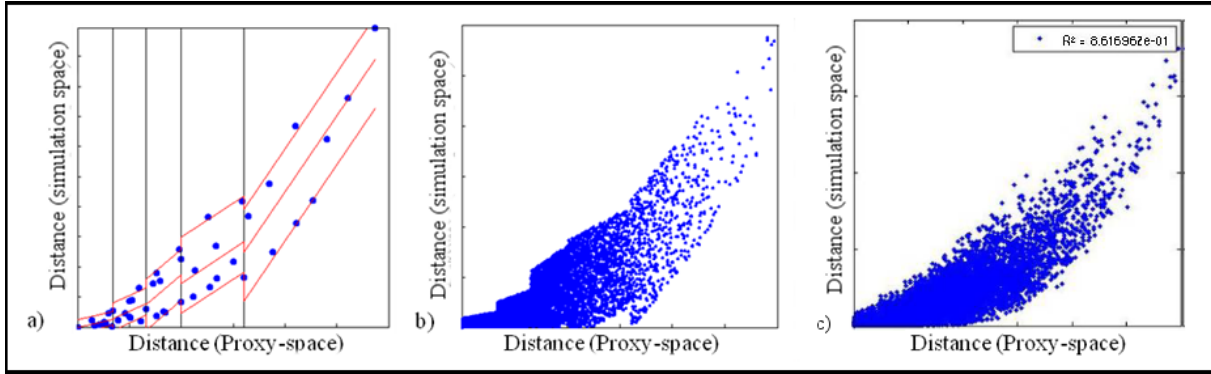


Fig. 4– (a) Cross plot of full flow simulation distances vs. proxy distances for \mathcal{S}_1 . Vertical lines represent domain borders. Red lines represent domain regression surrounded by the quantiles Q10 and Q90 of the residuals. (b) Cross plot of the reconstructed full flow simulation distances for \mathcal{S}_2 (obtained by random draw as explained in the text) vs. proxy distances. (c) True cross plot of full flow simulation distance vs. proxy distance, used as benchmark

A linear regression between the two types of distances could be estimated but there is no theoretical reason to assume a linear relationship. A non-linear correlation is confirmed by our experiments (**Fig. 4c**), and therefore a piecewise linear correlation model on K intervals is used instead. We choose these K domains in order to contain approximately the same number of points (N/K or $1 + N/K$, **Fig. 4a**). We also take care of having enough points in each domain to be able to compute reasonable statistics. As we have 10 simulated models, this means $\frac{10 \cdot (10-1)}{2} = 45$ distinct differences which give 45 points, we choose 5 domains of 9 points each. For each domain, we use a simple linear model simulating the full flow simulation distance using a linear function of the proxy distance and a random error term $\varepsilon_{k,i,j}^*$:

$$d_{i,j}^{*|P} = a_k d_{i,j}^P + b_k + \varepsilon_{k,i,j}^* \text{ with } k \in [1; K]. \quad (7)$$

The coefficients a_k and b_k of this regression are given by ordinary least squares.

Using equation (7), we simulate all the missing values of $d_{i,j}^{*|P}$ for all the pair of models belonging to \mathcal{S} . The random noise $\varepsilon_{k,i,j}^*$ is randomly sampled from a normal distribution whose parameters are given by the linear regression model (centered Gaussian distribution) and corrected to ensure that only positive distances are obtained. Additionally, we force the coefficient b_1 to be 0 so that the regression passes through the origin in the first domain (a null distance corresponds to identical model response according to both proxy and full physics simulations). This piecewise regression allows for increasing variance with proxy distance, but may have discontinuities between domains and is not strictly consistent with regression theory.

In this work, we did not investigate more advanced regression models. Indeed, when applied to our test case (**Fig. 4b**), this method yields a set of acceptable possible full flow simulation distances which are consistent with the reference plot generated from the reference flow solution (**Fig. 4c**, Bardy et al. 2014) and which is sufficient for reasonable uncertainty estimates as shown below.

Note also that a perfect proxy, showing no error at all, would reproduce perfectly the true distances in the simulation space and would lead to a perfect straight diagonal line in **Fig. 4a**. However, a perfect proxy is very likely to be almost as computationally expensive as the accurate forward flow simulation and will have therefore no computational interest. Instead, a fast proxy will for sure display some discrepancies between the two distances as illustrated in **Fig. 4a**. The aim of the proposed method is therefore to model properly those errors for any kind of proxy and benefit from their numerical



1 efficiency. In the next sections, we explain how to produce confidence intervals around the forecasted
 2 production curves; in principle, the width of these confidence intervals should directly reflect the
 3 quality of the proxy.

4 7 Parametric model of oil production

5 Quantifying uncertainties about reservoir production calls for estimating oil production profiles for all
 6 the models belonging to \mathcal{S}_2 (for which full physics simulations are unavailable). The profiles of total
 7 oil produced at field scale (FOPT) represent the cumulative oil production of the field at each reported
 8 time-step. In a standard workflow, these profiles are obtained from flow simulations, but proxies do
 9 not necessarily generate this type of curves. For example, when using STOOIP as a proxy, we can
 10 compute the distances between the models but not the FOPT curves.

11 This is why, in this work, we suggest estimating the FOPT profiles from the set of distances (Eqs. 6
 12 and 7), the proxy simulations (Eq. 3) and the regression model (Eq. 7) using a parametric model (Eq.
 13 8). The aim of the current section is to present the parametric model that we propose to use. In section
 14 8, we will explain how the parameters of these curves are determined from the value of the distances.

15 A parametric model of the FOPT curve should respect three criteria imposed by the physics of oil
 16 recovery:

- 17 • It is a cumulative profile. The mathematical function must be monotonically increasing –
 18 with a positive or null derivative.
- 19 • The first part of the profile corresponds to the maximum production rate in which liquid rates
 20 are constrained by the production installation (e.g., pipe diameter) so the function must be
 21 linear at the beginning and cross the origin.
- 22 • The reservoir is considered at the beginning of its production, with no historical data and a
 23 constant number of production wells; after a while (time t_d), production declines so the
 24 function must have a strictly negative second derivative.

25 Rather than choosing between an exponential or hyperbolic analytical production decline model
 26 (Fetkovitch et al., 1996), we use a piecewise parametric function (**Fig. 5a**). It consists first in a linear
 27 part from the origin to a first control point $\mathbf{P}_d = (t_d, \varphi_d)$ corresponding to the beginning of the
 28 production decline, followed by a Cubic Hermite Spline from \mathbf{P}_d to a second final control point
 29 $\mathbf{P}_f = (t_f, \varphi_f)$.

$$\varphi^A(t) = \begin{cases} \alpha t & \text{if } t < t_d \text{ (before production decline)} \\ \tilde{\Gamma}(t) & \text{if } t > t_d \text{ (during production decline)} \end{cases}'$$

$$\text{where: } \tilde{\Gamma}(t) = \Gamma(\tau) = h_{00}(\tau)\mathbf{P}_d + h_{10}(\tau)\mathbf{S}_d + h_{01}(\tau)\mathbf{P}_f + h_{11}(\tau)\mathbf{S}_f \text{ for } \tau = \frac{t-t_d}{t_f-t_d},$$

$$h_{00}(\tau) = 2\tau^3 - 3\tau^2 + 1, \tag{8}$$

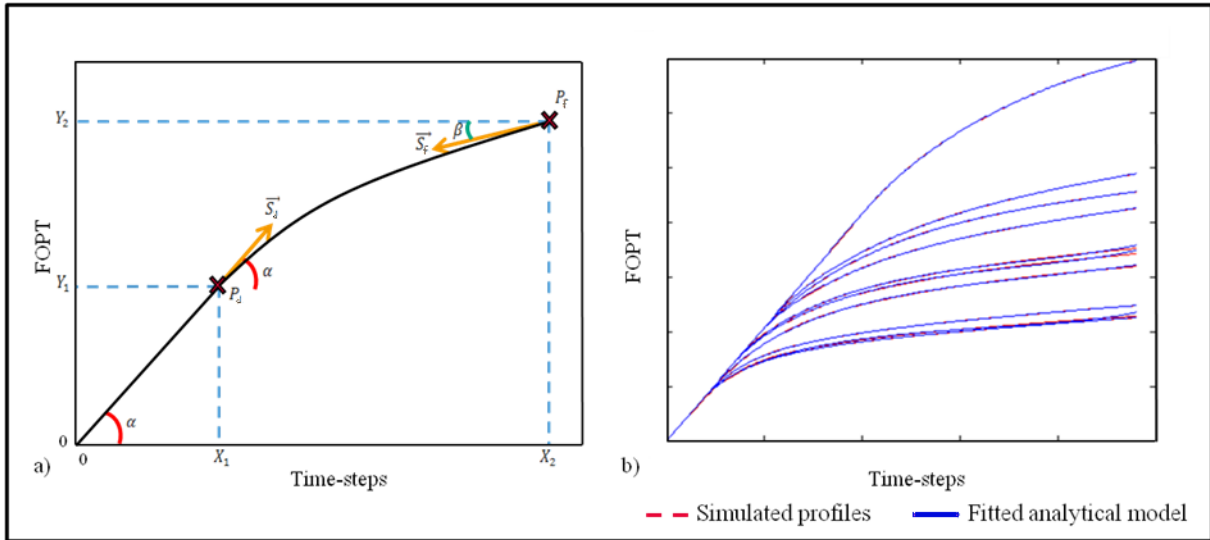
$$h_{10}(\tau) = \tau^3 - 2\tau^2\tau,$$

$$h_{01}(\tau) = 2\tau^3 + 3\tau^2,$$

$$h_{11}(\tau) = \tau^3 - \tau^2,$$



1 where h_{00}, h_{10}, h_{01} and h_{11} are the Hermite functions. Further symbols are defined below.



2
3 **Fig. 5 – a) Graphical view of the analytical model and parameters of analytical oil recovery profiles. b) Result of least-**
4 **squares fit of the analytical model (red curves) to production profiles obtained by flow simulation (blue curves).**

5 As shown in **Fig. 5a**, the slope α in Eq. (8) corresponds to the peak production rate and can be
6 identified using the linear part on the m simulated models during the early time steps. The final time
7 t_d corresponds to the duration of production so it is the same for all models. Therefore, five unknown
8 parameters remain in Eq. (8) to fully define the analytical production profile $\varphi^A(t)$:

- 9 - t_d - the time after which production starts declining. The cumulated production at this time is
10 given by $\varphi_d = \alpha t_d$.
11 - φ_f - the total production at the final simulation time ($t = t_f$).
12 - $\|\mathbf{S}_d\|$ - the norm of the tangent vector \mathbf{S}_d at the start of production decline ($t = t_d$). We
13 assume a gradual production decline so the direction of the tangent vector \mathbf{S}_d is the same as in
14 the linear part, i.e., $\mathbf{S}_d = \begin{pmatrix} \|\mathbf{S}_d\| * \cos(\alpha) \\ \|\mathbf{S}_d\| * \sin(\alpha) \end{pmatrix}$. In the Hermite polynomial, this norm controls how
15 fast the production rate decreases after t_d .
16 - β and $\|\mathbf{S}_f\|$ - respectively the slope and the norm of the tangent vector \mathbf{S}_f at the end of
17 simulation ($t = t_f$), i.e.: $\mathbf{S}_f = \begin{pmatrix} \|\mathbf{S}_f\| * \cos(\beta) \\ \|\mathbf{S}_f\| * \sin(\beta) \end{pmatrix}$. The slope β corresponds to the production
18 rate at the final simulation time; and the norm controls how quickly this rate decreased in the
19 late phases of production.

20 To test the relevance of this analytical production model, we considered the set of FOPT curves
21 obtained by flow simulation on the $m = 10$ reservoir models of \mathcal{S}_1 , then we used least-squares
22 minimization to fit the parameters of the analytical model (Eq. (8)). Results (**Fig. 5b**) suggest that the
23 proposed function is able to adequately represent FOPT under our simulation settings (continuous
24 field production without interruption). This step will also provide an ensemble of valid sets of
25 parameters that we will use in the next part of the workflow.



8 Reconstruction of profiles using the proxy distance

Let us consider a reservoir model \mathcal{M}_i belonging to \mathcal{S}_2 . The aim of this section is to reconstruct a possible FOPT curve for this reservoir model without running the flow simulation. We propose to do this by identifying the five parameters of the analytical function given in Eq (8) and described in the previous section. For this, we introduce a new squared distance $d_{i,j}^A$ between the analytical production profile $\varphi_i^A(t)$ for model \mathcal{M}_i and the known profile $\varphi_j(t)$ of each of the m models \mathcal{M}_j of \mathcal{S}_1 :

$$d_{i,j}^A = \sum_{t=0}^{t_f} (\varphi_i^A(t) - \varphi_j(t))^2 \quad (9)$$

Ideally, the analytical profiles $\varphi_i^A(t)$ should be such that their distance $d_{i,j}^A$ to the known models \mathcal{M}_j of \mathcal{S}_1 be equal to the distance $d_{i,j}^{*|P}$ simulated from the piecewise linear regression (Eq. 7). Note that we assume that $d_{i,j}^{*|P}$ is a reasonable approximation of the unknown distance between the true profiles. Therefore, we propose to find the five unknown parameters of each profile $\varphi_i^A(t)$ (Eqs. 3-4) by minimizing the following objective function:

$$F_i = \sum_{j=1}^m (d_{i,j}^A - d_{i,j}^{*|P})^2, \quad (10)$$

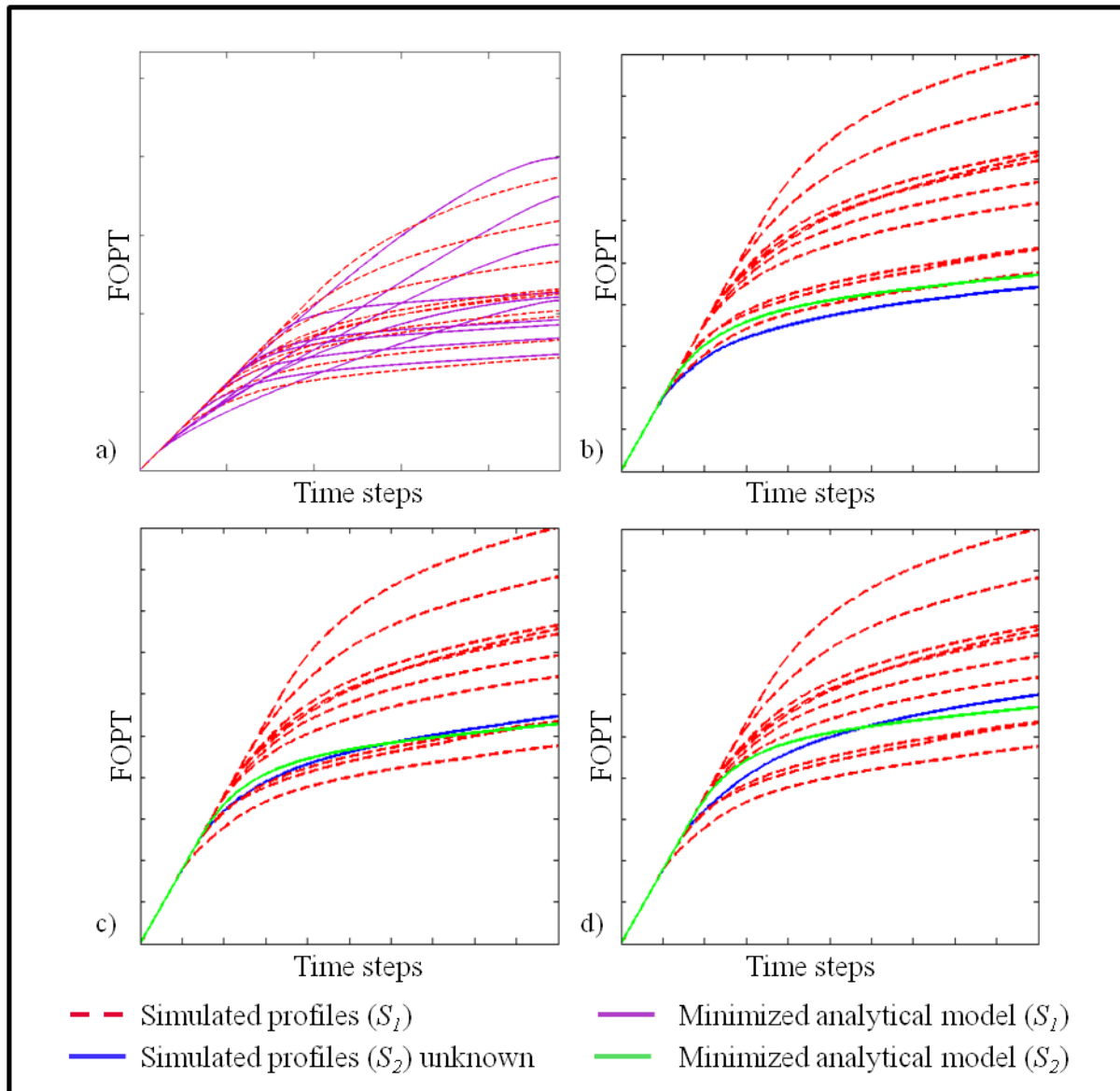
This minimization is a non-linear problem; it can be solved using the interior point method (Nocedal and Wright 2006). In practice, to help convergence, the curves are normalized so that the highest known profile $\varphi_i(t_f)$ is equal to 1.

Furthermore, to ensure an acceptable result, we prescribe bounds on some parameters and add some non-linear constraints. To force the production decline to start before the end of the simulation period, we constrain t_d to be in the interval $[0; \frac{1}{\alpha}]$. To reflect that production rate is lower at the end of the simulation than during peak reservoir performance, we constrain β to be within $[0, \alpha]$. We also assume that the best model in terms of final production does not produce more than two times the production of the best simulated models: $\varphi_f \in [0; 2]$. Last, $\|\mathcal{S}_d\|$ and $\|\mathcal{S}_f\|$ are bound within the interval $[0; 5 * t_f]$ to avoid sharp variations of the production curve between times t_d and t_f . In addition, we set a constraint to respect the physics of the simulation by forcing the second derivative of the mathematical model (Eq. (8)) to be negative ($\partial^2 \varphi_i^A(t) / \partial t^2 \leq 0$).

In our study, we first applied this method to the set of ten reservoir models of \mathcal{S}_1 , for which we know both the FOPT curve obtained by simulation and proxy results. For each model, we minimized Eq. (10) to estimate the analytical FOPT curve from the distances to the nine other models ($m = 9$). **Fig. 6a** shows that the ranking of the estimated analytical curves is the same as the reference ranking for most time steps (especially for long simulation times). However, some rank inversions are visible for intermediate time steps. Also, the linear part of the obtained production curves is often too short as compared to the true profiles. This could be due to the relatively small number of samples, or to the existence of local minima in the non-linear optimization. To reduce these errors we propose an optional kriging-based correction method described in Appendix.

We then ran the non-linear minimization of Eq. (10) for all the models of \mathcal{S}_2 based on the distance to all ten models of set \mathcal{S}_1 ($m = 10$). To speed up this optimization, we selected as starting point of the minimization the parameters of the known curve for which the objective function value was lowest.

1 Three individual comparisons between the estimated analytical profiles and the true profiles from the
 2 reference simulations are presented in **Fig. 6b-d**. In these cases, also, the analytical profiles are
 3 correlated to the reference ones, but display some variations. We think these deviations occur partly
 4 because of errors in the proxy-based distances and partly because distances are unable to uniquely
 5 characterize the shape of the production curves. Further research is needed to characterize these
 6 sources of error and to characterize the numerical properties of the minimization of Eq (10).



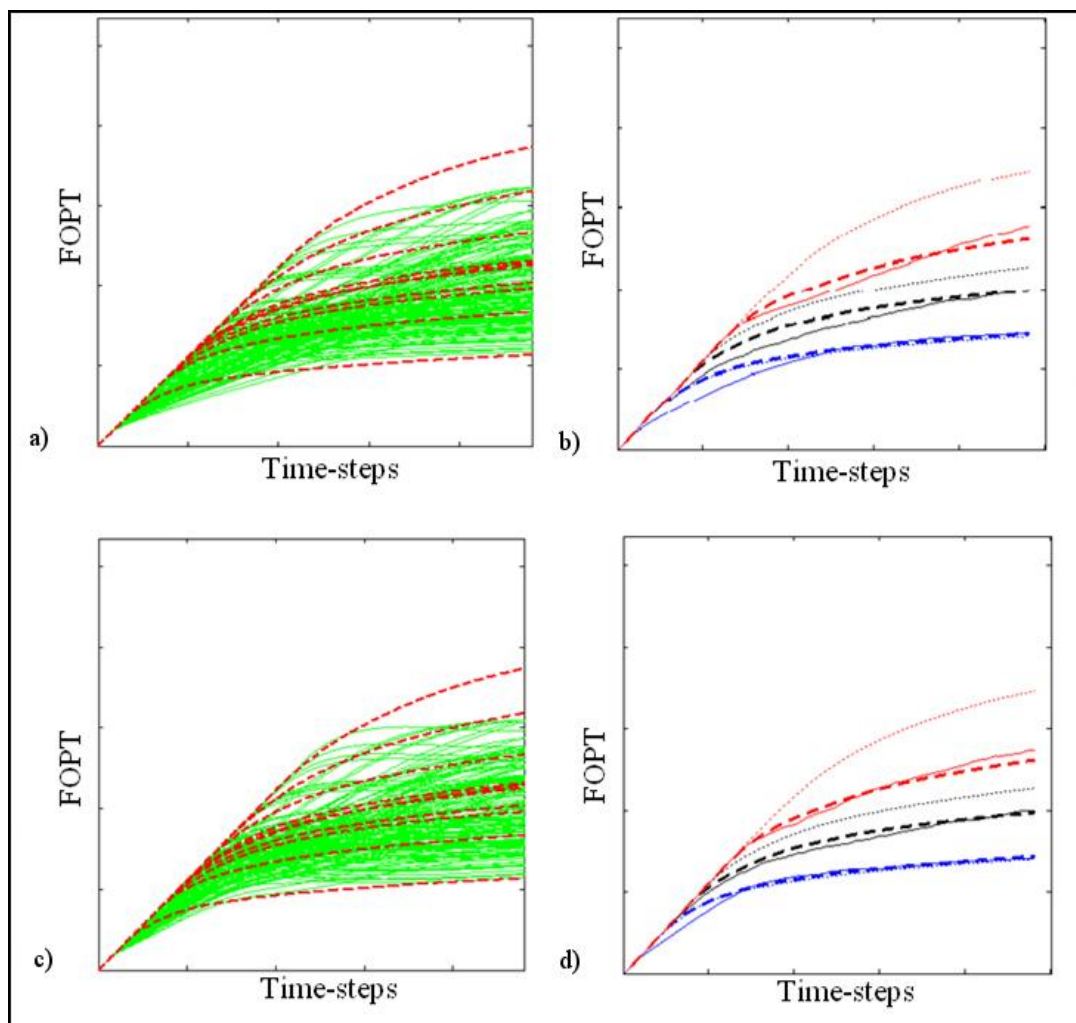
8 **Fig. 6 – (a)** Minimization of the analytical model parameters for the models belonging to \mathcal{S}_1 , dashed curves correspond
 9 to the reference profiles obtained with a full physics flow simulator and the purple ones are obtained after
 10 minimization. **(b-d)** Examples of minimization procedure results. Red dashed curves correspond to the real response
 11 profile of models belonging to \mathcal{S}_1 , green curves to the analytical curve of a model belonging to \mathcal{S}_2 and blue curves to
 12 corresponding real profile (computed with full physics flow simulator and not known in operational context).

13 9 Production curve quantiles and confidence intervals

14 The above procedures produce a set of m true production profiles (from flow simulation on models
 15 belonging to \mathcal{S}_1) and $n - m$ estimated analytical profiles $\varphi_i^A(t)$ (from the minimization of Eq. (10) on
 16 all models of \mathcal{S}_2). The corresponding curves are shown on **Fig. 7a**. For each time step, we create a

1 discrete probability distribution function from these N profiles. Final surrogate quantile curves are
 2 obtained by linking the quantiles of these distributions across all the time steps.

3 **Fig. 7b** shows the results of quantile curves estimation for our case study. The curves estimated with
 4 our methodology are closer to the reference than the production profiles of the models in \mathcal{S}_1 . This
 5 confirms experimentally that proxies integrated with the proposed method reduce uncertainty
 6 quantification errors as compared to uncertainties estimated only from a few representative models.
 7 However, the errors discussed in Section 8 are still visible when comparing the computed quantiles
 8 and the reference quantiles obtained from flow simulation on all models. In particular, the Q10 and
 9 Q50 curves underestimate the time when production starts declining, yielding more conservative
 10 production scenarios.



11

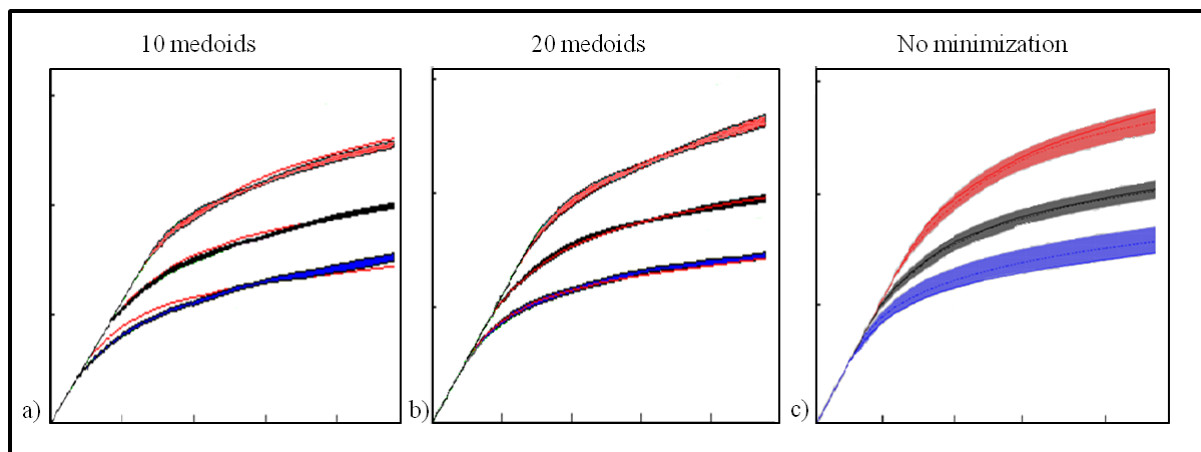
12 **Fig. 7** – Final analytical production curves and corresponding quantiles (a) Analytical curves obtained after
 13 minimization applied to each model of \mathcal{S}_2 (full green lines) and accurate curves obtained by flow simulation on \mathcal{S}_1
 14 models (dashed red lines). (b) Quantiles (Q10 in blue, Q50 in black and Q90 in red) computed with all models (\mathcal{S})
 15 (reference, dashed lines); quantiles computed only with models belonging to \mathcal{S}_1 (dotted lines); quantiles computed with
 16 our methodology ($\mathcal{S}_1 + \mathcal{S}_2$) (full lines). (c) Minimized profiles after correction (see Appendix). (d) Quantiles updated with
 17 corrected profiles for our methodology (full lines).

18 To remove these systematic errors, we applied the correction described in Appendix. As shown on
 19 **Fig. 7c**, this correction introduces some non-admissible behaviors in individual curves. For instance,
 20 some cumulative production curves locally have a negative slope. However, these local
 21 inconsistencies are filtered by the quantile estimation process and the final quantile curve estimates
 22 (**Fig. 7d**) are closer to the reference curves.

1 In all cases, quantile curve estimations are prone to several sources of uncertainty. First, there is no
 2 guarantee about the representativeness of the subset of models \mathcal{S}_1 selected for full flow simulation.
 3 This could ideally be addressed by increasing the number m of models in \mathcal{S}_1 until the output statistics
 4 converge. Because this can be computationally infeasible, another option is to repeat the k-medoid
 5 clustering process (Section 5) starting from different initial models. Indeed, the k-medoid algorithm is
 6 sensible to the choice initial models, as mentioned by Josset and Lunati (2013). This option is
 7 computationally expensive but can provide a practical way to assess confidence intervals around the
 8 quantile curves for a fixed number of models in \mathcal{S}_1 .

9 A second source of uncertainty relates to the nature of the chosen proxy which, as the name suggests,
 10 only approximates the actual flow response. Indeed, more accurate proxies should in principle lead to
 11 better confidence in the quantile curve estimates. In our method, this proxy error is sampled during the
 12 missing distance reconstruction (Section 6). Resampling alternative distances from the piecewise
 13 regression model provides an efficient way to address this source of uncertainty without need for
 14 additional flow simulations.

15 In this study, we first applied our methodology 200 times. Due to the stochasticity of the k-medoid
 16 algorithm used in the DKM (Section 5), we were able to generate 200 different sets of quantile profiles
 17 (Q10, Q50 and Q90) and then computed the confidence interval around them with ten representative
 18 models (**Fig. 8a**). The confidence intervals obtained with our method do not always include the actual
 19 quantile curves generated from the reference simulations due to proxy error or workflow limitations.
 20 We suspect that the correction step described in the appendix could explain why these confidence
 21 intervals are too low. Indeed, this correction may artificially lower the profiles. Therefore, we chose to
 22 increase the number of representative models to twenty and ran the randomized workflow 200 times to
 23 obtain new confidence intervals (**Fig. 8b**). The results in this case are better in the sense that
 24 confidence intervals are slightly narrower and include the reference quantile curves. In contrast, we
 25 applied 200 random selections of representative models, and aggregated the true production curves to
 26 compute the associated confidence intervals (**Fig. 8c**). When our methodology is used, the confidence
 27 intervals are narrower due to the integration of proxy knowledge.



28
 29 **Fig. 8 – Quantiles dispersion over 200 iterations of the methodology. (a) With a \mathcal{S}_1 subset of 10 models. (b) With a \mathcal{S}_1**
 30 **subset of 20 models. (c) With 10 models in \mathcal{S}_1 but without using proxies from \mathcal{S}_2 in the quantile computation.**

31 10 Conclusions

32 Estimating the uncertainties on the flow response of a reservoir by performing full physics flow
 33 simulation on a large ensemble of models implies often prohibitive simulation time. To overcome this



1 problem, some approaches consider an approximation of the dynamic simulation response which is
2 obtained faster by simplification of the physics of the simulator. The use of these proxies leads to a
3 delicate issue of balancing acceleration with accuracy of the response. Another problem is that proxies
4 are very often case-dependent; choosing the proxy which is best suited to a given case study can be a
5 long procedure (Bardy et al. 2014). Other approaches consider selecting only a few models from the
6 initial set and run the accurate flow simulation only on this small subset. Our approach aims at
7 integrating the dynamic behavior of the whole set of models to capture the uncertainties, as also
8 proposed by Josset and Lunati (2013). The method proposed in this paper allows for a larger class of
9 proxies to be considered, but it calls for the explicit definition of analytical production profiles, which
10 depend on the type of reservoir simulation output. From a reservoir management standpoint, it could
11 be interesting to use other types of production curves such as water production or pressure, because it
12 can also be important to adjust surface facilities. Looking at the well by well response would also be
13 important as this is where proxies lack accuracy. Applying this methodology to any other kind of
14 curve like water profile is possible but relies on the availability of compact analytical models for the
15 considered reservoir response curve.

16 In this paper, the model selection was performed using the Distance Kernel Method (Scheidt and
17 Caers, 2009) but other selection techniques could be considered. For field development studies, further
18 research is also needed to assess how to address more complex development scenarios where multiple
19 well locations are considered and well schedule changes though time.

20 Nonetheless, the results obtained show that the methodology can assess reservoir oil production
21 uncertainty even with a very simple proxy. It would be interesting to assess the value of this method to
22 accelerate history matching tasks in the same spirit as Scheidt et al (2011) and Josset et al. (2015). It
23 would also be interesting to test this methodology on a case study where the full flow simulation
24 responses are more chaotic and with a less regular dispersion.

25 **Acknowledgments**

26 The authors of this paper, would like to thank Total S.A. for funding this research and providing the
27 case study, Celine Scheidt for providing the initial distance kernel method code, as well as Peter King
28 (Imperial College) and Vicent Corpel (Total S.A) for their unconditioned help. We also thank Pedram
29 Masoudi, and the other reviewers whose comments helped us to improve this paper.

30 **References**

- 31 Abrahamsen, P., 1993, Bayesian kriging for seismic depth conversion of a multi-layer
32 reservoir, in Geostatistics Troia'92: Springer, p. 385–398.
- 33 Alabert, F.G., Modot, V., 1992. Stochastic Models of Reservoir Heterogeneity: Impact on
34 Connectivity and Average Permeabilities, in: SPE Annual Technical Conference and
35 Exhibition. Society of Petroleum Engineers, Washington, D.C. doi:10.2118/24893-MS
- 36 Ballin, P.R., Journel, A.G., Aziz, K., 1992. Prediction Of Uncertainty In Reservoir
37 Performance Forecast. J Can Pet Tech 31. doi:10.2118/92-04-05
- 38 Bardy, G., 2015. Intégration de modèles approchés pour mieux transmettre l'impact des
39 incertitudes statiques sur les courbes de réponse des simulateurs d'écoulements. PhD Thesis,
40 Université de Lorraine.
- 41 Bardy, G., Biver, P., Caumon, G., Renard, P., Corpel, V., King, P.R., 2014. Proxy
42 Comparison for Sorting Models and Assessing Uncertainty on Oil Recovery Profiles, in:



- 1 ECMOR XIV - 14th European Conference on the Mathematics of Oil Recovery. EAGE.
- 2 doi:10.3997/2214-4609.20141901
- 3 Caers J (2011) Modeling Uncertainty in the Earth Sciences. Wiley-Blackwell,
- 4 Corre, B., Thore, P., de Feraudy, V., Vincent, G., 2000. Integrated Uncertainty Assessment
- 5 For Project Evaluation and Risk Analysis, in: SPE European Petroleum Conference. Society
- 6 of Petroleum Engineers, Paris, France. doi:10.2118/65205-MS
- 7 Cox, T.F., Cox, M.A., 2001. Multidimensional scaling, 2nd ed, Monographs on statistics and
- 8 applied probability. Chapman & Hall/CRC press.
- 9 Deutsch, C.V., 1998. Fortran programs for calculating connectivity of three-dimensional
- 10 numerical models and for ranking multiple realizations. Computers & Geosciences 24, 69–76.
- 11 doi:10.1016/S0098-3004(97)00085-X
- 12 Doherty, J., Christensen, S., 2011. Use of paired simple and complex models to reduce
- 13 predictive bias and quantify uncertainty. Water Resources Research 47, n/a-n/a.
- 14 doi:10.1029/2011WR010763
- 15 Durlofsky LJ, 2005. Upscaling and gridding of fine scale geological models for flow
- 16 simulation. In: 8th International Forum on Reservoir Simulation, Stresa, Italy, 2005.
- 17 Fenwick, D., Scheidt, C., Caers, J., 2014. Quantifying Asymmetric Parameter Interactions in
- 18 Sensitivity Analysis: Application to Reservoir Modeling. Mathematical Geosciences 46, 493–
- 19 511. doi:10.1007/s11004-014-9530-5
- 20 Fetel, E., Caumon, G., 2008. Reservoir flow uncertainty assessment using response surface
- 21 constrained by secondary information. Journal of Petroleum Science and Engineering 60,
- 22 170–182. doi:10.1016/j.petrol.2007.06.003
- 23 Fetkovich, M.J., Fetkovich, E.J., Fetkovich, M.D., 1996. Useful Concepts for Decline Curve
- 24 Forecasting, Reserve Estimation, and Analysis. SPE Reservoir Engineering 11, 13–22.
- 25 doi:10.2118/28628-PA
- 26 Ginsbourger, D., Rossopoff, B., Piro, G., Durrande, N., Renard, P., 2013. Distance-based
- 27 kriging relying on proxy simulations for inverse conditioning. Advances in Water Resources
- 28 52, 275–291. doi:10.1016/j.advwatres.2012.11.019
- 29 Gross, H., 2012. Response surface approaches for large decision trees: decision making under
- 30 uncertainty, in: ECMOR XIII-13th European Conference on the Mathematics of Oil
- 31 Recovery. doi:10.3997/2214-4609.20143195
- 32 Hovadik, J., Larue, D., 2011. Predicting waterflood behavior by simulating earth models with
- 33 no or limited dynamic data: From model ranking to simulating a billion-cell model, in:
- 34 Uncertainty Analysis and Reservoir Modeling, AAPG Memoir. pp. 29–55.
- 35 doi:10.1306/13301406M961028
- 36 Josset, L., Lunati, I., 2013. Local and Global Error Models to Improve Uncertainty
- 37 Quantification. Mathematical Geosciences 45, 601–620. doi:10.1007/s11004-013-9471-4
- 38 Josset, L., Ginsbourger, D., Lunati, I., 2015a. Functional error modeling for uncertainty
- 39 quantification in hydrogeology. Water Resources Research 51, 1050–1068.
- 40 doi:10.1002/2014WR016028
- 41 Josset, L., Demyanov, V., Elsheikh, A.H., Lunati, I., 2015b. Accelerating Monte Carlo
- 42 Markov chains with proxy and error models. Computers & Geosciences 85, 38–48.
- 43 doi:10.1016/j.cageo.2015.07.003



- 1 Kaufman, L., Rousseeuw, P., 1987. Clustering by means of medoids. North-Holland.
- 2 Majdi Yazdi, M., Jensen, J.L., 2014. Fast screening of geostatistical realizations for SAGD
3 reservoir simulation. Journal of Petroleum Science and Engineering 124, 264–274.
4 doi:10.1016/j.petrol.2014.09.030
- 5 Manceau, E., Zabalza-Mezghani, I., Roggero, F., 2002. Use of experimental design
6 methodology to make decisions in an uncertain reservoir environment from reservoir
7 uncertainties to economic risk analysis, in: 17th World Petroleum Congress, Rio de Janeiro,
8 Brazil.
- 9 Maschio, C., de Carvalho, C.P.V., Schiozer, D.J., 2010. A new methodology to reduce
10 uncertainties in reservoir simulation models using observed data and sampling techniques.
11 Journal of Petroleum Science and Engineering 72, 110–119. doi:10.1016/j.petrol.2010.03.008
- 12 Maschio, C., Schiozer, D.J., 2014. Bayesian history matching using artificial neural network
13 and Markov Chain Monte Carlo. Journal of Petroleum Science and Engineering 123, 62–71.
14 doi:10.1016/j.petrol.2014.05.016
- 15 Nocedal, J., Wright, S.J., 2006. Numerical optimization, 2nd edn., Springer series in
16 operations research. Springer, New York.
- 17 Obi, E., Eberle, N., Fil, A., Cao, H., 2014. Giga Cell Compositional Simulation, in: IPTC-
18 17648-MS. Presented at the International Petroleum Technology Conference, Doha, Qatar.
19 doi:10.2523/IPTC-17648-MS.
- 20 Rahim, S., Li, Z., Trivedi, J., 2015. Reservoir Geological Uncertainty Reduction: an
21 Optimization-Based Method Using Multiple Static Measures. Mathematical Geosciences 47,
22 373–396. doi:10.1007/s11004-014-9575-5
- 23 Renard, P., Allard, D., 2013. Connectivity metrics for subsurface flow and transport.
24 Advances in Water Resources 51, 168–196. doi:10.1016/j.advwatres.2011.12.001
- 25 Scheidt, C., Caers, J., 2009. Representing Spatial Uncertainty Using Distances and Kernels.
26 Mathematical Geosciences 41, 397–419. doi:10.1007/s11004-008-9186-0
- 27 Scheidt, C., Caers, J., 2010. Bootstrap confidence intervals for reservoir model selection
28 techniques. Computational Geosciences 14, 369–382. doi:10.1007/s10596-009-9156-8
- 29 Scheidt, C., Caers, J., Chen, Y., Durlofsky, L.J., 2011. A multi-resolution workflow to
30 generate high-resolution models constrained to dynamic data. Computational Geosciences 15,
31 545–563. doi:10.1007/s10596-011-9223-9
- 32 Scheidt, C., Li, V., Caers, J., 2018. Bayesian Evidential Learning, in: Quantifying Uncertainty
33 in Subsurface Systems, Geophysical Monograph Series. John Wiley & Sons, Inc., Hoboken,
34 NJ, USA, pp. 193–215. <https://doi.org/10.1002/9781119325888.ch7>
- 35 Schiozer, D.J., Ligerio, E.L., Suslick, S.B., Costa, A.P.A., Santos, J.A.M., 2004. Use of
36 representative models in the integration of risk analysis and production strategy definition.
37 Journal of Petroleum Science and Engineering 44, 131–141. doi:10.1016/j.petrol.2004.02.010
- 38 Subbey, S., Christie, M., Sambridge, M., 2004. Prediction under uncertainty in reservoir
39 modeling. Journal of Petroleum Science and Engineering 44, 143–153.
40 doi:10.1016/j.petrol.2004.02.011



- 1 Suzuki, S., Caers, J., 2008. A Distance-based Prior Model Parameterization for Constraining
2 Solutions of Spatial Inverse Problems. *Mathematical Geosciences* 40,445-469.
3 doi:10.1007/s11004-008-9154-8
- 4 Suzuki, S., Caumon, G., Caers, J., 2008. Dynamic data integration for structural modeling:
5 model screening approach using a distance-based model parameterization. *Computational*
6 *Geosciences* 12, 105–119. doi:10.1007/s10596-007-9063-9
- 7 Thiele, M.R., Batycky, R.P., Blunt, M.J., Orr Jr, F.M., 1996. Simulating flow in
8 heterogeneous systems using streamtubes and streamlines. *SPE Reservoir Engineering* 11, 5–
9 12. doi:10.2118/27834-PA
- 10 Wang, K., Liu, H., Chen, Z., 2015. A scalable parallel black oil simulator on distributed
11 memory parallel computers. *Journal of Computational Physics* 301, 19–34.
12 doi:10.1016/j.jcp.2015.08.016
- 13 Xie, J., Yang, C., Gupta, N., King, M.J., Datta-Gupta, A., 2015. Integration of Shale-Gas-
14 Production Data and Microseismic for Fracture and Reservoir Properties With the Fast
15 Marching Method. *SPE-161357-PA* 20, 347–359. doi:10.2118/161357-PA
- 16 Zabalza-Mezghani, I., Manceau, E., Feraille, M., Jourdan, A., 2004. Uncertainty management:
17 From geological scenarios to production scheme optimization. *Journal of Petroleum Science*
18 *and Engineering* 44, 11–25. doi:10.1016/j.petrol.2004.02.002
- 19
- 20



11 Appendix: Error correction

We propose an optional correction stage to reduce the errors between the surrogate curves and the true curves. This correction makes the assumption that errors have zero mean, are smooth, and can be characterized only from the models of set \mathcal{S}_1 . It is similar in spirit to the “gobal error model” introduced by Josset and Lunati (2013), but it considers production profiles instead of breakthrough curves and uses time-dependent weighting instead of a global distance-based weighting. This correction takes the form of a vertical shift map (Section 8), computed as follows:

- For every known model (belonging to \mathcal{S}_1) the difference between the true profile and the reconstructed one is computed (**Fig. 9a**).
- These differences are projected on the analytical curves in the plot space (**Fig. 9b**)
- The differences are then interpolated for each time step using simple kriging with a 0 mean. The juxtaposition of all the 1D corrections over all the time steps covers the entire domain FOPT versus time (**Fig. 9c**).

The final corrected production profiles are finally obtained by subtracting the error estimate from the reconstructed analytical profiles.

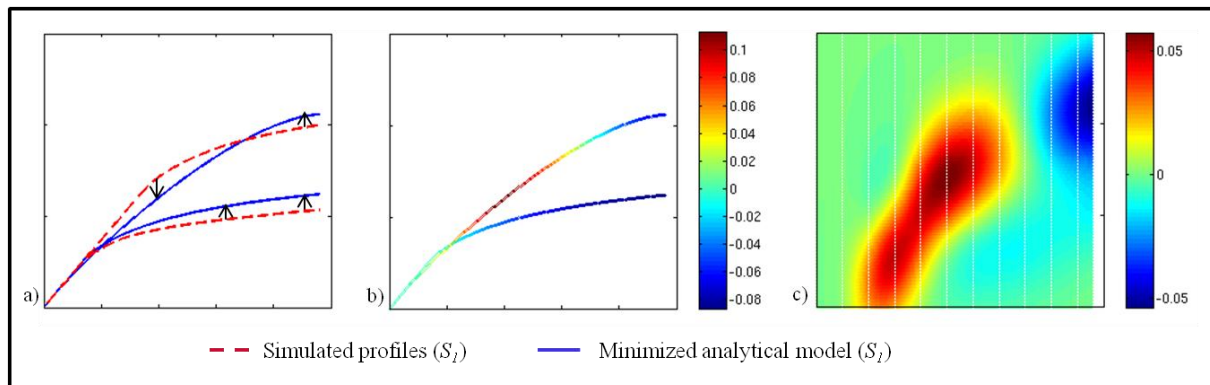


Fig. 9 – (a) Error computed between simulation profiles and analytical profile for two models belonging to \mathcal{S}_1 . (b) Mapping error on analytical model. (c) Interpolation of the error on the entire plot domain to create an error map.