Pet.Sci.(2014)11:265-271
DOI 10.1007/s12182-014-0339-0

On the uncertainty of interwell connectivity estimations from the capacitance-resistance model

Gustavo A Moreno[1, 2, 3]* and Larry W Lake[3]

[1] YPF, Macacha Güemes 515, Buenos Aires, Argentina
[2] YPF-Tecnología, Baradero S/N, (1925), Ensenada, Buenos Aires, Argentina
[3] Center for Petroleum and Geosystems Engineering, University of Texas at Austin, Austin, TX, USA

Abstract: The capacitance-resistance model (CRM) is an alternative to conventional reservoir simulation. CRM, a simplification of complex numerical models, uses production and injection rates to infer a reservoir description. There is no prior geologic model. The principal output of CRM fitting is the fraction of injected fluid (usually water) that is produced at a producer at steady-state. These fractions are interwell connectivities. Interwell connectivities are fundamental information needed to manage waterfloods in oil reservoirs. The data-driven CRM is a fast tool to estimate these parameters in mature fields and allows one to make full use of the dynamic data available. This paper considers the problem of setting an upper bound on the uncertainty of interwell connectivities for linear-constrained models. Using analytical bounds and numerical simulations, we derive a consistent upper limit on the uncertainty of interwell connections that can be used to quantify the information content of a given dataset.

Key words: Data-driven models, capacitance-resistance model, secondary recovery, waterflood optimization, interwell connectivities

1 Introduction

For the better part of a half-century, modeling of fluid flow through hydrocarbon-bearing reservoirs has dominated reservoir simulation. This is a process whereby the complex non-linear dynamics of fluid flow, described by partial differential equations (PDEs) of fluids, is approximated on a grid block or cell scale. This process is extremely data intensive, requiring as it does a prior geologic model and a host of other inputs that are either known or unknowable. The objective, in most cases, of numerical simulation is to determine the input-output relationship that represents the state of the reservoir system. Here we address an alternate approach, one that directly captures the relationship with a much smaller (and simpler) model. The approach is the capacitance-resistance model (CRM). For mature fields, where water saturations are typically large, the PDEs become approximately linear, guaranteeing a linear response at least for stationary rates (Chierici, 1994). The coefficients of such a linear relation between the total produced rate and the injection rates of neighboring wells is an intrinsic property of the reservoir (Yousef et al, 2006), which we call interwell connectivities. In the general case of a non-constant injection rate and compressible media and fluids one has to incorporate new features in the model, other than simple connectivities, to fully describe the dynamics of the system. To the best of our knowledge the simplest model that can incorporate compressibility effects is the CRM (Yousef et al, 2006). This model is complex enough to capture the main features of the system but simpler enough to be calibrated using hard dynamic measurements such as injection and production rates. There are many ways to fit the parameters in the model (Weber, 2009; Liang, 2010), all involving a minimization of a non-linear objective function (an associated least-square problem) with constraints in the variables. Despite many successful applications of this technique (Sayarpour et al, 2009, Nguyen et al, 2011), the limitations in the analysis imposed by available data have only be partially analyzed (Weber, 2009; Yousef et al, 2006) in this context. This work studies the deviation of the parameters estimates with noise and also their dependence with the time fluctuations of the input to understand the problems in the fit and possible solutions. In particular, we show that the unconstrained optimization problem for the connectivities can be used to set upper bounds on uncertainty of the constrained case. We explore the implications of these bounds to avoid spurious correlations.

*Corresponding author. email: gustavo.moreno@ypf.com
Received April 9, 2013

2 Theory and bounds

We will assume that the system of interest satisfies the

conditions for which the simplified description given by CRM is valid. For this system we have measurements of injection rates $I_i(t_k)$ in injector i and total liquid production rates $Q_j(t_k)$ in producer j in certain time periods t_k (typically monthly). Injection rates have small uncertainties compared to production rates and thus their uncertainties will be neglected here. Then the CRM-predicted total liquid production rate $q_j(t)$ ($Q_j(t_k)$ is calculated; $q_j(t)$ is the model prediction) will evolve according to:

$$\tau_j \dot{q}(t) + q_j(t) = \sum_{\alpha} f_{i(j,\alpha)j} I_{i(j,\alpha)}(t) \tag{1}$$

This evolution assumes a system of constant compressibility, a linear relationship between well pressure and flow, and constant bottom hole flowing pressures. Here α= 1, 2, …, n_j, where n_j is the number of neighbors of the j-th producer and the mapping $i(j, \alpha)$ stands for the i-th injector in the field that is connected to the j-th producer, uniquely identified also as the α-th neighbor of the j-th producer. This notation may seem cumbersome at first, but it will be simpler to span the indexes listing the neighbors per producer because here only neighboring interactions to j-th producer matter (of course we still do not define a priori the neighbor list). The time constant τ_j is a measure of the response time in the producer and $f_{i(j,\alpha)j}$ are the connectivities to the neighboring injectors. This equation can be easily integrated:

$$q_j(t) = q_j(0)\mathrm{e}^{-t/\tau_j} + \sum_{\alpha} f_{i(j,\alpha)j} \left\langle I_{i(j,\alpha)} \right\rangle_j (t) \tag{2}$$

where $\left\langle I_{i(j,\alpha)} \right\rangle_j (t)$ is a function of time obtained by averaging $I_{i(j,\alpha)}(t)$

$$\left\langle I_{i(j,\alpha)} \right\rangle_j (t) = \int_0^t \frac{\mathrm{e}^{-(t-s)/\tau_j}}{\tau_j} I_{i(j,\alpha)}(s)\,\mathrm{d}s \tag{3}$$

In the following we will restrict the analysis to the connectivities, which are the most sensitive parameters to the fit (Kaviani, 2012), moreover time constants in a well-behaved case should be small (on the order of months) and thus the conclusions derived below for τ_j constant will hold approximately also in the general case. Furthermore, to find analytical estimates we assume that the time window where data is provided is much larger than the typical response time τ_j thus for almost all times we have $t/\tau_j >> 1$ and Eq. (2) takes the approximate form:

$$q_j(t) \cong \sum_{\alpha} f_{i(j,\alpha)j} \left\langle I_{i(j,\alpha)} \right\rangle_j (t) \tag{4}$$

Thus the problem including the response time τ_j has exactly the same linear form as the incompressible (τ_j=0) case, but instead of using $I_{i(j,\alpha)}$ the injection rate is replaced by the τ_j-averaged $\left\langle I_{i(j,\alpha)} \right\rangle_j (t_1)$. This replacement represents no loss of utility because "rate" measured in the oilfield is always cumulative taken over a finite time interval.

The connectivity estimates are derived from minimizing the difference between the prediction $q_j(t)$ and the data $Q_j(t_k)$. If the liquid production rate is affected by random uncorrelated zero-mean noise sources, the appropriate objective function for this problem (Box et al, 2008) is:

$$E\left(\left\{f_{i(j,\alpha)j}\right\}\right) = \sum_{j,k} \left(q_j(t_k) - Q_j(t_k)\right)^2 \tag{5}$$

where $E(\{f_{i(j,\alpha)j}\})$ stands for a function E that depends on the complete set of connectivities $\{f_{i(j,\alpha)j}\}$. The identification of the statistical expectation operator with temporal data is a form of the ergodic assumption (Jensen et al, 2000). As is usual, ergodicity cannot be proven, only assumed. However, its use here opens up the use of statistical procedures and insights that have been developed over the years. Least squares are obtained from the assumption of maximum likelihood for uncorrelated Gaussian noise. In what follows we will always use this objective to calculate f_{ij} and test the effects of different noise sources.

The minimization of the objective as given by Eq. (5) is a quadratic programming problem for which there exists a unique global optimum when the quadratic form is non-degenerate. However, note that variables are constrained by $f_{ij}>0$ (the non-negativity constraint, or NNC) and $\sum_j f_{ij} \leq 1$ (when equal to one corresponds to a balanced field or block, denoted in what follows by BC). One of the objectives of this paper is to quantify the deviations of such an optimum point with the noise in $Q_j(t_k)$ and to understand how the degeneracy of the solutions is associated with the input signal (injection rates). Error analysis of the coefficients in linear regression is not new (for example (Jensen et al, 2000)). What is new here is error analysis of the coefficients in constrained linear regression for this model structure and the associated input design problem.

The constraints on the fitting parameters impose limitations on the possibilities for explicit solutions, so we will show how to handle the problem incorporating them at different stages and solving numerical cases when analytical estimates are not possible. Let us first consider the simplest case, where the system is approximately described by non-zero connectivities, with real values $F_{i(j,\alpha)j}$, inside the open set associated to the feasible region (that is $\sum_j F_{i(j,\alpha)j} \leq 1-\theta$ and $F_{i(j,\alpha)j} \geq \theta$ for some small θ) and that the values of $Q_j(t_k)$ are affected by a small noise term

$$Q_j(t_k) = g_j(t_k) + \xi_j(t_k) \tag{6}$$

with $\xi_j(t_k)$ a stochastic variable (which will be assumed uncorrelated) and $g_j(t_k)$ being calculated using the real values that we want to reveal

$$g_j(t_k) = \sum_{\alpha} F_{i(j,\alpha)j} \left\langle I_{i(j,\alpha)} \right\rangle_j (t_k) \tag{7}$$

This non-zero connectivities will exist between certain neighbors (listed by α) of each producer. Then the optimum point can be calculated analytically as:

$$\sum_{\alpha=1}^{n_j} A_{\alpha\beta}^{(j)} f_{i(j,\alpha)j} = IQ_{j\alpha} \tag{8}$$

where the matrix $A_{\alpha\beta}^{(j)}$ is given by

$$A_{\alpha\beta}^{(j)} = \sum_k \left\langle I_{i(j,\alpha)} \right\rangle_j (t_k) \left\langle I_{i(j,\beta)} \right\rangle_j (t_k) \tag{9}$$

and the matrix $IQ_{j\alpha}$ is calculated as a sum of products between injection and production rates

$$IQ_{j\alpha} = \sum_k \left\langle I_{i(j,\alpha)} \right\rangle_j (t_k) Q_j (t_k) \tag{10}$$

The set of matrices $A^{(j)}$ determines the values of the connectivity estimates, but more interestingly they will play an essential role in setting upper bounds on deviations in the general case. These matrices, which are symmetric and positive semi-definite, are formed by inner-products of the neighboring injection history vectors of W components

$$\left(\left\langle I_{i(j,\alpha)} \right\rangle_j (t_1),\ \left\langle I_{i(j,\alpha)} \right\rangle_j (t_2),\ \cdots,\ \left\langle I_{i(j,\alpha)} \right\rangle_j (t_W) \right)$$

and thus they encode information about the structure of the temporal series and its correlations. This vector is no more than the component-wise τ_j-averaged injection rate history $\left\langle I_{i(j,\alpha)} \right\rangle_j (t)$ for each time period t_1, …, t_W. Now, if these matrices are invertible (that is, if the n_j injection vectors which are neighbors to producer j are linear independent), one finds a unique solution:

$$f_{i(j,\alpha)j} = \sum_{\beta=1}^{n_j} A_{\alpha\beta}^{-1(j)} IQ_{j\beta} \tag{11}$$

which of course coincides with $F_{i(j,\alpha)j} = f_{i(j,\alpha)j}$ in the case of zero noise $\xi_j(t_k)=0$.

However when the matrix is singular, the solution to the problem is non-unique. The matrix becomes singular when the injection rate vectors are linearly dependent, a situation that can happen in field-wide injections where fluids are injected through common surface manifolds. Albertoni and Lake (2003) used a uniqueness index to identify this problem; Yousef et al (2006) tested the so-called Ridge regression.

As can be expected, the transition is not sharp, and when a particular matrix becomes closer to being singular the uncertainty in the estimate, which is directly related to this ambiguity, will become more sensitive to noise. To show this, let us assume that the random uncorrelated errors $\xi_j(t_k)$ satisfy $\xi_j(t_k)/[\xi_j(t_k)\xi_{j'}(t_{k'})]=\delta_{kk'}\delta_{jj'}\sigma_j^2$, here [.] denotes average over noise realizations. These errors will yield to deviations $\delta f_{i(j,\alpha)j} = f_{i(j,\alpha)j} - [f_{i(j,\alpha)j}]$ in the estimated values of the connectivities. Under the previous assumptions (and assuming that the errors are small enough to maintain all the constraints inactive) we have:

$$\left[\delta f_{i(j,\alpha)j}{}^2 \right] = A_{\alpha\alpha}^{-1(j)} \sigma_j^2 \tag{12}$$

Eq. (12) clarifies the relation between uncertainty and data structure. Of course, the noise in total production rates σ_j^2 generates indeterminacy in the connectivities but, more importantly, the data structure, encoded in the inner product matrix $A_{\alpha\alpha}^{(j)}$ of the neighboring injector history, defines the amplification factor for such a relation. In particular when the matrix is closer to an ill-defined condition, where injection vectors are linear dependent, the coefficient will diverge. In contrast, for certain injection signals the matrix $A_{\alpha\beta}^{(j)}$ will be better ($A_{\alpha\alpha}^{-1(j)}$ smaller), making the dataset more informative because the estimation will be more robust to noise. In the general framework of time series analysis $A_{\alpha\beta}^{(j)}$ are the blocks of the so-called Fisher information matrix (Box et al, 2008). Most of the literature on the subject is restricted to unconstrained cases; thus in the following we analyze possible deviations for this model structure in the presence of active constraints. Finally, note that we can determine the error estimated in Eq. (12) without actually fitting the model to data. In the following we use analytical estimates when possible and compare the results with numerical simulations that span different types of active constraints.

3 Numerical simulations

3.1 Unconstrained case

We will show in the following that even when constraints are active, the set of matrices $A_{\alpha\beta}^{(j)}$ will set upper bounds on uncertainty and guide the optimization of future injection with the objective of minimizing uncertainty. To contrast these ideas with analytical bounds, we compare the estimates with numerical simulations that include, progressively, the effect of constraints. Our numerical examples typically involve inverted 5-spot patterns generated with 10 lines of 10 producers and 9 lines of 9 injectors with connectivities populated randomly between the first neighbors in the inverted 5-spot. Injection rates were varied randomly with 15% probability between 50 and 100 m^3/d (units here will be irrelevant because of the linearity of the underlying model) and producers were assumed to evolve according a CRM evolution. Time constants were fixed in 1 month and monthly data were fitted for each realization of the stochastic noise $\{\xi_j(t_k)\}$ in the original production rates. Gaussian and white noise were used; the same results described below hold for both cases if the proper measure of noise level is chosen. In what follows we describe explicitly the white noise examples, in these cases the stochastic variables are taken in the range $\xi_j(t_k)$ in $[-L/2, L/2]$ thus the parameter L will be referred as noise amplitude or noise level (analogously for Gaussian noise the width of the Gaussian distribution will measure the noise level) which also defines the mean deviation $[\xi_j(t_k)\xi_{j'}(t_k)]=\delta_{kk'}\delta_{jj'}L^2/12$. Fig. 1 shows a typical single producer in the synthetic field and its associated fit. To study the dependence of the fitting parameters with noise level we used a sequence of increasing noise levels L. For each L we have generated 100 realizations of the set $\{\xi_j(t_k)\}$. Each of those realizations was fitted and the mean connectivities and root mean square deviations were calculated for each noise level. Thus, for each noise level L we end up with a distribution function for the fitted connectivities (that is represented in these simulations by 100 samples). The fitting uses a (least-squares) quadratic programming problem for which there are many possible numerical schemes (Bjorck, 1996); our implementation is a gradient-based scheme that takes advantage of the model structure (e. g. sparse connections).

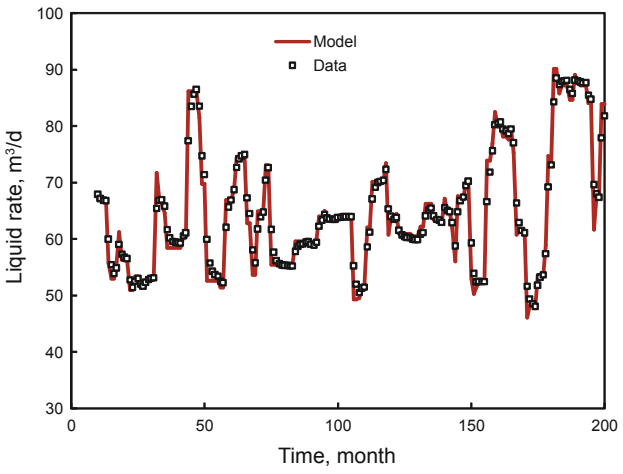

Fig. 1 Typical liquid production rate data and fit in synthetic simulations. This is a typical response of one producer in the field that has 100 producers and 81 injectors.

Fig. 2 shows a typical response for the mean connectivities of one injector as a function of the noise level L in the case described above where constraints (NNC and BC) are inactive. Now the mean values remain unchanged and deviations agree with the analytical estimates, as shown in Fig. 3. The comparison between calculated and estimated deviations for all the connectivities in the field is visualized in Fig. 4, where we plot a histogram for the ratio $r=\delta f_{\text{measured}}/\delta f_{\text{estimated}}$ for all the values of L and possible pairs. This clearly shows a distribution centered at $r=1$. The finite width is generated by the finite size of the sampling (100 noise realizations of each experiment). Now let us consider the cases where one or more constraints are active.

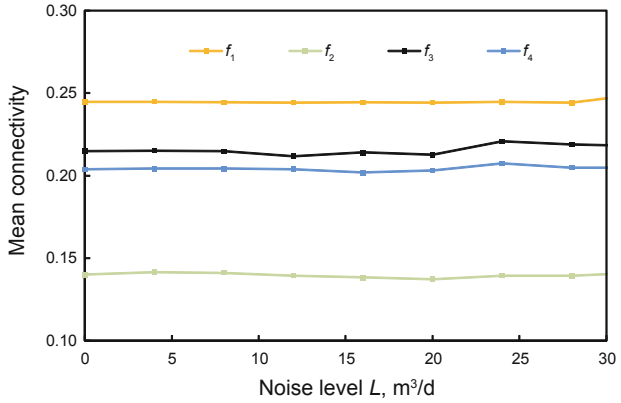

Fig. 2 Mean connectivities vs. noise level (L) for a typical case when BC and NNC are not active. The parameters f_1-f_4 are the connectivities to four producers.

3.2 Constrained cases

To be able to work out further the analytical estimates in constrained cases other assumptions must be taken. Let us start by assuming that we have only positive connectivities between all the pairs in the problem and the producers have the same time constant $\tau_j=\tau$, in this case, and when deviations

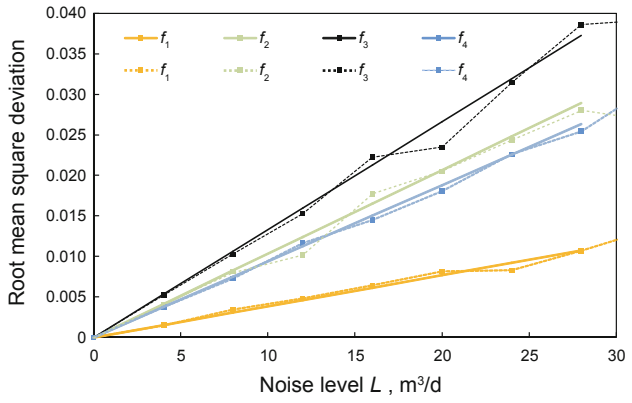

Fig. 3 Root mean square deviation (dimensionless) from the mean values of connectivities vs. noise level for a typical case when BC and NNC are not present. Solid lines are calculated analytically. Numerical calculations fluctuate because of finite sampling.

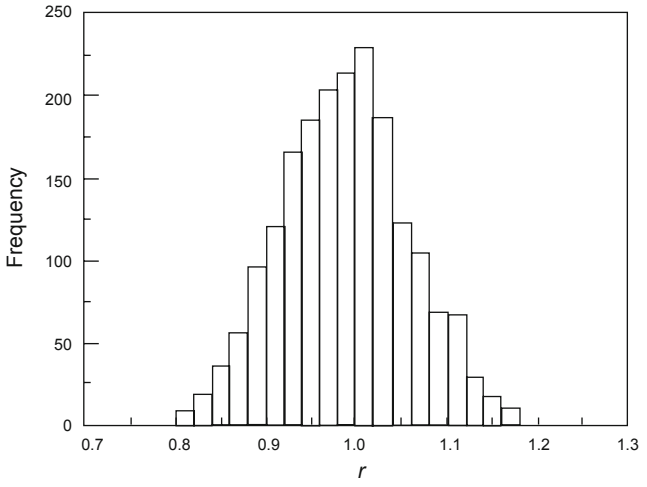

Fig. 4 Frequency of the ratio $r=\delta f_{\text{measured}}/\delta f_{\text{estimated}}$ for all the connectivities and noise levels (L). Neither BC nor NNC are active. Here there are only non-zero nearest-neighbor connections.

of the data from the physical values are not large, we can calculate the minimum of the objective under the BC

$$\sum_j f_{ij} = 1 \tag{13}$$

As we are assuming now that, in principle, any pair can be connected we do not need to distinguish between neighbors here (that is, the mapping $i= i(j, \alpha)$ is unnecessary here for this particular analytical estimate). Thus, using Lagrange multipliers it is easy to show that

$$f_{ij} = \sum_{i'} A^{-1}_{ii'} \left[IQ_{i'j} + \frac{1}{N}\left(\sum_l A_{i'l} - \sum_s IQ_{i's} \right) \right] \tag{14}$$

where N is the number of producers. Taking into account the deviations associated with the fluctuations $\{\xi_j(t_k)\}$ one can use the previous equation to obtain the covariance matrix of the fluctuations from the mean value $\delta f_{i(j,\alpha)j}= f_{i(j,\alpha)j} -[f_{i(j,\alpha)j}]$, namely

$$\left[\delta f_{ij} \delta f_{i'j'} \right] = A^{-1}_{ii'} \sum_s \left(\delta_{sj} - 1/N \right)\left(\delta_{sj'} - 1/N \right)\sigma_s^2 \tag{15}$$

In this case fluctuations for different producers are correlated through the BC. To understand the typical

behavior we take two particular cases. For example, if $N\to\infty$ the unconstrained case is recovered as can be expected intuitively; when $N\to\infty$ the constraint for each injector has no effect on producers and fluctuations become uncorrelated. For the other case assume $\sigma_s^2=\sigma^2$ to compute the autocorrelation $\left[\delta f_{ij}^2\right]=A_{ii}^{-1}\sigma^2(1-1/N)$. So the BC suppresses part of the fluctuations because of the presence of an extra relation between the connectivities. In typical situations some of the connections are eliminated a priori thus yielding a problem similar to the previous one; however, in such a case equations cannot be decoupled systematically unless the network structure (the neighbor list) is defined. Nevertheless it can be seen from numerical simulations that when BC is active and only first neighbor connections are positive, the resulting fluctuations are always smaller than the unconstrained analytical bound, including the cases where some connectivities are eliminated a priori (this is the case, for example, in a balanced nearest neighbor interacting field). Again, the mean values of the connectivities do not change with noise, and root mean square deviation from the mean is always upper-bounded by the unconstrained estimate. So in this case, where some connectivities are eliminated a priori and the field is balanced, the uncertainty using the unconstrained case is always an upper bound.

A more complicated case is the general situation, where some of the gains are (physically) very small or even null and no a priori elimination was carried out and/or when fluctuations are large enough to activate NNC. In these cases both the BC and NNC will be active for the noisy dataset. A typical case of evolution of the fitting with noise level for one of the injectors in the field is in Fig. 5. As can be seen, in contrast to the previous cases the mean values deviate from the correct answer (at L=0). The reason is that when the noise is large enough small gains will activate its NNC and thus fluctuations of small connectivities will be asymmetric, drifting the mean to larger values (pull up). When BC is not active the latter drift is the only effect but when BC becomes active some of the others non-small mean gains (a priori not identifiable) for the same injector will drift downwards. Remarkably, deviations from the new mean at each noise level still remain smaller than the unconstrained bound, as shown in Fig. 6, however this can be expected from intuition but it is not easy to prove. The latter observation holds for all the gains in the problem, Fig. 7(a) shows the statistics of ratios $\delta f_{\text{measured}}/\delta f_{\text{estimated}}$ for two different values of the noise level. There is a clear shift to values less than one. Moreover, larger values of the noise level tend to spread the distribution and skew it toward smaller fluctuations, as can be seen in the density histogram of Fig. 7(b) where all values of noise level are considered. This allows estimation not only of an upper bound on the deviations from the mean but also gives a measure of the fluctuations in δf, and thus whether a certain NNC is being activated or not can be estimated. When a NNC is activated the associated connectivity mean will increase and some of the other connectivities for the same injector will be subject to possible drifts (mean value) in addition to their own deviations.

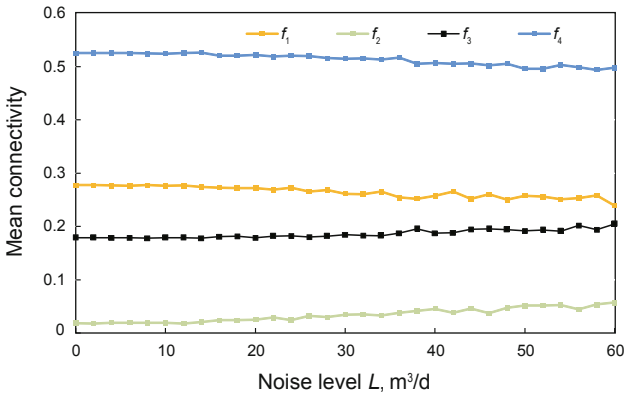

Fig. 5 Evolution of mean connectivities with noise level for a balanced reservoir. The small connectivity f_2 (green) drift from its real value (at L=0) when the noise is large enough to activate the NNC constraint. When this happens other elements (e.g. f_1 and f_4) will deviate downwards because of BC.

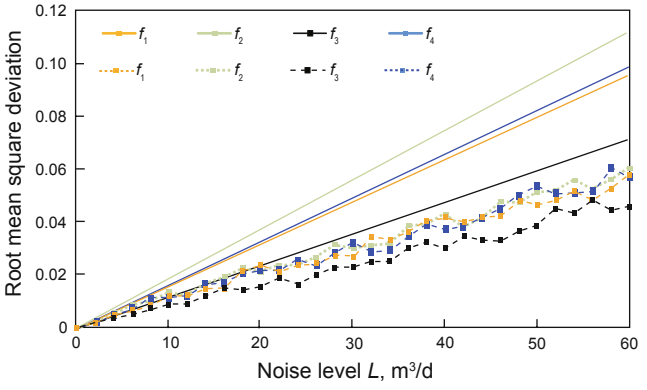

Fig. 6 Evolution of the root mean square deviation of the connectivities from their mean values as a function of noise level (balanced reservoir). In this case both NNC and BC are active. The drift of the means in Fig. 5 starts when the noise displayed in this plot is large enough to activate the NNC for the small gains.

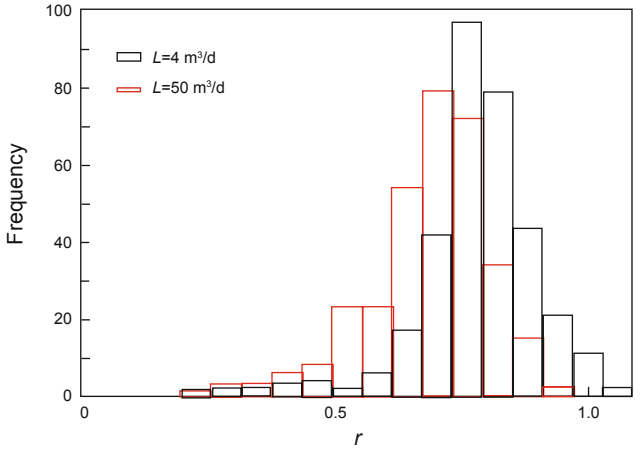

Fig. 7(a) Frequency of the ratio $\delta f_{\text{measured}}/\delta f_{\text{estimated}}$ for all the connectivities at two different noise levels (L) in the case of a balanced reservoir. Fluctuations are bounded from above by the unconstrained case (r <1).

To the best of our knowledge this general case, where both BC and NNC are active, cannot be treated in closed

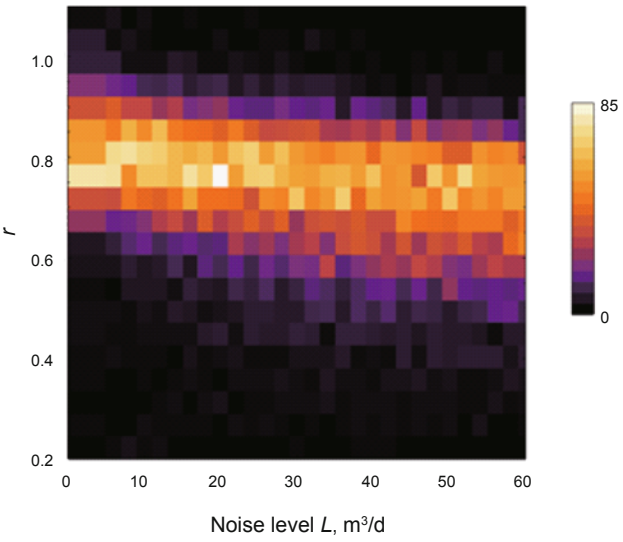

Fig. 7(b) Density histogram for $r=\delta f_{measured}/\delta f_{estimated}$ and for all the values of the noise level L. The color represents frequency (number of counts).

form. However the previous analysis based on linear response can be used to set bounds on uncertainty and understand numerical findings.

Similar problems appear when non-physical connectivities are not eliminated a priori. If the dataset is absolutely noise free and the matrices invertible, there will be no spurious connectivities. However, when the noise is added to the data small gains will be pulled up drifting the means of the physical interactions and thus distorting the fit. This supports the idea of setting to zero connectivities below a threshold value in a pre-processing step to mitigate the effect of spurious correlations on the fit of physical connectivities (Weber, 2009). The following section explores some practical applications of the previous discussions.

4 Practical aspects and discussion

This section discusses applications of the previous ideas that are of practical interest, based on the use of the matrices $A_{\alpha\beta}$ and $A^{-1}_{\alpha\beta}$ as a measure of linear independence in the context of CRM, specifically giving a direct estimate of the deviation growth rate of uncertainty via its diagonal elements $A^{-1}_{\alpha\alpha}$. This matrix imposes restrictions in the conclusions that can be extracted from the data available which in this framework can be estimated without a previous fit.

Let us consider as a first example the case of a single producer connected to a pair of injectors, in this situation the information matrix is 2×2 and the inverse can be easily handled. The bounds for the deviation growth rate are

$$A^{-1}_{11}=\frac{A_{22}}{A_{22}A_{11}-\left(A_{12}\right)^2} \tag{16}$$

$$A^{-1}_{22}=\frac{A_{11}}{A_{22}A_{11}-\left(A_{12}\right)^2} \tag{17}$$

When the injection histories are non-overlapping $A_{12}=0$ and $A^{-1}_{ii}=1/A_{ii}$. Thus the more the injection is measured the lower the bound will be, as expected measurement decreases uncertainty (A_{ii} is an increasing function of the number of measurements). Moreover, for two given signals $I_1(t)$ and $I_2(t)$ of fixed shape the coefficients A_{11} and A_{22} are determined once the measurement time steps t_k are determined. In contrast, the coefficient A_{12} depends on the overlap of these signals. Eqs. (16) and (17) tell us that overlapping two signals of fixed shape ($A_{12}\neq 0$) always degrade our conclusions (the denominator decreases for increasing overlap).

To further illustrate how these ideas can be used we consider a more general example of a single producer and many injectors. We want to know if this producer is interacting or not with a certain neighboring injector, let us say I_1. Of course, one would like to consider all possible interactions from the very beginning but this may, in principle, not yield to a well-defined problem because enough data-points must be available to unambiguously solve the fitting problem (for example more time periods than possible connections; that is fitting parameters, must be available but this is not sufficient). To study to what extent this is possible, one can construct the matrix $A_{\alpha\beta}$ of neighboring injectors including an increasing number of injectors and see how the coefficient A^{-1}_{ii} changes both with this increasing number of neighbors and with the number of periods T available. When we add data-points (increase T) $A_{\alpha\beta}$ does not change dimension; in contrast the number of neighbors N defines the dimension of the matrix $A_{\alpha\beta}$, so we write this explicitly as $A_{\alpha\beta}|_N$. The degradation of the coefficient $A^{-1}_{11}|_N$ will measure how sensitive this inter-well connectivity estimate is to the existence of other possible interactions around. Fig. 8 shows the value of $A^{-1}_{11}|_N$ as a function of the time periods T in the data set and the possible neighboring interacting elements N. As can be expected, the more possible neighbors in the problem, the more sensitive the estimate will be to noise in the data. In fact, using second-order perturbation theory one can find that when a new possible neighbor is included in the problem of order N with matrix $A^{-1}|_N$, the new problem of order N+1 with matrix $A^{-1}|_{N+1}$ satisfies:

$$A^{-1}_{11}\Big|_{N+1}=A^{-1}_{11}\Big|_N+\frac{1}{\sum_k\left\langle I_{N+1}\right\rangle_j\left(t_k\right)\left\langle I_{N+1}\right\rangle_j\left(t_k\right)}\left(\sum_{l=1}^{N}A^{-1}_{1l}\Big|_N v_l\right)^2 \tag{18}$$

with

$$v_l=\sum_k\left\langle I_{N+1}\right\rangle_j\left(t_k\right)\left\langle I_l\right\rangle_j\left(t_k\right) \tag{19}$$

Eq. (19) shows explicitly that when the new neighbor has history overlapping with the previous injection signals ($v_l>0$) it always increases the uncertainty ($A^{-1}_{11}\Big|_{N+1}>A^{-1}_{11}\Big|_N$). And when it is temporally uncorrelated (for example because non-zero values of the averages do not overlap) with the other injectors this results in no harm to our estimation of the connectivity. The previous formula agrees with intuition (that is, a new well in the problem never can help) and gives an explicit approximation for the perturbed uncertainty.

Fig. 8 shows a useful way to select an uncertainty tolerance level a priori and explore how many interactions we will be able to decouple (that is a contour line of fixed A^{-1}_{11}) from the dataset. Of course this plot depends on the particular injection history, which defines the matrix $A_{\alpha\beta}|_N$, but there are

some generic features: more possible interactions increase uncertainty (as shown before), more time periods always decrease uncertainty, and finally larger response time in producers results in larger uncertainty. The latter effect is because the injection fluctuations below the producer response time τ_j are washed out, giving rise to a smoother set of vectors. Of course the same reasoning applies to study the existence of spurious correlations in more general cases, but when more than one producing well is present one has to take into account possible drifts as was described in the previous section.

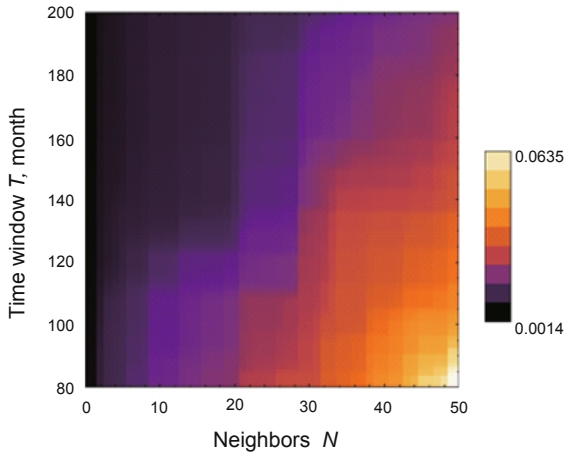

Fig. 8 Coefficient sqrt(A_{11}^{-1}) (in 1/(m^3/d) units) as a function of the time window T in the fit and number of possible interactions (neighbors) N in the problem. This construction allows the establishment of a safe region to a given tolerance level. For example 25 neighbors with a time window of 110 months will have a connectivity error of 0.02 per m^3/d of noise.

5 Summary and conclusions

This work has analyzed how the temporal structure in injection (input) dataset and noise in liquid production rate impose limitations on the information that can be extracted from correlations of injection/production rates in mature oil fields where CRM is applicable. Using analytical estimates and numerical simulations we show that uncertainty for the unconstrained parameter estimates can be used to set upper bounds on the fluctuations from the means in the constrained case. Using the previous observation maximum deviations can be estimated and the information content of the dataset studied. In particular, we show that input temporal series dominates the uncertainty estimates; a general framework to study possible spurious correlations is possible. Moreover, second-order perturbation theory applied to this problem yields analytical bounds that show explicitly the effect of degradation in parameter estimates when more possible well-to-well interactions are allowed.

Acknowledgements

Gustavo A. Moreno is grateful to YPF for financial support and to the Center for Petroleum Asset Risk Management of the University of Texas at Austin for hospitality and an exciting research environment. The author is also grateful to Dr. Carlos A. Glandt for setting challenging objectives and constant support on initiatives.

Larry W. Lake holds the W.A. (Monty) Moncrief Centennial Chair at the University of Texas.

References

Albertoni A and Lake L W. Inferring interwell connectivity only from well-rate fluctuations in waterfloods. SPE Reservoir Evaluation & Engineering. 2003. 6(1): 6-16 (Paper SPE 83381 PA)

Bjorck A. Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics. 1996

Box G E P, Jenkins G M and Reinsel G C. Time Series Analysis: Forecasting and Control. Wiley, 4th ed. 2008

Chierici G L. Principles of Petroleum Reservoir Engineering. Springer-Verlag Vol. II. 1994

Jensen J L, Lake L W, Corbett P W M and Goggin D J. Statistics for Petroleum Engineers and Geoscientists. Elsevier. 2000

Kaviani D. Personal Communication, 2012

Liang X. A simple model to infer interwell connectivity only from well-rate fluctuations in waterfloods. Journal of Petroleum Science and Engineering. 2010. 70(12): 35-43

Nguyen A P, Lasdon L S, Lake L W and Edgar T F. Capacitance resistive model application to optimize waterflood in a West Texas field. Paper SPE 146984 presented at SPE Annual Technical Conference and Exhibition, 30 October-2 November 2011, Denver, Colorado, USA

Sayarpour M, Kabir C S and Lake L W. Field applications of capacitance-resistance models in waterfloods. SPE Reservoir Evaluation & Engineering. 2009. 12(6): 853-864 (Paper SPE 114983)

Weber D. The Use of Capacitance-Resistance Models to Optimize Injection Allocation and Well Location in Water Floods. Ph.D Dissertation. The University of Texas at Austin. 2009

Yousef A A, Gentil P, Jensen J L and Lake L W. A capacitance model to infer inter-well connectivity from production and injection rate fluctuations. SPE Reservoir Evaluation & Engineering. 2006. 9(6): 630-646 (Paper SPE 95322)

(Edited by Sun Yanhua)