Journal of Petroleum Science and Engineering 208 (2022) 109728

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

Information content in 4D seismic data: Effect of correlated noise

Dean S. Oliver

NORCE Norwegian Research Centre, Norway

ARTICLE INFO

Keywords:
Time-lapse seismic
4D seismic
History matching

ABSTRACT

Several types of data used in history matching for subsurface reservoir characterization have errors that are spatially or temporally correlated. Although it is often assumed that correlated observation errors decrease information content, using a simplified flow problem, we show that for data that are spatially dense (such as 4D seismic data), correlated observation errors result in higher information content than data with similar levels of errors, but without correlation. Unfortunately, correlations in the observation error are often unrecognized and difficult to estimate, especially if the correlation scale of the errors is similar to a characteristic length of the signal. In addition, many history matching algorithms are incapable of accounting properly for correlated observation error, so data are either thinned or the observation error is inflated to partially account for the lack of proper treatment. We show that neglecting correlations in the observation errors or inflating the variance, result in loss of information content. Finally, we show that it is possible to iteratively estimate the correlated observation errors through analysis of residuals after history matching.

1. Introduction

Correlated observation errors are ubiquitous in model-based data assimilation if for no other reason than the fact that models are never completely accurate representations of reality and deficiencies in a model generally result in errors that are correlated. Common general sources of model errors in data assimilation include under-parameterization, missing physics, and bias introduced in processing of data prior to assimilation. Time-lapse or 4D seismic data suffers in this regard as the attributes that are commonly used as data for history matching are derived from data that have been heavily processed. Roach et al. (2015) show that the effect of every step in the processing of 4D seismic data is to decrease the variability among surveys. The result, however, is an increase in bias which has long range correlation. When the “data” for data assimilation are attributes that have been inverted from the actual seismic data, the potential for spatially correlated errors increases (Thore, 2015). History matching of seismic data also requires coupling of more than one complex model, each of which is necessarily simplified. A relatively thorough discussion of sources of correlated observation and model errors in 4D seismic data can be found in a recent review paper on 4D seismic history matching (Oliver et al., 2021).

The term ‘4D seismic’ data refers to spatial 3D seismic data sets that are measured multiple times (hence the 4th dimension). One attractive feature of 4D seismic data is that the density of observations in the spatial domain is high compared to other types of data typically available for reservoir characterization. Consequently, the data are used to monitor changes in fluid saturation and pressure in regions between wells where other measurements are not available. Because of the spatial density and the ability to interpret changes in reservoir properties, it is often assumed that the 4D seismic data must contain large amounts of information for reservoir characterization and for identification of locations of bypassed oil. Yet the actual improvements in reservoir characterization resulting from assimilation of 4D seismic data have been modest, implying a lower level of information content than commonly assumed. It is natural to enquire as to the source of the mismatch between expectation and experience. As 4D seismic data have several distinctive characteristics (high spatial density, sharp saturation fronts, and long range correlated observation error), we focus on the effect of spatially correlated errors on information content for this type of data. In particular, we investigate the possible influence of correlated observation error in the reduction in 4D seismic data information.

Because of the extensive processing of seismic data, the errors in seismic attributes are generally more complex than described by a additive Gaussian model and information on reservoir properties will always be lost when a simplified model of observation error is chosen over a more accurate one. In this manuscript, we have focussed on the information loss when the errors are assumed to be uncorrelated because this is a nearly universal assumption in history matching; it results in large information loss and yet including observation error correlations for history matching is straightforward.

We begin in Section 2 with a discussion of various methods of defining information content for data, including Shannon information

E-mail address: dean.oliver@norceresearch.no.

https://doi.org/10.1016/j.petrol.2021.109728
Received 6 September 2021; Received in revised form 21 October 2021; Accepted 23 October 2021
Available online 1 November 2021

and relative entropy. We then introduce a numerical flow example in Section 2.3, which is subsequently used to illustrate the quantification of information content for estimation of log-permeability and saturation from repeated noisy observations of saturation. Despite the fact that we believe that correlated observation errors are the norm in most subsurface data assimilation problems, they are usually ignored and often not recognized. It is common to assume an independent noise model because it simplifies the calibration or history matching step in the data assimilation. It is also common to estimate the noise from the data themselves (without comparison with the model predictions). In Section 3 we investigate the consequences of neglecting correlations in observation error in the data assimilation, and the effect of inflating the observation error variance to possibly compensate for neglected correlations. Then, in Section 4 we discuss characterization of correlated observation errors and a method for utilizing the estimate of observation error covariance to extract more information from the data.

Although this investigation focusses on the effect of correlated observation error on information content on 4D seismic data, the general conclusions will apply to other types of closely spaced data (temporally or spatially) with correlated observation error. This includes data such as well production and injection rates and pressures from wireline formation testers. In many offshore locations, well production rates are measured using a test separator, but these well tests are infrequent — in some cases once per year. At intermediate times, the well production is allocated based on choke setting or on multiphase flowmeters with lower level of accuracy hence errors tend to be correlated in time (Folgerø et al., 2013; Sadri and Shariatipour, 2020).

2. Information content in data

A qualitative interpretation of information content is that it is a measure of how surprising an observation is, based on what is already known about the system. Data with high information content force us to substantially alter our assessment of uncertainty about quantities of interest, whereas data with low information content will have little effect on our assessment of uncertainty.

The information provided by an observation is sometimes defined simply in terms of the subsequent reduction in the expected error of the variable of interest relative to prior uncertainty in the variable (Neuman et al., 2012). The information content will thus have different values depending on the variable whose value we wish to estimate. In typical applications of history matching, the goal might be to calibrate a model of the reservoir that can subsequently be used to make predictions. In that case, we might define the information content in terms of the reduction in uncertainty in the parameters of the model (e.g. permeability, porosity, and fault transmissibility). Frequently, however, we may be more interested in the uncertainty in predictions from the model, such as locations of remaining oil in a petroleum reservoir or locations of contaminants in an aquifer, in which case we may wish to define the information content in terms of the reduction in uncertainty in the predictions.

2.1. Measures of information content

When we take a Bayesian approach to the quantification of uncertainty, the information content must sometimes be defined more broadly than simply the reduction in uncertainty, as it is possible for assimilation of data to result in an increase in uncertainty. A more general approach is to define the information content as the difference between the prior probability, p^{pr}, and the posterior probability, p^{po}, after conditioning to data (Majda et al., 2002; Xu, 2007; Petty, 2018; Chen, 2020). The Kullback–Leibler directed divergence (or relative entropy) from the prior to the posterior for model parameters $\mathbf{m}$,

$$D_{\text{KL}}(p^{\text{po}} \parallel p^{\text{pr}}) = \int d\mathbf{m}\, p^{\text{po}}(\mathbf{m}) \ln \left(\frac{p^{\text{po}}(\mathbf{m})}{p^{\text{pr}}(\mathbf{m})} \right)$$

can be used to quantify the information content in the observations. The information content measured this way has two advantages: (1) the information content is 0 if $p^{\text{po}}(\mathbf{m}) = p^{\text{pr}}(\mathbf{m})$, and (2) information content is positive if the two distributions are different. For the special case in which the prior distribution (perhaps after transformation of variables) is approximated as Gaussian with prior mean $\mathbf{m}_{\text{pr}}$ and prior model covariance $\mathbf{C}_{m,\text{pr}}$,

$$p^{\text{pr}}(\mathbf{m}) = (2\pi)^{-M/2} |\mathbf{C}_{m,\text{pr}}^{-1}|^{1/2} \exp\Big(-\frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^{\text{T}} \mathbf{C}_{m,\text{pr}}^{-1} (\mathbf{m} - \mathbf{m}_{\text{pr}}) \Big).$$

and the posterior distribution can be similarly approximated as Gaussian with posterior mean $\mathbf{m}_{\text{po}}$ and posterior covariance $\mathbf{C}_{m,\text{po}}$, the relative entropy can be written as (Xu, 2007)

$$D_{\text{KL}}\left(p^{\text{po}} \parallel p^{\text{pr}}\right) = \frac{1}{2}\left((\bar{\mathbf{m}}_{\text{pr}} - \bar{\mathbf{m}}_{\text{po}})^{\text{T}} \mathbf{C}_{m,\text{pr}}^{-1} (\bar{\mathbf{m}}_{\text{pr}} - \bar{\mathbf{m}}_{\text{po}}) \right) + \frac{1}{2}\left(\text{Tr}\left(\mathbf{C}_{m,\text{pr}}^{-1} \mathbf{C}_{m,\text{po}} \right) - k + \ln\left(\frac{\det \mathbf{C}_{m,\text{pr}}}{\det \mathbf{C}_{m,\text{po}}} \right) \right). \quad (1)$$

The first term on the right is a measure of the change in the mean of the distribution, while the second term is a result of the change in the covariance.

Despite the advantages of relative entropy as a measure of information content, the Shannon information content (SIC) is the more common measure. For problems with Gaussian prior and posterior distributions, the Shannon information content (in nats) is

$$\text{SIC} = \frac{1}{2} \ln \left(\frac{\det \mathbf{C}_{m,\text{pr}}}{\det \mathbf{C}_{m,\text{po}}} \right). \quad (2)$$

Questions about the usefulness of the SIC as a measure of information content have been raised several times in the data assimilation literature (Majda et al., 2002; Xu, 2007; Fowler and Van Leeuwen, 2012; Petty, 2018). One disadvantage of SIC is that it is not invariant to nonlinear transformations of the variables of interest so, for example, we would obtain different estimates of information content of data for estimation of log-permeability compared to estimation of permeability. In this paper, we will generally report the information content in terms of relative entropy.

Note that for Gauss-linear data assimilation problems the value of the relative entropy is sensitive to the actual observations, while SIC is a function only of the prior and posterior covariance matrices (Eq. (1)). For Gauss-linear problems, the posterior covariance is a function of *what* is measured, but it does not depend on the realization of the data. If $\mathbf{G}$ denotes the linear observation operator relating the model parameters to the data, the posterior covariance is

$$\mathbf{C}_{m,\text{po}} = \left(\mathbf{C}_{m,\text{pr}}^{-1} + \mathbf{G}^{\text{T}} \mathbf{C}_{d}^{-1} \mathbf{G} \right)^{-1}.$$

On the other hand, the relative entropy depends on the posterior mean of the estimated quantity, so it is a function of the actual realization of the data (with observation error). Thus, while it is possible to quantify the expected information gain using relative entropy, it is not possible to quantify the information content of a particular realization of the data prior to data acquisition using relative entropy. The consequence is that comparisons of information content based on relative entropy will be "noisier" than computations based on SIC and it may be necessary to average over several realizations of the truth to draw meaningful conclusions.

2.2. Effect of correlated observation errors on information content

Our primary objective in this paper is not to determine the information content of actual 4D seismic data, as an investigation of that type would be highly case dependent and would require specification of data acquisition details, model details, petro-elastic models and either forward or inverse modeling (Oliver et al., 2021). Instead, we focus on the impact of correlated noise in an idealized problem that captures the key elements of the 4D seismic history matching problem, i.e. the

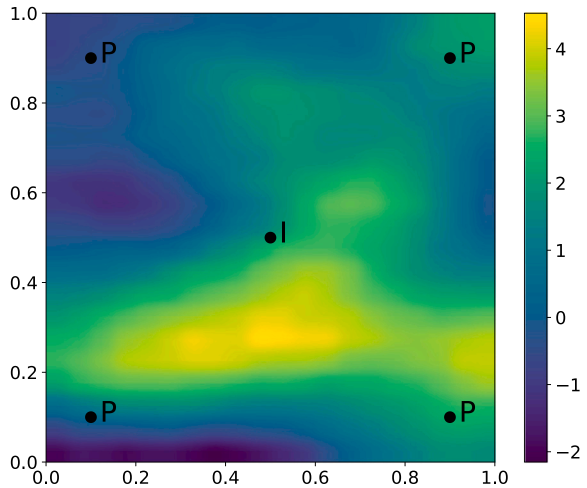


Fig. 1. True $\log_{10}$ permeability field and well locations for a numerical 2-phase flow test problem. Water is injected in a well (I) at a fixed rate at the center of the grid. Fluids are produced at equal rates from producing wells (P) at the four corners.

effect of spatially and temporally correlated errors in observations of saturation maps on information content with respect to estimation of permeability and remaining oil saturation.

It is clear that in some cases, correlated observation errors (when properly accounted for) reduce the information content of data. If, for example, one were to repeatedly measure the length of an object with the same (inaccurate) ruler, additional measurements would not increase the accuracy. Similarly, for observations with spatially correlated error, increasing the observation density beyond a threshold value will yield little or no improvement in analysis accuracy even when the data assimilation is performed optimally (Liu and Rabier, 2002). The assumption that correlations in observation error generally reduce information content are fairly common (Daley, 1992; Aanonsen et al., 2003), but the actual effect of correlation on information content is complex and could result in either increase or decrease in information content. Notably, correlations in observation error may increase the accuracy of estimates of gradients of the observed field (Seaman, 1977; Stewart et al., 2008). Similarly, Stewart et al. (2008) have shown that positive error correlations reduce the weight on averages of observations, but increase the weight on differences. Based on those findings, one might speculate that estimation of saturation fronts (boundaries between regions of differing saturation) from dense seismic data may benefit from correlations in observation error. Importantly for assimilation of seismic data, Rainwater et al. (2015) reports that spatially dense observations with correlated error contain more information about small scale phenomena than similarly dense observations with uncorrelated error.

2.3. Numerical flow example

We use an idealized numerical flow experiment that shares many key characteristics of the 4D seismic history matching problem to investigate the effect of correlated observation error. In particular, the numerical model has been designed to allow simulation of repeated spatially dense observations related to displacement of oil by water in a heterogeneous porous media. The problem was also chosen to be small enough to allow numerical experiments to be performed multiple times with different "truths" and different observation error characteristics.

In the numerical model, water is injected at a constant rate into the center of a 2D $[0,1] \times [0,1]$ reservoir grid (properties are uniform in the third direction). Fluids are produced from wells near the four corners of the model (locations marked with P in Fig. 1). The production rates at the four producing wells are identical despite spatial variability in permeability. Porosity is assumed to be uniform, but the logarithm of permeability is multi-Gaussian with mean 0 and isotropic covariance $\mathbf{C}_m(h) = 5^2 \exp(-3|h/0.8|^{1.9})$ that depends only on the distance h between points. The "true" log-permeability field, sampled from this distribution, is shown in Fig. 1 along with well locations.[1]

Water saturations are then "observed" on the 2D reservoir grid at 4 different times (0.25, 0.50, 0.75, and 1.0). The observations are noisy with both short and long range correlations in the observation error. Short range observation errors are uncorrelated with standard deviation of 0.1. Long range observation errors are correlated spatially with mean 0 and covariance $\mathbf{C}_D = 0.3^2 \exp(-3|h/r|^{1.9})$, where h is the distance between observation locations and the parameter r is varied to investigate the effect of correlation length. The total variance in the observation error is the same in all experiments. Fig. 2 shows the true saturations at three observation times, the noisy saturations, and estimates of the flooded region obtained by direct thresholding of observations. Note that the noisy observations (middle and bottom rows) have observed values that are outside the physically plausible range. This is a consequence of additive noise.

Insight into the effect of correlated observation error on information content can be obtained from visual examination of dense spatially distributed data with varying error correlation lengths. Fig. 3 shows observations with relatively short correlation length in observation errors (bottom row) and observations with long correlation (top row) at three different observation times. The variance in the noise is identical in both cases yet the region of increased water saturation is more easily identified when the observation error has long correlation range. One might guess from the figure that in this case, where the information is largely determined by the location of the "front", the observations for which the errors were more highly correlated will have the higher information content.

2.3.1. Assimilation of the data

To quantitatively compute the information content from the data requires estimation of the posterior distribution for the quantities to be estimated. In reservoir characterization, one may be interested in the permeability distribution conditioned on the observations of data. This estimate could be used to forecast the future behavior of the reservoir. One may also be interested specifically in estimation of water saturation at some particular time as this can be used to determine the location of remaining oil. We should expect that the information content in the data for estimation of permeability will be different from the information content for estimation of saturation, as has been observed by Tiedeman and Green (2013) in the context of groundwater modeling.

Our estimates of the posterior distributions for log-permeability and for saturation are obtained using an iterative ensemble smoother (Chen and Oliver, 2012) to generate an ensemble of samples of permeability from an approximation to the posterior. Although the iterative ensemble smoother is limited as a sampling method to problems for which the posterior is approximately Gaussian, that appears to be the case for the relatively simple reservoir model used in this investigation. Well injection rates and well production rates are fixed. The data are the observed (noisy) saturations fields at 4 times (400 observations at each survey). The only unknowns are the 400 log-permeability values on the grid. Posterior samples of saturation are obtained by running the reservoir simulator with permeability samples from the posterior as input. When performing data assimilation, we did not localize updates (a technique used to reduce the effect of spatial correlations and limited number of degrees of freedom) because the range of the correlation of permeability was approximately as large as the simulation domain, making the viability of localization questionable. To reduce the effects

[1] The simulator used for the numerical experiments can be obtained from https://github.com/patricknraanes/HistoryMatching.

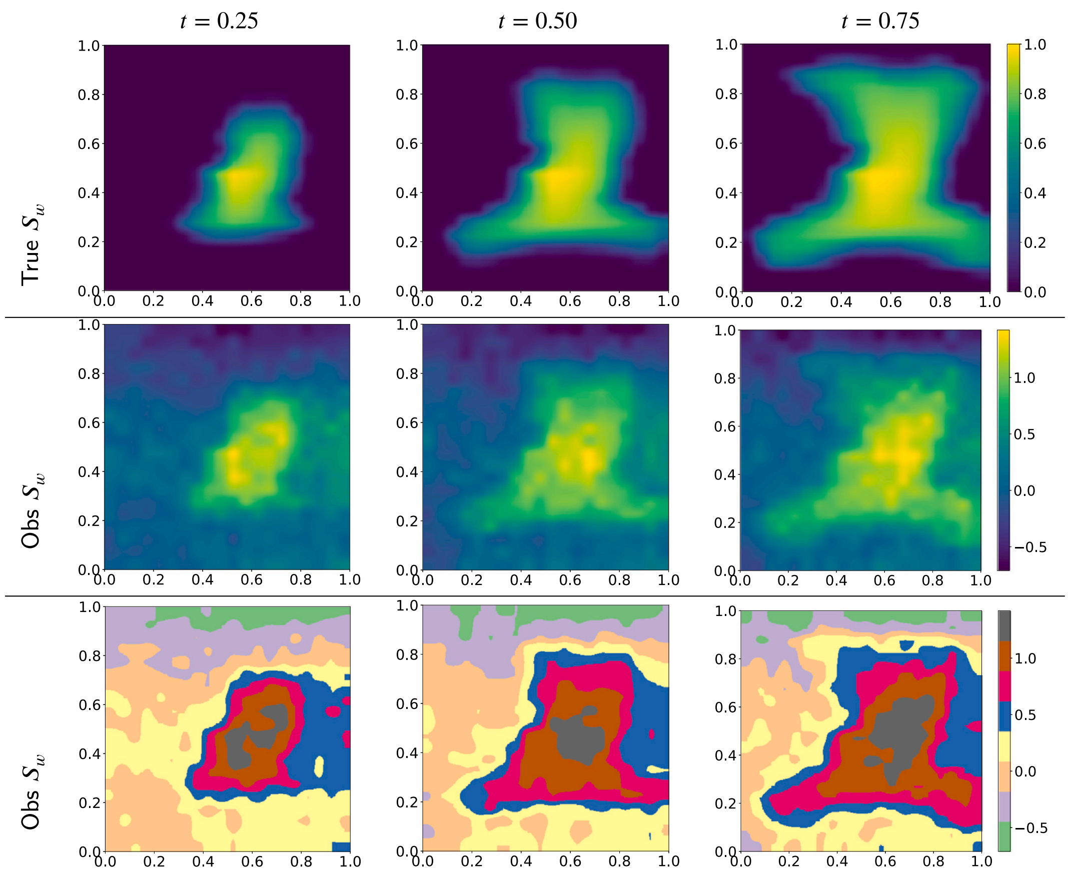


Fig. 2. Water saturation (S_w) is observed at 4 dimensionless times: 0.25, 0.5, 0.75 and 1.0 (only the first three observations are shown). The observations are contaminated with additive Gaussian noise consisting of white noise with standard deviation 0.1 and correlated Gaussian noise with standard deviation of 0.3 and correlation range of 1.2. The noise is added to the true saturation (top row) to generate the observed saturations (middle row). The bottom row shows result of thresholding of observations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

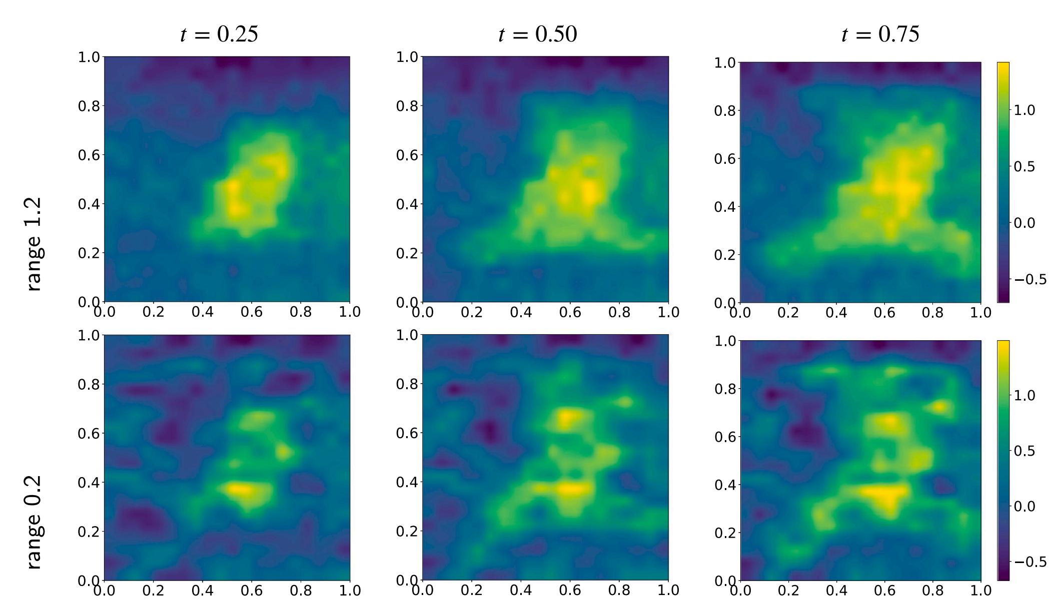


Fig. 3. Compare observations with relatively short correlation length (bottom row) with observations with long correlation (top row) at three different observation times. The variance in the noise is identical in both cases.

of sampling error and spurious correlations without localization, we used a fairly large (800-member) ensemble size.

As the observation errors are correlated and the number of data is fairly large (1600), the update equation was modified to allow a low-rank approximation of the observation error covariance matrix (Evensen, 2009; Alfonzo and Oliver, 2020). The data mismatch was monitored and iterations were stopped when the data mismatch stopped decreasing — typically after 10 iterations.

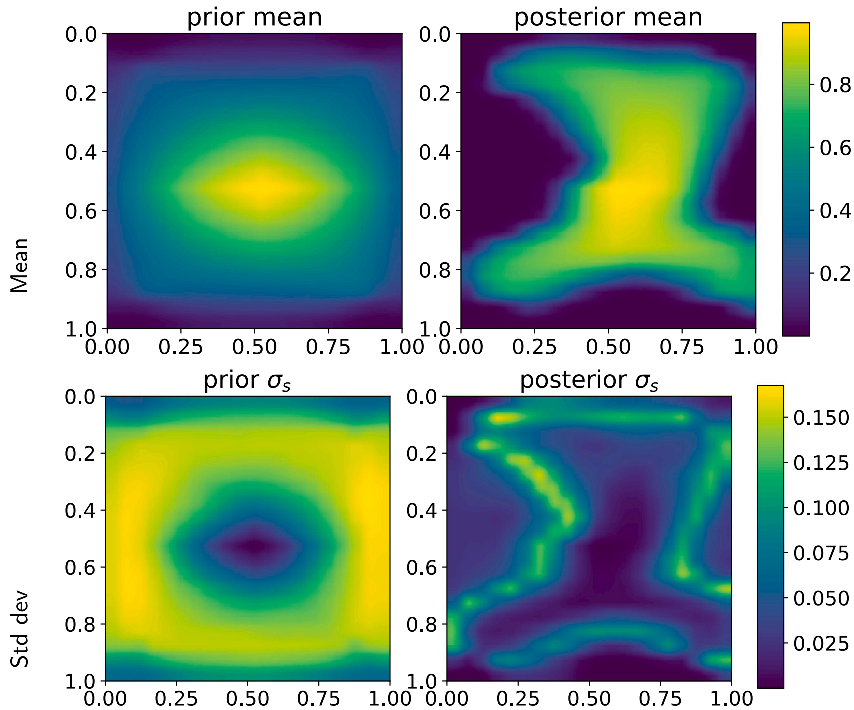


Fig. 4. Compare prior and posterior estimates of the change in water saturation, ΔS_w. The top row shows the prior and posterior means, while the bottom row shows the prior and posterior standard deviations for change in saturation.

After performing data assimilation, the information content can be computed either from the change in the posterior pdf (relative entropy) or in the change in the variance (SIC). Because we use an iterative ensemble smoother, we are largely limited to a Gaussian approximation of the relative entropy as in (1), in which case information content is a measure of the change in the mean and the change in the covariance from prior to posterior. Fig. 4 shows the change in the mean and in the standard deviation for saturation at time = 1 after assimilation of all noisy saturation observations with error correlation range 1.2. The posteriori mean for saturation is much different from the prior mean and the mean of the posteriori distribution is quite close to the true saturation at the prediction time. Note also that the uncertainty in saturation (bottom row of Fig. 4) is greatly reduced after assimilation of observations with correlated error. The uncertainty in the posteriori saturation is largely limited to the edges of the swept region.

The effect of correlation length in the observation error was quantified by performing numerical experiments in which observations were generated and assimilated as described above. These experiments were repeated multiple times with different correlation lengths in the observation error. For each correlation range in the observation error, we performed the data assimilation with seven different true models. Each data assimilation required approximately 11 iterations to converge so we ultimately performed approximately 400,000 simulations. Results are summarized in Fig. 5.

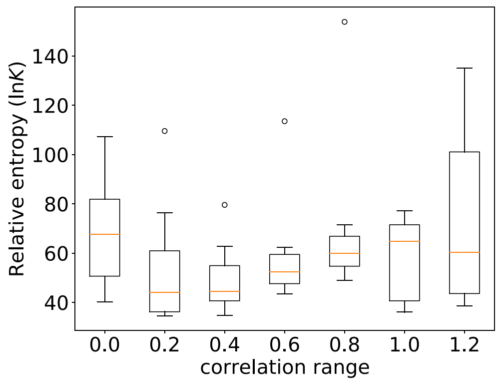
In Fig. 5a we see a minimum in the information content (relative entropy) for estimation of $\ln K$ (log-permeability) when the correlation range in observation error is about 0.2. This is similar to the characteristic length scale of the saturation field that is obtained from an analysis of the correlation of tangent lines to the saturation front at $t = 0.75$ (Zhang and Lu, 2004), confirming quantitatively the qualitative observation that when the noise and the signal have similar length scales, the information content is minimized (Fig. 3). The effect of observation error correlation length on estimation of saturation is similar (Fig. 5b)

We also examined the accuracy of the estimated properties in comparison to the values of the true properties. For this comparison, it is necessary to repeat the experiments several times to reduce the effect of sampling error on results. Once again, we see that for fixed variance, long correlation length in the observation error is beneficial for estimating the true properties (Fig. 5c). The root-mean-square error (RMSE) in the estimate of $\ln K$ is generally lower for larger correlation lengths in the observation error, although the spread in RMSE is large. The increase in accuracy of estimation is more substantial for saturation (Fig. 5d); we see a reduction in RMSE from approximately 0.11 when the correlation range of observation error and signal are approximately the same, to approximately 0.05 when the correlation length in the observation error is increased to 1.0. The reduction in RMSE for estimation of log-permeability is less because even noise-free data do not carry a large amount of information for identifying permeability.

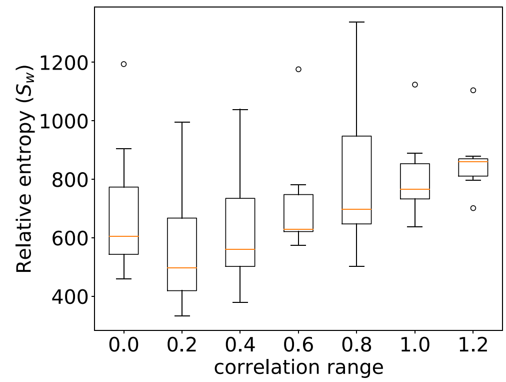
3. Neglecting correlation in observation error

Although data with correlated observation errors may, in some cases, have higher information content than similar data with uncorrelated observation errors, the extraction of the additional information might be difficult with typical data assimilation methods that either ignore correlations in observation error or deal with the correlations in an approximate way, such as by thinning observations or inflating the error. In general, two types of problems have been observed when the correlations are neglected: biased estimates of parameters or predictions, and overconfidence in results. If the off-diagonal elements of the observation error covariance are not simply neglected, but the variance is inflated or the data are thinned to compensate for the neglect of correlations, the results are much more varied.

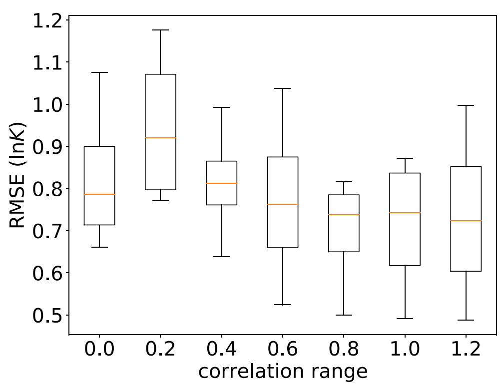
Stewart et al. (2006, 2008) report that ignoring correlations in observation errors results in a loss of information. Of course, if relative entropy is used to measure information content, then this finding should not be surprising as it simply says that the posterior pdf is sensitive to off-diagonal elements in $\mathbf{C}_D$; when relative entropy is used to measure information content any difference in the approximate pdf from the correct pdf translates to a loss of information. When Shannon entropy is used to measure loss in information content, however, it is only the effect on the confidence intervals that is important. Cooley and



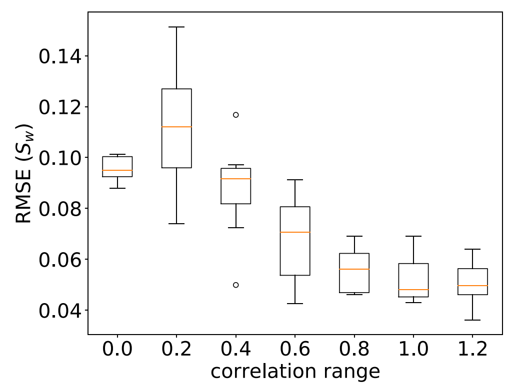
(a) Effect of correlation in observation error on information content for $\ln k$.



(b) Effect of correlation in observation error on information content for S_w.



(c) Effect of correlation in observation error on accuracy of mean for $\ln k$.



(d) Effect of correlation in observation error on accuracy of mean for S_w.

Fig. 5. Summary of the effect of correlated observation errors on estimates of $\ln K$ and S_w. Each box and whisker summarizes results from 7 instances of the truth.

Christensen (2006) found that data assimilation using simple weighted least squares in which $\mathbf{C}_D$ was replaced by a diagonal approximation, resulted in confidence intervals on predictions from calibrated groundwater models that were several times too small (overconfidence). In contrast, Tiedeman and Green (2013) reported that neglecting the correlations in observation errors caused both increases and decreases of parameter variances.

The largest effect of neglecting correlations in observation error is typically observed on estimates of small-scale features. When seismic data are used to monitor fluid movement or to identify fluid saturation fronts,[2] the front itself may be a small-scale feature of the flow. Rainwater et al. (2015) compared optimal data assimilation using the full observation error covariance matrix with an approximate method of data assimilation that neglected off-diagonal elements of the observation error covariance matrix for a linear data assimilation problem applied to a multiscale stochastic model. Inflating the variance was able to provide reasonable estimates for large-scale errors, but reduced accuracy for small-scale features. Both Stewart et al. (2008) and Rainwater et al. (2015) noted that the effect of failure to account for correlations in observation error can lead to poor estimates of small-scale features. Liu and Rabier (2002) concluded that a data assimilation scheme that neglects correlations in observation errors, limits the usefulness of high-density observations. This conclusion will apply to closely spaced seismic data, closely spaced formation test pressure data, and to frequent production rate data.

[2] A saturation front is a location at the interface between regions of much different saturations, such as might occur if water is used to displace oil.

3.1. Quantifying information loss

Similar to the definition of information content, we define the information loss resulting from the use of a diagonal approximation of $\mathbf{C}_D$ as the Kullback–Leibler directed divergence (or relative entropy) from the approximate posterior, p^{app}, to the correct posterior, p^{po}. If both pdfs are approximated as Gaussian the information loss is

$$D_{\text{KL}}\left(p^{\text{po}} \parallel p^{\text{app}}\right) = \overbrace{\frac{1}{2}\left((\bar{\mathbf{m}}_{\text{app}} - \bar{\mathbf{m}}_{\text{po}})^{\text{T}} \mathbf{C}_{m,\text{app}}^{-1} (\bar{\mathbf{m}}_{\text{app}} - \bar{\mathbf{m}}_{\text{po}})\right)}^{\text{difference in the mean}}$$

$$+ \overbrace{\frac{1}{2}\left(\text{Tr}\left(\mathbf{C}_{m,\text{app}}^{-1}\mathbf{C}_{m,\text{po}}\right) - k + \ln\left(\frac{\det \mathbf{C}_{m,\text{app}}}{\det \mathbf{C}_{m,\text{po}}}\right)\right)}^{\text{difference in dispersion}}. \quad (3)$$

Note that information loss can be attributed to bias in the mean estimate (the first term) and to miss-estimation of the dispersion (the second term). Computation of the determinant of large posteriori covariance matrices for model parameters will generally be problematic, but low-rank approximations are feasible (Xu, 2007).

3.2. Numerical flow example

Using the same simplified version of the 4D seismic data assimilation problem introduced in Section 2.3, we investigate the loss of information resulting from an approximate treatment of the error covariance matrix. In all experiments the observation errors are the sum of two types of Gaussian errors: the first is independent errors with standard deviation of 0.1, and the second is correlated errors with standard deviation 0.3 and correlation range 1.2. We numerically examine the effect of ignoring the correlation in the observation errors, but possibly inflating the variance of the diagonal elements in the approximation to $\mathbf{C}_D$. Results from the experiments for estimation of $\ln K$ are summarized in Fig. 6. The smallest values of σ_D in Fig. 6 (0.1) correspond to neglect of the correlated observation error. This would be a plausible value to use in data assimilation if an image denoising approach was used to estimate the noise level (Olsen, 1993). In that case, the low frequency contribution to the noise would not be separated from the signal. The red dashed line indicates the level of uncorrelated noise that would be used if one were able to correctly estimate the total noise (independent and correlated) but neglected off-diagonal terms in the observation error covariance matrix when performing data assimilation.

Madsen et al. (2017) argue that noise in certain seismic attributes is usually correlated and that correctly accounting for correlations in the noise is necessary for accurate inversion. They apply a hierarchical Bayesian inversion approach to estimation of the noise, which assumes knowledge of the characteristics of the signal and the stationary noise correlation matrix, but not the variance of the noise. By estimating the correct total variance, they were able to avoid overfitting of results and obtain improved estimates of the posterior variance. Madsen et al. (2018) discuss in more detail why correlated (total) observation errors might occur in seismic data and the consequence of ignoring the correlation in seismic inversion.

The relative entropy for estimation of $\ln K$ from saturation data with correlated observation errors is quite large if the off-diagonal elements of $\mathbf{C}_D$ are neglected. It can be reduced substantially, however, by inflating the magnitude of the diagonal terms (Fig. 6a). It is not possible to determine the reason for the reduction in relative entropy, or the reason that the relative entropy remains relatively large at large inflation, by examining only the values of relative entropy. However, examination of the two components (difference in means (Fig. 6b) and difference in dispersion (Fig. 6c)), shows that the reduction in relative entropy is largely a result of the reduction in the error in the mean estimate, measured with respect to the approximate covariance. Distance between approximate posteriori mean (neglecting correlations) and 'correct' posteriori mean is reduced when the assumed observation error is inflated. Note, however, that the difference is defined as follows

$$\Delta \ln K = (\bar{\mathbf{m}}_{\text{app}} - \bar{\mathbf{m}}_{\text{po}})^{\mathrm{T}} \mathbf{C}_{m,\text{app}}^{-1} (\bar{\mathbf{m}}_{\text{app}} - \bar{\mathbf{m}}_{\text{po}})$$

so the reduction may be a result of the increase in $\mathbf{C}_{m,\text{app}}$, not to a reduction in the Euclidean distance between $\bar{\mathbf{m}}_{\text{app}}$ and $\bar{\mathbf{m}}_{\text{po}}$. Consequently, even if the estimated mean was unchanged by inflation, the difference would appear to reduce as the variance in the observation error is inflated. Inflating $\mathbf{C}_D$ causes the difference between the approximate posterior covariance for $\ln K$ and the correct posterior covariance to increase (Fig. 6c).

In many cases, the estimation of saturation is of greater importance than the estimation of permeability. We again look at the effect of neglecting correlations in the observation errors and the effect of inflating the approximate $\mathbf{C}_D$, but this time focussing on the estimation of S_w. Recall that the posterior samples of S_w are obtained by running the reservoir simulator on the posterior samples of $\ln K$ so that estimation of $\ln K$ is still required. The effect of neglecting off-diagonal elements of $\mathbf{C}_D$ on information content for the estimation of S_w (Fig. 7a) is larger than the effect on estimation of $\ln K$. Although inflation appears to be beneficial in preventing information loss (and reducing the tendency for overconfidence), the effect of variance inflation on the accuracy of the estimation of saturation is negligible; when correlations in the observation error are neglected in the data assimilation, the RMSE in the mean estimate of the saturation field is more than 2.5 times larger than the error obtained using the full observation covariance and not improved at all by variance inflation (Fig. 7c).

3.3. Summary information loss

The numerical flow example with assimilation of saturation data illustrated two effects of neglect of correlated observation errors: biased estimates and overconfidence in estimation. Both effects result in loss of information. We saw that it was possible to eliminate the tendency for overconfidence by inflating the variance in the observation error, but the RMSE measure of accuracy of the mean estimate did not improve. Neglect of off-diagonal terms in $\mathbf{C}_D$ has been reported to increase the tendency for the uncertainty in parameters to be underestimated in problems with model error (Brynjarsdóttir and O'Hagan, 2014; Vink et al., 2015; Oliver and Alfonzo, 2018),

4. History matching with correlated observation error

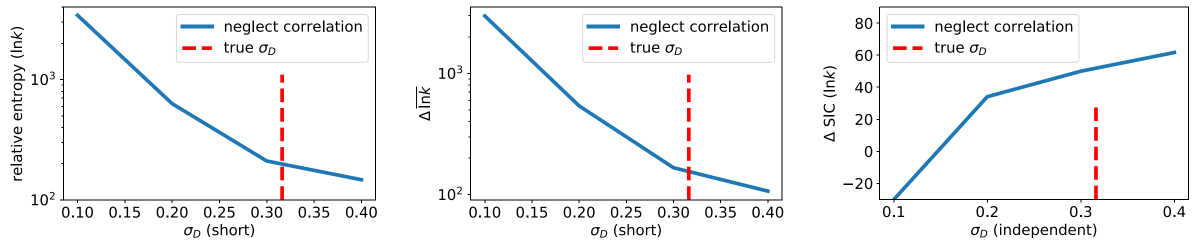
4.1. Estimation of $\mathbf{C}_D$

In the previous section, we saw that ignoring off-diagonal elements of the observation error covariance matrix $\mathbf{C}_D$ results in information loss, which manifests itself as biased and overconfident estimation. Inflation of the variance reduced the tendency for overconfidence, but did not improve the quality of the mean estimate. Two challenges then present themselves: how to use the correct observation error covariance if it is known and how to estimate it if it is not known. We examine how much information can be recovered by estimation of the full approximate $\mathbf{C}_D$ from the data and the model predictions, and how the full $\mathbf{C}_D$ can be used efficiently in large models.

First, we note that it is not possible to estimate noise or observation error in a single image without knowledge of the difference in the characteristics of the signal and the noise. If, for example, the noise is known to be uncorrelated and the variance stationary while the signal is spatially correlated then the magnitude of the noise can be effectively estimated (Olsen, 1993). Or, if the characteristics of the signal covariance or the noise covariance are available, then a sensible estimation of the other can be obtained from an image (Oppermann et al., 2011). The situation with 4D seismic data is difficult in this regard, as characterization of the covariance of the saturation signal is challenging and very little information is available on the noise except perhaps in regions of the formation where the signal is not expected to occur.

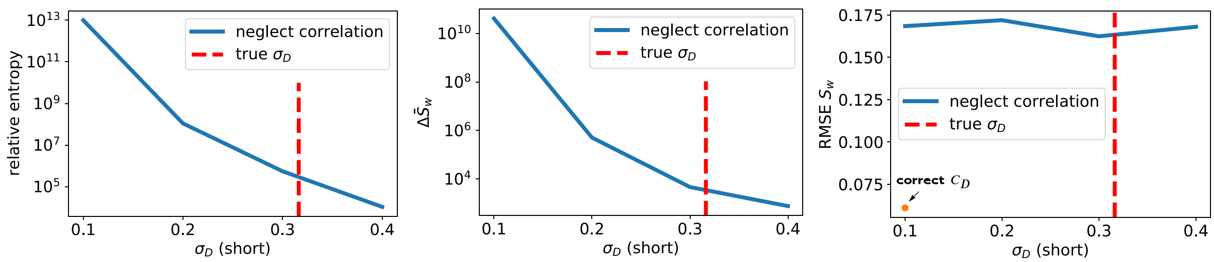
It has been common to assume that the noise in the 4D seismic data is either uncorrelated or has shorter correlation range than the saturation signal when estimating the noise covariance for history matching (Aanonsen et al., 2003; Emerick, 2016; Luo and Bhakta, 2017; Zhao et al., 2007), although this assumption does not appear to be supported by other methods of estimation that used repeated measurements in locations where no signal was expected (Abreu et al., 2005; Alfonzo and Oliver, 2020; Nivlet et al., 2017). Alternatively, the observation error covariance matrix can be estimated using data mismatch statistics from before and after model calibration (Desroziers et al., 2005). This statistic has been used both to select an optimal inflation for diagonal approximations of $\mathbf{C}_D$ (Li et al., 2009) and to estimate an approximation of the full $\mathbf{C}_D$ matrix (Miyoshi et al., 2013). In those cases, the observation operators were linear and the data assimilation scheme used was non-iterative. Oliver and Alfonzo (2018) derived a residual statistic based only on the posterior data mismatch statistics that can be used for iterative estimation of $\mathbf{C}_D$.

In the approach described by Oliver and Alfonzo (2018), one must specify a prior model for the model parameters and an initial guess for the covariance of observation errors. The first guess will typically be obtained using a filtering technique that assumes the observation errors are uncorrelated. The estimated $\mathbf{C}_D$ is then used to assimilate data, resulting in an ensemble of synthetic data predictions that can then be compared to the actual data. The process can be repeated and a new



(a) Relative entropy ($\ln k$). (b) Difference in mean $\ln k$. (c) ΔSIC ($\ln k$).

Fig. 6. Error in posterior pdf for $\ln K$ estimate measured by relative entropy, and components of relative entropy.



(a) Relative entropy (S_w). (b) Diff in mean S_w. (c) RMSE in S_w.

Fig. 7. Information loss for S_w resulting from diagonal approximation of C_D.

estimate of $\mathbf{C}_D$ produced. Using predicted data from the ith posterior predictions, the $(i+1)$th estimate of $\mathbf{C}_D$ is estimated as

$$\mathbf{C}_D^{(i+1)} = (\mathbf{d}_{\text{obs}}\mathbf{1}^{\text{T}} - \mathbf{D}_{\text{post}}^{(i)})(\mathbf{d}_{\text{obs}}\mathbf{1}^{\text{T}} - \mathbf{D}_{\text{post}}^{(i)})^{\text{T}}/N_e \tag{4}$$

where $\mathbf{D}$ is the $N_d \times N_e$ matrix whose columns are posteriori samples of the data and $\mathbf{1}$ is the column vector of length N_d, each of whose elements are equal to 1. When the number of samples in the ensemble is relatively small compared to the number of observations, the estimate of $\mathbf{C}_D^{(i+1)}$ from Eq. (4) will benefit from some type of regularization. Unless there is reason to suspect that the variance is not stationary, it is common to regularize the estimate by assuming that the observation error covariance depends only on the distance between the observations (e.g., Miyoshi et al., 2013; Alfonzo and Oliver, 2020).

Most data assimilation or history matching methods attempt to find a set of parameters $\mathbf{m}$ that minimize an objective function containing a term of the form $(\mathbf{g}(\mathbf{m}) - \mathbf{d}^{\text{obs}})^T \mathbf{C}_D^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}^{\text{obs}})$, which adjusts values of parameters $\mathbf{m}$ in the model to be consistent with observations, i.e. $\mathbf{g}(\mathbf{m}) \approx \mathbf{d}^{\text{obs}}$. When the number of data is large, it is common to assume that $\mathbf{C}_D$ is diagonal (i.e. that the observation errors are independent), in which case computing the value of the objective function is trivial. In some methods, such as the perturbed observation form of ensemble-Kalman based data assimilation (Burgers et al., 1998; Chen and Oliver, 2012), it is also necessary to generate perturbed observations from the multivariate Gaussian distribution $N[0, \mathbf{C}_D]$.

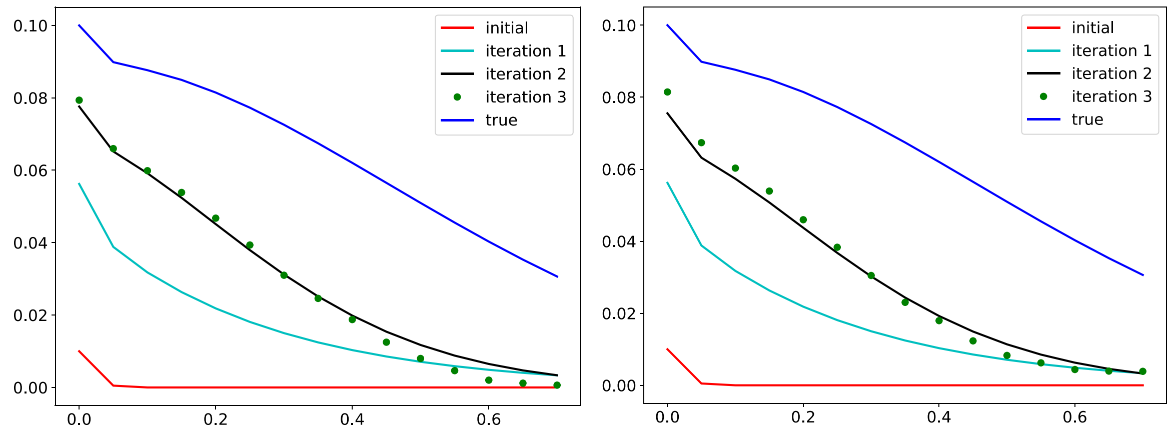
We illustrate the approach with application to the numerical flow experiment introduced in Section 2.3. As in Section 3.2 we generate saturation observations at four survey times with combined short and long range correlations in the observation error. The white noise has standard deviation 0.1; the correlated noise has standard deviation 0.3 and range 1.2 (dark blue curve in Fig. 8). Setting the initial guess for $\mathbf{C}_D$ to be consistent with the white noise component of the observation error (red curves in Fig. 8), the algorithm appeared to converge after two iterations (black curves in Fig. 8). A third iteration did not change the result substantially (green dotted curves). The experiment was run a second time with a different realization of the true permeability and a different realization of the observation error (Fig. 8b). Results in this case were nearly identical to the first case (Fig. 8a).

The means of the ensemble of posteriori saturation realizations for three of the iterative estimates of $\mathbf{C}_D$ are shown in Fig. 9. The saturation plot on the left labeled 'iter 0' corresponds to the use of the initial guess for $\mathbf{C}_D$, which in our case we took to be diagonal. The estimated saturation distribution obtained using a diagonal $\mathbf{C}_D$ is similar in appearance to the saturation estimate that would be obtained from selection of a saturation front directly from the data (see the blue level contour at bottom right in Fig. 2) – when correlations in observation error are neglected it is not possible to distinguish correlated noise from signal. After a single iteration for updating of $\mathbf{C}_D$, the estimated saturation mean (center of Fig. 9) is similar to the true saturation (top right corner of Fig. 2).

We quantify the effect of iterative estimation of $\mathbf{C}_D$ through computation of relative entropy of the data for log-permeability and for saturation. Fig. 10a shows the information loss when performing data assimilation to estimate S_w and $\ln K$ using an incorrect value of $\mathbf{C}_D$ relative to a data assimilation that uses the correct $\mathbf{C}_D$. The information loss decreases dramatically following a single update (iteration 1) of the estimate of $\mathbf{C}_D$. Further iterations had almost no impact on information content. The accuracy of the estimate os S_w as quantified by RMSE also showed a dramatic improvement after a single update of the estimate of $\mathbf{C}_D$ (Fig. 10b). In this case, continued iteration did result in a small additional decrease in RMSE to a value of 0.08. For comparison, the value of RMSE obtained when using the correct $\mathbf{C}_D$ is 0.06. The fact that a large improvement in information content is obtained by a single update to the estimate of $\mathbf{C}_D$, and that the gains in subsequent updates are relatively minor is consistent with the findings of Stewart et al. (2013) and Bédard and Buehner (2020), both of whom report that it is generally better to include an approximate correlation structure in $\mathbf{C}_D$ than to incorrectly assume that the observation errors are independent.

5. Summary and recommendations

The results from history matching of 4D seismic data have generally been disappointing — despite the clear benefits from 4D seismic monitoring, 4D seismic data often appear to add little to predictability of models (Oliver et al., 2021). The motivation for this investigation was partially to determine possible reasons for the mismatch in expectation for information and experience and to investigate if the apparent lack of information could be a consequence of choices in history matching methodology.



(a) Seed = 1. (b) Seed = 4.

Fig. 8. Iterative estimation of $\mathbf{C}_D$ using Eq. (4). Each curve shows the estimated observation error covariance (y-axis) as a function of spatial separation between data point locations (x-axis). The numerical experiment was performed several times with different random seeds to verify repeatability of results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

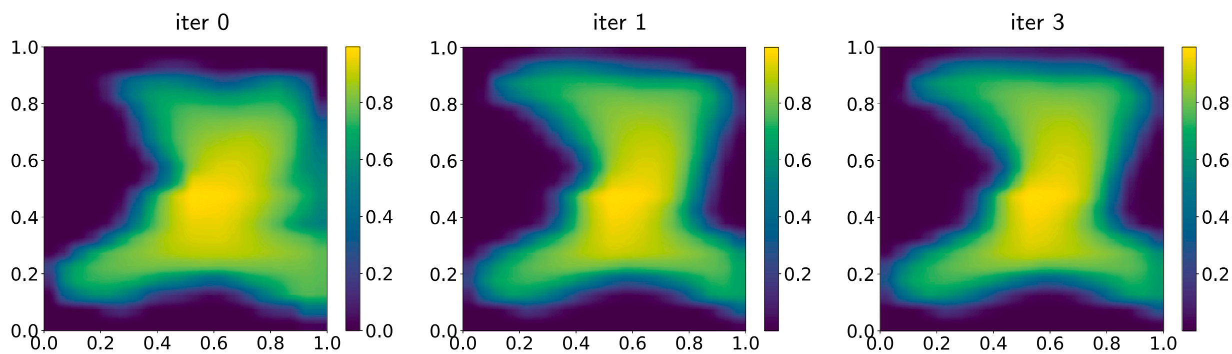
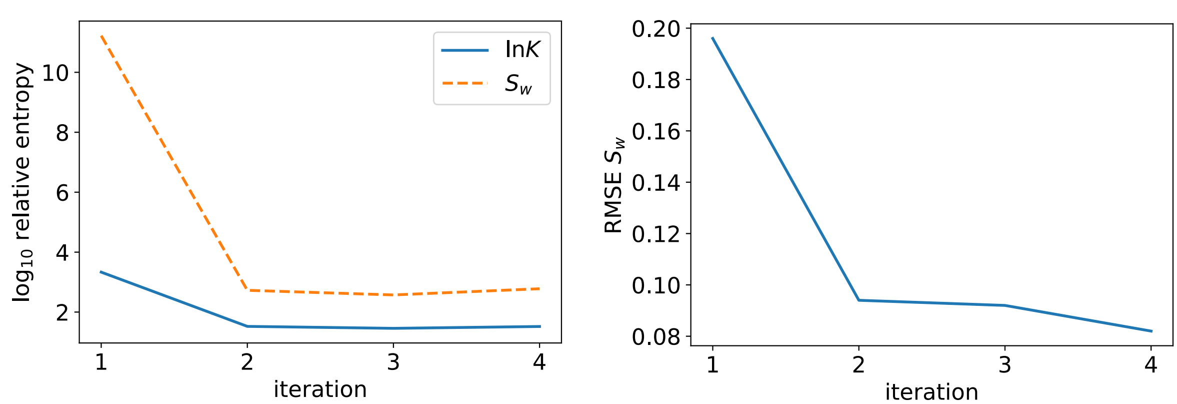


Fig. 9. Improved estimates of water saturation at time 1 from iterative estimation of $\mathbf{C}_D$. The first plot (iter 0) uses the initial guess for $\mathbf{C}_D$.



(a) Evolution of relative entropy for saturation and log-permeability. (b) Evolution of RMSE for saturation field.

Fig. 10. Increase in information content (relative entropy) and improvement in accuracy of estimated saturation resulting from iterative estimation of C_D.

Information content in data can be measured a number of ways. We focussed on the use of relative entropy (also known as the Kullback–Leibler divergence) to measure information content in the data. When defined this way, any change in the parameter pdf from the prior to the posterior due to the assimilation of data results in positive information content. Since the relative entropy is measuring the change in the probability distribution, the same data may have different information content for estimation of different quantities. We examined information content for estimation of log-permeability and for estimation of saturation at a specific forecast time. Although correlated observation errors are often assumed to reduce information content, we found that information content in densely spaced saturation data is not necessarily reduced when the observation error is spatially correlated. Specifically, we found in our numerical experiments that if observation error is correctly characterized and correctly used in assimilation, information content is lowest when the signal and the observation error have similar length scales so that the two are difficult to separate. Highest information content occurs when the scales of error and permeability field are very different. Large range correlation error in observations is not harmful to saturation estimation if the correlation errors are correctly characterized and handled correctly in the data assimilation

On the other hand, information is reduced if actual correlations are neglected in data assimilation. The result is incorrect estimates of the variable mean and poor quantification of uncertainty. One common approximate approach to dealing with correlated observation errors is to inflate the variance, but ignore correlations. For dense spatial data like 4D seismic data, inflation of the observation error is not equivalent to data assimilation that accounts correctly for correlations in observation error. Inflation of variance in the observation error covariance matrix reduces the tendency to be overconfident, but may not improve mean estimates when errors are correlated.

It is not possible to estimate correlated observation errors directly from an image without characterization of statistics of the signal, which in the case of time-lapse saturation observations must be obtained from the model. We showed that the total observation error (measurement error plus modeling error) for saturation could be estimated by comparing actual observations to the posterior ensemble of saturation predictions. Iterative estimation of $\mathbf{C}_D$ greatly improves the estimate of the saturation and the quantification of uncertainty with corresponding increase in information content in the data. Extracting large amounts of information from the observed saturation data was possible, even when the noise level was fairly large. It did require, however, that the correlation length for the noise was much longer than a characteristic scale of the saturation distribution and that it was properly estimated.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Dean Oliver acknowledges financial support from the NORCE research cooperative research project “Assimilating 4D Seismic Data: Big Data Into Big Models” which is funded by industry partners, Equinor Energy AS, Lundin Energy Norway AS, Repsol Norge AS, Shell Global Solutions International B.V., TotalEnergies E&P Norge AS and Wintershall Dea Norge AS, as well as the Research Council of Norway through the Petromaks2 program (NFR project number: 295002).

References

Aanonsen, S.I., Aavatsmark, I., Barkve, T., Cominelli, A., Gonard, R., Gosselin, O., Kolasinski, M., Reme, H., 2003. Effect of scale dependent data correlations in an integrated history matching loop combining production data and 4D seismic data (SPE 79665). In: Proceedings of the SPE Reservoir Simulation Symposium, The Woodlands.

Abreu, C.E., Lucet, N., Nivlet, P., Royer, J.-J., 2005. Improving 4D seismic data interpretation using geostatistical filtering. In: 9th International Congress of the Brazilian Geophysical Society.

Alfonzo, M., Oliver, D.S., 2020. Seismic data assimilation with an imperfect model. Comput. Geosci. 24 (2), 889–905.

Bédard, J., Buehner, M., 2020. A practical assimilation approach to extract smaller-scale information from observations with spatially correlated errors: An idealized study. Q. J. R. Meteorol. Soc. 146 (726), 468–482.

Brynjarsdóttir, J., O’Hagan, A., 2014. Learning about physical parameters: the importance of model discrepancy. Inverse Problems 30 (11), 114007.

Burgers, G., van Leeuwen, P.J., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. Mon. Weather Rev. 126 (6), 1719–1724.

Chen, N., 2020. An information criterion for choosing observation locations in data assimilation and prediction. SIAM/ASA J. Uncertainty Quantif. 8 (4), 1548–1573.

Chen, Y., Oliver, D.S., 2012. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. Math. Geosci. 44 (1), 1–26.

Cooley, R.L., Christensen, S., 2006. Bias and uncertainty in regression-calibrated models of groundwater flow in heterogeneous media. Adv. Water Resour. 29 (5), 639–656.

Daley, R., 1992. The effect of serially correlated observation and model error on atmospheric data assimilation. Mon. Weather Rev. 120 (1), 164–177.

Desroziers, G., Berre, L., Chapnik, B., Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc. 131 (613, C), 3385–3396.

Emerick, A.A., 2016. Analysis of the performance of ensemble-based assimilation of production and seismic data. J. Pet. Sci. Eng. 139, 219–239.

Evensen, G., 2009. Data Assimilation: The Ensemble Kalman Filter, second ed. Springer Verlag.

Folgerø, K., Soldal, E.L., Kocbach, J., Frøysa, K.-E., Kleppe, K., Åbro, E., (2013). Uncertainty analysis of multiphase flow meters used for allocation measurements: Field experiences and future challenges. In: Proceedings of the 31st International North Sea Flow Measurement Workshop, 22—25 October, Tønsberg, Norway.

Fowler, A., Van Leeuwen, P.J., 2012. Measures of observation impact in non-Gaussian data assimilation. Tellus A: Dyn. Meteorol. Oceanogr. 64 (1), 17192.

Li, H., Kalnay, E., Miyoshi, T., 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. Q. J. R. Meteorol. Soc. 135 (639), 523–533.

Liu, Z.-Q., Rabier, F., 2002. The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. Q. J. R. Meteorol. Soc. 128 (582), 1367–1386.

Luo, X., Bhakta, T., 2017. Estimating observation error covariance matrix of seismic data from a perspective of image denoising. Comput. Geosci. 21 (2), 205–222.

Madsen, R.B., Nørmark, E., Hansen, T.M., 2018. Accounting for processing errors in AVO/AVA data. In: 80th EAGE Conference & Exhibition. European Association of Geoscientists & Engineers, Copenhagen, Denmark.

Madsen, R.B., Zunino, A., Hansen, T.M., 2017. On inferring the noise in probabilistic seismic AVO inversion using hierarchical Bayes. In: SEG International Exposition and Annual Meeting, 24-29 September, Houston, Texas.

Majda, A., Kleeman, R., Cai, D., 2002. A mathematical framework for quantifying predictability through relative entropy. Methods Appl. Anal. 9 (3), 425–444.

Miyoshi, T., Kalnay, E., Li, H., 2013. Estimating and including observation-error correlations in data assimilation. Inverse Prob. Sci. Eng. 21 (3), 387–398.

Neuman, S.P., Xue, L., Ye, M., Lu, D., 2012. Bayesian analysis of data-worth considering model and parameter uncertainties. Adv. Water Resour. 36, 75–85, Special Issue on Uncertainty Quantification and Risk Assessment.

Nivlet, P., Smith, R., Jervis, M.A., Bakulin, A., 2017. Toward an integrated and realistic interpretation of continuous 4D seismic data for a CO2 EOR and sequestration project. In: SPE Middle East Oil & Gas Show and Conference. Society of Petroleum Engineers, Manama, Kingdom of Bahrain.

Oliver, D.S., Alfonzo, M., 2018. Calibration of imperfect models to biased observations. Comput. Geosci. 22 (1), 145–161.

Oliver, D.S., Fossum, K., Bhakta, T., Sandø, I., Nævdal, G., Lorentzen, R.J., 2021. 4D seismic history matching. J. Pet. Sci. Eng. 207, 109119.

Olsen, S.I., 1993. Estimation of noise in images: An evaluation. CVGIP: Graph. Models Image Process. 55 (4), 319–323.

Oppermann, N., Robbers, G., Enßlin, T.A., 2011. Reconstructing signals from noisy data with unknown signal and noise covariance. Phys. Rev. E 84, 041118.

Petty, G.W., 2018. On some shortcomings of Shannon entropy as a measure of information content in indirect measurements of continuous variables. J. Atmos. Ocean. Technol. 35 (5), 1011–1021.

Rainwater, S., Bishop, C.H., Campbell, W.F., 2015. The benefits of correlated observation errors for small scales. Q. J. R. Meteorol. Soc. 141 (693), 3439–3445.

Roach, L.A.N., White, D.J., Roberts, B., 2015. Assessment of 4D seismic repeatability and CO2 detection limits using a sparse permanent land array at the aquistore CO2 storage site. Geophysics 80 (2), WA1–WA13.

Sadri, M., Shariatipour, S.M., 2020. Mitigating allocation and hydrocarbon accounting uncertainty using more frequent flow test data. J. Energy Resour. Technol. 142 (4), 043003.

Seaman, R.S., 1977. Absolute and differential accuracy of analyses achievable with specified observational network characteristics. Mon. Weather Rev. 105 (10), 1211–1222.

Stewart, L.M., Dance, S.L., Nichols, N.K., 2006. Information Content of Spatially Correlated Observation Errors. Technical Report Numerical Analysis Report 4/06, Department of Mathematics, The University of Reading.

Stewart, L.M., Dance, S.L., Nichols, N.K., 2008. Correlated observation errors in data assimilation. Internat. J. Numer. Methods Fluids 56 (8), 1521–1527.

Stewart, L.M., Dance, S.L., Nichols, N.K., 2013. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. Tellus A (65).

Thore, P., 2015. Uncertainty in seismic inversion: What really matters? Leading Edge 34 (9), 1000–1004.

Tiedeman, C.R., Green, C.T., 2013. Effect of correlated observation error on parameters, predictions, and uncertainty. Water Resour. Res. 49 (10), 6339–6355.

Vink, J.C., Gao, G., Chen, C., 2015. Bayesian style history matching: Another way to under-estimate forecast uncertainty? In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Xu, Q., 2007. Measuring information content from observations for data assimilation: relative entropy versus Shannon entropy difference. Tellus A: Dyn. Meteorol. Oceanogr. 59 (2), 198–209.

Zhang, D., Lu, G., 2004. Review of shape representation and description techniques. Pattern Recognit. 37 (1), 1–19.

Zhao, Y., Li, G., Reynolds, A.C., 2007. Characterization of the measurement error in time-lapse seismic data and production data with an EM algorithm. Oil Gas Sci. Technol. 62 (2, Sp. Iss. SI), 181–193.