

# *Assimilation of multiple linearly dependent data vectors*

**Trond Mannseth**

**Computational Geosciences**

Modeling, Simulation and Data Analysis

ISSN 1420-0597

Volume 24

Number 1

Comput Geosci (2020) 24:349-354

DOI 10.1007/s10596-019-09924-6

**Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.**



# Assimilation of multiple linearly dependent data vectors

Trond Mannseth<sup>1</sup>

Received: 28 November 2018 / Accepted: 21 November 2019 / Published online: 18 December 2019  
© The Author(s) 2019

## Abstract

Assimilation of a sequence of linearly dependent data vectors,  $\{d_l\}_{l=1}^L$  such that  $\{d_l = B_l d_L\}_{l=1}^{L-1}$ , is considered for a parameter estimation problem. Such a data sequence can occur, for example, in the context of multilevel data assimilation. Since some information is used several times when linearly dependent data vectors are assimilated, the associated data-error covariances must be modified. I develop a condition that the modified covariances must satisfy in order to sample correctly from the posterior probability density function of the uncertain parameter in the linear-Gaussian case. It is shown that this condition is a generalization of the well-known condition that must be satisfied when assimilating the same data vector multiple times. I also briefly discuss some qualitative and computational issues related to practical use of the developed condition.

**Keywords** Ensemble-based methods · Linearly dependent data vectors · Generalized MDA condition

## 1 Introduction

Consider a parameter estimation problem defined by  $d = g(m) + \eta$ , where  $d$  denotes the data vector,  $g$  the forward-model operator,  $m$  the unknown parameter vector and  $\eta$  a realization of the measurement error. The Bayesian framework for parameter estimation includes uncertainty quantification as an integral part. Markov-chain Monte Carlo (MCMC) methods can sample correctly from the posterior probability density function (PDF) for the parameter vector also for nonlinear problems, but are prohibitively computationally expensive for many realistic problems, such as reservoir history matching. Ensemble-based methods, such as the ensemble Kalman filter (EnKF) [4] and the ensemble smoother (ES) [12], sample correctly from the posterior PDF when the forward-model operator is linear, the prior PDF is Gaussian and the ensemble size approaches infinity (linear-Gaussian case). In the general case, these methods can be seen as computationally tractable, approximate alternatives to MCMC methods. A practical disadvantage of the EnKF with respect to the ES for nonlinear

parameter estimation problems is that it requires a rerun of the forward simulator from initial time to the current assimilation time to avoid statistical inconsistencies [18]. If such reruns are performed, however, the EnKF can be expected to outperform the ES, at least for weakly nonlinear problems [5, 6].

The nonlinearity of the reservoir history matching problem has triggered the development of popular data assimilation (DA) methods such as iterative ES [1] and ES with multiple data assimilations (ES-MDA) [3]. As shown in [15], ES-MDA is related to annealed importance sampling [14]. With ES-MDA,  $d$  is assimilated  $L$  times ( $L$  is a predetermined number), and a compensation for utilizing the same information  $L$  times must then be devised. In assimilation cycle no.  $l$ , the original data-error covariance matrix,  $C$ , is therefore replaced by an inflated data-error covariance matrix,  $\alpha_l C$ , where  $\alpha_l > 1$ .

The motivation for introducing methods such as the ES-MDA is to better handle nonlinearities by taking  $L$  smaller steps instead of a single large step. Still, it is highly desirable for any ensemble-based method to sample correctly from the posterior PDF in the linear-Gaussian case. If the inflation coefficients  $\{\alpha_l\}_{l=1}^L$  satisfy a certain equation—termed the MDA condition in this paper—the ES-MDA will sample correctly in the linear-Gaussian case [3].

I consider assimilation of multiple linearly dependent data vectors, which can be seen as a generalization of MDA. To illustrate this, let  $\{d_l\}_{l=1}^L$  denote a sequence of linearly dependent data vectors to be assimilated, that is,  $\{d_l =$

✉ Trond Mannseth  
trma@norce-research.no

<sup>1</sup> NORCE Norwegian Research Centre,  
P.O. Box 22 Nygårdstangen, 5838 Bergen,  
Norway

$B_l d_L\}_{l=1}^{L-1}$ , where  $\{B_l\}_{l=1}^{L-1}$  denotes a sequence of matrices. In the special case when  $\{B_l = I_L\}_{l=1}^{L-1}$ , where  $I_L$  denotes the identity matrix, all  $L$  data vectors equal  $d_L$ . Hence, MDA is a special case of assimilation of multiple linearly dependent data vectors. In addition to being of interest as a generalization of MDA, assimilation of multiple linearly dependent data vectors occur, for example, in the context of multilevel data assimilation (see the [Appendix](#)).

In this paper, I develop a condition that any data assimilation method has to fulfill in order to sample correctly in the linear-Gaussian case when multiple linearly dependent data vectors are assimilated. This condition reduces to the MDA condition in the special case where all the involved data vectors are identical. While MDA utilizes exactly the same information multiple times, the situation is more subtle when considering the generalization to assimilation of multiple linearly dependent data vectors. Clearly, part of, but not necessarily all of, the available information is used multiple times. For this reason, and to emphasize the relation to MDA, I denote assimilation of multiple linearly dependent data vectors *partially multiple data assimilation* (PMDA). The condition that has to be fulfilled in order to sample correctly in the linear-Gaussian case when multiple linearly dependent data vectors are assimilated will be denoted the *PMDA condition*.

The main emphasis in the paper is on the derivation of the PMDA condition, leading up to the main result, (19) and (20), and on the relation to the MDA condition, but I will also briefly discuss practical use of the PMDA condition. More work in this direction is, however, required, but that should preferably be performed when applying the PMDA condition to a nonlinear model in a real application with linearly dependent data vectors, which is outside the scope of the current paper.

I first (Section 2.1) consider the standard linear-Gaussian case with a single data vector, and write up expressions for the mean and covariance of the (Gaussian) posterior PDF. Next (Section 2.2), I introduce multiple data vectors, and write up expressions for the mean and covariance of the posterior PDF for that variant of the linear-Gaussian case. I also derive some expressions that will serve as a common starting point for the sketch of the derivation of the well-known MDA condition, where the data vectors are identical (Section 3), and for the derivation of the novel PMDA condition, where the data vectors are linearly dependent (Section 4). In addition to the derivation of the PMDA condition, Section 4 also discusses its practical use.

Finally, my motivation for deriving the PMDA condition for linearly dependent data vectors stems from my interest in multilevel data assimilation. I therefore provide a brief motivation for multilevel data assimilation in the [Appendix](#), where I also point out the connection to linearly dependent data vectors.

## 2 Linear-Gaussian case

In [2, Section 4.1 and Section 4.2], the authors gave two alternative derivations of the MDA condition in the setting of a joint parameter-state estimation problem, both assuming equal inflation coefficients. The derivation in [2, Section 4.2] is sample based, and explicitly shows that one needs to resample the measurement perturbations in each update cycle to obtain the correct posterior. This derivation was later generalized to allow for unequal inflation coefficients [3] in the setting of a parameter estimation problem. The derivation in [2, Section 4.1] utilizes only the Kalman filter equations and linear algebra.

The purpose of Section 2 is to facilitate a unified derivation of the MDA condition (Section 3) and the PMDA condition (Section 4). The main steps in the analysis of the linear-Gaussian case in Sections 2, 3, and 4 follow those in [2, Section 4.1], but in the setting of a parameter estimation problem, and allowing for unequal modification of the data-error covariances in the different update cycles.

### 2.1 Single data vector and model operator

Consider a Bayesian parameter estimation problem with parameter,  $m \in \mathbb{R}^M$ , data vector  $d_L \in \mathbb{R}^{N_L}$  and linear forward operator  $G_L \in \mathbb{R}^{N_L \times M}$ . Assume that the prior PDF is  $\mathcal{N}(m_{\text{pr}}, C_{\text{pr}})$  and the data-error PDF is  $\mathcal{N}(0, C_L)$ , where  $\mathcal{N}$  denotes the multi-Gaussian distribution. It is well known (see, e.g., [16, Chap. 3.2.2]) that the posterior PDF is then  $\mathcal{N}(m_{\text{po}}, C_{\text{po}})$ , where

$$C_{\text{po}} = \left( C_{\text{pr}}^{-1} + G_L^T C_L^{-1} G_L \right)^{-1}, \quad (1)$$

$$m_{\text{po}} = C_{\text{po}} \left( C_{\text{pr}}^{-1} m_{\text{pr}} + G_L^T C_L^{-1} d_L \right). \quad (2)$$

### 2.2 Multiple data vectors and model operators

Let  $\{d_l \in \mathbb{R}^{N_l}\}_{l=1}^L$  denote a sequence of data vectors with data-error covariance matrices  $\{C_l\}_{l=1}^L$ , and  $\{G_l \in \mathbb{R}^{N_l \times M}\}_{l=1}^L$  a sequence of linear forward operators, such that  $d_l$  and  $G_l m$  denote data and simulated output for entry number  $l$ , respectively. Furthermore, let  $N_{\Sigma} = \sum_{l=1}^L N_l$ , and define the quantities  $\delta \in \mathbb{R}^{N_{\Sigma}}$ ,  $\Gamma \in \mathbb{R}^{N_{\Sigma} \times M}$ , and  $\Xi \in \mathbb{R}^{N_{\Sigma} \times N_{\Sigma}}$ , by

$$\delta = \begin{pmatrix} d_1 \\ \vdots \\ d_L \end{pmatrix}, \quad \Gamma = \begin{pmatrix} G_1 \\ \vdots \\ G_L \end{pmatrix}, \quad \Xi = \begin{pmatrix} C_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_L \end{pmatrix}. \quad (3)$$

Consider a Bayesian parameter estimation problem with forward operator  $\Gamma$ , data vector  $\delta$ , prior PDF  $\mathcal{N}(m_{\text{pr}}, C_{\text{pr}})$ ,

and data-error PDF  $\mathcal{N}(0, \Xi)$ . The posterior PDF for  $m$  is then  $\mathcal{N}(m_{\text{mu}}, C_{\text{mu}})$ , where

$$C_{\text{mu}} = \left( C_{\text{pr}}^{-1} + F \right)^{-1}, \tag{4}$$

$$m_{\text{mu}} = C_{\text{mu}} \left( C_{\text{pr}}^{-1} m_{\text{pr}} + f \right), \tag{5}$$

and where  $F \stackrel{\text{def}}{=} \Gamma^T \Xi^{-1} \Gamma$  and  $f \stackrel{\text{def}}{=} \Gamma^T \Xi^{-1} \delta$ . Comparing (4) and (5) with (1) and (2), respectively, it is seen that

$$C_{\text{mu}} = C_{\text{po}} \Leftrightarrow F = G_L^T C_L^{-1} G_L, \tag{6}$$

$$m_{\text{mu}} = m_{\text{po}} \Leftrightarrow f = G_L^T C_L^{-1} d_L, \tag{7}$$

where (7) assumes that (6) is fulfilled.

Inserting from (3) into the definitions for  $F$  and  $f$  gives

$$F = \sum_{l=1}^L G_l^T C_l^{-1} G_l, \tag{8}$$

$$f = \sum_{l=1}^L G_l^T C_l^{-1} d_l. \tag{9}$$

Jointly, (6), (7), (8), and (9) serve as a common starting point for the sketch of the derivation of the well-known MDA condition, where the data vectors are identical (Section 3), and for the derivation of the novel PMDA condition, where the data vectors are linearly dependent (Section 4).

### 3 Multiple data assimilation—MDA condition

In [2] and [3], the authors considered  $L$  assimilations of the same data vector with an inflated data-error covariance matrix in each assimilation cycle. In the linear-Gaussian case, this situation can be described by (3) with the following additional definitions,

$$\left. \begin{aligned} d_l &= d_L \\ G_l &= G_L \\ C_l &= C_L \end{aligned} \right\}; l \in [1, L], \tag{10}$$

and then inflating the data-error covariances by  $C_l \rightarrow \alpha_l C_l$ , where  $\alpha_l > 1$ ;  $l \in [1, L]$ . The inflated assembled data-error covariance for MDA can then be written

$$\Xi_{\text{MDA}}^{[1,L]} = \begin{pmatrix} \alpha_1 C_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \alpha_L C_L \end{pmatrix}. \tag{11}$$

For reasons that will become evident in Section 4.1.3, I have not substituted  $C_L$  for  $C_l$ , and I have introduced a superscript on  $\Xi_{\text{MDA}}$  indicating the involved diagonal blocks.

Note that the inflated covariance,  $\alpha_l C_l$ , is the true covariance of the stretched measurement perturbation in  $d_l$ .

Inserting (10) (i.e., with inflated covariances) in (8) and (9), it is straightforward to show (see also [2] and [3]) that

$$F = s G_L^T C_L^{-1} G_L, \tag{12}$$

$$f = s G_L^T C_L^{-1} d_L, \tag{13}$$

where  $s \stackrel{\text{def}}{=} \sum_{l=1}^L \alpha_l^{-1}$ . From (6), (7), (12), and (13), it is evident that  $C_{\text{mu}} = C_{\text{po}}$  and  $m_{\text{mu}} = m_{\text{po}}$  iff

$$s = 1. \tag{14}$$

Hence, MDA with inflated data-error covariance matrices yields the correct posterior mean and covariance matrix for  $m$  iff the MDA condition, (14), is satisfied. As commented in [3], there are infinitely many sequences  $\{\alpha_l\}_{l=1}^L$  that satisfies (14).

### 4 Multiple linearly dependent data vectors—PMDA condition

Consider the case where data vectors as well as forward models are linearly dependent,

$$\left. \begin{aligned} d_l &= B_l d_L \\ G_l &= B_l G_L \end{aligned} \right\}; l \in [1, L], \tag{15}$$

where  $B_L$  equals the identity matrix,  $I_L$ . It follows that  $C_l = B_l C_L B_l^T$ ;  $l \in [1, L]$ . Since the data vectors  $\{d_l\}_{l=1}^L$  are dependent, use of the entire sequence as data requires some modification of  $\{C_l\}_{l=1}^L$ . Inspired by the use of inflated covariance matrices in Section 3, one could replace the expression for  $C_l$  by  $\alpha_l B_l C_L B_l^T$ . Inserting the latter expression and (15) into (8) and (9) yields

$$F = G_L^T \left( \sum_{l=1}^L \alpha_l^{-1} B_l^T \left( B_l C_L B_l^T \right)^{-1} B_l \right) G_L, \tag{16}$$

$$f = G_L^T \left( \sum_{l=1}^L \alpha_l^{-1} B_l^T \left( B_l C_L B_l^T \right)^{-1} B_l \right) d_L. \tag{17}$$

From (6), (7), (16), and (17), it follows that  $C_{\text{mu}} = C_{\text{po}}$  and  $m_{\text{mu}} = m_{\text{po}}$  iff

$$\sum_{l=1}^L \alpha_l^{-1} B_l^T \left( B_l C_L B_l^T \right)^{-1} B_l = C_L^{-1}. \tag{18}$$

In general, there is, however, no sequence  $\{\alpha_l\}_{l=1}^L$  that satisfies this matrix equation.

Replacing the inflation of covariances,  $C_l \rightarrow \alpha_l C_l = \alpha_l^{1/2} C_l \alpha_l^{1/2}$ , by  $C_l \rightarrow \Lambda_l^{1/2} C_l \Lambda_l^{T/2} \stackrel{\text{def}}{=} A_l C_l A_l^T$ , where  $\Lambda_l$  is an  $N_l \times N_l$  matrix, leads to (18) being replaced by the PMDA condition

$$\sum_{l=1}^L B_l^T \left( (A_l B_l) C_L (A_l B_l)^T \right)^{-1} B_l = C_L^{-1}. \tag{19}$$

Since the PMDA condition reduces to the MDA condition when  $\{B_l = I_L\}_{l=1}^L$  and  $\{A_l = \alpha_l^{1/2} I_L\}_{l=1}^L$ , it is a generalization of the latter. The transformed data-error covariance with multiple linearly dependent data vectors becomes

$$\Xi_{\text{PMDA}} = A \Xi A^T = \begin{pmatrix} A_1 C_1 A_1^T & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_L C_L A_L^T \end{pmatrix}, \quad (20)$$

where  $A = \text{diag}(A_1 \dots A_L)$ , and where one should recall that  $\{C_l = B_l C_L B_l^T\}_{l=1}^L$ . Note that the transformed covariance,  $A_l C_l A_l^T$ , is the true covariance of the transformed measurement perturbation in  $d_l$ .

### 4.1 Specification of $\{A_l\}_{l=1}^L$

There are more scalar unknowns in the PMDA condition than there are scalar equations, which allows for free specification of some of the unknown before the PMDA condition is invoked to determine the remaining unknowns. In this section, a certain choice will be made regarding which unknowns that will be specified freely. I emphasize that this particular choice, while advantageous for the present exposition, is certainly not mandatory when applying the PMDA condition.

#### 4.1.1 Existence

It is not clear if there exist solutions of (19) that will ensure that all the diagonal blocks in  $\Xi_{\text{PMDA}}$  become proper covariance matrices, that is symmetric and positive semi-definite. It is easy to show that they are symmetric, so the real issue is positive semi-definiteness. Before discussing this issue for the PMDA condition, it is perhaps useful to reiterate the condition ensuring that all the diagonal blocks in  $\Xi_{\text{MDA}}$  become positive semi-definite. To this end, the MDA condition will be expressed in a form dictated by what is advantageous for the discussion of the PMDA condition.

**MDA condition** The MDA condition can be reformulated as

$$\lambda_L C_L \lambda_L = \left( C_L^{-1} - \sum_{l=1}^{L-1} (\lambda_l C_L \lambda_l)^{-1} \right)^{-1}, \quad (21)$$

where  $\lambda \stackrel{\text{def}}{=} \alpha^{1/2}$ . It is evident that for  $\{\lambda_l C_L \lambda_l\}_{l=1}^L$  to be positive semi-definite,  $\{\lambda_l\}_{l=1}^L$  must be real numbers. One may select  $\{\lambda_l\}_{l=1}^{L-1}$  such that  $\{\lambda_l C_L \lambda_l\}_{l=1}^{L-1}$  become positive semi-definite, but then  $\lambda_L$  cannot be selected freely since (21) must be satisfied. To ensure that  $\lambda_L C_L \lambda_L$  becomes positive semi-definite, it is therefore necessary to further restrict the selection of  $\{\lambda_l\}_{l=1}^{L-1}$ . Cancelling out  $C_L$  from (21), it is evident that  $\lambda_L$  becomes a real number iff  $\{\lambda_l\}_{l=1}^{L-1}$  is a sequence of real numbers fulfilling  $\sum_{l=1}^{L-1} \lambda_l^{-2} < 1$ .

**PMDA condition** Since  $B_L = I_L$ , (19) can be reformulated as

$$A_L C_L A_L^T = \left( C_L^{-1} - \sum_{l=1}^{L-1} B_l^T (A_l C_l A_l^T)^{-1} B_l \right)^{-1}. \quad (22)$$

The situation is similar as for the MDA condition in that one can select  $\{A_l\}_{l=1}^{L-1}$  such that  $\{A_l C_l A_l\}_{l=1}^{L-1}$  become positive semi-definite, while  $A_L$  cannot be selected freely since (22) must be satisfied. But the situation also differs from that of the MDA condition in that it is not possible to cancel out  $C_L$  from (22). This contributes to making it very challenging to derive a condition that the matrices in  $\{A_l\}_{l=1}^{L-1}$  should satisfy to ensure that  $A_L C_L A_L^T$  becomes positive semi-definite. The issue is, of course, important, and researchers applying the PMDA condition will have to consider it for their particular application (i.e., when  $C_L$  and  $\{B_l\}_{l=1}^{L-1}$  are fully specified), but it is beyond the scope of this paper.

#### 4.1.2 Computation of $A_L C_L A_L^T$

Defining the matrix sequence  $\{X_l\}_{l=1}^{L-1}$  by

$$X_l = C_L B_l^T (A_l C_l A_l^T)^{-1} B_l, \quad (23)$$

and rearranging the right-hand side of (22), one may write

$$A_L C_L A_L^T = (I_L - Q_L)^{-1} C_L, \quad (24)$$

where  $Q_L \stackrel{\text{def}}{=} \sum_{l=1}^{L-1} X_l$ . Calculating  $A_L C_L A_L^T$  thus entails the inversion of the  $N_L \times N_L$ -matrix  $I_L - Q_L$ , in addition to inversion of  $L - 1$  matrices of dimension  $\{N_l\}_{l=1}^{L-1}$  in order to calculate  $\{X_l\}_{l=1}^{L-1}$ . Altogether, this constitutes a computationally challenging task for large  $N_L$ .

To alleviate the computational burden in a practical application of the PMDA condition to a nonlinear problem where  $N_L$  is large, it is an option to utilize approximate matrix factorization for the matrix inversions. It is also an option to use truncated Neumann series to approximately evaluate  $(I_L - Q_L)^{-1}$  if  $\|Q_L\| < 1$ . (Here  $\|\cdot\|$  denotes an operator norm.) Obviously, invoking such options will result in only approximately correct sampling in the linear-Gaussian case.

#### 4.1.3 Selection of $\{A_l\}_{l=1}^{L-1}$

There are altogether  $\sum_{l=1}^{L-1} N_l^2$  elements in the matrices  $\{A_l\}_{l=1}^{L-1}$ , and it is not clear how to use all these degrees of freedom optimally, for example with respect to the existence issue discussed in Section 4.1.1. Specifying  $\sum_{l=1}^{L-1} N_l^2$  numbers is also manually cumbersome for all but small problems. It can therefore be of interest to consider a reduction of the degrees of freedom.

**Simplistic alternative** If one lets  $\{A_l = \alpha_l^{1/2} I_l\}_{l=1}^{L-1}$ , the manual selection is no more cumbersome than with the MDA condition, and one may write

$$\Xi_{\text{PMDA}} = \begin{pmatrix} \Xi_{\text{MDA}}^{[1, L-1]} & 0 \\ 0 & (I_L - Q_L)^{-1} C_L \end{pmatrix}, \tag{25}$$

where the matrix sequence  $\{X_l\}_{l=1}^{L-1}$  is now given by

$$X_l = \alpha_l^{-1} C_L B_l^T C_l^{-1} B_l. \tag{26}$$

Hence, the first  $L - 1$  diagonal blocks are inflated as for the MDA condition, while the PMDA condition is resolved by calculating block number  $L$ .

It is perhaps useful to illustrate the simplistic alternative with a small, simple example. To this end, let  $L = 2$ ,  $N_1 = 1$ ,  $N_2 = 2$ ,  $C_2 = I_2 \in \mathbb{R}^{2 \times 2}$ , and let  $B_1 = \frac{1}{2}(1, 1)$ , that is,  $B_1$  is the arithmetic averaging operator on  $\mathbb{R}^2$  (see Appendix A.2 for why averaging operators are relevant). It follows that

$$(I_2 - Q_2)^{-1} C_2 = \frac{\alpha_1}{\alpha_1 - 1} \begin{pmatrix} 1 - \frac{1}{2\alpha_1} & \frac{1}{2\alpha_1} \\ \frac{1}{2\alpha_1} & 1 - \frac{1}{2\alpha_1} \end{pmatrix}, \tag{27}$$

which is a proper covariance matrix iff  $\alpha_1 > 1$ , that is, if  $C_1$  is inflated.

Finally, I will briefly discuss what needs to be considered when specifying  $\{\alpha_l\}_{l=1}^{L-1}$  for the simplistic alternative in general. Selecting large entries in  $\{\alpha_l\}_{l=1}^{L-1}$  will be beneficial with respect to the existence issue, since  $(I_L - Q_L)^{-1} C_L \approx C_L$  for large  $\alpha_l$ 's. It would also facilitate use of a truncated Neumann series to approximate  $(I_L - Q_L)^{-1}$ . Selecting too large  $\alpha_l$ 's would, however, result in  $Q_L \approx 0$ , and

$$\Xi_{\text{PMDA}}^{-1} \approx \begin{pmatrix} 0 & 0 \\ 0 & C_L^{-1} \end{pmatrix}, \tag{28}$$

so that only  $d_L$  would influence the estimation results. In that case there would be no point in including  $\{d_l\}_{l=1}^{L-1}$  and the corresponding forward-model sequence in the modeling. Further investigations into these issues should be conducted for each application separately, and such investigations are beyond the scope of this paper.

**Funding information** Partial financial support was provided by the finalized research project ‘‘4D Seismic History Matching,’’ funded by Eni, Petromaks 2), and Total EP NORGE, and by the ongoing research project ‘‘Assimilating 4D seismic data: big data into big models,’’ funded by Aker BP, Equinor, Lundin, Repsol, the Research Council of Norway (Petromaks 2), and Total EP NORGE.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not

included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Multilevel data assimilation—motivation

Multilevel data assimilation has two main components; multilevel simulations and multilevel data. I briefly motivate use of multilevel simulations (Appendix A.1) and multilevel data (Appendix A.2), and point out that multilevel data vectors is a special case of linearly dependent data vectors.

### A.1 Multilevel simulations

Statistical sampling quality and computational cost of ensemble-based methods both increase with ensemble size. To alleviate negative sampling effects of using a small or moderately large ensemble, localization [11] is routinely applied for field cases. It can, however, be challenging to design proper localization regions for porous-media flow problems. Alternative methods for alleviating effects of sampling errors, which can also be used in conjunction with localization, thereby making localization less challenging, has started to appear. One type of methods performs forward simulations on a single coarsened grid [7, 13] or on a multilevel sequence of coarsened temporal [10] or spatial [8, 9] grids. Another alternative is to perform most of the forward simulations with a linearized forward model [17]. The main idea behind all these methods is to allow for a substantially larger ensemble size (thereby increasing sampling quality) without increasing the total computational cost. Roughly speaking, numerical accuracy in the forward simulations is traded for statistical accuracy in the final results.

### A.2 Multilevel data

Various types of spatial data appear in many areas of geoscience, for example, rainfall and temperature in climate and land-use models and inverted seismic data in reservoir modeling. When spatial data are applied with a numerical simulator, some kind of averaging has already taken place since subgrid variations in the data are not represented. When spatial data are used in conjunction with multilevel simulations, multilevel averaging of the observed data, that is the data vector on the finest level, will take place in order for the averaged data to ‘‘fit’’ the simulated output at grids of multiple scales. Clearly, the elements in the resulting sequence of data at different scales will not be independent. With arithmetic averaging of data from

one scale to the next coarser scale, there will be linear dependencies between data vectors at the different scales, and  $\text{rank}(B_1) < \text{rank}(B_2) < \dots < \text{rank}(B_L)$ .

The PMDA condition derived in Section 4 (and the MDA condition derived in Section 3) builds on the assumption that  $\Xi$  is block diagonal (see (3)). Use of the PMDA condition in the context of assimilation of multilevel data therefore requires that the data-error vectors in the different assimilation cycles are uncorrelated. This is the same requirement as with use of the MDA condition when performing multiple data assimilation (see the first paragraph of Section 2).

## References

- Chen, Y., Oliver, D.S.: Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.* **17**(4), 689–703 (2013)
- Emerick, A.A., Reynolds, A.C.: History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilation. *Comput. Geosci.* **16**(3), 639–659 (2012)
- Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. *Comput. Geosci.* **55**, 3–15 (2013)
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10,143–10,162 (1994)
- Fossum, K., Mannseth, T.: Parameter sampling capabilities of sequential and simultaneous data assimilation. Part I: analytical comparison. *Inverse Probl.* **30**(25pp), 114,002 (2014)
- Fossum, K., Mannseth, T.: Parameter sampling capabilities of sequential and simultaneous data assimilation. Part II: statistical analysis of numerical results. *Inverse Probl.* **30**(28pp), 114,003 (2014)
- Fossum, K., Mannseth, T.: Coarse-scale data assimilation as an alternative to localization. *Comput. Geosci.* **21**(1), 167–186 (2017)
- Fossum, K., Mannseth, T.: A novel multilevel method for assimilating spatially dense data. In: *Proceedings of ECMOR XVI – 16<sup>th</sup> European Conference on the Mathematics of Oil Recovery*. Barcelona, Spain (2018)
- Fossum, K., Mannseth, T., Stordal, A.S.: Assessment of multilevel ensemble based data assimilation for reservoir history matching. *Comput. Geosci.* In press
- Hoel, H., Law, K.J.H., Tempone, R.: Multilevel ensemble Kalman filtering. *SIAM J. Numer. Anal.* **54**(3), 1813–1839 (2016)
- Houtekamer, P.L., Mitchell, H.L.: Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**(3), 796–811 (1998)
- van Leeuwen, P.J., Evensen, G.: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev.* **124**(12), 2898–913 (1996)
- Mannseth, T., Fossum, K.: Assimilating spatially dense data for subsurface applications—balancing information and degrees of freedom. *Comput. Geosci.* **22**(5), 1323–1349 (2018)
- Neal, R.M.: Annealed importance sampling. *Stat. Comput.* **11**(2), 125–139 (2001)
- Stordal, A.S., Elsheikh, A.H.: Iterative ensemble smoothers in the annealed importance sampling framework. *Adv. Water Resour.* **86**, 231–239 (2015)
- Tarantola, A.: *Inverse Problem Theory and Model Parameter Estimation*. SIAM Publications, Philadelphia (2005)
- Tarrahi, M., Elahi, S.H., Jafarpour, B.: Fast linearized forecasts for subsurface flow data assimilation with ensemble Kalman filter. *Comput. Geosci.* **20**(5), 929–952 (2016)
- Wang, Y., Li, G., Reynolds, A.C.: Estimation of depths of fluid contacts by history matching using iterative ensemble smoothers. *SPE J.* **15**(2), 509–525 (2010)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.