Computational Geosciences
https://doi.org/10.1007/s10596-024-10289-8

ORIGINAL PAPER

Multiscale model diagnostics

Trond Mannseth[1]

Received: 18 August 2023 / Accepted: 11 April 2024

Abstract
I consider the problem of model diagnostics, that is, the problem of criticizing a model prior to history matching by comparing data to an ensemble of simulated data based on the prior model (prior predictions). If the data are not deemed as a credible prior prediction by the model diagnostics, some settings of the model should be changed before history matching is attempted. I particularly target methodologies that are computationally feasible for large models with large amounts of data. A multiscale methodology, that can be applied to analyze differences between data and prior predictions in a scale-by-scale fashion, is proposed for this purpose. The methodology is computationally inexpensive, straightforward to apply, and can handle correlated observation errors without making approximations. The multiscale methodology is tested on a set of toy models, on two simplistic reservoir models with synthetic data, and on real data and prior predictions from the Norne field. The tests include comparisons with a previously published method (termed the Mahalanobis methodology in this paper). For the Norne case, both methodologies led to the same decisions regarding whether to accept or discard the data as a credible prior prediction. The multiscale methodology led to correct decisions for the toy models and the simplistic reservoir models. For these models, the Mahalanobis methodology either led to incorrect decisions, and/or was unstable with respect to selection of the ensemble of prior predictions.

Keywords Prior predictive distribution · Model criticism · Multiscale test vectors · Synthetic data · Norne field data

1 Introduction

Reliable reservoir simulation forecasts, with quantification of the uncertainty in the forecasts, are instrumental for sound reservoir management. The Bayesian framework allows for all types of information to be incorporated and facilitates uncertainty quantification as part of the history matching (HM) procedure. Ensemble-based data assimilation (DA) (see, e.g., [2, 9]) is an increasingly popular family of approximate solutions to the Bayes problem, which is computationally feasible for reservoir HM.

Even if ensemble-based DA is computationally feasible, the computational work is still very large for field-scale cases. This comes in addition to the substantial manual work required for computer assisted HM of a real field case. Part of the manual work is to prepare the reservoir model for DA so that the DA has the possibility of being successful. That is, the reservoir model should pass a check (*model diagnostics*) where samples of simulated output with measurement errors added (*prior predictions*) are compared to observed data to assess if the observed data could be a credible realization from the distribution of prior predictions. If so, the model is ready for DA to be performed. If not, obstacles that are expected to prevent successful DA should be removed and model settings that are not compatible with observed data should be adjusted before DA is conducted.

Several geological scenarios are usually viable for a reservoir, and the scenario uncertainty should be taken into account in the reservoir description. This can be done by combining several scenarios in the uncertainty quantification, for example through Bayesian model averaging (see, e.g., [15]) or Bayesian stacking (see, e.g., [1, 27]). Working with more than a single scenario will, however, add significantly to the total work load. Model diagnostics can then be useful in a screening process to decide if some scenarios can be discarded before the remaining scenarios are combined [14, 21, 24].

It is standard in ensemble-based data assimilation to visually compare the ensemble of prior predictions to the

✉ Trond Mannseth
trma@norceresearch.no

[1] NORCE Norwegian Research Centre, Nygårdsgaten 112, Bergen 5008, Norway

Published online: 28 May 2024

observed data. If there are observations that are not contained within the spread of the prior predictions, the model setup is deemed inadequate [5, 8]. The converse is, however, not necessarily true: coverage of observed data from the Norne field by the prior predictions was not sufficient to facilitate a good match after DA [6].

More advanced methodologies for model diagnostics can be found, for example, in [3, 10, 13, 14, 21, 24, 26, 28]. Support vector machine and tree-based regression are utilized, for example, in [10] and described also in [26, Chapter 2]. The methodologies in [14, 21, 24] utilize a type of Bayesian model averaging to compute model probabilities. Hence, it is implicitly assumed that one of the model setups is adequate. In [3], the authors propose a methodology that directly compares prior predictions to observed data without such an assumption. A toy example in [3] (see, their Fig. 1) illustrates that coverage of individual observations by the prior predictions is insufficient as a tool for model diagnostics since it does not take into account trends and shapes of the data and prior predictions. To account for trends and shapes, they propose a technique utilizing a certain (approximate) Mahalanobis distance as discrepancy measure. They illustrate use of the methodology on some toy examples, before they apply it to the Norne data and show that certain settings in the Norne simulation model should be changed in order to facilitate a good data match with DA. A Mahalanobis distance was also applied in [13] on problems that did not require the Mahalanobis distance to be approximated. In [28], the authors utlize robust Mahalanobis distances, that is, robust estimates of mean and covariance matrix [16], when evaluating Mahalanobis distances. When using robust Mahalanobis distances, it is required that $E \geq 2N$ and recommended that $E > 5N$ [16], where E denotes the ensemble size and N denotes the data dimension.

The aim of this paper is to present a methodology for model diagnostics applicable to large reservoir models with large amounts of data. In this situation, N (often of the order of 10^3–10^5) will be much larger than E (typically of the order of 10^2), so that utilization of robust Mahalanobis distances is not a viable alternative. Hence, I address the same type of problem as in [3], but with a different methodology for model diagnostics. A key difference is that I will apply a multiscale discrepancy measure that analyzes trends and shapes in a scale-by-scale fashion, as opposed to previously proposed methodologies that lumps all scales into a single measure. An advantage of the multiscale methodology with respect to the methodology in [3], is that the multiscale methodology can handle problems with correlated measurement errors where the number of data is larger than the ensemble size without making approximations.

The novel multiscale methodology, and a brief summary of the methodology in [3], will be presented in Section 2, while results illustrating their performances will be shown in Section 3. The numerical examples will include toy models (Section 3.1), reservoir models with synthetic data (Section 3.2), and the Norne G-segment model with real data (Section 3.3).

2 Methodologies

It is not straightforward to directly compare data to prior predictions, particularly in high-dimensional spaces. In this paper, I will utilize selected test functionals on the data space for the comparison.

Let $\mathcal{F}$ denote a test functional, $W := \{w_e\}_{e=1}^E$ the set of prior predictions, and d the observed data. With real data, $\mathcal{F}(d)$ will be compared to statistics based on $\mathcal{F}(W) := \{\mathcal{F}(w_e)\}_{e=1}^E$. To test the robustness of the methodologies considered with respect to potential dependency of their results on a particular data vector, it would be desirable to have a set of data vectors available. With synthetic data, one may utilize the data-error distribution to generate the data set, $D := \{d_e\}_{e=1}^E$. For such data, I will therefore compare statistics based on $\mathcal{F}(D) := \{\mathcal{F}(d_e)\}_{e=1}^E$ to statistics based on $\mathcal{F}(W)$.

Before proceeding with the methodology descriptions, it will be convenient to label the methodologies. I will denote the methodology proposed in [3] the *Mahalanobis methodology*, and the novel methodology the *multiscale methodology*.

2.1 Summary of the Mahalanobis methodology

In the numerical examples in Section 3, I will compare the performance of the multiscale methodology to that of the Mahalanobis methodology. In Sections 2.1.1 and 2.1.2, I therefore briefly summarize the Mahalanobis methodology for the convenience of the reader.

2.1.1 Mahalanobis distance

Let $\mu \in \mathbb{R}^N$ denote the mean and $\Sigma \in \mathbb{R}^{N\times N}$ the covariance matrix of a sample with size E. The squared Mahalanobis distance, $\mathcal{M}(y)$, from an arbitrarily selected vector, $y \in \mathbb{R}^N$, to the probability distribution underlying the sample is then

$$\mathcal{M}(y) = (y-\mu)^T \Sigma^{-1} (y-\mu). \tag{1}$$

For the problems considered here, Σ will be singular since typically $E \ll N$, and hence, the Mahalanobis distance will not exist. An approximation, $\tilde{\mathcal{M}}(y)$ to $\mathcal{M}(y)$ can, however, be computed utilizing a regularized covariance matrix, $\tilde{\Sigma}$. In [3, 4] Σ was regularized using a shrinkage covariance estimate [17], using Target B of [23] to shrink the sample

covariance towards a diagonal matrix with a constant value equal to the average variance,

$$\tilde{\Sigma} = \delta \nu I + (1 - \delta)\, \Sigma, \tag{2}$$

where δ denotes the shrinkage parameter, ν the sample variance, and I the identity matrix. Following [3], Σ will be computed from the sample using the maximum likelihood estimate, while the shrinkage parameter will be set equal to $2/(E+2)$ (an estimate of its optimal value [18]). A computationally more efficient form of $\tilde{\mathcal{M}}(y)$, requiring Cholesky decomposition of an $E \times E$ (as opposed to $N \times N$) matrix, can then be derived utilizing the Sherman-Morrison-Woodbury formula to reformulate the expression for $\tilde{\Sigma}^{-1}$ [3].

2.1.2 Leave-one-out strategy

To build statistics for the Mahalanobis distance, [3] utilized a leave-one-out strategy. Looping over the prior predictions, the prediction for which they compute the approximate Mahalanobis distance, w_e, is left out when calculating the mean, μ_{-e}, and covariance, $\tilde{\Sigma}_{-e}$,

$$\tilde{\mathcal{M}}_{-e}(w_e) = (w_e - \mu_{-e})^T\, \tilde{\Sigma}_{-e}^{-1}\, (w_e - \mu_{-e})\,. \tag{3}$$

With real data, the empirical cumulative distribution function (CDF) for $\tilde{\mathcal{M}}(d) := \{\tilde{\mathcal{M}}_{-e}(d)\}_{e=1}^{E}$ is compared to the empirical CDF for $\tilde{\mathcal{M}}(W) := \{\tilde{\mathcal{M}}_{-e}(w_e)\}_{e=1}^{E}$.

With synthetic data, the empirical CDF for $\rho(\tilde{\mathcal{M}}(D)) := \{\rho(\tilde{\mathcal{M}}(d_e))\}_{e=1}^{E}$, where ρ denotes the median, is compared to the empirical CDF for $\tilde{\mathcal{M}}(W)$.

A $\pi\%$ *credible interval* is defined based on the empirical CDF for $\tilde{\mathcal{M}}(W)$. I will follow [3] and let $\pi = 96$ (which corresponds approximately to ± 2 standard deviations from the mean for a Gaussian probability distribution). With real data, d is deemed as a credible prior prediction if a sufficiently large percentage, λ_{md}, of the Mahalanobis distances in $\tilde{\mathcal{M}}(d)$ is contained within the credible interval. With synthetic data, one can assess the robustness (towards effects caused by a particular data vector) of the Mahalanobis methodology by calculating λ_{md} for the members of D in cases where it is known *a priori* that the members of D should be rejected. The authors of [3] do not propose an upper bound for λ_{md} in order for the Mahalanobis methodology to be deemed successful in this respect, so this is left to the judgment of the user.

2.2 Multiscale methodology

The multiscale methodology utilizes projections onto a set of multiscale test vectors, $H := \{h_l\}_{l=0}^{L}$, where $h_l \in \mathbb{R}^{N \times M}$. The set H is ordered so that the characteristic length of variation of h_l decreases monotonically when l increases. Define a set of multiscale projections of the set of prior predictions by

$$\mathcal{H}(W) := \{\mathcal{H}_l(W)\}_{l=0}^{L}, \tag{4}$$

$$\mathcal{H}_l(W) := \{\mathcal{H}_l(w_e)\}_{e=1}^{E}, \tag{5}$$

$$\mathcal{H}_l(w_e) := \frac{1}{NM} h_l^T\, w_e. \tag{6}$$

With a set of data vectors, which will be the case for synthetic data, the multiscale projections can be applied to the set of data vectors by replacing W by D and w_e by d_e in (4)–(6).

With a single data vector, which will be the case for real data, (4)–(6) reduces to

$$\mathcal{H}(d) := \{\mathcal{H}_l(d)\}_{l=0}^{L}, \tag{7}$$

$$\mathcal{H}_l(d) := \frac{1}{NM} h_l^T\, d. \tag{8}$$

Let $\xi_l(V)$ and $\zeta_l(V)$ denote the mean and standard deviation of $\mathcal{H}_l(V) := \{\mathcal{H}_l(v_e)\}_{e=1}^{E}$, respectively, where v is wildcard notation for w and d. Algorithm 1 summarizes the multiscale methodology. In addition, anyone using the multiscale methodology needs to decide when d should be discarded as a credible prior prediction (for synthetic data; when D should be discarded as credible prior predictions) based on it's results. In this paper, I will discard d as a credible prior prediction if at least one $\mathcal{H}_l(d)$ falls clearly outside the interval $B_l := [\xi_l(W) - 2\zeta_l(W), \xi_l(W) + 2\zeta_l(W)]$, or if several of the $\mathcal{H}_l(d)$'s falls at least slightly outside their corresponding B_l's. With synthetic data, I will discard D as credible prior predictions if at least one $\xi_l(D)$ falls clearly outside B_l, or if several of the $\xi_l(D)$'s falls at least slightly outside their corresponding B_l's. In addition, I will illustrate the robustness of the multiscale methodology by presenting the percentage, $\lambda_{\text{ml}}(l)$, of $\mathcal{H}_l(d_e)$ in B_l for the members of D. The vector $[\lambda_{\text{ml}}(0), \ldots, \lambda_{\text{ml}}(L)]$ will be denoted by λ_{ml}.

In Section 3.3.3, I will apply a particular set of test vectors that will be described there. Otherwise, I will let $M = 1$, and apply the set of test vectors described in Section 2.2.1. These vectors are derived from the Haar wavelet basis [7, 12]. There are, however, numerous other options for test vectors. It is the multiscale feature of H, that is, the ability to facilitate discrimination between shapes at multiple scales, that is important. One may, for example, utilize other wavelet bases (see, e.g., [7, Chapter 5]) than the Haar basis to construct alternative test vectors.

Algorithm 1 Multiscale methodododology.

procedure MULTISCALE SYNTHETIC DATA(H, W, D, M) ▷ H: Haar vectors, W: Prior predictions, D: Data vectors, M: Spatial dimension
  $E_w \leftarrow W$.shape[1] ▷ Python style
  $E_d \leftarrow D$.shape[1]
  for $e \leftarrow 0, E_w$ **do**
    $\mathcal{H}(w_e) \leftarrow$ PROJECTIONS($H, W[:, e], M$)
  end for
  for $e \leftarrow 0, E_d$ **do**
    $\mathcal{H}(d_e) \leftarrow$ PROJECTIONS($H, D[:, e], M$)
  end for
  for $l \leftarrow 0, L$ **do**
    $\xi_l(W) \leftarrow$ mean $(\mathcal{H}_l(W))$ ▷ Mean for each scale
    $\zeta_l(W) \leftarrow$ std $(\mathcal{H}_l(W))$ ▷ Standard deviation for each scale
    $\xi_l(D) \leftarrow$ mean $(\mathcal{H}_l(D))$
    $\zeta_l(D) \leftarrow$ std $(\mathcal{H}_l(D))$
  end for
  return $\xi(W), \zeta(W), \xi(D), \zeta(D)$
end procedure

procedure MULTISCALE REAL DATA(H, W, d, M) ▷ H: Haar vectors, W: Prior predictions, d: Data vector, M: Spatial dimension
  $E_w \leftarrow W$.shape[1]
  for $e \leftarrow 0, E_w$ **do**
    $\mathcal{H}(w_e) \leftarrow$ PROJECTIONS($H, W[:, e], M$)
  end for
  $\mathcal{H}(d) \leftarrow$ PROJECTIONS(H, d, M)
  for $l \leftarrow 0, L$ **do**
    $\xi_l(W) \leftarrow$ mean $(\mathcal{H}_l(W))$
    $\zeta_l(W) \leftarrow$ std $(\mathcal{H}_l(W))$
  end for
  return $\xi(W), \zeta(W), \mathcal{H}(d)$
end procedure

procedure PROJECTIONS(H, v, M) ▷ H: Haar vectors, v: Vector, M: Spatial dimension
  $L \leftarrow H$.shape[1] $- 1$
  $N \leftarrow v$.shape[0]
  for $l \leftarrow 0, L$ **do**
    $q_l \leftarrow \frac{1}{NM} v^T H[:, l]$
  end for
  return q
end procedure

2.2.1 Haar vectors

I will utilize multiscale test vectors constructed from the normalized Haar basis [7, 12], and refer to these as *Haar vectors*. The Haar basis has elements $\phi(x)$ and $\psi_k^l(x)$, where $l = 0, 1, 2, \ldots$ and $k = 1, 2, \ldots, 2^l$. Without loss of generality, the presentation of the Haar vectors will assume that $x \in [0, 1]$, and while the applications will utilize discrete Haar vectors defined on $\{x_n\}_{n=1}^N$, the presentation of the Haar vectors will assume that x is continuous. The normalized Haar basis may then be written as

$$\phi(x) := 1, \tag{9}$$

$$\psi(x) := \begin{cases} 1, & x \in \left[0, \frac{1}{2}\right], \\ -1, & x \in \left(\frac{1}{2}, 1\right], \end{cases} \tag{10}$$

$$\psi_k^l(x) := 2^{l/2} \psi\left(2^l x - (k-1)\right). \tag{11}$$

I now define the multiscale test vectors in H by

$$h_l(x) := \begin{cases} \phi(x), & l = 0, \\ \sum_{k=1}^{2^{(l-1)}} \psi_k^{(l-1)}(x), & l \in [1, L]. \end{cases} \tag{12}$$

The Haar vectors $\{h_l\}_{l=0}^3$ are shown in Fig. 1.

3 Applications

The multiscale methodology will now be demonstrated, and compared to the Mahalanobis methodology, on some toy examples, on two reservoir models with synthetic data, and on real data from the G-segment of the Norne field.

Denote the empirical CDF for $\tilde{\mathcal{M}}(W)$ by CDF_W, and the empirical CDFs for $\rho(\tilde{\mathcal{M}}(D))$ and $\tilde{\mathcal{M}}(d)$ by CDF_D and CDF_d, respectively. For the Mahalanobis methodology, I will show plots of CDF_W and CDF_D (synthetic data), and CDF_W and CDF_d (real data).

With synthetic data, I could show plots of CDFs for the multiscale methodology as well, but then I would have to show two CDFs for each l. That would lead to too many plots, and collecting the CDFs for all scales in a single plot would not give good readability. I will therefore present the results obtained with the multiscale methodology in an alternative manner. For the multiscale methodology with synthetic data, I will show $\xi_l(W), \xi_l(D), \xi_l(W) \pm 2\zeta_l(W)$ and $\xi_l(D) \pm 2\zeta_l(D)$ for all scales in a single plot. With real data, I will show $\xi_l(W)$, $\xi_l(d)$ and $\xi_l(W) \pm 2\zeta_l(W)$ for all scales in a single plot. In Section 3.1.2, I will show $\zeta_l(W)$ and $\zeta_l(D)$ for all scales in a single plot.

3.1 Toy examples

For all examples in Section 3.1, $N = 1024$, while W is drawn from $\mathcal{N}(o, \Sigma_w)$, where $\mathcal{N}$ denotes the multivariate Gaussian distribution, $o := (0, \ldots, 0)^T \in \mathbb{R}^N$, and Σ_w is based on a cubical variogram with sill 1 and range 25. It is assumed that there is a set of data vectors, $D := \{d_e\}_{e=1}^E$, available, and that $E = 100$ for both W and D. D is drawn from $\mathcal{N}(\mu, \Sigma_d)$, where Σ_d is based on a variogram with sill s and range r. In the different subsections, I will show results for selected D's, where μ, s, r, and variogram type, may differ from case to case.

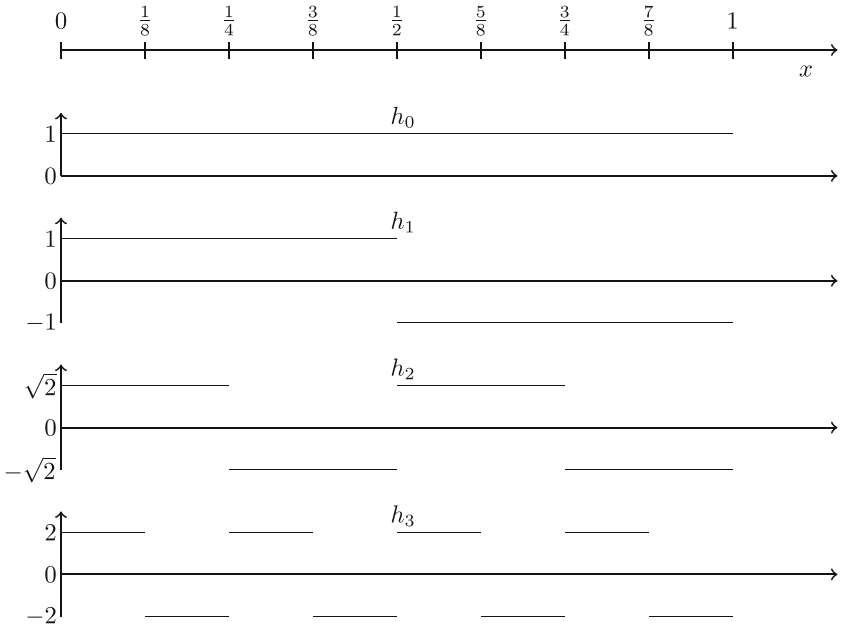

Fig. 1 The four coarsest Haar vectors on [0, 1] with a ruler above. The vertical scaling is different for the different h_l's because they are constructed from the *normalized* Haar basis

3.1.1 Identical distributions

I start by presenting plot types and color coding that will occur repeatedly throughout Section 3. For this, I will show results obtained when D is drawn from the same distribution as W.

Figure 2(a) shows the prior predictions and an arbitrarily selected data realization. Figure 2(b) shows the same data realization as in Fig. 2(a) and an arbitrarily selected prior prediction. Figure 2(c) shows results obtained with the Mahalanobis methodology. The black horizontal lines mark the limits of the 96% credible interval based on

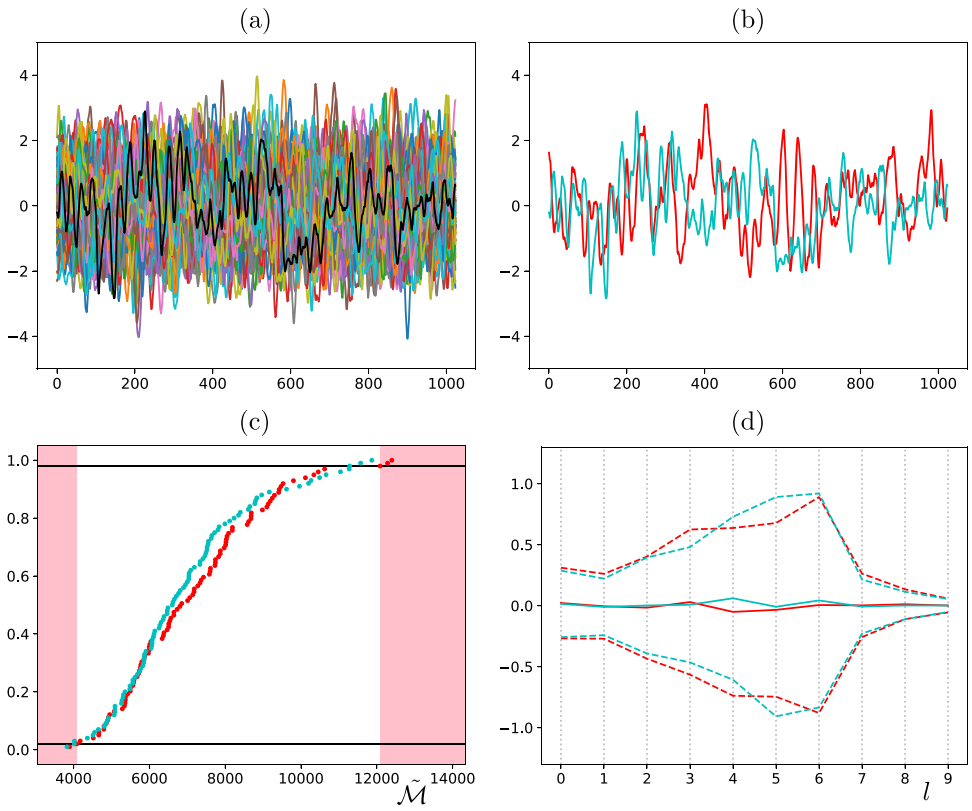

Fig. 2 **(a)**: Prior predictions (colored) and arbitrarily selected data realization (black). **(b)**: Arbitrarily selected prior prediction (red) and the data realization depicted in (a) (cyan). **(c)**: CDF_W (red) and CDF_D (cyan). **(d)**: $\xi_l(W)$ (red solid), $\xi_l(D)$ (cyan solid), $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed), and $\xi_l(D) \pm 2\zeta_l(D)$ (cyan dashed).

CDF_W, so that data realizations with Mahalanobis distances in the pink shaded regions will not be deemed as credible prior predictions. For this example, $\lambda_{\text{md}} = 96$. Figure 2(d) shows results obtained with the multiscale methodology. Note that even though I (to secure good readability) select to plot (solid and dashed) *curves*, results are obtained for integer values of l only, as indicated by the vertical light gray dotted lines. For this example, $\lambda_{\text{ml}} = [96, 98, 99, 100, 96, 86, 95, 98, 98, 96]$.

Since D and W are drawn from the same distribution, the minor differences seen between results obtained for data and prior predictions would of course disappear if E had been increased sufficiently.

3.1.2 Different correlation lengths

D is drawn from $\mathcal{N}(o, \Sigma_d)$ with a cubical variogram with $s = 1$.

Figure 3(a) shows the prior predictions and an arbitrarily selected data realization when $r = 50$. Figure 3(c) shows corresponding results obtained with the Mahalanobis methodology. It is seen that the Mahalanobis methodology would accept more than half ($\lambda_{\text{md}} = 58$) of the data realizations as credible prior predictions even though the range of the underlying variogram is twice that of the variogram for the prior predictions. Figure 3(e) shows corresponding results obtained with the multiscale methodology. It is seen

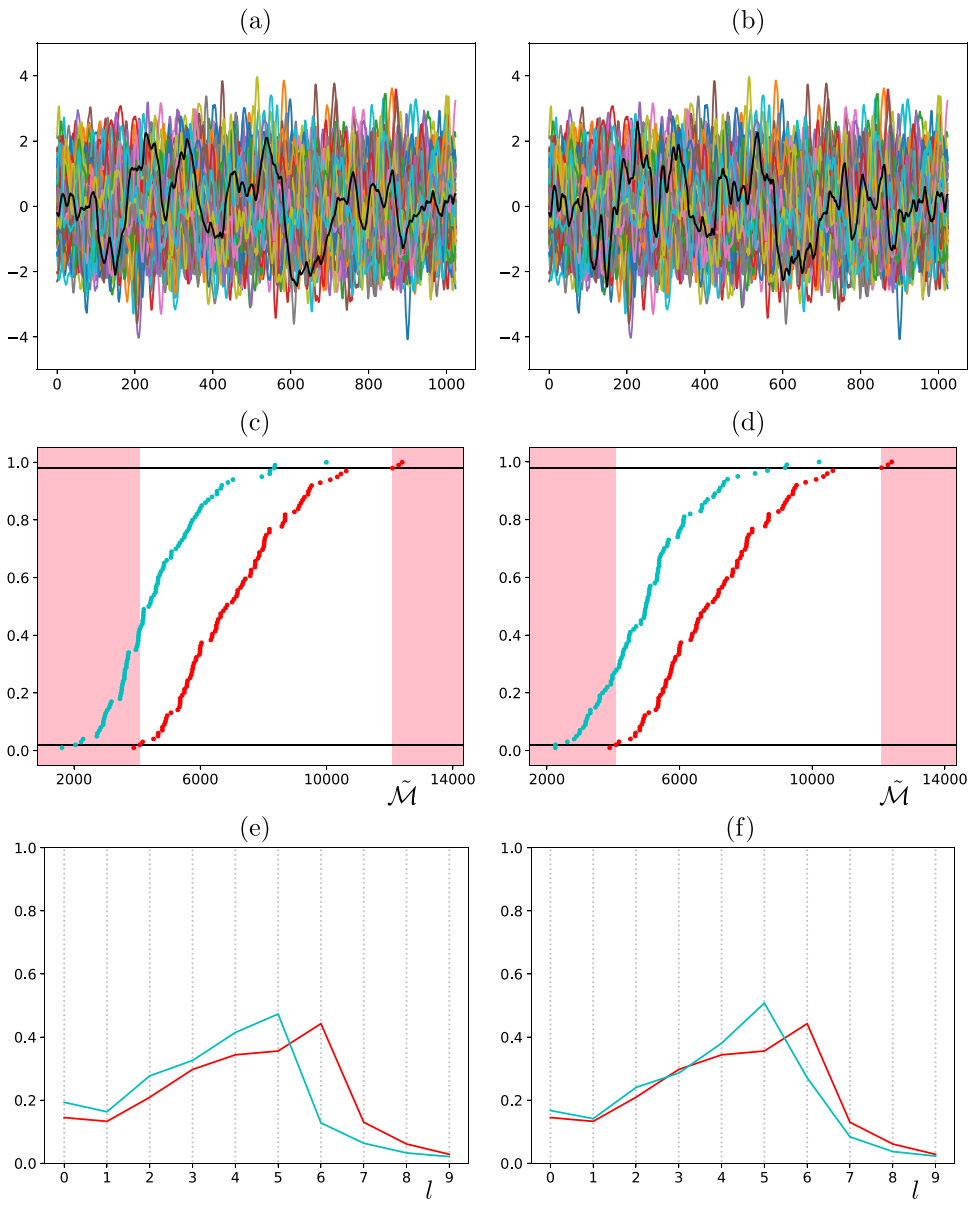

Fig. 3 Left column: $r = 50$. Right column: $r = 37.5$. **(a, b)**: Prior predictions (colored) and arbitrarily selected data realization (black). **(c, d)**: CDF_W (red) and CDF_D (cyan). **(e, f)**: $\zeta_l(W)$ (red) and $\zeta_l(D)$ (cyan)

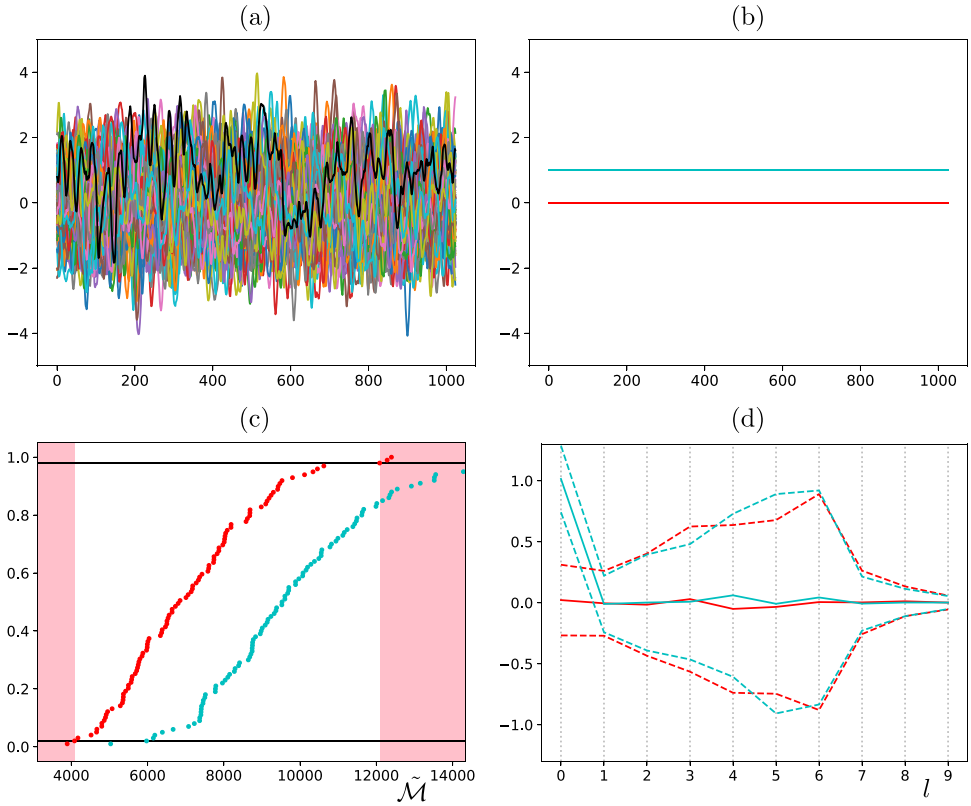

Fig. 4 **(a)**: Prior predictions (colored) and arbitrarily selected data realization (black). **(b)**: $\overline{W}$ (red) and $\overline{D} = u$ (cyan). **(c)**: CDF_W (red) and CDF_D (cyan). **(d)**: $\xi_l(W)$ (red solid), $\xi_l(D)$ (cyan solid), $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed), and $\xi_l(D) \pm 2\zeta_l(D)$ (cyan dashed)

that while both $\zeta_l(W)$ and $\zeta_l(D)$ approaches zero for the finest scales, $\zeta_l(D)$ attains its maximum value at a coarser scale than what $\zeta_l(W)$ does. It is a general feature of $\zeta_l(\cdot)$ that it attains its maximum value at the scale where the characteristic length of variation of h_l is closest to the characteristic length of variation of the argument of $\zeta_l(\cdot)$. The results seen on Fig. 3(e) are therefore caused by the correlation length of D being longer than that of W.

Figure 3(b) shows the prior predictions and an arbitrarily selected data realization when $r = 37.5$. Figure 3(d) shows corresponding results obtained with the Mahalanobis methodology. In this case, the Mahalanobis methodology would accept about three quarters ($\lambda_{\text{md}} = 73$) of the data realizations as credible prior predictions even though the range of the underlying variogram is significantly larger than that of the variogram for the prior predictions. Figure 3(f) shows corresponding results obtained with the multiscale methodology. It is seen that also in this case, the multiscale methodology is able to discriminate between the data realizations and the prior predictions.

3.1.3 Different stationary means

D is drawn from $\mathcal{N}(u, \Sigma_d)$ with a cubical variogram with $s = 1$ and $r = 25$, where $u := (1, \ldots, 1)^T \in \mathbb{R}^N$.

Figure 4(a) shows the prior predictions and an arbitrarily selected data realization. Figure 4(b) shows u and $\overline{W}$. Figure 4(c) shows results obtained with the Mahalanobis methodology. It is seen that the Mahalanobis methodology would accept most ($\lambda_{\text{md}} = 85$) of the data realizations as credible prior predictions. Figure 4(d) shows results obtained with the multiscale methodology. It is seen that while $\xi_l(D)$ and $\xi_l(W)$ attain very similar values for $l \geq 1$, $\xi_0(D)$ is clearly outside B_0 ($\lambda_{\text{ml}} = [0, 98, 99, 100, 96, 86, 95, 98, 98, 96]$), so that the multiscale methodology is able to successfully discard the data realizations as credible prior predictions.

3.1.4 Stationary versus non-stationary means

D is drawn from $\mathcal{N}(\mu(i), \Sigma_d)$ with a cubical variogram with $r = 25$. The non-stationary mean is given by $\mu(i) := \frac{A}{N-1}(2i - (N-1))$, where i is an integer in $[0, N-1]$.

Let $s = 0.01$. Figure 5(a) shows the prior predictions and an arbitrarily selected data realization when $A = 1.5$. Figure 5(c) shows corresponding results obtained with the Mahalanobis methodology, while Fig. 5(e) shows corresponding results obtained with the multiscale methodology. It is seems reasonable to state that both methodologies would correctly discard the data realizations as credible prior predictions. With the Mahalanobis methodology, $\lambda_{\text{md}} = 12$, while $\xi_1(D)$ is clearly outside, and $\xi_2(D)$ is slightly outside, the corresponding B_l's ($\lambda_{\text{ml}} = [100, 0, 0, 100, 100, 100, 100, 100, 100, 100]$).

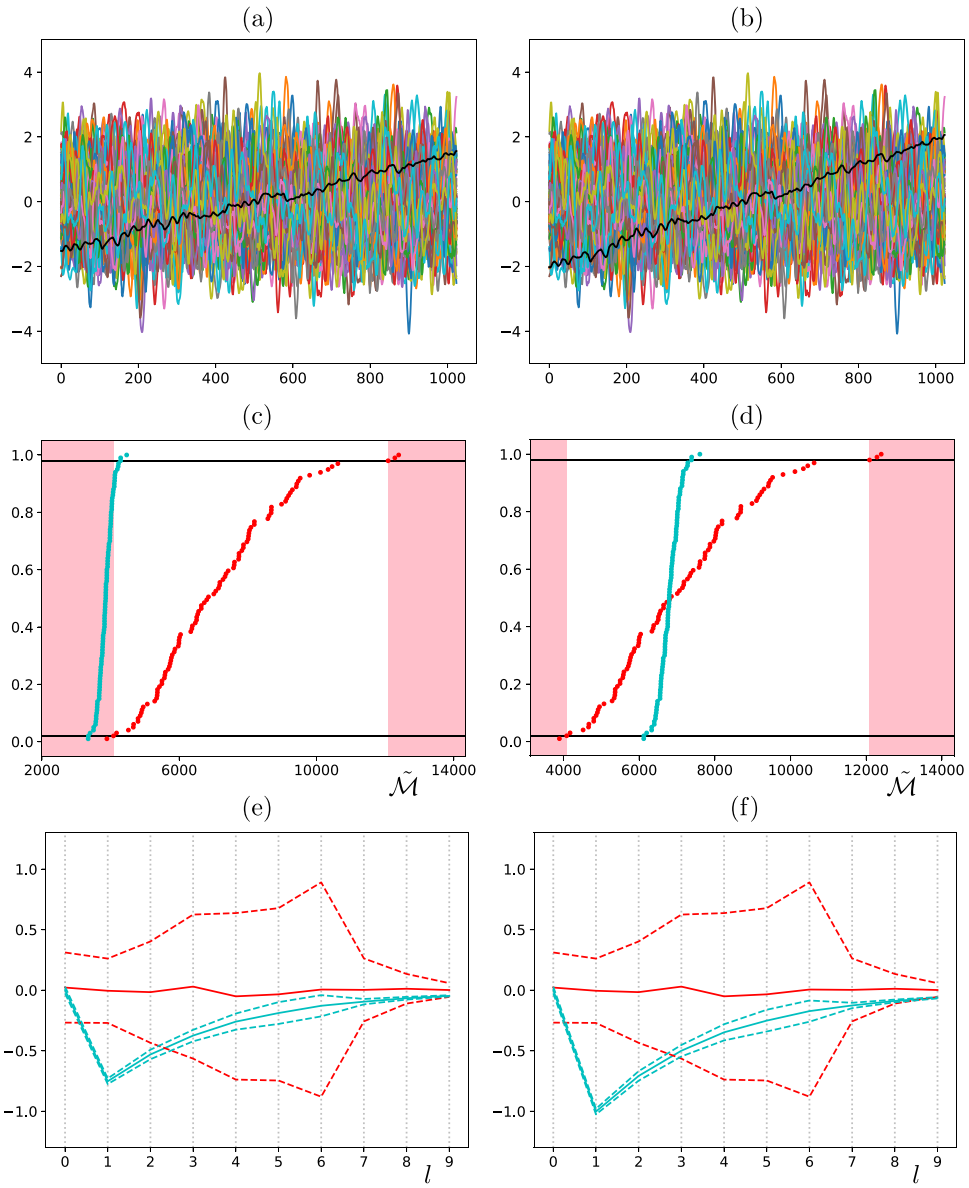

Fig. 5 Left column: $A = 1.5$. Right column: $A = 2$. **(a, b)**: Prior predictions (colored) and arbitrarily selected data realization (black). **(c, d)**: CDF_W (red) and CDF_D (cyan). **(e, f)**: $\xi_l(W)$ (red solid), $\xi_l(D)$ (cyan solid), $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed), and $\xi_l(D) \pm 2\zeta_l(D)$ (cyan dashed)

Figure 5(b) shows the prior predictions and an arbitrarily selected data realization when $s = 0.01$ and $A = 2$. Figure 5(d) shows corresponding results obtained with the Mahalanobis methodology, while Fig. 5(f) shows results obtained with the multiscale methodology. While the multiscale methodology, as expected, shows even bigger differences between the $\xi_l(D)$'s and the corresponding $\xi_l(W)$'s ($\lambda_{\text{ml}} = [100, 0, 0, 100, 100, 100, 100, 100, 100, 0]$), the Mahalanobis methodology now surprisingly accepts all data realizations as credible prior predictions. Changing the initialization of the random number generator applied to generate the realizations had no significant effect on the results of the multiscale methodology, while it made λ_{md} change from 100 to 0. This behavior of the Mahalanobis methodology could be due to its reliance of use of a regularized covariance based on a particular set of prior predictions.

Different variances

I will now consider two different values for s than $s = 0.01$. Figure 6(a) shows the prior predictions and an arbitrarily selected data realization when $s = 0.1$ and $A = 2$. Figure 6(c) shows corresponding results obtained with the Mahalanobis methodology, while Fig. 6(e) shows results obtained with the multiscale methodology. Both the Mahalanobis methodology and the multiscale methodology show similar behavior as when $s = 0.01$ (Fig. 5(d, f)) with respect to discarding the data realizations as credible prior predictions. The Mahalanobis methodology accepts

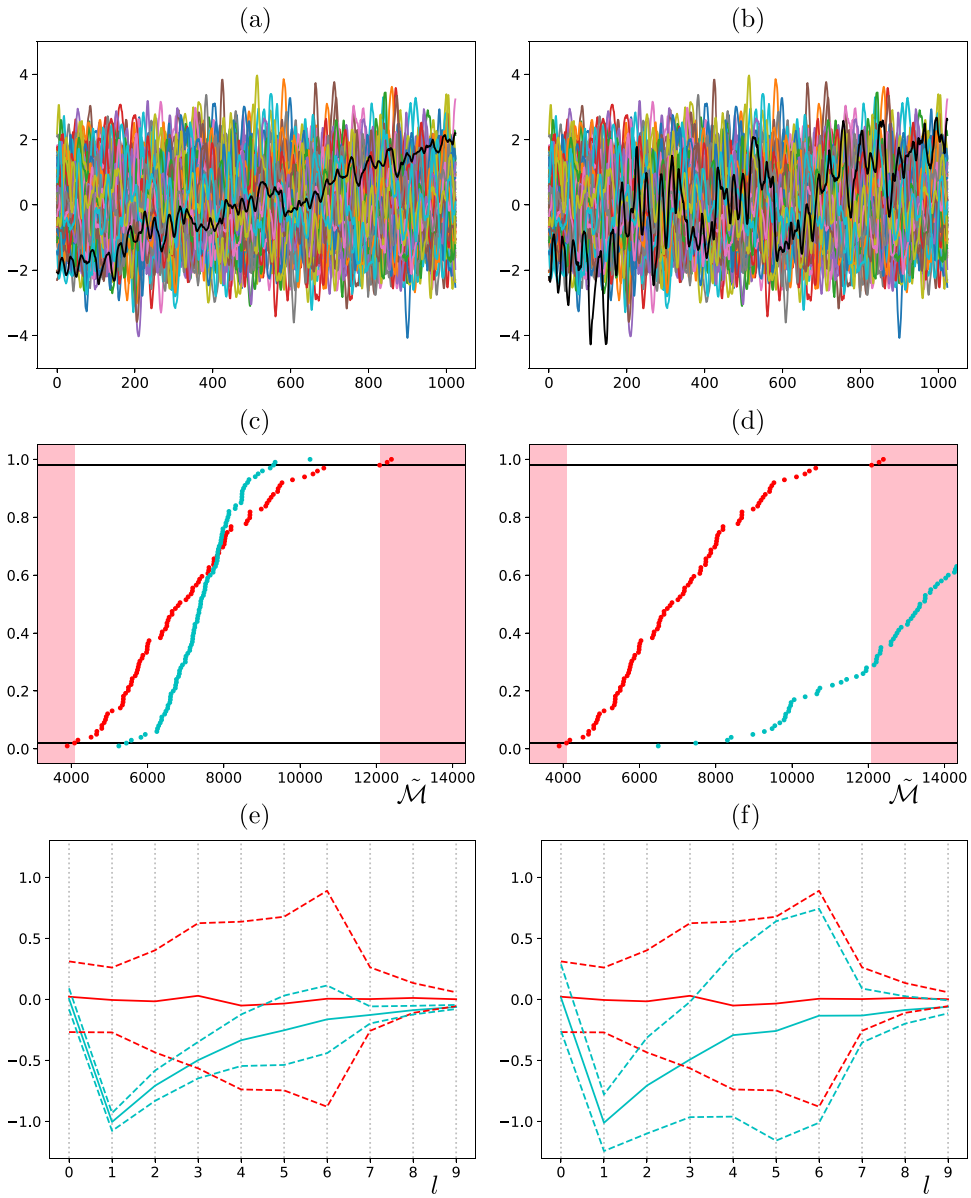

Fig. 6 Left column: $s = 0.1$. Right column: $s = 1$. **(a, b)**: Prior predictions (colored) and arbitrarily selected data realization (black). **(c, d)**: CDF_W (red) and CDF_D (cyan). **(e, f)**: $\xi_l(W)$ (red solid), $\xi_l(D)$ (cyan solid), $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed), and $\xi_l(D) \pm 2\zeta_l(D)$ (cyan dashed)

all data realizations as credible prior predictions ($\lambda_{\text{md}} = 100$), while $\xi_l(D)$ is clearly outside B_l for $l \in [1, 2]$ ($\lambda_{\text{ml}} = [100, 0, 0, 79, 100, 100, 100, 100, 84, 19]$), which would correctly make us discard the data realizations as credible prior predictions when applying the multiscale methodology.

Figure 6(b) shows the prior predictions and an arbitrarily selected data realization when $s = 1$ and $A = 2$. Figure 6(d) shows corresponding results obtained with the Mahalanobis methodology, while Fig. 6(f) shows results obtained with the multiscale methodology. Now, the Mahalanobis methodology would probably make us discard the data realizations as credible prior predictions ($\lambda_{\text{md}} = 28$). As when $s = 0.01$ (Fig. 5(f) and when $s = 0.1$ Fig. 6(e)), the multiscale methodology would correctly make us discard the data realizations as credible prior predictions ($\lambda_{\text{ml}} = [96, 0, 9, 60, 88, 82, 92, 86, 69, 40]$).

At the outset, prior predictions and data realizations may vary on several length scales. The multiscale methodology provides scale-by-scale discrepancy measures. The Mahalanobis methodology, on the other hand, lumps all scales into a single discrepancy measure, but certain scales may dominate this measure. The results in this section indicate that small-scale variations dominated the Mahalanobis methodology measure. The Mahalanobis methodology was not able to consistently discriminate well between prior predictions with a stationary zero mean and data realizations with a slowly varying non-stationary mean, while small-scale changes influenced the measure significantly (confer the results for $s = 0.01$ and $s = 0.1$ versus those for $s = 1$).

3.1.5 Different variogram types

D is drawn from $\mathcal{N}(o, \Sigma_d)$ with $s = 1$ and $r = 25$, but with other data-variogram types than cubic.

Figure 7(a, b) show the prior predictions and an arbitrarily selected data realization when the data-variogram type is spherical and exponential, respectively. Figure 7(c, d) show corresponding results obtained with the Mahalanobis methodology. For the spherical variogram, it is seen that the Mahalanobis methodology would accept most ($\lambda_{\text{md}} = 84$) of the data realizations as credible prior predictions. For the exponential variogram, however, the Mahalanobis methodology would not accept any ($\lambda_{\text{md}} = 0$) of the data realizations as credible prior predictions. The fact that there are virtually no coarse-scale differences between prior predictions and data realizations is not reflected in the results of the Mahalanobis methodology. These results supports earlier result in Section 3 in that the Mahalanobis methodology is very sensitive to differences in small-scale variations in the prior predictions and the data realizations. This becomes even more apparent when comparing with the results for a cubic data variogram (Fig. 2(c)) (confer also the last paragraph in Section 3.1.4).

Figure 7(e, f) show corresponding results obtained with the multiscale methodology. There are some minor differences in the behavior of $\zeta_l(D)$ for the finer scales ($l \in [7, 9]$) between the two variogram types ($\lambda_{\text{ml}} = [96, 98, 99, 100, 95, 89, 98, 71, 55, 44]$ (spherical) and $\lambda_{\text{ml}} = [98, 99, 99, 100, 97, 95, 100, 60, 43, 26]$ (exponential)). For the coarser scales there are no differences that would not disappear if E had

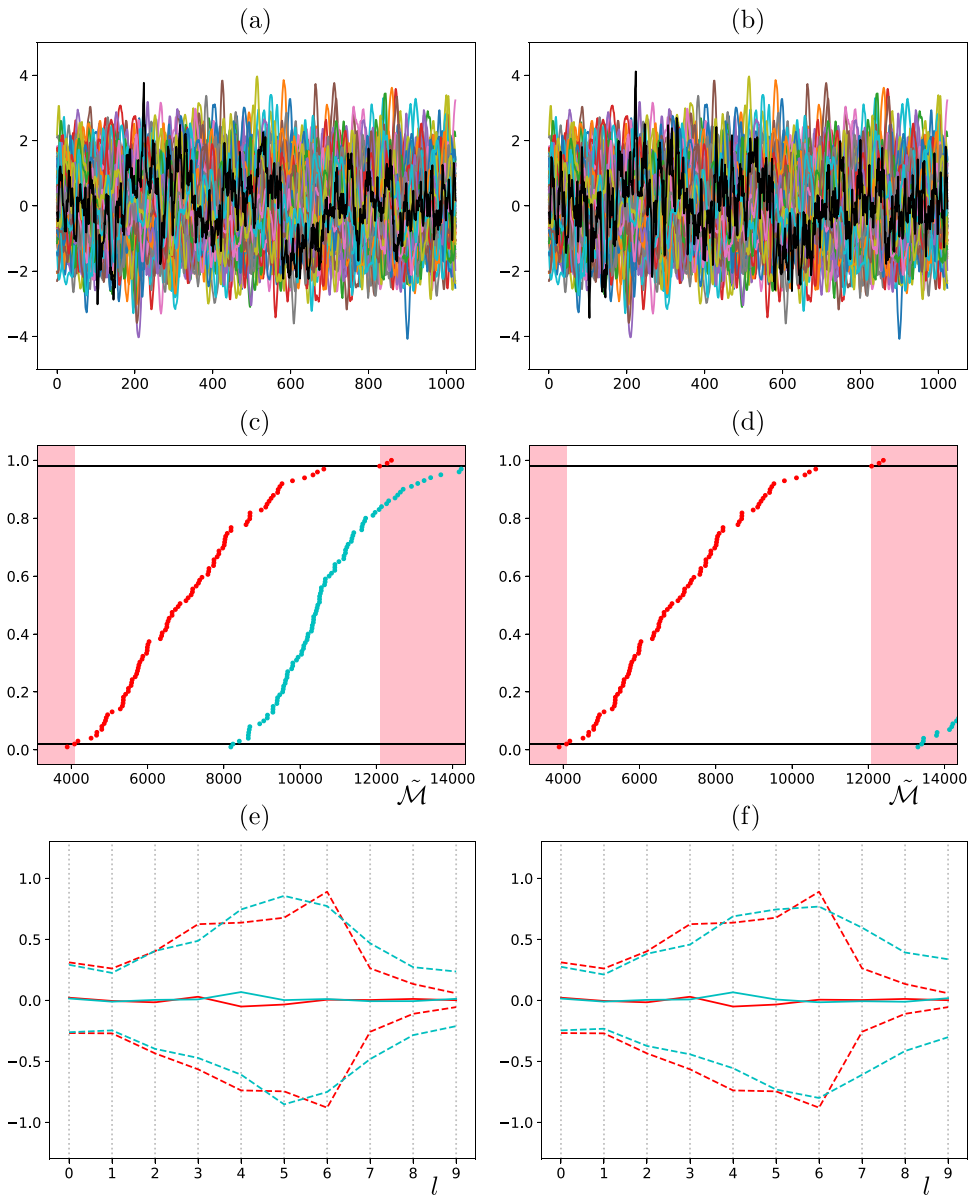

Fig. 7 Left column: spherical variogram. Right column: exponential variogram. **(a, b)**: Prior predictions (colored) and arbitrarily selected data realization (black). **(c, d)**: CDF_W (red) and CDF_D (cyan). **(e, f)**: $\xi_l(W)$ (red solid), $\xi_l(D)$ (cyan solid), $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed), and $\xi_l(D) \pm 2\zeta_l(D)$ (cyan dashed)

been increased sufficiently. This holds also when comparing with the results obtained for a cubic data variogram (Figure 2d).

3.2 Reservoir models

The purpose with the reservoir models considered is to provide test cases where the data realizations reflect a large-scale subsurface feature that is not known to the modeler when generating the prior predictions. A 64×64 grid is applied for both reservoir models. A production well is located in the cell with indices $(i, j) = (31, 63)$ and a water injector is located in the cell with indices $(31, 0)$. A constant porosity of 0.2 is applied. The logarithms of the permeability fields applied to generate the prior predictions are drawn from $\mathcal{N}(4u, \Sigma_d)$, and an isotropic, spherical variogram with sill s and range r is applied. The permeability fields applied to generate the data realizations are made by perturbing cell values of the permeability field underlying the prior predictions.

The reservoir models will produce results that are not dimensionless. To be able to present results obtained with the multiscale methodology in a concise manner (i.e., results for all l in a single plot) with good visibility, these results will be normalized separately for each l. The normalization is performed by dividing $\xi_l(W)$, $\xi_l(D)$, $\xi_l(W) \pm 2\zeta_l(W)$ and $\xi_l(D) \pm 2\zeta_l(D)$ by the maximum absolute value of $\xi_l(W) \pm 2\zeta_l(W)$ and $\xi_l(D) \pm 2\zeta_l(D)$, so that all results will be contained in $[-1, 1]$.

3.2.1 Channel

Here, $s = 0.2$ and $r = 10$. Figure 8(a) shows the logarithm of the permeability field applied to generate the data realizations. That field is generated from the field applied to generate the prior predictions by multiplying the permeability values in the cells in the region, Ω_c, defined by $i \in [30, 32] \times j \in [4, 59]$, by 10, and the permeability values in the cells outside Ω_c by $1/5$. Hence, the permeability field applied to generate the data realizations has lower values than the field applied to generate the prior predictions, except in the high-permeable channel, Ω_c.

An error, $\epsilon_c \sim \mathcal{N}(o, \sigma_c^2 I,)$, where $\sigma_c := 0.03 \cdot \overline{W_c \cup D_c}$, is added to the prior predictions and the data realizations. Figure 8(b) shows the resulting prior predictions and data realizations for the oil production, ω_o. Admittedly, simpler methods, like lack of coverage of the data by the prior predictions, would in this case have been sufficient to discard the data realizations as credible prior predictions, but it is still of interest to assess the performances of the Mahalanobis methodology and the multiscale methodology on this example.

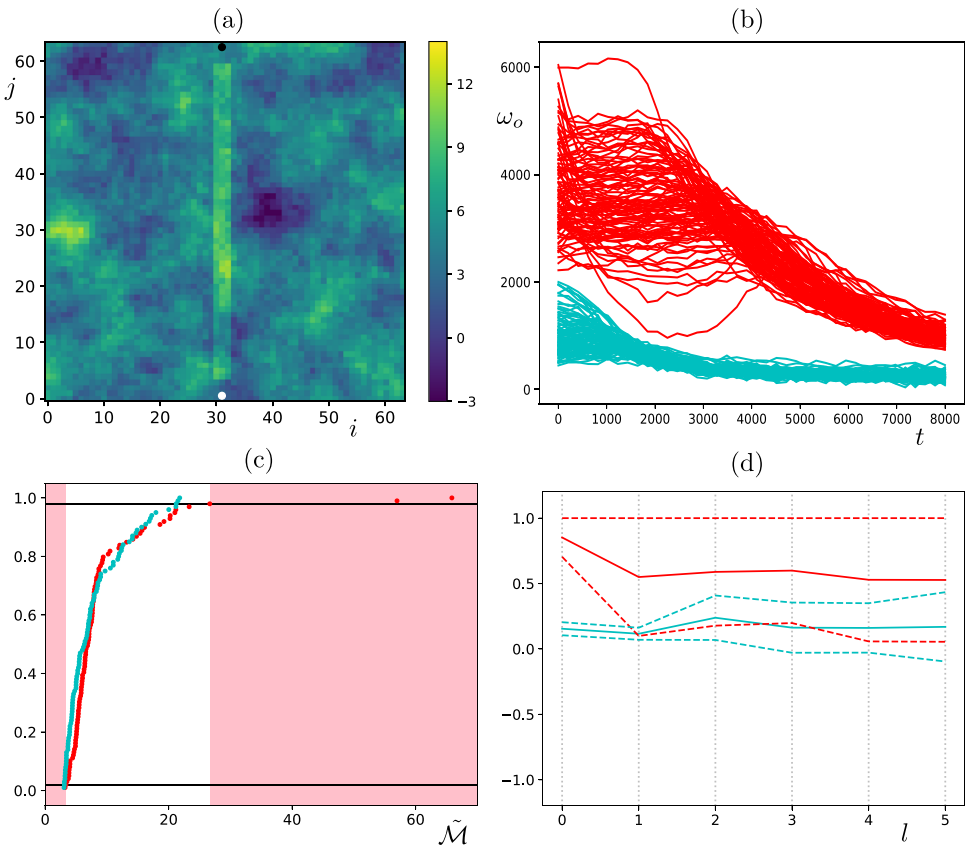

Fig. 8 **(a)**: Logarithm of permeability field (permeability in mDarcy) used to generate the data realizations with injector (white dot) and producer (black dot). **(b)**: Prior predictions (red) and data realizations (cyan) for the oil production, ω_o (m^3/day), vs. time, t (days). **(c)**: CDF_W (red) and CDF_D (cyan). **(d)**: $\xi_l(W)$ (red solid), $\xi_l(D)$ (cyan solid), $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed), and $\xi_l(D) \pm 2\zeta_l(D)$ (cyan dashed). The results obtained with the multiscale methodology have been normalized

Figure 8(c) shows the results obtained with the Mahalanobis methodology. CDF_D is very similar to CDF_W, and almost all data realizations are accepted as credible prior predictions ($\lambda_{\text{md}} = 92$).

Figure 8(d) shows the results obtained with the multiscale methodology. $\xi_0(D)$ is clearly outside B_0 (and $\xi_3(D)$ is slightly outside B_3), so that the multiscale methodology would correctly make us discard the data realizations as credible prior predictions ($\lambda_{\text{ml}} = [0, 79, 77, 29, 88, 84]$).

3.2.2 Barrier

Here, $r = 10$. Figure 8(a) shows the logarithm of the permeability field applied to generate the data realizations. That field is generated from the field applied to generate the prior predictions by multiplying the permeability values in the cells in the region, Ω_b, defined by $i \in [0, 62] \times j = 31$ by $1 \cdot 10^{-3}$. Hence, the permeability field applied to generate the data realizations has the same values as the field applied to generate the prior predictions except in the low-permeable barrier, Ω_b.

An error, $\epsilon_b \sim \mathcal{N}(o, \sigma_b^2 I,)$, where $\sigma_b := 0.05 \cdot \overline{W_b \cup D_b}$, is added to the prior predictions and the data realizations. Figure 9(b) shows the resulting prior predictions and data realizations for the water production, ω_w.

Figure 9(c) shows the results obtained with the Mahalanobis methodology. About half of the data realizations are accepted as credible prior predictions ($\lambda_{\text{md}} = 53$).

Figure 9(d) shows the results obtained with the multiscale methodology. For scales 1 and 2, $\xi_l(D)$ is clearly outside B_l, and only $\xi_5(D)$ is inside its corresponding interval, so the multiscale methodology would correctly discard the data realizations as credible prior predictions ($\lambda_{\text{ml}} = [18, 16, 15, 32, 29, 62]$).

3.3 Norne G-segment

In [3], application of the Mahalanobis methodology to the G-segment of the Norne field (Fig. 10) was the major topic, and the results obtained for the initial reservoir model description, developed in [6], were utilized to improve that description in several ways through manual adjustments. It was shown that these adjustments removed apparent inconsistencies between observed data and prior predictions based on the initial reservoir description.

In the current paper, the Norne field does not play the same dominant role. I will apply the Mahalanobis methodology and the multiscale methodology to data and prior predictions based on the initial reservoir model description, only, and I will not attempt to improve that description. The purpose of the investigation is to assess the performance of the

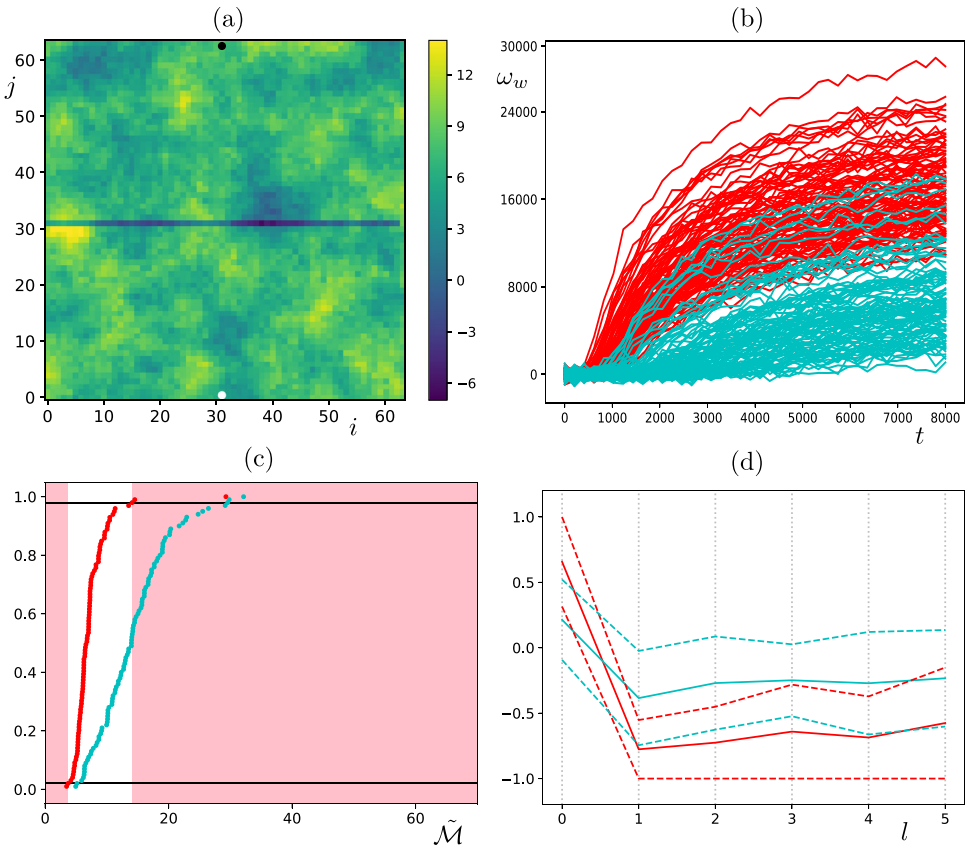

Fig. 9 **(a)**: Logarithm of permeability field (permeability in mDarcy) used to generate the data realizations with injector (white dot) and producer (black dot). **(b)**: Prior predictions (red) and data realizations (cyan) for the oil production, ω_o (m^3/day), vs. time, t (days). **(c)**: CDF_W (red) and CDF_D (cyan). **(d)**: $\xi_l(W)$ (red solid), $\xi_l(D)$ (cyan solid), $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed), and $\xi_l(D) \pm 2\zeta_l(D)$ (cyan dashed). The results obtained with the multiscale methodology have been normalized

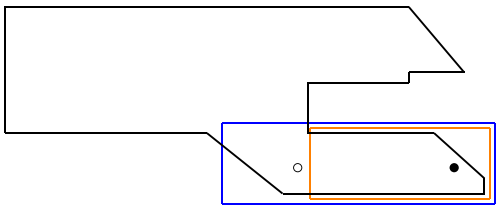

Fig. 10 Schematic of horizontal slice of the boundaries for the simulation model for the Norne field (black). Region inside blue rectangle: G-segment. Region inside orange rectangle: Study region for time-lapse seismics. Producer E4 (•) and water injector F4 (○)

multiscale methodology, and to compare it to that of the Mahalanobis methodology.

The Norne oil-and-gas field is located in the Norwegian Sea, approximately 200 km from the coast of Nordland county, Norway. It consists of the formations (from top to bottom) Garn, Not (shale), Ile, Tofte and Tilje.

Data from production start in 1997 to 2006 (production rates, RFT pressures, time-lapse seismics), a reservoir simulation model, and geological reports, were released by Equinor and partners in 2010. The release has since been used to test various history-matching methods, see, e.g., [6, 11, 14, 19, 20, 22, 25, 29]. We refer to these publications for a more thorough description of the Norne field than what is presented here.

Data-error models will be the same as those applied in [3], that is, $\epsilon_{\text{pro}} \sim \mathcal{N}(0, \sigma_{\text{pro}}^2 I)$, where $\sigma_{\text{pro}} = 50$ m^3/day, for production rates, and $\epsilon_{\text{rft}} \sim \mathcal{N}(0, \sigma_{\text{rft}}^2 I)$, where $\sigma_{\text{rft}} = 0.01$ bar, for RFT pressures. Time-lapse acoustic impedance data-errors were estimated using factorial co-Kriging, result-

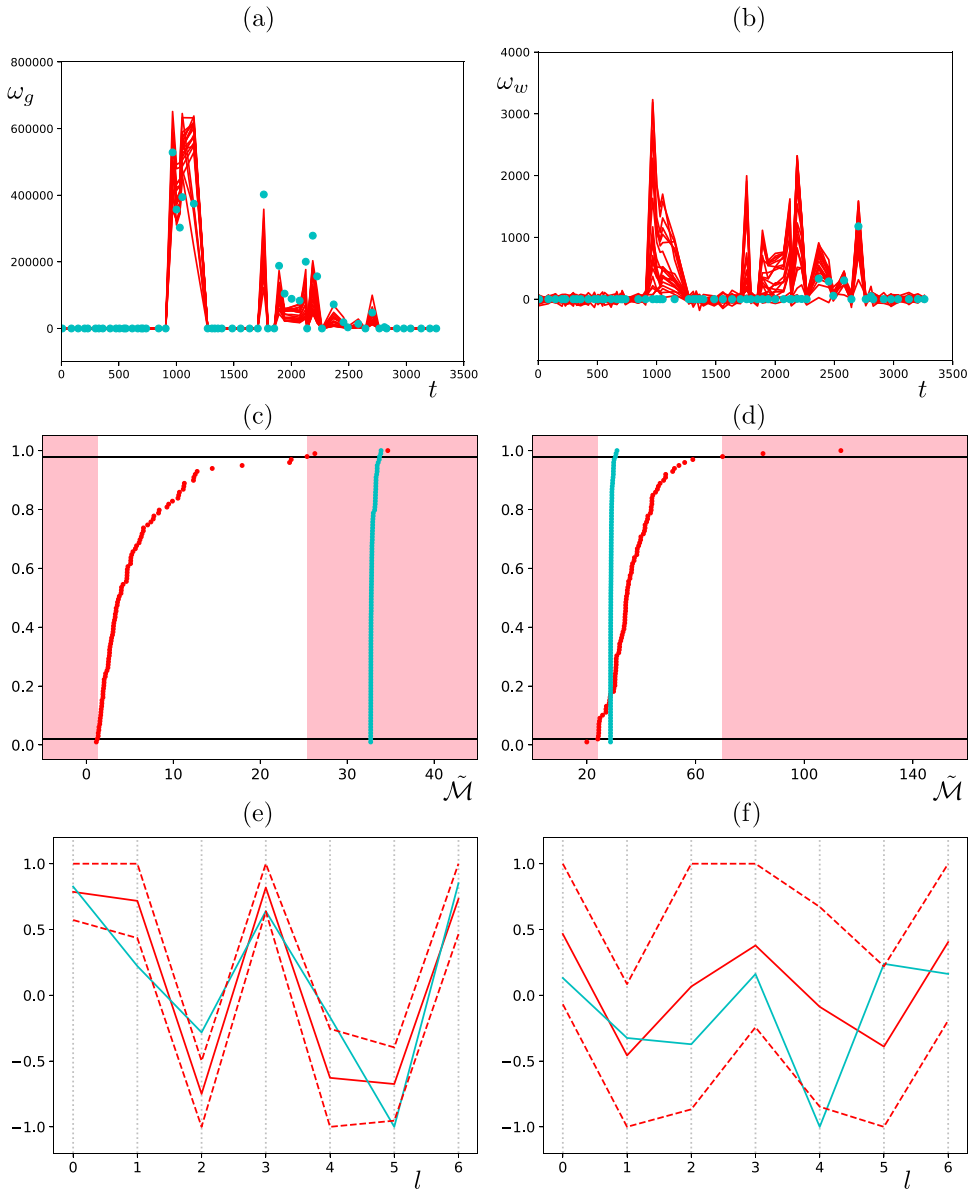

Fig. 11 **(a)** Gas production, ω_g (m^3/day), in well E4 vs. time, t (days). Prior predictions (red solid) and observed data (cyan dots). **(c)** Corresponding results obtained with the Mahalanobis methodology. CDF_W (red) and CDF_d (cyan). **(e)** Corresponding results obtained with the multiscale methodology. $\xi_l(W)$ (red solid), $\xi_l(d)$ (cyan solid), and $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed). **(b)** Water production, ω_w (m^3/day), in well E4 vs. time, t (days). Prior predictions (red solid) and observed data (cyan dots). **(d)** Corresponding results obtained with the Mahalanobis methodology. CDF_W (red) and CDF_d (cyan). **(f)** Corresponding results obtained with the multiscale methodology. $\xi_l(W)$ (red solid), $\xi_l(d)$ (cyan solid), and $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed). The results obtained with the multiscale methodology have been normalized

ing in $\epsilon_{\text{imp}} \sim \mathcal{N}(0, \Sigma_{\text{imp}})$, where Σ_{imp} is based on a cubic variogram with nugget amplitude 1.236 $(\text{m/s} \cdot \text{g/cc})^2$, sill 27.260 $(\text{m/s} \cdot \text{g/cc})^2$, and ranges of 10 and 9 grid cells in the i and j directions, respectively [3, 4].

The Norne model will produce results that are not dimensionless. To be able to present results obtained with the multiscale methodology in a concise manner with good visibility, these results will be normalized separately for each l. The normalization is performed by dividing $\xi_l(W)$, $\xi_l(d)$ and $\xi_l(W) \pm 2\zeta_l(W)$, by the maximum absolute value of $\xi_l(W) \pm 2\zeta_l(W)$, and $\xi_l(d)$, so that all results will be contained in $[-1, 1]$.

3.3.1 Production data

Figure 11(a) shows 20 of the prior predictions (100 were used in the analysis) and the data for the gas production in well E4. Figure 11(c) shows corresponding results obtained with the Mahalanobis methodology, while Fig. 11(e) shows results obtained with the multiscale methodology. The Mahalanobis methodology discards the data as a credible prior prediction. $\xi_1(d)$ and $\xi_2(d)$ are clearly outside the intervals B_1 and B_2, respectively, while $\xi_4(d)$ and $\xi_5(d)$ are marginally outside the intervals B_4 and B_5, respectively. Therefore, also the multiscale methodology discards the data as a credible prior prediction.

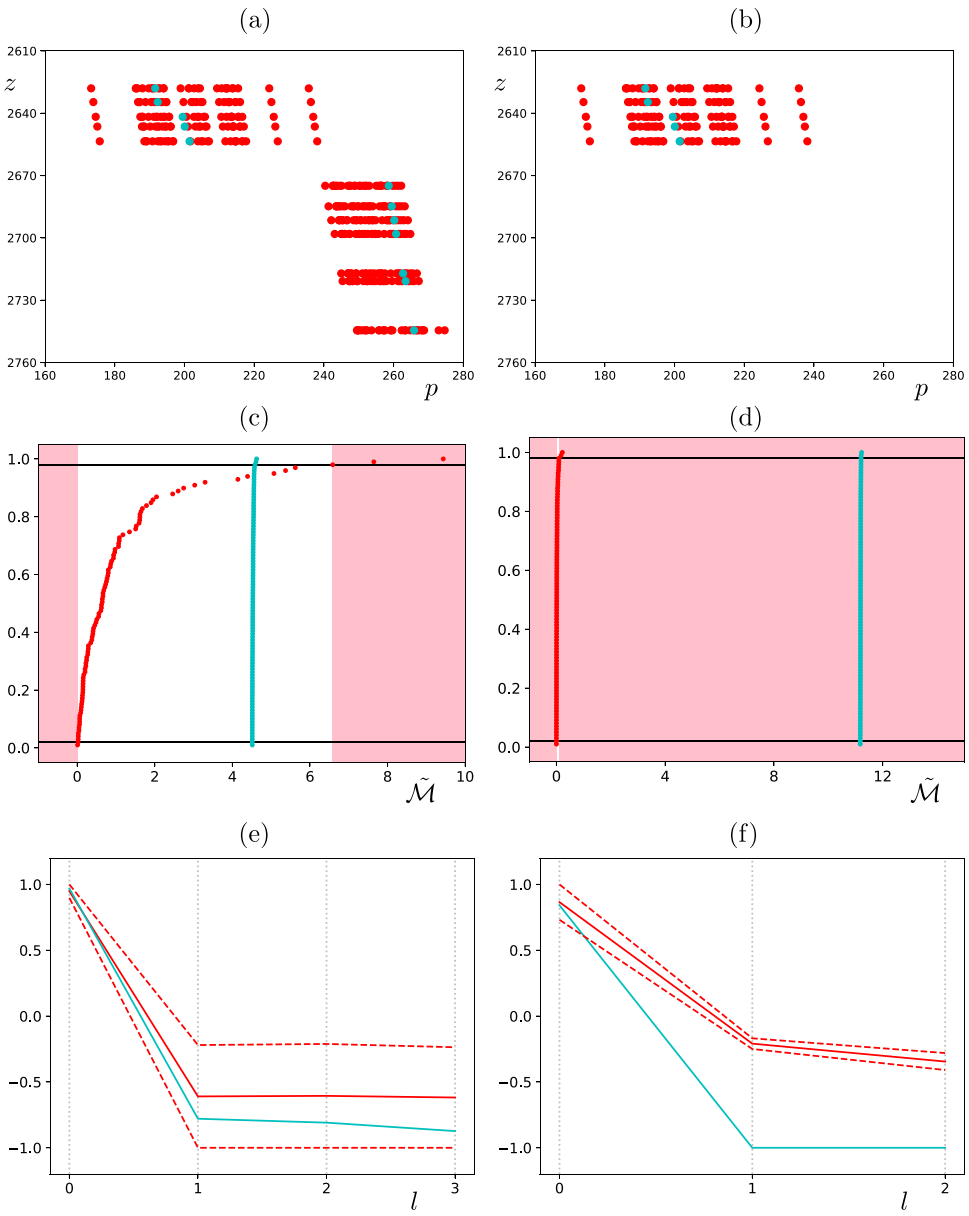

Fig. 12 **(a)** Pressure, p (bar), vs. depth (m) in well F4. Prior predictions (red dots) and data (cyan dots). **(c)** Corresponding results obtained with the Mahalanobis methodology. CDF_W (red) and CDF_d (cyan). **(e)** Corresponding results obtained with the multiscale methodology. $\xi_l(W)$ (red solid), $\xi_l(d)$ (cyan solid), and $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed). **(b)** Prior predictions (red dots) and data (cyan dots) in the Garn formation. **(d)** Corresponding results obtained with the Mahalanobis methodology. CDF_W (red) and CDF_d (cyan). **(f)** Corresponding results obtained with the multiscale methodology. $\xi_l(W)$ (red solid), $\xi_l(d)$ (cyan solid), and $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed). The results obtained with the multiscale methodology have been normalized

With real data there is, of course, only a single data vector. This means that the elements of λ_{ml} will take values either 0 or 100, while CDF_d will be extremely narrow, so that λ_{md} becomes either very close to 0 or very close to 100. There is therefore no extra information conveyed by reporting the values of λ_{md} and λ_{ml}, and I will refrain from doing so in Section 3.3.

Figure 11(b) shows 20 of the prior predictions and the data for the water production in well E4. Figure 11(d) shows corresponding results obtained with the Mahalanobis methodology, while Fig. 11(f) shows results obtained with the multiscale methodology. The Mahalanobis methodology accepts the data as a credible prior prediction. The results obtained with the multiscale methodology does not lead to an obvious conclusion, but one would perhaps tend to accept the data as a credible prior prediction, although $\xi_4(d)$ is marginally outside the interval B_4 and $\xi_5(d)$ is at the upper boundary of B_5.

3.3.2 RFT data

Figure 12(a) shows 30 of the prior predictions (100 were used in the analysis) and the data for the RFT pressures in well F4. Note the jump in pressure values from the Garn formation above the Not shale (above approximately 2650 m) to the formations below the Not shale (below approximately 2670 m). Note also the smaller jump in pressure values within the Garn formation (at approximately 2640 m) seen only in the data. This feature was central in [3] when seeking to improve the initial reservoir description.

Figure 12(c) shows corresponding results obtained with the Mahalanobis methodology, while Fig. 12(e) shows results obtained with the multiscale methodology. Both the Mahalanobis methodology and the multiscale methodology accept the data as a credible prior prediction

Figure 12(b) shows 30 of the prior predictions and the data for the RFT pressures in well F4 within the Garn formation. Figure 12(d) shows corresponding results obtained with the Mahalanobis methodology, while Fig. 12(f) shows results obtained with the multiscale methodology. The Mahalanobis methodology discards the data as a credible prior prediction, while $\xi_l(d)$ is clearly outside B_l for $l \in [1, 2]$, which would make us discard the data as a credible prior prediction also when applying the multiscale methodology.

3.3.3 Time-lapse impedance data

The Norne simulation model represents the Garn formation with three grid-cell layers in the vertical direction. I will consider prior predictions and data for the acoustic impedance change in Layer 2 from year 2001 to year 2006.

The maximum number of grid cells in Layer 2 in the i and j directions are 36 and 12, respectively, and most of the 36 j-columns contain 12 cells. I therefore let H consist of two-dimensional test vectors of size 36×12. Since Layer 2 is not rectangular, I extend it to a rectangle with dummy cells to facilitate multiplication with the rectangular test vectors in H. The acoustic impedance change in each dummy cell is set to zero to ensure that these cells will not influence the results obtained with the multiscale methodology. Four of the five vectors in H are illustrated in Fig. 13. The fifth vector, h_0, takes the value 1 in each grid cell.

Figure 14(a) shows a scatter plot of the time-lapse acoustic impedance change in 25 of the prior predictions, Z_W (100 were used in the analysis), versus that of the data, Z_d, for all grid cells in the study region. In all grid cells, the majority of the values of the prior predictions are significantly lower than the corresponding data values, but in most of the cells, there are also several prior predictions with higher values than the corresponding data value. Figure 14(b) shows the data in the study region.

Figure 14(c) shows corresponding results obtained with the Mahalanobis methodology, which accepts the data as a credible prior prediction. Figure 14(d) shows results obtained with the multiscale methodology. Also the mul-

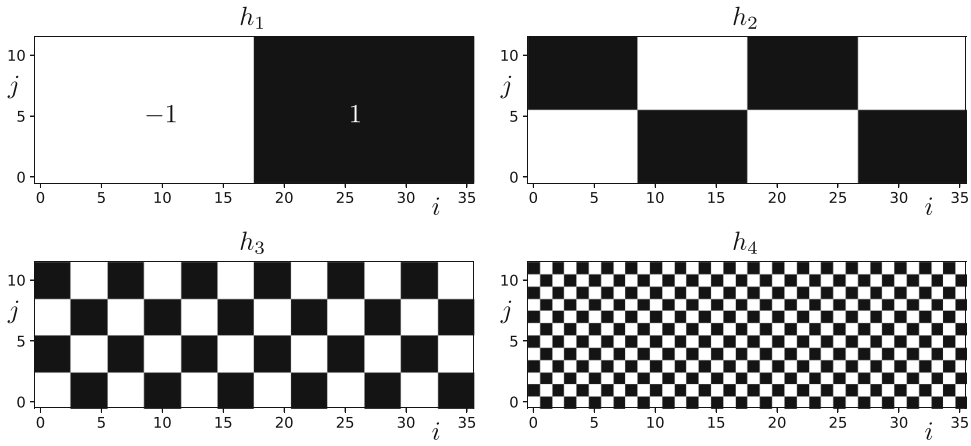

Fig. 13 Test vectors $\{h_l\}_{l=1}^4$

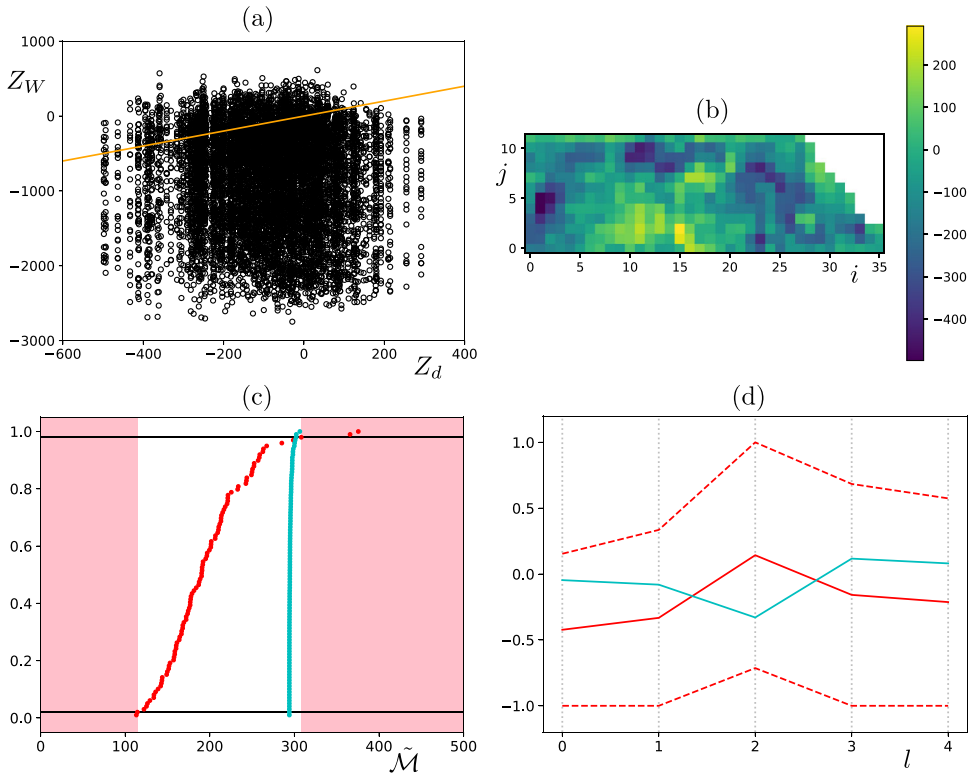

Fig. 14 **(a)** Time-lapse acoustic impedance change in the prior predictions, Z_W (m/s · g/cc), vs. that of the data, Z_d (m/s · g/cc). Orange line indicates equal values for Z_w and Z_d. **(b)** Z_d in study area. **(c)** CDF_W (red) and CDF_d (cyan). **(d)** $\xi_l(W)$ (red solid), $\xi_l(d)$ (cyan solid), and $\xi_l(W) \pm 2\zeta_l(W)$ (red dashed). The results obtained with the multiscale methodology have been normalized

tiscale methodology accepts the data as a credible prior prediction.

3.4 Discussion

The definition of $\{B_l\}_{l=0}^L$ is important for the outcome of the multiscale methodology. The current definition is convenient for the comparison with the Mahalanobis methodology (see, last paragraph in Section 2.1.2). It is, of course, possible to apply the multiscale methodology with an alternative definition of $\{B_l\}_{l=0}^L$ if one believes that would be more suitable for the problem at hand.

Furthermore, the criteria (see, second paragraph in Section 2.2) for when to accept and when to discard d as a credible prior prediction (or similarly, when to accept and when to discard D as credible prior predictions), given the definition of $\{B_l\}_{l=0}^L$, are not completely precise. The situation regarding lack of precision is similar for the Mahalanobis methodology (see, the last paragraph in Section 2.1.2). It would, of course, be possible to be more precise, but I think it is better to leave this issue to the judgment of potential users of the methodology on their particular problems.

Finally, there are numerous options for the multiscale test vectors in H, and I do not claim that my choice is optimal. On the contrary, it is the multiscale feature of H that is important, rather than the particular definition of its members.

4 Summary

The problem of criticizing a model prior to history matching by comparing data to prior predictions from the model was considered. A multiscale methodology, that can be applied to analyze differences between data and prior predictions in a scale-by-scale fashion, was proposed. The methodology is computationally inexpensive, straightforward to apply, and can handle correlated observation errors without making any approximations.

The multiscale methodology was tested on a set of toy models, on two simplistic reservoir models with synthetic data, and on (production, RFT, and time-lapse acoustic impedance) data and prior predictions from the G-segment of the Norne field. The tests included comparisons with a previously published method [3], denoted the Mahalanobis methodology in this paper. For the toy models and the reservoir models, it was also possible to test the robustness of the methodologies towards effects caused by a particular data vector.

The multiscale methodology and the Mahalanobis methodology behaved similarly for the Norne case. Both methodologies led to the same decisions regarding whether to accept or discard the data as a credible prior prediction for the different data types. This is a desirable feature of the multiscale methodology, since the Mahalanobis methodology

induced changes in the Norne reservoir model that led to a better data match [3]. The multiscale methodology led to correct decisions for the toy models and the reservoir models. For these models, the Mahalanobis methodology either led to incorrect decisions, and/or was unstable with respect to ensemble selection. These problems seem to be related to the need for regularization of the covariance matrix entering the Mahalanobis distance, and to a too large sensitivity towards small-scale differences between data and prior predictions, while important coarse-scale differences were not disclosed. There is no need for a covariance matrix or any regularization with the multiscale methodology, and this may be the reason for it's superiority.

Acknowledgements The author thanks Equinor (operator of the Norne field) and its license partners Eni Norge (now, Vår Energi) and Petoro for the release of the Norne data. The author acknowledges the (former) Center for Integrated Operations at NTNU for cooperation and coordination of the Norne cases. The views expressed in this paper are the views of the author and do not necessarily reflect the views of Equinor and the Norne license partners.

Funding information Open access funding provided by NORCE Norwegian Research Centre AS. Partial financial support was provided by "NCS2030 - National Centre for Sustainable Subsurface Utilization of the Norwegian Continental Shelf", which is funded by industry partners Aker BP, DNO Norge, Equinor, Landmark Graphics, OKEA, Schlumberger Information Solutions, Vår Energi, and Wintershall DEA, as well as the Research Council of Norway (PETROMAKS2)

Data Availability Python code that will reproduce the results with the Toy models (Figs. 2-7 in the paper) is on GitHub:https://github.com/TrondMannseth/MMD. The other datasets generated during and/or analysed during the current study are available from the author on reasonable request.

Declarations

Ethical standards The manuscript is written in full compliance with the ethical standards of the Journal.

Competing interests The author has no competing interests to declare that are relevant to the content of this article.

References

1. Aanonsen, S.I., Fossum, K., Mannseth, T.: Bayesian model evaluation for multiple scenarios. Comput. Geosci. 27(6), 1001–1021 (2023)
2. Aanonsen, S.I., Naevdal, G., Oliver, D.S., Reynolds, A.C., Brice, V.: The ensemble Kalman filter in reservoir engineering-a review. 14(3):393–412 (2009)
3. Alfonzo, M., Oliver, D.S.: Evaluating prior predictions of production and seismic data. Comput. Geosci. 23(6), 1331–1347 (2019)
4. Alfonzo, M., Oliver, D.S.: Seismic data assimilation with an imperfect model. Comput. Geosci. 24(2), 889–905 (2020)
5. Chen, Y., Oliver, D.S: Ensemble-based closed-loop optimization applied to Brugge field. SPE Reserv. Eval. Engin. 13(1), 56–71 (2010)
6. Yan, C., Dean, S.O.: History matching of the Norne full-field model with an iterative ensemble smoother. SPE Reserv. Eval. Engin. **17**(2), 244–256 (2014)
7. Daubechies, I.: Ten lectures on wavelets. SIAM, (1992)
8. Emerick, A.A., Reynolds, A.C.: History matching a field case using the ensemble Kalman filter with covariance localization. SPE Reserv. Eval. Engin. **14**(4), 443–452 (2011)
9. Evensen, G., Vossepoel, F.C., Leeuwen, P.J.V.: Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem. Springer Nature, (2022)
10. Fang, J., Gong, B., Caers, J.: Data-driven model falsification and uncertainty quantification for fractured reservoirs. Engineering **18**, 116–128 (2022)
11. Ferreira, C.J., Alessandra, D.J., Schiozer, D.J.: Use of a probabilistic and multi-objective history matching for uncertainty reduction for the Norne benchmark case. In SPE Europec featured at EAGE Conference and Exhibition, page D121S011R008. SPE, (2017)
12. Haar, A.: Zur theorie der orthogonalen funktionensysteme. Math. Ann. **69**(3), 331–371 (1910)
13. He, J., Tanaka, S., Wen, X., Kamath, J.: Rapid S-curve update using ensemble variance analysis with model validation. In SPE Western Regional Meeting, OnePetro (2017)
14. Hermans, T., Nguyen, F., Caers, J.: Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data: Workflow and case study. Water Resour. Res. **51**(7), 5332–5352 (2015)
15. Hoeting, J.A, Madigan, D., Raftery, A.E., Volinsky, C.T.,:Bayesian model averaging: a tutorial. Stat. Sci.14 (4):382–417, (1999)
16. Hubert, M., Debruyne, M.: Minimum covariance determinant. Wiley interdisciplinary reviews: Computational statistics **2**(1), 36–43 (2010)
17. Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis **88**(2), 365–411 (2004)
18. Leung, P.L., Chan, W.Y.: Estimation of the scale matrix and its eigenvalues in the Wishart and the multivariate F distributions. Ann. Inst. Stat. Math. **50**(3), 523–530 (1998)
19. Lorentzen, R.J., Bhakta, T., Grana, D., Luo, X., Valestrand, R., Naevdal, G.: History matching of real production and seismic data in the Norne field. In ECMOR XVI-16th European conference on the mathematics of oil recovery, European Association of Geoscientists & Engineers vol. 2018, pp. 1–16 (2018)
20. Morell, E.: History matching of the Norne field. Master's thesis, Norges teknisk-naturvitenskapelige universitet (2010)
21. Park, H., Scheidt, C., Fenwick, D., Boucher, A., Caers, J.: History matching and uncertainty quantification of facies models with multiple geological interpretations. Comput. Geosci. 17, 609–621 (2013)

22. Rwechungura, R., Dadashpour, M., Kleppe, J.: Application of particle swarm optimization for parameter estimation integrating production and time lapse seismic data. In SPE Offshore Europe Conference and Exhibition, pp. SPE–146199. SPE (2011)
23. Schafer, J., Strimmer, K.: A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat. Appl. Genet. Mol. Biol. 4(1) (2005)
24. Scheidt, C., Jeong, C., Mukerji, T., Caers, J.: Probabilistic falsification of prior geologic uncertainty with seismic amplitude data: Application to a turbidite reservoir case. Geophysics 80(5), M89–M100 (2015)
25. Schulze-Riegert, R., Nwakile, M., Skripkin, S., Willen, Y.: Scalability and performance efficiency of history matching workflows using MCMC and adjoint techniques applied to the Norne North Sea reservoir case study. In 78th EAGE Conference and Exhibition 2016, European Association of Geoscientists Engineers vol. 2016, pp. 1–21 (2016)
26. Wang, L., Yin, D.Z., Caers, J.: Data Science for the Geosciences. Cambridge University Press, (2023)
27. Yao, Y., Vehtari, A., Simpson, D., Gelman, A.: Using stacking to average Bayesian predictive distributions (with discussion). Bayesian Anal. 13(3), 917–944 (2018)
28. Zhen, Yin., Strebelle, S., Caers, Jef.: Automated Monte Carlo-based quantification and updating of geological uncertainty with borehole data (AutoBEL v1.0). Geoscientific Model Development, 13(2), 651–672 (2020)
29. Zhang, Y., Leeuwenburgh, O., Carpentier, S., Steeghs, P.: 4D seismic history matching of the Norne field model using ensemble-based methods with distance parameterization. In IOR 2017-19th European Symposium on Improved Oil Recovery, European Association of Engineers vol. 2017, pp. 1–12 (2017)