MULTISCALE MODEL. SIMUL.
Vol. 18, No. 4, pp. 1565–1594

ENSEMBLE KALMAN FILTER FOR MULTISCALE INVERSE PROBLEMS*

ASSYR ABDULLE†, GIACOMO GAREGNANI†, AND ANDREA ZANONI†

Abstract. We present a novel algorithm based on the ensemble Kalman filter to solve inverse problems involving multiscale elliptic partial differential equations. Our method is based on numerical homogenization and finite element discretization and allows us to recover a highly oscillatory tensor from measurements of the multiscale solution in a computationally inexpensive manner. The properties of the approximate solution are analyzed with respect to the multiscale and discretization parameters, and a convergence result is shown to hold. A reinterpretation of the solution from a Bayesian perspective is provided, and convergence of the approximate conditional posterior distribution is proved with respect to the Wasserstein distance. A numerical experiment validates our methodology, with a particular emphasis on modeling error and computational cost.

Key words. inverse problems, multiscale modeling, homogenization, ensemble Kalman filter, Bayesian inference, modeling error

AMS subject classifications. 65N21, 65N30, 62F15, 74Q05

DOI. 10.1137/20M1348431

1. Introduction.

In this work we consider the application of techniques derived from the Kalman filter to inverse problems involving multiscale phenomena which can be modeled by means of partial differential equations (PDEs). Inverse problems arise in many fields, such as seismography, meteorology, and tomography, all physical domains with a multiscale nature. Our reference mathematical model is given by multiscale elliptic PDEs of the form

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f & \text{in } D, \\ p^\varepsilon = 0 & \text{on } \partial D, \end{cases} \tag{1}$$

where $D \subset \mathbb{R}^d$ is the physical domain, A_u^ε is a tensor oscillating with an amplitude described by the parameter ε, and u is a possibly infinite-dimensional unknown which parametrizes the tensor A_u^ε. We are then interested in the solution of inverse problems involving the retrieval of the parameter u given noisy observations derived from the solution p^ε.

Multiscale inverse problems of this form have been recently introduced in [14] and analyzed extensively in [2, 3]. In particular, in [2] Abdulle and Di Blasio build a coarse-graining approach to solve the inverse problem regularized with a Tikhonov technique. The main idea is replacing the computationally expensive solution of the highly oscillating multiscale problem with an homogenized surrogate, which eliminates the fast variables and is therefore cheaper. In particular, the theory of homogenization guarantees under certain assumptions, which will be specified throughout this work,

*Received by the editors June 26, 2020; accepted for publication (in revised form) September 25, 2020; published electronically December 8, 2020.
https://doi.org/10.1137/20M1348431
Funding: The authors are partially supported by the Swiss National Science Foundation, under grant 200020_172710.

†ANMC, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (assyr.abdulle@epfl.ch, giacomo.garegnani@epfl.ch, andrea.zanoni@epfl.ch).

that there exists a PDE of the form

$$\begin{cases} -\nabla \cdot (A_u^0 \nabla p^0) = f & \text{in } D, \\ p^0 = 0 & \text{on } \partial D, \end{cases} \tag{2}$$

such that the solution p^0 is the weak limit of the functions p^ε in the vanishing limit for ε, and such that A_u^0 is independent of ε. In [2], the authors showed that employing this homogenized model to the multiscale inverse problem guarantees a good approximation to its solution if a Tikhonov regularization is employed. This framework has been successively enlarged by the same authors to the Bayesian case in [3], where the analysis involves posterior distributions arising from both the multiscale and the homogenized model. In the same work, a technique for estimating the modeling error which was developed in [5, 6] is successfully applied to multiscale inverse problems to account for the homogenization and discretization errors.

The ensemble Kalman filter (EnKF), first introduced in [10], is an algorithm which is widely employed in the engineering community for the estimation of the state of partially observed dynamical systems whose dynamics are governed by a nonlinear agent. In particular, Kalman filters have long been used successfully in meteorology, oceanography, and automation applications. In [11], Iglesias, Law, and Stuart propose the application of the EnKF method to obtain a pointwise solution to inverse problems involving PDEs, and an extension of their analysis giving a Bayesian interpretation of the filtering solution is presented in [18].

In this work, we present a combination of the well-established techniques of homogenization and filtering to build a novel scheme for solving multiscale inverse problems in an efficient and reliable manner. In the same spirit as [2, 3], we prove that it is possible to eliminate the fast scales from the PDE appearing in the inverse problem relying on the theory of homogenization by introducing an artificial homogenized surrogate forward model, thus obtaining a solution which is accurate in the vanishing limit for the multiscale parameter ε. In our analysis, we both consider pointwise estimations as in [11] and Bayesian solutions as in [18], thus showing convergence results which are endowed with decay rates under special assumptions on the problem. Inspired by [3, 5, 6], we then consider offline and online techniques for estimating the modeling error and prove a novel result indicating the computational cost which is required for such an estimation for any given multiscale problem.

In general, the EnKF has two main advantages with respect to other approaches. First, a Bayesian interpretation of the solution to the inverse problem is obtained from the algorithm without any additional cost. The Bayesian paradigm, frequently adopted in the context of inverse problems involving PDEs, provides a full uncertainty quantification on the solution and is therefore preferable to a pointwise estimation. Second, the EnKF is easily parallelizable, thus allowing one, in practice, to solve complex inverse problems faster than employing, e.g., Markov chain Monte Carlo methods.

The main contributions of this paper are

- to introduce a new method based on filtering techniques and numerical homogenization, which is computationally efficient and easily parallelizable to solve multiscale inverse problems;
- to analyze theoretically the convergence properties of our method both from a pointwise and a Bayesian perspectives, proving the results of convergence of the EnKF scheme in the multiscale setting;
- to estimate the modeling error caused by the discrepancy between the artificial homogenized surrogate forward model and the true multiscale data and

prove a novel theoretical result related to the optimal number of additional solves required by this technique.

The outline of the work is the following. In section 2 we briefly summarize the technique of ensemble Kalman inversion, show how it can be applied to multiscale inverse problems, and state our main theoretical results. In section 3 we present the analysis of our theoretical results, and section 4 is dedicated to the estimation of the modeling error. Finally, in section 5 we present a series of numerical experiments which corroborate our analysis.

2. Ensemble Kalman inversion for multiscale problems. In this section, we present the ensemble Kalman inversion technique for multiscale inverse problems. First, we introduce a generic framework and illustrate how the EnKF is employed to solve an inverse problem. Then we particularize to inverse problems involving multiscale elliptic PDEs, and we conclude this section by announcing our main theoretical results. For a more exhaustive treatment of the EnKF in a generic PDE context, we refer the reader to [11, 18].

2.1. Ensemble Kalman inversion. We first give a brief summary of the ensemble Kalman inversion for problems of the form

$$\text{find } u \in X \text{ given observations } y = \mathcal{G}(u) + \eta \in Y, \tag{3}$$

where X and Y are Hilbert spaces, the operator $\mathcal{G}\colon X \to Y$ is a generic forward map, and the noise η follows the Gaussian distribution $\eta \sim \mathcal{N}(0, \Gamma)$ with a symmetric positive definite covariance Γ. Kalman filters are traditionally employed to estimate the state of a dynamical system given partial and noisy observations of its state. In order to approximate the solution of the otherwise static problem (3), it is therefore natural to introduce some artificial dynamics. Let us consider the space $Z = X \times Y$ and the map $\Xi\colon Z \to Z$ given by

$$\Xi(z) = \begin{bmatrix} u \\ \mathcal{G}(u) \end{bmatrix} \quad \text{for} \quad z = \begin{bmatrix} u \\ v \end{bmatrix} \in Z. \tag{4}$$

Given an initial value $z_0 \in Z$, we define artificial discrete dynamics on Z through the recursion

$$z_{n+1} = \Xi(z_n), \quad n = 0, 1, \dots. \tag{5}$$

The dynamics on Z are completed consistently with the problem (3) by the observation equation

$$y_{n+1} = H z_{n+1} + \eta_{n+1}, \tag{6}$$

where $H\colon Z \to Y$ is the projection operator defined by $H = \begin{bmatrix} 0 & I \end{bmatrix}$ and $\{\eta_n\}_{n\in\mathbb{N}}$ is an independent and identically distributed (i.i.d.) sequence of random variables distributed identically to the noise of the inverse problem (3), i.e., $\eta_n \sim \mathcal{N}(0, \Gamma)$. In fact, let us remark that combining (5) and (6) one gets $y_{n+1} = \mathcal{G}(u_n) + \eta_{n+1}$, which is in law equivalent to the equality appearing in (3).

Kalman filters proceed recursively to estimate the state of dynamics of the form (5) when observations are provided by the model (6). At each time n, the estimation is performed in two steps. First, (5) is employed in the so-called *prediction* step, and then (6) is employed to correct the prediction in the *update* or *analysis* step. In case Ξ is a linear map, both prediction and update steps admit a closed-form expression,

often referred to in literature as the Kalman formulas. Conversely, in the case Ξ is nonlinear, there exists no explicit solution to the estimation problem and one has to recur to an approximation such as the EnKF method, which we briefly describe here.

Given a positive integer J, the EnKF method proceeds by propagating and updating an ensemble $\{z_n^{(j)}\}_{j=1}^J \subset Z$ of particles with discrete approximations of the Kalman formulas. Let $\mathcal{A} \subset X$ be such that $\dim(\mathcal{A}) \leq J$, and let the initial ensemble $\{z_0^{(j)}\}_{j=1}^J$ be given by

$$z_0^{(j)} = \begin{bmatrix} \psi^{(j)} \\ \mathcal{G}(\psi^{(j)}) \end{bmatrix},$$

where $\{\psi^{(j)}\}_{j=1}^J \subset \mathcal{A}$. At each time $n = 0, 1, \dots, N-1$, and for each $j = 1, \dots, J$, the prediction step is simply given by

$$\hat{z}_{n+1}^{(j)} = \Xi(z_n^{(j)}). \tag{7}$$

In the analysis step, this partially updated ensemble is updated given knowledge of the data y. For better exploring the space Y, the data are randomized and each particle $z_{n+1}^{(j)}$ is compared to i.i.d. versions of the data given by $y_{n+1}^{(j)} = y + \eta_{n+1}^{(j)}$, where $\eta_{n+1}^{(j)} \sim \mathcal{N}(0, \Gamma)$. The analysis step is then given by

$$z_{n+1}^{(j)} = \hat{z}_{n+1}^{(j)} + K_{n+1}(y_{n+1}^{(j)} - H\hat{z}_{n+1}^{(j)}). \tag{8}$$

The operator $K_{n+1} \colon Y \to Z$, the Kalman gain, weighs the effects of dynamics and observations in this two-step procedure, and is defined as

$$K_{n+1} = C_{n+1}H^*R_{n+1}, \qquad R_{n+1} = (HC_{n+1}H^* + \Gamma)^{-1}, \tag{9}$$

where $C_{n+1} \colon Z \to Z$ is the empirical covariance of the partially updated ensemble $\{\hat{z}_{n+1}^{(j)}\}_{j=1}^J$, the operator $H^* \colon Y \to Z$ is the adjoint of H, which is given in (6), and we recall Γ to be the covariance of the noise y, so that $R_{n+1} \colon Y \to Y$. Intuitively, one can notice that when the ensemble's covariance C_{n+1} is large with respect to the noise covariance Γ, i.e., the observation model is more precise than the dynamics, we will have $z_{n+1}^{(j)} \approx y_{n+1}^{(j)}$, while in the opposite case we will have $z_{n+1}^{(j)} \approx \hat{z}_{n+1}^{(j)}$. A more precise definition of the operators appearing above will be given in section 3. At the final step N, we project the particles on the space X and average the result to obtain the estimate

$$u_{\text{EnKF}} = \frac{1}{J}\sum_{j=1}^J H^\perp z_N^{(j)} = \frac{1}{J}\sum_{j=1}^J u_N^{(j)},$$

where $H^\perp \colon Z \to X$ is defined by $H^\perp = \begin{bmatrix} I & 0 \end{bmatrix}$. The last detail missing to fully define the EnKF is its initialization, i.e., the choice of the space $\mathcal{A}$ defining the initial ensemble. We assume prior knowledge is available on the parameter $u \in X$ and that it is summarized by a probability measure μ_0 on X. In this case, one can draw J i.i.d. samples $\psi^{(j)}$ from μ_0 and fix $\mathcal{A} = \text{span}\{\psi^{(j)}\}_{j=1}^J$.

Remark 1. The computational cost of the EnKF method is approximately equal to the number of evaluations of the forward operator, which in a PDE framework dominates with respect to the algebraic operations needed in the analysis step. Therefore, the complexity of the algorithm is $\mathcal{O}(JN)$. Nonetheless, let us remark that the prediction step (7) can be easily parallelized, since the forward operator is applied independently to each particle. Hence, for a reasonable number of particles (or a high number of computing units), we have that the overall cost is of order $\mathcal{O}(N)$.

As shown in [18], a slight modification of the EnKF algorithm allows us to obtain with no additional cost a Bayesian solution to (3) from the evolving ensemble. Let μ_0 be, as above, a prior probability measure on X and let the initial ensemble $\{\psi^{(j)}\}_{j=1}^J$ consist of i.i.d. samples from μ_0. Given a number of steps N, let $\Delta = 1/N$ be a "step size." Let us modify the algorithm above by taking, instead of the covariance Γ of the noise, its scaled version $\Delta^{-1}\Gamma$ in formula (9). Moreover, let us define the empirical measure $\hat{\mu}_n$ on X induced by the ensemble at the nth step, i.e.,

$$\hat{\mu}_n(du) = \frac{1}{J}\sum_{j=1}^{J} \delta_{u_n^{(j)}}(du), \tag{10}$$

where δ_x is the Dirac mass concentrated in $x \in U$. Then, it has been shown in [18] that $\hat{\mu}_n$ is a good approximation of the measure μ_n defined by

$$\mu_n(du) = \frac{1}{Z_n} e^{-n\Delta\Phi(u;y)}\mu_0(du), \tag{11}$$

where Z_n is the normalization constant and $\Phi(u;y)$ is the least squares functional

$$\Phi(u;y) = \frac{1}{2}\left\|\Gamma^{-1/2}(y - \mathcal{G}(u))\right\|_2^2.$$

For $n = N$, we have by definition $N\Delta = 1$ and the measure $\mu \coloneqq \mu_N$ given by

$$\mu(du) = \frac{1}{Z} e^{-\Phi(u;y)}\mu_0(du), \tag{12}$$

where Z is the normalization constant, is exactly the posterior measure of the parameter u given the prior μ_0 in the Bayesian sense (see, e.g., [19]). Summarizing, if one carefully modifies formula (9) for the Kalman gain, it is sufficient to run the EnKF method for N steps and the empirical measure given by the particles is an approximation to the Bayesian posterior.

2.2. Multiscale ensemble Kalman inversion. In this work, we consider the application of ensemble Kalman inversion to a multiscale inverse problem of the form

$$\text{find } u \in X \text{ given observations } y = \mathcal{G}^\varepsilon(u) + \eta \in Y, \tag{13}$$

where $\varepsilon > 0$ is the multiscale parameter, which often is $\varepsilon \ll 1$, the operator $\mathcal{G}^\varepsilon \colon X \to Y$ is the multiscale forward map, and where, as above, $\eta \sim \mathcal{N}(0,\Gamma)$ for some symmetric positive definite covariance Γ on Y. Let $D \subset \mathbb{R}^d$ be an open bounded domain and let $H_0^1(D)$ denote the space of functions $v \colon D \to \mathbb{R}$ in $L^2(D)$ with first order weak derivatives in $L^2(D)$ and whose trace on ∂D vanishes. We consider the forward map $\mathcal{G}^\varepsilon$ to be the composition $\mathcal{G}^\varepsilon = \mathcal{O} \circ \mathcal{S}^\varepsilon$ of an observation operator $\mathcal{O} \colon H_0^1(D) \to Y$ and a multiscale solution operator $\mathcal{S}^\varepsilon \colon X \to H_0^1(D)$. In particular, for $u \in X$, the operator $\mathcal{S}^\varepsilon \colon u \mapsto p^\varepsilon \in H_0^1(D)$, where p^ε is the weak solution of the elliptic PDE

$$\begin{cases} -\nabla \cdot (A_u^\varepsilon \nabla p^\varepsilon) = f & \text{in } D, \\ p^\varepsilon = 0 & \text{on } \partial D, \end{cases} \tag{14}$$

for a right-hand side $f \in L^2(D)$. We assume that the tensor $A_u^\varepsilon \colon D \to \mathbb{R}^{d\times d}$ is a parametrized multiscale tensor admitting explicit scale separation between slow and fast spatial variables, i.e.,

$$A_u^\varepsilon(x) = A\left(u(x), \frac{x}{\varepsilon}\right),$$

where the map $(t, x) \mapsto A(t, x/\varepsilon)$ is assumed to be known and where A is periodic in its second argument. In other words, the unknown u of the inverse problem (13) governs the slow-scale variations of the rapidly oscillating tensor A_u^ε.

Let us consider now the application of ensemble Kalman inversion to the inverse problem (13). Since the PDE (14) does not in general admit a closed-form solution, one has to employ a numerical approximation to evaluate the forward map $\mathcal{G}^\varepsilon$. If ε is small and we employ the finite element method (FEM), a fine discretization is needed to resolve the smallest scale and thus evaluate the forward operator $\mathcal{G}^\varepsilon$, which clearly leads to a high computational cost. Indeed, as for Remark 1, a run of the EnKF algorithm would lead to $\mathcal{O}(N)$ solutions of (14), which is indeed unfeasible.

In order to approach the multiscale problem more efficiently we recur to the theory of homogenization (see, e.g., [8]), which ensures the existence of a nonoscillating homogenized tensor A_u^0, such that for $\varepsilon \to 0$ the solution p^ε of (14) tends weakly in $H_0^1(D)$ to the solution p^0 of the problem

$$\begin{cases} -\nabla \cdot (A_u^0 \nabla p^0) = f & \text{in } D, \\ p^0 = 0 & \text{on } \partial D. \end{cases} \tag{15}$$

Hence, this homogenized problem is a good surrogate of (14) when $\varepsilon \ll 1$, and its nonoscillating nature allows us to discretize it with FEM on an arbitrarily coarse mesh, whose maximum diameter is denoted by h. Therefore, denoting by $\mathcal{G}_h^0 \colon \mathcal{O} \circ \mathcal{S}_h^0$, where $\mathcal{S}_h^0 \colon u \mapsto p_h^0$, the numerical solution of (15), we study in this paper the behavior of the EnKF when $\mathcal{G}^\varepsilon$ is replaced by its cheap approximation $\mathcal{G}_h^0$. Let us denote by $\{u_{n,h}^{0,(j)}\}_{j=1}^J$ the ensemble obtained after n iterations of the EnKF algorithm with the forward operators $\mathcal{G}_h^0$ in the prediction step (7). With this notation, given an initial ensemble $\{u_{0,h}^{0,(j)}\}_{j=1}^J$, at each step $n = 0, 1, \dots, N-1$, our algorithm proceeds as

(i) for each $u_{n,h}^{0,(j)}$, compute the homogenized tensor $A^0_{u_{n,h}^{0,(j)}}$ and build the forward map $\mathcal{G}_h^0$;

(ii) perform the prediction step (7) with $\mathcal{G}_h^0$ and the analyis step (8) to obtain the updated ensemble $\{u_{n+1,h}^{0,(j)}\}_{j=1}^J$.

The computation of the homogenized tensor relies as well on numerical procedures; here we use the finite element heterogeneous multiscale method (FE-HMM) [1, 4]. Let us finally remark that similar analyses have been carried out in [2, 3, 14] for different methodologies in the solution of (13).

2.3. Statement of main results. Let us first introduce some assumptions and notation which will be employed in the analysis. First, we introduce a regularity assumption on tensors which will be fulfilled by A_u^ε and A_u^0.

Assumption 1. The tensor $A_u \colon D \to \mathbb{R}^{d\times d}$ satisfies for all $u, u_1, u_2 \in X$ and $\xi \in \mathbb{R}^d$

$$\|A_{u_1} - A_{u_2}\|_{L^\infty(D;\mathbb{R}^{d\times d})} \le M \|u_1 - u_2\|_X, \qquad A_u \xi \cdot \xi \ge \alpha_0 \|\xi\|_2^2, \tag{16}$$

where M and α_0 are positive constants.

We now introduce a regularity assumption on the observation operator.

Assumption 2. The observation operator $\mathcal{O}\colon H_0^1(D) \to Y$ satisfies for all $p_1, p_2 \in H_0^1(D)$

$$\|\mathcal{O}(p_1) - \mathcal{O}(p_2)\|_Y \leq C_{\mathcal{O}} \|p_1 - p_2\|_{L^2(D)},$$

where $C_{\mathcal{O}}$ is a positive constant.

Note that since $\mathcal{O}$ is defined on $H_0^1(D) \subset L^2(D)$, Assumption 2 is stronger than Lipschitz continuity. Finally, we introduce an assumption on the algorithm which will be employed in the analysis.

Assumption 3. All the particles in the ensemble lie at each iteration in a ball $B_R(u^*)$ for some $R > 0$ sufficiently big, where u^* is the true value of the unknown.

Remark 2. Let us remark that Assumption 3 reduces the possible outcomes of the algorithm even if R can be chosen arbitrarily big. Therefore, in the following all the expectations in the statements and in the proofs have to be intended as conditional expectations given that all particles lie in a ball $B_R(u^*)$ centered in the true value of the unknown. For example, when we write the expectation of the norm (see (18)) of an ensemble $u_N = \{u_N^{(j)}\}_{j=1}^J$ of particles at the Nth step of the algorithm what we mean is

$$\mathbb{E}\left[\|u_N\|\right] = \mathbb{E}\left[\|u_N\| \mid u_n^{(j)} \in B_R(u^*) \ \ \forall j = 1, \dots, J,\, n = 0, \dots, N\right]. \tag{17}$$

This abuse of notation is repeated throughout our text, and expectations should be thought as the above any time Assumption 3 holds.

For clarity, we present the analysis in the finite-dimensional setting $X = \mathbb{R}^M$ and $Y = \mathbb{R}^L$ but claim that it can be readily generalized to the infinite-dimensional case. For an ensemble $u = \{u^{(j)}\}_{j=1}^J$ of particles in $\mathbb{R}^M$, we introduce the ensemble norm

$$\|u\| \coloneqq \frac{1}{J} \sum_{j=1}^{J} \left\| u^{(j)} \right\|_2, \tag{18}$$

which is indeed a norm and where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^M$. Moreover, given a scalar α, we define the linear combination $w = u + \alpha v$ between two ensembles u and v with the same number of particles J as $\{w^{(j)} = u^{(j)} + \alpha v^{(j)}\}_{j=1}^J$.

We can now present the first main result of this work, in which we show the convergence of the ensemble obtained by the EnKF employing $\mathcal{G}_h^0$ to the one obtained employing the exact operator $\mathcal{G}^\varepsilon$ linked to the PDE (14).

THEOREM 1. *Let $u_{N,h}^0 = \{u_{N,h}^{0,(j)}\}_{j=1}^J$, $u_N^\varepsilon = \{u_N^{\varepsilon,(j)}\}_{j=1}^J$ be the ensembles after N iterations of the EnKF method with forward operators $\mathcal{G}_h^0$ and $\mathcal{G}^\varepsilon$, respectively. Then, if A_u^ε and A_u^0 satisfy Assumption* 1 *and if Assumptions* 2 *and* 3 *hold, we have*

$$\mathbb{E}\left[\left\|u_N^\varepsilon - u_{N,h}^0\right\|\right] \to 0 \qquad \text{as } \varepsilon, h \to 0.$$

In particular, if the exact solution p^0 of the homogenized problem (15) *is in $H^{q+1}(D)$ with $q \geq 1$ and we employ polynomials of degree r for the finite element basis, then*

$$\mathbb{E}\left[\left\|u_N^\varepsilon - u_{N,h}^0\right\|\right] \leq C(\varepsilon + h^{s+1}),$$

where $s = \min\{r, q\}$ and $C > 0$ is a constant independent of h and ε.

The proof of this result is the main focus of section 3.1. The second main theoretical result concerns the Bayesian interpretation of the EnKF methodology for inverse problems in the multiscale setting. Let μ_0 be a prior measure on X and the ensem-

bles $u^0_{N,h} = \{u^{0,(j)}_{N,h}\}_{j=1}^J$, $u^\varepsilon_N = \{u^{\varepsilon,(j)}_N\}_{j=1}^J$ resulting from the EnKF algorithms as in Theorem 1 both initialized with an i.i.d. sample from μ_0. We consider the discrete probability measures

$$\mu^\varepsilon = \frac{1}{J}\sum_{j=1}^J \delta_{u^{\varepsilon,(j)}_N} \qquad \text{and} \qquad \mu^0_h = \frac{1}{J}\sum_{j=1}^J \delta_{u^{0,(j)}_{N,h}}, \tag{19}$$

i.e., the EnKF approximations of the posterior μ on u defined in (12). Our goal is providing a metric on how far the two measures are from each other with respect to ε and h. Let us remark that due to the randomization of the data at each step of the EnKF algorithm, both μ^ε and μ^0_h are random probability measures. We now introduce the metric we consider for comparing the two measures.

DEFINITION 1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. A sequence of random measures $\{\mu_n\}_{n\in\mathbb{N}}$ on a metric space $(E, \mathcal{B}(E))$ dependent on a random variable ξ on $(\Omega, \mathcal{A}, P)$ is said to weakly converge in $L^1(\Omega)$ to a random measure μ on the same metric space if for all bounded continuous functions $f \in C^0_B(E)$ we have*

$$\mathbb{E}_\xi\left[\left|\int_E f\, d\mu_n - \int_E f\, d\mu\right|\right] \to 0.$$

In this case we write $\mu_n \xrightarrow{L^1} \mu$.

We can now state our second main result, whose proof is the main focus of section 3.2.

THEOREM 2. *Let the hypotheses of Theorem 1 be satisfied. Then the sequence of random measures $\{\mu^\varepsilon - \mu^0_h\}_{\varepsilon,h}$, where μ^ε and μ^0_h are defined in (19), satisfies*

$$\{\mu^\varepsilon - \mu^0_h\}_{\varepsilon,h} \xrightarrow{L^1} 0 \qquad \text{as } \varepsilon, h \to 0.$$

Remark 3. It is possible to verify that in both Theorems 1 and 2 the limits with respect to ε and h can be interchanged.

3. Convergence analysis. In this section we prove Theorems 1 and 2, the main results of this work. As announced above, the analysis is carried out in the finite-dimensional case $X = \mathbb{R}^M$ and $Y = \mathbb{R}^L$, but it can be generalized to the infinite-dimensional setting. For the purpose of the analysis, we introduce on top of the forward maps $\mathcal{G}^\varepsilon$ and $\mathcal{G}^0_h$, which have been introduced in section 2.2, the operator $\mathcal{G}^0 = \mathcal{O} \circ \mathcal{S}^0$, where $\mathcal{S}^0 \colon X \to H^1_0(D)$ is the exact solution operator associated with the homogenized PDE (15).

3.1. Convergence of the point estimate. We now focus on Theorem 1. It is clear from the desired bound that the effects of homogenization and discretization can be analyzed separately. In particular, we first show the convergence of the ensemble generated employing the forward operator $\mathcal{G}^\varepsilon$ to the one generated employing the exact homogenized operator $\mathcal{G}^0$ for $\varepsilon \to 0$. Then, in an analogous fashion, we prove the convergence of the ensemble generated with $\mathcal{G}^0_h$ to the ensemble generated employing $\mathcal{G}^0$. In order to introduce a compact notation, we denote by $\mathcal{U}_{J,M}$ the set of ensembles of dimension J with elements in $\mathbb{R}^M$ and we consider the homogenization error function $e \colon \mathbb{R} \times \mathcal{U}_{J,M} \to \mathbb{R}$, which is defined for a generic ensemble u as

$$e(\varepsilon, u) = \frac{1}{J}\sum_{j=1}^J \left\| \mathcal{G}^\varepsilon(u^{(j)}) - \mathcal{G}^0(u^{(j)}) \right\|_2, \tag{20}$$

and a discretization error function $\tilde{e}\colon \mathbb{R}\times\mathcal{U}_{J,M}\to\mathbb{R}$ as

$$\tilde{e}(h,u)=\frac{1}{J}\sum_{j=1}^{J}\left\|\mathcal{G}_h^0(u^{(j)})-\mathcal{G}^0(u^{(j)})\right\|_2. \tag{21}$$

Before proving the main theorem, we introduce some preliminary results.

Let us first consider a generic forward operator involving an elliptic PDE and show that the associated forward map is Lipschitz continuous.

LEMMA 1. *Let $\mathcal{G}\colon\mathbb{R}^M\to\mathbb{R}^L$, $\mathcal{G}=\mathcal{O}\circ\mathcal{S}$ be a forward operator such that $\mathcal{O}\colon H_0^1(D)\to\mathbb{R}^L$ is Lipschitz and $\mathcal{S}\colon\mathbb{R}^M\to H_0^1(D)$, $\mathcal{S}\colon u\mapsto p$ is defined by the solution of*

$$\begin{cases}-\nabla\cdot(A_u\nabla p)=f & \text{in } D,\\ p=0 & \text{on } \partial D,\end{cases} \tag{22}$$

where $D\subset\mathbb{R}^d$ is an open bounded set, the right-hand side $f\in L^2(D)$, and the tensor A_u satisfies Assumption 1. *Then $\mathcal{G}$ is Lipschitz with a constant depending only on the Poincaré constant of D, on the constants M and α_0 appearing in Assumption* 1, *on the right-hand side f and on the Lipschitz constant of the operator $\mathcal{O}$.*

The proof of Lemma 1 is given in the appendix. In the following lemma, whose proof is also given in the appendix, we consider the homogenization error defined in (20) and show that it vanishes in the limit $\varepsilon\to 0$.

LEMMA 2. *Let e be defined as in* (20). *Under Assumption* 2, *we have for all $u\in\mathcal{U}_{J,M}$*

$$e(\varepsilon,u)\to 0 \quad \text{as } \varepsilon\to 0.$$

Moreover, if the solution of the homogenized problem (15) *is in $H^2(D)$ independently of u, then there exists $K>0$ independent of ε and u such that*

$$e(\varepsilon,u)\le K\varepsilon.$$

Finally, we consider the particle empirical covariances of ensembles given by the EnKF algorithm, thus proving their boundedness and Lipschitz continuity. The proof of this lemma can be found in the appendix.

LEMMA 3. *Let $C^{up}(u)\in\mathbb{R}^{M\times L}$ and $C^{pp}(u)\in\mathbb{R}^{L\times L}$ be defined as*

$$C^{up}(u)=\frac{1}{J}\sum_{j=1}^{J}\big(u^{(j)}-\bar{u}\big)\big(\mathcal{G}(u^{(j)})-\bar{\mathcal{G}}\big)^T,\quad C^{pp}(u)=\frac{1}{J}\sum_{j=1}^{J}\big(\mathcal{G}(u^{(j)})-\bar{\mathcal{G}}\big)\big(\mathcal{G}(u^{(j)})-\bar{\mathcal{G}}\big)^T,$$

where $\bar{u}\in\mathbb{R}^M$ and $\bar{\mathcal{G}}\in\mathbb{R}^L$ are the empirical averages

$$\bar{u}=\frac{1}{J}\sum_{j=1}^{J}u^{(j)},\qquad \bar{\mathcal{G}}=\frac{1}{J}\sum_{j=1}^{J}\mathcal{G}(u^{(j)}),$$

and let $\mathcal{G}\colon\mathbb{R}^M\to\mathbb{R}^L$ be Lipschitz with constant $C_{\mathcal{G}}$. Then, there exist four constants $C_i>0$, $i=1,\dots,4$, such that

(i) $\|C^{up}(u)\|_2\le C_1$,
(ii) $\|C^{pp}(u)\|_2\le C_2$,
(iii) $\|C^{up}(u_1)-C^{up}(u_2)\|_2\le C_3\|u_1-u_2\|$,
(iv) $\|C^{pp}(u_1)-C^{pp}(u_2)\|_2\le C_4\|u_1-u_2\|$

for all ensembles $u,u_1,u_2\in\mathcal{U}_{J,M}$ which are stable in the sense of Assumption 3.

In order to clarify the exposition, we first consider the amplification of the error over one step between the EnKF algorithms employing the multiscale and the homogenized forward operators, respectively, which is summarized in the following lemma.

LEMMA 4. *For all $n = 0, \dots, N-1$, let $u_n^0 = \{u_n^{0,(j)}\}_{j=1}^J, u_n^\varepsilon = \{u_n^{\varepsilon,(j)}\}_{j=1}^J$ be the ensembles of particles at the nth iteration of the EnKF for the forward operators $\mathcal{G}^0$ and $\mathcal{G}^\varepsilon$, respectively. Then, under Assumptions 1, 2, and 3, there exist positive constants α and γ such that*

$$\mathbb{E}\left[\left\|u_{n+1}^\varepsilon - u_{n+1}^0\right\|\right] \le \alpha \mathbb{E}\left[\left\|u_n^\varepsilon - u_n^0\right\|\right] + \gamma \mathbb{E}\left[e(\varepsilon, u_n^0)\right], \tag{23}$$

where $e(\varepsilon, u)$ is given in (20).

Proof. First, due to Assumption 2 and the Poincaré inequality with constant C_p we have

$$\|\mathcal{O}(p_1) - \mathcal{O}(p_2)\|_2 \le C_{\mathcal{O}} \|p_1 - p_2\|_{L^2(D)} \le C_{\mathcal{O}} C_p \|\nabla p_1 - \nabla p_2\|_{L^2(D;\mathbb{R}^d)},$$

which shows that $\mathcal{O}$ is Lipschitz with constant $C_{\mathcal{O}} C_p$. Therefore, applying Lemma 1, we deduce that both $\mathcal{G}^0$ and $\mathcal{G}^\varepsilon$ are Lipschitz with constant $C_{\mathcal{G}}$ independent of ε. The Kalman update formulas (8) restricted to the u variable read (see [11])

$$u_{n+1}^{\varepsilon,(j)} = u_n^{\varepsilon,(j)} + C^{up}(u_n^\varepsilon)(C^{pp}(u_n^\varepsilon) + \Gamma)^{-1}(y_{n+1} - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})), \tag{24}$$

$$u_{n+1}^{0,(j)} = u_n^{0,(j)} + C^{up}(u_n^0)(C^{pp}(u_n^0) + \Gamma)^{-1}(y_{n+1} - \mathcal{G}^0(u_n^{0,(j)})). \tag{25}$$

Combining (24) and (25), we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|u_{n+1}^\varepsilon - u_{n+1}^0\right\|\right] \\
&\quad = \frac{1}{J}\sum_{j=1}^J \mathbb{E}\left[\left\| u_n^{\varepsilon,(j)} + C^{up}(u_n^\varepsilon)(C^{pp}(u_n^\varepsilon) + \Gamma)^{-1}(y_{n+1}^{(j)} - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})) \right.\right. \\
&\qquad\qquad \left.\left. - u_n^{0,(j)} - C^{up}(u_n^0)(C^{pp}(u_n^0) + \Gamma)^{-1}(y_{n+1}^{(j)} - \mathcal{G}^0(u_n^{0,(j)}))\right\|_2\right],
\end{aligned}$$

and using the triangle inequality we obtain

$$\mathbb{E}\left[\left\|u_{n+1}^\varepsilon - u_{n+1}^0\right\|\right] \le \mathbb{E}\left[\left\|u_n^\varepsilon - u_n^0\right\|\right] + S_1 + S_2 + S_3, \tag{26}$$

where

$$S_1 = \frac{1}{J}\sum_{j=1}^J \mathbb{E}\left[\left\|C^{up}(u_n^\varepsilon) - C^{up}(u_n^0)\right\|_2 \left\|(C^{pp}(u_n^\varepsilon) + \Gamma)^{-1}\right\|_2 \left\|y_{n+1}^{(j)} - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})\right\|_2\right], \tag{27}$$

$$\begin{aligned}
S_2 = \frac{1}{J}\sum_{j=1}^J \mathbb{E}\Big[&\left\|C^{up}(u_n^0)\right\|_2 \left\|(C^{pp}(u_n^\varepsilon) + \Gamma)^{-1} - (C^{pp}(u_n^0) + \Gamma)^{-1}\right\|_2 \\
&\times \left\|y_{n+1}^{(j)} - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})\right\|_2\Big],
\end{aligned} \tag{28}$$

$$S_3 = \frac{1}{J}\sum_{j=1}^J \mathbb{E}\left[\left\|C^{up}(u_n^0)\right\|_2 \left\|(C^{pp}(u_n^0) + \Gamma)^{-1}\right\|_2 \left\|\mathcal{G}^0(u_n^{0,(j)}) - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})\right\|_2\right]. \tag{29}$$

Let us introduce two useful inequalities which will be employed in the following. Given A and B square invertible matrices of the same size, it holds

$$\left\|A^{-1} - B^{-1}\right\|_2 \leq \left\|A^{-1}\right\|_2 \left\|B^{-1}\right\|_2 \|A - B\|_2 . \tag{30}$$

Moreover, if A is positive semidefinite and B is positive definite, it holds

$$\left\|(A+B)^{-1}\right\|_2 \leq \left\|B^{-1}\right\|_2 . \tag{31}$$

Let us first consider S_1. Applying Lemma 3 and (31) to the first two factors gives

$$S_1 \leq \frac{C_3}{J} \sum_{j=1}^{J} \mathbb{E}\left[\left\|u_n^\varepsilon - u_n^0\right\| \left\|\Gamma^{-1}\right\|_2 \left\|y_{n+1}^{(j)} - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})\right\|_2\right]. \tag{32}$$

Moreover, since $y_{n+1}^{(j)} = y + \eta_{n+1}^{(j)}$ and since $y = \mathcal{G}^\varepsilon(u^*) + \eta$, where u^* is the true value of the unknown and η is the true realization of the noise, the triangle inequality yields

$$\left\|y_{n+1}^{(j)} - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})\right\|_2 \leq \left\|\mathcal{G}^\varepsilon(u^*) - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})\right\|_2 + \left\|\eta_{n+1}^{(j)} + \eta\right\|_2 , \tag{33}$$

which, since $\mathcal{G}^\varepsilon$ is Lipschitz and due to Assumption 3, implies

$$\left\|y_{n+1}^{(j)} - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})\right\|_2 \leq C_{\mathcal{G}} R + \left\|\eta_{n+1}^{(j)} + \eta\right\|_2 . \tag{34}$$

Hence, we get

$$S_1 \leq \frac{1}{J} C_3 \left\|\Gamma^{-1}\right\|_2 \sum_{j=1}^{J} \mathbb{E}\left[\left\|u_n^\varepsilon - u_n^0\right\| \left(C_{\mathcal{G}} R + \left\|\eta_{n+1}^{(j)} + \eta\right\|_2\right)\right].$$

Finally, the random variables $\zeta_{n+1}^{(j)} \coloneqq \eta_{n+1}^{(j)} + \eta$ are i.i.d., distributed as $\zeta \sim \mathcal{N}(0, 2\Gamma)$ and independent of u_n^ε and u_n^0, which implies first

$$\mathbb{E}[\|\zeta\|_2] \leq \sqrt{\mathbb{E}[\|\zeta\|_2^2]} \leq \sqrt{2\mathrm{tr}(\Gamma)}$$

and, second, defining $\alpha_1 \coloneqq C_3 \|\Gamma^{-1}\|_2 (C_{\mathcal{G}} R + \sqrt{2\mathrm{tr}(\Gamma)})$, yields the final bound

$$S_1 \leq \alpha_1 \mathbb{E}\left[\left\|u_n^\varepsilon - u_n^0\right\|\right]. \tag{35}$$

Let us now consider the second term S_2. We apply Lemma 3 to the norm of $C^{up}(u_n^0)$. Moreover, applying the inequalities (30), (31) and Lemma 3 gives

$$\left\|(C^{pp}(u_n^\varepsilon) + \Gamma)^{-1} - (C^{pp}(u_n^0) + \Gamma)^{-1}\right\|_2 \leq C_4 \left\|\Gamma^{-1}\right\|_2^2 \left\|u_n^\varepsilon - u_n^0\right\|. \tag{36}$$

Reasoning as for S_1 for the third factor appearing in (28) finally yields

$$S_2 \leq \alpha_2 \mathbb{E}\left[\left\|u_n^\varepsilon - u_n^0\right\|\right], \tag{37}$$

where $\alpha_2 \coloneqq C_1 C_4 \|\Gamma^{-1}\|_2^2 (C_{\mathcal{G}} R + \sqrt{2\mathrm{tr}(\Gamma)})$. We now consider the last term S_3. The first factor appearing in (29) can be bounded by Lemma 3 and for the second factor we use (31), thus obtaining

$$\left\|(C^{pp}(u_n^0) + \Gamma)^{-1}\right\|_2 \leq \left\|\Gamma^{-1}\right\|_2 .$$

Regarding the third factor of (29), we apply the triangle inequality and the Lipschitz continuity of the forward operator $\mathcal{G}^\varepsilon$, which yield

$$\left\|\mathcal{G}^0(u_n^{0,(j)}) - \mathcal{G}^\varepsilon(u_n^{\varepsilon,(j)})\right\|_2 \leq \left\|\mathcal{G}^0(u_n^{0,(j)}) - \mathcal{G}^\varepsilon(u_n^{0,(j)})\right\|_2 + C_{\mathcal{G}} \left\|u_n^{0,(j)} - u_n^{\varepsilon,(j)}\right\|_2. \tag{38}$$

Substituting back into S_3 and by definition of $e(\varepsilon, u_n^0)$ and of the ensemble norm we obtain

$$S_3 \leq C_1 \left\|\Gamma^{-1}\right\|_2 \mathbb{E}\left[e(\varepsilon, u_n^0)\right] + C_1 \left\|\Gamma^{-1}\right\|_2 C_{\mathcal{G}} \mathbb{E}\left[\left\|u_n^0 - u_n^\varepsilon\right\|\right].$$

Therefore, defining $\alpha_3 = C_1 \|\Gamma^{-1}\|_2 C_{\mathcal{G}}$ and $\gamma = C_1 \left\|\Gamma^{-1}\right\|_2$ we have the bound

$$S_3 \leq \alpha_3 \mathbb{E}\left[\left\|u_n^0 - u_n^\varepsilon\right\|\right] + \gamma \mathbb{E}\left[e(\varepsilon, u_n^0)\right]. \tag{39}$$

Finally, defining $\alpha := 1 + \alpha_1 + \alpha_2 + \alpha_3$, and using the results (26), (35), (37), and (39), we obtain the desired result. ∎

We now present the main result about global multiscale convergence of the EnKF algorithm.

PROPOSITION 1. *Under the notation and assumptions of Lemma* 4, *letting* $u_0^\varepsilon = u_0^0$ *be the same initial ensemble, we have*

$$\mathbb{E}\left[\left\|u_N^\varepsilon - u_N^0\right\|\right] \to 0 \qquad \text{as } \varepsilon \to 0.$$

Moreover, if the solution of the homogenized problem (15) *is sufficiently regular, namely,* $p^0 \in H^2(D)$, *then there exists* $K_1 > 0$ *independent of* ε *such that*

$$\mathbb{E}\left[\left\|u_N^\varepsilon - u_N^0\right\|\right] \leq K_1 \varepsilon.$$

Proof. Since $u_0^\varepsilon = u_0^0$, iterating the estimate of Lemma 4 yields

$$\mathbb{E}\left[\left\|u_N^\varepsilon - u_N^0\right\|\right] \leq \gamma \sum_{i=0}^{N-1} \alpha^{N-1-i} \mathbb{E}\left[e(\varepsilon, u_i^0)\right].$$

Applying Lemma 2, we have $e(\varepsilon, u_i^0) \to 0$ for all $i = 0, \ldots, N-1$, hence as $\varepsilon \to 0$

$$\mathbb{E}\left[\left\|u_N^\varepsilon - u_N^0\right\|\right] \to 0.$$

Moreover, if p^0 belongs to $H^2(D)$, applying Lemma 2 gives

$$\mathbb{E}\left[\left\|u_N^\varepsilon - u_N^0\right\|\right] \leq K_1 \varepsilon, \tag{40}$$

where $K_1 = \gamma(\alpha^N - 1)K/(\alpha - 1)$, which is the desired result. ∎

We now consider convergence with respect to the FEM discretization of the homogenized problem. First, we introduce a preliminary result, which plays the role of Lemma 2 in the context of numerical convergence and whose proof is given in the appendix.

LEMMA 5. *Let* $\tilde{e}$ *be defined in* (21) *and let Assumption* 2 *hold. If the exact solution* p^0 *of the homogenized problem* (22) *is in* $H^{q+1}(D)$, *the right-hand side* f *is in* $H^{q-1}(D)$ *and we employ polynomials of degree* r *for the finite element basis, then*

$$\tilde{e}(h, u) \leq \tilde{K} h^{s+1},$$

where $s = \min\{r, q\}$.

We can now state the main result concerning convergence with respect to the numerical discretization of the homogenized problem.

PROPOSITION 2. *Let* $u_N^0 = \{u_N^{0,(j)}\}_{j=1}^J$, $u_{N,h}^0 = \{u_{N,h}^{0,(j)}\}_{j=1}^J$ *be the ensembles of particles at the last iteration of the iterative EnKF for the forward operators* $\mathcal{G}^0$ *and* $\mathcal{G}_h^0$*, respectively. Then, under Assumptions* 1, 2, 3 *and if the exact solution* p^0 *of the homogenized problem* (22) *is in* $H^{q+1}(D)$ *and we use polynomials of degree* r *for the finite element basis, we have*

$$\mathbb{E}\left[\left\|u_{N,h}^0 - u_N^0\right\|\right] \leq K_2 h^{s+1},$$

where $s = \min\{r, q\}$ *and* K_2 *is a positive constant independent of* h.

Proof. The proof of Proposition 2 is identical to the proof of Proposition 1, except that all the ensembles $\{u_n^\varepsilon\}_{n=1}^N$ obtained by the multiscale operator $\mathcal{G}^\varepsilon$ have to be replaced by the ensembles $\{u_{n,h}^0\}_{n=1}^N$ obtained by the finite element discretization of the homogenized operator $\mathcal{G}_h^0$. Moreover Lemma 2 for the error e has to be replaced by Lemma 5 for the error $\tilde{e}$. ∎

Applying Propositions 1 and 2, we finally prove Theorem 1.

Proof of Theorem 1. An application of the triangle inequality yields

$$\mathbb{E}[\left\|u_N^\varepsilon - u_{N,h}^0\right\|] \leq \mathbb{E}[\left\|u_N^\varepsilon - u_N^0\right\|] + \mathbb{E}[\left\|u_N^0 - u_{N,h}^0\right\|].$$

The two addends can be bounded applying Propositions 1 and 2, thus obtaining the desired result for $C = \max\{K_1, K_2\}$. ∎

3.2. Convergence of the posterior distributions. In this section, we give the proof of Theorem 2, i.e., the convergence of the discrete posterior measures μ^ε to μ_h^0 introduced in (19) as $\varepsilon, h \to 0$. Let $u^* \in \mathbb{R}^M$ and let $B_R(u^*)$ be the ball of radius R centered in u^* with respect to the norm $\|\cdot\|_s$ with $s \in [1, \infty]$. Due to the discrete nature of these distributions, we study convergence with respect to the Wasserstein metrics, for which we report its standard definition in the metric spaces $(B_R(u^*), \|\cdot\|_s)$, which can be found, e.g., in [17].

DEFINITION 2. *Let* μ *and* ν *be two probability measures on the metric space* $(B_R(u^*), \|\cdot\|_s)$. *The Wasserstein distance between* μ *and* ν *is defined for all* $p \in [1, \infty)$ *as*

$$W_{p,s}(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{B_R(u^*) \times B_R(u^*)} \|u - v\|_s^p \, d\gamma(u, v) \right)^{1/p}, \tag{41}$$

where $\Gamma(\mu, \nu)$ *denotes the collection of all joint distributions on* $B_R(u^*) \times B_R(u^*)$ *with marginals* μ *and* ν *on the first and second factors respectively.*

Remark 4. If μ and ν are two discrete distributions on finite state spaces, respectively, $\Omega_1 = \{u_1, \ldots, u_{K_1}\}$ and $\Omega_2 = \{v_1, \ldots, v_{K_2}\}$ included in $B_R(u^*)$, then (41) can be written as

$$W_{p,s}(\mu, \nu) = \left(\inf_{\gamma \in \mathbb{R}^{K_1 \times K_2}} \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \|u_i - v_j\|_s^p \, \gamma_{ij} \right)^{1/p}, \tag{42}$$

where the matrix γ has to satisfy the following constraints,

$$\sum_{j=1}^{K_2} \gamma_{ij} = \mu(u_i) \quad \text{for all } i = 1, \ldots K_1, \qquad \sum_{i=1}^{K_1} \gamma_{ij} = \nu(v_j) \quad \text{for all } j = 1, \ldots K_2. \tag{43}$$

We now show that the distance $W_{1,2}$ is bounded by the distance induced by the ensemble norm defined in (18). This result will be crucial later to prove Theorem 1.

LEMMA 6. *Let $u_1 = \{u_1^{(j)}\}_{j=1}^J$, $u_2 = \{u_2^{(j)}\}_{j=1}^J$ be two ensembles of particles and let μ_1, μ_2 be the corresponding distributions defined as the sum of Dirac masses*

$$\mu_1 = \frac{1}{J}\sum_{j=1}^J \delta_{u_1^{(j)}}, \qquad \mu_2 = \frac{1}{J}\sum_{j=1}^J \delta_{u_2^{(j)}}.$$

Then for all $s \in [1,\infty]$ and $p \in [1,\infty)$ it holds

$$W_{p,s}(\mu_1,\mu_2) \le \left(\frac{1}{J}\sum_{j=1}^J \left\|u_1^{(j)} - u_2^{(j)}\right\|_s^p\right)^{\frac{1}{p}}$$

and, in particular,

$$W_{1,2}(\mu_1,\mu_2) \le \|u_1 - u_2\|.$$

Proof. Take γ^* defined as

$$\gamma^*(u_1^{(j)}, u_2^{(i)}) = \begin{cases} \frac{1}{J} & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

which satisfies the constraints (43), and note that

$$\sum_{j=1}^J\sum_{i=1}^J \left\|u_1^{(j)} - u_2^{(i)}\right\|_s^p \gamma^*(u_1^{(j)}, u_2^{(i)}) = \frac{1}{J}\sum_{j=1}^J \left\|u_1^{(j)} - u_2^{(j)}\right\|_s^p.$$

Therefore, by the definition of Wasserstein distance for discrete distributions on finite spaces (42), we deduce that

$$W_{p,s}(\mu_1,\mu_2) \le \left(\frac{1}{J}\sum_{j=1}^J \left\|u_1^{(j)} - u_2^{(j)}\right\|_s^p\right)^{\frac{1}{p}},$$

which is the desired result. Finally, taking $p = 1$ and $s = 2$ and recalling the ensemble norm defined in (18), we obtain the second inequality. ∎

We now analyze the relationship between the weak L^1 convergence introduce in Definition 1 and the convergence with respect to the expectation of the Wasserstein distance for random probability measures. In particular, we prove that the latter implies the former, which was already proved in [17] for nonrandom measures. Here, we extend the result to random probability measures. The proof of the following lemma is given in the appendix.

LEMMA 7. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let the sequence $\{\mu_n\}_{n\in\mathbb{N}}$ and μ be random probability measures on the metric space $(B_R(u^*), \|\cdot\|_s)$ dependent on the random variable ξ on $(\Omega, \mathcal{A}, P)$. If*

$$\mathbb{E}_\xi[W_{1,s}(\mu_n,\mu)] \to 0,$$

then $\mu_n \xrightarrow{L^1} \mu$.

We can now complete the proof of Theorem 2.

Proof of Theorem 2. Applying Lemma 6 and due to Theorem 1, we deduce that for $\varepsilon, h \to 0$ it holds

$$\mathbb{E}[W_{1,2}(\mu^\varepsilon, \mu_h^0)] \to 0.$$

Note that the only difference in the update step of the EnKF when used for a point estimate and in the Bayesian framework is that Γ is replaced by $\Delta^{-1}\Gamma$, where $\Delta = 1/N$. The constants of the proof of Theorem 1 depend on $\|\Gamma^{-1}\|_2$, which is now replaced by $\|(\Delta^{-1}\Gamma)^{-1}\|_2$, which can be bounded by $\|\Gamma^{-1}\|_2$ as

$$\left\|(\Delta^{-1}\Gamma)^{-1}\right\|_2 = \Delta \left\|\Gamma^{-1}\right\|_2 \leq \left\|\Gamma^{-1}\right\|_2.$$

Finally, applying Lemma 7, we obtain the desired result. ∎

4. Modeling error. In this section, we consider the effects of model misspecification due to the homogenization and discretization error. All the results presented in section 3 deal with the asymptotic case $h, \varepsilon \to 0$, which is unrealistic in applications. Let us recall that the original inverse problem involves predicting the exact unknown u^* from observations originated by the model

$$y = \mathcal{G}^\varepsilon(u^*) + \eta, \tag{44}$$

where $\eta \sim \mathcal{N}(0, \Gamma)$ is the noise. Since evaluating $\mathcal{G}^\varepsilon$ is too expensive and in many applications unfeasible, we wish to employ the cheaper forward operator $\mathcal{G}_h^0$. Hence, we rewrite (44) as

$$y = \mathcal{G}_h^0(u^*) + \mathcal{E}(u^*) + \eta, \tag{45}$$

where

$$\mathcal{E}(u^*) \coloneqq \mathcal{G}^\varepsilon(u^*) - \mathcal{G}_h^0(u^*).$$

The quantity $\mathcal{E}(u^*)$ represents the error introduced by misspecification of the forward model. Equation (45) shows that the observed data y can be seen as data originating from the discrete homogenized model which are affected by two sources of errors, the original noise and the modeling error. This formulation of modeling error was originally presented in [6], and then applied to multiscale inverse problems in [3]. Following [3, 6], we assume that the modeling error is a Gaussian random variable independent of the noise η, so that $\mathcal{E} \sim \mathcal{N}(m, \Sigma)$ for all u, and write

$$y = \mathcal{G}_h^0(u^*) + m + \zeta + \eta, \tag{46}$$

where $\zeta \sim \mathcal{N}(0, \Sigma)$. There is no theoretical guarantee for the modeling error to be distributed as a Gaussian in this framework. Nevertheless, it has been shown in [13] that in the one-dimensional case a Gaussian assumption can be employed effectively for the modeling error, thus partially justifying our choice. Then we define

$$\tilde{y} = y - m \qquad \text{and} \qquad \tilde{\eta} = \eta + \zeta \sim \mathcal{N}(0, \Gamma + \Sigma)$$

and, from (46), we obtain

$$\tilde{y} = \mathcal{G}_h^0(u^*) + \tilde{\eta}. \tag{47}$$

Therefore, if the mean m and covariance Σ of the modeling error are known, a more reliable approximation of the unknown u^* can be obtained applying the EnKF to

(47). The modeling error distribution, by assumption fully determined by its mean and covariance, is approximated offline. We sample $N_{\mathcal{E}}$ unknowns $\{u_i\}_{i=1}^{N_{\mathcal{E}}}$ from μ_0 and, for all $i = 1, \dots, N_{\mathcal{E}}$, we apply both the forward operators $\mathcal{G}^{\varepsilon}(u_i)$ and $\mathcal{G}_h^0(u_i)$. Then we compute

$$\mathcal{E}_i = \mathcal{G}^{\varepsilon}(u_i) - \mathcal{G}_h^0(u_i),$$

and the mean m and the covariance Σ are obtained as the empirical mean and covariance of the sample $\{\mathcal{E}_i\}_{i=1}^{N_{\mathcal{E}}}$. This procedure is computationally involved due to the multiple evaluations of $\mathcal{G}^{\varepsilon}$, but it has to be performed only once and can then be applied to different sets of observations and true values u^*. Let us also remark that on the one hand, due to the theory of homogenization, the modeling error can be considered negligible when ε is very small, and the expensive estimation of $\mathcal{E}$ may not be necessary. On the other hand, when ε is larger, the homogenized equation does not provide a good approximation of the multiscale problem, and an estimation of $\mathcal{E}$ is required. One may rightfully argue that in the case $\varepsilon = \mathcal{O}(1)$, it is possible to evaluate the forward operator $\mathcal{G}^{\varepsilon}$ without a large computational effort. Hence, the techniques presented in this section are relevant for midrange values of ε, for which $\mathcal{E}$ is significant with respect to the noise η. Moreover, we remarked in practice via numerical experiments that a small number $N_{\mathcal{E}}$ can be employed to obtain a satisfactory approximation of the modeling error. A theoretical justification of this property is provided by Theorems 3 and 4.

In order to obtain a more reliable approximation of the distribution of the modeling error, we can follow a dynamic approach based on the estimation of the mean m and the covariance Σ online, i.e., during the run of the EnKF algorithm. This methodology has been developed in [5], and it consists of splitting the EnKF run on $\mathcal{L}$ levels, thus obtaining a new estimation of the modeling error sequentially at the end of each level. In practice, given a prior μ_0, and initializing $\mu_0^0 \equiv \mu_0$ and $\ell = 0$, the procedure can be algorithmically summarized as

1. approximate the distribution $\nu^{\ell} = \mathcal{N}(m^{\ell}, \Sigma^{\ell})$ of the modeling error with samples of $\mathcal{E}$ obtained from μ_0^{ℓ};
2. run the EnKF corrected by the modeling error ν^{ℓ} for N^{ℓ} steps and obtain the discrete approximation of the posterior

$$\mu_{N^{\ell}}^{\ell} = \frac{1}{J} \sum_{j=1}^{J} \delta_{u_n^{\ell(j)}};$$

3. set $\mu_0^{\ell+1} = \mu_{N^{\ell}}^{\ell}$, $\ell \leftarrow \ell + 1$ and if $\ell < \mathcal{L}$ return to 1.

This approach provides indeed a better approximation of the modeling error as instead of taking the samples from the prior distribution, they are drawn from distributions which are progressively closer to the true posterior. On the other hand, this procedure has to be done online and it is computationally expensive because it requires the resolution of $N_{\mathcal{E}} = \sum_{\ell=1}^{\mathcal{L}} N_{\mathcal{E}}^{\ell}$ full multiscale problems.

Finally, we are interested in studying whether the simple offline method for estimating the modeling error provides indeed a good approximation. In this direction, we give in Theorems 3 and 4 a criterion on how to choose the number $N_{\mathcal{E}}$ of full multiscale problems which have to be solved in order to have a reliable approximation of the true mean m^* and covariance Σ^* of the modeling error with respect to ε and h. Before stating Theorems 3 and 4, let us recall Hoeffding's inequality, which will be used in the proofs. Let $\{Y_i\}_{i=1}^{N}$ be independent random variables with values in $[a, b]$,

and let $\bar{Y}$ be the sample average of $\{Y_i\}_{i=1}^N$. Then, the Hoeffding's inequality states that for all $\eta \in \mathbb{R}$ it holds

$$\mathbb{P}(\left|\bar{Y} - \mathbb{E}[Y]\right| \geq \eta) \leq 2\exp\left\{-\frac{2\eta^2 N}{(b-a)^2}\right\}.$$

THEOREM 3. *Let $\alpha \in (0,1)$, $\eta > 0$, and $C_{\mathcal{E}} = \max\{K, \tilde{K}\}$, where K and $\tilde{K}$ are the constants of Lemmas 2 and 5. Let $\{\mathcal{E}_i\}_{i=1}^{N_\mathcal{E}} \subset \mathbb{R}^L$ be given by*

$$\mathcal{E}_i = \mathcal{G}^\varepsilon(u_i) - \mathcal{G}_h^0(u_i) \qquad \text{for all } i = 1, \ldots, N_\mathcal{E}$$

for a sample of realizations $\{u_i\}_{i=1}^{N_\mathcal{E}}$ drawn from a prior distribution μ_0, let m be the sample mean of $\{\mathcal{E}_i\}_{i=1}^{N_\mathcal{E}}$, and $m^* = \mathbb{E}[\mathcal{E}_i]$. If

$$N_\mathcal{E} \geq 4C_\mathcal{E}^2 \frac{L}{\eta^2} \log\left(\frac{2L}{\alpha}\right)\left[\varepsilon^2 + h^{2(s+1)}\right],$$

where s is given by Lemma 5, then

$$\mathbb{P}\left(\|m - m^*\|_2 \leq \eta\right) \geq 1 - \alpha.$$

Proof. First, note that the modeling error is indeed bounded by Lemmas 2 and 5, we have for each $i = 1, \ldots, N_\mathcal{E}$

$$\|\mathcal{E}_i\|_2 = \left\|\mathcal{G}^\varepsilon(u_i) - \mathcal{G}_h^0(u_i)\right\|_2 \leq \left\|\mathcal{G}^\varepsilon(u_i) - \mathcal{G}^0(u_i)\right\|_2 + \left\|\mathcal{G}^0(u_i) - \mathcal{G}_h^0(u_i)\right\|_2 \leq K\varepsilon + \tilde{K}h^{s+1},$$

so each component $(\mathcal{E}_i)_l$, for $l = 1, \ldots, L$, is bounded by the same constant

$$|(\mathcal{E}_i)_l| \leq \|\mathcal{E}_i\|_2 \leq K\varepsilon + \tilde{K}h^{s+1} \leq C_\mathcal{E}(\varepsilon + h^{s+1}). \tag{48}$$

Observe that if

$$|m_l - m_l^*| \leq \frac{\eta}{\sqrt{L}} \qquad \text{for each } l = 1, \ldots, L, \tag{49}$$

then

$$\|m - m^*\|_2 = \left(\sum_{l=1}^L |m_l - m_l^*|^2\right)^{\frac{1}{2}} \leq \eta, \tag{50}$$

which implies that

$$\mathbb{P}(\|m - m^*\|_2 \leq \eta) \geq \mathbb{P}\left(|m_l - m_l^*| \leq \frac{\eta}{\sqrt{L}} \quad \forall\, l = 1, \ldots, L\right). \tag{51}$$

Using (48) and applying Hoeffding's inequality we have

$$\begin{aligned}\mathbb{P}\left(|m_l - m_l^*| \geq \frac{\eta}{\sqrt{L}}\right) &\leq 2\exp\left\{-\frac{2\eta^2 N_\mathcal{E}}{4LC_\mathcal{E}^2(\varepsilon + h^{s+1})^2}\right\} \\ &\leq 2\exp\left\{-\frac{\eta^2 N_\mathcal{E}}{4LC_\mathcal{E}^2(\varepsilon^2 + h^{2(s+1)})}\right\}.\end{aligned} \tag{52}$$

Define the events $A_l = \{|m_l - m_l^*| \leq \frac{\eta}{\sqrt{L}}\}$ for each $l = 1, \ldots, L$, then we have

$$\mathbb{P}\left(|m_l - m_l^*| \leq \frac{\eta}{\sqrt{L}} \quad \forall\, l = 1, \ldots, L\right) = \mathbb{P}\left(\bigcap_{l=1}^L A_l\right),$$

and, applying De Morgan's laws and the union bound, we obtain

$$\mathbb{P}\left(\bigcap_{l=1}^{L} A_l\right) = 1 - \mathbb{P}\left(\left(\bigcap_{l=1}^{L} A_l\right)^C\right) = 1 - \mathbb{P}\left(\bigcup_{l=1}^{L} A_l^C\right) \geq 1 - \sum_{l=1}^{L} \mathbb{P}(A_l^C). \tag{53}$$

Therefore, thanks to (51), (52), and (53), we have

$$\begin{aligned}\mathbb{P}(\|m - m^*\|_2 \leq \eta) &\geq 1 - L \max_{l=1,\dots,L} \mathbb{P}\left(|m_l - m_l^*| \geq \frac{\eta}{\sqrt{L}}\right) \\ &\geq 1 - 2L \exp\left\{-\frac{\eta^2 N_{\mathcal{E}}}{4LC_{\mathcal{E}}^2(\varepsilon^2 + h^{2(s+1)})}\right\},\end{aligned} \tag{54}$$

and if $N_{\mathcal{E}}$ satisfies the hypothesis we obtain the desired result. ∎

THEOREM 4. *Let $\alpha \in (0,1)$, $\eta > 0$ and $C_{\mathcal{E}} = \max\{K, \tilde{K}\}$, where K and $\tilde{K}$ are the constants of Lemmas 2 and 5. Let $\{\mathcal{E}_i\}_{i=1}^{N_{\mathcal{E}}} \subset \mathbb{R}^L$ be given by*

$$\mathcal{E}_i = \mathcal{G}^{\varepsilon}(u_i) - \mathcal{G}_h^0(u_i) \qquad \text{for all } i = 1, \dots, N_{\mathcal{E}}$$

for a sample of realizations $\{u_i\}_{i=1}^{N_{\mathcal{E}}}$ drawn from a prior distribution μ_0, let m and Σ be the sample mean and covariance of $\{\mathcal{E}_i\}_{i=1}^{N_{\mathcal{E}}}$ and $m^* = \mathbb{E}[\mathcal{E}_i]$, and

$$\Sigma^* = \mathbb{E}[(\mathcal{E}_i - m^*)(\mathcal{E}_i - m^*)^T].$$

If

$$N_{\mathcal{E}} \geq \widehat{C} C_{\mathcal{E}}^4 \frac{L^2}{\eta^2} \log\left(\frac{2L(L+1)}{\alpha}\right)\left[\varepsilon^4 + h^{4(s+1)}\right],$$

where s is given by Lemma 5 and $\widehat{C}$ is specified in the proof, then

$$\mathbb{P}\left(\|\Sigma - \Sigma^*\|_2 \leq \eta\right) \geq 1 - \alpha.$$

Proof. First, repeating verbatim the first part of the proof of Theorem 3 we have

$$|(\mathcal{E}_i)_l| \leq \|\mathcal{E}_i\|_2 \leq K\varepsilon + \tilde{K} h^{s+1} \leq C_{\mathcal{E}}(\varepsilon + h^{s+1}). \tag{55}$$

Let us now rewrite the covariance matrix Σ^* and its estimator Σ as

$$\Sigma^* = \mathbb{E}[(\mathcal{E}_i - m^*)(\mathcal{E}_i - m^*)^T] = \mathbb{E}[\mathcal{E}_i \mathcal{E}_i^T] - m^*(m^*)^T \tag{56}$$

and

$$\Sigma = \frac{1}{N_{\mathcal{E}}} \sum_{i=1}^{N_{\mathcal{E}}} (\mathcal{E}_i - m)(\mathcal{E}_i - m)^T = \frac{1}{N_{\mathcal{E}}} \sum_{i=1}^{N_{\mathcal{E}}} \mathcal{E}_i \mathcal{E}_i^T - mm^T. \tag{57}$$

Then by the triangle inequality it holds

$$\|\Sigma - \Sigma^*\|_2 \leq \left\| \frac{1}{N_{\mathcal{E}}} \sum_{i=1}^{N_{\mathcal{E}}} \mathcal{E}_i \mathcal{E}_i^T - \mathbb{E}[\mathcal{E}_i \mathcal{E}_i^T] \right\|_2 + \left\| mm^T - m^*(m^*)^T \right\|_2, \tag{58}$$

and due to the elementary inequality $\|ab^T\|_2 \leq \|a\|_2 \|b\|_2$ and the bound (55) we have

$$\left\| mm^T - m^*(m^*)^T \right\|_2 = \left\| (m - m^*)m^T + m^*(m - m^*)^T \right\|_2 \leq 2C_{\mathcal{E}}(\varepsilon + h^{s+1}) \|m - m^*\|_2, \tag{59}$$

which implies

$$\|\Sigma - \Sigma^*\|_2 \le \left\| \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} \mathcal{E}_i \mathcal{E}_i^T - \mathbb{E}[\mathcal{E}_i \mathcal{E}_i^T] \right\|_2 + 2C_\mathcal{E}(\varepsilon + h^{s+1}) \|m - m^*\|_2 . \tag{60}$$

Therefore we obtain

$$\mathbb{P}\left(\|\Sigma - \Sigma^*\|_2 \le \eta\right) \ge \mathbb{P}\Bigg(\left\| \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} \mathcal{E}_i \mathcal{E}_i^T - \mathbb{E}[\mathcal{E}_i \mathcal{E}_i^T] \right\|_2 \le \frac{\eta}{2}, \quad \|m - m^*\|_2 \le \frac{\eta}{4C_\mathcal{E}(\varepsilon + h^{s+1})} \Bigg), \tag{61}$$

which yields

$$\mathbb{P}\left(\|\Sigma - \Sigma^*\|_2 \le \eta\right) \ge 1 - \mathbb{P}\left(\left\| \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} \mathcal{E}_i \mathcal{E}_i^T - \mathbb{E}[\mathcal{E}_i \mathcal{E}_i^T] \right\|_2 \ge \frac{\eta}{2} \right) - \mathbb{P}\left(\|m - m^*\|_2 \ge \frac{\eta}{4C_\mathcal{E}(\varepsilon + h^{s+1})} \right), \tag{62}$$

and then we bound the two terms in the right-hand side separately. By (54) in the proof of Theorem 3 we first have

$$\mathbb{P}\left(\|m - m^*\|_2 \ge \frac{\eta}{4C_\mathcal{E}(\varepsilon + h^{s+1})} \right) \le 2L \exp\left\{ -\frac{\eta^2 N_\mathcal{E}}{32 L C_\mathcal{E}^4 (\varepsilon + h^{s+1})^4} \right\}. \tag{63}$$

Then, similarly to the last part of the proof of Theorem 3, since $\|\cdot\|_2 \le \|\cdot\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm, we have

$$\mathbb{P}\left(\left\| \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} \mathcal{E}_i \mathcal{E}_i^T - \mathbb{E}[\mathcal{E}_i \mathcal{E}_i^T] \right\|_2 \ge \frac{\eta}{2} \right) \le L^2 \max_{j,k=1,\dots,L} \mathbb{P}\left(\left| \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} (\mathcal{E}_i)_j (\mathcal{E}_i)_k - \mathbb{E}[(\mathcal{E}_i)_j (\mathcal{E}_i)_k] \right| \ge \frac{\eta}{2L} \right), \tag{64}$$

and applying Hoeffding's inequality to the random variables $(\mathcal{E}_i)_j (\mathcal{E}_i)_k$ for all $j, k = 1, \dots, L$, which are bounded by $C_\mathcal{E}^2 (\varepsilon + h^{s+1})^2$ due to (55), we obtain

$$\mathbb{P}\left(\left\| \frac{1}{N_\mathcal{E}} \sum_{i=1}^{N_\mathcal{E}} \mathcal{E}_i \mathcal{E}_i^T - \mathbb{E}[\mathcal{E}_i \mathcal{E}_i^T] \right\|_2 \ge \frac{\eta}{2} \right) \le 2L^2 \exp\left\{ -\frac{\eta^2 N_\mathcal{E}}{8 L^2 C_\mathcal{E}^4 (\varepsilon + h^{s+1})^4} \right\}. \tag{65}$$

Finally, (63) and (65) together with (62) imply

$$\mathbb{P}\left(\|\Sigma - \Sigma^*\|_2 \le \eta\right) \ge 1 - 2L(L+1) \exp\left\{ -\frac{\eta^2 N_\mathcal{E}}{\widehat{C} L^2 C_\mathcal{E}^4 (\varepsilon^4 + h^{4(s+1)})} \right\}, \tag{66}$$

where $\widehat{C} = 256$ and if $N_\mathcal{E}$ satisfies the hypothesis we obtain the desired result. ∎

Remark 5. Note that, in Theorems 3 and 4, as expected, the number $N_{\mathcal{E}}$ of full multiscale problems tends to infinity if we require no error between the sample and the true mean and covariance ($\eta \to 0$) or certainty that the error is below a certain value ($\alpha \to 0$). Moreover, observe that for any given accuracy the number of samples required $N_{\mathcal{E}}$ is an increasing function of ε and h, so that if the model $\mathcal{G}_h^0$ is a good approximation of $\mathcal{G}$, thus computationally expensive, then only a few samples are needed. In particular, notice that in order to obtain a good approximation of the true mean, the number of full multiscale problems is

$$N_{\mathcal{E}} = \mathcal{O}\left(\eta^{-2}\log(\alpha^{-1})\left(\varepsilon^2 + h^{2(s+1)}\right)\right), \tag{67}$$

while to have a reliable approximation of the covariance matrix it is required that

$$N_{\mathcal{E}} = \mathcal{O}\left(\eta^{-2}\log(\alpha^{-1})\left(\varepsilon^4 + h^{4(s+1)}\right)\right). \tag{68}$$

5. Numerical experiments. In this section, using the setting of [3], we present some numerical experiments to illustrate the iterative ensemble Kalman method to solve multiscale inverse problems.

Let D be a bounded open domain. We consider a class of parametrized multiscale locally periodic tensors of the type $A^\varepsilon_{\sigma^*}(x) = A(\sigma^*(x), x/\varepsilon)$, where $\sigma^*\colon D \to \mathbb{R}$. We assume to know the map $(t,x) \to A(t, x/\varepsilon)$ for all $x \in D$ and $t \in \mathbb{R}$ and we want to estimate the function σ^* given measurements computed from the model

$$\begin{cases} -\nabla\cdot(A^\varepsilon_{\sigma^*}\nabla p^\varepsilon) = 0 & \text{in } D,\\ p^\varepsilon = g & \text{on } \partial D. \end{cases} \tag{69}$$

Remark 6. Note that the theory has been developed for Dirichlet homogeneous boundary conditions, but it can be applied to the nonhomogeneous case by considering an extension of the function at the boundary and slightly modifying the PDE. For more details we refer to [16, Remark 8.10].

For the unknown σ^* we consider the following admissible set

$$\Sigma = \{\sigma \in L^\infty(D)\colon \sigma^- \le \sigma(x) \le \sigma^+\},$$

where σ^- and σ^+ are two given values.

The measurements, which we take into account, are the integrals of the normal flux multiplied by some functions with compact support in a portion of the boundary of the domain. More precisely, we consider $I \in \mathbb{N}$ disjoint portions of D, which we denote by $\Gamma_i \in \partial D$, $i = 1,\dots,I$, $\Gamma_i \cap \Gamma_j = \emptyset$ for $i \neq j$, and I functions $\varphi_i \in H^{1/2}(\partial D)$ with compact support supp $(\varphi_i) \subset \Gamma_i$ for all $i = 1,\dots,I$. Moreover, we solve (69) for $K \in \mathbb{N}$ Dirichlet data g_k, $k = 1,\dots,K$, and we denote by p^ε_k the solution of the problem. Let $\Lambda_{A^\varepsilon_\sigma}\colon H^{1/2}(\partial D) \to H^{-1/2}(\partial D)$ be the operator which maps the Dirichlet data g to the normal flux of the solution p^ε of (69),

$$\Lambda_{A^\varepsilon_\sigma} g = A^\varepsilon_\sigma \nabla p^\varepsilon \cdot \nu, \tag{70}$$

where ν is the exterior unit normal vector to ∂D. Then we define the multiscale operator $\mathcal{F}^\varepsilon\colon \Sigma \to \mathbb{R}^L$, where $L = IK$ by components

$$\mathcal{F}^\varepsilon(\sigma)_{ik} = \mathcal{F}^\varepsilon(\sigma)_l = \left\langle \Lambda_{A^\varepsilon_\sigma} g_k, \varphi_i \right\rangle_{H^{-1/2}(\partial D), H^{1/2}(\partial D)}, \qquad i = 1,\dots,I, \quad k = 1,\dots,K, \tag{71}$$

which, with an abuse of notation, can be written

$$\mathcal{F}^{\varepsilon}(\sigma)_{ik} = \int_{\Gamma_i} A^{\varepsilon} \nabla p_k^{\varepsilon} \cdot \nu \varphi_i ds. \tag{72}$$

The final vector of observations y is given by the sum of the operator $\mathcal{F}^{\varepsilon}$ and a noise

$$y = \mathcal{F}^{\varepsilon}(\sigma^*) + \eta,$$

where $\eta \sim \mathcal{N}(0, \Gamma)$ and Γ is a given symmetric positive definite covariance matrix, which, in our experiments, is a multiple of the identity $\Gamma = \gamma^2 I$ and γ is a given value. Observations are computed with a refined FEM with mesh size $h_{\text{obs}} \ll \varepsilon$, while the homogenized version of problem (69) is solved using a macro mesh size $h \gg h_{\text{obs}}$. We call $\mathcal{T}_h$ the macro triangulation and N_h the total number of nodes defining $\mathcal{T}_h$. We assume that the prior distribution for the discretization of the unknown σ^* on the macro triangulation $\mathcal{T}_h$ is given by $\mathcal{N}(\sigma_0, C)$, where σ_0 is a given discretization of a function in Σ and $C \in \mathbb{R}^{N_h \times N_h}$ is defined by

$$C_{ij} = \delta \exp\left(-\frac{\|x_i - x_j\|_2}{\lambda}\right),$$

where $\delta, \lambda \in \mathbb{R}^+$ and $\{x_i\}_{i=1}^{N_h}$ are the nodes of the macro triangulation $\mathcal{T}_h$. The parameter λ is a correlation length that describes how the values at different positions of the functions supported by the prior measure are related, while the parameter δ is an amplitude scaling factor. Regarding the prior modeling, we need to take into account that even if in the homogenized problem the coarse and fine scales have been separated, functions drawn from the prior distribution on the coarse scale can exhibit multiple scales, including the fine scale of our multiscale model, depending on the rate of decay of the prior covariance. This issue can thus be controlled by setting the parameters δ and λ. Even though this does not ensure a clear separation between coarse and fine scales, our numerical results illustrate that it is sufficient in practice.

In order to reduce the dimensionality of the unknown we use a truncated Karhunen–Loève expansion. Any sample from the prior distribution $\mathcal{N}(\sigma_0, C)$ can be represented as

$$\sigma = \sigma_0 + \sum_{m=1}^{N_h} \sqrt{\lambda_m} u_m \psi_m, \tag{73}$$

where $\{\psi_m\}_{m=1}^{N_h}$ is an orthonormal set of eigenvectors of C with corresponding eigenvalues $\{\lambda_m\}_{m=1}^{N_h}$ in decreasing order, and $\{u_m\}_{m=1}^{N_h}$ is an i.i.d. sequence with $u_m \sim \mathcal{N}(0,1)$. Note that the Karhunen–Loève expansion works also in the infinite-dimensional setting, where $\sigma_0 \in \Sigma$, C is a covariance operator, and $\{\lambda_m, \psi_m\}_{m=1}^{\infty}$ is an orthonormal set of eigenvalues-eigenfunctions with respect to the scalar product in $L^2(D)$. Then the truncated Karhunen–Loève expansion of the discretization of σ consists of taking the first M components of the series in (73),

$$\sigma \simeq \sigma_0 + \sum_{m=1}^{M} \sqrt{\lambda_m} u_m \psi_m, \tag{74}$$

and the actual unknown becomes the vector $u \in \mathbb{R}^M$, whose components are the coefficients u_m in (74). Then we define the multiscale forward operator $\mathcal{G}^{\varepsilon} : \mathbb{R}^M \to \mathbb{R}^L$

as the composition of $\mathcal{F}^\varepsilon$ with the truncated Karhunen–Loève expansion

$$\mathcal{G}^\varepsilon(u) = \mathcal{F}^\varepsilon\left(\sigma_0 + \sum_{m=1}^{M} \sqrt{\lambda_m} u_m \psi_m\right).$$

In the iterative ensemble Kalman method we do not compute the exact solution of problem (69), but we solve its homogenized version numerically using the macro triangulation $\mathcal{T}_h$, therefore, we obtain the homogenized discrete solution p_h^0. The problem is solved applying the FE-HMM, which is described in [1, 4]. Hence, analogously to the multiscale case, we define the discrete homogenized operator $\mathcal{F}_h^0 \colon \Sigma \to \mathbb{R}^L$ with an abuse of notation as

$$\mathcal{F}_h^0(\sigma)_l = \mathcal{F}_h^0(\sigma)_{ik} = \int_{\Gamma_i} A^0 \nabla p_{h_k}^0 \cdot \nu \varphi_i ds, \qquad i = 1, \dots, I, \quad k = 1, \dots, K, \tag{75}$$

and the discrete homogenized forward operator $\mathcal{G}_h^0 \colon \mathbb{R}^M \to \mathbb{R}^L$, which is actually used in the algorithm, as

$$\mathcal{G}_h^0(u) = \mathcal{F}_h^0\left(\sigma_0 + \sum_{m=1}^{M} \sqrt{\lambda_m} u_m \psi_m\right).$$

Finally, we call u_{EnKF} the solution of the iterative ensemble Kalman algorithm and the estimated σ_{EnKF} is obtained from the truncated Karhunen–Loève expansion

$$\sigma_{\mathrm{EnKF}} = \sigma_0 + \sum_{m=1}^{M} \sqrt{\lambda_m} u_{\mathrm{EnKF}_m} \psi_m.$$

5.1. Data. In the numerical results presented in the following section the computational domain is the unit square

$$D = (0,1)^2 \subset \mathbb{R}^2.$$

For the discretization parameters we set $\varepsilon = 1/64$ and $h_{\mathrm{obs}} = 1/4096$ and for the forward homogenized problem we use a macro mesh size $h = 1/32$, which is much larger than h_{obs} and reduces the computational cost significantly. We solve the problem for $K = 3$ Dirichlet conditions $\{g_k\}_{k=1}^3$ and $g_k = \sqrt{\mu_k}\vartheta_k$, where $\{(\mu_k, \vartheta_k)\}_{k=1}^3$ are couples of eigenvalues and eigenfunctions of the one dimensional discrete Laplacian operator corresponding to the first $K = 3$ smallest eigenvalues. For each g_k we consider its restriction to the boundary ∂D in order to obtain a Dirichlet condition. These functions are orthonormal with respect to the scalar product in $L^2(D)$ and this ensures that each function gives independent information.

To compute the boundary integrals in (71) and (75), we consider $I = 12$ boundary portions, three for each side of the square D. In particular, for each side, all Γ_i have length equal to 0.2 and they consist of the intervals $(0.1, 0.3), (0.4, 0.6)$, and $(0.7, 0.9)$. The functions $\{\varphi_i\}_{i=1}^{12}$ are hat functions with $\mathrm{supp}(\varphi_i) = \Gamma_i$, which take value 1 at the midpoint and value 0 at the extremes of Γ_i. Then the parameter of the noise, which perturbs the observations, is $\gamma = 0.01$.

Moreover, regarding the prior distribution for the unknown, we consider $\sigma_0 = 0$ and the parameters of the covariance matrices are $\delta = 0.05$ and $\lambda = 0.5$. In the truncated Karhunen–Loève expansion we take $M = 100$. Finally, concerning the ensemble Kalman method, we consider $J = 1000$ particles for each ensemble and 500 iterations.

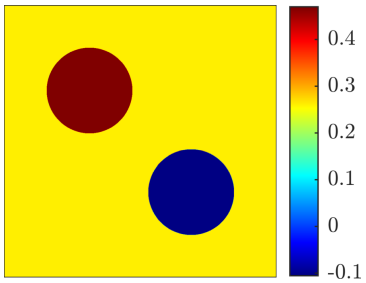

FIG. 1. *Exact unknown σ^* employed for numerical experiments.*

The exact tensor $A^{\varepsilon}_{\sigma^*}$ is given by

$$
\begin{aligned}
a_{11}\left(\sigma^*(x), \frac{x}{\varepsilon}\right) &= e^{\sigma^*(x)}\left(\cos^2\left(\frac{2\pi x_1}{\varepsilon}\right)+1\right)+\cos^2\left(2\pi\frac{x_2}{\varepsilon}\right),\\
a_{12}\left(\sigma^*(x), \frac{x}{\varepsilon}\right) &= 0,\\
a_{21}\left(\sigma^*(x), \frac{x}{\varepsilon}\right) &= 0,\\
a_{22}\left(\sigma^*(x), \frac{x}{\varepsilon}\right) &= e^{\sigma^*(x)}\left(\sin\left(\frac{2\pi x_2}{\varepsilon}\right)+2\right)+\cos^2\left(2\pi\frac{x_1}{\varepsilon}\right),
\end{aligned}
$$

where

$$
\sigma^*(x) = \log(1.3 + 0.3\mathbb{1}_{D_1} - 0.4\mathbb{1}_{D_2})
$$

and

$$
\begin{aligned}
D_1 &= \left\{x=(x_1,x_2)\colon \left(x_1-\frac{5}{16}\right)^2+\left(x_2-\frac{11}{16}\right)^2 \le 0.025\right\},\\
D_2 &= \left\{x=(x_1,x_2)\colon \left(x_1-\frac{11}{16}\right)^2+\left(x_2-\frac{5}{16}\right)^2 \le 0.025\right\}.
\end{aligned}
$$

Figure 1 shows the exact unknown σ^*. Note that σ^* is a noncontinuous function, but, in order to approximate it, we are using a truncated Karhunen–Loève expansion, where the eigenfunctions are smooth.

One can verify that the tensor A^{ε}_{σ} satisfies Assumption 1. In particular, for $\xi \in \mathbb{R}^2$ we have

$$
A^{\varepsilon}_{\sigma}\xi\cdot\xi = a_{1,1}\left(\sigma(x),\frac{x}{\varepsilon}\right)\xi_1^2 + a_{2,2}\left(\sigma(x),\frac{x}{\varepsilon}\right)\xi_2^2 \ge e^{\sigma(x)}(\xi_1^2+\xi_2^2) \ge e^{\sigma_-}\left\|\xi\right\|_2^2.
$$

Moreover, since the EnKF algorithm estimates the coefficients $\{u_m\}_{m=1}^M$ of the truncated Karhunen–Loève expansion, we show that $A^{\varepsilon}(u)\colon \mathbb{R}^M \to L^{\infty}(D,\mathbb{R}^{d\times d})$, which maps u into $A^{\varepsilon}_{\sigma_u}$, is Lipschitz. In fact we first have

$$
\left\|A^{\varepsilon}(u_1)-A^{\varepsilon}(u_2)\right\|_{L^{\infty}(D,\mathbb{R}^{d\times d})} \le \sqrt{13}e^{\sigma^+}\sup_{x\in D}\left|\sigma_{u_1}(x)-\sigma_{u_2}(x)\right|,
$$

then using the truncated Karhunen–Loève expansion and the Cauchy–Schwarz inequality we obtain

$$
\left\|A^{\varepsilon}(u_1)-A^{\varepsilon}(u_2)\right\|_{L^{\infty}(D,\mathbb{R}^{d\times d})} \le \sqrt{13}e^{\sigma^+}\sup_{x\in D}\left(\sum_{m=1}^{M}\lambda_m\psi_m^2(x)\right)^{1/2}\left\|u_1-u_2\right\|_2,
$$

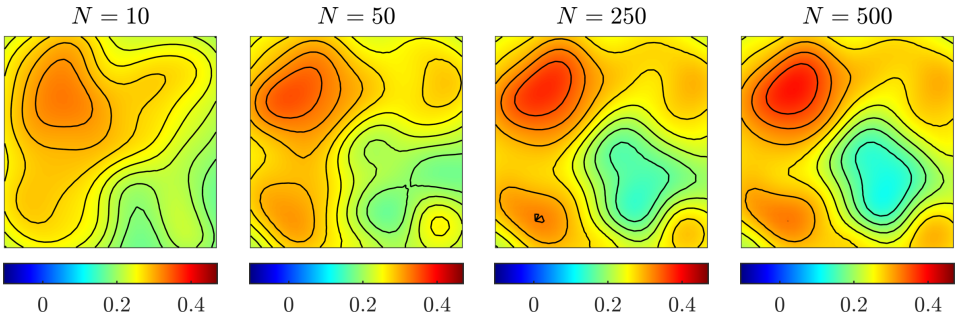

FIG. 2. *EnKF estimation after* $N = \{10, 50, 250, 500\}$ *iterations.*

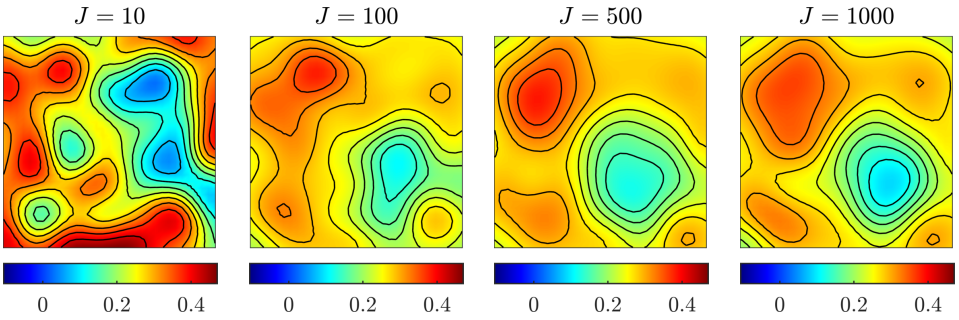

FIG. 3. *EnKF estimation after* $N = 500$ *iterations with ensemble size* $J = \{10, 100, 500, 1000\}$.

which shows that $A^\varepsilon(u)$ is Lipschitz with constant equal to

$$\sqrt{13}e^{\sigma^+} \sup_{x\in D} \left(\sum_{m=1}^{M} \lambda_m \psi_m^2(x) \right)^{1/2}.$$

5.2. Results. We first fix the multiscale parameter $\varepsilon = 1/32$ and the ensemble size $J = 500$ and study the evolution with respect to the number of steps. In Figure 2 we plot the estimation σ_{EnKF} after $10, 50, 250$, and 500 iterations of the ensemble Kalman algorithm. We clearly see that the approximation gets better as the number of iterations increases and that convergence has been reached. In particular, already after $N = 250$ iterations the algorithm seem to have reached convergence. We point out that we obtain a quite good approximation of the real unknown σ^* , indeed we are trying to recover a noncontinuous function in the whole domain given only some observations at the boundary.

We now perform a sensitivity analysis with respect to the ensemble size. In Figure 3 we vary the number of particles J and we compare the results obtained at the end of the algorithm after 500 iterations for $\varepsilon = 1/32$. As expected, the approximation becomes better when the ensemble contains more particles. In particular, note that if the number of particles is too small, e.g., $J = 10$, then the approximation is not satisfying.

Further, we fix the ensemble size $J = 500$ and we perform $N = 500$ iterations of the EnKF for different values of the multiscale parameter. Results, shown in Figure 4, highlight how the approximation becomes worse when ε is bigger, indeed the homogenized problem becomes too different with respect to the multiscale one and, if ε is too big, the solution does not approximate the true unknown.

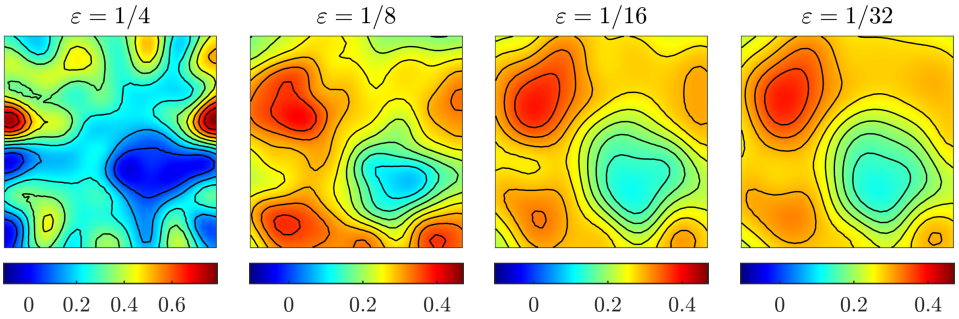

FIG. 4. *EnKF estimation after* $N = 500$ *iterations for the multiscale parameter* $\varepsilon = \{1/4, 1/8, 1/16, 1/32\}$.

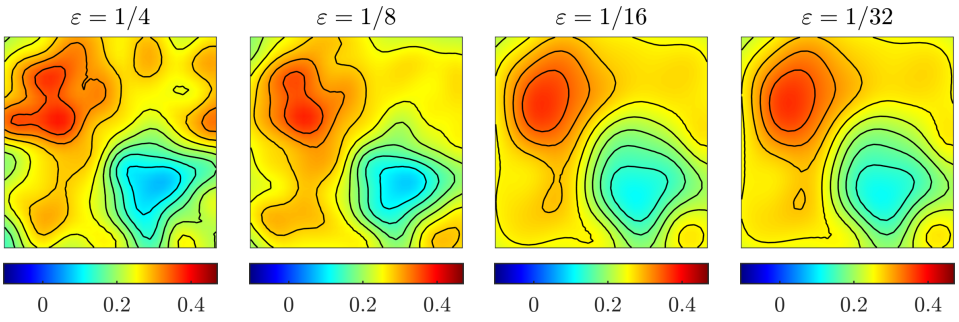

FIG. 5. *EnKF with offline modeling error estimation after* 500 *iterations for the multiscale parameter* $\varepsilon = \{1/4, 1/8, 1/16, 1/32\}$.

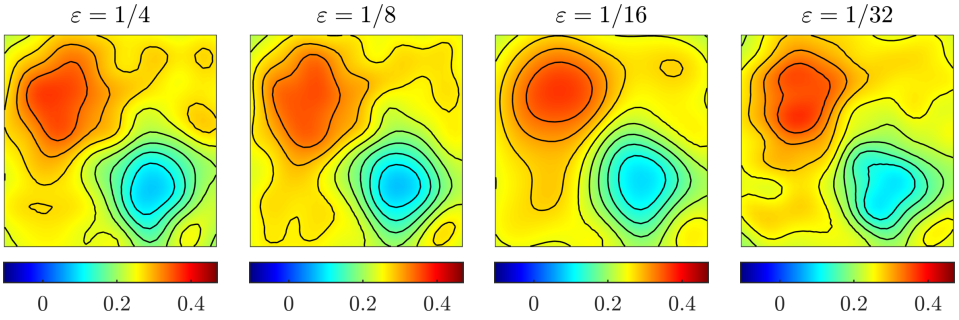

FIG. 6. *EnKF with online iterative modeling error estimation after* 500 *iterations for the multiscale parameter* $\varepsilon = \{1/4, 1/8, 1/16, 1/32\}$.

Moreover, in order to obtain good results even in the case ε is not close to the asymptotic limit $\varepsilon \to 0$, in Figure 5 we apply offline modeling error estimation with $N_{\mathcal{E}} = 20$ and we plot the solution of the inverse problem (47) for different values of the multiscale parameter ε. Comparing these plots with the ones in Figure 4, in particular for $\varepsilon = 1/4$, we observe that the modeling error estimation significantly improves the results.

Finally, in Figure 6 we show the results obtained by applying the ensemble Kalman method with dynamic updating of the modeling error distribution with $\mathcal{L} = 5$ levels,

$N_{\mathcal{E}}^{\ell} = 4$ samples, and $N^{\ell} = 100$ iterations at each level $\ell = 1, \dots, \mathcal{L}$. The number of resolutions of the full multiscale problem is 20 and the total number of iterations is 500, which are equal to the previous approach, where the distribution of the modeling error was approximated offline. Comparing these plots with the ones in Figure 5, we note that updating the distribution of the modeling error dynamically still improves the results.

6. Conclusion. In this paper we analyzed the ensemble Kalman inversion methodology in the context of inverse problems for multiscale elliptic PDEs with tensors highly oscillatory at a scale $\varepsilon \ll 1$. The multiscale algorithm we propose relies on the EnKF, on a surrogate homogenized forward operator, and on numerical homogenization techniques such as the FE-HMM. It guarantees a significant reduction in computational cost for problems which would be otherwise computationally involved or unfeasible. In Theorem 1 we have shown that the ensemble of particles approximating the unknown parameter generated by our multiscale algorithm converges to the ensemble generated by the true model as the small scale parameter ε and the numerical discretization parameter h go to zero. Furthermore in a Bayesian framework, we have shown in Theorem 2 that the discrete probability measure based on the ensemble originating from our multiscale algorithm converges to the measure generated by the true model, again as ε and h go to zero. Hence when $\varepsilon \ll 1$ and the full model is expensive to solve, the multiscale numerical method we propose is both accurate and efficient to recover an unknown parameter in multiscale elliptic PDEs. Moreover, we equipped our method with a technique which allows us to account for the discrepancy between the artificial homogenized surrogate forward model and the true multiscale data, thus alleviating the effects of model misspecification. This technique requires additional offline or online computations involving the numerical solution of the full multiscale problem. The optimal number of such additional solves is quantified in Theorems 3 and 4. In particular, we have proved that the number of solves needed to reach any required accuracy tends to zero when the small scale parameter ε and the numerical discretization parameter h vanish. Hence, we can conclude that accounting for model misspecification is particularly beneficial for midrange values of ε, when a small number of full solves should be computationally affordable. The efficiency and usefulness of the multiscale algorithm have been further demonstrated through a series of numerical experiments.

Appendix.

Proof of Lemma 1. Let $u_1, u_2 \in \mathbb{R}^M$, and $p_1 = \mathcal{S}(u_1)$, $p_2 = \mathcal{S}(u_2)$. From the weak formulations of (22) we get that

$$\int_D \left(A_{u_1} \nabla p_1 - A_{u_2} \nabla p_2\right) \cdot \nabla v = 0 \qquad \text{for all } v \in H_0^1(D),$$

which yields

$$\int_D A_{u_1} (\nabla p_1 - \nabla p_2) \cdot \nabla v = -\int_D (A_{u_1} - A_{u_2}) \nabla p_2 \cdot \nabla v.$$

Then choosing $v = p_1 - p_2$, by the hypotheses on A_u and applying the Hölder inequality, we obtain

$$\alpha_0 \left\|\nabla p_1 - \nabla p_2\right\|^2_{L^2(D;\mathbb{R}^d)} \leq M \left\|u_1 - u_2\right\|_2 \left\|\nabla p_2\right\|_{L^2(D;\mathbb{R}^d)} \left\|\nabla p_1 - \nabla p_2\right\|_{L^2(D;\mathbb{R}^d)},$$

which due to a standard coercivity argument implies

$$\|\nabla p_1 - \nabla p_2\|_{L^2(D;\mathbb{R}^d)} \leq \frac{MC_p}{\alpha_0^2} \|f\|_{L^2(D)} \|u_1 - u_2\|_2 , \tag{76}$$

where C_p is the Poincaré constant associated with the domain D. Hence (76) shows that $\mathcal{S}$ is Lipschitz with constant

$$L_{\mathcal{S}} = \frac{MC_p}{\alpha_0^2} \|f\|_{L^2(D)} .$$

Finally, since $\mathcal{G}$ is the composition of two Lipschitz operators, we deduce that it is also Lipschitz with constant $L_{\mathcal{G}} = L_{\mathcal{O}} L_{\mathcal{S}}$. ∎

Proof of Lemma 2. Let us consider an ensemble $u \in \mathcal{U}_{J,M}$ with particles $u^{(j)} \in \mathbb{R}^M$ for $j = 1, \ldots, J$. For each particle we have

$$\begin{aligned} \left\| \mathcal{G}^\varepsilon(u^{(j)}) - \mathcal{G}^0(u^{(j)}) \right\|_2 &= \left\| \mathcal{O}(\mathcal{S}^\varepsilon(u^{(j)})) - \mathcal{O}(\mathcal{S}^0(u^{(j)})) \right\|_2 \\ &\leq C_{\mathcal{O}} \left\| p^\varepsilon(u^{(j)}) - p^0(u^{(j)}) \right\|_{L^2(D)} , \end{aligned} \tag{77}$$

where we write explicitly the dependence of the solutions p^ε and p^0 on the particle they are generated by. Due to homogenization theory (see, e.g., [15, Theorem 19.1]), we have that $p^\varepsilon(u^{(j)}) \to p^0(u^{(j)})$ in $L^2(D)$ for all $j = 1, \ldots, J$, which implies

$$e(\varepsilon, u) = \frac{1}{J} \sum_{j=1}^{J} \left\| \mathcal{G}^\varepsilon(u^{(j)}) - \mathcal{G}^0(u^{(j)}) \right\|_2 \leq \frac{C_{\mathcal{O}}}{J} \sum_{j=1}^{J} \left\| p^\varepsilon(u^{(j)}) - p^0(u^{(j)}) \right\|_{L^2(D)} \to 0.$$

Moreover, if the solution of the homogenized problem p^0 is sufficiently smooth independently of u, namely, $p^0 \in H^2(D)$, letting $C > 0$ be a constant independent of ε, we have by [12] for all $j = 1, \ldots, J$,

$$\left\| p^\varepsilon(u^{(j)}) - p^0(u^{(j)}) \right\|_{L^2(D)} \leq C\varepsilon,$$

which implies

$$e(\varepsilon, u) = \frac{1}{J} \sum_{j=1}^{J} \left\| \mathcal{G}^\varepsilon(u^{(j)}) - \mathcal{G}^0(u^{(j)}) \right\|_2 \leq \frac{C_{\mathcal{O}}}{J} \sum_{j=1}^{J} \left\| p^\varepsilon(u^{(j)}) - p^0(u^{(j)}) \right\|_{L^2(D)} \leq C_{\mathcal{O}} C\varepsilon,$$

and defining $K = C_{\mathcal{O}} C$ gives the desired result. ∎

Proof of Lemma 3. First, for all $x \in B_R(u^*)$ we have

$$\begin{aligned} \|x\|_2 &\leq \|x - u^*\|_2 + \|u^*\|_2 \leq R + \|u^*\|_2 =: m, \\ \|\mathcal{G}(x)\|_2 &\leq \|\mathcal{G}(x) - \mathcal{G}(u^*)\|_2 + \|\mathcal{G}(u^*)\|_2 \\ &\leq C_{\mathcal{G}} \|x - u^*\|_2 + \|\mathcal{G}(u^*)\|_2 \leq C_{\mathcal{G}} R + \|\mathcal{G}(u^*)\|_2 =: M. \end{aligned} \tag{78}$$

We can also deduce the same bounds for the mean values

$$\|\bar{u}\|_2 \leq \frac{1}{J} \sum_{j=1}^{J} \left\| u^{(j)} \right\|_2 \leq m \qquad \text{and} \qquad \left\| \bar{\mathcal{G}} \right\|_2 \leq \frac{1}{J} \sum_{j=1}^{J} \left\| \mathcal{G}(u^{(j)}) \right\|_2 \leq M. \tag{79}$$

Then by (78) and (79) we get

$$\begin{aligned}\|C^{up}(u)\|_2 &= \sup_{x\in\mathbb{R}^L:\ \|x\|_2=1}\left\|\frac{1}{J}\sum_{j=1}^{J}(u^{(j)}-\bar{u})(\mathcal{G}(u^{(j)})-\bar{\mathcal{G}})^T x\right\|_2\\ &\le \frac{1}{J}\sum_{j=1}^{J}\left(\left\|\mathcal{G}(u^{(j)})\right\|_2+\|\bar{\mathcal{G}}\|_2\right)\left(\left\|u^{(j)}\right\|_2+\|\bar{u}\|_2\right)\\ &\le 4Mm,\end{aligned}$$

and defining $C_1 = 4Mm$ we get (i). The argument is similar for the matrix $C^{pp}(u)$, for which we have

$$\|C^{pp}(u)\|_2 \le \frac{1}{J}\sum_{j=1}^{J}\left(\left\|\mathcal{G}(u^{(j)})\right\|_2+\|\bar{\mathcal{G}}\|_2\right)^2 \le 4M^2,$$

and defining $C_2 = 4M^2$ we get (ii). Before proving (iii) and (iv), we need the following estimates for two ensemble of particles u_1 and u_2:

(80)
$$\begin{aligned}\|\bar{u}_1-\bar{u}_2\|_2 &= \left\|\frac{1}{J}\sum_{j=1}^{J}(u_1^{(j)}-u_2^{(j)})\right\|_2 \le \frac{1}{J}\sum_{j=1}^{J}\left\|u_1^{(j)}-u_2^{(j)}\right\|_2 = \|u_1-u_2\|,\\ \left\|\bar{\mathcal{G}}_1-\bar{\mathcal{G}}_2\right\|_2 &= \left\|\frac{1}{J}\sum_{j=1}^{J}(\mathcal{G}(u_1^{(j)})-\mathcal{G}(u_2^{(j)}))\right\|_2 \le \frac{C_{\mathcal{G}}}{J}\sum_{j=1}^{J}\left\|u_1^{(j)}-u_2^{(j)}\right\|_2 = C_{\mathcal{G}}\|u_1-u_2\|.\end{aligned}$$

Then we have

(81)
$$\begin{aligned}&\|C^{up}(u_1)-C^{up}(u_2)\|_2\\ &= \sup_{x\in\mathbb{R}^L:\ \|x\|_2=1}\left\|\frac{1}{J}\sum_{j=1}^{J}\left[(u_1^{(j)}-\bar{u}_1)(\mathcal{G}(u_1^{(j)})-\bar{\mathcal{G}}_1)^T x-(u_2^{(j)}-\bar{u}_2)(\mathcal{G}(u_2^{(j)})-\bar{\mathcal{G}}_2)^T x\right]\right\|_2\\ &\le \frac{1}{J}\sum_{j=1}^{J}\left(\left\|u_1^{(j)}\right\|_2+\|\bar{u}_1\|_2\right)\left(\left\|\mathcal{G}(u_1^{(j)})-\mathcal{G}(u_2^{(j)})\right\|_2+\left\|\bar{\mathcal{G}}_2-\bar{\mathcal{G}}_1\right\|_2\right)\\ &\quad+\frac{1}{J}\sum_{j=1}^{J}\left(\left\|u_1^{(j)}-u_2^{(j)}\right\|_2+\|\bar{u}_2-\bar{u}_1\|_2\right)\left(\left\|\mathcal{G}(u_2^{(j)})\right\|_2+\|\bar{\mathcal{G}}_2\|_2\right),\end{aligned}$$

and since $\mathcal{G}$ is Lipschitz and due to (78), (79), (80), we obtain

$$\begin{aligned}\|C^{up}(u_1)-C^{up}(u_2)\|_2 &\le 2m(C_{\mathcal{G}}J\|u_1-u_2\|+C_{\mathcal{G}}\|u_1-u_2\|)\\ &\quad+(J\|u_1-u_2\|+\|u_1-u_2\|)2M\\ &\le 2(J+1)(mC_{\mathcal{G}}+M)\|u_1-u_2\|,\end{aligned}$$

and defining $C_3 = 2(J+1)(mC_{\mathcal{G}}+M)$ we get (iii). The argument is similar for the

matrix $C^{pp}(u)$, for which we have

$$
\begin{aligned}
&\|C^{pp}(u_1) - C^{pp}(u_2)\|_2 \\
&\quad \le \frac{1}{J}\sum_{j=1}^{J}\left(\left\|\mathcal{G}(u_1^{(j)})\right\| + \|\bar{\mathcal{G}}_1\|_2\right)\left(\left\|\mathcal{G}(u_1^{(j)}) - \mathcal{G}(u_2^{(j)})\right\|_2 + \|\bar{\mathcal{G}}_2 - \bar{\mathcal{G}}_1\|_2\right) \\
&\quad\quad + \frac{1}{J}\sum_{j=1}^{J}\left(\left\|\mathcal{G}(u_1^{(j)}) - \mathcal{G}(u_2^{(j)})\right\|_2 + \|\bar{\mathcal{G}}_2 - \bar{\mathcal{G}}_1\|_2\right)\left(\left\|\mathcal{G}(u_2^{(j)})\right\| + \|\bar{\mathcal{G}}_2\|\right) \\
&\quad \le 4(J+1)MC_{\mathcal{G}},
\end{aligned} \tag{82}
$$

and defining $C_4 = 4(J+1)MC_{\mathcal{G}}$ we get (iv), which concludes the proof. ∎

Proof of Lemma 5. Let us consider an ensemble $u \in \mathcal{U}_{J,M}$ with particles $u^{(j)} \in \mathbb{R}^M$, for $j = 1, \dots, J$. For each particle we have

$$
\begin{aligned}
\left\|\mathcal{G}_h^0(u^{(j)}) - \mathcal{G}^0(u^{(j)})\right\|_2 &= \left\|\mathcal{O}(\mathcal{S}_h^0(u^{(j)})) - \mathcal{O}(\mathcal{S}^0(u^{(j)}))\right\|_2 \\
&\le C_{\mathcal{O}}\left\|p_h^0(u^{(j)}) - p^0(u^{(j)})\right\|_{L^2(D)},
\end{aligned} \tag{83}
$$

where we write explicitly the dependence of the solutions p^0 and p_h^0 on the particle they are generated by. Then due to standard a priori error estimates of FEM (see, e.g., [7, Theorem 3.2.5]) and higher order boundary regularity results for elliptic PDEs (see, e.g., [9, Theorem 6.3.5]) we have for all $j = 1, \dots, J$,

$$
\left\|p_h^0(u^{(j)}) - p^0(u^{(j)})\right\|_{L^2(D)} \le C\left|p^0(u^{(j)})\right|_{H^{s+1}(D)} h^{s+1} \le C\|f\|_{H^{q-1}(D)} h^{s+1},
$$

where $C > 0$ is a constant independent of h. Therefore, we obtain

$$
\tilde{e}(h,u) = \frac{1}{J}\sum_{j=1}^{J}\left\|\mathcal{G}_h^0(u^{(j)}) - \mathcal{G}^0(u^{(j)})\right\|_2 \le C_{\mathcal{O}}C\|f\|_{H^{q-1}(D)} h^{s+1}, \tag{84}
$$

and defining $\tilde{K} = C_{\mathcal{O}}C\|f\|_{H^{q-1}(D)}$ gives the desired result. ∎

Proof of Lemma 7. We follow the same steps as the proof of [17, Theorem 5.9]. Let us first recall the duality formula for the Wasserstein distance with $p = 1$:

$$
W_{1,s}(\mu_n, \mu) = \sup_{\varphi\in\Phi}\left\{\int_{B_R(u^*)} \varphi d(\mu_n - \mu)\right\},
$$

where Φ is the set of all globally Lipschitz continuous functions $\varphi\colon B_R(u^*) \to \mathbb{R}$ with Lipschitz constant $C_{\text{Lip}} \le 1$. Note that if $\varphi \in \Phi$, then also $-\varphi \in \Phi$. Hence we deduce that

$$
W_{1,s}(\mu_n, \mu) = \sup_{\varphi\in\Phi}\left\{\left|\int_{B_R(u^*)} \varphi d(\mu_n - \mu)\right|\right\}. \tag{85}
$$

Then we have

$$
\sup_{\varphi\in\Phi}\mathbb{E}_\xi\left[\left|\int_{B_R(u^*)} \varphi d(\mu_n - \mu)\right|\right] \le \mathbb{E}_\xi\left[\sup_{\varphi\in\Phi}\left\{\left|\int_{B_R(u^*)} \varphi d(\mu_n - \mu)\right|\right\}\right] = \mathbb{E}_\xi[W_{1,s}(\mu_n, \mu)], \tag{86}
$$

where the right-hand side vanishes by hypothesis. Therefore we obtain

$$\mathbb{E}_\xi\left[\left|\int_{B_R(u^*)}\varphi d\mu_n - \int_{B_R(u^*)}\varphi d\mu\right|\right] \to 0 \tag{87}$$

for all $\varphi \in \Phi$. Finally, we extend (87) to all Lipschitz functions by linearity and to all bounded continuous functions by density, thus proving the desired result. ∎

Acknowledgment. The authors thank the referee for the valuable input that in particular improved the proof of Theorem 4.

REFERENCES

[1] A. Abdulle, *A priori and a posteriori error analysis for numerical homogenization: A unified framework*, in Multiscale Problems, Ser. Contemp. Appl. Math. CAM 16, World Scientific, Singapore, 2011, pp. 280–305.

[2] A. Abdulle and A. Di Blasio, *Numerical homogenization and model order reduction for multiscale inverse problems*, Multiscale Model. Simul., 17 (2019), pp. 399–433.

[3] A. Abdulle and A. Di Blasio, *A Bayesian numerical homogenization method for elliptic multiscale inverse problems*, SIAM/ASA J. Uncertain. Quantif., 8 (2020), pp. 414–450.

[4] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden, *The heterogeneous multiscale method*, Acta Numer., 21 (2012), pp. 1–87.

[5] D. Calvetti, M. Dunlop, E. Somersalo, and A. Stuart, *Iterative updating of model error for Bayesian inversion*, Inverse Problems, 34 (2018), 025008.

[6] D. Calvetti, O. Ernst, and E. Somersalo, *Dynamic updating of numerical model discrepancy using sequential sampling*, Inverse Problems, 30 (2014), 114019.

[7] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Classics Appl. Math. 40, SIAM, Philadelphia, 2002.

[8] D. Cioranescu and P. Donato, *An Introduction to Homogenization*, Oxford Lecture Ser. Math. Appl. 17, Oxford University Press, New York, 1999.

[9] L. C. Evans, *Partial Differential Equations*, 2nd ed., Grad. Stud. Math. 19, American Mathematical Society, Providence, RI, 2010.

[10] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res., 99 (1994), pp. 10143–10162.

[11] M. A. Iglesias, K. J. H. Law, and A. M. Stuart, *Ensemble Kalman methods for inverse problems*, Inverse Problems, 29 (2013), 045001.

[12] S. Moskow and M. Vogelius, *First-order corrections to the homogenised eigenvalues of a periodic composite medium. A convergence proof*, Proc. Roy. Soc. Edinburgh Sect. A, 127 (1997), pp. 1263–1299.

[13] J. Nolen and G. Papanicolaou, *Fine scale uncertainty in parameter estimation for elliptic equations*, Inverse Problems, 25 (2009), 115021.

[14] J. Nolen, G. A. Pavliotis, and A. M. Stuart, *Multiscale modeling and inverse problems*, in Numerical Analysis of Multiscale Problems, Lect. Notes Comput. Sci. Eng. 83, Springer, Heidelberg, 2012, pp. 1–34.

[15] G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Texts Appl. Math. 53, Springer, New York, 2008.

[16] S. Salsa, *Partial Differential Equations in Action*, From Modelling to Theory, 3rd ed., Unitext Mat. 99, Springer, Cham, Switzerland, 2016.

[17] F. Santambrogio, *Optimal transport for applied mathematicians*, Progr. Nonlinear Differential Equations, 87, Birkhäuser, Cham, Switzerland, 2015.

[18] C. Schillings and A. M. Stuart, *Analysis of the ensemble Kalman filter for inverse problems*, SIAM J. Numer. Anal., 55 (2017), pp. 1264–1290.

[19] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.