Mathematical Biosciences 276 (2016) 133–144

Contents lists available at ScienceDirect

Mathematical Biosciences

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

Kalman filter parameter estimation for a nonlinear diffusion model of epithelial cell migration using stochastic collocation and the Karhunen–Loeve expansion☆

Jared Barber, Roxana Tanase, Ivan Yotov*

Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA

ARTICLE INFO

Article history:
Received 27 July 2015
Revised 16 March 2016
Accepted 25 March 2016
Available online 13 April 2016

Keywords:
Kalman filter
Stochastic collocation
Karhunen–Loeve expansion
Parameter estimation
Epithelial cell migration

ABSTRACT

Several Kalman filter algorithms are presented for data assimilation and parameter estimation for a nonlinear diffusion model of epithelial cell migration. These include the ensemble Kalman filter with Monte Carlo sampling and a stochastic collocation (SC) Kalman filter with structured sampling. Further, two types of noise are considered —uncorrelated noise resulting in one stochastic dimension for each element of the spatial grid and correlated noise parameterized by the Karhunen–Loeve (KL) expansion resulting in one stochastic dimension for each KL term. The efficiency and accuracy of the four methods are investigated for two cases with synthetic data with and without noise, as well as data from a laboratory experiment. While it is observed that all algorithms perform reasonably well in matching the target solution and estimating the diffusion coefficient and the growth rate, it is illustrated that the algorithms that employ SC and KL expansion are computationally more efficient, as they require fewer ensemble members for comparable accuracy. In the case of SC methods, this is due to improved approximation in stochastic space compared to Monte Carlo sampling. In the case of KL methods, the parameterization of the noise results in a stochastic space of smaller dimension. The most efficient method is the one combining SC and KL expansion.

1. Introduction

Parameter estimation is an important field in the area of modeling physical or biological processes [21]. The set of parameters that maximize the model's agreement with experimental data can be used to yield important insight into a given system. It can help scientists more clearly describe the behavior of the system, predict behavioral changes in the system during pathological situations, and assess the efficacy of various treatment options [6,26]. In addition, once those optimal parameters have been found, other mathematical techniques can be used to obtain further insight into the system's behavior. Local sensitivity analysis [5] at the optimal parameter set can be used to assess the local importance of the parameters. Also, the optimal parameter set can be used as a starting point for obtaining, via, e.g., Markov-Chain Monte Carlo (MCMC) methods, distributions of the parameters that produce computational estimates that agree reasonably well with experiment. These distributions can be used to assess the global importance of each parameter.

As computational models become more complex in order to describe systems in more detail, parameter estimation can become more difficult and costly because of increased numbers of parameters and simulation runtime. Obtaining parameter estimates in a reasonable amount of time has begun to depend more and more on efficient methods of parameter estimation.

Traditional approaches to obtaining ideal parameter sets include least-squares or maximum likelihood approaches in which a cost functional (usually a sum of weighted squared differences between model and experimental values) is minimized. Direct optimization methods such as the Nelder–Mead simplex method or the conjugate gradient method are often used to find the corresponding minimum of the cost functional, which serves as the ideal parameter set [20]. Probability based methods, such as the MCMC method, have also been used to explore parameter space and search for optimal parameters. Direct optimization techniques may get stuck in local minima and may require large amounts of time to find minima in high dimensional space. While improved implementations of the MCMC algorithms exist, see [2,7,16–18],

☆ This work was partially supported by NSF grants DMS 1115856 and DMS 1418947 and by DOE grant DE-FG02-04ER25618.

* Corresponding author. Tel.: +1 4126248338.
E-mail addresses: jaredb@pitt.edu (J. Barber), rot26@pitt.edu (R. Tanase), yotov@math.pitt.edu (I. Yotov).

http://dx.doi.org/10.1016/j.mbs.2016.03.018
0025-5564/

MCMC generally suffers from the need of a large number of simulations before optimal parameter sets are obtained.

Usually, in the above methods, a simulation is run from the start of an experiment to the end of an experiment before the cost function is evaluated and the best guess for the optimal parameter set is adjusted. In contrast, sequential data assimilation techniques adjust parameter sets at every time at which experimental data is available. Adjusting parameters more often usually allows for quicker convergence to desired parameter estimates. In this paper we study several Kalman filter (KF) techniques [8] that update the parameter estimate multiple times per simulation run and compare their performance for parameter estimation for a partial differential equation (PDE) model of cell migration in an *in vitro* experiment [1].

Kalman filter methods usually take a running temporal model and periodically update it by using experimental measurements. While traditionally these methods have been used to update the values of the dependent variables for a given model of a physical system, they can also be used to update estimates of the parameter values of the model if an initial guess for those parameter values is given [15].

The original linear Kalman filter can only be used for models with linear dynamics. The extended Kalman filter uses the Jacobian to linearize and deal with nonlinear dynamics. Both the linear and extended Kalman filters track the underlying distributions of the dependent variables and unknown parameters through time by evolving and tracking the mean and variance of the variables in the model. While the extended Kalman filter can be used on moderately nonlinear problems, it can suffer when presented with highly nonlinear problems. This is because the Jacobians provide only local information.

Other Kalman filters alleviate the problem of potentially misleading local information by using a global sampling of points, rather than a local mean and Jacobian-derived variance, to represent the underlying distribution. The ensemble Kalman filter [8,11] tracks the evolution of the variable distributions by using a Monte Carlo sampling of the variable space that is evolved through time. Recently other Kalman filters have been introduced that use structured samplings of stochastic space that are based upon quadrature rules [15,23,25]. Our stochastic collocation Kalman filter (SCKF) is of this type as it uses sparse grid collocation or quadrature in order to estimate the mean and variance resulting when the model is propagated in time. For moderately sized problems, filters based on structured sampling can be more efficient than the ensemble Kalman filter, since they require fewer realizations to obtain comparable accuracy.

Additional gains can be realized for PDEs if we utilize the fact that the model errors associated with many PDE-based models (the errors introduced to the variables when using the model to evolve the PDE variables in time) tend to be correlated rather than uncorrelated. If we assume model errors are uncorrelated then there is one independent degree of freedom for each variable at each grid point/cell and the dimension of the stochastic space quickly increases when the grid is refined. Usually in PDE systems, however, the model errors at one grid location is correlated with the model error at nearby grid locations. To incorporate this correlation, we have assumed that the errors at the grid cells can be represented by a Karhunen–Loeve (KL) expansion, which is based on an eigenfunction expansion of the covariance [10]. The KL expansion can be truncated due the fast decay of the eigenvalues [10]. This results in a reduction in the dimension of the effective stochastic space which corresponds to fewer realizations needed for a desired parameter estimation. In particular, there is one stochastic dimension for each significant KL term and the dimension of the stochastic space is independent of the spatial grid.

Table 1.1
Parameter estimation techniques used.

Technique	Sequential data assimilation	Structured sampling	Karhunen–Loeve expansion
DO			
EnKF	X		
SCKF	X	X	
KLEnKF	X		X
KLSCKF	X	X	X

In this paper we investigate the efficiency and accuracy of using sequential data assimilation via KF methods, structured sampling via SC methods, and the KL expansion for parameter estimation in a model of intestinal epithelial cell migration. We do this by comparing parameter estimates obtained from five different parameter estimation techniques (see Table 1.1): direct optimization of a cost functional (DO), ensemble Kalman filter (EnKF), stochastic collocation Kalman filter (SCKF), ensemble Kalman filter with KL expansion (KLEnKF), and stochastic collocation Kalman filter with KL expansion (KLSCKF). In the SC algorithms, a new random ensemble is generated after each data assimilation step to avoid adding noise in the same stochastic direction. We present computational results for two cases with synthetic data with and without noise, as well as experimental data from the lab of David Hackam [1]. We observe that all algorithms are able to match the target solution or experimental data and to estimate the diffusion coefficient and the growth rate. However, the algorithms that employ SC acceleration and the KL expansion are computationally more efficient, as they require fewer ensemble members for comparable accuracy.

We note that the stability of the Kalman filter algorithms depends on the observability of the dynamical system, i.e., the ability to determine uniquely the state variables and parameters from the set of measurements, see e.g. [12]. Although a rigorous mathematical proof of observability is beyond the scope of this work, the numerical results indicate that our model and set of measurements give an observable dynamical system. In particular, in the most challenging setting of using experimental data, the direct optimization algorithm converges to parameter values that are very close to those obtained by the four KF methods, and the computed solutions match the experimental data very well. In addition, In the simulated data setting, the parameters estimates obtained by all four KF algorithms are very close to the true parameter values.

The remainder of the paper is organized as follows. The methods and algorithms are presented in Section 2. The computational results are described in Section 3. The results are discussed in Section 4.

2. Methods

2.1. Experiments

The experimental data was obtained in the Hackam Lab at the University of Pittsburgh and the experimental procedures have been presented in [1].

2.2. Model

2.2.1. Equations

The mathematical model consists of a two-dimensional domain representing a layer of epithelial cells that evolves in time according to the partial differential equation

$$\frac{\partial e_c}{\partial t} = D\nabla\cdot\left(\left(\frac{e_c^2}{e_c^2+(e_{c,max}-e_c)^2}\right)\nabla e_c\right)+k_p e_c(e_{c,max}-e_c). \tag{2.1}$$

This nonlinear diffusion equation for the epithelial cell concentration e_c has been used to model wound closure in necrotizing enterocolitis [3]. Here D is the diffusion coefficient, k_p is the growth rate, and $e_{c,max} = 1$ is the maximum concentration.

The S-shaped nonlinear diffusion term is chosen from the Buckley–Leverett model of two-phase flow in porous media [4]). It produces no cell migration when $e_c = 0$ and maximal migration when $e_c = e_{c,max}$. In addition the choice of the S-shape makes it so that all regions with low epithelial integrity (small e_c) exhibit near minimal cell migration (approximately proportional to e_c^2) while all regions of high epithelial integrity exhibit nearly maximal migration rates (approximately proportional to $e_{c,max}^2 - (e_{c,max} - e_c)^2$; see [3]).

2.2.2. Computational methods and domain

A standard cell-centered finite difference method was implemented in MATLAB to discretize this equation on a 10×10 grid (100 free state variables) including appropriate upwinding of the nonlinear diffusion term [14] and using Forward Euler in time with step sizes that do not violate the CFL condition, see [3] for details. The simulation domain is the rectangle $[-0.05, 0.05] \times [-0.035, 0.035]$ discretized on a 10×10 spatial mesh. The initial condition for all tests is obtained from the initial image from the experimental data. It corresponds to an initial wound with irregular shape that closes during the simulation. We take $e_c = 0$ inside the wound and $e_c = 1$ outside.

2.3. General overview of Kalman filter methods

The Kalman filter is a two step process that evolves the state and uncertainty/variance associated with a system optimally by using experimental data corresponding to that system. The first step (prediction or forecast step) uses a computational model and the uncertainty associated with that model to evolve both the system's mean and variance to the next time step at which experimental data is available. The second step (the analysis or assimilation step) uses experimental data and the uncertainty associated with the experiments (measurement error) to adjust the variable means and variances to more closely agree with the experimental data.

The first step for just the means is given, mathematically, by the following:

$$\mathbf{x}_n^f = \mathbf{f}(\mathbf{x}_{n-1}^a) + \mathbf{w}_n.$$

Here $\mathbf{x}_n^f$ and $\mathbf{x}_{n-1}^a \in \mathbf{R}^m$ are vectors of the state variables and parameters for the given system (state vectors) from the nth forecast and $n-1$st assimilation time steps, respectively, $\mathbf{w}_n \in \mathbf{R}^m$ is a stochastic normally distributed vector with mean zero and covariance matrix $Q_n \in \mathbf{R}^{m\times m}$ representing the noise or uncertainty associated with using the model over the nth time step, and $\mathbf{f} \in \mathbf{R}^m$ is a forecasting model that is used to evolve the state vector in time.

The assimilation step, also known as the adjustment, analysis, filtering, or model update step, is given by:

$$\mathbf{x}_n^a = \mathbf{x}_n^f + K_n(\mathbf{y}_n - \mathbf{h}(\mathbf{x}_n^f)).$$

Here $K_n \in \mathbf{R}^{m\times l}$ is the Kalman gain, $\mathbf{y}_n \in \mathbf{R}^l$ is the vector of experimental measurement values available at time t_n (measurement vector), and $\mathbf{h} \in \mathbf{R}^l$ is a measurement function that returns estimates of the measurement values corresponding to a given state vector. While not shown explicitly, the measurement vectors, like the state vectors are also assumed to be stochastic so that they can be represented by

$$\mathbf{y}_n = \bar{\mathbf{y}}_n + \mathbf{v}_n,$$

where $\bar{\mathbf{y}}_n$ holds the expected measurement values, and $\mathbf{v}_n \in \mathbf{R}^l$ is the measurement noise vector or uncertainty that is normally distributed with mean zero and covariance matrix $R_n \in \mathbf{R}^{l\times l}$. The Kalman gain is chosen to minimize the amount of uncertainty in the new estimate of the state vector for the system, $\mathbf{x}_n^a$, and depends on the covariances of the forecast state vectors and measurement values.

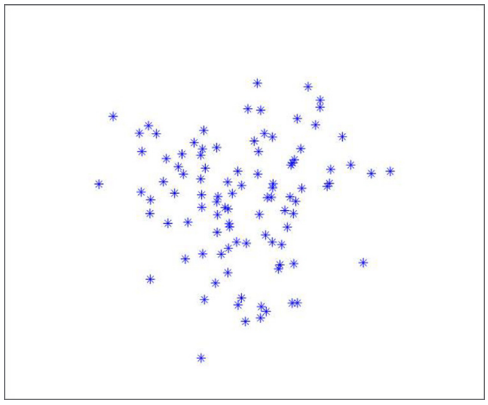

Fig. 1. EnKF random sampling of 2-d stochastic space.

There are two primary ways in which the means and variances of the variables are tracked in the Kalman filter. Traditionally, in both the linear and extended Kalman filter, the mean and covariance matrix of the state variables each have their own evolution equation and are explicitly tracked as time evolves. With the advent of the ensemble Kalman filter [8,11] it has become common to instead track the means and variances by evolving each member of a sampling, or ensemble, of stochastic variable space. In this latter case, if more specific information such as the mean or covariance of the variables in the actual system is desired, they can be estimated by calculating the mean and covariance matrix of the ensemble. Often the two approaches are mixed [15,22,23], as is the case here.

Finally we mention that while the Kalman filter has traditionally been used to correct just the state variables in a given model, it has become common to use the Kalman filter in a parameter estimation role. By appending guesses for the unknown parameter to the state vector, evolving those parameters with the identity function during the predict step, and then allowing the analysis step to adjust the parameter values so that the state variables more closely agree with experiment, the parameter values will tend to evolve towards the ideal values for the system, that is, towards a parameter set that reproduces the experimental data fairly well. In addition, it is often, though not always, the case that the first guess for the parameters need not be close to the ideal parameter set in order for the guesses to converge to that set.

2.3.1. Ensemble Kalman filter

The ensemble Kalman filter [8,11] tracks the underlying distributions of the state variables and measurements by representing the underlying distributions using an ensemble of state and measurement vectors and advancing those distributions over time by advancing each member of the ensemble independently.

The algorithm, along with a short description of each step, is listed in Table 2.1. In this and the other three algorithms, the dimension of the state/parameter vector $\mathbf{x}$ is $m = n_c + 2$, where n_c is the number of grid cells. In particular, we have one free state variable per grid cell and two parameters D and k_p. In this and the next algorithm, since the spatial noise is assumed uncorrelated, the dimension of the stochastic space is also m. Ensembles of state vectors, measurement noise, and model noise are of size q and are random, rather than structured, samplings of the stochastic space, see Fig. 1. When q is large enough, the ensembles should

Table 2.1
EnKF Algorithm.

Initialize	
$\mathbf{x}_0^a$	Initial best state vector
$P_{0,xx}^a$	Initial best state vector uncertainty
$\{\mathbf{x}_{0,k}^a\}_{k=1}^q$	Use $\mathbf{x}_0^a$ and $P_{0,xx}^a$ to obtain a random/unstructured sampling or ensemble of q vectors that correspond to/represent the underlying distribution
For $n = 1, \ldots, N$	
Prediction Step	
$\mathbf{x}_{n,k}^f = \mathbf{f}(\mathbf{x}_{n-1,k}^a) + \mathbf{w}_{n,k}$	Predict new state for each ensemble member
$\bar{\mathbf{x}}_n^f = \frac{1}{q}\sum_{k=1}^q \mathbf{x}_{n,k}^f$	Mean new state, according to model
$\bar{\mathbf{y}}_n^f = \frac{1}{q}\sum_{k=1}^q \mathbf{h}(\mathbf{x}_{n,k}^f)$	Mean new measurement, *according to model*
$E_{x,k}^f = \mathbf{x}_{n,k}^f - \bar{\mathbf{x}}_n^f$	Deviation of kth forecast ensemble member from mean
$E_{y,k}^f = \mathbf{h}(\mathbf{x}_{n,k}^f) - \bar{\mathbf{y}}_n^f$	Deviation of kth forecast measurement of ensemble member from mean measurement
$P_{n,xx}^f = \frac{1}{q-1}\sum_{k=1}^q E_{x,k}^f (E_{x,k}^f)^T$	New xx-covariance
$P_{n,xy}^f = \frac{1}{q-1}\sum_{k=1}^q E_{x,k}^f (E_{y,k}^f)^T$	New xy-covariance
$P_{n,yy}^f = \frac{1}{q-1}\sum_{k=1}^q E_{y,k}^f (E_{y,k}^f)^T$	New yy-covariance
Adjustment Step	
$K_n = P_{n,xy}^f (P_{n,yy}^f)^{-1}$	Find Kalman gain
$\mathbf{x}_{n,k}^a = \mathbf{x}_{n,k}^f + K_n(\mathbf{y}_n + \mathbf{v}_{n,k} - \mathbf{h}(\mathbf{x}_{n,k}^f))$	Find analyzed state for each ensemble member
$\bar{\mathbf{x}}_n^a = \frac{1}{q}\sum_{k=1}^q \mathbf{x}_{n,k}^a$	Mean best estimate state, after measurement adjustment
$E_{x,k}^a = \mathbf{x}_{n,k}^a - \bar{\mathbf{x}}_n^a$	Deviation of kth analyzed ensemble member from mean
$P_{n,xx}^a = \frac{1}{q-1}\sum_{k=1}^q E_{x,k}^a (E_{x,k}^a)^T$	Find new covariance (not needed to continue to next time)
end	

have a mean and variance that is approximately equal to the mean and variance of the underlying distributions. The index k corresponds to the kth ensemble member. The ensembles of noise vectors $\{\mathbf{w}_{n,k}\}_{k=1}^q$ and $\{\mathbf{v}_{n,k}\}_{k=1}^q$ are drawn from the normal distributions with covariance matrices Q_n and R_n, respectively.

2.3.2. Stochastic collocation Kalman filter

In the ensemble Kalman filter, the mean and variance of the ensemble converge to the true mean of the ensemble (according to Monte Carlo sampling) as $1/\sqrt{q}$. Because of this, a large number of ensemble members is often required if the ensemble Kalman filter is going to effectively track the true underlying distribution as it evolves in time. When the model function $\mathbf{f}$ is costly to evaluate, this would result in a slow algorithm.

To alleviate this problem, it has become practice (unscented Kalman filter, sigma point Kalman filter, Gaussian filters, stochastic collocation Kalman filter) to build ensembles that consist of points strategically chosen from the underlying stochastic space [15,22,23,25]. This is in contrast to the ensemble Kalman filter where the stochastic space is randomly sampled. When points are strategically chosen, numerical integration techniques on the corresponding structured grid can be used to obtain good estimates of the evolving mean and covariance of the underlying distribution (Fig. 2).

The stochastic collocation method builds an interpolant in the stochastic space using solution values at q_{SC} collocation points. Therefore, its computational complexity is q_{SC} times that of a deterministic problem. Thus, we need to choose a nodal set Θ with fewest possible number of points under a prescribed accuracy requirement. There are several choices of such collocation points, using either tensor products of one-dimensional nodal sets, or sparse grids constructed by the Smolyak algorithm [19,24]. The Smolyak approximation is a linear combination of product formulas, and the linear combination is chosen in such a way that an interpolation property for one-dimensional spaces is preserved for multidimensional spaces. Only products with a relatively small number of points are used and the resulting nodal set has significantly fewer number of nodes compared to the tensor product rule. In this paper, we use Smolyak formulas that are based on a linear combination of one-dimensional polynomial interpolants at the extrema of the Hermite polynomials, which are the orthogonal polynomials with a weight given by the probability density function of the normal distribution, i.e., Gaussian abscissas. Other choices, such as the extrema of the Chebyshev polynomials, i.e., Clenshaw–Curtis abscissas, can be considered as well. Let q_{SC} be the size of the ensemble, let $\{\mathbf{r}_{SC,k}\}_{k=1}^{q_{SC}} \in \mathbf{R}^m$ be the collocation points, and let $\{c_{SC,k}\}_{k=1}^{q_{SC}}$ be the collocation weights. The algorithm for calculating $\{\mathbf{r}_{SC,k}\}_{k=1}^{q_{SC}}$ and $\{c_{SC,k}\}_{k=1}^{q_{SC}}$ is given, e.g., in [19,24]. In our implementation we use Smolyak level-one sparse grid, which has two collocation points in each dimension plus the origin, resulting in $q_{SC} = 2m + 1$. The SCKF algorithm is given in Table 2.2.

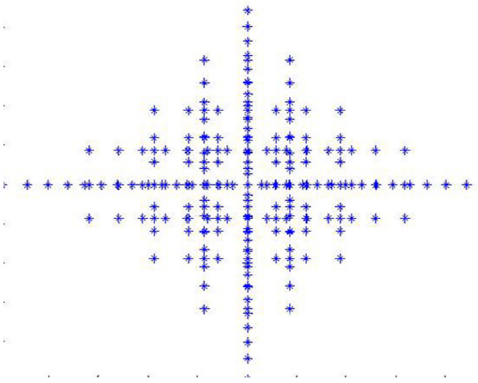

Fig. 2. SCKF structured sampling of 2-d stochastic space.

Remark 2.1. We note that, since the set of collocation points is fixed, sampling the noise at these points at each data assimilation step would result in adding noise to the model and measurements in the same stochastic direction. To avoid this, at each data assimilation step the Kalman gain is used to adjust the mean and a new ensemble is generated using the new mean, the vector of collocation points and the new covariance matrix.

2.3.3. Karhunen–Loeve stochastic collocation Kalman filter, Karhunen–Loeve ensemble Kalman filter

The most costly portion of the Kalman filter is the functional evaluation of $\mathbf{f}$, which corresponds to advancing the computational model in time. As such, the fewer ensemble members a method

Table 2.2
SCKF Algorithm.

Initialize	
$\bar{\mathbf{x}}_0^a$	Initial best state vector (corresponds to a mean)
$P_{0,xx}^a$	Initial best state vector uncertainty/covariance
$\{\mathbf{r}_{SC,k}\}_{k=1}^{q_{sc}}$	Collocation points
$\{c_{SC,k}\}_{k=1}^{q_{sc}}$	Weights for stochastic collocation
For $n = 1 \ldots N$	
Prediction Step	
$\mathbf{x}_{n-1,k}^a = \bar{\mathbf{x}}_{n-1}^a + \sqrt{P_{n-1,xx}^a}\mathbf{r}_{SC,k}$	Use variance associated with each component to create a structured ensemble
$\mathbf{x}_{n,k}^f = \mathbf{f}(\mathbf{x}_{n-1,k}^a)$	Predict new state
$\bar{\mathbf{x}}_n^f = \sum_{k=1}^{q_{sc}} c_{SC,k}\mathbf{x}_{n,k}^f$	Mean new state, according to model
$\bar{\mathbf{y}}_n^f = \sum_{k=1}^{q_{sc}} c_{SC,k}\mathbf{h}(\mathbf{x}_{n,k}^f)$	Mean new measurement, *according to model*
$E_{x,k}^f = \mathbf{x}_{n,k}^f - \bar{\mathbf{x}}_n^f$	Deviation of kth ensemble member from mean
$E_{y,k}^f = \mathbf{h}(\mathbf{x}_{n,k}^f) - \bar{\mathbf{y}}_n^f$	Deviation of measurement of kth ensemble member from mean measurement
$P_{n,xx}^f = \sum_{k=1}^{q_{sc}} c_{SC,k}E_{x,k}^f(E_{x,k}^f)^T + Q_n$	Find new covariance
$P_{n,xy}^f = \sum_{k=1}^{q_{sc}} c_{SC,k}E_{x,k}^f(E_{y,k}^f)^T$	Find new covariance
$P_{n,yy}^f = \sum_{k=1}^{q_{sc}} c_{SC,k}E_{y,k}^f(E_{y,k}^f)^T + R_n$	Find new covariance
Adjustment Step	
$K_n = P_{n,xy}^f(P_{n,yy}^f)^{-1}$	Find Kalman gain
$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^f + K_n(\mathbf{y}_n - \bar{\mathbf{y}}_n^f)$	Find adjusted mean state
$P_{n,xx}^a = P_{n,xx}^f - K_n P_{n,yy}^f K_n^T$	Find new covariance
end	

needs to obtain satisfactory results, the faster the method is. For low-dimensional systems, the stochastic collocation Kalman filter needs few ensemble members in its organized ensemble, while the Ensemble Kalman filter needs many in its randomly chosen ensemble. As the dimension is increased, however, the stochastic collocation Kalman filter suffers from the curse of dimensionality. For instance, for one spatially dependent variable on a coarse $10 \times 10 \times 10$ computational grid, the dimension of the stochastic space is one thousand, therefore over one thousand ensemble members are required to run the stochastic collocation Kalman filter. A $20 \times 20 \times 20$ grid would require over eight thousand ensemble members. The ensemble Kalman filter often requires only around one thousand ensemble members for similarly sized grids.

To address this problem, one can explore a parameterized noise representation, such as the Karhunen–Loeve (KL) expansion. The uncertainties associated with each component of the state vector are often correlated with each other. This is especially true when the components correspond to spatially dependent variables on computational grids. We use the KL expansion to represent these spatially correlated uncertainties. It is very similar to a Fourier expansion with the KL eigenfunctions looking somewhat sinusoidal in shape. On a discrete grid of size $p \times p \times p$, one needs p^3 KL eigenfunctions to completely represent a given discrete correlation function on the grid. Like a Fourier series, however, it can be shown that in continuous space the eigenvalues decay fast and the KL expansion of a given function converges to that function as more terms are included in the expansion [10]. As such, using just the first few terms of the KL expansion in discrete space, instead of p^3 terms, should sufficiently represent the distribution of the possible state of the system. Doing so reduces the effective stochastic space and allows one to use a much smaller ensemble to represent the underlying distributions. This corresponds to fewer necessary evaluations of the model function $\mathbf{f}$ and faster Kalman filtering.

Given a correlation function in two dimensions $C_v(\vec{x}_\alpha, \vec{x}_\beta)$ for a stochastic variable v, the corresponding Karhunen–Loeve expansion for that variable is given by [10]:

$$v(\vec{x}, \omega) = E[v](\vec{x}) + \sum_{i=1}^{\infty} \xi_i(\omega)\sqrt{\lambda_i} f_i(\vec{x})$$

where the corresponding eigenfunctions $f_i(\vec{x})$ satisfy the following integral equation

$$\iint_D C(\vec{x}_\alpha, \vec{x}_\beta) f_i(\vec{x}_\alpha) d\vec{x}_\alpha = \lambda_i f_i(\vec{x}_\beta).$$

Due to the symmetry and positive definiteness of the covariance function, the corresponding eigenfunctions are mutually orthogonal. In addition, since we assume the noise being represented is normally distributed at each point, $\xi_i(\omega)$ must be uncorrelated normal distributions with mean zero and standard deviation one.

As we discretize the model, it is useful to discuss the corresponding discrete version of the KL expansion, which is just an eigenfunction expansion. Recall that we consider a cell-centered finite difference method on a two-dimensional rectangular grid with n_c grid cells. Let $\mathbf{C} \in \mathbf{R}^{n_c \times n_c}$ be the covariance matrix where $\mathbf{C}_{ij}$ is the covariance between the noise at the two cell centers (x_i, y_i) and (x_j, y_j). The corresponding expansion is

$$\vec{v}(\omega) = E[\vec{v}] + \sum_{i=1}^{n_c} \xi_i(\omega)\sqrt{\lambda_i}\vec{e}_i$$

where $\mathbf{C}$ the eigenvectors $\vec{e}_i \in \mathbf{R}^{n_c}$ satisfy

$$\mathbf{C}\vec{e}_i = \lambda_i \vec{e}_i$$

Since the λ_i decay to zero relatively rapidly as i grows, there are only a few dominant eigenvectors. The eigenvectors are mutually orthogonal because of the symmetry and positive definiteness of the covariance matrix. In our computations we use the covariance matrix [9]

$$\mathbf{C}_{ij} = \sigma^2 e^{-|x_i - x_j|/L_x - |y_i - y_j|/L_y},$$

where σ is the variance and L_x, L_y are the correlation lengths. The above function is widely used for modeling diffusion processes in porous media [9,27]. We assume that it is also suitable for cell migration processes. The better our estimate of the covariance function/matrix of the process, the better the corresponding estimate of the KL expansion for the process will be and the better the KL version of the KF will perform [27,28]. We also note that the variance σ above and the corresponding $\sqrt{\lambda_i}$'s may evolve as time evolves and are easily rescaled as appropriate.

Table 2.3
KLSCKF Algorithm.

Initialize	
$\bar{\mathbf{x}}_0^a$	Initial best state vector
$P_{0,xx}^a$	Initial best state vector uncertainty
$\mathbf{E}$	Matrix of orthonormalized eigenvectors
$\{\mathbf{r}_{KL,k}\}_{k=1}^{q_{KL}}$	Collocation points
$\{c_{KL,k}\}_{k=1}^{q_{KL}}$	Weights for stochastic collocation
For $n = 1 \ldots N$	
Prediction Step	
$P_E = \mathbf{E}^T P_{n-1,xx}^a \mathbf{E}$	Project covariance onto the KL eigenspace
$\mathbf{x}_{n-1,k}^a = \bar{\mathbf{x}}_{n-1}^a + \mathbf{E}\sqrt{P_E}\mathbf{r}_{KL,k}$	Use projected covariance to create a structured ensemble
$\mathbf{x}_{n,k}^f = \mathbf{f}(\mathbf{x}_{n-1,k}^a)$	Predict new state
$\bar{\mathbf{x}}_n^f = \sum_{k=1}^{q_{KL}} c_{KL,k}\mathbf{x}_{n,k}^f$	Mean new state, according to model
$\bar{\mathbf{y}}_n^f = \sum_{k=1}^{q_{KL}} c_{KL,k}\mathbf{h}(\mathbf{x}_{n,k}^f)$	Mean new measurement, *according to model*
$E_{x,k}^f = \mathbf{x}_{n,k}^f - \bar{\mathbf{x}}_n^f$	Deviation of kth ensemble member from mean
$E_{y,k}^f = \mathbf{h}(\mathbf{x}_{n,k}^f) - \bar{\mathbf{y}}_n^f$	Deviation of measurement of kth ensemble member from mean measurement
$P_{n,xx}^f = \sum_{k=1}^{q_{KL}} c_{KL,k}E_{x,k}^f(E_{x,k}^f)^T + Q_n$	Find new covariance
$P_{n,xy}^f = \sum_{k=1}^{q_{KL}} c_{KL,k}E_{x,k}^f(E_{y,k}^f)^T$	Find new covariance
$P_{n,yy}^f = \sum_{k=1}^{q_{KL}} c_{KL,k}E_{y,k}^f(E_{y,k}^f)^T + R_n$	Find new covariance
Adjustment Step	
$K_n = P_{n,xy}^f(P_{n,yy}^f)^{-1}$	Find Kalman gain
$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^f + K_n(\mathbf{y}_n - \bar{\mathbf{y}}_n^f)$	Find adjusted mean state
$P_{n,xx}^a = P_{n,xx}^f - K_n P_{n,yy}^f K_n^T$	Find new covariance
end	

Table 2.4
KLEnKF Algorithm.

Initialize	
$\mathbf{x}_0^a$	Initial best state vector
$P_{0,xx}^a$	Initial best state vector uncertainty
$\{\mathbf{x}_{0,k}^a\}_{k=1}^q$	Use $\mathbf{x}_0^a$ and $P_{0,xx}^a$ to obtain a sampling or ensemble of q vectors that correspond to/represent the underlying distribution
For $n = 1, \ldots, N$	
Prediction Step	
$\mathbf{x}_{n-1,k}^a = \bar{\mathbf{x}}_{n-1}^a + \mathbf{E}\mathbf{E}^T(\mathbf{x}_{n-1,k}^a - \bar{\mathbf{x}}_{n-1}^a)$	Project the ensemble's distance from the mean onto the KL eigenspace and use this to construct an ensemble
$\mathbf{x}_{n,k}^f = \mathbf{f}(\mathbf{x}_{n-1,k}^a) + \mathbf{w}_{n,k}$	Predict new state for each ensemble member
$\bar{\mathbf{x}}_n^f = \frac{1}{q}\sum_{k=1}^q \mathbf{x}_{n,k}^f$	Mean new state, according to model
$\bar{\mathbf{y}}_n^f = \frac{1}{q}\sum_{k=1}^q \mathbf{h}(\mathbf{x}_{n,k}^f)$	Mean new measurement, *according to model*
$E_{x,k}^f = \mathbf{x}_{n,k}^f - \bar{\mathbf{x}}_n^f$	Deviation of kth forecast ensemble member from mean
$E_{y,k}^f = \mathbf{h}(\mathbf{x}_{n,k}^f) - \bar{\mathbf{y}}_n^f$	Deviation of kth forecast measurement of ensemble member from mean measurement
$P_{n,xx}^f = \frac{1}{q-1}\sum_{k=1}^q E_{x,k}^f(E_{x,k}^f)^T$	New xx-covariance
$P_{n,xy}^f = \frac{1}{q-1}\sum_{k=1}^q E_{x,k}^f(E_{y,k}^f)^T$	New xy-covariance
$P_{n,yy}^f = \frac{1}{q-1}\sum_{k=1}^q E_{y,k}^f(E_{y,k}^f)^T$	New yy-covariance
Adjustment Step	
$K_n = P_{n,xy}^f(P_{n,yy}^f)^{-1}$	Find Kalman gain
$\mathbf{x}_{n,k}^a = \mathbf{x}_{n,k}^f + K_n(\mathbf{y}_n + \mathbf{v}_{n,k} - \mathbf{h}(\mathbf{x}_{n,k}^f))$	Find analyzed state for each ensemble member
$\bar{\mathbf{x}}_n^a = \frac{1}{q}\sum_{k=1}^q \mathbf{x}_{n,k}^a$	Mean best estimate state, after measurement adjustment
$E_{x,k}^a = \mathbf{x}_{n,k}^a - \bar{\mathbf{x}}_n^a$	Deviation of kth analyzed ensemble member from mean
$P_{n,xx}^a = \frac{1}{q-1}\sum_{k=1}^q E_{x,k}^a(E_{x,k}^a)^T$	Find new covariance (not needed to continue to next time)
end	

The algorithm in Table 2.3 shows how the stochastic collocation Kalman filter is altered when a truncated set of n_{KL} KL eigenvectors are used to represent the stochastic space. Note that in this case the dimension of the stochastic space is $m_{KL} = n_{KL} + 2$, where we have included the two parameters D and k_p. We use the same Smolyak level-one sparse grid algorithm as in SCKF, described in Section 2.3.2, but in a smaller stochastic space of dimension m_{KL}. In this case the size of the ensemble is $q_{KL} = 2m_{KL} + 1$. The vectors $\{\mathbf{r}_{KL,k}\}_{k=1}^{q_{KL}} \in \mathbf{R}^{m_{KL}}$ contain the collocation points, and $\{c_{KL,k}\}_{k=1}^{q_{KL}}$ are the collocation weights. The matrix $\mathbf{E} \in \mathbf{R}^{m\times m_{KL}}$ is a 2×2 block-diagonal matrix with diagonal blocks $[\vec{e}_1, \ldots, \vec{e}_{n_{KL}}] \in \mathbf{R}^{n_c \times n_{KL}}$ and the 2×2 identity matrix. At each assimilation step the covariance matrix $P_{n-1,xx}^a \in \mathbf{R}^{m\times m}$ is projected onto the KL eigenspace via $\mathbf{E}^T P_{n-1,xx}^a \mathbf{E} \in \mathbf{R}^{m_{KL}\times m_{KL}}$.

Finally, Table 2.4 presents the KL version of the EnKF algorithm. The only difference is that the ensemble is projected onto the KL eigenspace at each assimilation step. We have included this KF for comparison with its En and KL cohorts.

2.4. Measurements

To compare the efficiency and accuracy of the four methods for this particular model, we consider three separate sets of measurements for the given system. The first set of measurements is manufactured by running the model for 3.75 h with $k_p = 1/\text{h}$ and $D = 3 \times 10^{-6}$ cm^2/h. The second set of measurements is obtained by adding white noise with variance 3×10^{-3} to the values in the first set of measurements. The third set of measurements is taken directly from the *in vitro* experiment mentioned in Section 2. We

use a time series of images in order to determine the edge of the wound, see Fig. 9, and the measured value of e_c in a grid cell is equal to the fraction of the cell that resides outside the wound edge yielding 0 for cells entirely inside the wound edge and 1 for those entirely outside the edge. We refer to the three sets of measurements as noiseless simulated measurements, noisy simulated measurements, and real measurements. Both the simulated and real measurements are assimilated every fifteen minutes.

2.5. Comparisons

For both the noiseless and noisy simulated measurements, we calculate the parameters errors, i.e., the difference between the actual value of the parameter and the estimated value. For the real measurements case, we compare the parameter values obtained via the KF methods with parameter values obtained using a direct optimization simplex method [20]. The latter is based on minimizing the residual error

$$R(t) = \sqrt{\int_\Omega (\vec{m}^{model}(t) - \vec{m}^{experiment}(t))^2 dxdy},$$

where $\vec{m}^{model}$ is the vector of estimated measurements of state variables and parameters according to the model and $\vec{m}^{experiment}$ is the vector of actual measurements. $D^{DirectOptimization}(t_{n+1})$ and $k_p^{DirectOptimization}(t_{n+1})$ are defined as the values of D and k_p that minimize $R(t_{n+1})$ when we start with $\vec{m}^{model}(t_n) = \vec{m}^{experiment}(t_n)$. The direct optimization solution is used as a reference solution, i.e., the closer a KF result is to the direct optimization result, the more accurately the KF method estimates the parameters.

3. Results

The parameter values used for producing the simulated measurements are $D = 3 \times 10^{-6}$ for the diffusion coefficient and $k_p = 1.0$ for the growth rate. The parameter estimation methods for all three measurement types are ran with initial guesses $D = 1.5 \times 10^{-6}$ and $k_p = .5$. In addition, in Section 3.6 we present a study on the sensitivity of the results to the choice of initial parameter guesses. For each of the three types of aforementioned measurements, we use each of the four parameter estimation techniques presented in Section 2: the EnKF, SCKF, KLSCKF and KLEnKF methods. We take the model covariance matrix $Q_n \in \mathbf{R}^{m\times m}$ to be a diagonal matrix. Recall that $m = n_c + 2$, where n_c is the number of state variables, one per each grid cell, and there are two parameters. The diagonal elements of Q_n corresponding to the state variables are $0.003 \cdot s_{max}^2$, where s_{max} is an estimated maximal state variable value, which in our case is $s_{max} = e_{c,max} = 1$. Because there are no measurements of the actual parameter values, we assume a slightly larger relative uncertainty for the parameter values, setting the diagonal elements of Q_n corresponding to the parameters to $0.01 \cdot p_{init}^2$, where p_{init} is the initial parameter value. Similarly, the measurement covariance matrix $R_n \in \mathbf{R}^{l\times l}$ is taken to be diagonal. We have measurements for all state variables, so $l = n_c$. For R_n we use the same covariance values as for the state variables in Q_n, $0.003 \cdot s_{max}^2$. We note that the choice of covariance values is problem dependent and it is related to the uncertainty in the model and the measurements. In practice one often performs offline parameter tuning by computing converged values of the error covariance P_n for a range of Q_n and R_n values, see e.g. [13].

3.1. Computational cost estimates

Since the function evaluation to advance the model in the prediction step is the dominant computational cost, and each ensemble member requires one function evaluation at each data assimilation step, for comparison purpose we define the computational cost to be the size of the ensemble. Recall that for the EnKF the dimension of the stochastic space is $m = n_c + 2$, where n_c is the number of grid cells. Here $n_c = 10 \times 10 = 100$, so $m = 102$. We choose for the size of the ensemble $q = 1000$, which corresponds to 10 ensemble members per grid cell. In the SCKF, the dimension of the stochastic space is also $m = n_c + 2 = 102$, but the size of the ensemble for level one Smolyak sparse grid is $q_{SC} = 2m + 1 = 205$. In the KL-based methods the dimension of the stochastic space is independent of the number of cell in the physical grid, but depends on the number of terms in the KL expansion. In our simulations we choose $n_{KL} = 7 \times 7 = 49$ KL terms, using 7 eigenfunctions in each x and y directions. Since the KL eigenvalues decay exponentially fast, the truncated series provides a highly accurate approximation of the full one, see Section 2.3.3. The dimension of the stochastic space is $m_{KL} = n_{KL} + 2 = 51$ and the size of the SC ensemble is $q_{KL} = 2m_{KL} + 1 = 103$. Finally, in the KLEnKF, the dimension of the stochastic space is as in the KLSCKF, $m_{KL} = 51$, but the size of the ensemble needs to be chosen to provide an accurate Monte Carlo sampling. For a fair comparison to the EnKF where $q = 10n_c$, we choose here $q = 10n_{KL} = 10 \times 49 = 490$. These dimensions are summarized in Table 3.1. Note that the ensemble size of EnKF is approximately twice the cost of the KLEnKF, five times the cost of the SCKF, and ten times the cost of KLSCKF. In Table 3.1 we also report the CPU times for the real data simulations using a four core 1.73 GHz processor and note that they scale roughly linearly with ensemble number, as expected. As model complexity increases, the CPU time-ensemble number relationship will become more linear.

3.2. Noiseless simulated measurements

Fig. 3 shows a time sequence of surfaces obtained by running the model using SCKF and noiseless simulated data. The plots for the other three versions of the KF algorithm are similar and are not included. Fig. 4a and c show the Kalman filter parameter estimates as a function of time for noiseless simulated measurements. It can be seen that as time goes on, all methods converge to the actual parameter values (horizontal lines) used to produce the noiseless simulated measurements. Fig. 4b and d show the error associated with the parameter estimates. The most accurate parameter estimate is produced by the SCKF (green), but overall the accuracy in all four methods is comparable. This is also evident from the time-averaged estimates and relative errors for the parameters k_p and D given in Table 3.2. Note that the averaging is done over the time interval [2,3] in order to minimize the effect of the incorrect initial guess and the possible singular behavior at the end of the simulation when the wound is closing.

3.3. Noisy simulated measurements

The time sequences of surfaces obtained by the four KF algorithms using the noisy simulated data are similar to those obtained using the noiseless simulated data shown in Fig. 3 and are not included. Fig. 5a and c show the Kalman filter parameter estimates as a function of time for the noisy simulated measurements. All methods converge to the actual parameter values used. The convergence, however, is not nearly as tight as in the cases with noiseless simulated measurements, Fig. 5b and d show the error associated with the parameter estimates. The errors for all Kalman filter techniques are approximately the same. In this case, the accuracy of the parameter estimation techniques is limited by the noise in the measurements. As it can be seen in Table 3.3, the relative errors in the time-averaged mean estimates are slightly larger than in the noiseless measurement case.

Table 3.1
Number of parameters, stochastic space dimension, ensemble size, and CPU time for real data simulations for the four methods.

	Number of spatial parameters	Dimension of stochastic space	Ensemble size	CPU time
EnKF	$n_c = 10 \times 10 = 100$	$m = n_c + 2 = 102$	$q = 10n_c = 1000$	4.41 s
SCKF	$n_c = 10 \times 10 = 100$	$m = n_c + 2 = 102$	$q_{SC} = 2m + 1 = 205$	1.27 s
KLSCKF	$n_{KL} = 7 \times 7 = 49$	$m_{KL} = n_{KL} + 2 = 51$	$q_{KL} = 2m_{KL} + 1 = 103$	0.95 s
KLEnKF	$n_{KL} = 7 \times 7 = 49$	$m_{KL} = n_{KL} + 2 = 51$	$q = 10n_{KL} = 490$	2.44 s

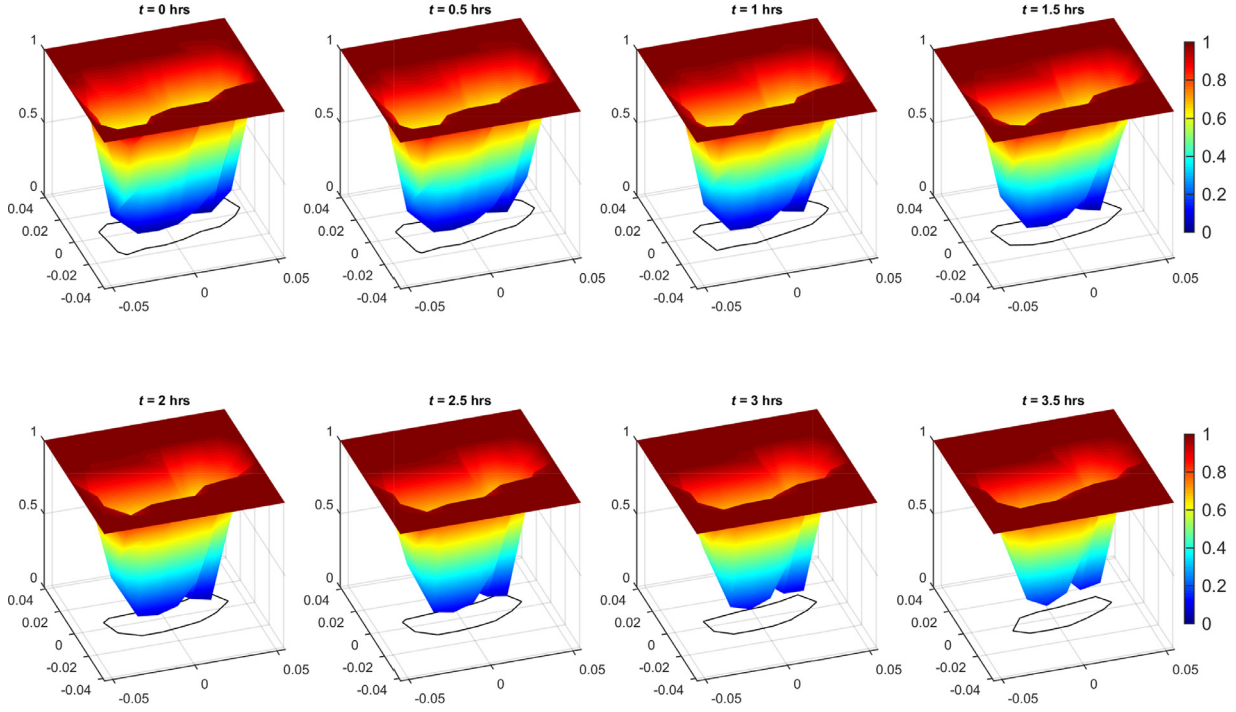

Fig. 3. Time sequence of surfaces obtained by running the model using SCKF and noiseless simulated data.

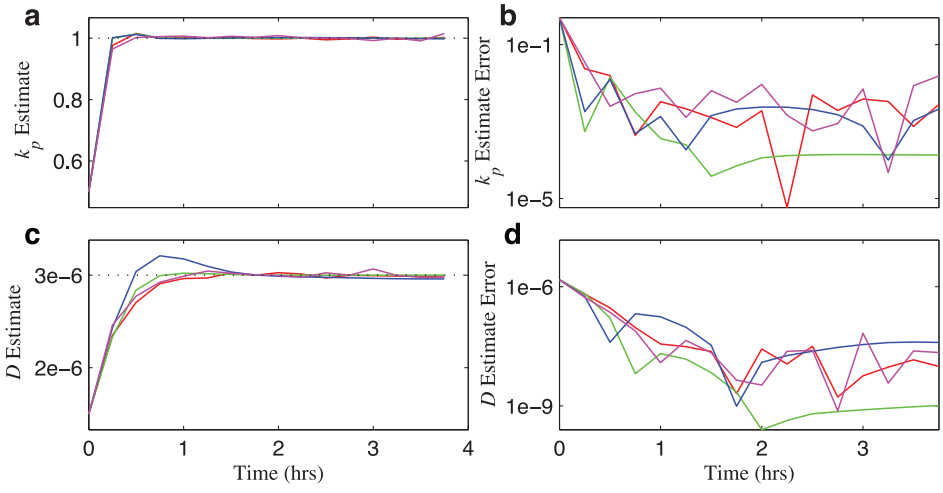

Fig. 4. Parameter estimates and errors using noiseless simulated data for the EnKF(red), SCKF(green), KLSCKF(blue) and KLEnKF(magenta). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Table 3.2
Time-averaged estimates on interval [2,3] for k_p and D using noiseless simulated data.

	k_p			D		
	Mean	Rel.Error	Std.Dev.	Mean	Rel.Error	Std.Dev.
EnKF	0.9990	0.10%	0.0028	2.9999e-06	0.003%	1.9551e-08
SCKF	1.0001	0.01%	9.2122e-06	2.9995e-06	0.010%	3.8195e-10
KLSCKF	1.0018	0.18%	6.1535e-04	2.9759e-06	0.803%	8.0139e-09
KLEnKF	1.0006	0.06%	0.0052	3.0128e-06	0.426%	3.1334e-08

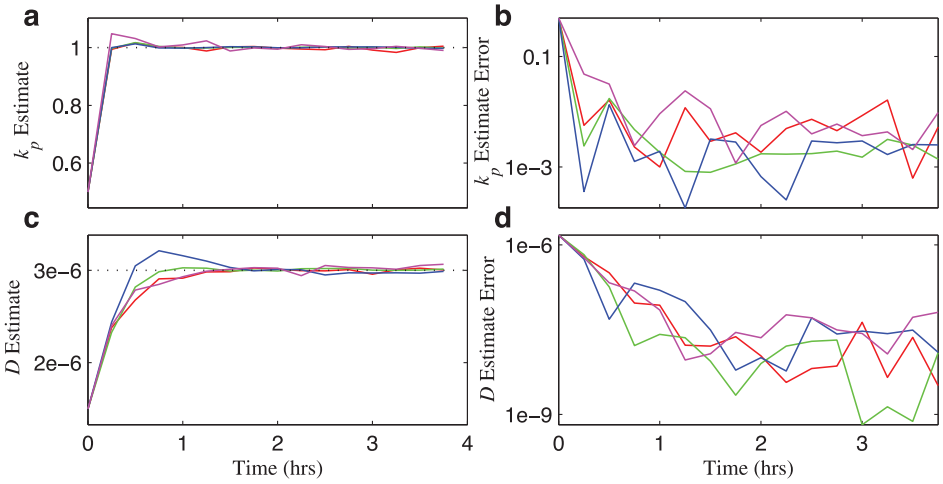

Fig. 5. Parameter estimates and errors using noisy simulated data for the EnKF(red), SCKF(green), KLSCKF(blue) and KLEnKF(magenta). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3.3
Time-averaged estimates on interval [2,3] for k_p and D using noisy simulated data.

	k_p			D		
	Mean	Rel.Error	Std.Dev.	Mean	Rel.Error	Std.Dev.
EnKF	0.9963	0.37 %	0.0046	2.9930e-06	0.23 %	1.8997e-08
SCKF	1.0002	0.02 %	0.0017	3.0098e-06	0.32 %	1.1484e-08
KLSCKF	1.0015	0.15 %	0.0016	2.9817e-06	0.61 %	2.3084e-08
KLEnKF	0.9997	0.03 %	0.0063	3.0149e-06	0.49 %	3.7723e-08

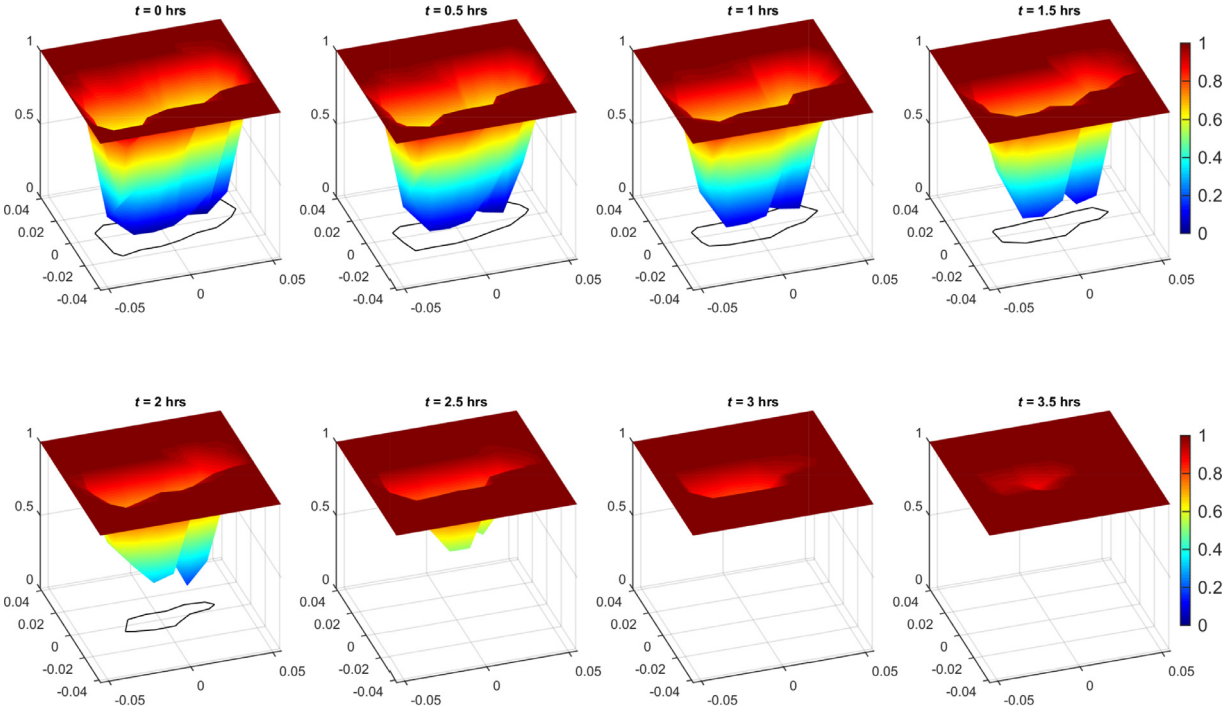

Fig. 6. Time sequence of surfaces obtained by running the model using SCKF and real data.

3.4. Real measurements

Fig. 6 shows a time sequence of surfaces obtained by running the model using SCKF and real data. We note that the wound closes faster when compared to using simulated data, see Fig. 3. This is consistent with the higher estimated values of the growth rate k_p and the diffusion coefficient D, as seen in Fig. 7 and Table 3.4. In particular, Fig. 7a and b show the Kalman filter parameter estimates as a function of time for real data measurements. For comparison we have used the parameters obtained by direct optimization (dashed line) as a best estimate. The time-averaged estimates over time interval [2,3] for all five techniques are given in Table 3.4.

We observe that the KL techniques are able to obtain parameter estimates that are in agreement with the direct parameter estimation technique for both the proliferation rate and the

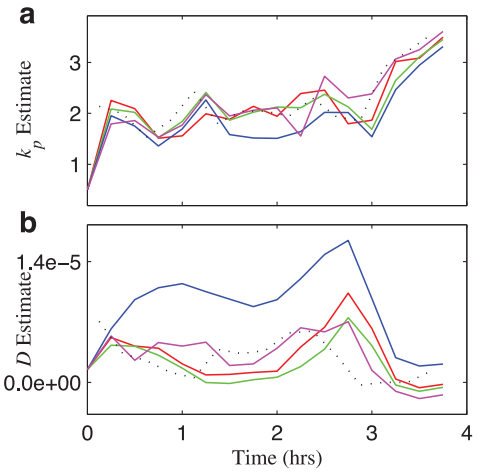

Fig. 7. Parameter estimation using real data for the EnKF(red), SCKF(green), KLSCKF(blue), KLEnKF(magenta), and Direct Optimization (dashed line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3.4
Time-averaged estimates on interval [2,3] for k_p and D using real data.

	k_p		D	
	Mean	Std.Dev.	Mean	Std.Dev.
En KF	2.0275	0.2511	5.2521e-06	2.0617e-06
SC KF	2.0848	0.2230	3.6097e-06	2.3661e-06
KL SC KF	1.7446	0.2261	1.2520e-05	2.7342e-06
KL En KF	2.0176	0.2258	5.9341e-06	2.9548e-06
Direct optimization	2.0077	0.1715	3.6211e-06	2.5226e-06

effective diffusion rate. As opposed to the two previous simulated measurement cases, the parameter estimates obtained using the parameter estimation techniques with the real measurements do not converge to one value. Rather, they appear to depend on time. There is clearly a significant variation toward the end of the simulation. This could be explained by the fact that the wound is almost closed at the end, see Figs. 8 and 9, and the parameter estimation problem becomes more ill-posed. It also suggests a possible need to adjust or add terms to the model equation in order to more closely match dynamics when the wound is nearing full closure.

3.5. Matching the experimental results

Here we demonstrate that, using the estimated parameter values, the model produces simulation results that match the *in vitro* experiment very well. We use the parameters estimated by the SCKF since the temporal variance of the SCKF estimates is closest to the temporal variance of the direct optimization technique.

We note that the parameter estimate at $t = 0$ corresponds to the initial guess for the parameters and does not incorporate any data information into that parameter estimate. Additionally, as seen in Fig. 7, as the wound closes, the parameter estimates begin to change more rapidly with respect to time. For these reasons, to obtain a single parameter estimate for the entire time course of the simulation, we take the time-averaged values of the parameter estimates appearing in Fig. 7 for $2 \le t \le 3$. This averaging on the SCKF gives parameter estimates of $D = 3.30 \times 10^{-6} \pm 1.51$ cm^2/h and $k_p = 1.99 \pm 0.25$/h.

The model is then run, without filtering, with $D = 3.30 \times 10^{-6}$ cm^2/h and $k_p = 1.99$/h. The resulting simulation produces Fig. 8. The figure shows a time sequence of the values of $e_c(x, y, t)$. To obtain an estimate of where the wound edge is, we take the $e_c = 50\%$ contour and project it into the $x - y$ plane.

To compare this wound edge estimate with the actual wound edge seen in experiment, we take the contours obtained and overlay them on the images from the *in vitro* wound healing experiment in Fig. 9. It can be seen that using the parameter estimates obtained from the SCKF parameter estimation technique produces results that are in a very good agreement with the experiment. Furthermore, comparing Fig. 8 to Figs. 3 and6, we note that using the simulated data results in a slower rate of wound closure, while the initial SCKF run with real data results in faster rate of wound closure. Running the unfiltered algorithm with the parameters estimated by the SCKF run with real data results in a simulation that best matches the rate of wound closure in the real data.

3.6. Sensitivity to initial parameter guesses

In this section we present a study on the sensitivity of the results to the choice of initial parameter guesses. Recall that all results in the previous sections are obtained with initial guesses $D = 1.5 \times 10^{-6}$ and $k_p = .5$. In Figs. 10–12 we present the parameter estimates and errors for the four algorithms with the three types of data for four cases of initial parameter guesses: 1) $D = 1.5 \times 10^{-6}$, $k_p = .5$, 2) $D = 4.5 \times 10^{-6}$, $k_p = .5$, 3) $D = 1.5 \times 10^{-6}$, $k_p = 1.5$, and 4) $D = 4.5 \times 10^{-6}$, $k_p = 1.5$. These choices are symmetrical with respect to the actual parameter values used in obtaining the simulated data, $D = 3 \times 10^{-6}$ and $k_p = 1.0$. As seen from the figures, the results indicate that all four algorithms are very robust with respect to variations in the initial parameter guesses.

4. Discussion and conclusions

We have developed four Kalman filter algorithms for data assimilation and parameter estimation for time dependent nonlinear diffusion equations and compared their performance for a model of epithelial cell migration. The methods are based on either random Monte Carlo sampling (ensemble methods) or structured stochastic collocation sampling. In addition, either uncorrelated random noise or correlated noise parameterized by the Karhunen–Loeve expansion is considered. This results in the methods EnKF, SCKF, KLSCKF, and KLEnKF. The SC methods with sparse grid collocation points provide improved approximation in stochastic space compared to Monte Carlo sampling, and thus result in comparable accuracy with a smaller ensemble size. Furthermore, KL parameterization of the noise results in a stochastic space of smaller dimension (one stochastic dimension per KL term) compared to uncorrelated noise (one stochastic dimension per element of the spatial grid). Consequently, the most efficient method is KLSCKF, followed by SCKF, KLEnKF, and EnKF, see Table 3.1.

We compared the performance of the four methods for two cases of simulated measurements, with and without noise, as well as data from *in vitro* experiment of epithelial cell migration. In all cases the four methods exhibited similar accuracy, making the more efficient methods preferable. In the simulated data cases, all methods converged to the correct parameter values for the growth rate k_p and the diffusion D, with small relative errors for the time-averaged mean value estimates. In the real measurements case, all four methods performed comparably to a much more expensive direct optimization simplex

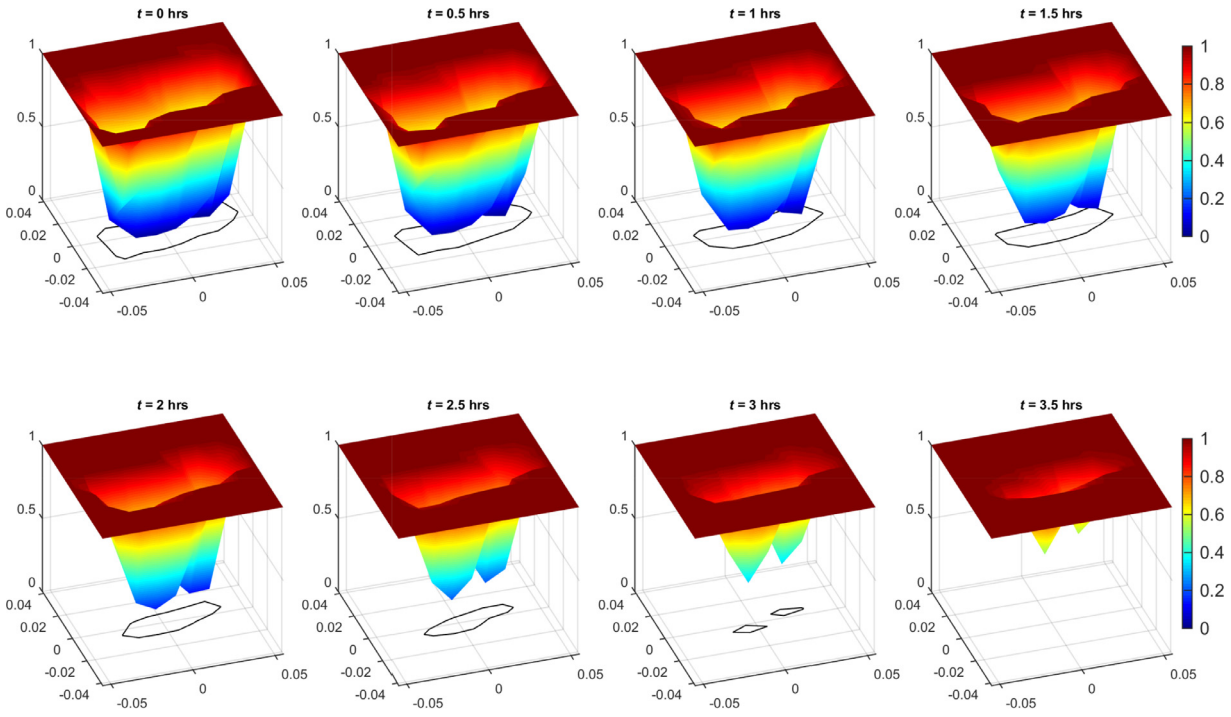

Fig. 8. Time sequence of surfaces obtained by running the model without any filtering using the parameter estimates of $D = 3.30\text{e-}6$ cm^2/h and k_p=1.99/h.

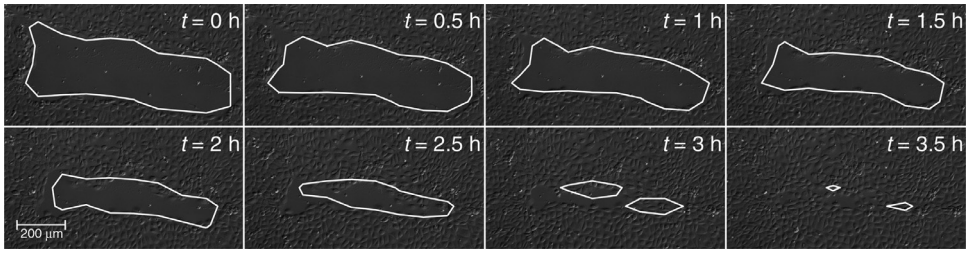

Fig. 9. Overlay of the 50% contours for e_c obtained by the model and the experimental images.

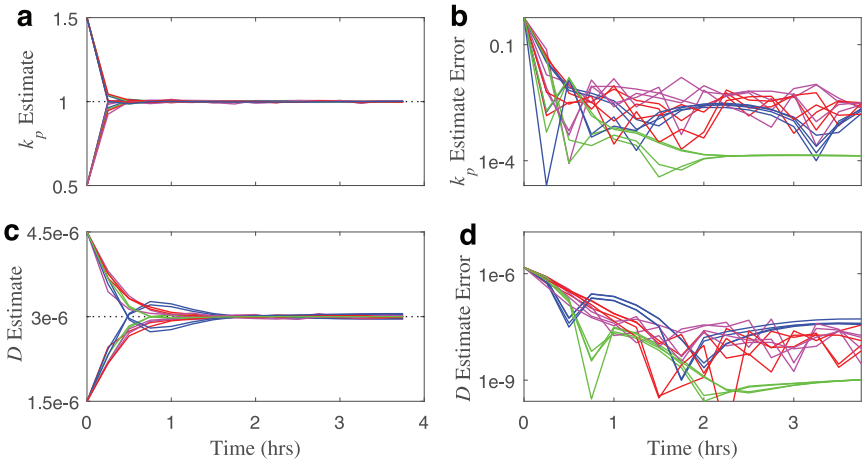

Fig. 10. Parameter estimates and errors for a range of initial parameter guesses using noiseless simulated data for the EnKF(red), SCKF(green), KLSCKF(blue) and KLEnKF(magenta). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

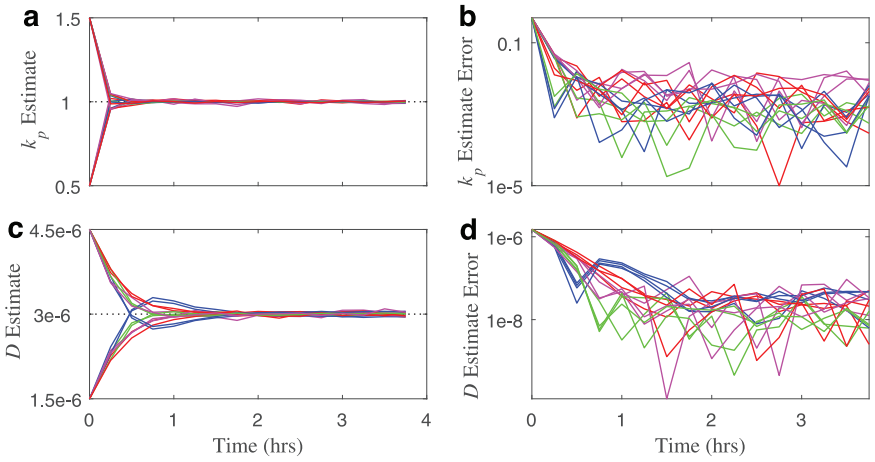

Fig. 11. Parameter estimates and errors for a range of initial parameter guesses using noisy simulated data for the EnKF(red), SCKF(green), KLSCKF(blue) and KLEnKF(magenta). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

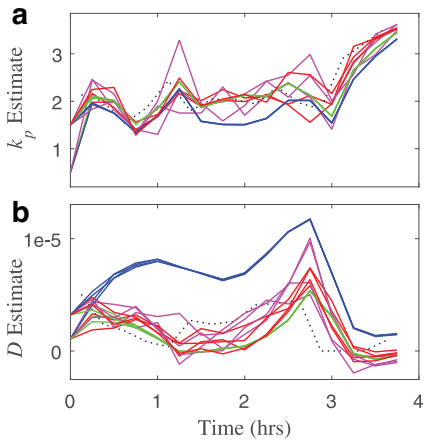

Fig. 12. Parameter estimation for a range of initial parameter guesses using real data for the EnKF(red), SCKF(green), KLSCKF(blue), KLEnKF(magenta), and Direct Optimization (dashed line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

method. The methods exhibited certain time variation in the estimated parameters, especially near the end of the simulation. This could be due to singularity in the data when the wound is almost closed, but could also indicate the need to consider a more complex model. Nevertheless using the estimated parameters provided an excellent match of the computed wound shape to the experimental data, as evident from the series of images in Fig. 9.

References

[1] J.C. Arciero, Q. Mi, M.F. Branca, D.J. Hackam, D. Swigon, Continuum model of collective cell migration in wound healing and colony expansion, Biophys. J. 100 (3) (2011) 535–543.

[2] Y.F. Atchadé, G.O. Roberts, J.S. Rosenthal, Towards optimal scaling of Metropolis-coupled Markov Chain Monte Carlo, Stat. Comput. 21 (4) (2011) 555–568.

[3] J. Barber, M. Tronzo, C. Horvat, G. Clermont, J. Upperman, Y. Vodovotz, I. Yotov, A three-dimensional mathematical and computational model of necrotizing enterocolitis, J. Theor. Biol. 322 (2013) 17–32.

[4] G. Chavent, J. Jaffre, Mathematical models and finite elements for reservoir simulation, North-Holland, Amsterdam, 1986.

[5] N. Chitnis, J.M. Hyman, J.M. Cushing, Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model, Bull. Math. Biol. 70 (5) (2008) 1272–1296.

[6] G. Clermont, S. Zenker, The inverse problem in mathematical biology, Math. Biosci. 260 (2015) 11–15.

[7] D.J. Earl, M.W. Deem, Parallel tempering: theory, applications, and new perspectives, Phys. Chem. Chem. Phys. 7 (23) (2005) 3910–3916.

[8] G. Evensen, Data assimilation: The ensemble Kalman filter, second ed., Springer-Verlag, Berlin, 2009.

[9] B. Ganis, H. Klie, M.F. Wheeler, T. Wildey, I. Yotov, D. Zhang, Stochastic collocation and mixed finite elements for flow in porous media, Comput. Methods Appl. Mech. Eng. 197 (43–44) (2008) 3547–3559.

[10] R.G. Ghanem, P.D. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag, New York, 1991.

[11] S. Gillijns, O.B. Mendoza, J. Chandrasekar, B.L.R. De Moor, D.S. Bernstein, A. Ridley, What is the Ensemble Kalman filter and how well does it work? in: Proceedings of the 2006 American Control Conference, 2006, pp. 4448–4453.

[12] G.C. Goodwin, S.F. Graebe, M.E. Salgado, Control System Design, Prentice Hall PTR, 2000.

[13] M.S. Grewal, A.P. Andrews, Kalman filtering.Theory and practice using MATLAB, fourth edition, John Wiley & Sons, Inc., Hoboken, NJ, 2015.

[14] C. Hall, T. Porsching, Numerical Analysis of Partial Differential Equations, Prentice Hall, Englewood, N.J., 1990.

[15] S.J. Julier, J.K. Uhlmann, New extension of the Kalman Filter to nonlinear systems, AeroSense'97, International Society for Optics and Photonics, 1997, pp. 182–193.

[16] X. Ma, N. Zabaras, An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method, Inverse Probl. 25 (3) (2009) 035013.

[17] Y. Marzouk, D. Xiu, A stochastic collocation approach to Bayesian inference in inverse problems, Commun. Comput. Phys. 6 (4) (2009) 826–847.

[18] Y.M. Marzouk, H.N. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, J. Comput. Phys. 228 (6) (2009) 1862–1902.

[19] F. Nobile, R. Tempone, C.G. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (5) (2008) 2309–2345.

[20] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, 2007.

[21] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, 2004.

[22] R. van der Merwe, E.A. Wan, Sigma-point Kalman Filters for probabilistic inference in dynamic state-space models, in: Proceedings of the Workshop on Advances in Machine Learning, 2003.

[23] E.A. Wan, R. van der Merwe, The Unscented Kalman Filter, in: Kalman Filtering and Neural Networks. John Wiley and Sons, 2002.

[24] D. Xiu, J.S. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27 (3) (2005) 1118–1139.

[25] L. Zeng, D. Zhang, A stochastic collocation based Kalman filter for data assimilation, Comput. Geosci. 14 (4) (2010) 721–744.

[26] S. Zenker, J. Rubin, G. Clermont, From inverse problems in mathematical physiology to quantitative differential diagnoses, PLoS Comput. Biol. 3 (11) (2007) e204.

[27] D. Zhang, Z. Lu, An efficient, high-order perturbation approach for flow in random porous media via karhunen–loève and polynomial expansions, J. Comput. Physics 194 (2) (2004) 773–794.

[28] D. Zhang, Z. Lu, Y. Chen, Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter, SPE J. 12 (1) (2007) 108–117.