Characterizing viscoelastic materials via ensemble-based data assimilation of bubble collapse observations

Jean-Sebastien Spratt[a,*], Mauro Rodriguez[a], Kevin Schmidmayer[a], Spencer H. Bryngelson[a], Jin Yang[b], Christian Franck[b], Tim Colonius[a]

[a]*Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA*
[b]*Department of Mechanical Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA*

Abstract

Viscoelastic material properties at high strain rates are needed to model many biological and medical systems. Bubble cavitation can induce such strain rates, and the resulting bubble dynamics are sensitive to the material properties. Thus, in principle, these properties can be inferred via measurements of the bubble dynamics. Estrada et al. (2018) demonstrated such bubble-dynamic high-strain-rate rheometry by using least-squares shooting to minimize the difference between simulated and experimental bubble radius histories. We generalize their technique to account for additional uncertainties in the model, initial conditions, and material properties needed to uniquely simulate the bubble dynamics. Ensemble-based data assimilation minimizes the computational expense associated with the bubble cavitation model. We test an ensemble Kalman filter (EnKF), an iterative ensemble Kalman smoother (IEnKS), and a hybrid ensemble-based 4D–Var method (En4D–Var) on synthetic data, assessing their estimations of the viscosity and shear modulus of a Kelvin–Voigt material. Results show that En4D–Var and IEnKS provide better moduli estimates than EnKF. Applying these methods to the experimental data of Estrada et al. (2018) yields similar material property estimates to those they obtained, but provides additional information about uncertainties. In particular, the En4D–Var yields lower viscosity estimates for some experiments, and the dynamic estimators reveal a potential mechanism that is unaccounted for in the model, whereby the viscosity is reduced in some cases due to material damage occurring at bubble collapse.

Keywords: A. dynamics; B. constitutive behaviour; B. viscoelastic material; C. numerical algorithms; data assimilation

1. Introduction

Measuring the mechanical properties of soft viscoelastic materials at high strain rates (exceeding $10^3 \, \mathrm{s}^{-1}$) is a challenging goal of rheometry. These measurements are of particular interest in biological and medical engineering, where high strain rates occur during impact and blast exposure (Bar-Kochba et al., 2016; Sarntinoranont et al., 2012; Meaney and Smith, 2011; Nyein et al., 2010) and during therapeutic ultrasound (Maxwell et al., 2009; Xu et al., 2007; Mancia et al., 2017; Vlaisavljevich et al., 2015; Bailey et al., 2003). Cavitation, which can take place on exposure to

---

*Corresponding author; jspratt@caltech.edu

tensile waves, induces high strain rates in surrounding materials, and the resulting bubble dynamics are sensitive to the adjacent material properties (Estrada et al., 2018). This observation led Estrada et al. (2018) to propose a high-strain rate rheometer to estimate the viscoelastic properties of polyacrylamide gels through observation of the bubble radius time history during a laser-generated cavitation event in a sample of the material. To infer the viscosity and shear modulus, they developed a least-square fitting technique which minimizes the difference between the measurements and the bubble radius history predicted through a mechanical model of a spherical bubble in an assumed neo-Hookean Kelvin–Voigt viscoelastic material.

In this paper, we generalize bubble-dynamic rheometry by considering data assimilation (DA) techniques that can potentially improve predictions in uncertainty-prone high-strain-rate regimes. Provided the material properties are observable and the dynamics are sufficiently sensitive to their values, DA provides a solution to this inverse problem that accounts for uncertainty in both the model and data (Evensen, 2009a,b; Bocquet and Sakov, 2013b; Schillings and Stuart, 2017). Thus, for bubble-dynamics-based rheometry, DA can address additional uncertainties beyond the unknown viscosity and shear modulus, including those associated with measurement noise, additional material properties, modeling assumptions, and initial conditions. DA techniques are characterized as *filters* if the state (and parameters) are updated at each moment based on the prior trajectory or *smoothers* if the state history and parameters are estimated over a horizon. By considering both filters and smoothers, we can gain additional insights into whether constant or time-varying parameters best fit the observed behavior.

Three data assimilation methods are considered: an ensemble Kalman filter (EnKF), an iterative ensemble Kalman smoother (IEnKS), and a hybrid ensemble-based 4D–Var method (En4D–Var). EnKF and IEnKS are variations on the classical Kalman Filter (KF) (Kalman, 1960). Their algorithms follow the same structure as the KF, which assimilates data at each step of a discrete time series. These are Monte Carlo methods that represent the system via an ensemble. The assimilation uses statistics of the ensemble to calculate a sample covariance. This replaces the covariance matrix of the KF and thus the covariance forecast operator, reducing computational cost. En4D–Var is a different ensemble-based approach to the assimilation problem. It is a variant of offline 4D variational data assimilation (4D–Var) (Caya et al., 2005; Trémolet, 2007), where a guess for the initial condition is iterated upon to improve the fit to the data over the entire time domain.

In section 2, we describe the specific material–bubble-dynamic model, which matches the one considered by Estrada et al. (2018). In section 3, details of the data assimilation algorithms and their implementation are provided. We then examine the relative merits of the estimators in section 4 using synthetic data generated by running the model (with additional simulated noise). This allows us to gauge their relative performance for cases with no modeling uncertainties. Next, in section 5, we apply these estimators to established experimental data for polyacrylamide gels. We compare and contrast results with each method and with the the estimates of Estrada et al. (2018). We summarize the conclusions in section 6.

2. Bubble dynamics model

A physical model for the collapsing bubbles is required to characterize the viscoelastic properties of surrounding materials. Many spherical bubble dynamics models exist. Of particular relevance here are those for cavitation in soft materials (Gaudron et al., 2015; Yang and Church, 2005) and specific numerical methods for solving them (Warnez and Johnsen, 2015; Barajas and Johnsen,

2

2017). We use the model of Estrada et al. (2018), which adopts approximations validated in previous spherical-bubble models (Prosperetti and Lezzi, 1986; Prosperetti et al., 1988; Akhatov et al., 2001; Epstein and Keller, 1972; Keller and Miksis, 1980; Preston et al., 2007). Key assumptions of this model are that the motion of the bubble and its contents are spherically symmetric, the bubble pressure is spatially uniform (homobaricity), the temperature of the surrounding material is constant, and that there is no mass transfer of the non-condensible gas across the bubble wall.

The Keller–Miksis equation models the radius evolution (Keller and Miksis, 1980),

$$\left(1 - \frac{\dot{R}}{c}\right) R\ddot{R} + \frac{3}{2}\left(1 - \frac{\dot{R}}{3c}\right)\dot{R}^2 = \frac{1}{\rho}\left(1 + \frac{\dot{R}}{c}\right)\left(p_b - \frac{2\gamma}{R} + S - p_\infty\right) + \frac{1}{\rho}\frac{R}{c}\overline{\left(p_b - \frac{2\gamma}{R} + S\right)}, \quad (1)$$

where R is the bubble radius, c the material speed of sound, ρ the material density, p_b the bubble internal pressure, γ the bubble-wall surface tension, S the stress integral (see (6)), and p_∞ the far-field pressure. Under the model assumptions, no mass or energy conservation equations are needed outside the bubble. Furthermore, the conservation of momentum simplifies to an ordinary differential equation for the bubble pressure

$$\dot{p}_b = \frac{3}{R}\left[-\kappa p_b \dot{R} + (\kappa - 1)K(T(R))\left.\frac{\partial T}{\partial r}\right|_{r=R} + \kappa p_b \frac{C_{p,v}}{C_p(k_v(R))}\frac{D}{1 - k_v(R)}\left.\frac{\partial k_v}{\partial r}\right|_{r=R}\right], \quad (2)$$

where κ is the specific heat ratio, K the thermal conductivity, T the gas temperature, C_p the specific heat, D the binary diffusion coefficient, and k_v the vapor mass fraction. Subscripts g, v, and m refer to gas, vapor, and mixture properties. Conservation of energy in the bubble interior yields an equation for the bubble temperature:

$$\rho_m C_p\left(\frac{\partial T}{\partial t} + v_m\frac{\partial T}{\partial r}\right) = \dot{p}_b + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 K\frac{\partial T}{\partial r}\right) + \rho_m(C_{p,v} - C_{p,g})D\frac{\partial k_v}{\partial r}\frac{\partial T}{\partial r}, \quad (3)$$

where v_m is the radial mixture velocity and ρ_m the mixture density. The boundary condition $T(R) = T_\infty$ follows from the model assumptions. The radial mixture velocities are computed as

$$v_m(r,t) = \frac{1}{\kappa p_b}\left[(\kappa - 1)K\frac{\partial T}{\partial R} - \frac{1}{3}r\dot{p}_b\right] + \frac{C_{p,v} - C_{p,g}}{C_{p,m}}D\frac{\partial k}{\partial r}. \quad (4)$$

Fick's law describes the mass diffusion process in the bubble. Casting the conservation of mass inside the bubble in terms of the mixture density, the vapor mass fraction inside the bubble is

$$\frac{\partial k_v}{\partial t} + v_m\frac{\partial k_v}{\partial r} = \frac{1}{\rho_m}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\rho_m D\frac{\partial k_v}{\partial r}\right). \quad (5)$$

Under the assumption of equilibrium phase change at the bubble wall, the associated boundary condition at the wall is $p_{v,sat}(T(R)) = \mathcal{R}_v k(R)\rho_m(R)T(R)$, where $p_{v,sat}$ is the saturation pressure of the vapor and $\mathcal{R}_v$ is the gas constant of the vapor.

Equations (1), (2), (3), and (5) form a system of equations. This system is evolved in time with an implicit Runge–Kutta algorithm that uses the trapezoidal rule and backwards differentiation at each step (TR–BDF2) (Hosea and Shampine, 1996). The partial differential equations for temperature and vapor mass fraction are discretized in space via a uniform grid and computed using second-order-accurate central finite differences. Estrada et al. (2018) showed that the finite-deformation neo-Hookean Kelvin–Voigt model can represent the material response at high strain

3

rates. In this framework, the material is modeled with a parallel spring (neo-Hookean elastic response with shear modulus G) and dashpot (linear viscous response with viscosity μ). The stress integral in (1) is

$$S = -\frac{G}{2}\left[5 - \left(\frac{R_0}{R}\right)^4 - 4\frac{R_0}{R}\right] - \frac{4\mu\dot{R}}{R}, \tag{6}$$

where R_0 is the equilibrium bubble radius Gaudron et al. (2015).

3. Data assimilation methods

Two difficulties that drive the choice of data assimilation method are the nonlinearity of the dynamics and large state vector required to discretize the partial differential equations adequately. The former rules out the standard linearized Kalman filter (EKF) (Kalman, 1960) and the latter renders its direct nonlinear extensions (e.g. the unscented Kalman filter, UKF) computationally prohibitive. Instead, ensemble-based methods (Evensen, 1994) are considered. They combine computational efficiency with nonlinear dynamics by approximating the state covariance via statistics of a finite (and typically small) ensemble. We consider three specific ensemble methods: an ensemble Kalman filter (EnKF), an iterative ensemble Kalman smoother (IEnKS) and a hybrid ensemble-based 4D–Var method.

The discretized equations of section 2 are re-written as a nonlinear operator F, and we define the linear observation function H that maps the state $\boldsymbol{x}$ to measurement space. This yields the discrete-time dynamical system

$$\boldsymbol{x}_{k+1} = F(\boldsymbol{x}_k) + \boldsymbol{\eta}_k, \tag{7}$$

$$\boldsymbol{y}_k = H(\boldsymbol{x}_k) + \boldsymbol{\nu}_k, \tag{8}$$

where

$$\boldsymbol{x}_k \in \mathbb{R}^d ~,~ \boldsymbol{y}_k \in \mathbb{R}^n,$$
$$\boldsymbol{\eta}_k \sim \mathcal{N}(0, \Sigma) ~,~ \boldsymbol{\nu}_k \sim \mathcal{N}(0, \Gamma),$$
$$F : \mathbb{R}^d \to \mathbb{R}^d ~,~ H : \mathbb{R}^d \to \mathbb{R}^n.$$

$\boldsymbol{x}_k$ is the d-dimensional state comprised of all the dependent variables plus the unknown parameters

$$\boldsymbol{x} = \{R, \dot{R}, p_b, S, \mathbf{T}, \mathbf{C}, \log(\mathrm{Ca}), \log(\mathrm{Re})\}, \tag{9}$$

which are the bubble-wall radius, velocity, bubble pressure, stress integral, the discretized temperature and vapor concentration fields inside the bubble, and the log-Cauchy and log-Reynolds numbers, respectively. The Cauchy and Reynolds numbers are defined as

$$\mathrm{Ca} \equiv \frac{p_\infty}{G} \quad \text{and} \quad \mathrm{Re} \equiv \frac{\sqrt{\rho p_\infty} R_{\max}}{\mu}. \tag{10}$$

These quantities appear in the nondimensionalized model equations of section 2 and the shear modulus G and viscosity μ can be computed via (10). The forecast operator F maps $\log(\mathrm{Ca})$ and $\log(\mathrm{Re})$ to themselves because they are constant in the physical model. Using the logarithm avoids

negative (and thus non-physical) values during the analysis step of the assimilation algorithms (described in sections 3.1, 3.2, and 3.4).

The variable $\boldsymbol{y}_k$ is the n-dimensional observation (data) at time k. $\boldsymbol{\eta}_k$ is the unknown process noise (or model error) added to $H(\boldsymbol{x}_k)$ to retrieve $\boldsymbol{y}_k$. It is assumed to be Gaussian with zero mean and standard deviation Σ. Similarly, $\boldsymbol{\nu}_k$ is the assumed Gaussian measurement noise added to $F(\boldsymbol{x}_k)$ to obtain $\boldsymbol{x}_{k+1}$, with zero mean and unknown standard deviation Γ. Throughout this study, the only available measurement is the bubble radius. This means that $\boldsymbol{y}_k$ is the radius only, and the observation operator H is the linear map from the state vector to its first element R. In the following, the linear operator H is sometimes represented as the matrix $\boldsymbol{H}$ for clarity ($\boldsymbol{Hx} = H(\boldsymbol{x})$).

The following methods estimate the full state vector $\boldsymbol{x}$ (including parameters of interest $\log(\mathrm{Ca})$ and $\log(\mathrm{Re})$) based on observations of $\boldsymbol{y}$. The EnKF and IEnKS are online (or quasi-online) methods—they optimize the value of $\boldsymbol{x}$ at each time through the simulation. The IEnKS is deemed quasi-online because it uses data from future times as well. The estimation trails the simulation time by a fixed number of time steps called the lag. Alternatively, the En4D–Var is an offline method, which only optimizes the initial condition for $\boldsymbol{x}$, taking into account data from the entire time-domain.

3.1. The ensemble Kalman filter

The ensemble Kalman filter (Evensen, 1994) represents the probability density function (PDF) for the state of the dynamics through the statistics of an ensemble of q state vectors. It does not require an adjoint, or deriving a tangent linear operator to the physical model (Evensen, 2003, 2009a). Starting with suitably randomized initial conditions, each ensemble member is propagated through the physical model, and the predictions are then corrected using the ensemble statistics. The ensemble is initialized with a guess for the initial condition $\boldsymbol{x}_0$ as the mean, and a given covariance corresponding to the expected error covariance. In practice, each ensemble member is independently sampled from a normal distribution with mean $\boldsymbol{x}_0$ and the assumed covariance matrix. Several initialization strategies exist depending on the system and its dynamics. In the present case, the nonlinear dynamics render a systematic approach difficult. Instead, the ensemble is initialized with a covariance corresponding to our best estimate based on simulations, and adjusted through trial and error to optimize results. An ensemble size of $q = 48$ is used for the tests. This has shown to give accurate results while keeping computational costs modest. At any given time, the estimated value for the state vector is then taken to be the ensemble average.

$$\overline{\boldsymbol{x}}_k = \frac{1}{q} \sum_{j=1}^{q} \boldsymbol{x}_k^{(j)}. \tag{11}$$

The filter is broken down into a forecast and an analysis step. In the forecast step, the physical model is used to step the state forward in time with (7). Each representation of the state vector $\boldsymbol{x}_k^{(j)}$ in the ensemble at time k is propagated through F with $\hat{\boldsymbol{x}}_{k+1}^{(j)} = F(\boldsymbol{x}_k^{(j)})$. Next, in the analysis step, if an experimental measurement $\boldsymbol{y}_{k+1}$ is available at the current time step $k+1$, then it is used to correct the forecast. As described in the dynamical system equations, each ensemble member is mapped to measurement space $H(\boldsymbol{x}_{k+1})$. The analysis proceeds by minimizing a cost function involving the difference between $H(\boldsymbol{x})$ and the data point $\boldsymbol{y}$, while accounting for measurement noise and model error. This cost function marks the key difference between the EnKF and other

5

ensemble Kalman methods. The EnKF cost function is given by

$$J(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y}_k - H(\boldsymbol{x})\|_{\boldsymbol{R}}^2 + \frac{1}{2}\|\boldsymbol{x} - \hat{\boldsymbol{x}}_k\|_{\boldsymbol{\mathcal{C}}_k}^2 \tag{12}$$

$$= \frac{1}{2}[\boldsymbol{y}_k - H(\boldsymbol{x})]^{\mathrm{T}}\boldsymbol{R}^{-1}[\boldsymbol{y}_k - H(\boldsymbol{x})] + \frac{1}{2}[\boldsymbol{x} - \hat{\boldsymbol{x}}_k]^{\mathrm{T}}\boldsymbol{\mathcal{C}}_k^{-1}[\boldsymbol{x} - \hat{\boldsymbol{x}}_k], \tag{13}$$

where $\boldsymbol{R}$ is the measurement noise covariance matrix, which is an input to the algorithm, and $\boldsymbol{\mathcal{C}}_k$ is the ensemble covariance at time step k. This covariance is defined as

$$\boldsymbol{\mathcal{C}}_k = \boldsymbol{A}_k(\boldsymbol{A}_k)^{\mathrm{T}}, \tag{14}$$

where $\boldsymbol{A}_k$ is the state perturbation matrix

$$\boldsymbol{A}_k = \frac{1}{\sqrt{q-1}}\left[\boldsymbol{x}_k^{(1)} - \overline{\boldsymbol{x}}_k, \ \dots\ , \ \boldsymbol{x}_k^{(q)} - \overline{\boldsymbol{x}}_k\right]. \tag{15}$$

In fact, the minimization does not make use of the covariance matrix directly, but instead uses the state perturbation matrix and scaled output perturbation matrix $\boldsymbol{H}\boldsymbol{A}_k$ defined as

$$\boldsymbol{H}\boldsymbol{A}_k = \frac{1}{\sqrt{q-1}}\left[\boldsymbol{y}_k^{(1)} - \overline{\boldsymbol{y}}_k, \ \dots\ , \ \boldsymbol{y}_k^{(q)} - \overline{\boldsymbol{y}}_k\right]. \tag{16}$$

The optimization is carried out by finding the minimizer $\boldsymbol{x}_k$ satisfying

$$\boldsymbol{x}_k = \hat{\boldsymbol{x}}_k + \boldsymbol{A}_k \cdot \boldsymbol{w}_k, \tag{17}$$

with $\boldsymbol{w}_k$ a correction coefficient. This restricts the solution to the subspace spanned by the scaled perturbation matrix around the prior estimate $\hat{\boldsymbol{x}}_k$. The optimization can be restated as

$$\boldsymbol{w}_k = \underset{\boldsymbol{w} \in \mathbb{R}^q}{\mathrm{argmin}}\ J(\boldsymbol{w}), \tag{18}$$

where

$$J(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{1}{2}\|\boldsymbol{y}_k - H(\hat{\boldsymbol{x}}_k) - \boldsymbol{H}\boldsymbol{A}_k(\boldsymbol{w})\|_{\boldsymbol{R}}^2. \tag{19}$$

The solution is unique, and using the Woodbury matrix identity to write the inversion in measurement space, can be written as

$$\boldsymbol{w}_k = (\boldsymbol{H}\boldsymbol{A}_k)^{\mathrm{T}}[\boldsymbol{R} + (\boldsymbol{H}\boldsymbol{A}_k)(\boldsymbol{H}\boldsymbol{A}_k)^{\mathrm{T}}]^{-1}(\boldsymbol{y}_k - H(\hat{\boldsymbol{x}}_k)). \tag{20}$$

Performing this inversion in the measurement space is in most cases more computationally efficient. Here this is clear, as the measurement space is comprised of only one variable (bubble radius). Once the minimizer is found and the analysis step complete, covariance inflation is applied to the ensemble to correct for the (typical) underestimation of the variance with finite (typically small) ensembles (see section 3.3 for details on covariance inflation). Finally, the forecast step can be repeated.

3.2. The iterative ensemble Kalman smoother

Minimizing deviation from data at future times can help to smooth out estimation and focus on longer-term trends. The IEnKS uses information from one or multiple future time steps in its assimilation, and can thus be an effective tool. While the ensemble initialization and forecast step are the same as that of the EnKF, the difference in the analysis step is twofold. First, the cost function is modified to minimize difference with data at a single or multiple future times (Evensen and van Leeuwen, 2000). The assimilation thus trails the simulation by a number of time steps (called the lag). Second, it is no longer minimized analytically but iteratively using a Gauss–Newton algorithm.

The IEnKS method used here is from Bocquet and Sakov (2013a) and Sakov et al. (2012). Bocquet and Sakov (2013b) have shown it to be effective for state and parameter estimation problems with highly nonlinear dynamics. Its cost function can take two forms referred to as 'single data assimilation' (SDA) or 'multiple data assimilation' (MDA) (Bocquet and Sakov, 2013a). The IEnKS–SDA cost function penalizes difference with measurements at a single time step $k+L$, where L corresponds to the lag of the smoother. It is given by

$$J(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y}_{k+L} - H \circ F_{k \to (k+L)}(\boldsymbol{x})\|_{\boldsymbol{R}}^2 + \frac{1}{2}\|\boldsymbol{x} - \hat{\boldsymbol{x}}_k\|_{\boldsymbol{\mathcal{C}}_k}^2. \tag{21}$$

On the other hand, the IEnKS–MDA cost function minimizes this difference over a data assimilation window (DAW) from $k+1$ to $k+L$, and is expressed as

$$J(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{L} \beta_i \|\boldsymbol{y}_{k+i} - H \circ F_{k \to (k+i)}(\boldsymbol{x})\|_{\boldsymbol{R}}^2 + \frac{1}{2}\|\boldsymbol{x} - \hat{\boldsymbol{x}}_k\|_{\boldsymbol{\mathcal{C}}_k}^2, \tag{22}$$

where β_i are weights attributed to given time steps with $\sum \beta_i = 1$. Again, a solution of the form $\boldsymbol{x} = \hat{\boldsymbol{x}}_k + \boldsymbol{A}_k \cdot \boldsymbol{w}$ is sought, but a Gauss–Newton method is used (Bocquet and Sakov, 2013a). The minimizer $\boldsymbol{w}$ is found by iterating following

$$\boldsymbol{w}_{(i+1)} = \boldsymbol{w}_{(i)} - \mathcal{H}_{(i)}^{-1}\Delta J_{(i)}(\boldsymbol{w}_{(i)}), \tag{23}$$

where i is the iteration number, and $\mathcal{H}$ is the approximate Hessian

$$\mathcal{H}_{(j)} = (q-1)\boldsymbol{I} + \boldsymbol{H}\boldsymbol{A}_{(j)}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{A}_{(j)}, \tag{24}$$

where $\boldsymbol{I}$ is the $q \times q$ identity matrix. The gradient is given by

$$\Delta J_{(j)} = -\boldsymbol{H}\boldsymbol{A}_{(j)}^{\mathrm{T}}\boldsymbol{R}^{-1}[\boldsymbol{y}_{k+L} - H \circ F_{k+L \leftarrow k}(\boldsymbol{x}_k)] + (q-1)\boldsymbol{w}_{(j)}. \tag{25}$$

For the smoother, $\boldsymbol{H}\boldsymbol{A}$ is more complicated than it is for the filter as it involves differences with measurements at future time steps. This quantity is akin to a tangent linear operator from ensemble to measurement space and has to be estimated. Following Bocquet and Sakov (2013a), a finite-difference estimate is used:

$$\boldsymbol{H}\boldsymbol{A}_{(j)} \approx \frac{1}{\alpha}H \circ F_{k+L \leftarrow k}(\boldsymbol{x}_k^{(j)}\boldsymbol{1}^{\mathrm{T}} + \alpha\boldsymbol{A}_k)\left(\boldsymbol{I} - \frac{\boldsymbol{1} \cdot \boldsymbol{1}^{\mathrm{T}}}{q}\right), \tag{26}$$

with scaling factor $\alpha \ll 1$ and $\boldsymbol{1} = (1 \ \cdots \ 1)^{\mathrm{T}}$ a vector of length q. The iteration is repeated until a threshold $\boldsymbol{w}_{(i+1)} - \boldsymbol{w}_{(i)} < \epsilon$, or a fixed number of iterations is reached. Once the optimal value

$\boldsymbol{w}_{\mathrm{opt}}$ is obtained, $\boldsymbol{x}_{\mathrm{opt}} = \hat{\boldsymbol{x}} + \boldsymbol{A}_k \cdot \boldsymbol{w}_{\mathrm{opt}}$ is calculated and a new ensemble $\boldsymbol{E}_k$ is sampled at time step k with

$$\boldsymbol{E}_k = \boldsymbol{x}_{\mathrm{opt}}\boldsymbol{1}^{\mathrm{T}} + \sqrt{q-1}\boldsymbol{A}_k\mathcal{H}_{\mathrm{opt}}^{-1/2}\boldsymbol{I}. \tag{27}$$

This completes the analysis step. When using the MDA variant, the Hessian and gradient of J are found with

$$\mathcal{H}_{(j)} = (q-1)\boldsymbol{I} + \sum_{i=1}^{L} \boldsymbol{H}\boldsymbol{A}_i^{\mathrm{T}}\beta_i\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{A}_i \tag{28}$$

$$\Delta J_{(j)} = -\sum_{i=1}^{L} \boldsymbol{H}\boldsymbol{A}_i^{\mathrm{T}}\beta_i\boldsymbol{R}^{-1}[\boldsymbol{y}_{k+i} - H \circ F_{k+i\leftarrow k}(\boldsymbol{x}_i)] + (q-1)\boldsymbol{w}_{(j)}. \tag{29}$$

3.3. Covariance inflation

While the EnKF and IEnKS may converge, ensemble methods are subject to intrinsic sampling error (Bocquet, 2011; Luo and Hoteit, 2011). This sampling error results from the finite ensemble size q used to represent the statistics of a system of often much higher dimension. As van Leeuwen (1999) explains, the EnKF tends to underestimate error variances, particularly for small ensemble sizes. There exist different ways to address this sampling error, but a simple approach is covariance inflation (Whitaker and Hamill, 2012), where we correct

$$\boldsymbol{x}^{(j)} = \overline{\boldsymbol{x}} + \alpha(\boldsymbol{x}^{(j)} - \overline{\boldsymbol{x}}) + \boldsymbol{\lambda}^{(j)}. \tag{30}$$

Here, $\overline{\boldsymbol{x}}$ denotes the ensemble average, as defined in (11), after the analysis step. Parameters α and $\boldsymbol{\lambda}$ correspond to multiplicative and additive inflation parameters, respectively.

There exist many schemes for multiplicative inflation, the most simple of which is picking a scalar α (usually $1.005 \leq \alpha \leq 1.05$). This can work well but requires extensive tuning to optimize the value for each run or data set. Instead, Whitaker and Hamill (2012) propose a scheme they call 'Relaxation Prior to Spread' (RTPS). Here, the value for α is found at each time step using

$$\alpha_i = 1 + \theta\left(\frac{\sigma_i^b - \sigma_i^a}{\sigma_i^a}\right), \tag{31}$$

where σ_i^a and σ_i^b are the prior and posterior ensemble standard deviation for the ith element of the state vector (α is a vector here), and θ is a scalar (usually $0.5 \leq \theta \leq 0.95$). As this expression for α shows, this scheme inflates the covariance more in regions where the analysis led to a large correction. Whitaker and Hamill (2012) test this method and compare it to other approaches, showing that it performs well. We similarly find that this performs as well or better than a simple scalar α for our tested cases. This RTPS model was used with $\theta = 0.7$. Additive covariance inflation was not found to significantly affect results and introduced some stability issues with larger magnitudes of $\boldsymbol{\lambda}$. Therefore, $\boldsymbol{\lambda} = \boldsymbol{0}$ is used.

3.4. A hybrid ensemble-based 4D–var method

As with ensemble Kalman methods, ensembles can be used with 4D-Var to estimate covariance empirically, thus reducing computational cost (Gustafsson and Bojarova, 2014; Liu et al., 2008).

The present method (En4D-Var) is a fully offline extension of the IEnKS–MDA method. Again, the ensemble is initialized in the same way as EnKF, but the cost function is here

$$J(\boldsymbol{x}) = \frac{1}{2}\sum_k \beta_k \|\boldsymbol{y}_k - H \circ F_{k\leftarrow 0}(\boldsymbol{x})\|_{\boldsymbol{R}}^2 + \frac{1}{2}\|\boldsymbol{x} - \hat{\boldsymbol{x}}_k\|_{\tilde{\boldsymbol{c}}_0}^2. \tag{32}$$

The difference with the IEnKS–MDA cost function is the data assimilation window size. Rather than minimizing over a few time steps forward and then stepping through time, the minimization is done over the entire time domain and only the initial state vector is corrected. Each new iteration is initialized with the corrected initial state (including parameters to estimate). The same minimization procedure as described in section 3.2 is used. When the minimization has converged, a final simulation is run with the forecast model only. In cases where only a few iterations are necessary, this method reduces computational cost as compared to the IEnKS-MDA. The time dimension is still full included, but each point in time is only assimilated once per iteration. Furthermore, this retains the advantage of ensemble methods. As opposed to classical 4D–Var, there is no need to linearize the state function and find the tangent linear adjoint operator. This novel adaptation of the IEnKS method is well suited to the present problems given that our interest is the estimation of material properties which are, at the outset, assumed to be constant.

4. Testing with synthetic data

4.1. Parameter estimation results

Synthetic data where the true shear modulus and viscosity are known is generated from the model (section 2) and used to test the data assimilation methods in a setting where there is no modeling error. Bubble radius time-history data from the simulation is sampled at 270,000 frames per second to match available experiments. Random Gaussian noise is added to these samples to mimic experimental data. The standard deviation of this noise is set at $\sigma = 0.02$, which is greater than the estimated noise of the experiments. Two polyacrylamide gels were examined with nominal values of shear modulus and viscosity determined by Estrada et al. (2018). For the stiff gel: $G_{\text{stiff}} = 7.69\,\text{kPa}$, $\mu_{\text{stiff}} = 0.101\,\text{Pa s}$, and for the soft gel: $G_{\text{soft}} = 2.12\,\text{kPa}$, $\mu_{\text{soft}} = 0.118\,\text{Pa s}$. Since similar estimation accuracy was achieved in both cases, we report results for the stiff gel only. The other material properties used are taken from Estrada et al. (2018) and given in table 1. No uncertainty is added to these parameters in the present study to match their conditions and focus on estimating G and μ.

Parameter	Value	Parameter	Value
ρ	$1060\,\text{kg/m}^3$	c	$1430\,\text{m/s}$
p_∞	$101.3\,\text{kPa}$	γ	$5.6 \times 10^{-2}\,\text{N/m}$
D	$24.2 \times 10^{-6}\,\text{m}^2/\text{s}$	κ	1.4
$C_{p,g}$	$1.62\,\text{kJ/kg K}$	$C_{p,v}$	$1.00\,\text{kJ/kg K}$
A	$5.3 \times 10^{-5}\,\text{W/m K}^2$	B	$1.17 \times 10^{-2}\,\text{W/m K}^2$
p_{ref}	$1.17 \times 10^8\,\text{kPa}$	T_{ref}	$5200\,\text{K}$
T_∞	$298.15\,\text{K}$		

Table 1: Model parameters as they follow from Estrada et al. (2018).

9

An example simulated radius curve and sampled surrogate measurements (with noise added) with these parameters is shown in figure 1a, plotted against non-dimensional time

$$t^* = \frac{t}{R_{\max}} \sqrt{\frac{p_\infty}{\rho}}. \tag{33}$$

With the simulated data, the evolution over time of all variables in the state vector is known. For example, bubble-wall velocity, bubble pressure and stress integral are plotted in figure 1b.
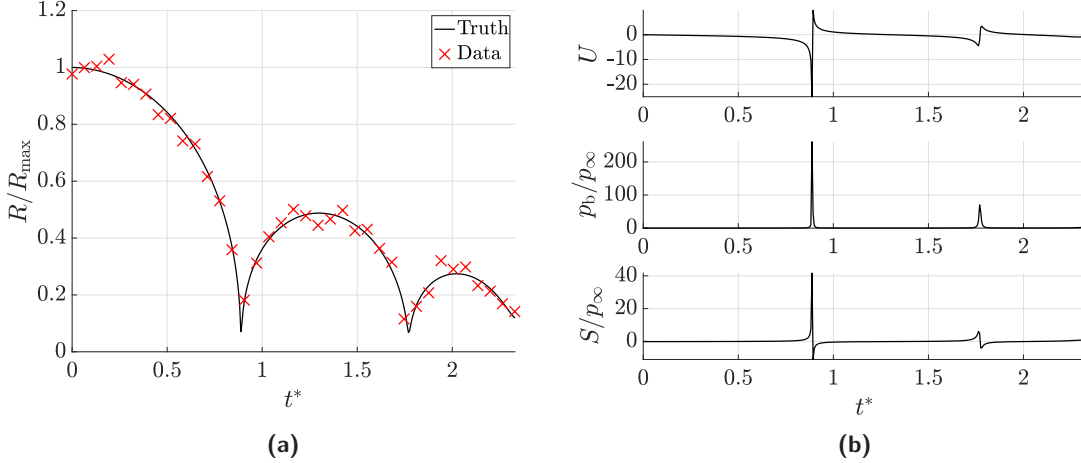

(a)   **(b)**

Figure 1: Simulated bubble radius and noisy sampled data used to test data assimilation methods **(a)**, alongside simulated bubble-wall velocity, normalized bubble pressure and stress integral **(b)**, plotted over non-dimensional time t^*.

A set of initial guesses for the shear modulus and viscosity, ranging from 10% to 100% error from the true values, were used to test each method. Table 2 summarizes results for a subset of these cases, representing 10, 50 and 100% initial error in G and μ. In each case, ensembles were initialized as Gaussian with these erroneous material properties as the mean, and standard deviation increasing with increased error. That is, the spread of the initial ensemble was made wider for cases with more error, to account for the increased uncertainty in the initial guess. To match the tests on experimental data in the next section, simulated data is limited to the first three peaks of the bubble collapse. This corresponds to approximately 35 points given the initial conditions and frame rate. Estrada et al. (2018) found that limiting the data to this region led to better parameter estimation. Similarly, we find that the model fails to fit the radius measurements after this time. Reasons for this reduced accuracy at later times are discussed in section 5.3.

10

Method	G estimate (%error) [kPa]	μ estimate (%error) [Pa s]	Run time [s]
Guess 1	8.50 (+10%)	0.09 (-10%)	–
EnKF	7.234 (5.93%)	0.098 (2.61%)	428
IEnKS–SDA (lag 1)	7.364 (4.24%)	0.110 (8.58%)	852
IEnKS–MDA (lag 3)	6.682 (13.11%)	0.100 (0.92%)	8076
En4D–Var	7.150 (7.03%)	0.099 (1.80%)	679
Guess 2	3.80 (-50%)	0.05 (-50%)	–
EnKF	3.988 (48.1%)	0.057 (43.1%)	375
IEnKS–SDA (lag 1)	4.203 (45.4%)	0.080 (20.9%)	904
IEnKS–MDA (lag 3)	7.390 (3.90%)	0.086 (15.2%)	9755
En4D–Var	7.396 (3.82%)	0.100 (0.52%)	690
Guess 3	15.0 (+100%)	0.20 (+100%)	–
EnKF	13.649 (77.5%)	0.175 (73.1%)	495
IEnKS–SDA (lag 1)	10.272 (33.6%)	0.142 (40.7%)	800
IEnKS–MDA (lag 3)	10.078 (31.1%)	0.121 (19.9%)	9802
En4D–Var	10.210 (32.7%)	0.118 (16.6%)	611

Table 2: Comparing accuracy of estimation with 3 different initial guesses for the parameters. Runs were performed on a machine with dual 12-core 2.3Ghz processors

Table 2 shows that with a relatively good initial guess with 10% error, the assimilation methods perform adequately. For example, the EnKF tracks the correct values for shear modulus and viscosity within 6% and 3% respectively. With a moderate initial error of 50%, however, the EnKF looses accuracy and barely improves on the initial guess. In some cases, the EnKF was observed to be unstable, and the initial ensemble covariance had to be limited to prevent divergence. This limited the ability of the filter to estimate the parameters of interest, and thus despite its computational efficiency, the EnKF is eliminated from further consideration.

The IEnKS and En4D–Var performed better than the EnKF for the 50% error case. The estimation was stable while varying initial conditions and covariance. Still, the lag 1 IEnKS–SDA only resulted in marginal improvements in the parameter values. The lag 3 IEnKS–MDA, on the other hand, resulted in further improvement, but at a high computational cost. This cost is associated with the calculation of the Hessian (see equation (28)) and gradient of the cost function (see equation (29)), which now involves three future time steps. The En4D–Var performs best in this test case, achieving good estimation with a comparably fast computational time. We note that while the En4D–Var was run for fifteen iterations in each case, the material property estimation converged by the fifth iterations. Thus, results and run time after five iterations are reported.

Estimation results in the case with 50% error are presented in figure 2. Figure 2d shows the suitability of the En4D–Var: both parameters converge to accurate estimates within a few iterations. Overall, figure 2 also highlights the value of looking over a time horizon. While the EnKF and lag 1 IEnKS appear to disbelieve the data too much throughout the run, taking into account multiple times enables the lag 3 IEnKS–MDA to adjust to new information well, notably around collapse. Indeed, the IEnKS–MDA significantly corrects the viscosity estimate around each collapse, and the shear modulus estimate during the second collapse. Assimilating data from single time-steps appears to be insufficient given the short time scales of bubble cavitation and limited data. Smoothing over multiple times far improves performance around collapse points, which, given the IEnKS–MDA

results, appear to hold the most pertinent data to make the necessary corrections.
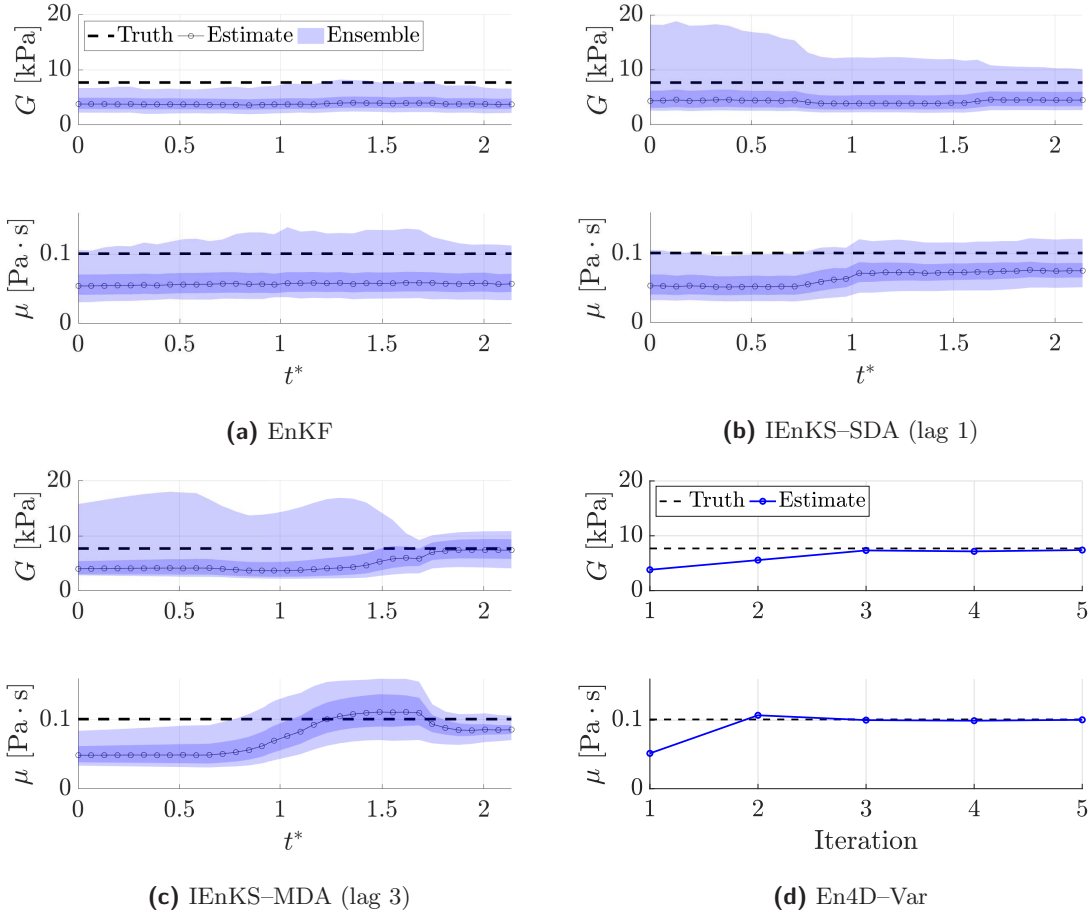

Figure 2: Estimation of shear modulus and viscosity with initial guesses of $G = 3.8\,\mathrm{kPa}$ and $\mu = 0.05\,\mathrm{Pa\,s}$ (both at 50% error). The estimation is plotted over non-dimensional time t for the EnKF and IEnKS methods, and over iteration number for the En4D–Var.

In the case with 100% error in the initial guess, the relative performances of each method are similar to the 50% error case, but the three smoothers stagnate at 20 to 40% errors for μ and G. Weighted by computational expense, the En4D–Var performs best, but the IEnKS–MDA should not be discarded. Indeed, the time-varying estimation provides additional information about potentially time-dependent modeling uncertainties. While the physical model used assumes a constant shear modulus and viscosity, the quasi-online IEnKS–MDA can uncover potential limitations of this assumption. This issue is further examined in section 5.3.

4.2. Uncertainty

Ensemble methods carry information about error statistics of the estimated parameters in the final ensemble. One way to visualize ensembles is through a histogram, an example of which is shown
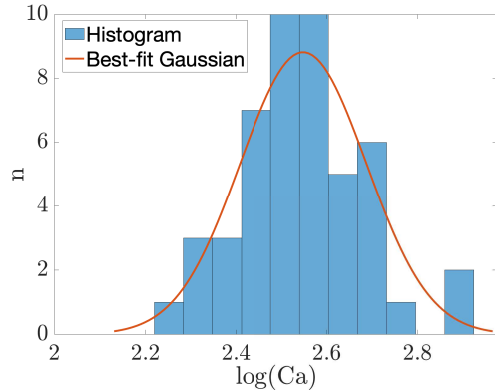
Figure 3: Histogram of the final estimate for log(Ca) with the lag 1 IEnKS and fitted normal curve, where n is the number of ensemble members at each value of log(Ca).

in figure 3 for the logarithm of the Cauchy number with the lag 1 IEnKS–SDA estimator. Despite the nonlinearity of the model, the tested methods track only the first two statistical moments of an assumed Gaussian filtering or smoothing PDF. Previous works (e.g., (Evensen and van Leeuwen, 2000; Yang et al., 2012; Katzfuss et al., 2016)) have discussed that adequate results can still be achieved with a nonlinear model where this assumption must break down to some degree. Our results for the IEnKS and En4D–Var results above confirm that this is the case in this example.

Figure 4 shows a comparison of the fitted histograms for the methods for the case with 50% initial error in both parameters. Despite imperfect estimation, the En4D–Var converges significantly more than other methods given the limited data. The IEnKS–MDA curve displays the least variance of the Kalman methods, as expected.

Anticipating experimental results, the En4D–Var was run for 10 simulated data sets with the same ground truth but different (random) noise. Results are shown in figure 5 for shear modulus and viscosity estimates over the data sets. The dashed black lines correspond to the truth, and the blue line to the mean estimate over the 10 runs. Results across these 10 data sets are fairly uniform (standard deviation of $0.98\,\text{kPa}$ for G, $0.009\,\text{Pa}\,\text{s}$ for μ), confirming that reliable estimates are obtained despite noisy measurements across data sets. Figure 6 shows a histogram combining final ensembles for shear modulus to visualize overall results. As each of the 10 ensembles should be approximately normal, a Gaussian curve is expected when combining them. Figure 6 indeed shows an approximately normal distribution, as does the equivalent histogram for viscosity (as shown in section 5.3 in figure 9a).
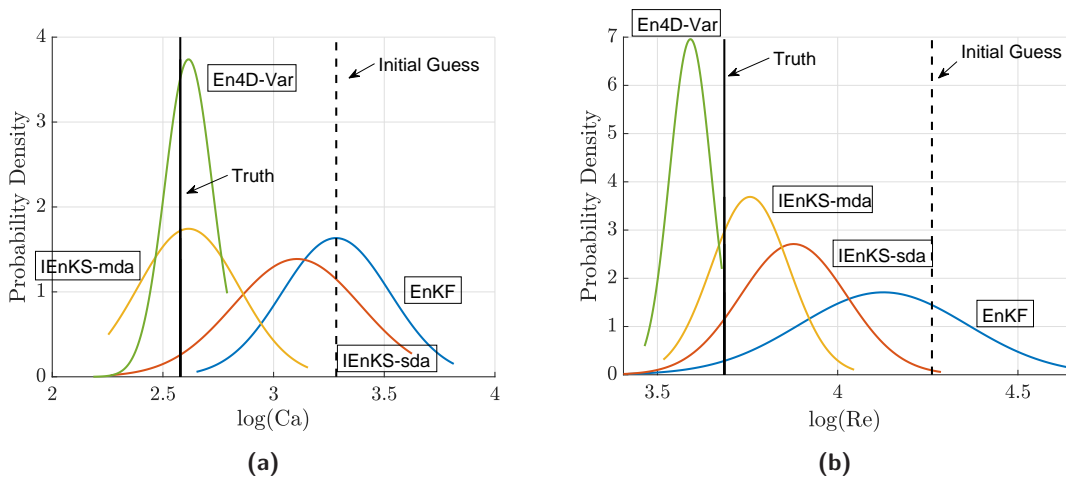
13

Figure 4: Comparing final ensembles for log(Ca) **(a)** and log(Re) **(b)** in the case with 50% initial error in both parameters.
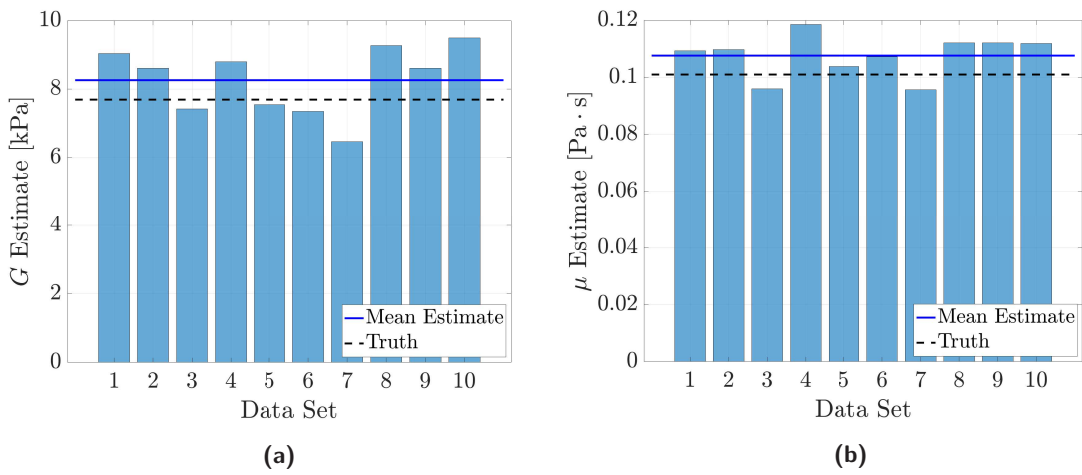

Figure 5: En4D–Var results for G **(a)** and μ **(b)**, for ten simulated data sets.
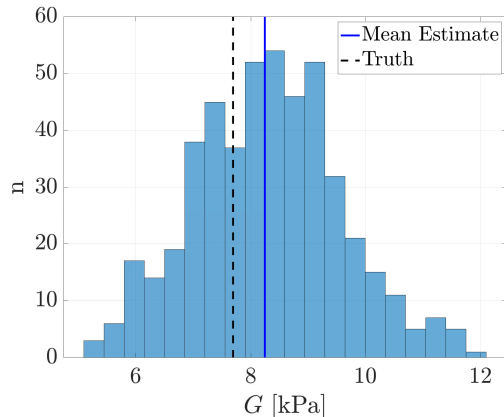
Figure 6: Histogram for G combining 10 final ensembles for simulated data runs with En4D-Var.

Based on these results with simulated data, given reasonable initial guesses as to the shear modulus and viscosity, we can confidently expect to estimate both parameters to within 5% using the 10 available data sets. Multiple initial guesses can be tested and their fits with experimental radius histories compared. Therefore, it is straightforward to formulate an initial guess with less than 50% error, and thus obtain results comparable to the second test case with simulated data. The En4D–Var is the baseline given its performance. IEnKS–MDA is also tested for its ability to estimate parameters quasi-online.

5. Applied to experimental data

5.1. Experimental setup

For a more detailed description of the experimental setup for data collection, see Estrada et al. (2018). After the polyacrylamide gel is prepared, each cavitation event is induced with a 6ns pulse of a "user-adjustable 1–50 mJ, frequency-doubled Q-switched 532 nm Nd:YAG laser". These cavitation events are triggered at different locations in the same large batch of polyacrylamide to maximize uniformity of material properties across experiments. Bubble radius is captured approximately every 3.7 μs, processing 270 000 fps high-speed camera output by subtracting a reference image from each frame and fitting a circle. A few sources of error may be present. Nonuniformity of the polyacrylamide gel or discrepancies across data sets could cause the bubble to lose spherical symmetry. Laser pulses may also vary slightly across runs, affecting the energy deposited in the system and thus initial growth conditions. In practice, a difference in maximum bubble radii was observed, with $R_{\max} = 388 \pm 35$ μm across experiments with the stiff gel, and $R_{\max} = 430 \pm 17$ μm with the soft gel. Ten experimental data sets in the stiff gel were used for the following results, in part to address this potential lack of uniformity in experimental conditions.

5.2. Initial estimates for G and μ with En4D–Var

The noise magnitude in the experimental data is smaller than what was used in the simulated data with the same data rate. Therefore, if the model is adequate and the noise accurately represented as Gaussian, the IEnKS and En4D–Var should yield comparable or better estimation results

15

with the experimental data. As with the simulated data, the assimilation window is limited to the initial collapse and two subsequent rebounds to match the setup used by Estrada et al. (2018), who estimated in the stiff-gel: $G_{\text{stiff}} = 7.69 \pm 1.12\,\text{kPa}$ and $\mu_{\text{stiff}} = 0.101 \pm 0.023\,\text{Pa\,s}$. The results are compared to theirs. Our estimation is initialized with three different initial guesses, detailed in table 3. Similarly to the surrogate truth data from the last section, initial guesses with 10%, 50% and 100% difference from the Estrada et al. (2018) estimates are chosen.

Method	G estimate $\pm\sigma$ [kPa]	μ estimate $\pm\sigma$ [Pa s]	Run time [s]
Guess 1	8.50 (+10% diff)	0.09 (-10% diff)	–
IEnKS–SDA (lag 1)	7.93 ± 1.68	0.096 ± 0.012	2751
IEnKS–MDA (lag 3)	7.51 ± 1.50	0.089 ± 0.016	9536
En4D–Var	7.41 ± 1.63	0.093 ± 0.014	609
Guess 2	3.80 (-50% diff)	0.05 (-50% diff)	–
IEnKS–SDA (lag 1)	4.32 ± 0.46	0.085 ± 0.013	2832
IEnKS–MDA (lag 3)	6.67 ± 1.43	0.083 ± 0.016	10052
En4D–Var	6.53 ± 1.58	0.090 ± 0.014	585
Guess 3	15.0 (+100% diff)	0.20 (+100% diff)	–
IEnKS–SDA (lag 1)	9.46 ± 2.76	0.114 ± 0.014	2871
IEnKS–MDA (lag 3)	8.57 ± 1.52	0.103 ± 0.015	9222
En4D–Var	8.24 ± 1.58	0.098 ± 0.016	535

Table 3: Comparing results of estimation with three different initial guesses for the parameters. Runs were again performed on a machine with dual 12-core 2.3Ghz processors

Table 3 summarizes the results from the three different initial material parameters guesses for the three methods. These estimates correspond to the mean estimate over all 10 experimental data sets. The standard deviation σ of the results is also reported.

The En4D–Var estimates for shear modulus and viscosity all fall within the error bounds provided by Estrada et al. (2018). Results are close to theirs in the test cases considered. The estimates are uniform, with only 8.5% difference between viscosity estimates, and 23% difference in the shear modulus results across the data sets. This larger difference in the shear modulus estimates and in the associated standard deviations is expected, as we have found the radius curves to be relatively more sensitive to μ than G. Finally, the average normalized root mean squared error (NRMSE) for bubble radius is low at 2.16×10^{-2} for guess 1, indicating that a good fit was achieved with this method (see equation (34) for NRMSE definition). Given that the guess–1 results lead to the smallest radius error, our initial shear and viscosity modulus estimates are $G = 7.41 \pm 1.63\,\text{kPa}$ and $\mu = 0.093 \pm 0.014\,\text{Pa\,s}$, respectively. An example bubble radius curve is shown in figure 7, for one of the experimental data sets (data set 10).
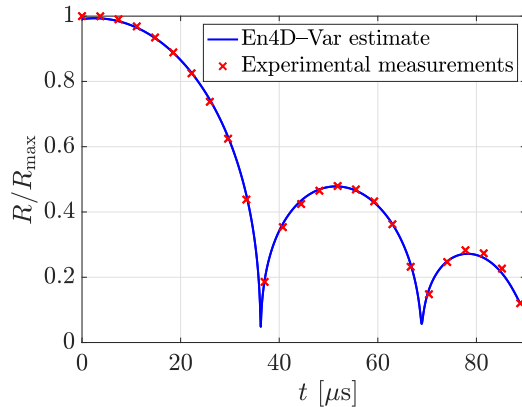
Figure 7: Radius curve given by En4D–Var estimates and experimental measurements for data set 10.

The standard deviation across the 10 IEnKS–SDA runs was comparable to the En4D–Var. However, the estimates varied significantly based on the initial guess. These ranged from $4.32\,\text{kPa}$ to $9.46\,\text{kPa}$ for shear modulus, and from $0.085\,\text{Pa s}$ to $0.114\,\text{Pa s}$ for viscosity. Except for estimates from guess 1, the shear modulus estimates are also outside the bounds given by Estrada et al. (2018), and the radius fit is significantly worse than that of the En4D–Var, with an average NRMSE of 9.10×10^{-2}.

On the other hand, while still worse than that of the En4D–Var, the IEnKS–MDA estimates are within the Estrada et al. (2018) margin, and the radius fit is better than that of the IEnKS–SDA (NRMSE $= 6.79 \times 10^{-2}$). The IEnKS–MDA thus represents the best tested quasi-online method, as expected from the simulated data results of section 4. It is important to note here that the bubble radius fits were all obtained by re-running simulations with final shear modulus and viscosity estimates and comparing to experimental measurements. This is a fair way to compare the ability of each method to estimate these parameters. However, the radius fit obtained online during assimilation with the IEnKS methods is better (and comparable with the En4D–Var estimates), given that the radius is also being directly corrected at each time-step as part of the state vector. For parameter estimation, the En4D–Var is the best tested method, but the IEnKS–MDA is a good quasi-online estimator. This is particularly useful for the discussion in section 5.3, where we make use of this time-varying estimation.

5.3. Refined estimates

The estimates obtained for the shear modulus and viscosity from the previous section show that ensemble data assimilation methods can be effectively used for estimation of viscoelastic material properties. A further look at the results, though, provides more information than simply this estimate. Examining estimates for each variable across the 10 tested experimental data sets, as shown in figure 8, there appears to be a discrepancy between data sets 3, 4, 5 and the rest for the viscosity. While the shear modulus estimation shows no discernible trend (despite the previously mentioned larger spread in results), the viscosity data appears to be split between two estimates. The red line in figure 8b shows the mean estimate of data sets 3 to 5 ($\mu = 0.074\,\text{Pa s}$) and the green line that of the rest ($\mu = 0.102\,\text{Pa s}$).

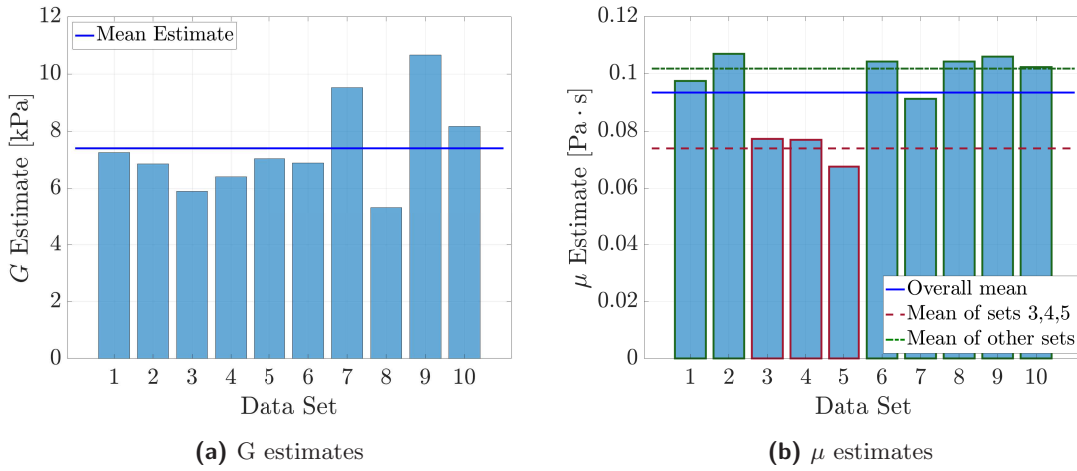**(a)** G estimates

(b) μ estimates

Figure 8: En4D–Var estimates for 10 experimental data sets.

Figure 9 compares the histogram obtained when collating the $10 \times q$ final ensemble members for viscosity from 10 runs with different simulated data but the same ground truth (Figure 9a) and the 10 runs done with experimental data (Figure 9b). As discussed in section 4.2, we expect to approximately retrieve a Gaussian distribution around the estimate, as is the case for the simulated run in figure 9a. However, figure 9b shows an apparent bimodal distribution. The lower viscosity peak corresponds to the mean estimate of data sets 3 to 5, and the higher peak to that of the rest.
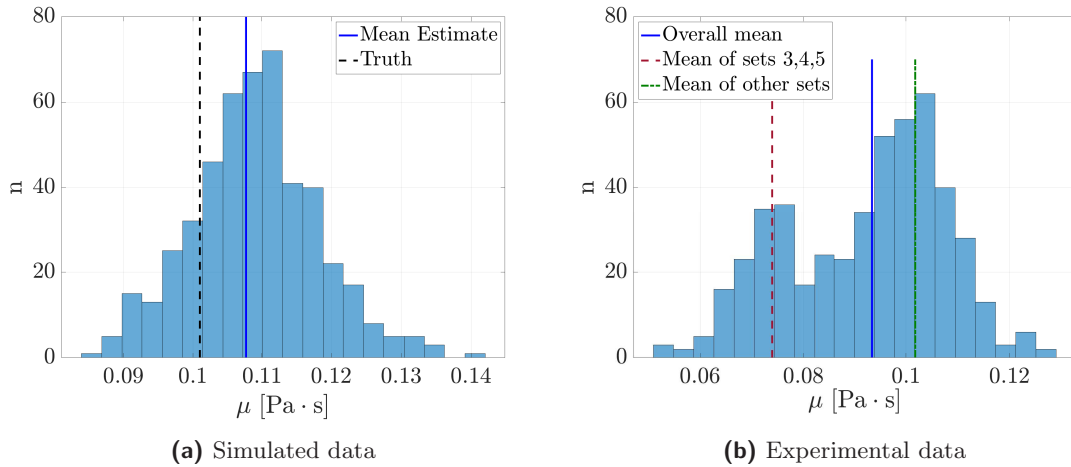

(a) Simulated data

(b) Experimental data

Figure 9: Comparing final combined ensembles for viscosity estimation in simulated and experimental data.

To understand what may be causing this discrepancy of results, it is useful to consider the IEnKS–MDA and its quasi-online estimation of viscosity. Figure 10 shows a comparison between

18

viscosity estimation for data set 2 (figure 10a) and data set 3 (figure 10b). These are representative of data sets with a high and low viscosity estimate respectively. They result in estimates of $\mu = 0.098\,\mathrm{Pa\,s}$ for data set 2, and $\mu = 0.077\,\mathrm{Pa\,s}$ for data set 3. Comparing the two data sets, it appears that the assimilation begins similarly, correcting to a higher viscosity estimate during the first collapse. However, there is a divergence between the behavior of the smoother after each collapse point, particularly the second one (around $t = 65\,\mu\mathrm{s}$). Figure 10a shows a slow decrease and convergence towards a higher viscosity value, with negligible change at the second collapse point. However, this estimate drops sharply after these collapse points in figure 10b. In fact, the data around each collapse causes the viscosity estimate to sharply drop in data set 3, which does not occur in data set 2. This behavior is representative of what is seen in data sets 3 through 5, but does not occur in the rest of the runs.
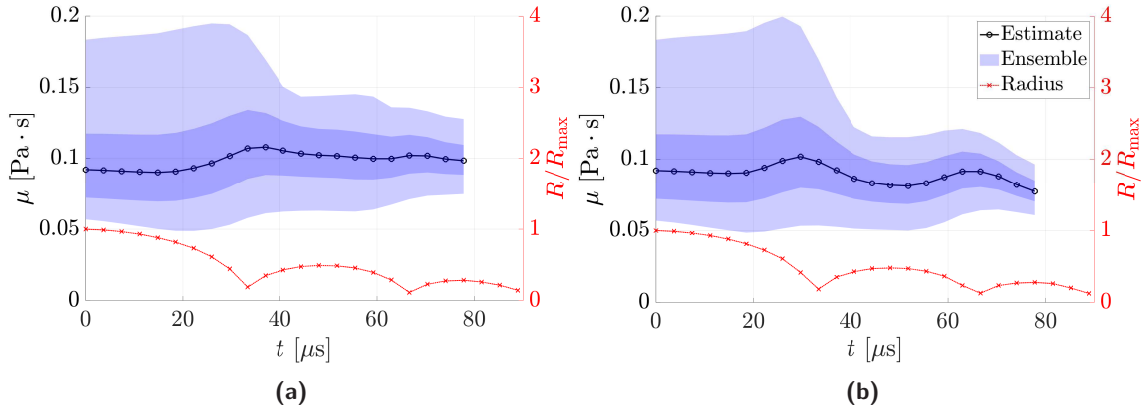

Figure 10: Comparing online estimation of viscosity in data sets 2 (**a**) and 3 (**b**) using the IEnKS–MDA.

Given the physical model used, the viscosity should be constant and such drops in the parameter are not expected. The model alone thus cannot adequately capture the behavior of the gel seen by the IEnKS in these data sets. We can posit that a violent collapse in these data sets is causing inelastic behavior in the material, and thereupon this perceived change in material properties (Yang et al., 2020). More work will be needed to determine the exact cause, but this could perhaps result from fracture, damage to the polymer network in the gel, or combustion in the gas phase (Movahed et al., 2016; Kundu and Crosby, 2009; Raayai-Ardakani et al., 2019). Regardless of physical cause, this time-dependent behavior is not accounted for in the model, but is captured by the IEnKS–MDA as a drop in the perceived viscosity.

Figure 11 compares the normalized root mean squared error (NRMSE) across all data sets, given by

$$\mathrm{NRMSE} = \frac{\sqrt{\overline{(\boldsymbol{y}_{\mathrm{sim}} - \boldsymbol{y}_{\mathrm{exp}})^2}}}{\overline{\boldsymbol{y}_{\mathrm{exp}}}}, \tag{34}$$

where $\boldsymbol{y}_{\mathrm{exp}}$ is the experimental bubble radius time history, and $\boldsymbol{y}_{\mathrm{sim}}$ the simulated time history given the estimated material properties (at the corresponding times).
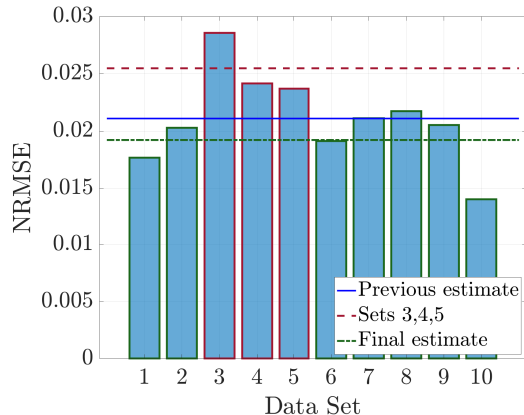
Figure 11: Bar plot of radius normalized root mean squared errors for each data set. Also plotted are the previous estimate mean NRMSE (mean of all sets), the mean NRMSE for sets 3 to 5, and the final estimate mean NRMSE (mean of all other sets).

Figure 11 shows a higher error in the estimated bubble radius curves fit for data sets 3 through 5, which is expected given the heightened model uncertainty in these data sets. Because of this uncertainty and higher error, we discard these three sets as outliers, which yields the final IEnKS–MDA-informed En4D–Var estimate reported in table 4. Notable are the drop in standard deviation for viscosity as compared to the previous En4D–Var estimate and the reduced NRMSE.

Estimate	$G \pm \sigma$ [kPa]	$\mu \pm \sigma$ [Pa s]	NRMSE
Estrada et al. (2018)	7.69 ± 1.12	0.101 ± 0.023	
Previous	7.41 ± 1.63	0.093 ± 0.014	2.16×10^{-2}
Final	$\mathbf{7.81 \pm 1.80}$	$\mathbf{0.102 \pm 0.006}$	1.95×10^{-2}

Table 4: Final En4D–Var estimates (discarding three outlier data sets) and standard deviation, along with the average radius normalized root mean squared error. The previous best estimate corresponds to the mean of all 10 data sets, outliers included.

6. Conclusions

Ensemble-based data assimilation was successfully used to estimate the mechanical properties of soft viscoelastic materials at high strain rates via observations of bubble collapse. In particular, the ensemble-based 4D–var method (En4D–Var) provided an accurate estimate efficiently, while the iterative ensemble Kalman smoother with multiple data assimilation (IEnKS–MDA) reliably estimated parameters quasi-online. Added benefits of these algorithms include adaptability to different numerical or viscoelastic models, and scalability to further parameter estimation, with negligible computational cost for additional parameters. These methods account for both model and experimental error, with noisy measurements or inaccuracies in the model having a limited impact on estimation. The ability to adjust the entire state vector rather than just the parameters to estimate limits inaccuracies from a poor initial guess. Overall, this represents a viable framework for estimation of mechanical properties of viscoelastic materials.

Using the En4D–Var and IEnKS–MDA together provided information, in this case, in the form of model error. It is hypothesized that the bubble collapses are damaging the polyacrylamide gel in certain test cases, leading to a reduced estimated viscosity after each subsequent collapse, a physical effect not accounted for in the model. Thus, while the En4D–Var provides the best estimates (especially given its relative computational efficiency), the IEnKS–MDA can provide additional information about time-dependent modeling errors.

Acknowledgements

References

I. Akhatov, O. Lindau, A. Topolnikov, R. Mettin, N. Vakhitova, and W. Lauterborn. Collapse and rebound of a laser-induced cavitation bubble. *Physics of Fluids*, 13(10):2805–2819, 2001. doi: 10.1063/1.1401810. URL https://doi.org/10.1063/1.1401810.

M. Bailey, V. Khokhlova, O. Sapozhnikov, S. Kargl, and L. Crum. Physical mechanisms of the therapeutic effect of ultrasound (a review). *Acoust Phys*, 49:369–388, 07 2003. doi: 10.1134/1.1591291. URL http://dx.doi.org/10.1134/1.1591291.

E. Bar-Kochba, M. T. Scimone, J. B. Estrada, and C. Franck. Strain and rate-dependent neuronal injury in a 3d in vitro compression model of traumatic brain injury. *Scientific Reports*, 6(1): 30550, 2016. doi: 10.1038/srep30550. URL https://doi.org/10.1038/srep30550.

C. Barajas and E. Johnsen. The effects of heat and mass diffusion on freely oscillating bubbles in a viscoelastic, tissue-like medium. *The Journal of the Acoustical Society of America*, 141(2):908–918, Feb 2017. ISSN 0001-4966. doi: 10.1121/1.4976081. URL http://dx.doi.org/10.1121/1.4976081.

M. Bocquet. Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 18(5):735–750, 2011. doi: 10.5194/npg-18-735-2011. URL https://www.nonlin-processes-geophys.net/18/735/2011/.

M. Bocquet and P. Sakov. An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1521–1535, Oct 2013a. doi: 10.1002/qj.2236. URL http://dx.doi.org/10.1002/qj.2236.

M. Bocquet and P. Sakov. Joint state and parameter estimation with an iterative ensemble Kalman smoother. *Nonlinear Processes in Geophysics*, 20(5):803–818, Oct 2013b. doi: 10.5194/npg-20-803-2013. URL http://dx.doi.org/10.5194/npg-20-803-2013.

A. Caya, J. Sun, and C. Snyder. A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Monthly Weather Review*, 133(11):3081–3094, 2005. doi: 10.1175/MWR3021.1. URL https://doi.org/10.1175/MWR3021.1.

D. Epstein and J. B. Keller. Expansion and contraction of planar, cylindrical, and spherical underwater gas bubbles. *The Journal of the Acoustical Society of America*, 52(3B):975–980, 1972. doi: 10.1121/1.1913203. URL https://doi.org/10.1121/1.1913203.

J. B. Estrada, C. Barajas, D. L. Henann, E. Johnsen, and C. Franck. High strainrate soft material characterization via inertial cavitation. *Journal of the Mechanics and Physics of Solids*, 112:291–317, 2018. doi: https://doi.org/10.1016/j.jmps.2017.12.006. URL http://www.sciencedirect.com/science/article/pii/S0022509617307585.

G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994. doi: 10.1029/94JC00572. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JC00572.

G. Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, Nov 2003. doi: 10.1007/s10236-003-0036-9. URL http://dx.doi.org/10.1007/s10236-003-0036-9.

G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2009a.

G. Evensen. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29(3):83–104, June 2009b. doi: 10.1109/MCS.2009.932223. URL http://dx.doi.org/10.1109/MCS.2009.932223.

G. Evensen and P. J. van Leeuwen. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128:1852–1867, Jun 2000.

R. Gaudron, M. T. Warnez, and E. Johnsen. Bubble dynamics in a viscoelastic medium with nonlinear elasticity. *Journal of Fluid Mechanics*, 766:54–75, Jan 2015. doi: 10.1017/jfm.2015.7. URL http://dx.doi.org/10.1017/jfm.2015.7.

N. Gustafsson and J. Bojarova. Four-dimensional ensemble variational (4d-en-var) data assimilation for the high resolution limited area model (hirlam). *Nonlinear processes in geophysics*, 21(4):745–762, 2014.

M. Hosea and L. Shampine. Analysis and implementation of TR–BDF2. *Applied Numerical Mathematics*, 20(1):21–37, 1996. doi: https://doi.org/10.1016/0168-9274(95)00115-8. URL http://www.sciencedirect.com/science/article/pii/0168927495001158.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

M. Katzfuss, J. R. Stroud, and C. K. Wikle. Understanding the ensemble Kalman filter. *The American Statistician*, 70(4):350–357, 2016. doi: 10.1080/00031305.2016.1141709. URL https://doi.org/10.1080/00031305.2016.1141709.

J. B. Keller and M. Miksis. Bubble oscillations of large amplitude. *The Journal of the Acoustical Society of America*, 68(2):628–633, Aug 1980. ISSN 0001-4966. doi: 10.1121/1.384720. URL http://dx.doi.org/10.1121/1.384720.

S. Kundu and A. J. Crosby. Cavitation and fracture behavior of polyacrylamide hydrogels. *Soft Matter*, 5:3963–3968, 2009. doi: 10.1039/B909237D. URL http://dx.doi.org/10.1039/B909237D.

C. Liu, Q. Xiao, and B. Wang. An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test. *Monthly Weather Review*, 136(9): 3363–3373, 09 2008.

X. Luo and I. Hoteit. Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Monthly Weather Review*, 139(12):3938–3953, 2011. doi: 10.1175/ MWR-D-10-05068.1. URL https://doi.org/10.1175/MWR-D-10-05068.1.

L. Mancia, E. Vlaisavljevich, Z. Xu, and E. Johnsen. Predicting tissue susceptibility to mechanical cavitation damage in therapeutic ultrasound. *Ultrasound in Medicine & Biology*, 43(7):1421 – 1440, 2017. doi: https://doi.org/10.1016/j.ultrasmedbio.2017.02.020. URL http://www.sciencedirect.com/science/article/pii/S030156291730100X.

A. D. Maxwell, C. A. Cain, A. P. Duryea, L. Yuan, H. S. Gurm, and Z. Xu. Noninvasive thrombolysis using pulsed ultrasound cavitation therapy – histotripsy. *Ultrasound in Medicine & Biology*, 35(12):1982 – 1994, 2009. doi: https://doi.org/10.1016/j.ultrasmedbio.2009.07.001. URL http://www.sciencedirect.com/science/article/pii/S0301562909014252.

D. F. Meaney and D. H. Smith. Biomechanics of concussion. *Clinics in Sports Medicine*, 30(1):19 – 31, 2011. doi: https://doi.org/10.1016/j.csm.2010.08.009. URL http://www.sciencedirect.com/science/article/pii/S027859191000061X. Concussion in Sports.

P. Movahed, W. Kreider, A. D. Maxwell, S. B. Hutchens, and J. B. Freund. Cavitation-induced damage of soft materials by focused ultrasound bursts: A fracture-based bubble dynamics model. *The Journal of the Acoustical Society of America*, 140(2):1374–1386, 2016. doi: 10.1121/1.4961364. URL https://doi.org/10.1121/1.4961364.

M. K. Nyein, A. M. Jason, L. Yu, C. M. Pita, J. D. Joannopoulos, D. F. Moore, and R. A. Radovitzky. In silico investigation of intracranial blast mitigation with relevance to military traumatic brain injury. *Proceedings of the National Academy of Sciences*, 107(48):20703–20708, 2010. doi: 10.1073/pnas.1014786107. URL https://www.pnas.org/content/107/48/20703.

A. T. Preston, T. Colonius, and C. E. Brennen. A reduced-order model of diffusive effects on the dynamics of bubbles. *Physics of Fluids*, 19(12):123302, 2007. doi: 10.1063/1.2825018. URL https://doi.org/10.1063/1.2825018.

A. Prosperetti and A. Lezzi. Bubble dynamics in a compressible liquid. part 1. first-order theory. *Journal of Fluid Mechanics*, 168:457–478, 1986. doi: 10.1017/S0022112086000460. URL http://dx.doi.org/10.1017/S0022112086000460.

A. Prosperetti, L. A. Crum, and K. W. Commander. Nonlinear bubble dynamics. *The Journal of the Acoustical Society of America*, 83(2):502–514, 1988. doi: 10.1121/1.396145. URL https://doi.org/10.1121/1.396145.

S. Raayai-Ardakani, D. R. Earl, and T. Cohen. The intimate relationship between cavitation and fracture. *Soft Matter*, 15:4999–5005, 2019. doi: 10.1039/C9SM00570F. URL http://dx.doi.org/10.1039/C9SM00570F.

P. Sakov, D. S. Oliver, and L. Bertino. An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, 140(6):1988–2004, Jun 2012. doi: 10.1175/mwr-d-11-00176.1. URL http://dx.doi.org/10.1175/MWR-D-11-00176.1.

M. Sarntinoranont, S. J. Lee, Y. Hong, M. A. King, G. Subhash, J. Kwon, and D. F. Moore. High-strain-rate brain injury model using submerged acute rat brain tissue slices. *Journal of Neurotrauma*, 29(2):418–429, 2012. doi: 10.1089/neu.2011.1772. URL https://doi.org/10.1089/neu.2011.1772. PMID: 21970544.

C. Schillings and A. M. Stuart. Analysis of the ensemble Kalman filter for inverse problems. *SIAM Journal on Numerical Analysis*, 55(3):1264–1290, 2017.

Y. Trémolet. Incremental 4D–Var convergence study. *Tellus A: Dynamic Meteorology and Oceanography*, 59(5):706–718, 2007. doi: 10.1111/j.1600-0870.2007.00271.x. URL https://doi.org/10.1111/j.1600-0870.2007.00271.x.

P. J. van Leeuwen. Comment on "Data assimilation using an ensemble Kalman filter technique". *Monthly Weather Review*, 127(6):1374–1377, 1999. doi: 10.1175/1520-0493(1999)127⟨1374:CODAUA⟩2.0.CO;2. URL https://doi.org/10.1175/1520-0493(1999)127<1374:CODAUA>2.0.CO;2.

E. Vlaisavljevich, K.-W. Lin, M. T. Warnez, R. Singh, L. Mancia, A. J. Putnam, E. Johnsen, C. Cain, and Z. Xu. Effects of tissue stiffness, ultrasound frequency, and pressure on histotripsy-induced cavitation bubble behavior. *Physics in medicine and biology*, 60(6):2271–2292, 2015. doi: 10.1088/0031-9155/60/6/2271. URL http://dx.doi.org/10.1088/0031-9155/60/6/2271.

M. T. Warnez and E. Johnsen. Numerical modeling of bubble dynamics in viscoelastic media with relaxation. *Physics of Fluids*, 27(6):063103, Jun 2015. doi: 10.1063/1.4922598. URL http://dx.doi.org/10.1063/1.4922598.

J. S. Whitaker and T. M. Hamill. Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140(9):3078–3089, 2012. doi: 10.1175/MWR-D-11-00276.1. URL https://doi.org/10.1175/MWR-D-11-00276.1.

Z. Xu, M. Raghavan, T. L. Hall, C. Chang, M. Mycek, J. B. Fowlkes, and C. A. Cain. High speed imaging of bubble clouds generated in pulsed ultrasound cavitational therapy - histotripsy. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 54(10):2091–2101, 2007.

J. Yang, H. C. Cramer, and C. Franck. Extracting non-linear viscoelastic material properties from violently-collapsing cavitation bubbles. *Extreme Mechanics Letters*, 39:100839, 2020. doi: 10.1016/j.eml.2020.100839. URL http://doi.org/10.1016/j.eml.2020.100839.

S.-C. Yang, E. Kalnay, and B. Hunt. Handling nonlinearity in an ensemble Kalman filter: Experiments with the tree-variable Lorenz model. *Montly Weather Review*, 140(8):2628–2646, 2012.

X. Yang and C. C. Church. A model for the dynamics of gas bubbles in soft tissue. *The Journal of the Acoustical Society of America*, 118(6):3595–3606, Dec 2005. doi: 10.1121/1.2118307. URL http://dx.doi.org/10.1121/1.2118307.