

Online model learning with data-assimilated reservoir computers

Andrea Nóvoa¹[0000–0003–0597–8326] and Luca Magri^{1–3}[0000–0002–0657–2611]

¹ Imperial College London, Aeronautics Dept., SW7 2AZ, London, UK.

² The Alan Turing Institute, NW1 2DB, London, UK.

³ Politecnico di Torino, DIMEAS, Corso Duca degli Abruzzi, 24 10129 Torino, Italy.
{a.novoa,l.magri}@imperial.ac.uk

Abstract. We propose an online learning framework for forecasting nonlinear spatio-temporal signals (fields). The method integrates (i) dimensionality reduction, here, a simple proper orthogonal decomposition (POD) projection; (ii) a generalized autoregressive model to forecast reduced dynamics, here, a reservoir computer; (iii) online adaptation to update the reservoir computer (the model), here, ensemble sequential data assimilation. We demonstrate the framework on a wake past a cylinder governed by the Navier-Stokes equations, exploring the assimilation of full flow fields (projected onto POD modes) and sparse sensors. Three scenarios are examined: a naïve physical state estimation; a two-fold estimation of physical and reservoir states; and a three-fold estimation that also adjusts the model parameters. The two-fold strategy significantly improves ensemble convergence and reduces reconstruction error compared to the naïve approach. The three-fold approach enables robust online training of partially-trained reservoir computers, overcoming limitations of *a priori* training. By unifying data-driven reduced order modelling with Bayesian data assimilation, this work opens new opportunities for scalable online model learning for nonlinear time series forecasting.

Keywords: Reduced order model · Ensemble Kalman filter · Echo state network · State and parameter estimation

1 Introduction

Echo State Networks (ESNs) are effective reduced-order modelling and nonlinear time series forecasting tools [11]. As reservoir computing systems, ESNs are generalized auto-regressive models that capture temporal correlations in data [1]. In fluid dynamics, ESNs have been successfully employed to model turbulent systems, capturing their dynamical and predicting extreme events [13]. However, ESNs struggle with high-dimensional spatio-temporal flows and tend to lose accuracy over long prediction horizons, thus hindering their real-time modelling capabilities [11, 4]. First, to address the dimensionality limitation, ESNs have been integrated with reduced order modelling techniques. These encode high-dimensional datasets into a reduced representation, either linearly, such as

proper orthogonal decomposition (POD) [15], nonlinearly such as convolutional autoencoders [2, 13, 10], then, the ESN predicts the evolution of the reduced system. Second, to improve the real-time predictability, ESNs have been coupled with sequential data assimilation (e.g., the ensemble Kalman filter) in a Bayesian framework to update their predictions with sparse and noisy data that become available in real time [3]. Current methods improve short-term forecasts, but do not update the ESN’s internal states or parameters, limiting their generalizability [4]. To overcome this limitation, we propose a method that extends the use of ensemble filters beyond state estimation in reservoir computers. We exploit the augmented state-space formulation [8] to seamlessly update the ESN’s state predictions and trainable parameters in real time.

2 Methodology

The test case is a two-dimensional unsteady laminar wake behind a cylinder at Reynolds number 100 [7] (see Fig.1a). We refer to the noise-free dataset as the *truth*, which is reserved for post-processing purposes only. We add Gaussian noise $\eta(\mathbf{x}) = (\epsilon * \mathbf{G})(\mathbf{x})$ to the truth, where $\epsilon \sim \mathcal{N}(0, 0.1^2)$ is convoluted with the Gaussian kernel \mathbf{G} with standard deviation of 0.1. The noisy data (Fig.1b) are used to train the POD-ESN, and for online model updates via data assimilation to simulate realistic conditions.

2.1 Non-intrusive reduced-order model: POD-ESN

We propose a POD-ESN model as a non-intrusive reduced order model. First, we apply snapshot POD¹ to our dataset $\mathbf{U} = [\mathbf{u}_{x,0}, \dots, \mathbf{u}_{x,N_t}; \mathbf{u}_{y,0}, \dots, \mathbf{u}_{y,N_t}]$, where N_t is the number of training snapshots, and $\mathbf{u}_x, \mathbf{u}_y$ are the velocity fields in the spatial coordinates $\mathbf{x} = [x; y]$. The snapshot POD decomposition is [12]

$$\mathbf{U}(\mathbf{x}, t) = \Phi(t) \Sigma \Psi^T(\mathbf{x}), \quad (1)$$

where Φ are the temporal coefficients, Σ is the matrix of singular values, and Ψ is the orthonormal basis of POD modes. To reduce the dimensionality of the system we retain the first $N_m = 4$ temporal modes, $\phi = [\phi_1; \dots; \phi_{N_m}]$, which contain 98% of the energy of the true field [7]. Figure 1 shows the reduced representation of the noisy field, which accurately reproduces the unseen true field. Second, we train an ESN to learn the time evolution of the temporal coefficients ϕ . The architecture is shown in Fig. 2. There are three key matrices in an ESN: (i) the input matrix \mathbf{W}_{in} , which is scaled by σ_{in} , maps the state, normalized by the factor \mathbf{g} , into the high-dimensional reservoir state \mathbf{r} ; (ii) the reservoir matrix \mathbf{W} , scaled by the spectral radius ρ , acts as a memory; and (iii) the output matrix

¹ In this work, we use POD to keep the model as simple as possible; however, the proposed method can be extended to nonlinear model order reduction techniques.

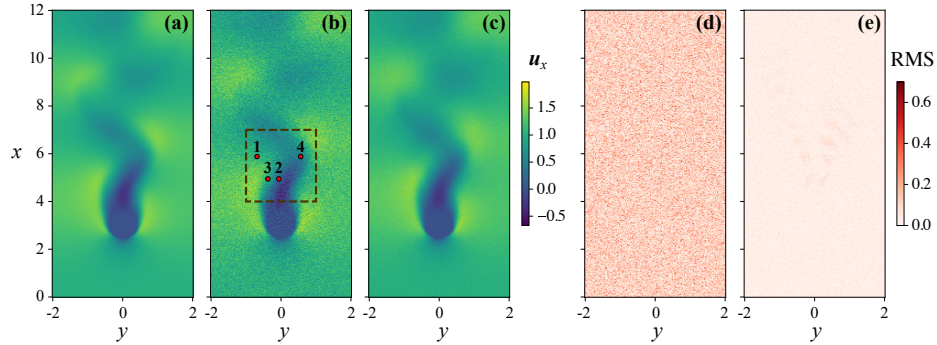


Fig. 1. Streamwise velocity field of cylinder flow. Snapshots of (a) the truth; (b) the noisy data, which are the observable fields; (c) the POD reconstruction with $N_m = 4$; and the instantaneous root mean squared (RMS) error field of panel (c) with respect to (d) the noisy data, and (e) the truth. The circles in (c) indicate the sensors, which are selected and numbered by the QR decomposition.

\mathbf{W}_{out} , which maps the reservoir state \mathbf{r} into the physical domain. With this, the ESN equations used to forecast the POD coefficients are

$$\begin{aligned}\phi_{k+1} &= \mathbf{W}_{\text{out}} [\mathbf{r}_{k+1}; 1] \\ \mathbf{r}_{k+1} &= \tanh(\sigma_{\text{in}} \mathbf{W}_{\text{in}} [\phi_k \odot \mathbf{g}; \delta] + \rho \mathbf{W} \mathbf{r}_k),\end{aligned}\quad (2)$$

where the $\tanh(\cdot)$ is performed element-wise, \odot is the Hadamard product, and $\delta = 0.1$ is a symmetry-breaking constant [5]. Training the ESN consists of finding the weights of the output matrix \mathbf{W}_{out} , while \mathbf{W}_{in} and \mathbf{W} are sparse and randomly generated matrices that are fixed [6]. The training consists of solving the linear regression problem

$$(\mathbf{R}\mathbf{R}^T + \lambda_T \mathbb{I}) \mathbf{W}_{\text{out}}^T = \mathbf{R} \mathbf{U}_{\text{train}}^T, \quad (3)$$

where $\mathbf{U}_{\text{train}}$ is the training dataset, \mathbf{R} are the reservoir states corresponding to the training set, \mathbb{I} is the identity matrix, and λ_T is the Tikhonov regularization factor. The hyperparameters ρ and σ_{in} are optimized with a recycle validation and Bayesian optimization. (For further details on the ESN training see [14, 9].)

2.2 Online learning with data assimilation

We wish to update the POD-ESN model any time that data from sensors, which may be noisy and sparse, become available. This calls for real-time data assimilation. We focus on ensemble Kalman filters because they (i) can handle nonlinearities, and (ii) assimilate data on the fly. This allows us to achieve real-time modelling using the nonlinear ESN equations. From a statistical perspective, real-time data assimilation finds the most likely state \mathbf{z}^a of the system that is consistent with both the observed data \mathbf{d} and the model prediction \mathbf{z}^f . We use

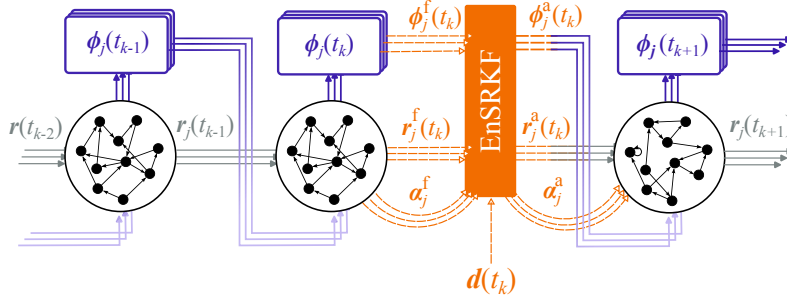


Fig. 2. Schematic of online learning with echo state networks. The ESN model forecasts autonomously the POD coefficients ϕ_k and the reservoir state \mathbf{r}_k . At the measurement time t_k , the EnSRKF updates ϕ_k^f and, optionally, \mathbf{r}_k^f and time-constant parameters α .

the superscripts ‘a’ and ‘f’ to denote analysis and forecast, respectively, following the data assimilation nomenclature. In the ensemble approach, \mathbf{z}^f is estimated by propagating an ensemble of states of size m , and computing the ensemble statistics $\mathbb{E}(\mathbf{z}) \approx \bar{\mathbf{z}} = \frac{1}{m} \sum_j \hat{\mathbf{z}}_j$, and $\mathbf{C}_{zz} = \frac{1}{m-1} \sum_j (\mathbf{z}_j - \bar{\mathbf{z}})(\mathbf{z}_j - \bar{\mathbf{z}})^T$. We define the augmented state $\hat{\mathbf{z}} = [\mathbf{z}; \mathcal{M}(\mathbf{z})]$, where $\mathcal{M} : \mathbb{R}^{N_z} \rightarrow \mathbb{R}^{N_d}$ is the measurement operator mapping the state space to the observation space of size N_d . The observations \mathbf{d} are measurements from the \mathbf{u}_x and \mathbf{u}_y fields. We compare the assimilation of the full flow field projected on the POD modes, i.e., the POD coefficients and $\mathcal{M}(\mathbf{z}) \triangleq \phi$, with the assimilation of sparse measurements such that $\mathcal{M}(\mathbf{z}) \triangleq [\mathbf{u}_x(\mathbf{x}_q); \mathbf{u}_y(\mathbf{x}_q)]$, where \mathbf{x}_q is the location of the sensors (Fig.1b). When observations become available, the ensemble is updated as [3]

$$\hat{\mathbf{z}}_j^a = \hat{\mathbf{z}}_j^f + \mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}}^f \mathbf{M}^T \left(\mathbf{C}_{dd} + \mathbf{M} \mathbf{C}_{\hat{\mathbf{z}}\hat{\mathbf{z}}}^f \mathbf{M}^T \right)^{-1} \left(\mathbf{d}_j - \mathbf{M} \hat{\mathbf{z}}_j^f \right), \quad (4)$$

where $\mathbf{M} = [\mathbf{0} \mid \mathbb{I}]$, and \mathbf{C}_{dd} is the prescribed measurement error covariance. We solve equations (4) via the square-root transform, i.e., with the ensemble square root Kalman Filter (EnSRKF). The EnSRKF can handle non-Gaussian error statistics and nonlinear dynamics and often requires smaller ensembles than traditional Kalman filters [16]. The online learning process is illustrated in Figure 2. We compare three approaches:

1. *Physical state estimation*: update the ESN prediction of the POD coefficients, $\mathbf{z} \triangleq \phi$. This naïve estimation is equivalent to performing an open-loop step.
2. *Two-fold state estimation*: update simultaneously the POD coefficients and the reservoir state (the full state) by augmenting the state vector $\mathbf{z} \triangleq [\phi; \mathbf{r}]$.
3. *Three-fold state and parameter estimation*: update simultaneously the full state and the trainable parameters of the ESN, i.e., the output matrix. We factorize \mathbf{W}_{out} using singular value decomposition to form $\mathbf{W}_{\text{out}} = \mathbf{X} \mathbf{A} \mathbf{V}^T$, where \mathbf{A} is the diagonal matrix of singular values of \mathbf{W}_{out} . With this, rather than modifying every entry in $\mathbf{W}_{\text{out}} \in \mathbb{R}^{N_r \times N_m}$, we update the N_m singular values α , and $\mathbf{z} \triangleq [\phi; \mathbf{r}; \alpha]$.

3 Results

We train a POD-ESN with $N_r = 40$ on $N_t = 250$ snapshots, which corresponds to six periods of the principal POD coefficient ϕ_1 . The ESN hyperparameters $\rho = 0.976$ and $\sigma_{\text{in}} = 0.890$, are optimized in $[0.5, 50] \times [0.2, 1.05]$ for $[\sigma_{\text{in}} \times \rho]$ (in logarithmic scale for σ_{in}). First, we compare in Figure 3 the physical state estimation and the two-fold estimation of both physical and reservoir states with an ensemble with $m = 10$, and observations from a single sensor (number 1 in Fig. 1) every $\Delta = \Delta t_d / \Delta t = 25$. We analyse the mean squared error (MSE) of the POD-ESN prediction \mathbf{u}^* with respect to the true velocity fields

$$\text{MSE} = \frac{1}{2N_x N_y} (\|\mathbf{u}_x^* - \mathbf{u}_x^{\text{truth}}\|^2 + \|\mathbf{u}_y^* - \mathbf{u}_y^{\text{truth}}\|^2) \quad (5)$$

where $\|\cdot\|$ denotes the L_2 -norm and $N_x N_y = 66177$ are the number of grid points in the domain. The ensemble converges faster with smaller uncertainty (represented by the ensemble spread) for full state assimilation (Fig. 3). This is because not only we perform an open loop step, but we also update the reservoir state. Hence, we focus on the full state for the remainder of this work. Figure 4

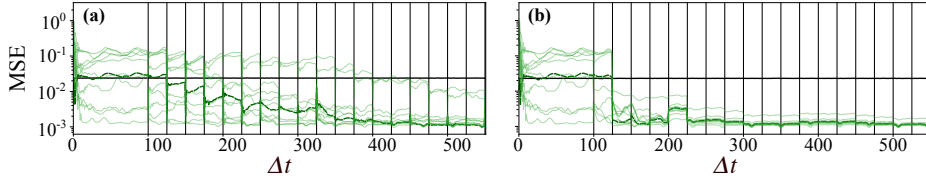


Fig. 3. Comparison between (a) physical state estimation, and (b) two-fold state estimation. Time evolution of the mean squared error (MSE) for each ensemble member (thin lines) and their mean (dashed line), and the noisy data (black).

shows the effect of the number of observations N_d in the reconstruction. With full observability, the error decreases suddenly with the first observation because we have direct information in the reduced dimension (Fig. 4a), and the projection onto the POD modes decreases the noise from the field measurement (Fig. 1). With partial observations, the more sensors, the faster the convergence, and, even when $N_d = 1$, the MSE is as small as it would be with full observability. This means that our trained POD-ESN is a *good* model of the system.

Next, we train a POD-ESN with parameters identical to earlier experiments but on a truncated training set of $N_t = 90$ (2.1 periods of ϕ_1). This yields a partially-trained POD-ESN, which does not fully capture the statistics of the system. To address this, we perform online parameter adaptation with the EnSRKF, i.e., we continue training the ESN on the fly. The EnSRKF converges to a consistently similar set of α with $m > 10$ (results not shown for brevity). A larger ensemble is needed due to the increased uncertainty and degrees of freedom. In Figure 5, we set $m = 50$ and compare the online learning with (Fig. 5 a-c)

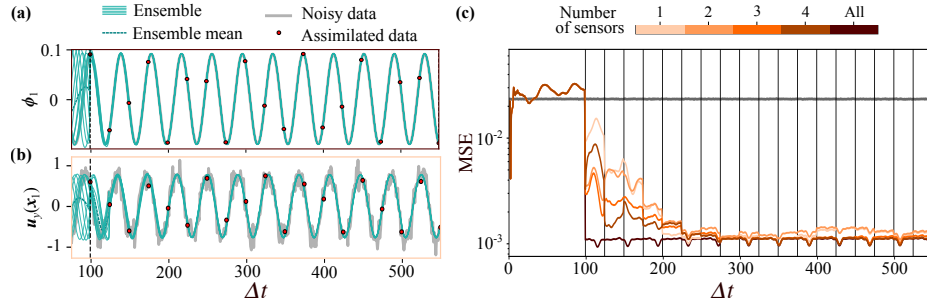


Fig. 4. Effect of the number of sensors on the full state estimation. Comparison between the evolution of the ensemble (blue) forecast and the reference data (gray) of (a) the first POD coefficient for assimilation of the full fields, and (b) the span-wise velocity for $N_d = 1$. (c) MSE of the ensembles (colormap), and the noisy data (black). The assimilation times are indicated with red dots (a-b) and vertical lines (c).

and without (Fig. 5 d-f) parameter estimation for varying Δ and N_d . On the one hand, the two-fold estimation often yields MSEs higher than the noisy data MSE. This is due to an early collapse of the ensemble covariance, i.e., the model is overfitting to the initial observations. On the other hand, the three-fold estimation avoids the covariance collapse by dynamically adapting the singular values of the ESN output matrix, thereby maintaining ensemble diversity and uncertainty quantification. This demonstrates that joint state-parameter estimation enables the POD-ESN to generalize beyond the training horizon, with the EnSRKF effectively updating \mathbf{W}_{out} to learn the system dynamics.

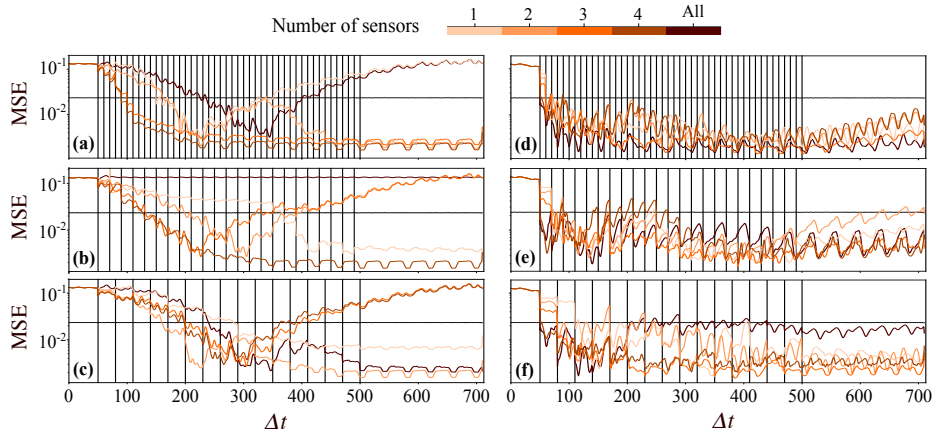


Fig. 5. Comparison between (a-c) two-fold and (d-f) three-fold estimation for varying number of sensors. Reconstruction MSE of the ensemble mean (colormap), and the noisy data (black). The vertical lines indicate the assimilation steps such that (a,d) $\Delta = 10$, (b,e) $\Delta = 20$, and (c,f) $\Delta = 30$.

4 Conclusions

We introduce an online learning approach for forecasting spatio-temporal fields that integrates real-time data assimilation with data-driven modelling. The proposed strategy combines: (i) a dimensionality reduction technique, (ii) a generalized autoregressive model, and (iii) an online data assimilation tool.

Specifically, we use snapshot proper orthogonal decomposition (POD) to reduce the dimensionality of the system; we train a reservoir computer—an echo state network (ESN)—to model the temporal evolution of the POD coefficients; and we update the POD-ESN on the fly via the ensemble square-root Kalman filter (EnSRKF). We demonstrate the method on a two-dimensional unsteady laminar wake behind a cylinder governed by the Navier-Stokes equations, using noisy velocity flow field measurements for both training and online learning. Two data assimilation strategies are explored: one that assimilates the full field projected onto the POD modes, and another that uses sparse sensors in the wake. For each strategy, we test three incremental online learning scenarios for the ESN: (i) a naïve estimation of the physical state (the POD coefficients), (ii) a two-fold estimation of the POD coefficients and the reservoir states (the memory of the network), and (iii) a three-fold estimation of the POD coefficients, reservoir states, and model parameters—specifically, by updating the singular values of the ESN’s trainable output matrix, \mathbf{W}_{out} .

We find that simultaneously updating both the physical and reservoir states reduces reconstruction error and enhances filter convergence, even with limited observability, which validates the robustness of the trained POD-ESN. Furthermore, the EnSRKF can optimally update \mathbf{W}_{out} in a Bayesian framework alongside the physical and reservoir states, therefore enabling online training of the ESN, which is of significant advantage when handling partially trained data-driven models. This work bridges the gap between offline training and real-time forecasting, opening new opportunities for online learning in nonlinear time series forecasting.

Acknowledgments. We acknowledge the support from the UKRI AI for Net Zero grant EP/Y005619/1, the ERC Starting Grant PhyCo 949388, and the grant EU-PNRR YoungResearcher TWIN ERC-PI_0000005.

References

1. Aggarwal, C.C.: Neural networks and deep learning. Springer **10**, 978–3 (2018)
2. Doan, N.A.K., Polifke, W., Magri, L.: Auto-encoded reservoir computing for turbulence learning. In: International Conference on Computational Science. pp. 344–351. Springer (2021)
3. Evensen, G.: Data assimilation: the ensemble Kalman filter. Springer Science & Business Media (2009)
4. Goswami, D., Wolek, A., Paley, D.A.: Data-driven estimation using an echo-state neural network equipped with an ensemble kalman filter. In: 2021 American Control Conference (ACC). pp. 2549–2554. IEEE (2021)

5. Huhn, F., Magri, L.: Learning ergodic averages in chaotic systems. In: International Conference on Computational Science. pp. 124–132. Springer (2020)
6. Lukoševičius, M.: A practical guide to applying echo state networks. In: Neural networks: Tricks of the trade, pp. 659–686. Springer (2012)
7. Mo, Y., Traverso, T., Magri, L.: Decoder decomposition for the analysis of the latent space of nonlinear autoencoders with wind-tunnel experimental data. *Data-Centric Engineering* **5**, e38 (2024). <https://doi.org/10.1017/dce.2024.31>
8. Nóvoa, A., Magri, L.: Real-time thermoacoustic data assimilation. *Journal of Fluid Mechanics* **948**, A35 (2022). <https://doi.org/10.1017/jfm.2022.653>
9. Nóvoa, A., Noiray, N., Dawson, J.R., Magri, L.: A real-time digital twin of azimuthal thermoacoustic instabilities. *Journal of Fluid Mechanics* **1001**, A49 (2024). <https://doi.org/10.1017/jfm.2024.1052>
10. Özalp, E., Magri, L.: Stability analysis of chaotic systems in latent spaces. *Nonlinear Dynamics* pp. 1–16 (2025)
11. Pathak, J., Hunt, B., Girvan, M., Lu, Z., Ott, E.: Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical review letters* **120**(2), 024102 (2018)
12. Poletti, R., Schena, L., Ninni, D., Mendez, M.A.: Modulo: A python toolbox for data-driven modal decomposition. *Journal of Open Source Software* **9**(102), 6753 (2024)
13. Racca, A., Doan, N.A.K., Magri, L.: Predicting turbulent dynamics with the convolutional autoencoder echo state network. *Journal of Fluid Mechanics* **975**, A2 (2023). <https://doi.org/10.1017/jfm.2023.716>
14. Racca, A., Magri, L.: Robust optimization and validation of echo state networks for learning chaotic dynamics. *Neural Networks* **142**, 252–268 (2021)
15. Sharifi Ghazijahani, M., Cierpka, C.: Echo state networks for modeling turbulent convection. *Scientific Reports* **14**(1), 1–12 (2024)
16. Tippett, M.K., Anderson, J.L., Bishop, C.H., Hamill, T.M., Whitaker, J.S.: Ensemble square root filters. *Monthly weather review* **131**(7), 1485–1490 (2003)