

Visualizing the Ensemble Kalman Filter

Sean Elvidge and David R. Themens

Abstract – The ensemble Kalman filter is a widely used technique across the many radio science domain areas. Despite this, conventional descriptions of Kalman filters can, at times, be somewhat opaque and difficult to conceptualize, making them difficult to approach for many users. This article presents a visualization of the underlying mathematical equations to provide a more intuitive understanding of the steps of the filter.

1. Introduction

The Kalman filter [1] is a powerful tool used to combine noisy measurements to obtain a better estimation of a system. Since its introduction in 1960, the filter has become of increasing importance across science and engineering. Part of its appeal is its broad applicability, from missile tracking [2] to economics [3], many disciplines have applied Kalman’s technique. Although the Kalman filter is very popular and used in a variety of applications, it can seem obscure to many people because of the mathematics.

Simplistically, the Kalman filter is a two-step process. First, an estimate of the current system (for example, in missile tracking this would be position and speed) and uncertainties are required. The filter then updates these estimates with measurements (which also have some level of uncertainty) by using a weighted average (more uncertainty means less weight). The updated estimate of the system is then propagated forward to the next time step by using a model of the underlying system.

One of the limitations of the Kalman filter is its computational expense for large systems. To address this problem, [4] developed the *ensemble Kalman filter* (EnKF). Essentially, the EnKF just replaces the term in the Kalman filter that tells us how the variables in our system are related to each other (which is called a *covariance matrix*), with an approximation (sample covariance matrix). This approximation greatly reduces the computational cost of the process that has allowed it to be applied to many more problems. The use of the EnKF is rapidly growing, and while it is simple enough to write down the mathematics, the equations alone do not provide much insight into this technique.

2. Visualization

Imagine a simple system with two variables ($x = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$). For example, this could be a position

Manuscript received 22 December 2021.

Sean Elvidge and David R. Themens are with the Space Environment (SERENE) Group, University of Birmingham, Birmingham B15 2TT, United Kingdom; e-mail: s.elvidge@bham.ac.uk, d.r.themens@bham.ac.uk.

(latitude and longitude) or the properties of a medium (temperature and pressure of a gas). Also, imagine we have a model that provides an estimate of these at a given time. Slightly tweaking how the model is run gives us slightly different estimates of the system. Such a model describes a space of possible outputs, called a *probability distribution function*.

The naive best guess of the true values of our system would be to run the model many times, using realistic inputs, and then take the average of the outputs. However, if we then obtain a measurement (observation) of the system, with a known measurement error, the EnKF can be used to provide an updated estimate of the system by using both the current model estimate and the observation.

The different estimates of the system (by repeatedly running the model; Figure 1a) are used to estimate the space that the model covers (estimating a probability distribution). We do this by seeing how far each of the model outputs are away from the average (Figure 1b).

Mathematically, we can describe this by

$$\mathbf{X} = [x_1, x_2, \dots, x_n] \quad (1)$$

where each x_i represents one run of the model (an ensemble member). \mathbf{X}^b is then defined as the matrix that represents how far each member is from the average (\bar{x}). We call this the *perturbation matrix*:

$$\mathbf{X}^b = [x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}]. \quad (2)$$

The covariance matrix (how the systems variables are related) can then be estimated by

$$\mathbf{B} = \frac{1}{n-1} \mathbf{X}^b (\mathbf{X}^b)^T. \quad (3)$$

As the number of ensemble members (model runs) increases the estimated covariance matrix approaches the true covariance matrix.

We now want to update this estimate with an observation of the system. Because our observation (y_o) is not completely reliable, the measurements also have an associated uncertainty (\mathbf{R} ; the error is assumed to be Gaussian). This observation and associated error

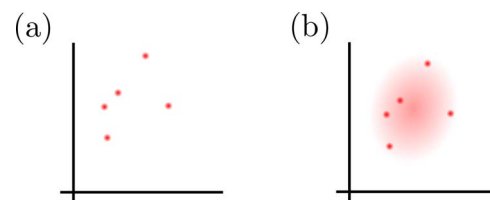


Figure 1. The individual estimates of the state are in red and are used to estimate the underlying probability distribution of our current estimate (shaded region).

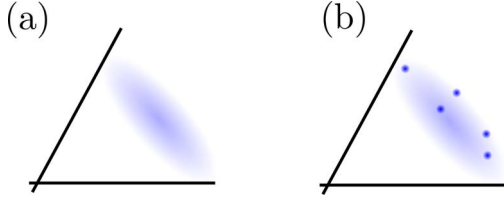


Figure 2. The underlying distribution of the observation is shaded blue. The blue circles are our (random) samples of this distribution. The nonperpendicular axes are used to highlight the difference between the model and observations.

describe a space of possible values (probability distribution; see Figure 2a, the nonperpendicular axes are to distinguish between the model variables and observation variables). We sample this distribution (Figure 2b) as many times as we have ensemble members. This can be accomplished by adding (Gaussian) noise to the observation to form the matrix \mathbf{Y}

$$\mathbf{Y} = [y_o + \varepsilon_1, y_o + \varepsilon_2, \dots, y_o + \varepsilon_n] \quad (4)$$

where each (ε_i) is drawn from the Gaussian distribution described by the observation uncertainties (\mathbf{R}) .

We do not need an observation of every variable in the system (because we are estimating in the EnKF the relationship between them) or even for the observation to be of exactly the same variables as our system. However, we do need a way of mapping between the variables in our model to the observation variables. This mapping function (called an *observation operator*, denoted by H), transforms our estimate into *observation space* (Figure 3; the mapping from model space to observation space is highlighted by the changing axes). This gives us the ensemble members $(H\mathbf{X})$ and model covariances $(\mathbf{B}H^T)$ in observation space.

We now look at the differences between the model values and observations. This is done by simply subtracting ensemble members from the observations (in observation space; Figure 4a), resulting in a set of differences (Figure 4b). These differences, \mathbf{D} , are given by

$$\mathbf{D} = \mathbf{Y} - H\mathbf{X} \quad (5)$$

We want to update our current estimate with these differences. Exactly how we update the estimate (as well as the uncertainties) depends on the relative errors

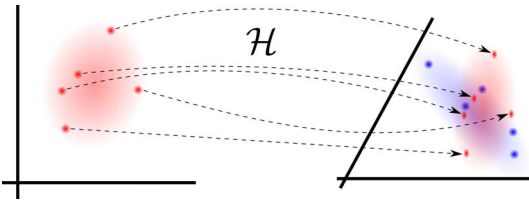


Figure 3. H maps the model values (red) into the same space as the observation values (blue).

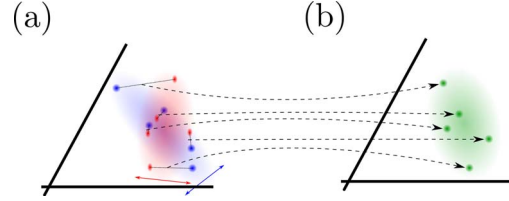


Figure 4. The green circles (\mathbf{D}) are found by subtracting the red from the blue. The red and blue arrows in (a), which represent the uncertainty, are used to calculate the weighting to be used in updating the estimate.

between the estimate and observation (the relative lengths of the red and blue arrows, shown in Figure 4a).

The differences (shown in green, Figure 4b) are added to the members (red) by using a weighting function. The weighting used (denoted by \mathbf{K} , called the *Kalman gain*) is based on how much the current estimate uncertainty contributes to the total uncertainty (the current plus measurement uncertainty). In a simple scalar case, for two values of uncertainty α and β , the contribution of α to the total uncertainty would be

$$\frac{\alpha}{\alpha + \beta} = \alpha(\alpha + \beta)^{-1}. \quad (6)$$

With the model errors given by \mathbf{B} (3) and the observation errors given by \mathbf{R} , the Kalman gain is calculated (in observation space) as

$$\mathbf{K} = \mathbf{B}H^T(\mathbf{H}\mathbf{B}H^T + \mathbf{R})^{-1} \quad (7)$$

which is the same as (6) but with the necessary observation operator H and a matrix transpose.

Each individual ensemble member is then updated (and the model uncertainty) by combining the differences \mathbf{D} and our original estimate \mathbf{X} with the calculated weighting \mathbf{K} (Figure 5). The weighted sum of the ensemble members and differences is given by

$$\mathbf{X}^a = \mathbf{X} + \mathbf{K}\mathbf{D} \quad (8)$$

where \mathbf{X}^a is our updated estimate.

Taking the average values of the updated members now provides a better estimate of our system than what we started with. After the update process, the uncertainty of the estimate is usually reduced. The final part of the EnKF is to move each individual ensemble member forward in time (Figure 6). To do this, we use a model propagator $(M_{t,t+1})$ that moves the estimate at time t to a new position at time $t + 1$.

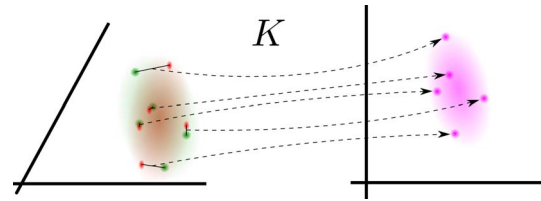


Figure 5. The pink circles are the weighted sum of the red and green. The pink shaded area represents the updated model error.

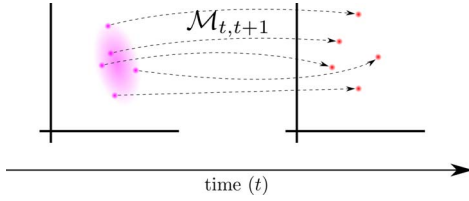


Figure 6. Our estimate at a given time (in pink) is propagated forward in time, and the process repeats.

$$\mathbf{X} = M_{t,t+1} \mathbf{X}^a \quad (9)$$

The estimate at the next time step is then ready for further measurements to be added. The process repeats exactly as previously mentioned.

Putting all of this together, we have the EnKF equations (where t is time and n the number of ensemble members)

$$\mathbf{X}_t^b = \mathbf{X}_t - \bar{x} \quad (10)$$

$$\mathbf{B}_t = \frac{1}{n-1} \mathbf{X}_t^b (\mathbf{X}_t^b)^T \quad (11)$$

$$\mathbf{K}_t = \mathbf{B}_t H_t^T (H_t \mathbf{B}_t H_t^T + \mathbf{R}_t)^{-1} \quad (12)$$

$$\mathbf{X}_t^a = \mathbf{X}_t + \mathbf{K}_t (\mathbf{Y}_t - H_t \mathbf{X}_t) \quad (13)$$

$$\mathbf{X}_{t+1}^b = M_{t,t+1} \mathbf{X}_t^a \quad (14)$$

where \mathbf{X}_t^b represents the average value removed from each column of \mathbf{X}_t (10). This process can be visualized

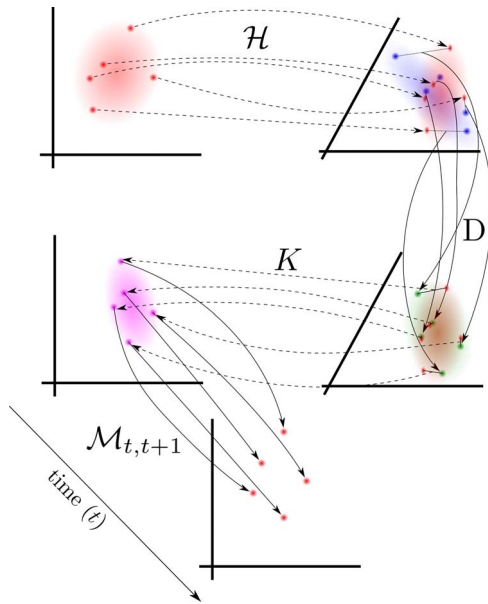


Figure 7. A graphical overview of the EnKF. The model values (red) are mapped to observation space and the observations (blue) are subtracted, giving a set of differences (green). A weighted sum between the model points and differences result in updated model points (pink) that are then propagated forward in time.

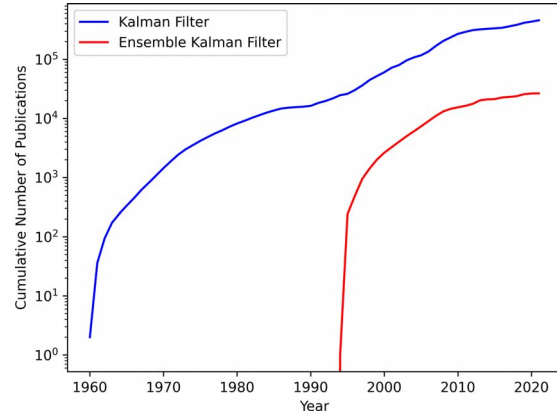


Figure 8. Cumulative number of papers, by year, that reference Kalman filter and EnKF, according to Google Scholar.

(by combining all the figures together, Figure 8) as mapping the model values (red) into observation space and subtracting the observations (blue) to obtain a set of differences (green). The uncertainties in the model and observations (estimated by the spread of points) are used to create a weighted sum between the model points and differences, resulting in updated model points (pink). Finally, these points are moved forward to the next time step, where the EnKF starts again (red).

3. Conclusions

The EnKF is a computationally efficient implementation of the Kalman filter. Its ease of use and effectiveness have resulted in its widespread adoption across all fields of radio science and beyond, as demonstrated in Figure 8, by the cumulative number of publications using the algorithms. The aim of this article has been to demystify the underlying mathematics of the EnKF through a series of simple visualizations.

4. References

1. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, **82**, 1, March 1960, pp. 35-45.
2. G. M. Siouris, G. Chen, and J. Wang, "Tracking an Incoming Ballistic Missile Using an Extended Interval Kalman Filter," *IEEE Transactions on Aerospace and Electronic Systems*, **33**, 1, January 1997, pp. 232-240.
3. M. Athans, "The Importance of Kalman Filtering Methods for Economic Systems," *Annals of Economic and Social Measurement*, **3**, 1, January 1974, pp. 49-64.
4. G. Evensen, "Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics," *Journal of Geophysical Research* **99**, C5, May 1994, pp. 10143-10162.