Stat Comput (2012) 22:1273–1276
DOI 10.1007/s11222-011-9300-x

The ensemble Kalman filter is an ABC algorithm

David J. Nott · Lucy Marshall · Tran Minh Ngoc

Received: 13 April 2011 / Accepted: 9 November 2011 / Published online: 23 November 2011
© Springer Science+Business Media, LLC 2011

Abstract The ensemble Kalman filter is the method of choice for many difficult high-dimensional filtering problems in meteorology, oceanography, hydrology and other fields. In this note we show that a common variant of the ensemble Kalman filter is an approximate Bayesian computation (ABC) algorithm. This is of interest for a number of reasons. First, the ensemble Kalman filter is an example of an ABC algorithm that predates the development of ABC algorithms. Second, the ensemble Kalman filter is used for very high-dimensional problems, whereas ABC methods are normally applied only in very low-dimensional problems. Third, recent state of the art extensions of the ensemble Kalman filter can also be understood within the ABC framework.

Keywords Approximate Bayesian computation · Data assimilation · Ensemble Kalman filter · Regression adjustment

1 Introduction

The ensemble Kalman filter (Evensen 1994, 2007) is perhaps the most widely used approach in very high-dimensional filtering problems in areas such as meteorology, oceanography and hydrology. In this note we show that the ensemble Kalman filter updates are an example of the regression adjustment ABC algorithm of Beaumont et al. (2002). This is interesting as the ensemble Kalman filter predates the development of ABC algorithms and is effective in very high dimensions, whereas ABC algorithms are used mostly in low-dimensional problems. Furthermore, a recent state of the art extension of the ensemble Kalman filter due to Lei and Bickel (2011) also corresponds to an ABC algorithm, namely the heteroscedastic nonlinear regression adjustment of Blum and François (2010). The regression perspective on the ensemble Kalman filter noted in Lei and Bickel (2011) and also discussed in the meteorological literature by Anderson (2003) is effectively the same as the method of Beaumont et al. (2002) applied to the filtering problem, although there has been no previous explicit link made with ABC methods. Lei and Bickel (2011) contains a number of other important innovations including some analysis of theoretical properties of their method and dimension reduction by localization, a very important issue in high-dimensional applications in the geosciences. There has been some recent interest in using ABC methods for filtering problems (Jasra et al. 2011) but there has been no study of regression adjustment methods in this context as far as we are aware. An anonymous referee has pointed out to us, however, that the convolution particle filter (Campillo and Rossi 2009) is in essence a regression approach to the filtering problem based on kernel conditional density estimation and simulation from a forward model. The purpose of this note is to make explicit the links between the method of Beaumont et al. (2002) and the ensemble Kalman filter, and between the method of Blum and François (2010) and that of Lei and Bickel (2011).

In Sect. 2 we briefly describe the regression adjustment ABC methods of Beaumont et al. (2002) and Blum and François (2010). Section 3 introduces the ensemble

D.J. Nott (✉) · T.M. Ngoc
Department of Statistics and Applied Probability,
National University of Singapore, Singapore 117546, Singapore
e-mail: standj@nus.edu.sg

L. Marshall
Department of Land Resources and Environmental Sciences,
Montana State University, 334 Leon Johnson Hall,
PO Box 173120, Bozeman, MT 59717-3120, USA

Kalman filter. Section 4 describes the connection between the method of Beaumont et al. (2002) and the ensemble Kalman filter updates. Section 5 discusses the method of Lei and Bickel (2011) and its relationship with Blum and François (2010) and Sect. 6 concludes.

2 Regression adjustment ABC algorithms

Approximate Bayesian computation (ABC) methods originated in statistical genetics with Tavaré et al. (1997), although the basic idea was also given by Rubin (1984) in the context of explaining the content of Bayes' theorem based on frequentist intuitions. Wilkinson (2008) shows how to interpret rejection type ABC algorithms as giving exact results when a model error term is taken into account. Here we will not be concerned with "local" rejection based ABC algorithms but only with regression adjustment ABC methods, which are described further below. For a review of state of the art ABC algorithms see Marin et al. (2011).

2.1 Method of Beaumont et al. (2002)

The regression adjustment method of Beaumont et al. (2002) is based on the following idea. Suppose we have a statistical model with parameter θ, that the data are denoted y and the likelihood is $p(y|\theta)$. We consider Bayesian inference with prior $p(\theta)$. We simulate $(\theta_i, y_i) \sim p(\theta)p(y|\theta)$, $i = 1, \ldots, n$ and then suppose that the regression model

$$\theta_i = \alpha + \beta^T(y_i - y) + \delta_i \tag{1}$$

holds, where the errors δ_i are independent and identically distributed. Estimating α and β to obtain $\hat{\alpha}$ and $\hat{\beta}$, the fitted regression model gives a density estimate of θ given any observed value y_i for the data. When the observed data is y, we obtain an estimate of $p(\theta|y)$ and the corresponding estimate of $E(\theta|y)$ is $\hat{\alpha}$. Adding the empirical residuals

$$\hat{\delta}_i = \theta_i - \hat{\alpha} - \hat{\beta}^T(y_i - y)$$

to this mean estimate gives an approximate sample from $p(\theta|y)$ if the model (1) holds. This gives a sample

$$\theta_i' = \hat{\alpha} + \hat{\delta}_i = \theta_i - \hat{\beta}^T(y_i - y).$$

In the method of Beaumont et al. (2002) model (1) is usually not regarded as globally valid, but rather kernel weights are assigned to the simulated (θ_i, y_i) in fitting a local linear model. However, in developing the connection with the ensemble Kalman filter we will not be concerned with this kind of weighting but instead assume that (1) is valid over the whole space.

2.2 Method of Blum and François (2010)

The nonlinear heteroscedastic regression method of Blum and François (2010) replaces (1) with

$$\theta_i = \mu(y_i) + \sigma(y_i)\delta_i \tag{2}$$

where $\mu(y)$ is a mean vector, $\sigma(y)$ is a diagonal matrix of variances and the δ_i are zero mean independent and identically distributed errors where the components all have variance one. If one is only interested in posterior marginals then a model such as (2) could be considered for each component separately, and it will also be clear in what follows that it would be possible in principle to take $\sigma(y)$ to be a square root of a covariance matrix where all elements are functions of y, although we do not consider this here. If (2) holds, then

$$\mu(y) + \sigma(y)\sigma(y_i)^{-1}\big(\theta_i - \mu(y_i)\big)$$

is a draw from $p(\theta|y)$. Blum and François (2010) suggest flexible approaches for estimating $\mu(y)$ and $\sigma(y)$, and there are also various other innovations in their work that relate to adaptive importance sampling refinements. While significant these other extensions don't relate to our present discussion.

3 The ensemble Kalman filter

Consider a linear Gaussian state space model (see, for instance, Harvey 1989)

$$y_t = H_t x_t + \epsilon_t, \quad \epsilon_t \sim N(0, \Omega_t) \tag{3}$$

$$x_t = S_t x_{t-1} + \eta_t, \quad \eta_t \sim N(0, Q_t), \tag{4}$$

$t = 1, \ldots, T$, where y_t is the observation vector at time t, x_t is the state vector, and ϵ_t and η_t are random disturbances. Here the matrices Ω_t, Q_t, H_t and S_t are assumed known. For simplicity assume also that the initial condition x_0 has a known Gaussian distribution. In the celebrated Kalman filter algorithm (Kalman 1960) the distribution for $x_t|y_1, \ldots, y_t$, $t = 1, \ldots, T$ is calculated recursively. With the Gaussian assumptions we have made the distribution of $x_t|y_1, \ldots, y_{t-1}$ (the forecast distribution) is normal, $N(\mu_{t|t-1}, \Sigma_{t|t-1})$ say, and hence the distribution of $x_t|y_1, \ldots, y_t$ is also normal, $N(\mu_{t|t}, \Sigma_{t|t})$ where

$$\mu_{t|t} = \mu_{t|t-1} + K_t(y_t - H_t\mu_{t|t-1}),$$

$$K_t = \Sigma_{t|t-1}H_t^T\big(H_t\Sigma_{t|t-1}H_t^T + \Omega_t\big)^{-1},$$

$$\Sigma_{t|t} = (I - K_tH_t)\Sigma_{t|t-1}.$$

K_t is called the Kalman gain matrix. Application of the Kalman filter in high-dimensional problems is difficult, due to the need to calculate and store large covariance matrices,

and nonlinear models in which the state equation (4) is replaced with

$$x_t = s(x_{t-1}, \eta_t) \tag{5}$$

are also problematic. The ensemble Kalman filter (Evensen 1994, 2007) attempts to address these issues. Nychka and Anderson (2010) is an accessible introduction to the ensemble Kalman filter for statisticians, indicating some of the practical issues in dealing with high-dimensional models which we ignore in the simple exposition we give here. In the ensemble Kalman filter we first generate a sample $(x_f^{(i)}, y_f^{(i)})$, $i = 1, \ldots, n$ from $\pi(x_t|y_{1:t-1})\pi(y_t|x_t)$, the distribution of (x_t, y_t) given $y_{1:t-1} = (y_1, \ldots, y_{t-1})$. This sample is called the forecast ensemble. If n samples from $\pi(x_{t-1}|y_{1:t-1})$ are already available then simulation of the forecast ensemble is performed by propagating each of the samples through (4) to obtain the values $x_f^{(i)}$ and then sampling $y_f^{(i)}$ from $\pi(y_t|x_t = x_f^{(i)})$. $\mu_{t|t-1}$ and $\Sigma_{t|t-1}$ empirically from the forecast ensemble to obtain $\hat{\mu}_{t|t-1}$ and $\hat{\Sigma}_{t|t-1}$ respectively. Computations involving $\hat{\Sigma}_{t|t-1}$ are tractable even with a very high-dimensional state if n is much less than the dimension of the state so that $\hat{\Sigma}_{t|t-1}$ has low rank. In particular, plugging in $\hat{\Sigma}_{t|t-1}$ to K_t to obtain $\hat{K}_t$, we can work with the subsequent $\hat{K}_t$ efficiently and storage requirements are also modest as in effect we only need to store the ensemble.

The next important idea in the ensemble Kalman filter is to adjust the forecast ensemble in order to make the sample covariance close to $\hat{\Sigma}_{t|t} = (I - \hat{K}_t H_t)\hat{\Sigma}_{t|t-1}$, the covariance of $x_t|y_1, \ldots, y_t$ obtained if we assume $x_t|y_1, \ldots, y_{t-1}$ is Gaussian. There are several ways to do this, but one common way that provides a connection with ABC methods is the following perturbed observations scheme. Simulate $\epsilon_t^{(i)} \sim N(0, \Omega_t)$, $i = 1, \ldots, n$ and then update $x_f^{(i)}$ to

$$x_a^{(i)} = x_f^{(i)} + \hat{K}_t\big(y_t - H_t x_f^{(i)} - \epsilon_t^{(i)}\big).$$

This so-called analysis ensemble at time t has the desired covariance structure as the ensemble size gets large. It should be noted that there are variants of the ensemble Kalman filter that do not rely on Monte Carlo methods for the covariance adjustment, but the connection with ABC methods relates to the Monte Carlo approach. Furrer and Bengtsson (2007) is a recent statistical discussion of some issues which arise in the Monte Carlo implementation.

4 Ensemble Kalman filter as ABC

We can now understand the connection between the ensemble Kalman filter and the regression adjustment of Beaumont et al. (2002). In the ensemble Kalman filter $x_f^{(i)}$, $i = 1, \ldots, n$ represents a sample from the "prior" for x_t (i.e. from $\pi(x_t|y_{1:t-1})$): the corresponding likelihood term comes from the observation equation at time t (i.e. from $\pi(y_t|x_t)$). Simulating data and parameters from the prior as in Beaumont et al. (2002) while using the forecast ensemble as the prior sample gives the corresponding simulated data

$$y_f^{(i)} = H_t x_f^{(i)} + \epsilon_t^{(i)}$$

where the $\epsilon_t^{(i)}$ are generated independently according to the noise in the measurement equation with covariance Ω_t. Here the sample $(x_f^{(i)}, y_f^{(i)})$ in the ensemble Kalman filter play the role as the sample (θ_i, y_i) in Sect. 2.1. Now considering the regression model (1) we have

$$x_f^{(i)} = \alpha + \beta^T\big(H_t x_f^{(i)} + \epsilon_t^{(i)} - y_t\big) + \delta_i$$

and fitting this multivariate linear model the estimated coefficient is $\hat{\beta} = \mathrm{Cov}(y_f)^{-1}\mathrm{Cov}(y_f, x_f)$ where $\mathrm{Cov}(y_f)$ and $\mathrm{Cov}(y_f, x_f)$ represent the sample covariance of the $y_f^{(i)}$ and the sample cross covariance for $(y_f^{(i)}, x_f^{(i)})$, $i = 1, \ldots, n$, respectively. This observation was also made by Lei and Bickel (2011). When the ensemble size is large $\mathrm{Cov}(y^f)$ tends to $H_t\Sigma_{t|t-1}H_t^T + \Omega_t$ and $\mathrm{Cov}(x_f, y_f)$ to $H_t\Sigma_{t|t-1}$ so that $\hat{\beta} \approx \hat{K}_t^T$. Hence

$$x_a^{(i)} = x_f^{(i)} + \hat{K}_t\big(y_t - H_t x_f^{(i)} - \epsilon_t^{(i)}\big),$$

$i = 1, \ldots, n$ represents an approximate sample from the posterior $x_t|y_1, \ldots, y_t$. That is, the ensemble KF update can be viewed as the regression adjustment of Beaumont et al. (2002).

5 The NLEAF algorithm

Recently Lei and Bickel (2011) have suggested a refinement of the ensemble Kalman filter, which they call the nonlinear ensemble adjustment (NLEAF) algorithm. If instead of (1) in the ensemble Kalman filter update we consider the model of Blum and François (2010) for this problem,

$$x_f^{(i)} = \mu\big(y_f^{(i)}\big) + \sigma\big(y_f^{(i)}\big)\delta_i$$

then this would suggest

$$x_a^{(i)} = \mu(y) + \sigma(y)\sigma(y_i)^{-1}\big(x_f^{(i)} - \mu\big(y_f^{(i)}\big)\big)$$

is a sample from $x_t|y_1, \ldots, y_t$. This is essentially the second order NLEAF algorithm of Lei and Bickel (2011) although there are differences in the way that Lei and Bickel (2011) and Blum and François (2010) would estimate $\mu(y)$ and $\sigma(y)$. In particular, Lei and Bickel (2011) consider both importance sampling approaches to estimating the mean and covariance matrix as well as regression based approaches. Lei and Bickel (2011) also consider theoretical aspects of the approximation, and the issue of dimension reduction by localization, which is important in the geosciences.

6 Discussion and conclusions

It is interesting that there is a connection between a state of the art method for high-dimensional filtering such as the ensemble Kalman filter and ABC methods which have typically been employed only in low dimensions. Some cross fertilization of ideas between these two fields may be helpful.

We believe that regression adjustment approaches may be currently underappreciated for working with a large number of parameters in the ABC context. In this regard, one extension of conventional ABC methods that we are currently pursuing involves the situation where the parameter θ is split into two parts—a low-dimensional parameter of primary interest η and a high-dimensional nuisance parameter of lesser interest ω. We can use ordinary "local" ABC methods to examine the marginal posterior distribution of η in the marginal model with ω integrated out, and then treat uncertainty in the high-dimensional ω more roughly by regression methods.

Suppose $\eta = (\eta_1, \ldots, \eta_{p_1})^T$ and $\omega = (\omega_1, \ldots, \omega_{p_2})^T$ where $p = p_1 + p_2$ is the dimension of θ. In the Bayesian framework inference about η is based on the marginal posterior distribution $p(\eta|y)$,

$$p(\eta|y) \propto p(\eta)p(y|\eta)$$

where the marginal likelihood $p(y|\eta)$ is

$$p(y|\eta) = \int p(y|\theta)p(\omega|\eta)d\omega.$$

Now suppose we have a set of statistics s_η that are marginally sufficient for η. If η is low-dimensional it should be possible to find a low-dimensional near sufficient statistic for η in the marginal model with ω integrated out and then ordinary ABC methods will work well to sample from $p(\eta|y)$. Write $\eta_i^{(1)}$, $i = 1, \ldots, n$ for the generated sample from $p(\eta|y)$ in this first stage. Next, in a second stage we use regression adjustment methods to generate samples from the conditional densities $p(\omega|\eta = \eta_i^{(1)}, y)$ to get samples from the joint posterior. Suppose that at the first stage we generated samples (θ_i, s_i), $i = 1, \ldots, n$ where the s_i are values of some sufficient statistic s for θ where $s = (s_\eta, s_\omega)^T$ (that is, s_η is a subset of s with s_ω providing additional information relating to the full parameter θ). Write $\theta_i = (\eta_i^T, \omega_i^T)^T$ and consider the regression model

$$\omega_i = \alpha + \beta^T(s_i - s) + \gamma^T\eta_i + \delta_i$$

where the errors δ_i are assumed independent and identically distributed. Using similar reasoning to before after estimating α, β and γ to obtain $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ we could obtain approximate samples from the conditional posterior by a mean estimate from the regression plus an empirical residual. This would give adjusted samples $\omega_i^{(2)}$ in our second stage for ω such as

$$\omega_i^{(2)} = \omega_i - \hat{\beta}^T(s_i - s) - \hat{\gamma}^T\left(\eta_i - \eta_i^{(1)}\right).$$

There are many possible variations on this basic idea, and the idea of decomposing the posterior sequentially is potentially very useful for applying regression adjustment ABC methods. In the filtering context if such an idea were applied recursively to estimation of the state this would represent a kind of fusion of the ABC filter of Jasra et al. (2011) and the ensemble Kalman filter. Work on these ideas is ongoing.

Acknowledgements David Nott gratefully acknowledges the support and contributions of the Singapore-Delft Water Alliance (SDWA). The research presented in this work was carried out as part of the SDWAs tropical reservoir research programme.

References

Anderson, J.L.: A local least squares framework for ensemble filtering. Mon. Weather Rev. **131**, 634–642 (2003)

Beaumont, M., Zhang, W., Balding, D.: Approximate Bayesian computation in population genetics. Genetics **162**, 2025–2035 (2002)

Blum, M., François, O.: Non-linear regression models for approximate Bayesian computation. Stat. Comput. **20**, 63–73 (2010)

Campillo, F., Rossi, V.: Convolution particle filter for parameter estimation in general state-space models. IEEE Trans. Aerosp. Electron. Syst. **45**, 1063–1072 (2009)

Evensen, G.: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. **99**(C5), 10143–10162 (1994)

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter. Springer, Berlin (2007)

Furrer, R., Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivar. Anal. **98**, 227–255 (2007)

Harvey, A.C.: Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge (1989)

Jasra, A., Singh, S.S., Martin, J., McCoy, E.: Filtering via approximate Bayesian computation. Stat. Comput. (2011). doi:10.1007/s11222-010-9185-0

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

Lei, J., Bickel, P.J.: Ensemble filtering for high dimensional non-linear state space models. Technical Report (2011). Available at http://www.stat.berkeley.edu/~bickel/LeiB09_NLEAF.pdf

Marin, J.-M., Pudlo, P., Robert, C.P., Ryder, R.: Approximate Bayesian computational methods. Technical Report (2011). Available at http://arxiv.org/abs/1101.0955

Nychka, D., Anderson, J.L.: Data assimilation. In: Gelfand, A.E., Diggle, P., Guttorp, P., Fuentes, M. (eds.) Handbook on Spatial Statistics, pp. 477–494. Chapman and Hall/CRC, London (2010)

Rubin, D.: Bayesianly justifiable and relevant frequency calculations for the applied statistician. Ann. Stat. **12**, 1151–1172 (1984)

Tavaré, S., Balding, D.J., Griffiths, R.C., Donnelly, P.: Inferring coalescence times from DNA sequence data. Genetics **145**, 505–518 (1997)

Wilkinson, R.D.: Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. Technical Report (2008). Available at http://arxiv.org/abs/0811.3355v1