



This is a repository copy of *Count time series prediction using particle filters*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/147642/>

Version: Accepted Version

---

**Article:**

Triantafyllopoulos, K. [orcid.org/0000-0002-4144-4092](https://orcid.org/0000-0002-4144-4092), Shakandli, M. and Campbell, M. (2019) Count time series prediction using particle filters. *Quality and Reliability Engineering International*. ISSN 0748-8017

<https://doi.org/10.1002/qre.2534>

---

This is the peer reviewed version of the following article: Triantafyllopoulos, K, Shakandli, M, Campbell, M. Count time series prediction using particle filters. *Qual Reliab Engng Int*. 2019; 1– 15., which has been published in final form at <https://doi.org/10.1002/qre.2534>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## ARTICLE TYPE

# Count time series prediction using particle filters

K. Triantafyllopoulos\*<sup>1</sup> | M. Shakandli<sup>1</sup> | M. J. Campbell<sup>2</sup>

<sup>1</sup>School of Mathematics and Statistics,  
University of Sheffield, Sheffield, UK

<sup>2</sup>School of Health and Related Research,  
University of Sheffield, Sheffield, UK

**Correspondence**

\*K. Triantafyllopoulos, School of  
Mathematics and Statistics, Hicks Building,  
University of Sheffield, Sheffield S3 7RH,  
UK. Email: kostas@sheffield.ac.uk

**Summary**

Non-Gaussian dynamic models are proposed to analyse time series of counts. Three models are proposed for responses generated by a Poisson, a negative binomial and a mixture of Poisson distributions. The parameters of these distributions are allowed to vary dynamically according to state space models. Particle filters or sequential Monte Carlo methods are used for inference and forecasting purposes. The performance of the proposed methodology is evaluated by two simulation studies for the Poisson and the negative binomial models. The methodology is illustrated by considering data consisting of medical contacts of schoolchildren suffering from asthma in England.

**KEYWORDS:**

Count time series, medical statistics, particle filter, non-normal time series, dynamic generalised linear model, Poisson distribution

## 1 | INTRODUCTION

Time series of counts are observed routinely in many disciplines, such as medicine and healthcare, sociology and economics, quality control and engineering. Observations of counts recorded over time offer an opportunity to study the dynamics of counts with the view of forecasting. Discrete-valued time series are studied in the literature as part of dynamic generalised linear modelling (DGLM), see e.g. Harvey and Fernandes<sup>1</sup>, West et al.<sup>2</sup>, Fahrmeir<sup>3</sup>, Kitagawa and Gersch<sup>4</sup> and others. For book-length discussions of DGLM, the reader is referred to West and Harrison<sup>5</sup>, Fahrmeir and Tutz<sup>6</sup> and to Kedem and Fokianos<sup>7</sup>; a review of DGLM is given in Triantafyllopoulos<sup>8</sup>.

In this paper we put forward three dynamic models suitable for count time series estimation and forecasting. Our motivation stems from the observation that count time series is somewhat less frequently discussed in medical statistics, although counts are routinely observed in medicine and healthcare (Gasparrini et al.<sup>9</sup> and He et al.<sup>10</sup>). A data set consisting of medical contacts of schoolchildren in England suffering from asthma serves as a motivating example; this data set is discussed in Julious et al.<sup>11</sup>, but the data was not analysed from a time-series perspective. Our aim in this paper is to set up suitable models for count time series and to discuss suitable estimation procedures. In particular, we propose three dynamic models: the first adopts a Poisson distribution for the observed counts, the second adopts a negative binomial distribution for the counts and the third adopts a finite mixture of Poisson distributions for the counts. For all three models, the dynamics of the models are driven by simple Gaussian state space models (Shumway and Stoffer<sup>12</sup>). Following modern Bayesian inference, we propose the use and utility of particle filtering, also known as sequential Monte Carlo; for a review of particle filtering see Doucet et al.<sup>13</sup> and Künsch<sup>14</sup>. We describe the Liu and West<sup>15</sup> (LW) and the Storvik<sup>16</sup> (STK) filters, which are able to estimate dynamic model components as well as static model components. In this paper we focus on these two parameter estimation methods, because they are general, require small computational cost and therefore are appealing to practitioners. More recent methodologies tackling this problem exist, but tend to be problem-specific or computationally expensive, see e.g. Andrieu *et al.*<sup>17</sup>, Carvalho *et al.*<sup>18</sup> and Nemeth *et al.*<sup>19</sup>.

Two simulation studies are used to illustrate the goodness of fit of the proposed filters, for the Poisson and the negative binomial models. The results suggest both Poisson and negative binomial models fit well. A comparison of the Poisson model with a standard Markov chain Monte Carlo (MCMC) approach provides further insight in the quality of the particle filter performance. Our findings also suggest that the Liu and West filter outperforms the Storvik filter for the time series considered in this paper. The models are fitted to the asthma data set mentioned above. The findings suggest that both dynamic models have good performance. Although the asthma data set is used for illustration, the methodology which is put forward in this paper is general and can be applied for a wide range of count time series data, including stationary and non-stationary and be aimed for real-time application.

The rest of the paper is organised as follows. Section 2.1 introduces the three models and Section 2.2 discusses the LW and STK algorithms. The simulation studies are placed in Section 3 and the asthma data are analysed in Section 4. The paper concludes with closing comments in Section 5.

## 2 | DYNAMIC MODELS FOR COUNT TIME SERIES

### 2.1 | Dynamic Poisson, negative binomial and Poisson mixture models

In this section we discuss three model specifications suitable for analysing count time series. The first model is referred to as dynamic Poisson model (or simply here as the Poisson model), the second model is referred to as dynamic negative binomial model (or negative binomial model) and the third model is referred to as two-component dynamic mixture Poisson model. The first two of these models consider Poisson or negative binomial distributions for the response distribution, while model parameters or states are allowed to follow a dynamic model, also known as a state space model. The third one considers a two-component Poisson mixture as a response distribution and each component follows individually a dynamic model.

Consider a time series of counts  $\{y_t\}_{t=1,\dots,T}$  so that, the observation model is

$$y_t | x_t \sim \text{Pois}[\exp(x_t)], \quad (1)$$

where  $\text{Pois}[\lambda_t]$  indicates a Poisson distribution with rate  $\lambda_t = \exp(x_t)$ . Here for each count  $y_t$ , the Poisson distribution has a different rate  $\lambda_t$ , with dynamics driven by the evolution of the state  $x_t$  according to

$$x_t - \mu = \phi(x_{t-1} - \mu) + \omega_t, \quad (2)$$

where  $\mu$  and  $\phi$  are time-invariant parameters (see below) and  $\omega_t$  is a white noise sequence, with  $\omega_t \sim N(0, W)$ , for some variance  $W$ . It is also assumed that  $\omega_t$  is independent of  $x_0$ , which is assumed to follow a Gaussian distribution with some mean  $m_0$  and some variance  $C_0$ .

The parameters  $\mu$  and  $\phi$  enables the model to cater for either a stationary model or for a non-stationary one. For the former,  $\mu$  is the constant mean of  $x_t$ , hence  $x_t - \mu$  is a zero mean process and, with the additional assumption  $-1 < \phi < 1$ , the process  $\{x_t\}$  is an autoregressive (AR) process with AR coefficient  $\phi$ . In this case it is easy to see that stationarity of  $\{x_t\}$ , implies stationarity of  $\{y_t\}$ . For the non-stationary model  $\mu$  is redundant (as now  $x_t$  does not have a constant mean) and  $\phi$  should satisfy  $|\phi| \geq 1$ ; the most typical value of  $\phi$  is  $\phi = 1$ . This set up corresponds to a random walk process for the state evolution. Other non-stationary models for  $x_t$  may be considered, for example allowing a trend or including time-varying covariates. Hence, one may consider a general linear state space model

$$x_t = \mathbf{F}_t^\top \mathbf{x}_t \quad \text{and} \quad \mathbf{x}_t = \mathbf{G}_t \mathbf{x}_{t-1} + \boldsymbol{\omega}_t,$$

where  $\mathbf{F}_t$  is a design column vector,  $\mathbf{x}_t$  is a vector of states,  $\mathbf{G}_t$  an evolution matrix and  $\top$  denotes transposition; more details for such models are described in West and Harrison<sup>5</sup> and in references therein. In this paper we focus on the simpler evolution (2), which is used in the subsequent sections.

Moving on to the negative binomial model, the observation model (1) is replaced by

$$y_t | \tau, x_t \sim \text{NegBin}[\tau, \pi_t] \equiv \text{NegBin} \left[ \tau, \frac{\exp(x_t)}{1 + \exp(x_t)} \right], \quad (3)$$

where  $\tau$  is the dispersion parameter ( $y_t$  is the number of successes until  $\tau$  failures) and  $\pi_t$  is the probability of success. In the above representation,  $\tau$  is considered time-invariant, while  $\pi_t$  is allowed to vary over time, through  $x_t$ . The reason we consider  $\tau$  not to change over time is that the shape of the distribution changes quite dramatically with a change of  $\tau$ , which is not reasonable to assume over time. In the above model we have used the standard logistic link, i.e.  $\log \pi_t(1 - \pi_t)^{-1} = x_t$ . The negative binomial

model is complete by adopting (2) for the evolution of  $x_t$ . Similar comments made about  $x_t$  for the Poisson model apply for the negative binomial model.

Next we describe the Poisson mixture dynamic model. Suppose  $\{y_{1t}\}_{t=1,\dots,T}$  and  $\{y_{2t}\}_{t=1,\dots,T}$  denote two different time series for count data where each follows a Poisson distribution with parameters  $\lambda_{1t}, \lambda_{2t}$ . A two-component dynamic mixture Poisson model is considered through mixing two single dynamic Poisson models. In particular, the observation and the state processes of the two-component dynamic mixture Poisson model can be represented as follows:

- Observation model:

$$p(y_t | \lambda_{1t}, \lambda_{2t}) = \vartheta p(y_{1t} | \lambda_{1t}) + (1 - \vartheta)p(y_{2t} | \lambda_{2t}) = \vartheta \text{Pois}(\lambda_{1t}) + (1 - \vartheta) \text{Pois}(\lambda_{2t}). \quad (4)$$

- State models (two components):

$$\begin{aligned} x_{1t} - \mu_1 &= \phi_1(x_{1,t-1} - \mu_1) + \omega_{1t}, & \omega_{1t} &\sim N(0, W_1), \\ x_{2t} - \mu_2 &= \phi_2(x_{2,t-1} - \mu_2) + \omega_{2t}, & \omega_{2t} &\sim N(0, W_2), \end{aligned}$$

where  $\phi_1, \phi_2$  are defined as first-order autoregressive coefficients for each state model,  $\omega_{1t}, \omega_{2t}$  are evolution errors of the two state models, which follow a normal distribution with mean zero and variances  $W_1, W_2$  respectively, and  $\vartheta$  is a mixing parameter which represents the probability of an observation belonging to the first distribution (component). Furthermore,  $\{\omega_{it}\}$  is assumed to be white noise ( $i = 1, 2$ ) and  $\omega_{1t}$  is independent of  $\omega_{2s}$ , for any  $t, s = 1, \dots, T$ .

As one referee has suggested, the relationship of the above three distributions (Poisson, negative binomial and Poisson mixture) is interesting. On one hand, clearly, the Poisson model is a special case of the mixture model, if we set  $\vartheta = 1$ . Moreover, the mixture model can accommodate a situation whereby  $\vartheta \approx 1$ . Hence, estimating  $\vartheta$  might provide information of validation for the Poisson model. On the other hand, the negative binomial distribution is obtained as a mixture of a Poisson distribution when the rate  $\lambda$  follows a gamma distribution. For example, the probability mass function of the negative binomial distribution is obtained from the gamma mixture

$$p_{NB}(y) = \int_0^{\infty} p_{Pois}(y; \lambda) p_{Ga}(\lambda; \alpha, \beta) d\lambda = \int_0^{\infty} \exp(-\lambda) \frac{\lambda^y}{y!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda) d\lambda = \frac{\Gamma(x + \alpha)}{y! \Gamma(\alpha)} \pi^\alpha (1 - \pi)^y,$$

where  $p_{Pois}(y; \lambda)$  denotes the probability mass function of the Poisson distribution with rate  $\lambda$ ,  $p_{Ga}(\lambda; \alpha, \beta)$  denotes the density function of the gamma distribution with parameters  $\alpha, \beta$  and  $\Gamma(\cdot)$  denotes the gamma function. Hence one can regard the negative binomial model as a Poisson model when  $\lambda \sim \text{Ga}(\tau, \pi(1 - \pi)^{-1})$ . It also follows that for large  $\tau$  the Poisson distribution is obtained by the negative binomial, as in this case the mean and the variance of the gamma distribution are approximately the same. Hence, if  $\tau$  is estimated this can provide information as to how suitable a Poisson model might be. For more details on the many Poisson mixture models the reader is referred to Karlis and Xekalaki<sup>20</sup> and to references therein.

## 2.2 | Particle filters based inference

Adopting Bayesian inference we aim to obtain sequentially the posterior distribution  $p(x_t | D_t)$ , where  $D_t = \{y_1, \dots, y_t\}$  includes the observed data up to time  $t = 1, \dots, T$ . This posterior distribution provides a summary of the available information at time  $t$  and can be used to extract the one-step ahead forecast distribution  $p(y_{t+1} | D_t)$  sequentially over time.

It is well known that  $p(x_t | D_t)$  is not analytically available for the models of Section 2.1, hence we need to resort to approximations or simulations. Markov chain Monte Carlo (MCMC) methods are available, see e.g. Gamerman<sup>21</sup>, but the posterior  $p(x_t | D_t)$  is approximated based on a single path of data; this time-consuming MCMC procedure prevents the use of computationally efficient sequential application. For sequential application and in particular with the view of forecasting, particle filters provide a natural alternative. Particle filters make use of sequential importance sampling and are briefly described below.

Usually it is not possible to sample from  $p(x_t | D_t)$ . The basic particle filter algorithm suggests at each time  $t$  to sample  $N$  particles  $x_t^{(1)}, \dots, x_t^{(N)}$  from a proposal distribution, known also as *importance function*,  $g(x_t | x_{t-1}, y_t)$ . Based on these sampled values the importance weights defined

$$w_t = \frac{p(x_1, \dots, x_t | y_1, \dots, y_t)}{g(x_1, \dots, x_t | y_1, \dots, y_t)}.$$

are computed for each particle  $i$  sequentially by

$$w_t^{(i)} = \frac{p(y_t | x_t^{(i)})p(x_t^{(i)} | x_{t-1}^{(i)})}{g(x_t^{(i)} | x_{t-1}^{(i-1)}, y_t)} w_{t-1}^{(i)}$$

These weights are standardised to sum to one, i.e.  $\tilde{w}_t^{(i)} = w_t^{(i)} \left( \sum_{i=1}^N w_t^{(i)} \right)^{-1}$ . Finally, the posterior distribution  $p(x_t | D_t)$  is approximated by

$$\hat{p}(x_t | D_t) = \sum_{i=1}^N \tilde{w}_t^{(i)} \delta(x_t - \hat{x}_t),$$

where  $\hat{x}_t = \sum_{i=1}^N \tilde{w}_t^{(i)} x_t^{(i)}$  is a weighted posterior estimate of  $x_t$  and  $\delta(\cdot)$  is the Dirac function.

A deficiency of the above procedure is that the weights have the tendency to be zero and very few to be large. This spoils the power of the Monte Carlo estimation and is overcome by a resampling step. The effective sampling size, defined as  $N_{eff} = \left( \sum_{i=1}^N \tilde{w}_t^{(i)2} \right)^{-1}$  and if  $N_{eff} < N/3$  the resampling step is warranted. The most popular is the multinomial resampling, according to which we resample  $x_t^{(i)}$  from the discrete distribution  $P(x_t = x_t^{(i)}) = \tilde{w}_t^{(i)}$  and then to update the weights by  $\tilde{w}_t^{(i)} = 1/N$ . The above procedure is applied sequentially over time for  $t = 1, \dots, T$ .

The basic particle filter as discussed above, provides an approximation of the posterior distribution  $p(x_t | D_t)$ . However, many models include time-invariant parameters subject to estimation; in the Poisson model such parameters are  $\mu$  and  $\phi$  and in the negative binomial model there is also the parameter  $\tau$ . In addition to those, the variance of  $\omega_t$  in the evolution equation 2 might be time-invariant. These parameters cannot be updated using the basic particle filter, because they do not change from one time to another. Indeed let  $\theta$  denote the vector comprising the static parameters subject to estimation and write  $\mathbf{x}_t^* = (x_t, \theta^\top)^\top$ , where  $\top$  denotes transposition. Then, sampling the ‘‘extended’’ state vector  $\mathbf{x}_t^{*(i)}$  from the importance function  $g(\mathbf{x}_t^* | \mathbf{x}_{t-1}^{*(i-1)}, y_t)$  updates only  $x_t^{(i)}$ , while  $\theta$  is already sampled at  $t - 1$ . After backtracking, it follows that  $\theta$  is sampled once at  $i = 0$  (from the prior of  $\theta$ ) and cannot be updated for any  $t \geq 1$ .

For the Poisson model  $\theta = (\mu, \phi, W)^\top$  and for the negative binomial model  $\theta = (\tau, \mu, \phi, W)^\top$ . To tackle the problem of estimation of  $\theta$ , Gordon et al.<sup>22</sup> have proposed incorporating an artificial state to the states,  $\theta_t = \theta_{t-1} + \nu_t$ , as if  $\theta$  were following a random walk, where  $\nu_t = (\nu_{1,t}, \dots, \nu_{k,t})^\top$  is a  $k$ -column vector whose  $i$ -th element  $\nu_{i,t}$  is a zero-mean random variable and  $\nu_{i,t}$  is independent of  $\nu_{i',t'}$ , for  $i \neq i'$  and  $t \neq t'$ . In this specification given  $D_{t-1}$ , the expectation of  $\theta_t$  is the same as the expectation of  $\theta_{t-1}$  but the variance / covariance matrix of  $\theta_t$  is increased compared to that of  $\theta_{t-1}$ . This approach has the disadvantage of introducing more uncertainty, which may not be realistic. For example in the negative binomial model it is difficult to motivate a random walk evolution for  $\tau$ .

Two variants of the basic particle filter are able to deal with this problem successfully and are described briefly below.

- Liu and West algorithm (LW). Liu and West<sup>15</sup> proposed an algorithm according to which  $\theta$  follows a Gaussian mixture and is estimated sequentially as follows. At each time  $t$ , conditionally on  $\theta$  the basic particle filter algorithm applies as described above. Suppose that at time  $t - 1$  we have sampled  $x_{t-1}^{(i)}$ , given sampled values of  $\theta$  at time  $t - 2$ . For  $\theta$  a Gaussian mixture is set, so that the posterior  $p(\theta | D_{t-1})$  is approximated by

$$\hat{p}(\theta | D_{t-1}) = \sum_{i=1}^N \tilde{w}_{t-1}^{(i)} f_N(\theta; c\theta_{t-1}^{(i)} + (1-c)\bar{\theta}_{t-1}, h^2 \mathbf{V}_{t-1}),$$

where  $f_N(\theta; \alpha, \mathbf{B})$  denotes the density of the multivariate normal distribution with mean vector  $\alpha$  and covariance matrix  $\mathbf{B}$ ,  $c \in (0, 1)$  is a shrinkage parameter,  $h^2 = 1 - c^2$  and  $\bar{\theta}_{t-1} = \sum_{i=1}^N \tilde{w}_{t-1}^{(i)} \theta_{t-1}^{(i)}$ . The choice of  $c$  and  $\bar{\theta}_{t-1}$  are such that  $E(\theta | D_{t-1}) = \bar{\theta}_{t-1}$  and  $\text{Var}(\theta | D_{t-1}) = \mathbf{V}_{t-1}$ , which are calculated at time  $t - 1$ . The value of  $h$  (or alternatively that of  $c$ ) is specified using a discount factor  $\beta$

$$h^2 = 1 - \left( \frac{3\beta - 1}{2\beta} \right)^2,$$

as proposed in Liu and West<sup>15</sup>. The idea is borrowed from the relationship of discounting in dynamic models and exponential smoothing. In particular,  $\beta = 1$  corresponds to a static evolution in dynamic models and this corresponds to  $c = 1$ , hence the mean of  $\theta$  does not depend on  $\bar{\theta}_{t-1}$  in the mixture. If on the other hand  $\beta < 1$ , then  $c < 1$  and the mean of  $\theta$  will be a weighted average of  $\theta_{t-1}^{(i)}$  and  $\bar{\theta}_{t-1}$ . Generally, the range of the discount factor  $\beta$  is  $(0.95 - 0.99)$ , which results in a range  $(0.974 - 0.995)$  for  $c$ ; for more discussion on discount factors the reader is referred to West and Harrison<sup>5</sup>. For the filter to run, a prior on each component of  $\theta$  is assumed. If the support of the prior distribution of each element of  $\theta$  is not

the real line, transformations may be used (Liu and West<sup>15</sup>). Hence here the variance  $W$  can be transformed to  $\log(W)$  and  $\tau$  to  $\log \tau$ . Alternatively, we can choose suitable priors (for  $W$  and  $\tau$  can be a gamma prior) which offers some more flexibility and is more natural to adopt from a Bayesian standpoint. In Sections 3 and 4 we discuss the chosen priors.

Moving on to time  $t$  the joint posterior distribution of  $x_t$  and  $\boldsymbol{\theta}$  is approximated by

$$\hat{p}(x_t, \boldsymbol{\theta} \mid D_t) \propto \sum_{i=1}^N \tilde{w}_{t-1}^{(i)} p(y_t \mid x_t, \boldsymbol{\theta}) p(x_t \mid x_{t-1}^{(i)}, \boldsymbol{\theta}) f_N(\boldsymbol{\theta}; c\boldsymbol{\theta}_{t-1}^{(i)} + (1-c)\bar{\boldsymbol{\theta}}_{t-1}, h^2\mathbf{V}_{t-1}) \delta(x_{t-1} - x_{t-1}^{(i)}).$$

Sampling  $x_t^{(i)}$  can be facilitated by adopting auxiliary variables as explained in the so called auxiliary particle filter of Pitt and Shephard<sup>23</sup>. For the application of this algorithm to the Poisson mixture model the issue of label switching should be dealt with. This issue reflects the invariance of the mixture likelihood function under relabelling of the components (here there are only two components). Failure to deal with label switching results in chains (in MCMC) or samples in particle filters (PF) that switch from one label to another. It turns out this is more of a problem for parameter estimation and even more in model comparison, rather than in forecasting, because in forecasting we simulate from the response distribution conditional on the mode of the estimated parameters. Some key references in the issue of label switching include Stephens<sup>24</sup>, Frühwirth-Schnatter<sup>25</sup>, Jasra *et al.*<sup>26</sup>; a extended coverage of the topic is discussion in Chapter 3 of Frühwirth-Schnatter<sup>27</sup>. In this paper we deal with this issue by adopting the random permutation sampling scheme proposed in Frühwirth-Schnatter<sup>25</sup> for MCMC, suitably adjusted for the needs of this paper; essentially we apply the algorithm off-line after we have obtained the PF samples. Examples of two-component mixtures, such as those considered here, are given in Chapter 3 of Frühwirth-Schnatter<sup>27</sup>, to which the reader is referred for further reading.

- Storvik filter (STK). Storvik<sup>16</sup> proposes an alternative particle-based technique to determine the hyper-parameters and the latent state variables. In this section we describe the Storvik filter because of its simplicity in application and its computational savings as opposed to other more complex filters (see also the introduction). However, it is somehow limited in its application and in this article we demonstrate that for the data we discuss its performance is not as good as the LW filter described above.

The concept of the Storvik particle filter algorithm is to use low dimensional sufficient statistics, denoted by  $s$ , rather than kernel smoothing densities as described above. Storvik<sup>16</sup> employs these sufficient statistics to simulate sequentially the unknown static parameters and the state variables. He utilises the fact that the vector of the sufficient statistics can be updated sequentially when a new state,  $x_t$ , is generated from its posterior density and a new observation,  $y_t$ , is created via mapping  $s_t = S(s_{t-1}, x_t, y_t)$ . Therefore, for the static parameters,  $\boldsymbol{\theta}$ , is also tracked by the sufficient statistics in a sequential manner. For the Gaussian system process, in particular, the particle set of the sufficient statistics at time instant  $t - 1$  can be updated sequentially to a new particle set at time point  $t$  by using a class of equations of the Kalman filter as detailed in Storvik<sup>16</sup>. Thus, the marginal posterior distribution of the static parameter  $p(\boldsymbol{\theta} \mid x_t, D_t)$  can be defined conditionally on sufficient statistics in the form  $p(\boldsymbol{\theta} \mid s_t)$ .

For the evolution (2) we can write

$$x_t = [1, x_{t-1}] \begin{bmatrix} \mu(1 - \phi) \\ \phi \end{bmatrix} + \omega_t = \mathbf{X}_t^\top \boldsymbol{\Psi} + \omega_t,$$

where  $\mathbf{X}_t^\top = [1, x_{t-1}]$ ,  $\omega_t \sim N(0, W)$  and  $\boldsymbol{\Psi}^\top = [\mu(1 - \phi), \phi]$  is a vector of static parameters. This model is a regression model, assuming we observe  $x_t$ . Assume that at time  $t - 1$ , we have sampled  $x_{t-1}^{(i)}$ . Thus, we can use the Kalman filter, in order to obtain sequentially the posterior distribution of  $\boldsymbol{\Psi} \mid x_{t-1}^{(i)}$  and  $W \mid x_{t-1}^{(i)}$ . These are provided by standard updating equations of the Kalman filter adjusted to cater for Bayesian estimation of the variance  $W$  (here for simplicity purposes we have provided them conditional on  $x_t$ ):

$$\boldsymbol{\Psi} \mid x_t, x_{t-1} \sim N(\mathbf{m}_t, W\mathbf{C}_t) \quad \text{and} \quad W^{-1} \mid x_t, x_{t-1} \sim \text{Ga}(n_t/2, d_t/2),$$

**TABLE 1** Summary of the performance of the Liu and West (LW) and Storvik (STK) algorithms with Poisson model. Shown are Monte Carlo mean and standard deviation of the AR coefficient  $\phi$ , state variance  $W$  and state mean  $\mu$ ; shown are also Monte Carlo means of the root mean square error and mean standard deviation of the forecast errors.

	LW	STK
Mean/SD ( $\phi = 0.75$ )	0.6293/0.0933	0.5638/0.1868
Mean/SD ( $W = 1.35$ )	1.229/0.1174	4.121/0.0998
Mean/SD ( $\mu = 0.85$ )	0.7487/0.1406	0.6271/0.1312
Mean RMSE Pred	1.4913	2.988
Mean SD Pred	1.373	2.8951

with recursions

$$\begin{aligned}
 q_t &= \mathbf{X}_t^\top \mathbf{C}_{t-1} \mathbf{X}_t + 1, \\
 \mathbf{m}_t &= \mathbf{m}_{t-1} + \mathbf{C}_{t-1} \mathbf{X}_t (x_t - \mathbf{X}_t^\top \mathbf{m}_{t-1}) / q_t, \\
 \mathbf{C}_t &= \mathbf{C}_{t-1} - \mathbf{C}_{t-1} \mathbf{X}_t \mathbf{X}_t^\top \mathbf{C}_{t-1} / q_t, \\
 d_t &= d_{t-1} + (x_t - \mathbf{X}_t^\top \mathbf{m}_{t-1})^2 / q_t, \\
 n_t &= n_{t-1} + 1,
 \end{aligned}$$

where  $\text{Ga}(n_t/2, d_t/2)$  denotes a gamma distribution with shape parameter  $n_t/2$  and scale parameter  $d_t/2$ . A point estimate of  $W$  may be obtained as the posterior mode of  $W \mid x_t, x_{t-1}$ , which is equal to  $d_t(n_t + 2)^{-1}$ . Hence we can sample  $\boldsymbol{\theta}^\top = [\boldsymbol{\psi}^\top, W]$  as  $\boldsymbol{\theta}^{(i)}$  and then use a standard particle filter to sample  $x_t^{(i)}$  from  $p(x_t \mid \boldsymbol{\theta}^{(i)}, D_t)$ . If we are not interested in sampling  $W$ , we can integrate out, hence sampling  $x_t$  from a Student  $t$  distribution with  $n_t$  degrees of freedom; details appear in West and Harrison<sup>5</sup>. The above recursions assume we start with some initial values of  $m_0, C_0, n_0, d_0$ , which follow standard Bayesian proposals and are reviewed in detail in West and Harrison<sup>5</sup>. Unfortunately, the above approach of the Storvik filter does not allow estimation of the dispersion parameter in the negative binomial model. In the simulations in Section 3 we have used the true value for  $\tau$  for the application of the Storvik filter. In real-data applications,  $\tau$  must be determined outside the algorithm, perhaps by looking at the shape of the empirical density of data or by employing a Metropolis step within the particle filter, see e.g. similar algorithms are proposed in Andrieu *et al.*<sup>17</sup>.

### 3 | SIMULATION STUDIES

In this section two simulated studies are conducted, with the aim to make a comparison between the Liu and West (LW) and the Storvik (STK) algorithms. Considering the dynamic Poisson model (1) - (2), data are generated by using true parameters  $\phi = 0.75$ ,  $W = 1.35$  and  $\mu = 0.85$ . For the two algorithms discussed above, the Liu and West and Storvik particle filters, we set the same priors for the hyperparameters. The prior distributions are represented as follows:  $W^{-1} \sim \text{Ga}(n_0/2, d_0/2)$ ,  $\boldsymbol{\psi} \sim N(\mathbf{m}_0, W\mathbf{C}_0)$ . We remark that in later sections of the paper we use a gamma prior for  $W$  for LW; here we have used the inverse gamma prior for  $W$  in order to have the same prior as in the Storvik filter. The sufficient statistics are setup as follows:  $n_0 = 2$ ,  $d_0 = 1$ ,  $\mathbf{m}_0 = [0, 0]^\top$ ,  $\mathbf{C}_0 = 10\mathbf{I}_2$ , where  $\mathbf{I}_2$  denotes the  $2 \times 2$  identity matrix. In addition, the prior distribution of the initial value of the state in both algorithms is  $x_0 \sim N(0, 4)$ . For the use of the Liu and West algorithm, the value of the shrinkage parameter is determined as  $c = 0.975$ ; this is chosen with the aid of the discount factor  $\beta$ , using the formula  $c = 2^{-1}\beta^{-1}(3\beta - 1)$  (see Section 2.2 above). Typically values of  $\beta$  close to 1 should be picked, here  $c = 0.975$  corresponds to  $\beta = 0.952381$ .

A Monte Carlo study with 100 runs is performed to obtain a new set of 100 observations at each run. The proposed algorithms are demonstrated by the simulated data from a Poisson distribution in order to study the algorithm performance in tracking the latent variable (state). Also observed is how accurately the parameters are estimated. A multinomial resampling step is used if needed for each particle of the state and hyperparameters at each time point of time. A sample of 1000 particles ( $N = 1000$ ) is used at each time point. The square root of the mean square error (RMSE) is used to judge the forecast performance of the proposed algorithms.

Table 1 provides a summary comparison of the performance accuracy of the Liu and West particle filter with a Poisson model (LW) and the Storvik particle filter with a Poisson model (STK). According to the estimation of the hyperparameters, we can

see the Monte Carlo average of the posterior mean for the autoregressive coefficient by the LW algorithm ( $\hat{\phi} = 0.6293$ ) is closer to the true value than that of the STK algorithm ( $\hat{\phi} = 0.5638$ ). In addition, the LW algorithm provides a considerably better estimate of the state variance ( $\hat{W} = 1.229$ ) compared to the STK algorithm ( $\hat{W} = 4.121$ ). In terms of one-step-ahead forecasting, the Monte Carlo average of the RMSE of prediction by the LW algorithm (RMSE = 1.4913) is closer to one than that of the STK algorithm (RMSE=2.988); in fact RMSE of the Storvik filter is more than twice that of LW, both being larger than 1. In addition, the LW algorithm provides smaller Monte Carlo standard deviation of the 1-step ahead prediction errors (1.373) than the STK algorithm (2.8951); here for each Monte Carlo run time series the 1-step ahead prediction error is calculated and the above figures refer to the standard deviation over all Monte Carlo runs. Therefore, based on these results we conclude that the Liu and West particle filter outperforms the Storvik filter as far as forecast accuracy is concerned. Considering the negative binomial model similar results are confirmed for the LW algorithm, not shown here. The Storvik filter cannot estimate the dispersion parameter  $\tau$  of the negative binomial model, but conditional on  $\tau$ , it is possible to apply this filter. Because of the performance of the LW algorithm, for the rest of the paper we choose focus on the implementation and performance of this algorithm, but we do provide some comparative results with the STK algorithm.

The above analysis is based on Pearson residuals. Residuals based on the  $P$ -score statistics may also be used to judge the goodness of fit of non-Gaussian dynamic models, see Smith<sup>28</sup>, Frühwirth-Schnatter<sup>29</sup> and Czado et al.<sup>30</sup>. In this approach the predictive cumulative distribution is formed

$$U_t = \Pr(Y_t \leq y_t \mid D_{t-1}),$$

for any  $t = 1, \dots, T$ . For continuous-valued time series  $U_t$  follows a Uniform distribution and hence the test is to see whether  $U_1, \dots, U_T$  follows independently a  $U(0, 1)$  distribution, or alternatively whether  $B_t = \Phi^{-1}(U_t)$  follows a standard normal distribution. If  $\{y_t\}$  is discrete, such as the counts we consider in this paper, then  $U_t$  does not follow a uniform distribution. Smith<sup>28</sup> replaces  $U_t$  by

$$\tilde{U}_t = (1 - z_t)\Pr(Y_t \leq y_t - 1 \mid D_{t-1}) + z_t\Pr(Y_t \leq y_t \mid D_{t-1}), \quad (5)$$

where  $z_t$  is a sequence of i.i.d. random draws from a  $U(0, 1)$  distribution. Then if the goodness of fit is good, the statistic  $\tilde{U}_t$  follows a  $U(0, 1)$  distribution, or  $\tilde{B}_t = \Phi^{-1}(\tilde{U}_t)$  follows a  $N(0, 1)$  distribution. Standard tests of the normal distribution may be used to judge the goodness of fit formally. Informal checks are also available and include QQ-plots and other summaries as discussed in Smith<sup>28</sup>. For the application of this approach one needs to have the predictive cumulative distribution function. This is routinely available for basic generalised linear models, but is not analytically available for dynamic generalised linear models and in particular for those considered in this paper. As a result Frühwirth-Schnatter<sup>29</sup> has suggested to use (5), with

$$\Pr(Y_t \leq y_t \mid D_{t-1}) = \sum_{i=1}^N \tilde{w}_t^{(i)} \Pr(Y_t \leq y_t \mid x_t^{(i)}),$$

where  $x_t^{(i)}$  is the  $i$ -th draw of the state  $x_t$  and  $\tilde{w}_t^{(i)}$  is the  $i$ -th standardised weight. Following the suggestions of Smith<sup>28</sup> and Frühwirth-Schnatter<sup>29</sup>, we report on both  $\tilde{U}_t$  and  $\tilde{B}_t$ , although one is enough to make a judgement. For example,  $\tilde{U}_t$  might be good for informal checks but  $\tilde{B}_t$  might be more suitable for formal tests, such a checking for normality.

In the Monte Carlo experiment above, simulated data are obtained from a Poisson distribution. In this case, the Liu and West (LW.Pois) algorithm is applied to fit this simulated data as a correct model; we have also fitted the LW with negative binomial model (LW.NB) and the LW with Poisson mixture model (LW.MPois) as misspecified models (assuming the wrong model).

The summary comparison of results, based on a Monte Carlo study, is given in Table 2. The Monte Carlo average of the means of the  $P$ -score residuals  $\tilde{U}_t$  (AMPSR) by the correct model and wrong models are not too far apart (0.509, 0.478, 0.545); however, we can see that under the correct Poisson model LW.Pois provides the closest value to 0.5, which is the mean of the  $U(0, 1)$  distribution. The Monte Carlo average of the variances of the  $P$ -score residuals (AVPSR) by both the correct model and misspecified models are quite similar (0.077, 0.071, 0.062), and they are close to the theoretical variance of the uniform distribution, which is equal to 0.08. To judge the distributional assumptions of the  $U_t$  and the  $B_t$ , the Kolmogorov-Smirnov (K-S) test for a uniform assumption and the Shapiro-Wilk (S-W) for a normality test are used. The average of  $P$ -values of the K-S test (APVPSR) over Monte Carlo runs by correct and wrong models were (0.469, 0.281, 0.211). This suggests there is no evidence to reject the hypothesis that the PSR for all models follows a uniform distribution. The figures of the average of  $P$ -value (APVINTPSR) over Monte Carlo samples obtained from the S-W test (0.557, 0.312, 0.257) confirm that the  $\tilde{B}_t$  for all models has a standard normal distribution. From the results reported in Table 2 it can be concluded that when we look at the overall shape of the goodness of fit the Poisson model fits the data very well. The interesting observation is that the goodness of fit when applying the LW algorithm under the wrong model is still good. This suggests that the LW algorithm is not very sensitive to the choice of distribution, if we are looking at the overall goodness of fit. However, when we look at forecasting from one time to



**TABLE 2** Summary comparison of a Monte Carlo study of three proposed Liu West (LW) algorithms for simulated Poisson data. For each of the three algorithms, shown are the average mean P-score residuals (AMPSR), the average variance of the PSR (AVPSR), the average P-value of the Kolmogorov-Smirnov test (APVPSR) and the average P-value of the Shapiro Wilk test (APVINTPSR). Shown are also Monte Carlo standard deviations.

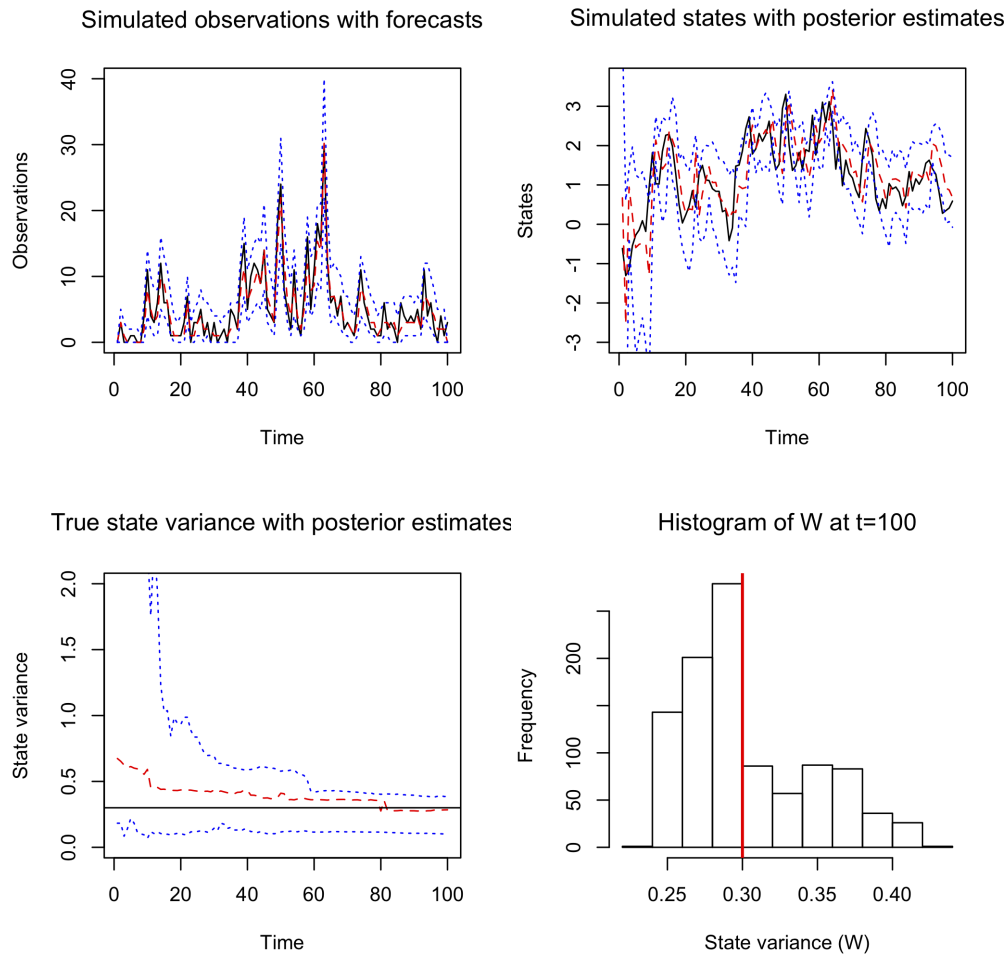
	LW.Pois	LW.NB	LW.MPois
AMPSR/SD	0.5092/0.024	0.478/0.056	0.545/0.087
AVPSR/SD	0.077/0.008	0.071/0.006	0.062/0.074
APVPSR/SD	0.469/0.304	0.281/0.211	0.2106/0.207
APVINTPSR/SD	0.557/0.225	0.312/0.271	0.257/0.232

another there might be differences between the distributions which we choose. This is explored in more detail with the aid of some plots.

Considering now a single non-stationary time series we simulate 100 values from model (1) - (2), with  $\mu = 0$ ,  $\phi = 1$  and  $W = 0.3$ . Figure 1 plots graphs to illustrate the goodness of fit for this Poisson model. We observe that the one-step forecasts (top left panel of the plot) are very close to the original simulated Poisson observations. A plot of the states  $x_t$  and its estimates is given in the top right panel. Again we remark the close fit of the states. The bottom panels of Figure 1 show the posterior model with 95% credible bounds of the variance  $W$ . We remark that as more data are coming in the posterior mode converge to the true value of  $W = 0.3$  (indicated in the plot with the horizontal line); the credible bounds are narrowing as  $t$  increases. The bottom right panel shows a histogram of the posterior distribution of  $W$  at time  $t = 100$ ; again we remark on the accuracy of the estimation with the mode of the histogram being very close to the true value of  $W$ , depicted here by the vertical line.

A similar study was performed on a single Monte Carlo run for the negative binomial model (stationary time series). The true parameters here are  $\phi = 0.8$ ,  $\tau = 2.5$  and  $W = 1.35$ . For a single run the LW algorithm is applied for 100 observations and the results are shown in Figure 2. As in the Poisson case, we remark that the estimates converge to static values quite quickly and this is depicted in Figure 2. The 95% credible bounds are essentially very narrow as compared to those considering non-stationary time series. This has the advantage of very quick convergence to the true value, but the disadvantage that the estimates do not improve over time. Here we have picked a small dispersion parameter; following the discussion in the introduction a larger dispersion parameter would be a reasonable extension to consider, but in this case it may be better to fit a Poisson model directly (here this is not considered due to space limits).

In order to further explore the performance of the Poisson model we make a comparison with a standard MCMC estimation procedure. Due to lack of space and because the models we use in Section 4 aimed at non-stationary time series prediction, we only illustrate the non-stationary dynamic Poisson model, given by (1) - (2) with  $\mu = 0$  and  $\phi = 1$ . The parameters of interest are the dynamic states  $x_t$  and the static variance  $W$ . We simulate a single run of 70 observations from this model with  $x_0 \sim N(0, 10)$  and  $W = 2.5$ . The evolution variance is relatively low for not causing  $x_t$  to be too large which causes numerical issues in the rate of the Poisson  $\lambda_t = \exp(x_t)$ . The Liu and West Particle filter described above is applied; we have used the prior  $x_0 \sim N(0, 10)$  used for the simulation and a weakly information gamma prior for  $W$ ,  $W \sim \text{Ga}(2, 0.1)$ . For the MCMC we have used a standard blocked hybrid scheme, as described in Gamerman and Lopes<sup>31</sup> for dynamic generalised linear models: this involves a Metropolis step in order to draw the states at each point of time and state variance  $W$  is drawn from an inverse gamma distribution, following a standard Gibbs step. A random walk chain and a Gaussian proposal are used: experimentation has led to set the variance of the proposal distribution equal to 0.6 in order to achieve an acceptance probability equal to 28.53% (the small variance allows for small moves of the chain, which works here as  $\beta_t$  is univariate and is unimodal). The chain is run for 10000 iterations, the first 1000 are used as burn-in and the last 1000 are used for illustration purposes. Trace plots and their correlograms (not shown here) indicate convergence of the chain. Figure 3 shows some comparative results of the PF and the MCMC. The top left panel shows modes of the states under each method (PF is indicated by the dashed red line and MCMC is indicated by the dotted blue line) together with the actual values of the states (solid black line). We note that the PF produces posterior estimates of the states, while the MCMC gives smoothed estimates (estimates of the states given all data). Here we see that the MCMC seems to provide a good trend estimation of the states. The top right panel shows the posterior mode of the variance  $W$  using the PF (solid red line) together with the MCMC (mode and credible intervals); the true value of  $W$  is indicated by the horizontal line. Here we see that MCMC is closer to the true value and is more stable than that of the PF. The PF underestimates the true variance, while MCMC slightly overestimates. The middle panels of Figure 3 show histograms of the states for  $t = 48$  and the bottom panels show respective histograms of  $W$ , for  $t = 48$ . This figure indicates that the PF does



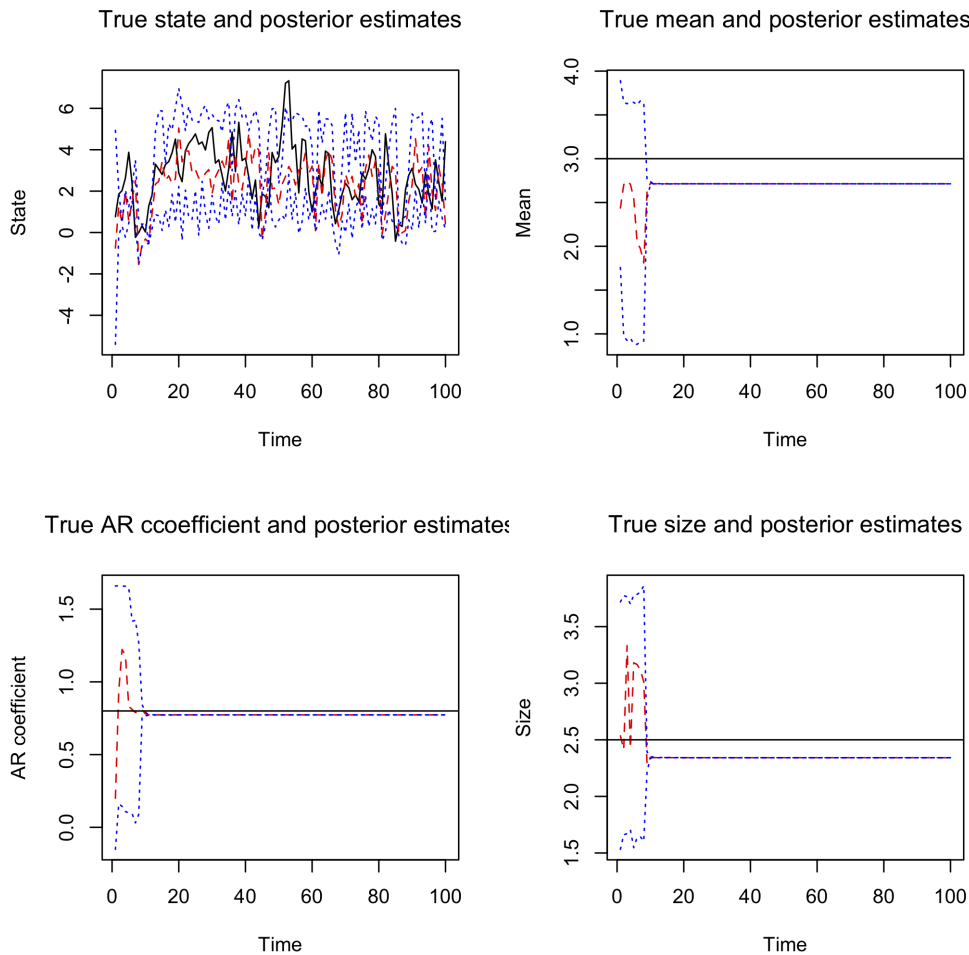
**FIGURE 1** Plot of the performance of the Poisson model for 100 simulated counts. Shown are simulated data and one-step forecasts (top left panel), simulated states and their estimates (top right panel), estimate of the state variance the true value (posterior mode with credible bounds, bottom left panel) and estimate of the state variance ( $t = 100$ ) with the true value (histogram, bottom right panel).

a good job and compares well with the MCMC. A further more in depth comparison could consider comparing the MCMC output to a smoothing-based particle filter.

#### 4 | ASTHMA DATA

In this section we consider data consisting of daily medical contacts for schoolchildren aged between 5 and 16 years old who suffered from asthma over a seven-year period between 1999 and 2005 in England. These data, reported in Julious et al.<sup>11</sup> are depicted in Figure 4 (top panel). The lower panel of this figure shows weekly counts of medical contacts and this is the primary data set we consider in this section. The main reason for this aggregation is to account for the weekend effect, but as a referee has suggested it might be beneficial to consider the daily data and to aggregate the forecasts as necessary. For a recent study on aggregation in time series the reader is referred to Nason et al.<sup>32</sup> A primary interest related to these data involves short-term forecasting of the count of asthma patients. This can provide vital input in hospital bed availability and requirements as well as hospital staff availability and planning of resources.

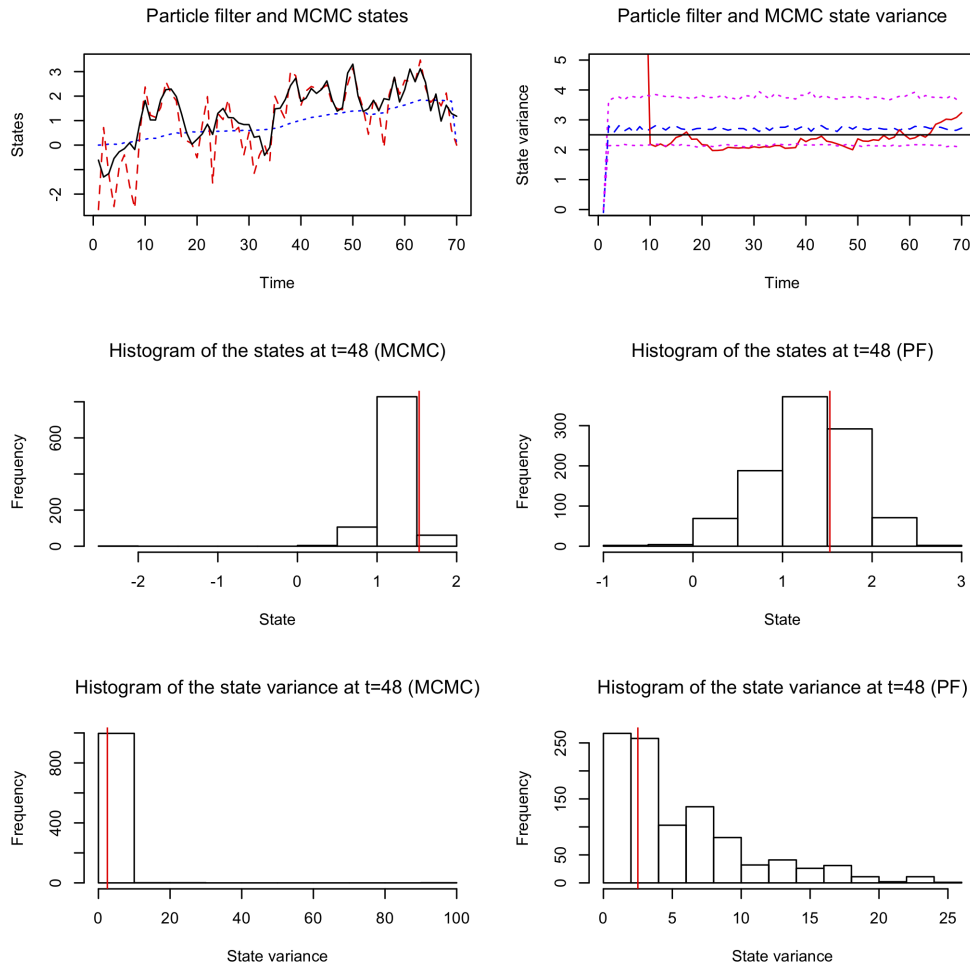
Figure 4 suggests that the weekly data appear to be a non-stationary time series. There appears to be some evidence of seasonality, but this is not persistent and modelling it in the dynamic model did not provide an improvement. Stationarity was



**FIGURE 2** Plot of the performance of the negative binomial model for 100 simulated counts. Shown are simulated states and posterior estimates (top left panel), estimates of the mean  $\mu = 3$  (top right panel), estimates of the AR coefficient  $\phi = 0.8$  (bottom left panel) and estimates of the dispersion parameter  $\tau = 2.5$  (bottom right panel). In each panel plot shown are the posterior mode together with 95% credible bounds (the true value is indicated by the horizontal solid line).

rejected and a correlogram (not shown here) confirmed this. Because the dynamic model should reflect upon the aforementioned non-stationarity, we have set  $\mu = 0$  and  $\phi = 1$  in the evolution (2). Hence, for the Poisson model the hyperparameter is the state variance  $W$  and for the negative binomial model the parameters are the size  $\tau$  and the state variance  $W$ . For the mixture model (4) we set  $\mu_1 = \mu_2 = 0$  and  $\phi_1 = \phi_2 = 1$ .

The Liu and West and Storvik particle filter algorithms are applied to the asthma weekly time series data with the proposed models as in the simulation study without including the process mean (within the proposal models ( $\mu = 0$  and  $\phi = 1$ )). We have used throughout  $N = 1000$  particles. Liu and West<sup>15</sup> discuss Gaussian mixtures for the prior of each of the hyperparameters (see also Section 2.1. For parameters where support is not the real line these authors make use of transformations to map the support of these parameters to the real line, e.g. for  $W$ , one can work with  $\log W$ . In this section we consider a gamma prior for  $W$  (for both Poisson and negative binomial), i.e.  $W \sim \text{Ga}(2, 0.1)$ . For the size  $\tau$  of the negative binomial we consider three possibilities (a) a gamma prior  $\tau \sim \text{Ga}(2, 0.1)$ , (b) a uniform prior  $\tau \sim \text{U}(0, 50)$  and (c) a uniform prior for  $\tau^{-1}$ ,  $\tau^{-1} \sim \text{U}(0, 1)$ . In (a) the gamma prior is unbounded from above to allow large values of  $\tau$ ; moreover, this gamma prior is a weakly informative prior with prior mode 10 and prior variance 200. In (b) the non-informative prior is bounded above, but a large value 50 is chosen; here the prior mean is 25 and the variance is 208.33. In (c) the non-informative prior for  $\tau^{-1}$  gives the Poisson model when  $\tau^{-1} = 0$  in the boundary. With respect to the mixture Poisson model (see equation (4)), the prior distribution for the static

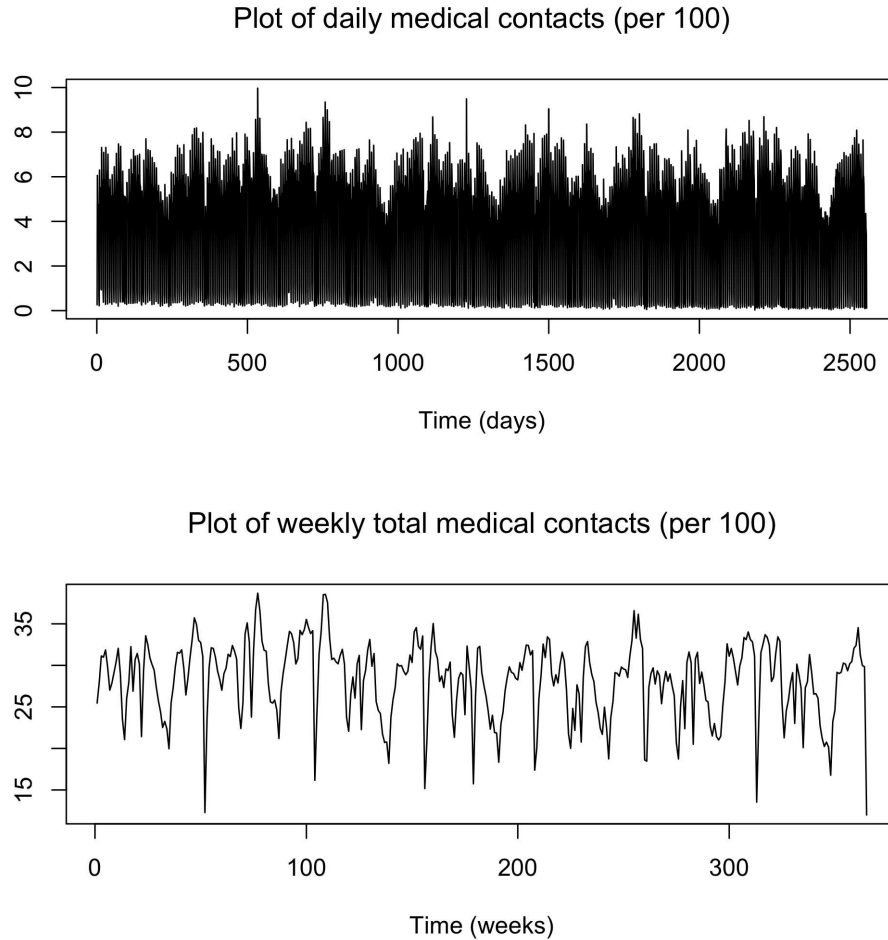


**FIGURE 3** Comparison between PF algorithm with MCMC, for the dynamic Poisson model. Top left panel (estimation of  $x_t$ ): posterior mode using the PF (dashed red line), smoothed mode using MCMC (dotted blue line) and true states (solid black line). Top right panel (estimation of  $W$ ): posterior mode using the PF algorithm (solid red line), posterior mode using MCMC (dashed blue line), 95% credible bounds using MCMC (dotted purple line) and true value (horizontal line). Middle panels: histograms of  $x_{48}$  using PF and MCMC (shown is the true value of  $x_{48} = 1.5284$ , depicted by the vertical lines). Bottom panels: histograms of estimates of  $W$  at  $t = 48$  using PF and MCMC; the vertical lines indicate the true value  $W = 2.5$ .

parameters and the initial states are set up as follows:  $W_i \sim \text{Ga}(2, 0.1)$ ,  $\theta \sim \text{Beta}(2, 1)$  and  $x_{0,i} \sim \text{N}(0, 10)$ , for  $i = 1, 2$ . The value of the shrinkage parameter is determined as  $c = 0.975$ , for all proposal models.

The different proposed algorithms are evaluated based on a root mean square one-step ahead forecast error (RMSE), the mean of the squared standardised residuals (MSSR) and the mean absolute deviation (MAD); see Table 3. According to these figures, the LW (Poisson) algorithm exhibits the smallest RMSE (2.3982) followed by the Storvik algorithm (6.440). According to the MSSE all but one of the LW models produce an impressive performance (we compare MSSR to 1); the best performer is the negative binomial model with gamma prior for  $\tau$  (0.9999). The Storvik model produces very large MSSE, which indicates this model underestimates the forecast variance. According to the MAD the best model is the Poisson (1.6272). Overall the performance of the negative binomial model with gamma prior is similar to that of the mixture of two Poisson distributions.

Figure 5 exhibits the final 105 observations of the real data together with one-week-ahead forecasting by using the Liu and West algorithm for the Poisson and the negative binomial models; the Poisson mixture forecasts are not shown for clarity (as mentioned above the forecast performance of this model is very similar to that of the negative binomial with a gamma prior of  $\tau$ ). The LW with Poisson model gives a number of forecasts closer to the real data than does the negative binomial, but for



**FIGURE 4** Daily and weekly total medical contacts for asthmatic children.

**TABLE 3** Comparison of four algorithms: Liu and West (LW) with Poisson, LW with negative binomial, LW with a Poisson mixture and Storvik (STK) with a Poisson distribution. Shown are the root mean squared error (RMSE), the mean of squared standardized residuals (MSSR) and the mean absolute deviation (MAD).

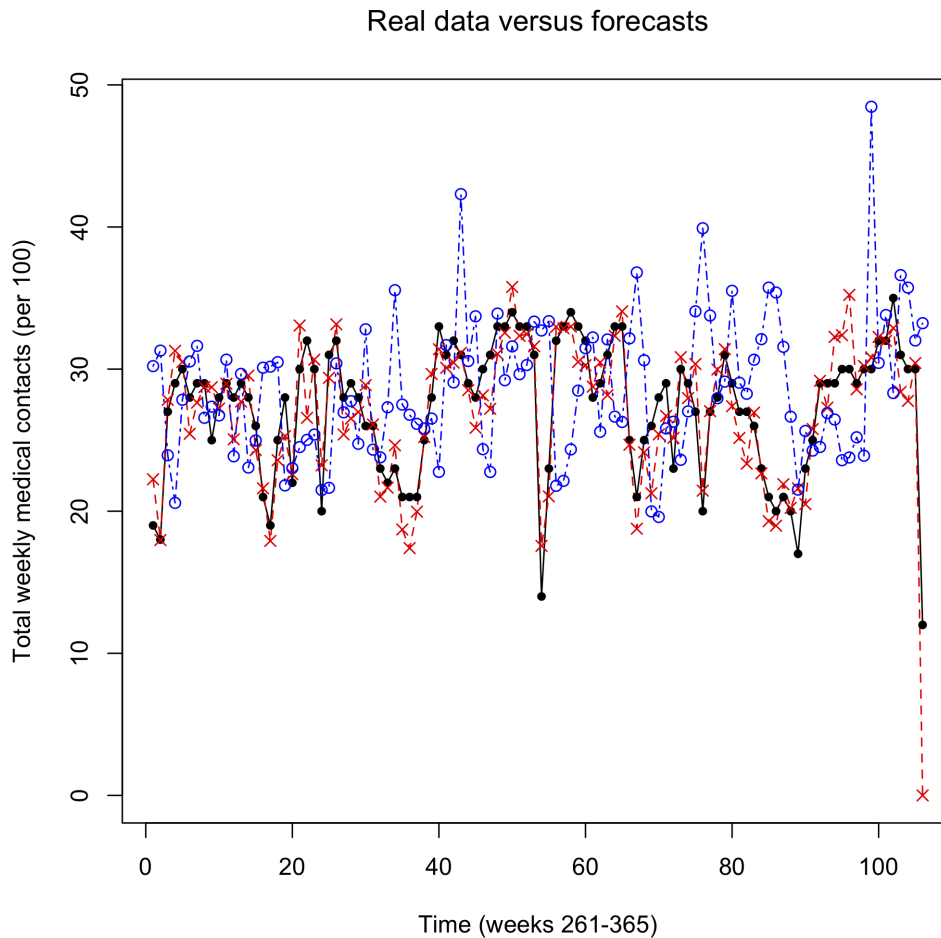
Filter	RMSE	MSSR	MAD
LW (Poisson)	2.3982	1.0299	1.6272
LW (negative binomial, $\tau \sim \text{Ga}(2, 0.1)$ )	18.7317	0.9999	17.6130
LW (negative binomial, $\tau \sim U(0, 50)$ )	2104.6700	6.6959	137.3215
LW (negative binomial, $\tau^{-1} \sim U(0, 1)$ )	1596.4400	0.9996	828.6936
LW (Poisson mixture)	26.9208	1.0087	17.2247
STK (Poisson)	6.440	6.141	5.108

some observations the negative binomial model outperforms the Poisson model. This is in agreement with the overall model performance depicted in Table 3 and highlights the good performance of LW algorithm for both models.

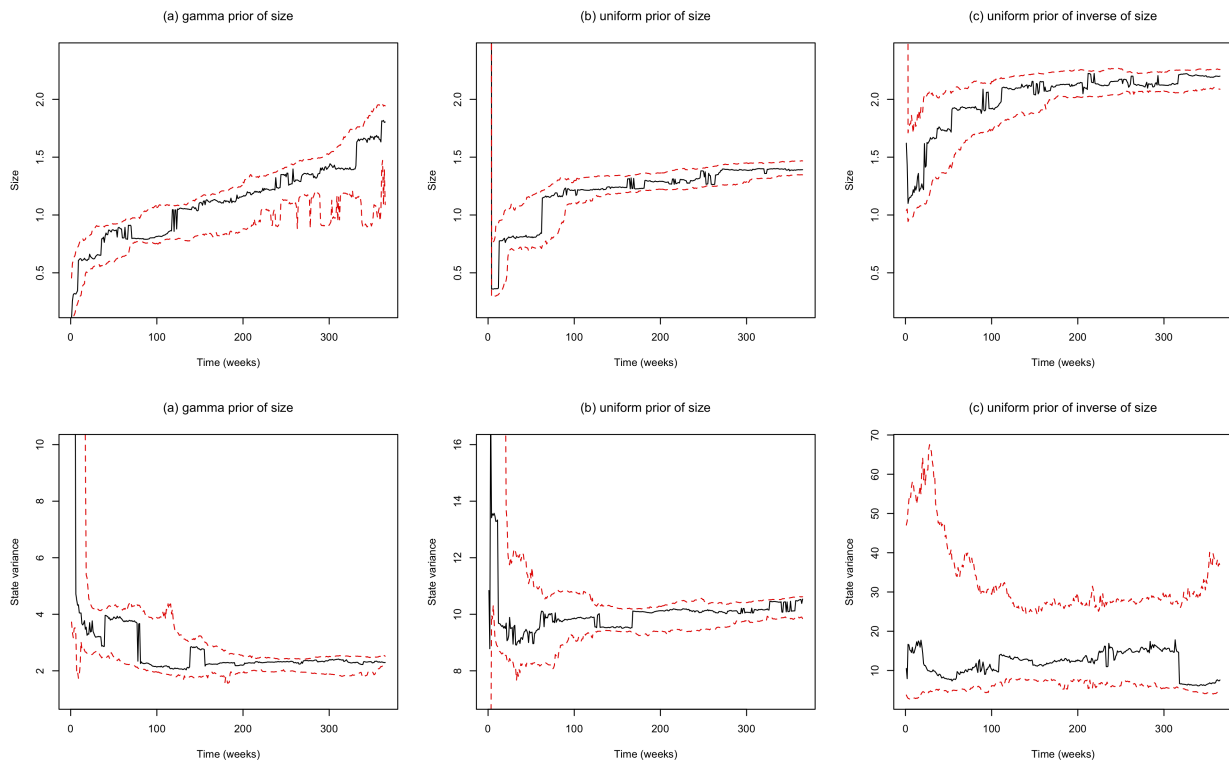
For the three negative binomial models, Figure 6 shows the estimates of the parameters  $\tau$  and  $W$  (here all data points considered 1-365). We see that all three priors for  $\tau$  considered here produce estimates of  $\tau$  in the bound  $(0, 2.5)$ , for all  $t$ . This indicates that forecasts generated by the negative binomial model have left skewed distributions with the variance being larger than the mean. We remark that the uniform prior of  $\tau$  in the middle panel of Figure 6 produces small posterior estimates for  $\tau$ , which is

responsible for the large variance depicted in Table 3 by the large values of RMSE and MAD, for this model. We also observe from Figure 6 that the state variance  $W$  (gamma prior model) is more stable after about 200 time points compared to the other two models. Moreover, the credible bounds of  $W$  (gamma prior model) are the most narrow with their values not exceeding the value of 5 and be consistently less than 4 after 200 time points. Figure 7 shows posterior mode and credible bounds of the state variance  $W$  under the Poisson model. We observe that after 200 time points the mode is quite stable and the credible bounds do not exceed the value of 2. The low estimated values of the dispersion parameter  $\tau$  in Figure 6 put forward the negative binomial model and provide evidence against the Poisson (we would expect  $\tau$  to be large to favour the Poisson). This appears to be surprising as the Poisson model forecasts in Figure 5 seem reasonable and Table 3 indicating the Poisson is a good fit to the data. A close look at Figure 5 reveals that at the start of the series some of the negative binomial forecasts are poor, while towards the end the negative binomial provides some impressive forecasts ( $t = 105$ ).

Finally, Figure 8 show histograms of the count for both the Poisson and the negative binomial models. The histograms are picked at three points of time ( $t = 8, 70, 105$ ) to reflect on the performance of the two models at different times; plotted are the true observations (vertical lines). We remark that for some points of time the Poisson model is better (e.g. for  $t = 8$ , corresponding to week 267) and at some points the negative binomial model is better (e.g. for  $t = 105$ , corresponding to week 365). Both models are reasonable and provide good forecast performance, but there is little support for symmetric histograms for the data (we split the data in time-intervals of length 40 and we found that they were skewed).



**FIGURE 5** Real data (black solid line with solid points) against one step ahead forecasts for the LW.poisson model (red dashed line with cross), and LW.negative binomial model (blue dotted line with circle.)



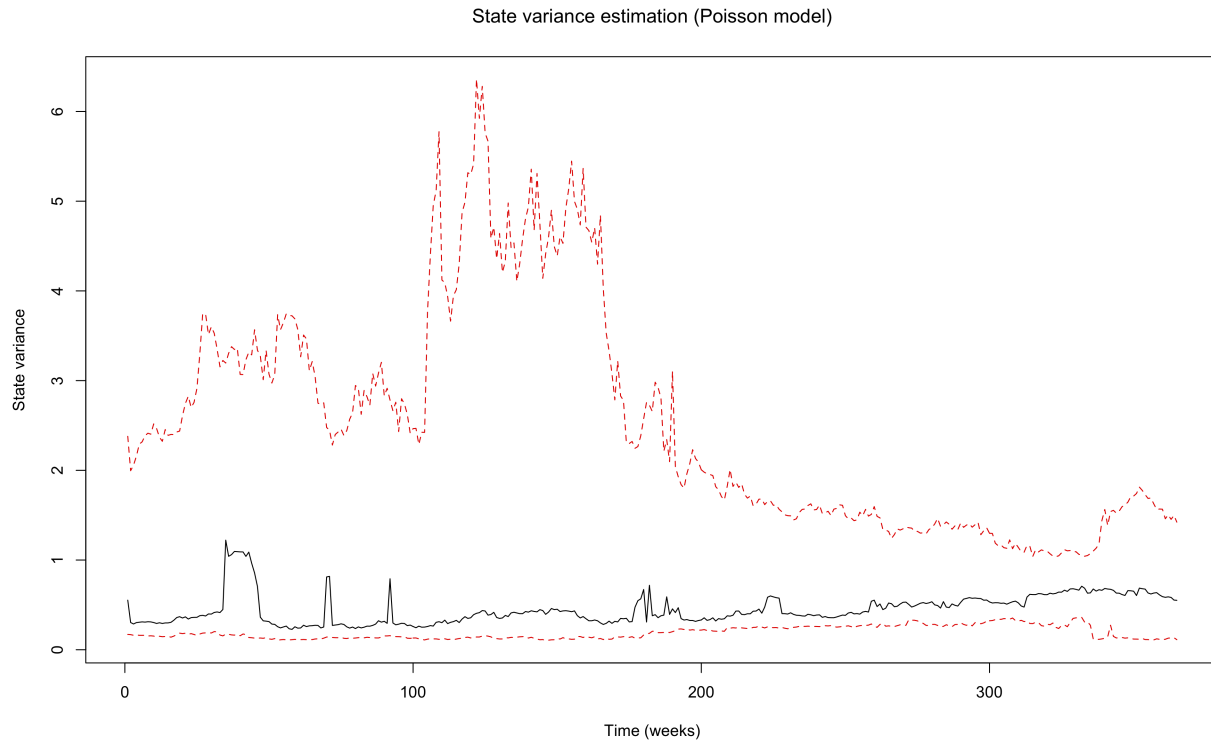
**FIGURE 6** Posterior estimates of the size  $\tau$  and state variance  $W$  under the three priors of  $\tau$ : (a) gamma prior, (b) uniform prior and (c) uniform prior for  $\tau^{-1}$ . Shown are posterior modes with 95% credible bounds.

## 5 | CONCLUSIONS

In this paper three dynamic models are proposed for the modelling and forecasting of count time series. We have put forward sequential Monte Carlo methodologies and in particular the Liu and West<sup>15</sup> (LW) and the Storvik<sup>16</sup> (STK) estimation procedures. Both of these algorithms are able to deal with the estimation of static parameters, such as autoregressive and variance parameters. The proposed models are built around Poisson and negative binomial models, which are attractive default models for count data; the negative binomial model is more versatile as it can incorporate the Poisson model as a special case. Simulation results indicate that this methodology works well. Data of counts of medical contacts of schoolchildren suffering from asthma is used in order to illustrate the utility and the application of the proposed estimation. It is found that the Poisson and negative binomial dynamic model fit well to the data. The proposed algorithms are general and can deal with a host of other response distributions, as long as we are able to sample from these distributions. Examples may include zero-inflated Poisson and finite Poisson mixtures of higher order than that considered in this paper.

## References

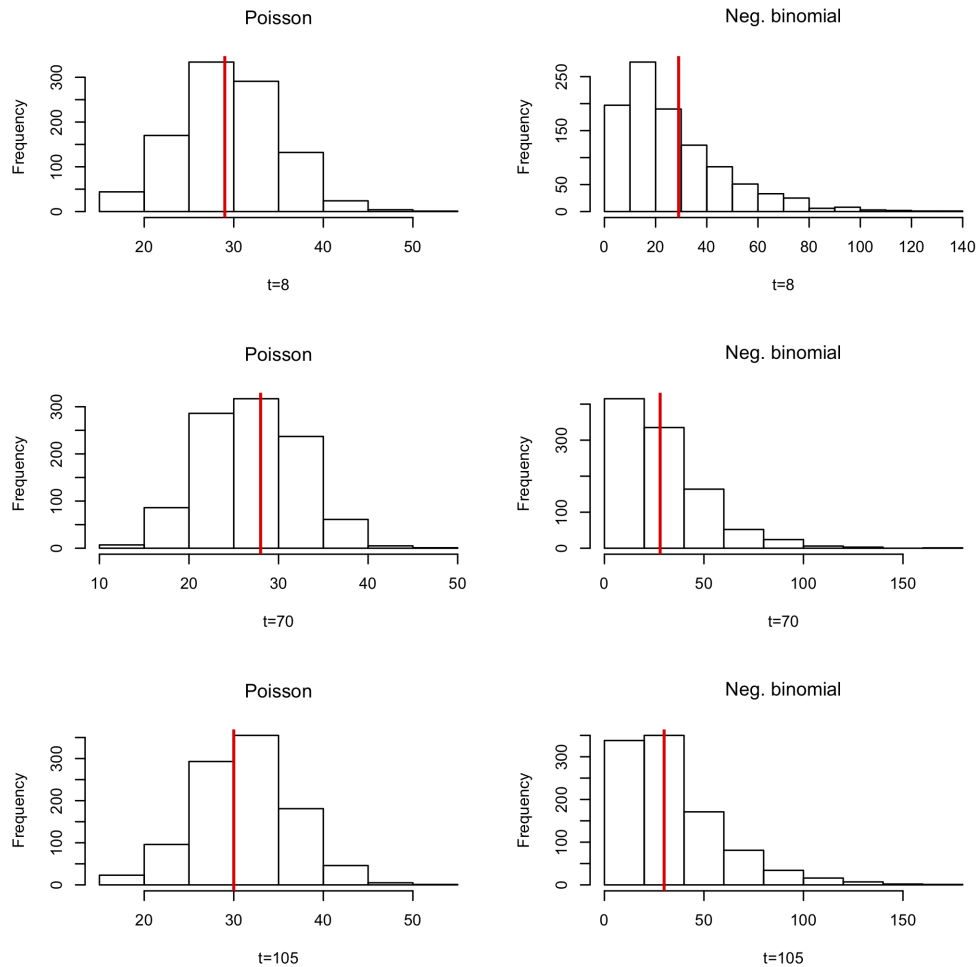
1. Harvey AC, Fernandes C. Time series models for count or qualitative observations. *Journal of Business and Economic Statistics* 1989; 7(4): 407–417.
2. West M, Harrison PJ, Migon HS. Dynamic generalised linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association* 1985; 80: 73–97.
3. Fahrmeir L. Posterior mode estimation by extended Kalman filtering for multivariate generalised linear models. *Journal of the American Statistical Association* 1992; 87: 501–509.
4. Kitagawa G, Gersch W. *Smoothness Priors Analysis of Time Series*. Springer, New York . 1996.



**FIGURE 7** Posterior estimates of the state variance  $W$  using the Poisson model; shown are the posterior mode and 95% credible bounds.

5. West M, Harrison PJ. *Bayesian Forecasting and Dynamic Models*. Springer, New York. 2nd ed. 1997.
6. Fahrmeir L, Tutz G. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer, New York . 2001.
7. Kedem B, Fokianos K. *Regression Models for Time Series analysis*. John Wiley . 2002.
8. Triantafyllopoulos K. Inference of dynamic generalized linear models: on-line computation and appraisal. *International Statistical Review* 2009; 77(3): 430–450.
9. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Statistics in Medicine* 2010; 29: 2224–2234.
10. He J, McGee DL, Niu X. Application of the Bayesian dynamic survival model in medicine. *Statistics in Medicine* 2010; 29: 347–360.
11. Julious SA, Campbell MJ, Bianchi SM, Murray-Thomas T. Seasonality of medical contacts in school-aged children with asthma: association with school holidays. *Public Health* 2011; 125: 769–776.
12. Shumway RH, Stoffer DS. *Time Series Analysis and Its Applications: With R Examples*. Springer, New York. 4th ed. 2017.
13. Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 2000; 10: 197–208.
14. Künsch HR. Particle filters. *Bernoulli* 2013; 19(4): 1391–1403.
15. Liu J, West M. *Sequential Monte Carlo Methods in Practice*. Combined parameter and state estimation in simulation-based filtering; Springer, New York . 2001.
16. Storvik G. Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing* 2002; 50: 281–290.





**FIGURE 8** Empirical predictive densities of the Poisson model (left panel) and the negative binomial model (right panel). Densities are plotted at three points of time  $t = 8, 70, 105$  and the observed counts are depicted by the vertical lines.

17. Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B* 2010; 72(3): 269–342.
18. Carvalho CM, Johannes M, Lopes H, Polson N. Particle learning and smoothing. *Statistical Science* 2010; 25(1): 88–106.
19. Nemeth C, Fearnhead P, Mihaylova L. Sequential Monte Carlo Methods for state and parameter estimation in abruptly changing environments. *IEEE Transactions on Signal Processing* 2014; 62(2): 1245–1255.
20. Karlis D, Xekalaki E. Mixed Poisson distributions. *International Statistical Review* 2005; 73(1): 35–58.
21. Gamerman D. Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika* 1998; 85(1): 215–227.
22. Gordon NJ, Salmond DJ, Smith AFM. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F-Radar and Signal Processing* 1993; 140(2): 107–113.
23. Pitt MK, Shephard N. Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association* 1999; 94(446): 590–599.
24. Stephens M. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society Series B* 2000; 62(4): 795–809.

25. Frühwirth-Schnatter S. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* 2001; 96: 194–209.
26. Jasra A, Holmes CC, Stephens DA. Markov chain Monte Carlo methods and the label switching problem in Bayesian Mixture Modeling. *Statistical Science* 2005; 20(1): 50–67.
27. Frühwirth-Schnatter S. *Finite Mixture and Markov Switching Models*. Springer, New York . 2006.
28. Smith JQ. Diagnostic checks of non standard time series models. *Journal of Forecasting* 1985; 4: 283–291.
29. Frühwirth-Schnatter S. Recursive residuals and model diagnostics for normal and non-normal state space models. *Environmental and Ecological Statistics* 1996; 3(4): 291–309.
30. Czado C, Gneiting T, Held L. Predictive model assessment for count data. *Biometrics* 2009; 65: 1254–1261.
31. Gamerman D, Lopes HF. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, New York . 2006.
32. Nason GP, Powell B, Elliott D, Smith PA. Should we sample a time series more frequently?: decision support via multirate spectrum estimation (with discussion). *Journal of the Royal Statistical Society Series A* 2017; 180(2): 353–407.

