Fitting a spatial-temporal rainfall model using Approximate Bayesian Computation

Nanda Aryal[1], Owen D. Jones[2]

[1] University of Melbourne, Australia
[2] Cardiff University, UK

E-mail for correspondence: `joneso18@cardiff.ac.uk`

Abstract: We fit a stochastic spatial-temporal model to high-resolution rainfall radar data for a single rainfall event. Approximate Bayesian Computation (ABC) is used to fit a model of Cox, Isham and Northrop, previously fitted using the Generalised Method of Moments (GMM). We then show that ABC readily adapts to more general, and thus more realistic, variants of the model. The Simulated Method of Moments (SMM) is used to initialise the ABC fit.

Keywords: Spatial-temporal; spatiotemporal; rainfall; Approximate Bayesian Computation.

1 Introduction

The Cox-Isham-Northrop (C-I-N) rainfall model is a spatial-temporal stochastic model for a rainfall event, constructed using a cluster point process. The cluster process is constructed by taking a primary process, called the storm arrival process, and then attaching to each storm center a finite secondary point process, called a cell process. To each cell center we then attach a rain cell, with an associated area, duration and intensity. The storm and cell centers all share a common velocity. The total rainfall intensity at point (x, y) and time t is then the sum of the intensity at (x, y) of all cells active at time t. (Cox & Isham 1988, Northrop 1998.)
The storm arrival process is taken to be a Poisson process in $\mathbb{R}^2 \times [0, \infty)$ with homogeneous rate λ. Let $\mathbf{v} = (v_x, v_y)$ be the velocity of the rainfall event, so if a storm center arrives at $(\mathbf{u}, s)$ then at time $s + t$ it will be at $(\mathbf{u} + t\mathbf{v}, s + t)$. Storm durations are random with an $\exp(\gamma)$ distribution.
While a storm is active it produces cells at a rate β in time, starting with a cell at the moment the storm center begins. If the storm arrives at $(\mathbf{u}, s)$

This paper was published as a part of the proceedings of the 33rd International Workshop on Statistical Modelling (IWSM), University of Bristol, UK, 16-20 July 2018.

and produces a cell at time $s+t$, the cell will be centered at $\mathbf{u}+t\mathbf{v}+\mathbf{w}$, where $\mathbf{w}$ comes from a Gaussian distribution with mean $\mathbf{0}$ and covariance Σ. The cell centre then also moves with velocity $\mathbf{v}$. We parameterise Σ using its size σ^2, eccentricity e and orientation Θ. σ^2 is assumed to have an inverse-gamma distribution, where $1/\sigma^2$ has mean ξ_μ and coefficient of variation ξ_{CV}.
Individual cells have random durations, distributed as $\exp(\eta)$, and random sizes. Rain cells are elliptical, with the same eccentricity e and orientation Θ as the storms. The size is given by the major axis, which is distributed as a $\Gamma(\alpha_1, \alpha_2)$. Note that we can re-express α_2 using α_1, e, and μ_A (the mean area of a rain cell). The intensity of a rain cell is constant over the shape and duration of the cell, with an exponential distribution mean μ_X. The displacements, durations, sizes, and intensities of a cell are all independent, and independent of other cells.
The C-I-N model is stationary and is used to model the 'interior' of a rainfall event. We will suppose that we have observations of the rainfall in some finite space-time window $A \times [0, T]$, where T is chosen so that the leading and trailing edges of the rainfall event are not observed. For this study we used radar data collected at Laverton, Melbourne, on 24th September 2016, calibrated by the Australian Bureau of Meteorology using rain-gauge data. The data gives rainfall intensity averaged over 1 km square pixels every 6 minutes, over an area of 180×180 pixels for a period of 3 hours. A contour plot of the spatial rainfall intensity at a single time-point is given in Figure 1(a).
Because it has an intractible likelihood function, the C-I-N model has been fitted using the Generalized Method of Moments (GMM) (Wheater et al. 2006). The puprose of this paper is firstly to show that Approximate Bayesian Computation (ABC) can be used to fit a Bayesian version the C-I-N model, and secondly to use ABC to fit a generalisation of the C-I-N model that is too much for GMM to cope with. GMM fitting matches theoretical and observed moments of the process, and thus is restricted to moments for which you have an analytic expression. ABC fitting compares the observed process to simulations, and thus places no restrictions on the statistics used to compare them. The penalty we pay for this increased flexibility is an increase in computational time.

1.1 Approximate Bayesian Computation

ABC was introduced by Pritchard et al. (1999), and was later extended to incorporate Markov Chain Monte Carlo (MCMC) by Marjoram et al. (2003), or alternatively Sequential Monte Carlo (SMC) (Sisson et al. 2007 and 2009, Beaumont et al. 2009). We will use the ABC-MCMC methodology.
We suppose that we have an observation D from some model $f(\cdot|\boldsymbol{\theta})$, depending on parameters $\boldsymbol{\theta}$, and that we are able to simulate from f. Let π

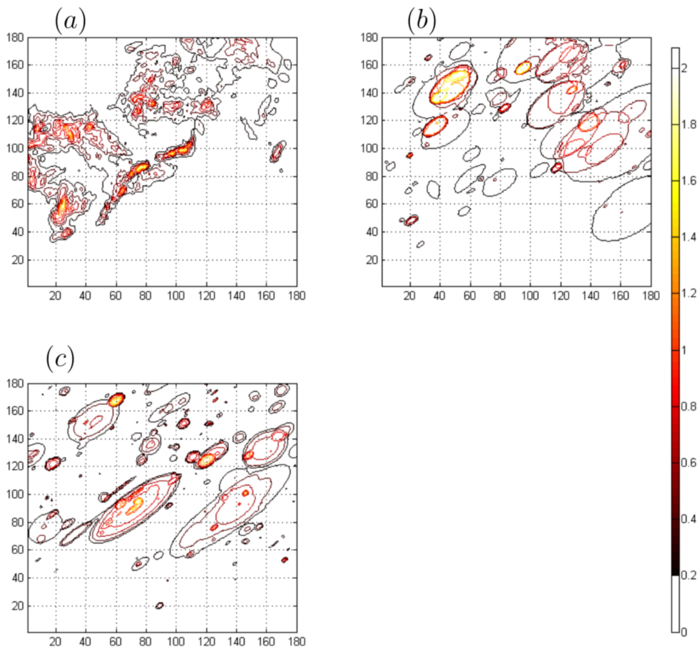

FIGURE 1. (a) Calibrated rainfall radar data, courtesy of the Australian Bureau of Meteorology. (b) Simulation from the C-I-N model fitted using ABC. (c) Simulation from the modified C-I-N model.

be the prior distribution for $\boldsymbol{\theta}$ and $S = S(D)$ a vector of summary statistics for D, then ABC generates samples from $f(\boldsymbol{\theta}|\rho(S(D^*), S(D)) < \epsilon)$, where $D^* \sim f(\cdot|\boldsymbol{\theta})$, $\boldsymbol{\theta} \sim \pi$, and ρ is some distance function. If S is a sufficient statistic, then as $\epsilon \to 0$ this will converge to the posterior $f(\boldsymbol{\theta}|D)$. ABC-MCMC adds a proposal chain with density q and a rejection step, to generate a sample $\{\boldsymbol{\theta}_i\}$. The algorithm is as follows:

FOR $i = 1$ to N

1 Given current state $\boldsymbol{\theta}_i$ propose a new state $\boldsymbol{\theta}^*$ using $q(\cdot|\boldsymbol{\theta}_i)$

2 Put $\alpha = \min\{1, (\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}_i|\boldsymbol{\theta}^*))/(\pi(\boldsymbol{\theta}_i)q(\boldsymbol{\theta}^*|\boldsymbol{\theta}_i))\}$

3 Go to 4 with probability α, otherwise set $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$ and return to 1

4 Simulate data $D^* \sim f(\cdot|\boldsymbol{\theta}^*)$

5 If $\rho(S(D^*), S(D)) \leq \epsilon$ then set $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}^*$, otherwise set $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$

END FOR

Note that the MCMC rejection at step 3 comes before the ABC comparison in step 5. This is to avoid unnecessarily running the simulation in step 4.

2 Fitting the C-I-N model using ABC

Following Wheater et al. (2006), the velocity $\mathbf{v}$, eccentricity e and orientation Θ were all estimated ad hoc using temporal and spatial autocovariance estimates, and then fixed.
The remaining parameters were transformed to reduce dependence and skewness. For the ABC step we used $\log(\lambda/\gamma)$, $\log(\lambda\gamma)$, $\log(\beta/\eta)$, $\log(\beta\eta)$, $\log(\mu_X/\mu_A)$, $\log(\mu_X\mu_A)$, $\log(\alpha_1)$, $\log(\xi_\mu)$, and $\log(\xi_{CV})$. Vague normal priors are used for all the transformed parameters, and for the proposal chain we used a random walk with $N(0, 0.2^2 I)$ steps.
The choice of summary statistics S and distance metric ρ plays a large part in the performance of ABC. Ideally S should be sufficient, but certainly it should reflect those aspects of the real process considered most important. However choosing S too large reduces the efficiency of ABC, though this can be mitigated to some extent using post hoc analysis of the significance of each component (Beaumont et al. 2002).
We used 23 summary statistics:

- The overall mean and standard deviation of rainfall, taken over all pixels and all times.
- The spatial-temporal auto-correlation, with lags of (x, y, t), where x and y are measured in pixels and t is in units of 6-minutes. We take $t = 0, x \in \{-1, 0, 1\}, y \in \{-1, 0, 1\}$, and $t = 1, x \in \{-1, 0, 1\} + v_x, y \in \{-1, 0, 1\} + v_y$. Here v_x and v_y are the velocity components, in units of pixels per 6-minutes. Note that the lag $(0, 0, 0)$ auto-correlation is not used because it is just variance.
- The probability of an arbitrary pixel and time being dry.
- The ratio of dry/wet area and mean and standard deviation of wet area, averaged over time.

For the distance function ρ we used a weighted sum of squares $\rho(S(D^*), S(D)) = \sum_i w_i(S^*(i) - S(i))^2$, where $S^*(i)$ and $S(i)$ are the i-th components of $S(D^*)$ and $S(D)$ respectively. We found empirically, as have other authors, that a good choice is to take w_i inversely proportional to the variance of $S^*(i)$ under the posterior.
Plots of our fitted posteriors are given in Figure 2, and a simulation from the fitted model is given in Figure 1(b), for a single time-point. For the simulation the posterior means were used as point estimates for the parameter values.

2.1 Starting ABC using SMM

The Simulated Method of Moments (SMM) is a variant of the Generalised Method of Moments (GMM) that uses Monte-Carlo estimates of moments, rather than analytic expressions (McFadden 1989). Thus, like ABC, using

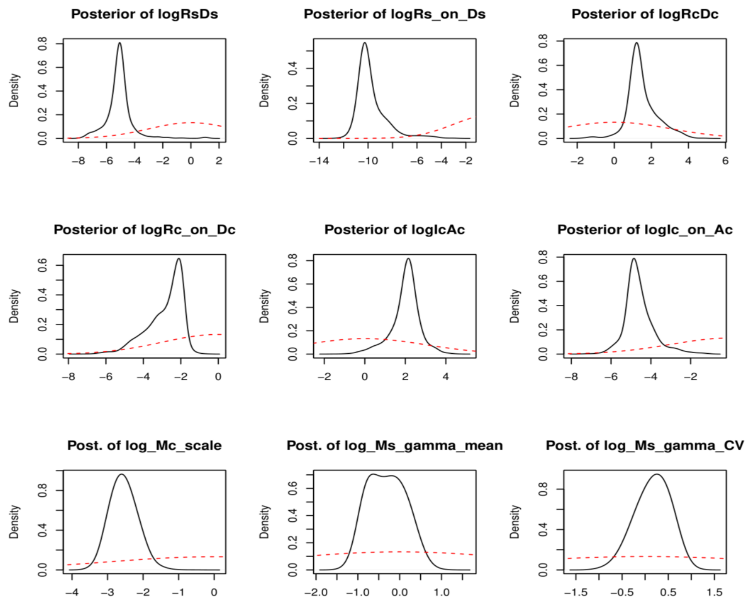

FIGURE 2. Some nice looking posteriors. Priors are given by the red dashed lines. See the text for details.

SMM we have much more freedom in the choice of moments used to fit the model to the data.

When applying ABC, we found it advantageous to 'jump-start' the algorithm by choosing the initial parameter selection $\boldsymbol{\theta}_0$ using an SMM fit, using the same summary statistics S that we chose for the ABC fitting. There are two main benefits to this step. The first is that if $\boldsymbol{\theta}_0$ has very small posterior probability, then ABC-MCMC requires a prohibitively large burn-in period. The second is that it gives us a distribution for $S(D^*)$ that we can use to estimate the weights w_i of the distance function ρ.

Previous authors have suggested using a separate ABC step to initialise $\boldsymbol{\theta}_0$; we found that using SMM instead requires much less computation time.

3 Extending the C-I-N model

There are many ways in which the C-I-N model can be extended. If you do so, however, GMM is no longer suitable for estimation, as it becomes too difficult to obtain analytic expressions for the moments. Fortunately this does not apply to ABC, which can be applied much as before. For the example below we did not even have to modify the set of summary statistics S.

We leave a comprehensive generalisation of the C-I-N model to future work, and just consider the following modifications:

- Randomised cell eccentricity.
- Rainfall intensity that increases continuously from the edge to the centre of each cell, rather than acting as a step function.
- Heavy tailed distributions for cell intensity and area.
- Correlated cell intensity and area.

Using the posterior distribution of $S(D^*)$ we can show that the modified model gives a better fit. A simulation from the fitted model is given in Figure 1(c), for a single time-point; qualitatively it also looks to be doing a better job

References

Beaumont, M.A., J.-M. Cornuet, J.-M. Marin, and C.P. Robert (2009). Adaptive approximate Bayesian computation. *Biometrika 96*(4):983–990.

Beaumont, M.A., W. Zhang, and D.J. Balding (2002). Approximate Bayesian computation in population genetics. *Genetics 162*(4):2025–2035.

Cox, D.R. and Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proc. Roy. Soc. A*, 415(1849):317–328

Northrop, P. (1998). A clustered spatial-temporal model of rainfall. *Proc. Roy. Soc. A*, 454(1975):1875–1888.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026.

Marjoram, P., J. Molitor, V. Plagnol, and S. Tavaré (2003). Markov Chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States 100*(26):15324–15328.

Pritchard, J.K., M.T. Seielstad, A. Perez-Lezaun, and M.W. Feldman (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution 16*:1791–1798.

Sisson, S., Y. Fan, and M.M. Tanaka (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States 104*:1760–1765.

Sisson, S., Y. Fan, and M.M. Tanaka (2009). Correction for "Sequential Monte Carlo without likelihoods". *Proceedings of the National Academy of Sciences of the United States 106*:16889.

Wheater, H.S., V.S. Isham, R.E. Chandler, C.J. Onof and E.J. Stewart (2006). Improved methods for national spatial-temporal rainfall and evaporation modelling for BSM. Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme, Technical Report F2105/TR.