Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2009)
58, *Part* 2, *pp.* 197–210

M-estimation of Boolean models for particle flow experiments

Jason A. Osborne

North Carolina State University, Raleigh, USA

and Tony E. Grift

University of Illinois, Urbana, USA

[Received June 2007. Revised August 2008]

Summary. Probability models are proposed for passage time data collected in experiments with a device that was designed to measure particle flow during aerial application of fertilizer. Maximum likelihood estimation of flow intensity is reviewed for the simple linear Boolean model, which arises with the assumption that each particle requires the same known passage time. *M*-estimation is developed for a generalization of the model in which passage times behave as a random sample from a distribution with a known mean. The generalized model improves the fit in these experiments. An estimator of total particle flow is constructed by conditioning on lengths of multiparticle clumps.

Keywords: Boolean models; Coverage processes; Infinite server queues; Likelihood; *M*-estimation

1. Introduction

Measuring the outflow of granular particles from an airborne spreader during the aerial application of fertilizer or pesticide presents agricultural engineers with a difficult problem. The goal of uniform distribution over a targeted field requires knowledge about the flow rate of the material as it is dropped from the aircraft. Wind speed, air speed, granule properties, humidity and temperature have been identified (Casady *et al.*, 1997) as factors which can lead to variability in these outflow rates and hence amounts of material that reach the target, so that they achieve an average target flow rate. In practice, pilots use a simple lever-operated gate to change the flow rate to account for extreme values of these factors. This adjustment is based on intuition, without any feedback from measurement of particle flow.

One approach to providing the pilot with more information uses an optical sensor device (Grift and Hofstee, 1997) which measures the velocity (in metres per second) and size of clumps of particles as they flow through the spreader duct. This device has two photosensitive arrays of optical sensors that receive a signal from a light source. As a particle passes an active area, it blocks this light thereby interrupting the signal that is received by the sensors. As long as all the sensors in the array are receiving a high signal, the channel is classified as unoccupied and this is taken as an indication that there are no particles flowing through at that instant. If the signal to any one of the sensors is interrupted, this is interpreted as the presence of at least one particle, constituting a clump, in flow. The two sensor arrays are 0.78 mm apart and it is possible

Address for correspondence: Jason A. Osborne, Department of Statistics, North Carolina State University, Campus Box 8203, Raleigh, NC 27695-8203, USA.
E-mail: jaosborn@stat.ncsu.edu

 0035–9254/09/58197

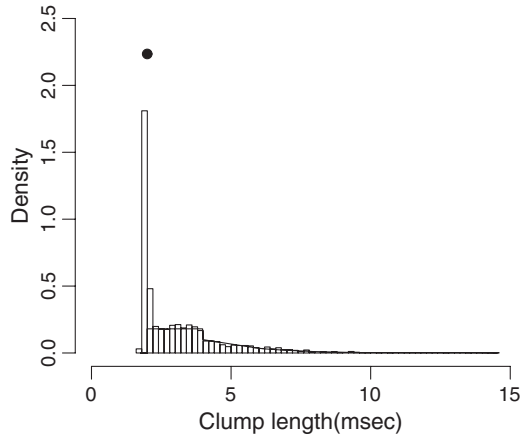

Fig. 1. Probability histogram of 1790 clump lengths, generated by 4000 particles (run 20): ———, estimated clump length density from deterministic segment length; •, estimated point mass at the passage time required by a single particle

to measure the time in milliseconds that it takes a clump to move from one array to the other, Δt_{f}. The total time that either array is blocked, Δt_{b}, is also measured, facilitating calculation of velocity $v = 0.78/\Delta t_{\mathrm{f}}$ and clump length in millimetres, $\mathrm{CL} = v\Delta t_{\mathrm{b}}$. These observable clump lengths, either in terms of physical length or time, are the basis for inference about particle flow in the system. Such a measurement device is called a type II counter (Pyke, 1958).

Grift *et al.* (2001) and Grift (2002) carried out bench scale experiments to evaluate the optical sensor device in situations that were designed to simulate the flow of fertilizer particles through an airborne spreader duct. In these experiments, a known number of steel spherical air rifle pellets was used. These particles have a known mean diameter of 4.45 mm, with little variance. They were dropped from predetermined heights through a duct on which the sensor device was installed. The heights from which the particles were dropped was controlled at several values to simulate a range of particle velocities and flow rates. A histogram representing the distribution of particle clump lengths in units of time, obtained from one run of these experiments, is shown in Fig. 1. The relative frequencies for clump lengths (in milliseconds) are based on dropping 4000 particles from a fixed height.

The goal of this paper is to obtain a probability model that is appropriate for data generated by the sensor device that was used in the Grift experiments and to develop statistical inference for particle flow from this kind of data. Specifically, the data are modelled by using simple linear Boolean models, and flow intensity is quantified by a single unknown rate parameter in these models. Maximum likelihood is reviewed in cases where particles require a fixed time for passage. An M-estimator of flow intensity based on a simple estimating equation is obtained in more general cases. Assessment of total particle flow utilizing this estimator is also developed.

Section 2 introduces the Boolean models and establishes notation and terminology. Results for the clump length distribution that was derived in Hall (1988) are used to develop M-estimation of flow intensity and the M-estimator is compared with maximum likelihood and other moment estimators by simulation. In Section 3, two estimators of total particle flow are proposed, including one which is obtained by derivation of the conditional expectation of the number of particles in a clump given clump length under the equal diameters model. Simulations are carried out to give an assessment of the performance of this estimator in the random passage times model. The methods are evaluated on the basis of their performance with the experimental data in Section 4. Section 5 concludes.

2. Estimation

To obtain a probability model for the clump length data, particles are assumed to be identically spherical with a known diameter d_0 and to arrive at the sensor according to a homogeneous Poisson process with unknown intensity λ. Passage of particles is assumed to continue unabated on arrival at the sensor. In one version of the model, the particles are travelling at a constant velocity, say v_0, and the *segment length* (Hall, 1988), or time required for any single particle to pass the sensor, is constant at $t_0 = d_0/v_0$. In a second version, velocities or diameters are assumed to vary in such a way that segment lengths behave as a random sample from a population with a known mean μ and an unknown variance σ^2. The two models will be referred to as deterministic segment length (DSL) or random segment length (RSL) models respectively.

Suppose that particle flow is observed for t time units. Let the number of particles arriving at the sensor in this time period be denoted by $A(t)$. Let $N(t)$ denote the number of complete particle clumps observed by time t. Let $Y_1, Y_2, \ldots, Y_{N(t)}$ denote the lengths of these clumps and $Z_1, Z_2, \ldots$ the spacings between them. Let the unobservable number of particles comprising clump i be called the clump *order* and be denoted by K_i.

Fig. 2 illustrates the clumping process by using an example with $A(t) = 7$ particles arriving at a sensor at times 1.9, 5.9, 6.8, 7.5, 11.6, 12.8 and 17.1 ms during an observation period of $t = 20$ ms. If particles are assumed to have diameter $d_0 = 4.45$ mm and to be travelling at a constant velocity of 2.225 mm ms^{-1}, the passage time required for each, or deterministic segment length, is $t_0 = 2$ ms, leading to four clumps of lengths $y_1 = 2$, $y_2 = 3.6$, $y_3 = 3.2$ and $y_4 = 2$ ms that exit the sensor at times 3.9, 9.5, 14.8 and 19.1 ms respectively. Spacings between clumps would be of length $z_1 = 1.9$, $z_2 = 2.0$, $z_3 = 2.1$ and $z_4 = 2.3$ ms and the four clump orders would be $k_1 = 1$, $k_2 = 3$, $k_3 = 2$ and $k_4 = 1$.

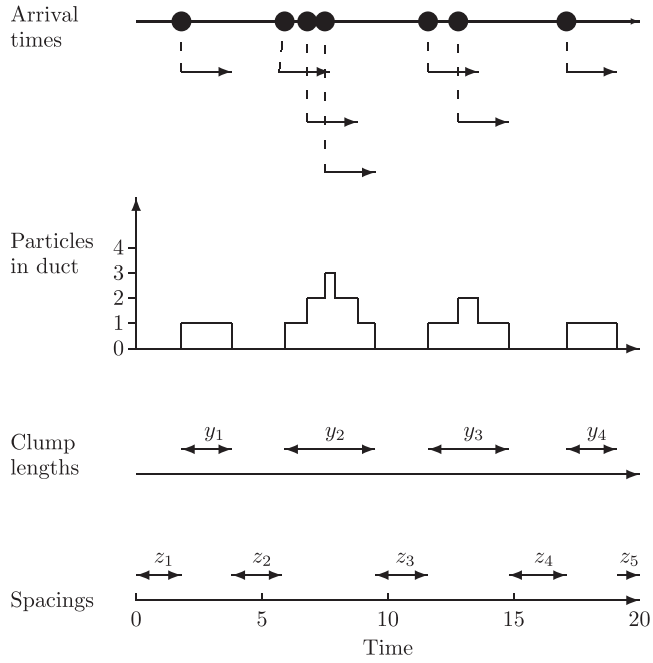

Fig. 2. Diagrammatic particle flow measurement by using a type II counter: the number of particles comprising each clump is apparent from the second set of axes, but only the segment lengths and spacings that are depicted in the bottom two sets of axes are observable

The particle clumps may be modelled by using a coverage process, or set process where the location of sets in space is driven by a random process. In this case, the sets are simply the passage times for individual particles, and the space where they occur is the timeline, with the locations driven by the Poisson process of particle arrivals. Hall (1988) described such a process as a simple linear Boolean model; simple because the clump lengths are line segments and linear because the events occur in one dimension, the timeline. The process is called Boolean because of the dichotomy of the channel being occupied or unoccupied; at any point in time, there is either a particle in passage at the location of the sensor array, or there is not. Linear Boolean models also arise as linear transects from higher dimensional convex grain Boolean models. In the language of queuing theory, the number of particles in a clump at the sensor at a given time forms an $M/D/\infty$ queue in the DSL model and an $M/G/\infty$ queue in the RSL model and clump lengths are called *busy periods*. There is much literature on these models from queuing theory (Daley, 2001). For statistical inference for the distribution of diameters or more complex quantities describing the grain process, or for Boolean models in higher dimensions, see Molchanov (1997). Handley (1999) derived a discrete approximation to the distribution of clump length in the linear Boolean model and used it for likelihood inference. Crespi *et al.* (2005) have employed the linear Boolean model for monitoring events of viral activity in human subjects.

2.1. Likelihood

Specification of the clump length density $f(y;\lambda)$ is difficult outside the case where particle diameters are deterministically constant. In the DSL model, Hall (1988) has shown that the density has point mass $\exp(-\lambda t_0)$ at $y = t_0$ and is otherwise given by

$$f(y;\lambda,t_0) = \lambda \frac{\exp(-\lambda t_0)}{1-\exp(-\lambda t_0)} \left(1 + \sum_{j=1}^{s-1} \frac{(-1)^j}{j!} [\lambda\{y-(j+1)t_0\}]^{j-1} \exp(-j\lambda t_0)[\lambda\{y-(j+1)t_0\}+j]\right) \quad (1)$$

where $y > t_0$ and s is the largest integer such that $t_0 < y/s$. The continuous part of the density is uniform over $(t_0, 2t_0)$, and decreasing for $y > 2t_0$. For small λ, $f(y;\lambda,t_0)$ can be approximated by the uniform distribution on $(t_0, 2t_0)$; for large λ it can be approximated by an exponential distribution. The fitted density $f(y;\lambda = 0.40, t_0 = 2.00$ ms) overlays the probability histogram of experimental clump lengths in Fig. 1.

For RSL models, likelihood inference is difficult because of the complexity of the clump length distributions (Handley, 2004). In the DSL model, an approximate likelihood function can be specified by ignoring the residual lifetime of the process. The residual lifetime is the duration of the last incomplete clump or spacing. A clump is a singleton if there are no arrivals within t_0 time units of the start of the clump, an event which occurs with probability $\exp(-\lambda t_0)$. Let $M_1 = \#\{y_i : y_i = t_0\}$ denote the number of singleton clumps. By independence of clump lengths, the approximate partial Boolean likelihood can be factored into components for singleton point masses and multiparticle clump length densities:

$$\tilde{\mathcal{L}}(\lambda; y_1, \ldots, y_{N(t)}) = \underbrace{\exp(-m_1\lambda t_0)}_{\text{singletons}} \underbrace{\prod_{i:y_i > t_0} f(y_i;\lambda,t_0)}_{\text{multiparticle lengths}}. \quad (2)$$

For cases where the spacings $\{z_i\}$ are available, an approximate complete Boolean likelihood may be obtained by multiplying the partial likelihood by the likelihood from an exponential random sample, $\lambda^N \exp(-\lambda \Sigma z_i)$.

For large t, the maximum likelihood estimator of λ based on $\tilde{\mathcal{L}}$ is approximately normally distributed. However, the analytic expression for the Fisher information is unwieldy, particularly for large clump lengths, where the degree of the polynomial components of the clump length density is high. Alternatively, approximate confidence regions can be constructed from the likelihood ratio test statistic, which has an approximate χ^2-distribution on 1 degree of freedom.

In the RSL model, where segment lengths are distributed as a random sample from a known distribution with distribution function $G(x;\theta)$, the clump length density and resulting likelihood are considerably more complex. Let $f_{\mathrm{RSL}}(y;\lambda,\theta)$ denote the clump length density, which depends on the unknown parameters, λ and θ. Ignoring the residual lifetime, the partial likelihood of the complete clumps is then

$$\tilde{\mathcal{L}}_{\mathrm{RSL}}(\lambda,\theta;y_1,\ldots,y_{N(t)}) = \lambda^N \exp(-\lambda \textstyle\sum z_i) \prod_{i=1}^{N(t)} f_{\mathrm{RSL}}(y_i;\lambda,\theta). \tag{3}$$

Stadje (1985) obtained the clump length distribution function $F_{\mathrm{RSL}}(y)$ by inversion of the Laplace transform of Y, but it is an infinite sum of self-convolutions of a function that may involve an integral with no analytic solution, making inference based on $\tilde{\mathcal{L}}_{\mathrm{RSL}}$ difficult.

2.2. *M-estimation*

An important issue in estimation of λ is robustness under model misspecification. Inspection of the clump lengths from the experimental data, such as the run that is depicted in Fig. 1, reveals that the number of clumps with lengths that are slightly in excess of t_0 is greater than expected, so the distribution between t_0 and $2t_0$ is not uniform. This can be caused by variability in diameter or velocity or by errors of measurement. A desirable property for any estimator is robustness to this departure from model assumptions.

For mean segment length μ, the mean clump length is given by

$$E(Y;\lambda) = \frac{\exp(\lambda\mu)-1}{\lambda} \tag{4}$$

in either the DSL or RSL model, regardless of the distribution of segment lengths (Hall, 1988). For known μ, consider the M-estimator $\tilde{\lambda}$ which satisfies

$$\bar{y} = \frac{\exp(\tilde{\lambda}\mu)-1}{\tilde{\lambda}}. \tag{5}$$

A solution exists by the mean value theorem with $E(Y;\lambda)$ increasing in λ. Though there is no analytic solution, the equation can be solved rapidly by using any root finding procedure, such as the `uniroot` function in the R statistical software package (R Development Core Team, 2007). A starting point that works in simulations is given by $\tilde{\lambda} = (\bar{y}-\mu)/2\mu^2$, which is the solution that is obtained by using a second-order expansion of $\exp(\tilde{\lambda}\mu)$ about 0. An interesting aspect of the sampling distribution of $\tilde{\lambda}$ is that it is negative whenever $\bar{y} < \mu$: an event whose probability is small as long as $\lambda\mu$ is not too small.

This estimating equation for λ can be written

$$\sum_i \psi(y_i,\lambda) = 0 \tag{6}$$

where $\psi(y,\lambda) = y - \lambda^{-1}\{\exp(\lambda\delta)-1\}$. Large sample theory for M-estimators (Stefanski and Boos, 2002) can be used for inference about λ. For a random sample of n clump lengths $y_1,\ldots,y_n$, the asymptotic distribution of $\tilde{\lambda}$ is given by

$$n^{1/2}(\tilde{\lambda}-\lambda) \xrightarrow{\mathcal{L}} N(0, C/B^2)$$

where B and C are functions of λ defined by

$$B(\lambda) = E\left\{-\frac{\partial}{\partial\lambda}\psi(Y_1, \lambda)\right\}, \tag{7}$$

$$C(\lambda) = E\{\psi^2(Y_1, \lambda)\}. \tag{8}$$

Since ψ is linear in Y, the expectation operations are straightforward:

$$B(\lambda) = \frac{\exp(\lambda\mu)(\lambda\mu-1)+1}{\lambda^2}, \tag{9}$$

$$C(\lambda) = \mathrm{var}(Y; \lambda). \tag{10}$$

The variance of Y depends on the distribution of segment lengths. In the DSL model with $t_0 = \mu$,

$$\mathrm{var}(Y) = \lambda^{-2}\{\exp(2\lambda\mu) - 2\lambda\mu\exp(\lambda\mu) - 1\}. \tag{11}$$

In the RSL model with segment lengths distributed according to the general distribution function $G(x)$, clump lengths have variance

$$\mathrm{var}(Y) = 2\lambda^{-1}\exp(\lambda\mu)\int_0^\infty \left(\exp\left[\lambda\int_t^\infty \{1-G(x)\}\,\mathrm{d}x\right] - 1\right)\mathrm{d}t - \lambda^{-2}\{\exp(\lambda\mu)-1\}^2, \tag{12}$$

which can be estimated using the sample variance of clump lengths, s_y^2. Estimators for the variance of $\tilde{\lambda}$ are then given by

$$\widehat{\mathrm{var}}(\tilde{\lambda}) = n^{-1}\frac{\tilde{\lambda}^2\{\exp(2\tilde{\lambda}\mu) - 2\tilde{\lambda}\mu\,\exp(\tilde{\lambda}\mu) - 1\}}{\{\exp(\tilde{\lambda}\mu)(\tilde{\lambda}\mu-1)+1\}^2} \tag{13}$$

in the DSL model and

$$\widehat{\mathrm{var}}_{\mathrm{G}}(\tilde{\lambda}) = n^{-1}\frac{\tilde{\lambda}^4 s_y^2}{\{\exp(\tilde{\lambda}\mu)(\tilde{\lambda}\mu-1)+1\}^2} \tag{14}$$

in either the DSL or RSL model. In large samples, approximate confidence intervals for λ can be constructed from these estimates along with the normal approximation for $\tilde{\lambda}$.

2.3. Other estimators

For the DSL model with common deterministic passage time t_0, other method-of-moments estimators can be constructed by using only the clump count ($N(t)$) and singleton count (M_1) statistics. The sequence of independent and identically distributed sums $\{Z_i + Y_i\}$ is a renewal process. Elementary renewal theory (Cox, 1962) yields that, as $t \to \infty$, the normalized clump count converges in distribution to that of a standard normal random variable,

$$\frac{N(t) - t/\mu_{\mathrm{R}}}{\sigma_{\mathrm{R}}\sqrt{(t/\mu_{\mathrm{R}}^3)}} \xrightarrow{\mathcal{L}} N(0,1)$$

where μ_{R} and σ_{R}^2 denote the mean and variance of a randomly sampled renewal period. In the DSL model with deterministic common passage time t_0,

$$\mu_R = \lambda^{-1} \exp(\lambda t_0), \tag{15}$$

$$\sigma_R^2 = \lambda^{-2}\{\exp(2\lambda t_0) - 2\lambda t_0 \exp(\lambda t_0)\}. \tag{16}$$

Moments for $N(t)$ are then

$$E\{N(t)\} \approx \lambda t \exp(-\lambda t_0), \tag{17}$$

$$\mathrm{var}\{N(t)\} \approx \lambda t\{\exp(-\lambda t_0) - 2\lambda t_0 \exp(-2\lambda t_0)\}. \tag{18}$$

The probability that a randomly selected clump is a singleton is $\exp(-\lambda t_0)$, so $E(M_1) = \lambda t \exp(-2\lambda t_0)$. A method-of-moments estimator based on the observed number of singletons is then

$$\tilde{\lambda}_S = -\frac{1}{t_0} \log\left\{\frac{M_1}{N(t)}\right\}. \tag{19}$$

Grift *et al.* (2001) and Grift (2002) based estimation of total mass flow on this estimator. Other estimators of λ can be constructed by consideration of *vacancy*, $V \approx \Sigma Z_i$, or total time that the sensor is unoccupied. Hall (1988) developed asymptotic theory for a number of vacancy-based estimators. Measurements of V were not available from the experiments that are discussed in Section 4, and vacancy-based estimators are not considered further.

2.4. Simulation

Simulations were undertaken to provide some information about the performance of these estimators, with three goals in particular: a comparison of the efficiency of the moment estimator $\tilde{\lambda}$ relative to the maximum likelihood estimator (MLE) under the DSL model, an investigation of the robustness of the MLE under the RSL model and a comparison of coverage probabilities of confidence intervals resulting from the two variance estimates of the asymptotically normal M-estimator $\tilde{\lambda}$. Particle arrivals were generated according to a Poisson process. Three cases with an increasing degree of clumping were simulated by using flow intensities of $\lambda = 0.1, 0.2, 0.3$. Two times were considered for the length of the total observation period: $t = 1000$ and $t = 10000$. Preliminary experiments with particles sufficiently far apart that there was no clumping indicated that measured passage times were normally distributed. So, passage times for individual particles were generated from a normal distribution with a mean of $\mu = 5$ with three different standard deviations: $\sigma = 0, 0.5, 1$. The first of these standard deviations leads to the DSL model; the others to RSL models. The approximate mean clump counts for the DSL model were $E(K) \approx 1.6, 2.7, 4.5$ for the three flow rates $\lambda = 0.1, 0.2, 0.3$ respectively. The simulation experiment then had a crossed $3 \times 2 \times 3$ design, with $n = 500$ independent data sets generated per combination of λ, t and σ. Normal plots and Kolmogorov–Smirnov statistics did not indicate any obvious non-normality for either the MLE or $\tilde{\lambda}$.

Table 1 summarizes the results of the simulation. The bias of the M-estimate relative to λ and the efficiency relative to the MLE are given in the middle section. Though the bias of the MLE that is formulated under the DSL model dissipates with increasing λ or t, it does not exhibit robustness to heterogeneous segment lengths, in the sense that it has larger variance than the M-estimate. Empirical coverage probabilities for 95% confidence intervals based on the likelihood ratio test statistic LRT and those of the form $\tilde{\lambda} \pm 1.96\,\mathrm{SE}$ where SE denotes the appropriate estimated asymptotic standard error from Section 2 are given in the right-hand section of Table 1. For the shorter simulations ($t = 1000$), there is a tendency for coverage prob-

Table 1. Simulation: relative efficiency and coverage probability of λ estimators

σ	t	λ	$\overline{N(t)}$	*Relative bias*	*Relative efficiency*	*Coverage probabilities*		
						LRT	$SE(\tilde{\lambda})$	$SE_G(\tilde{\lambda})$
0	1000	0.1	60.1	0.01	0.89	0.966	0.950	0.944
0	1000	0.2	73.2	−0.01	0.95	0.956	0.956	0.934
0	1000	0.3	66.7	−0.01	0.98	0.948	0.946	0.944
0	10000	0.1	605.8	0.00	0.83	0.940	0.942	0.942
0	10000	0.2	734.5	0.00	0.88	0.960	0.956	0.950
0	10000	0.3	669.6	0.00	0.97	0.942	0.938	0.942
0.5	1000	0.1	60.2	0.56	8.3	0.128	0.922	0.916
0.5	1000	0.2	73.0	0.15	2.2	0.768	0.910	0.902
0.5	1000	0.3	66.3	0.05	1.1	0.940	0.952	0.950
0.5	10000	0.1	606.1	0.57	82.4	0.000	0.942	0.946
0.5	10000	0.2	736.1	0.15	17.2	0.002	0.946	0.954
0.5	10000	0.3	670.1	0.05	3.3	0.646	0.938	0.940
1	1000	0.1	60.5	0.56	7.3	0.136	0.914	0.940
1	1000	0.2	73.5	0.13	1.9	0.798	0.922	0.920
1	1000	0.3	66.9	0.03	0.98	0.952	0.944	0.942
1	10000	0.1	605.3	0.56	71.9	0.000	0.922	0.952
1	10000	0.2	735.7	0.14	13.6	0.016	0.924	0.946
1	10000	0.3	668.6	0.04	3.0	0.694	0.944	0.954

abilities that are based on $\tilde{\lambda}$ to be low. For data sets with a larger number of clumps ($t = 10\,000$), the nominal coverages for intervals that are based on $\tilde{\lambda}$ are reached. With $n = 500$ simulations, the Monte Carlo standard error is such that any sample proportion that is less than 0.934 is significantly less than the nominal 0.95 with comparisonwise error rate 0.05. Additionally, the intervals around the M-estimate that use the standard error, SE_G, which is a function of the sample variance of the clump lengths, appear to do better for the RSL models with large $N(t)$, particularly for the noisy segment length $\sigma = 1$ case. The likelihood ratio interval gives coverages that are consistent with nominal levels in simulations with the DSL model but breaks down under the RSL model where the likelihood is misspecified. In summary, the recommendation that is based on these simulations is that the M-estimator is reasonably efficient under the DSL model and robust to the conditions of the RSL model. Confidence intervals that are based on the standard error SE_G meet nominal coverage probabilities in large samples under either model.

3. Estimation of total particle flow

In the case where either λ is known or variance in its estimation is negligible, the total particle flow may be estimated by $E\{A(t)\} = \lambda t$. When t is not available, another estimator can be formed by substitution of $t \approx \Sigma Y_i + \Sigma E(Z_i)$ into the expression, giving $E\{\widehat{A(t)}\} = \lambda \Sigma Y_i + N(t)$.

In the DSL model, clump orders $(K_1, K_2, \ldots)$ may be shown (Pippenger, 1998) to be geometrically distributed. A clump is of order 1 ($K_i = 1$) if there are no arrivals within t_0 time units of the start of the clump, which occurs with probability $\exp(-\lambda t_0)$. A clump is of order 2 if there is exactly 1 arrival within t_0 units and none in the next t_0 time units: an event which occurs with probability $\{1 - \exp(-\lambda)\}\exp(-\lambda)$ and so on. $K_1, K_2, \ldots$ are then independent geometric random variables with support on positive integers:

$$\Pr(K_i = k) = \{1 - \exp(-\lambda t_0)\}^{k-1} \exp(-\lambda t_0) \qquad \text{for } k = 1, 2, \ldots$$

with $E(K_i) = \exp(\lambda t_0)$ and $\mathrm{var}(K_i) = \exp(2\lambda t_0) - \exp(\lambda t_0)$. If the system is vacant when observation ends at time t, then the total particle flow may be expressed as the sum of these clump orders: $A(t) = K_1 + \ldots + K_{N(t)}$. If the system is occupied at time t, there is a partial clump that contributes a relatively small amount of particle flow for large t. Expressing the total particle flow $A(t)$ as the sum of clump orders each with mean $\exp(\lambda t_0)$ suggests the estimator $\hat{A}_1(t;\lambda) = N(t)\exp(\lambda t_0)$. When evaluated at the M-estimator $\tilde{\lambda}$, with mean passage time $\mu = t_0$, the two estimators of total particle flow become equivalent: $\tilde{A}_0(\tilde{\lambda}) = N(t)\exp(\tilde{\lambda}\mu) = \hat{A}_1(t;\tilde{\lambda})$.

More efficiency might be gained by conditioning on the clump lengths. The estimator $p(y)$ of an individual clump order which is a function of the clump length y and minimizes the mean-squared error $E[\{K - p(y)\}^2]$ is the *Bayes* estimate, or $p(y) = E(K|Y = y)$. An estimate of the mean total particle flow $E\{A(t)\} = E\{\Sigma K_i\}$ is then given by summing over clumps:

$$\hat{A}_{\mathrm{B}}(t;\lambda) = \sum_{i=1}^{N(t)} E(K_i|Y_i;\lambda). \tag{20}$$

Of course $E(K_i|Y_i = t_0;\lambda) = 1$. The approach that was used by Hall (1988) to derive the clump length density $f(y)$ in the DSL model may be extended to obtain the conditional mean of clump orders, $E(K|Y)$. Let the beginning of a clump be the origin and let k denote an integer greater than 1. The joint event $K = k$ and $Y \in (y, y + \mathrm{d}y)$ occurs if and only if there is a particle arrival at $(y - t_0, y - t_0 + \Delta y)$, no arrival in $(y - t_0 + \Delta y, y)$ and exactly $k - 2$ arrivals in $(0, y - t_0)$, and the nearest neighbour of each of these $k - 2$ arrival times is not further than t_0 time units away. Since the first three of these conditions are independent and the fourth is conditionally independent of the first two given the third, the joint probability of these four events is the product

$$\lambda\,\Delta y\,\exp(-\lambda t_0)\frac{\{\lambda(y - t_0)\}^{k-2}}{(k-2)!}\exp\{-\lambda(y - t_0)\}\;p_{k-2}\left(\frac{t_0}{y - t_0}\right)$$

where $p_n(u)$ denotes the chance that the largest division that is formed by a random sample of n points taken from the unit interval does not exceed u. This probability is given by

$$\begin{aligned} p_n(u) &= \sum_{j=0}^{[u^{-1}]} (-1)^j \binom{n+1}{j}(1 - ju)^n \\ &= 1 - (n+1)(1-u)^n + \binom{n+1}{2}(1-2u)^n - \ldots \end{aligned} \tag{21}$$

where $[\cdot]$ denotes the largest integer not exceeding the argument. Division by $f(y)$ and differentiation with respect to y yield the conditional density

$$\Pr(K = k|Y = y) = \frac{\lambda\exp(-\lambda y)}{f(y)}\frac{\{\lambda(y - t_0)\}^{k-2}}{(k-2)!}p_{k-2}\left(\frac{t_0}{y - t_0}\right). \tag{22}$$

If $s = [y/t_0]$, then summation over positive integers yields an exact expression for the conditional mean:

$$\begin{aligned} E(K|Y = y) &= \sum_{k=s+1}^{\infty} k\,\Pr(K = k|Y = y) \\ &= \frac{\lambda\exp(-\lambda t_0)}{f(y)}\sum_{k=s+1}^{\infty} k\frac{\{\lambda(y - t_0)\}^{k-2}}{(k-2)!}p_{k-2}\left(\frac{t_0}{y - t_0}\right) \end{aligned}$$

$$= \frac{\lambda \exp(-\lambda t_0)}{f(y)} \sum_{k=s+1}^{\infty} k \frac{\{\lambda(y-t_0)\}^{k-2}}{(k-2)!} \sum_{j=0}^{s-1} (-1)^j \binom{k-1}{j} \left(1 - \frac{jt_0}{y-t_0}\right)^{k-2}. \quad (23)$$

Inspection of $\Pr(K=k|Y=y;\lambda)$ reveals that, for $t_0 < y < 2t_0$, K has the translated Poisson distribution with mean and variance that are linear in y. For larger y, numerical evaluation of $E(K|Y=y)$ may be difficult. Inspection of plots for larger y and various values of λ indicates that, after a jump discontinuity of $\lambda t_0 \exp(-\lambda t_0)\{1-\exp(-\lambda t_0)\}^{-1}$ at $y=2t_0$, approximate linearity extends to $y > 2t_0$. For cases where $N(t)$ is large and there is heavy clumping, $E(K|Y=y)$ can be approximated by linear interpolation to save computational effort.

3.1. Simulation

The performances of these estimators of mean total particle flow are compared by using the simulated data from Section 2. Error for either $\hat{A}_1$ or $\hat{A}_B$, as a percentage of the mean particle flow, is assessed by using the relative root-mean-squared error RRMSE:

$$\mathrm{RRMSE}\{\hat{A}(t)\} = \frac{1}{\bar{A}(t)} \sqrt{[500^{-1} \sum_i \{\hat{A}_i(t) - A_i(t)\}^2]} \quad (24)$$

where i indexes the 500 simulated data sets, and $\bar{A}(t)$ denotes the average particle flow over these simulations. Table 2 summarizes the relative bias and RRMSE of estimates obtained by substitution of the M-estimates $\tilde{\lambda}$ into the expressions $\hat{A}_1(t) = N(t)\exp(\lambda t_0)$ and $\hat{A}_B(t;\lambda)$ for each simulated experimental condition. The estimation that is based on clumpwise estimated clump orders $\hat{A}_B$ is competitive under the DSL model ($\sigma_t = 0$) for smaller sample sizes ($t = 1000$). It suffers from some positive bias in RSL models that appears to decrease as the flow rate λ

Table 2. Error in estimation of the total particle flow from simulations

σ_t	t	λ	*Relative bias*		*RRMSE*	
			$\hat{A}_1$	$\hat{A}_B$	$\hat{A}_1$	$\hat{A}_B$
0	1000	0.1	−0.008	−0.007	0.042	0.033
0	1000	0.2	−0.012	−0.011	0.048	0.045
0	1000	0.3	−0.014	−0.013	0.049	0.048
0	10000	0.1	−0.001	−0.001	0.014	0.01
0	10000	0.2	0.000	0.000	0.015	0.014
0	10000	0.3	−0.002	−0.002	0.016	0.015
0.5	1000	0.1	−0.003	0.178	0.048	0.185
0.5	1000	0.2	−0.008	0.061	0.051	0.076
0.5	1000	0.3	−0.011	0.013	0.054	0.052
0.5	10000	0.1	0.000	0.184	0.014	0.184
0.5	10000	0.2	−0.001	0.067	0.015	0.068
0.5	10000	0.3	−0.003	0.022	0.015	0.026
1	1000	0.1	−0.002	0.189	0.058	0.198
1	1000	0.2	−0.010	0.063	0.053	0.079
1	1000	0.3	−0.011	0.018	0.052	0.052
1	10000	0.1	0.001	0.192	0.018	0.193
1	10000	0.2	0.000	0.072	0.017	0.074
1	10000	0.3	−0.001	0.026	0.016	0.031

increases, though it remains inferior to $\hat{A}_1$ despite smaller variance and higher correlation with $A(t)$. In the RSL model, many singleton clumps have clump lengths that are slightly in excess of the mean singleton passage time μ_t and so have estimated orders that are in excess of 1. This may lead to a positive bias for the clumpwise estimators which is particularly acute when support is high near $Y = t_0$. This theory is supported by the poor performance under light clumping, when $\lambda = 0.1$ and the density near $Y = t_0$ is highest among values of λ that were considered in the simulation.

In summary, for minimal relative error, these simulations suggest the use of the simple $\hat{A}_1(t)$ estimator, which is unbiased and involves less computation than the clumpwise estimator $\hat{A}_B(t)$. A slight loss of efficiency under the DSL model may be offset by the superior performance in the RSL model. Expressed relative to total particle flow, the root-mean-squared error was not larger than 5.8% in any of the conditions that were simulated here.

4. Experimental data

An optical sensor was used to measure clump lengths and clump velocities in experiments (Grift *et al.*, 2001; Grift, 2002) in which a known number of spherical particles were dropped through a device simulating an aerial spreader duct. Various quantities of several kinds of particles were dropped at several velocities. The data that are considered here include 10 runs with 4000 spherical steel air rifle pellets dropped from each of two heights and five runs with 2000 pellets dropped from a fixed height. The mean $\bar{y}$ and variance s_y^2 of physical lengths (in millimetres) appear in Table 3 along with other statistics from the experiments. Division by the mean velocity ($\bar{v} = 2.23$ mm ms^{-1}) was used to transform the measurements to the timeline (in milliseconds) to obtain Fig. 1. In general, the velocity was reasonably constant within a run of the experiment. Spacings between particle clumps were not available for this analysis.

In these experiments, the total particle flow is fixed and the total flow time varies with run and is not observed. The opposite is true for the application of mass flow measurement during aerial application of fertilizer particles. The theoretical results regarding inference for the random particle flow $A(t)$ for fixed t do not necessarily hold under the conditions of the experiment, where $A(t)$ is fixed and t varies and is not observed. However, Table 3 provides some indication that estimates for the total particle flow, $A(t)$, have good empirical performance when it is treated as random, at least in these experiments.

The observed value of the estimator $\hat{A}_1$ is given in the penultimate column of Table 3. It appears to perform reasonably well under these conditions. The average absolute error of $\hat{A}_1$ over runs 1–20 is 9 and the root-mean-squared error from 4000 is 35.5, which is 0.9% of the target. There is some evidence of positive bias in the high intensity runs 11–20. A two-sided t-test of the hypothesis that $E(\hat{A}_1) = 4000$ under the conditions of runs 11–20 yielded a p-value of 0.065 on 9 degrees of freedom.

Higher flow rates lead to more clumps per particle, fewer singletons and larger variance in estimation of clump order, either conditionally as in $\hat{A}_B$ or unconditionally as in $\hat{A}_1$. The standard deviations of $\hat{A}_1$ under the light (runs 1–10) and heavy (runs 11–20) clumping conditions with 4000 pellets were $s_l = 26.4$ and $s_h = 37.1$ respectively. The estimates $\hat{A}_B$, which are based on the DSL model, exhibit substantial positive bias, as they did for data that were simulated under the RSL model. The same is true for the MLE of λ.

The data were imperfect and some outlier removal was undertaken. For example, the counter returned several clumps with negative velocities or negative physical lengths, or sometimes both. This is most likely due to a problem with the sensor hardware. Particles that stay in the centre of the duct throughout passage across the sensor device cast a perfect hard shadow on the sensors

Table 3. Descriptive statistics and estimation from experiments with air rifle pellets

Run	N	$\bar{y}$	s_y^2	$\tilde{\lambda}$ *(SE)*	$\hat{A}_B$	$\hat{A}_1$	$\hat{A}_1^*$
1	2958	5.22	3.07	0.07 (0.003)	4921	4041	4038
2	2930	5.22	2.91	0.07 (0.003)	4946	3997	3992
3	2891	5.26	3.39	0.073 (0.003)	4874	4008	4004
4	2935	5.22	3	0.07 (0.003)	4944	4000	3995
5	2990	5.16	2.88	0.065 (0.003)	4941	3986	3980
6	2941	5.2	3	0.068 (0.003)	4883	3984	3980
7	2983	5.15	2.84	0.064 (0.003)	4900	3969	3958
8	2956	5.16	2.9	0.065 (0.003)	4846	3952	3946
9	2894	5.24	3.12	0.071 (0.003)	4831	3976	3969
10	2914	5.25	3.11	0.073 (0.003)	4931	4025	4024
11	1821	6.76	11.77	0.176 (0.005)	4299	3988	3982
12	1770	6.85	11.56	0.182 (0.005)	4303	3976	3963
13	1805	6.8	12.3	0.179 (0.005)	4321	4000	3988
14	1748	6.96	12.57	0.188 (0.005)	4333	4038	4027
15	1800	6.85	12.13	0.182 (0.005)	4340	4040	4027
16	1784	6.93	14.82	0.186 (0.005)	4403	4089	4040
17	1772	6.93	12.56	0.187 (0.005)	4341	4064	4058
18	1788	6.89	13.28	0.184 (0.005)	4346	4051	4020
19	1812	6.78	12.01	0.178 (0.005)	4316	3995	3983
20	1790	6.84	11.98	0.181 (0.005)	4330	4005	3996
21	746	7.54	17.1	0.219 (0.008)	2143	1981	1981
22	791	7.24	13.2	0.204 (0.007)	2141	1958	1957
23	777	7.46	13.15	0.215 (0.007)	2184	2024	2024
24	774	7.3	13.23	0.207 (0.007)	2102	1941	1941
25	745	7.57	13.39	0.221 (0.007)	2133	1989	1989

and do not present any problems for measurement. However, it is possible that a particle moves at such a sharp angle on passage through the sensor device as first to cast a soft (excessively early) shadow on the second array before it causes a soft (excessively late) shadow on the first sensor array, resulting in a negative value for the passage time Δt_f that was defined in Section 1. Additionally, each run contained a very small number of extremely short clumps, much less than the particle diameter. This could be due to the indirect passage described above or may be due to matter that is smaller than the particles of interest, but sufficiently large to cast a shadow on the sensor. The number of questionable negative or short clump measurements that were removed did not exceed 1% for any of the 25 runs Even on removal of these clumps from the data, the resulting estimates did not change by more than 0.5%. They appear as $\hat{A}_1^*$ in the last column of Table 3 and achieve better mean-squared error than the estimate that is based on data where these clumps have been removed.

To assess the goodness of fit of the linear Boolean models more generally, probability histograms of the clump length data were checked for agreement with the estimated density $f(y; \tilde{\lambda}, d)$. One such check appears in Fig. 1, which exhibits reasonable fit except for slightly lowered mass at the mean segment length, $\mu = 2$ ms and slightly more observations just above the mean segment length than expected under uniformity of this part of the density. All the other histograms exhibited the same three distinctive features of a spike near this fixed segment length, near uniformity between one and two of these lengths and a long right-hand tail. Quantile plots and Kolmogorov–Smirnov goodness-of-fit tests, for estimates in runs 1–10 or runs 11–20, do not indicate any non-normality in the distribution of N, $\tilde{\lambda}$ or $\hat{A}_1$.

5. Conclusion

Mass flow measurement presents a challenging problem for many agricultural applications. Existing measurement devices for monitoring mass flow frequently rely on periodic recalibration, as with the lever-operated gates that are used in agricultural aircraft. A recently developed sensory device to measure mass flow was introduced and described as a type II counter system. Two versions of a simple linear Boolean model were proposed for the particle clump flow data that were generated by the device. One version assumed deterministically equal passage times for all particles, whereas the other assumed these to be distributed about a known mean with unknown variance. M-estimation of the flow rate was developed. A less simplistic alternative, such as a Bayesian approach or working directly with some representation of the complex likelihood function that is specified under the DSL model, may lead to improvements in estimation, but the M-estimator was intuitively sensible and computationally feasible and was found to be robust to conditions where either particle velocity or diameters have substantial variability.

For the total mass flow, two estimators were developed. The first is the product of the number of clumps and the estimate of the mean number of particles per clump. The second more complex estimator is the clumpwise sum of conditional mean clump orders K, given individual clump lengths Y. In models where segment lengths were deterministically equal, the estimator utilizing the individual clump lengths had relative root-mean-squared error that did not exceed 5.0% and was always lower than that of the simpler estimator based only on the M-estimate of flow rate and the number of clumps. Under the most favourable conditions, with light particle flow and a long dispersal period, the relative error was as small as 1%. Whereas the Bayes estimator did well in data that were simulated from the DSL model, it was outperformed by the simpler estimator in simulations where the segment lengths vary according to a normal distribution and in the bench scale experiments where the total particle flow was known. The relative root-mean-square error when using the estimator that does not condition on individual clump lengths ranged between 1.4% and 5.1%. The relative root-mean-square errors for the experimental data were 0.6% and 1.1% in the low and high intensity runs with 4000 particles respectively and 1.8% in the runs with 2000 particles. Additionally, the effect of clumps with negative velocities on this estimator was minimal, suggesting robustness to the hardware issue.

The development of this sensoring technology together with statistical inference for flow rate data enable more accurate application of agricultural inputs such as fertilizer. In addition to improved efficiency, this has the potential to minimize the environmental effect of agricultural production. Beyond fertilizer, the approach could be used to develop granular flow sensors for other applications in agriculture, such as yields on small fruits and grains. In medicine, the approach could conceivably be used for counting blood cells in flux, though this application would require the development of hardware on a much smaller scale.

Acknowledgement

The authors thank Dr Cavell Brownie for feedback provided after a careful reading of the manuscript.

References

Casady, W., Downs, H. and Fishel, F. (1997) Calibrating granular pesticide applicators. *Technical Report G1273*. University of Missouri, Columbia. (Available from `http://extension.missour.edu/explore/agguides/agengin/g01273.htm`.)

Cox, D. (1962) *Renewal Theory*. London: Wiley.

Crespi, C., Cumberland, W. and Blower, S. (2005) A queueing model for chronic recurrent conditions under panel observation. *Biometrics*, **61**, 193–198.
Daley, D. (2001) The busy period of the $M/GI/\infty$ queue. *Queng Syst.*, **38**, 195–204.
Grift, T. (2002) Cluster criteria. *Resource: Engng Technol. Sustain. Wrld*, **9**, 9–10.
Grift, T. and Hofstee, J. (1997) Measurement of velocity and diameter of individual fertilizer particles by an optical method. *J. Agric. Engng Res.*, **66**, 235–238.
Grift, T., Walker, J. and Hofstee, J. (2001) Mass flow measurement of granular materials in aerial application—part 2: experimental model validation. *Trans. Am. Soc. Agric. Engng*, **44**, 27–34.
Hall, P. G. (1988) *Introduction to the Theory of Coverage Processes*. New York: Wiley.
Handley, J. (1999) Discrete approximation of the linear boolean model of heterogeneous materials. *Phys. Rev.* E, **60**, 6150–6152.
Handley, J. (2004) Computationally efficient approximate likelihood procedures for the boolean model. *Computnl Statist. Data Anal.*, **45**, 125–136.
Molchanov, I. (1997) *Statistics of the Boolean Model for Practitioners and Mathematicians*. New York: Wiley.
Pippenger, N. (1998) Random interval graphs. *Rand. Struct. Alg.*, **12**, 361–380.
Pyke, R. (1958) On renewal processes related to type-i and type-ii counter models. *Ann. Math. Statist.*, **29**, 737–754.
R Development Core Team (2007) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Stadje, W. (1985) The busy period of the queueing system $M/G/\infty$. *J. Appl. Probab.*, **22**, 697–704.
Stefanski, L. and Boos, D. (2002) The calculus of m-estimation. *Am. Statistn*, **56**, 29–38.