Using mathematical modelling to explore mRNA localisation

Jonathan Harrison, Richard Parton, Ilan Davis, Ruth Baker

University of Oxford

Abstract

Insert abstract information here

1 Introduction

The localization of mRNA is crucial in a variety of biological contexts for the targeting of proteins to their site of function. This method of gene expression is particularly relevant for polarized cells such as oocytes and early embryos. For example, the axes of *Drosophila Melongaster* are established through regulation of gradients in *bicoid* (*bcd*), *gurken* (*grk*), *oskar* (*osk*) and *nanos* (*nos*) mRNA (1). mRNA localization has been observed in a variety of species and cell types, including *Drosophila* and *Xenopus* oocytes, neurons, chicken fibroblasts, yeast and bacteria (2–6), thus demonstrating that this process is ubiquitous and not limited to large cells. *Drosophila* relies heavily on the asymmetric locatization of mRNAs to coordinate early development processes both spatially and temporally.

Recent advances in imaging techniques and image analysis technologies (7–9) have allowed an advancement of our understanding of the mechanisms of mRNA localization. The use of flourescence *in situ* hybridisation (FISH) enables single molecules of mRNA to be labelled with high levels of sensitivity and specificity. The development of the MS2-MCP system, which consists of a MS2 bacteriophage RNA stem loop bound by MS2 coat protein fusion to a flourescent protein (10), has provided great benefits for the imaging of live cells *in vivo*. The MS2-MCP system has been used successfully in the visualisation of *nos*, *grk*, *bcd* and *osk* mRNAs in Drosophila (11–14). Although there are limitations to these technologies dependent on cell type, they have permitted an improvement of our understanding of intracellular motility and structure. What has become clear is that a variety of different mechanisms are used to ensure localization of mRNAs.

Hint here at what is not yet fully understood.

1.1 Outline of mRNA localization process

Once mRNA has been transcribed from DNA in the nucleus, the process by which it reaches its final site of localization, where it is then translated into protein, can be divided into four main stages: particle formation; nuclear export; transport; and anchoring. We will address each of these processes in turn.

mRNA does not exist in isolation inside the cell. Instead the mRNA molecules bind to proteins to form particle complexes, which are also known as ribonuclearproteins (RNPs). These proteins perform a range of different functions, including translational regulation preventing translation while the mRNA is in transit and may determine the final destination of the particle (15). There is also evidence that RNPs are dynamically remodelled during the transport process, to support translational regulation (16).

After the mRNA has assembled into RNPs, it must diffuse through the crowded interchromatin spaces in the nucleus until it reaches a nuclear pore complex (NPC). The RNP is then exported out of the NPC into the cytoplasm via interactions with co-factors. Certain NPCs are more active than others at different times (16), but the overall direction of export out of the nucleus is unbiased (2).

The transport stage of mRNA localization occurs by different mechanisms for different mRNAs. The most common method observed is transport via molecular motors moving along the cytoskeleton (either microtubules or actin filaments). However, *nos* mRNA localizes using a diffusion and trapping technique (11) rather than by active transport. Active transport results in fast, directed motion with velocities of the order of $1\mu\text{ms}^{-1}$ (13, 14, 17), which is an order of magnitude faster than movement by free diffusion. Bidirectionality was observed by Vendra et al. (18) in the motion of RNP complexes on microtubules in *Drosophila* blastoderm embryos. It should be noted that, conventionally, individual types of molecular motors are thought to move unidirectionally, with Dynein directed towards the minus end of microtubules and Kinesin directed towards the plus end (19). The number of motors required for a given RNP complex can vary and may be governed by cis-acting localization elements in the mRNA (20). Possible explanations for the

0006-3495/08/09/2624/12 $2.00 doi:15

bidirectionality include: the bidirectional disorganized distribution of the microtubule network; a tug of war between different molecular motors moving in opposing directions on different microtubules; a tug of war between different motor species all bound to the same RNP complex; reversal of a molecular motor moving on a single microtubule, possibly due to regulation by microtubule associated proteins (MAPs) (21). The molecular motors must be correctly joined to their cargo and the motor cargo complex secured to the cytoskeleton. This is ensured by the linkers *Bicaudal D* (*BicD*) and *Egalitarian* (*Egl*) (10). Although the cytoskeleton was initially thought to be a stable static network, due to how it was analysed in fixed material, it has recently been revealed to be a dynamic network with a biased random orientation of microtubules (22). This underlying structure had been suggested by the work of Zimyanin et al. (14) who observed RNPs containing *osk* mRNA moving in a biased random walk.

Once the RNP complex has reached its destination, it must be maintained in position by some anchoring mechanism to keep the mRNA localized within the cell. *grk* mRNA is anchored by the molecular motor Dynein (23) in the *Drosophila* oocyte and *nos* is trapped by actin (11). Dynein was observed by Delanoue and Davis (23) to act as a static motor without requiring further energy in the form of ATP to function. An alternative hypothesis is that continuous active transport is required to ensure the localization of mRNA, since Weil et al. (13) found that Dynein motor activity is required to ensure *bcd* localization in the anterior cortex of the Drosophila embryo. The transport process may ensure that RNP complexes are kept near their targeted region of localization by directing them back if they move away.

1.2 Motivation

After transcription and transport out of the nucleus, many mRNAs are localized and translated in the distinct cytoplasmic domains where they function (10, 24). This process is highly regulated and involves active transport by intricate molecular motors. mRNA localization is crucial for the establishment of polarity of cells and the formation of the basic animal body plan during development (1). It may have importance for a wide range of functions involving memory and learning in the nervous system. However, we are a long way from a complete understanding of the mechanisms governing the mRNA cargo transport process.

In particular, the mRNA transcripts that set up the *Drosophila* body axes originate in the nurse cells (see figure 1) and are actively transported on specialized subpopulations of microtubules into the oocyte through the ring canals (25). We will focus on this maternal transfer of transcripts into the oocyte.

One of the challenges involved in live imaging of cells to examine this transport process is the need to scan a sample

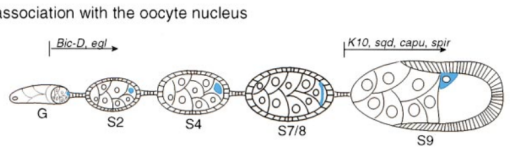


Figure 1: Stages of *Drosophila* development, showing nurse cells labelled as nc, follicle cells denoted by fc and the oocyte marked as oo. The localization of *grk* is shown in blue. Adapted from Lasko (26).

in three spatial dimensions as well as in time. This places restrictions on the resolution of data that can be obtained temporally and in the z direction (27). Although there are physical restrictions on what can be achieved by improvements in imaging technologies, these could be overcome by modelling of the transport of mRNA cargos, which may permit sparser sampling in time and finer sampling in space. This could be achieved by improvements in tracking of individual particles or by considering a more bulk flow approach. In this way, mathematical modelling could enable us to work with sparse data sets.

In this work, we present a simple model for transport of mRNA particles through nurse cells during localization and demonstrate how this model can be effectively parametrised. In Section 2, we construct the velocity jump process model and explain the approximate Bayesian computation approach taken in inference of the model parameters. We exhibit the posteriors over the model parameters generated via our inference approach for both *in silico* and *in vivo* data in Section 3. Finally, in Section 4, we discuss some of the implications of our modelling and parameter inference, before exploring possible future directions.

2 Materials and Methods

2.1 Modelling approach

In this context, modelling involves breaking down the mRNA localization process into its key components and identifying the biologically important aspects. We will put in place appropriate, simple modelling assumptions to describe the most important parts of the system biologically. These assumptions should as far as possible be supported by existing experimental observations.

Previous modelling approaches in this area have focused on descriptions of the mechanisms of molecular motors at smaller spatial and and temporal scales (28) or with continuous densities of mRNA (29) rather than representing single RNP particles which may be more appropriate at low densities. Other studies have generally considered mRNA dynamics in the oocyte in terms of cytoplasmic streaming or diffusion without direct reference to active transport (30**?**). Instead we shall describe the active transport of mRNA

in the nurse cells surrounding the ooctye, which are connected to each other and to the oocyte via ring canals, using an individual-based stochastic description.

We propose two different approaches to parameterise our model. The first will involve direct measurement of parameters, such as the average speed of cargo complexes in active transport. The second method requires application of an approximate Bayesian computation (ABC) framework (31–33) to parameterise the model via comparison with certain summary statistics of the data. This will enable comparison between the experimentally measured parameter values and the posterior for these parameters obtained via ABC.

We aim to produce a testable model in the sense that we are able to make predictions from the model of the result of varying certain parameters, such as the speed of the motor carrying the cargo, which could be tested experimentally using genetic mutants that have behaviour that corresponds to a different value of a given parameter. This should inform our understanding of how transcripts transport and localize.

2.2 Velocity jump process

The movement of RNP cargos is often described as a biased random walk (14). Rather than modelling this as a simple random walk in position, we let the direction of motion vary. This type of model is known as a velocity jump process and has been successfully applied to directed migration of animals and cells (34, 35).

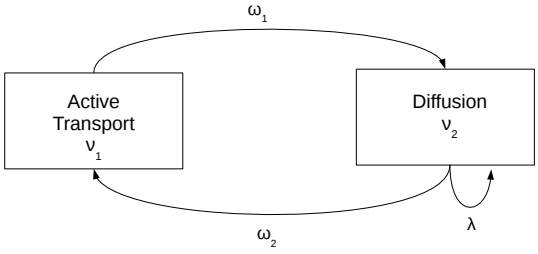


Figure 2: States and transitions between the states in the velocity jump process model

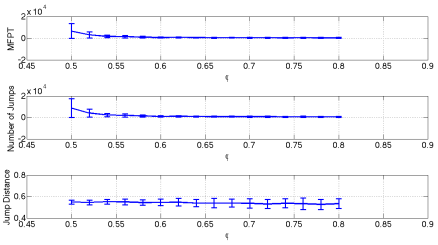
There are two phases of motion included in the model: an active transport phase when the motor is moving on the microtubules with constant speed ν_1 and a slower diffusive phase with constant speed ν_2. Switching occurs between these two phases of motion with exponential waiting times between events. Biologically this switching corresponds to the molecular motor falling off and reattaching to the microtubule. As illustrated in figure 2, switching occurs from active transport to diffusion with rate ω_1. From the diffusive phase, reorientations within the same phase occur with rate λ and switching to the active transport phase takes place at rate ω_2. After each switching event, the motor complex also changes direction by rotating to a new angle at random where the angle θ is drawn from some distribution $T(\theta)$. Here we chose to define $T(\theta)$ empirically as uniform on $[0, \pi]$ and $[\pi, 2\pi]$ with a bias in favour of $[0, \pi]$ based on biological data. We define a proportion of microtubules, ϕ, aligned in the posterior direction and thus take the following for $T(\theta)$:

$$T(\theta) = \begin{cases} \frac{\phi}{\pi} & : \theta \in [0, \pi] \\ \frac{1-\phi}{\pi} & : \theta \in [\pi, 2\pi). \end{cases}$$

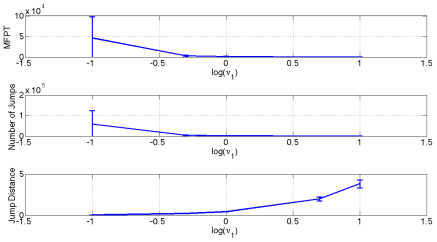
We model a two dimensional slice through a nurse cell as a rectangular region $[0, L_x] \times [-\frac{L_y}{2}, \frac{L_y}{2}]$. The cell nucleus is modelled as a circular region of given radius, R, placed centrally in the cell. This is excluding to the RNP complexes in the model which are reflected by it. Initially, cargoes are placed randomly at time $t = 0$ on the circumference of the nucleus to represent their export from nuclear pore complexes. In the model therefore, all mRNAs are produced simulataneously rather than modelling production over time. Cargoes move randomly as described until eventually they hit a samll target at the posterior end of nurse cell representing the ring canal joining to the oocyte, where they are absorbed and removed from the model. Throughout all simulations in this work, we fix $L_x = 52\mu$m, $L_y = 37\mu$m, $R = 10\mu$m and the size of the ring canal as 2μm.

We note that our model is dependent on 6 parameters: $\nu_1, \nu_2, \omega_1, \omega_2, \lambda, \phi$. We are able to explore the dependence of our model on these parameters by performing a sensitivity analysis, varying each paramters over several orders of magnitude. We show in figure 3 the effect of varying key parameters on the mean first passage time (MFPT), the mean over different particles of the number of jumps in one path up to absorption, and the mean over different particles of the median jump distance on a single path. The results of this demonstrate that the turning rate within diffusion, λ, has no effect on any of these summary statistics. Decreasing the speed parameters v_1, v_2 or the microtubule bias ϕ leads to a sharp rise in the MFPT. The transition rates ω_1 and ω_2 have opposite effects with an increase in ω_1 increasing the MFPT, while a decrease in ω_2 will increase the MFPT. Similar effects are seen on the mean number of jumps per simulation and the mean of the median jump distance.

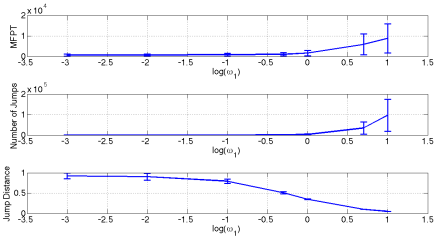
In addition, we exhibit the dependence of the spatial distribution of particles simulated from our model, as shown in figure 4. For increased motor speeds, ν_1 and ν_2, the particles spread out faster from the nucelus and are directed towards the posterior of the nurse cell. By increasing the microtubule bias, ϕ, the motion of the cargoes becomes much more



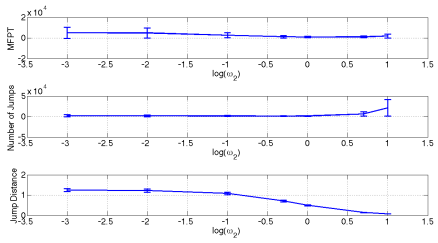
(a) Summary statistics as a function of ϕ



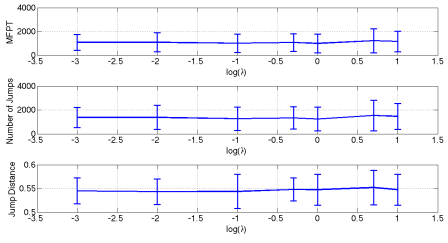
(b) Summary statistics as a function of ν_1 and ν_2



(c) Summary statistics as a function of ω_1

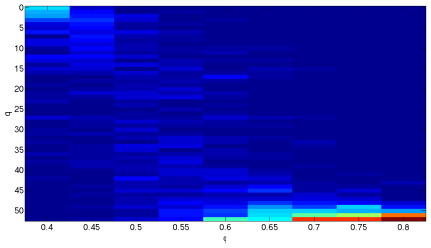


(d) Summary statistics as a function of ω_2

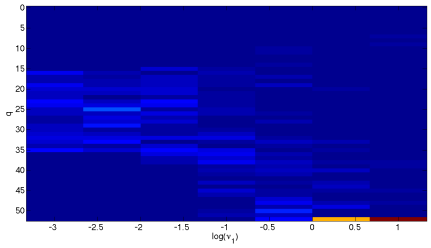


(e) Summary statistics as a function of λ

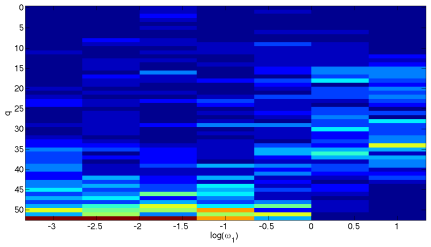
Figure 3: Three summary statistics (MFPT, number of jumps and jump distance) simulated for parameters $\phi = 0.58$, $\nu_1 = 1.16$, $\nu_2 = 0.8$, $\omega_1 = 0.42$, $\omega_2 = 0.84$, $\lambda = 0.11$. For each of 4a to 4e, we vary a parameter in turn. For 4b, $\nu_2 = \frac{1}{2}\nu_1$ while varying ν_1. Results are averaged over 100 particles with standard deviation shown on the error bars.



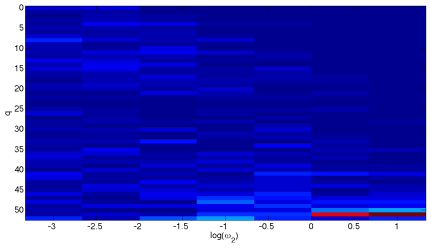
(a) Spatial distribution as a function of ϕ



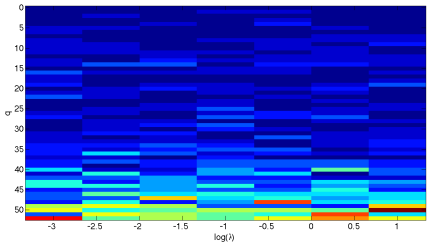
(b) Spatial distribution as a function of ν_1 and ν_2



(c) Spatial distribution as a function of ω_1



(d) Spatial distribution as a function of ω_2



(e) Spatial distribution as a function of λ

Figure 4: The spatial distribution of particles simulated for parameters as in figure 3. For each of 4a to 4e, we vary a parameter in turn. For 4b, $\nu_2 = \frac{1}{2}\nu_1$ while varying ν_1. Results are shown for 100 particles at time $t = 270s$. Note the final spatial box represents particles that have been absorbed and removed from the system.

directed towards to posterior with accumulation near the posterior and absoprtion from reaching the ring canal there. For high values of the switching rate ω_1, the distribution of particles is spread and has not reached the posterior, as little time will be spent in the active transport state. For low values of ω_1, the particles do not change direction often in the active transport state so greater accumulation is seen for intermediate values of ω_1. We note that for large values of ω_2, there is an accumulation of particles near the posterior, but that these have not been absorbed, since they spend little time in the diffusive phase of motion. The effect that the turning rate λ has on the spatial distribution appears to be unclear.

Based on direct measurements from manually tracked grk-GFP mRNA complexes in *Drosophila* nurse cells, we are able to make direct preliminary estimates of the parameters of our model. All data were obtained from Davidson (17). Particles are classed as static, paused or active. The number of total particles counted was $N = 340$ in a 40x40 μm area. The proportions of particles in each of the classes is 25% active, 50% paused and 21% static. Over the timescale observed, particles did not transition between states and these are proportions of particles. Particle speeds were assessed by observing a 10x10 μm area for 50 timepoints, with images taken in 1 z slice at 3 frames per second. Average speed in active transport ν_1 was $1.163 \pm 0.08 \mu\text{ms}^{-1}$ from $n = 33$. Average run length was $2.785 \pm 0.66 \mu$m. In the paused phase, an average speed, ν_2, of $0.798 \pm 0.6 \mu\text{ms}^{-1}$ was observed from 58 particles with an average run length of $0.84 \pm 0.06 \mu$m.

Considering the process as a continuous time Markov chain, with active transport and diffusive states, we obtain that the steady state of the continuous time markov chain is $[\frac{\omega_2}{\omega_1+\omega_2}, \frac{\omega_1}{\omega_1+\omega_2}]$. Then assuming ergodicity and neglecting static particles, this approach gives us $\frac{\omega_2}{\omega_1+\omega_2} = \frac{2}{3}$ and $\frac{\omega_1}{\omega_1+\omega_2} = \frac{1}{3}$ by comparing to the proportions of particles observed in each state. Therefore the rates of falling off and reattaching must be different.

Now assuming there are no internal transitions back to the active transport state, then we have $\omega_1 = 0.42 s^{-1}$ and hence $\omega_2 = 0.84 s^{-1}$. We can also deduce that the rate $\lambda = 0.95 - 0.84 = 0.11 s^{-1}$.

This gives parameter values to use of $\nu_1 = 1.16$, $\nu_2 = 0.80$, $\omega_1 = 0.42$, $\omega_2 = 0.84$, $\lambda = 0.11$.

Based on simulations from the model, we obtain a mean first passage time, averaged across 1000 particles, of $1070 \pm 840 s \approx 18$min. We plot the evolution of a typical path in figure 5. Given that stage 5 in *Drosophila* oocytes lasts 5 hours, this timescale for the passage time appears to be an order of magnitude shorter, suggesting that the limiting step in the transport and localization of the mRNAs is their production from the nucleus rather than the transport step. We are also able to make predictions on what would happen if we halved the speed of the molecular motors via a genetic

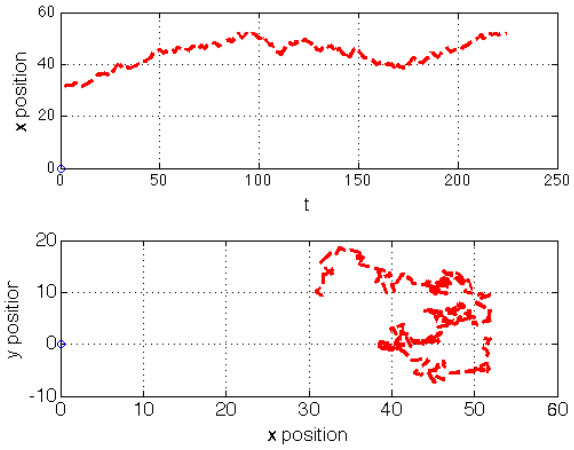


Figure 5: The evolution of a typical path until absorption at the posterior ring canal

mutation. We assume that altering the speed of the molecular motor, ν_1, would leave the speed in diffusive motion, ν_2, unchanged. The resulting distribution is shown in figure 6 and the MFPT would be $2150 \pm 1600 s \approx 36$min.

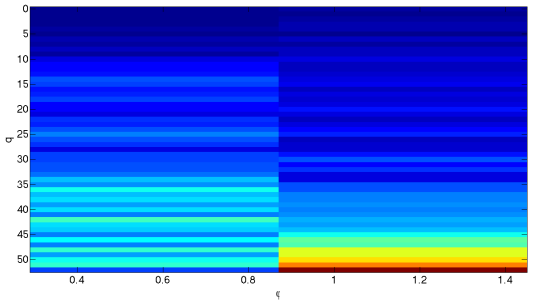


Figure 6: The spatial distribution resulting at time $t = 270 s$ from 1000 model simulations with parameters $\phi = 0.58$, $\nu_2 = 0.8$, $\omega_1 = 0.42$, $\omega_2 = 0.84$, $\lambda = 0.11$, varying ν_1 as $0.58 \mu\text{ms}^{-1}$ and $1.16 \mu\text{ms}^{-1}$. This represents the effect of introducing a genetic mutation to alter the speed of molecular motors.

2.2.1 Analytics

We may choose to include some analytics here at a later point in time.

2.3 Approximate Bayesian Computation (ABC)

In a statistical inference context, the likelihood of data given a certain set of parameters is a central quantity. In particular, it is essential in calculation of the posterior over the

parameters given certain data. The posterior is desirable as it offers more information than just point estimates of the parameters. Although for simple models it may be possible to evaluate the likelihood analytically, often for more complex models the likelihood is not tractable or is very expensive to compute. Approximate Bayesian Computation (ABC) techniques have been developed to address this issue (33). Instead of directly evaluating the likelihood, ABC techniques assume it is possible to cheaply simulate from the model and use this to approximate the likelihood.

In an ABC rejection sampling approach we follow the procedure shown in algorithm 1, where $\pi(\theta)$ is a prior on the parameters, ρ is a distance metric, $S(x)$ is a summary statistic used to summarize the data x and ϵ is a maximum tolerance for acceptance.

Algorithm 1 ABC Rejection Sampling

for $i = 1$ to n {
Sample parameters θ from a prior on those parameters $\pi(\theta)$
Simulate data x from the model $M(\theta)$ using those parameters
Calculate distance from observed data y.
if $\rho(S(x), S(y)) < \epsilon$ **then** accept the parameters θ
else reject the parameters θ.
}

One drawback of this algorithm is that it depends on appropriate choice of the distance metric ρ, the summary statistic $S(x)$ and the tolerance ϵ. Clearly the quality of the posterior will depend on the choice of these hyperparameters, as for example increasing ϵ will decrease the quality of the corresponding posterior. In some settings it may be possible to take the full data, rather than a summary statistic and to set $\epsilon = 0$ which would give an exact sample from the posterior, but in general this is not possible computationally. We can eliminate the importance of the tolerance to some extent by simulating N samples from the prior, storing all the distances and keeping the closest α quantile of the sampled parameters to the observed data. The quality of the posterior still depends on N and α but the choice of ϵ, which may be dependent on other model parameters, is removed.

The efficiency of the rejection sampling method is low, particularly for small values of ϵ which lead to tiny acceptance rates meaning many wasted samples. More efficient ways of sampling possible parameters θ have been suggested including Sequential Monte Carlo and Population Monte Carlo techniques $(36--39)$. We have opted to use an adaptive Population Monte Carlo (APMC) method presented in Lenormand et al. (39). This is implemented as shown in algorithm 2, where N is the number of parameter samples to simulate, $N_\alpha = \lfloor \alpha N \rfloor$ is the number to keep at each iteration, $p_{acc_{min}}$ is the minimum acceptance rate at which the algorithm stops, and $K_\sigma(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-y)^2}{2\sigma^2}}$ is a Gaussian kernel.

Algorithm 2 ABC Adaptive Population Monte Carlo

Initialise by setting N,α, N_α and $p_{acc_{min}}$.
Set $t = 1$.
for $i = 1$ to N {
Sample form prior: $\theta \sim \pi(\theta)$
Simulate data $x \sim f(x|\theta_i^{(0)})$
Calculate distance from observed data y via $\rho_i^{(0)} = \rho(S(x), S(y))$
Set $w_i^{(0)} = 1$
}
Take ϵ_1 as the first α-quantile of $\rho^{(0)}$.
Let $\{ (\theta_i^{(1)}, w_i^{(1)}, \rho_i^{(1)}) \} = \{ (\theta_i^{(0)}, w_i^{(0)}, \rho_i^{(0)}) | \rho_i^{(0)} \leq \epsilon_1, 1 \leq i \leq N \}$
Take σ_1^2 as twice the weighted empirical variance of $\{(\theta_i^{(1)}, w_i^(1))\}_{1 \leq i \leq N_\alpha}$
Set $p_{acc} = 1$
Update $t := t + 1$
while $p_{acc} > p_{acc_{min}}$ {
for $i = N_{\alpha+1}$ to N {
Pick θ_i^* from $\theta_j^{(t-1)}$ with probability $\frac{\omega_j^{(t-1)}}{\sum_{k=1}^{N_\alpha} w_k^{(t-1)}}$, $1 \leq j \leq N_\alpha$.
Generate new parameters $\theta_i^{(t-1)} | \theta_i^* \sim \mathcal{N}(\theta_i^*, \sigma^2_{(t-1)})$ and data $x \sim f(x|\theta_i^{(t-1)})$
Set$\rho_i^{(t-1)} = \rho(S(x), S(y))$
Set $w_i^{(t-1)} = \frac{\pi(\theta_i^{(t-1)})}{\sum_{j=1}^{N_\alpha} (w_j^{(t-1)} / \sum_{k=1}^{N_\alpha} w_k^{(t-1)}) K_{\sigma_{(t-1)}}(\theta_i^{(t-1)}, \theta_j^{(t-1)})}$
}
Set $p_{acc} = \frac{1}{N - N_\alpha} \sum_{k=N_\alpha+1}^{N} \mathbb{1}_{\rho_i^{(t-1)} < \epsilon_{t-1}}$
Take ϵ_1 as the first α-quantile of $\rho^{(t-1)}$.
Let $\{ (\theta_i^{(t)}, w_i^{(t)}, \rho_i^{(t)}) \} = \{ (\theta_i^{(t-1)}, w_i^{(t-1)}, \rho_i^{(t-1)}) | \rho_i^{(t-1)} \leq \epsilon_t, 1 \leq i \leq N \}$
Take σ_t^2 as twice the weighted empirical variance of $\{(\theta_i^{(t)}, w_i^(t))\}_{1 \leq i \leq N_\alpha}$
Set t=t+1
}

Inference on model parameters using ABC has been successfully applied to a variety of types of model in a biological context, often helping to offer biological insight into the model system. Cell migration models have used ABC to process data from scratch assays (31) and *in vivo* data (40), while individual-based ecological models with up to 14 parameters have also applied ABC for parameterisation (41).

When using ABC methods, it is important to be aware also of their limitations. It has been highlighted via Robert

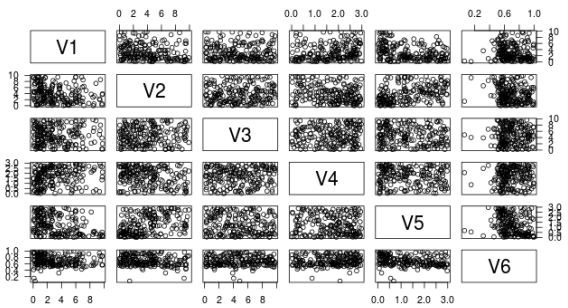
et al. (42) that although the error from the tolerance ϵ can be eliminated in the limit $\epsilon \to 0$, the use of summary statistics of the data may have larger effects, particularly when comparing models using ABC. Some more systematic methods for choice of summary statistics have been suggested, which use minimization of certain measures of entropy to determine which summary statistics are most informative (43).

2.3.1 Dependence of weights on distance

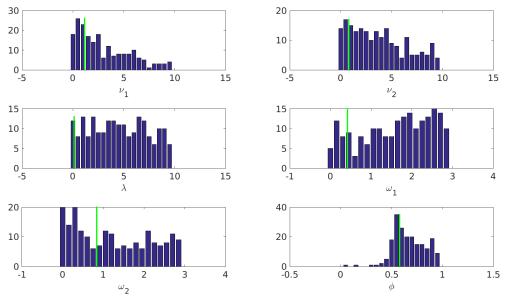
At each generation of the APMC algorithm, we calculate a distance for each set of parameters. These ought to be informative for which particles should contribute most heavily to the next generation, but are not used directly in the ABC-APMC algorithm. We have considered how we could use weights in the ABC-APMC algorithm that depend directly on the distances, using the update step:

$$w_i^{(t-1)} = \frac{\pi(\theta_i^{(t-1)})}{\rho_i^{(t-1)}} \frac{w_i^{(t-1)}}{\sum_{j=1}^{N_\alpha} w_j^{(t-1)}}. \quad (1)$$

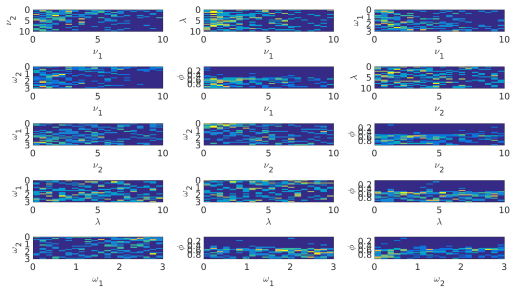
However, the weights used in the ABC-APMC algorithm are required to ensure that the importance sampling is performed correctly when sampling from the weighted distribution at the next time step (38). Without this exact choice, there is no guaruntee that the resulting posterior would be unbiased; a similar bias due to incorrect choice of weights was highlighted by Beaumont et al. (38) in the work of Sisson et al. (37). The appropriate weight for importance sampling to be effective uses the density $d_i^{(t)}$ from which parameter values are generated. Therefore the update step must be $w_i^{(t-1)} = \frac{\pi(\theta_i^{(t-1)})}{d_i^{(t-1)}}$, where $d_i^{(t)}$ is given by the probability to arrive at $\theta_i^{(t)}$ from one of the previous weighted parameter values. Although heuristically, weights directy dependent on distance seem an attractive option, statistically they may result in poor quality posteriors.



(a) Posterior obtained for $N = 2000$, $\alpha = 0.1$



(b) Posterior obtained for $N = 2000$, $\alpha = 0.1$



(c) Posterior obtained for $N = 2000$, $\alpha = 0.1$

Figure 7: Posterior for each parameter approximated via ABC rejection sampling, using $N = 2000$, $\alpha = 0.1$, $n_{repeats} = 1$. For 7a, a pairwise scatter plot is shown of all the posterior parameters. In 7b, histograms for each parameter a shown, with a green line for the real parameter values used for the simulated data. For 7c, heatmaps are shown pairwise for the posterior parameters.

3 Results and Discussion

3.1 Evaluation of ABC methods on in silico data

We demonstrate the effectiveness of the ABC inference techniques on in silico data, before applying them directly to real data sets. We present first the results of applying ABC methods to our model with a broad prior and known parameter values, as shown in figure 7.

For each experiment, we simulate a data set from chosen parameters and evaluate its summary statistic. That summary statistic is compared to summaries of data generated from proposed parameters, as described above. We average the results obtained across a number of experiments denoted $n_{repeats}$. As summary statistics we take the spatial distribution of the cargoes averaged in the x direction at given times $t_1, ..., t_\kappa$ and use a symmetric version of the Kullback-Leibler divergence as our distance to compare distributions. Uniform priors are used for each parameter between biologically realistic maximum and minimum values. In figure 7, scatterplots and heatmaps of the pairwise parameter distributions are shown, along with histograms for each parameter. In table (??), we present the results of varying the parameters $N, n_{repeats}, \alpha, p_{acc_{min}}$ on the quality of the posteriors.

We obtain most information from the posterior for the microtubule bias ϕ, while the speed parameters ν_1 and ν_2 are also identifiable. The remaining rate parameters have posteriors similar to the uniform prior and it is harder to recover the original parameter values used. This suggests that the microtubule bias and the speeds have most effect on the summary statistic chosen, which agrees with our earlier sensitivity analysis.

3.2 Application of ABC to Drosophila nurse cell data

Although comparative simulation results with ABC suggest that the detailed summary statistics using MFPT, number of jumps and jump distances result in a higher quality posterior, there are limitations to what data can be collected. These summary statistics would require detailed tracking data over an extended period of time, which would be difficult to obtain. Instead, we use spatial distribution of particles which is more tractable from single time points.

We observed grk mRNA nurse cells tagged using Dendra. These were imaged at 100x resolution. The age of embryos was determined by measuring oocyte size and comparing to a logistic growth model. Positions of mRNA particles were extracted via segmentation with Felzenswalb's algorithm (44) after denoising and deconvolution of the images. The centroids of appropriately sized segments were then taken to give the positions of particles at a time corresponding to the calculated age of that egg chamber.

Applying inference with ABC to these data, we obtain ...

4 Conclusion

Through use of a model based approach of mRNA transport and localization, we have been able to represent the dynamics of this process and reproduce behaviour observed experimentally. We constructed a velocity jump process with two phases of motion to represent motion in active transport and in diffusion. Since simulation from this model was accessible but the likelihood was intractable, we employed ABC methods to infer the parameters from our model from data collected via live *in vitro* imaging and compared the results of this with values obtained from direct measurement.

The timescales for mean first passage times based on simulations from the model are on the order of tens of minutes, rather than hours for the inferred parameters, suggesting that the transport step is not the rate limiting stage of the mRNA localization process. Based on the modelling work presented here, we postulate that the production of mRNA in the nucleus of the nurse cells is instead the rate limiting stage.

4.1 Further work

The subcellular environment is in some ways drastically different to that assumed in our spatially homogeneous two dimensional model. In reality, cells exist in three dimensions and are crowded inside with a hetrogeneous environment. One possible method of representing heterogeneity in the cell would be by use of a potential to account for volume exclusion effects, as used by Isaacson et al. (45) in the context of nuclear export of mRNAs. The three dimensional nature of the environment presents further challenges for tracking, as RNPs move between frames. Although some progress has already been made in this area (46), we hope that incorporation of modelling into a tracking framework could enable tracking of particles over longer timescales at increased resolution. Motivated by a Kalman filter type approach (47), we would like to investigate further using our model combined with regular inputs of data over time to assist particle tracking algorithms in the linkage step of tracking. Therefore in future the model should be extended to enable us to capture behaviour in three dimensions.

Finally, since we have hypothesised that production plays a critical role in this process, it would be informative to include production explicitly in our model. However, this may prove difficult to measure experimentally and to parametrise accurately.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

The authors would like to acknowedge valuable discussions from Dr S. Filippi and Dr J. Rittscher, in addition to beneficial input from Prof I. Davis.

References

1. Wolpert, L., R. Beddington, J. Brockes, T. Jessell, P. Lawrence, and E. Meyerowitz, 1998. Principles of Development. Oxford University Press, Oxford.
2. Wilkie, G. S., and I. Davis, 2001. Drosophila wingless and pair-rule transcripts localize apically by dynein-mediated transport of RNA particles. Cell 105:209–219.
3. Bobola, N., R.-P. Jansen, T. H. Shin, and K. Nasmyth, 1996. Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. Cell 84:699–709.
4. Mowry, K. L., and D. A. Melton, 1992. Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in Xenopus oocytes. Science 255:991–994.
5. Rosbash, M., and R. H. Singer, 1993. RNA travel: tracks from DNA to cytoplasm. Cell 75:399–401.
6. Nevo-Dinur, K., A. Nussbaum-Shochat, S. Ben-Yehuda, and O. Amster-Choder, 2011. Translation-independent localization of mRNA in E. coli. Science 331:1081–1084.

7. Jeffery, W. R., C. R. Tomlinson, and R. D. Brodeur, 1983. Localization of actin messenger RNA during early ascidian development. Developmental Biology 99:408–417.
8. Bertrand, E., P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long, 1998. Localization of ASH1 mRNA particles in living yeast. Molecular Cell 2:437–445.
9. Hamilton, R. S., R. M. Parton, R. A. Oliveira, G. Vendra, G. Ball, K. Nasmyth, and I. Davis, 2010. ParticleStats: open source software for the analysis of particle motility and cytoskeletal polarity. Nucleic Acids Research 38:W641–W646.
10. Parton, R. M., A. Davidson, I. Davis, and T. T. Weil, 2014. Subcellular mRNA localisation at a glance. Journal of Cell Science 127:2127–2133.
11. Forrest, K. M., and E. R. Gavis, 2003. Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. Current Biology 13:1159–1168.
12. Jaramillo, A. M., T. T. Weil, J. Goodhouse, E. R. Gavis, and T. Schupbach, 2008. The dynamics of fluorescently labeled endogenous gurken mRNA in Drosophila. Journal of Cell Science 121:887–894.
13. Weil, T. T., K. M. Forrest, and E. R. Gavis, 2006. Localization of bicoid mRNA in late oocytes is maintained by continual active transport. Developmental Cell 11:251–262.
14. Zimyanin, V. L., K. Belaya, J. Pecreaux, M. J. Gilchrist, A. Clark, I. Davis, and D. St Johnston, 2008. In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. Cell 134:843–853.
15. Hamilton, R. S., G. Ball, and I. Davis, 2013. A Multidisciplinary Approach to RNA Localisation. In Biophysical approaches to translational control of gene expression, Springer, 213–233.
16. Weil, T. T., R. M. Parton, B. Herpers, J. Soetaert, T. Veenendaal, D. Xanthakis, I. M. Dobbie, J. M. Halstead, R. Hayashi, C. Rabouille, et al., 2012. Drosophila patterning is established by differential association of mRNAs with P bodies. Nature Cell Biology 14:1305–1313.
17. Davidson, A., 2015. . Ph.D. thesis, University of Oxford.
18. Vendra, G., R. S. Hamilton, and I. Davis, 2007. Dynactin suppresses the retrograde movement of apically localized mRNA in Drosophila blastoderm embryos. RNA 13:1860–1867.
19. Howard, J., 2001. Mechanics of Motor Proteins and the Cytoskeleton. Sinauer, Sunderland, MA.
20. Amrute-Nayak, M., and S. L. Bullock, 2012. Single-molecule assays reveal that RNA localization signals regulate dynein-dynactin copy number on individual transcript cargoes. Nature Cell Biology 14:416–423.
21. Buxbaum, A. R., H. Gal, and S. R. H, 2015. In the right place at the right time: visualizing and understanding mRNA localization. Nature Reviews Molecular Cell Biology 16:95–109.
22. Parton, R. M., R. S. Hamilton, G. Ball, L. Yang, C. F. Cullen, W. Lu, H. Ohkura, and I. Davis, 2011. A PAR-1–dependent orientation gradient of dynamic microtubules directs posterior cargo transport in the Drosophila oocyte. The Journal of Cell Biology 194:121–135.
23. Delanoue, R., and I. Davis, 2005. Dynein anchors its mRNA cargo after apical transport in the Drosophila blastoderm embryo. Cell 122:97–106.
24. Jansen, R.-P., 2001. mRNA localization: message on the move. Nature Reviews Molecular Cell Biology 2:247–256.
25. Clark, A., C. Meignin, and I. Davis, 2007. A Dynein-dependent shortcut rapidly delivers axis determination transcripts into the Drosophila oocyte. Development 134:1955–1965.
26. Lasko, P., 1999. RNA sorting in Drosophila oocytes and embryos. The FASEB journal 13:421–433.
27. Weil, T. T., R. M. Parton, and I. Davis, 2010. Making the message clear: visualizing mRNA localization. Trends in Cell Biology 20:380–390.
28. Bressloff, P. C., and J. M. Newby, 2013. Stochastic models of intracellular transport. Reviews of Modern Physics 85:135.
29. Szymańska, Z., M. Parisot, and M. Lachowicz, 2014. Mathematical modeling of the intracellular protein dynamics: The importance of active transport along microtubules. Journal of Theoretical Biology 363:118–128.
. Ganguly, S., L. S. Williams, I. M. Palacios, and R. E. Goldstein, 2012. Cytoplasmic streaming in Drosophila oocytes varies with kinesin activity and correlates with the microtubule cytoskeleton architecture. Proceedings of the National Academy of Sciences 109:15109–15114.
30. Liu, W., and M. Niranjan, 2011. The role of regulated mRNA stability in establishing bicoid morphogen gradient in Drosophila embryonic development. PLOS ONE 6:e24896–e24896.
31. Johnston, S. T., M. J. Simpson, D. S. McElwain, B. J. Binder, and J. V. Ross, 2014. Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. Open Biology 4:140097.
32. Turner, B. M., and T. Van Zandt, 2012. A tutorial on approximate Bayesian computation. Journal of Mathematical Psychology 56:69–85.
33. Beaumont, M. A., W. Zhang, and D. J. Balding, 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.
34. Codling, E., and N. Hill, 2005. Calculating spatial statistics for velocity jump processes with experimentally observed reorientation parameters. Journal of Mathematical biology 51:527–556.
35. Taylor-King, J., E. van Loon, G. Rosser, and S. Chapman, 2015. From Birds to Bacteria: Generalised Velocity Jump Processes with Resting States. Bulletin of Mathematical Biology 1–24.
36. Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf, 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface 6:187–202.
37. Sisson, S. A., Y. Fan, and M. M. Tanaka, 2007. Sequential monte carlo without likelihoods. Proceedings of the National Academy of Sciences 104:1760–1765.
38. Beaumont, M. A., J.-M. Cornuet, J.-M. Marin, and C. P. Robert, 2009. Adaptive approximate Bayesian computation. Biometrika asp052.
39. Lenormand, M., F. Jabot, and G. Deffuant, 2013. Adaptive approximate Bayesian computation for complex models. Computational Statistics 28:2777–2796.
40. Liepe, J., H. Taylor, C. P. Barnes, M. Huvet, L. Bugeon, T. Thorne, J. R. Lamb, M. J. Dallman, and M. P. Stumpf, 2012. Calibrating spatio-temporal models of leukocyte dynamics

against in vivo live-imaging data using approximate Bayesian computation. Integrative Biology 4:335–345.

41. van der Vaart, E., M. A. Beaumont, A. S. Johnston, and R. M. Sibly, 2015. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. Ecological Modelling 312:182–190.
42. Robert, C. P., J.-M. Cornuet, J.-M. Marin, and N. S. Pillai, 2011. Lack of confidence in approximate Bayesian computation model choice. Proceedings of the National Academy of Sciences 108:15112–15117.
43. Nunes, M. A., and D. J. Balding, 2010. On optimal selection of summary statistics for approximate Bayesian computation. Statistical Applications in Genetics and Molecular Biology 9.
44. Felzenszwalb, P. F., and D. P. Huttenlocher, 2004. Efficient graph-based image segmentation. International Journal of Computer Vision 59:167–181.
45. Isaacson, S., D. McQueen, and C. S. Peskin, 2011. The influence of volume exclusion by chromatin on the time required to find specific DNA binding sites by diffusion. Proceedings of the National Academy of Sciences 108:3815–3820.
46. Thompson, M. A., J. M. Casolari, M. Badieirostami, P. O. Brown, and W. Moerner, 2010. Three-dimensional tracking of single mRNA particles in Saccharomyces cerevisiae using a double-helix point spread function. Proceedings of the National Academy of Sciences 107:17864–17871.
47. Faragher, R., et al., 2012. Understanding the basis of the Kalman filter via a simple and intuitive derivation. IEEE Signal Processing Magazine 29:128–132.

List of Figures

List of Tables