# sit-methods

Facundo Muñoz

11 April, 2024

## Contents

## 1 Introduction

First thing I want to do, is to have a simulator of releases of sterile mosquitoes and their interaction with wild individuals.

Doing a literature review about simulation of mosquito dynamics.

## 2 Simulation of SIT MRR trials

The R-package `nosoi` (Lequime et al. (2020)) might be of help. Although is designed for assessing epidemiological networks of contacts, and phylogenies, perhaps it could be *adapted* to our use case, where the transmissible pathogen is the sterility. However, I'm not sure whether we can easily represent concepts such as sterile or wild males, wild females, eggs and so on. In any event, the package itself can serve as an example of how to build and design an agent-based modelling framework in R.

Ocelet (Degenne et al. (2009)) can potentially be a suitable simulation platform for this use case, thanks to its agent-based spatially explicit approach. Still, it can be an overkill for a pilot study, since the learning curve is steep. It might be more convenient to start with a simple simulation approach in R and eventually evolve to something more powerful.

NetLogo (Wilensky (1999)) is another general purpose agent-based modelling platform. Looks powerful and stable, with many examples to take inspiration from.

Olivier Gimenez kindly shared[1] some code for simulation of Spatial Capture-Recapture (SCR) models. Based on Royle et al. (2014). The downside of this is that it does not take into account the dispersal dynamics.

Dufourd and Dumont (2012) propose a full simulator of dispersal dynamics of *Aedes albopictus* mosquitoes. But it contains many compartments for specific stages that I don't really need and that complicates the model greatly. Besides, the article is relatively old and I don't find the code available. Yet, it can also serve as inspiration.

Possibly the wisest solution is to develop our own simulator, tailored to our present needs for SIT studies. Stochastic diffusion processes can be simulated in R with the package `Sim.DiffProc` (Guidoum and Boukhetala (2020)). But for the simplest examples, I don't really need stochasticity.

Package `simecol` (http://simecol.r-forge.r-project.org/) (Petzoldt and Rinke (2007)) may possibly help. It provides a simulation framework for ecological models, and implements particle diffusion models.

Pilot functionality has worked successfully. See file `test_diffusion.R`. I need to integrate that into `sit` and make it generally available.

---

[1]https://oliviergimenez.github.io/basics_spatial_capturerecapture/

## 3 Design principles

- Integrated into `sit` in order to leverage classes and methods there.

- Simple models for dispersion, survival and trapping.

- Can be either individual-based or based on PDEs for modelling population behaviour, and coupled with some sampling variation based on probabilistic models (e.g. of death, capture, mating, etc.). The latter approach neglects variation in dispersion, which is more important for small-sized releases and populations. On the other hand, individual-based simulation is computationally intensive.

# 4 Likelihood of a finitely-observed dispersal model

Here we combine ideas from dispersal and time-to-event processes to derive a probabilistic model for the observations.

We know the number of individuals initially released, the release time, which can be set at $t = 0$ without loss of generality, and the location of the release-point, which can also be considered at $\mathbf{x} = (0, 0)$.

Individuals disperse and either die naturally in the wild or get captured in one of the $I$ traps at known locations. These are $1 + I$ *competing events* in the sense that they are mutually exclusive. Moreover, we only observe a limited period of time in which the traps are active, leading to censoring. Thus, the process can be framed as a time-to-event data under competing risks.

However, in contrast to standard survival models, we don't follow each individual separately. Instead, we only observe aggregated counts of individuals which got captured at the end of each observation period. Furthermore, we don't even observe the natural death process directly since we don't know at each time how many individuals remain alive. We only know that is an additional risk at play.

## 4.1 Dispersal model

We assume that in the absence of stimuli, mosquitoes disperse following a Brownian motion (Dumont et al. 2011), which means that they fly around randomly, independently from each other, changing directions continuously, with no preferential direction of flow. This is the classical model for the dynamics of gas molecules where changes in directions are due to random collisions between molecules.

The density $\rho$ of Brownian particles at point $\mathbf{x} \in \mathbb{R}^2$ and at time $t \in \mathbb{R}$ satisfies the *diffusion equation*

$$\frac{\partial \rho}{\partial t} = D\nabla^2\rho, \tag{1}$$

where $D$ is the *diffusivity* constant and represents the rate at which particles spread and $\nabla^2 = \nabla \cdot \nabla$ is the Laplace operator[2] (i.e., the divergence of the gradient of a scalar function).

Assuming that individuals are released from the origin of spatial coordinates at the initial time $t = 0$, the diffusion equation has the solution

$$\rho(\mathbf{x}, t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2}\frac{\|\mathbf{x}\|^2}{\sigma^2} \right\} = \frac{1}{\sqrt{8\pi Dt}} \exp\left\{ -\frac{\|\mathbf{x}\|^2}{8Dt} \right\} \tag{2}$$

---

[2]https://en.wikipedia.org/wiki/Laplace_operator

with $\sigma^2 = 2nDt$, and $n$ is the dimension of the space. This is a Gaussian density function centred at the release point and with a variance equal to $4Dt$ in $\mathbb{R}^2$.
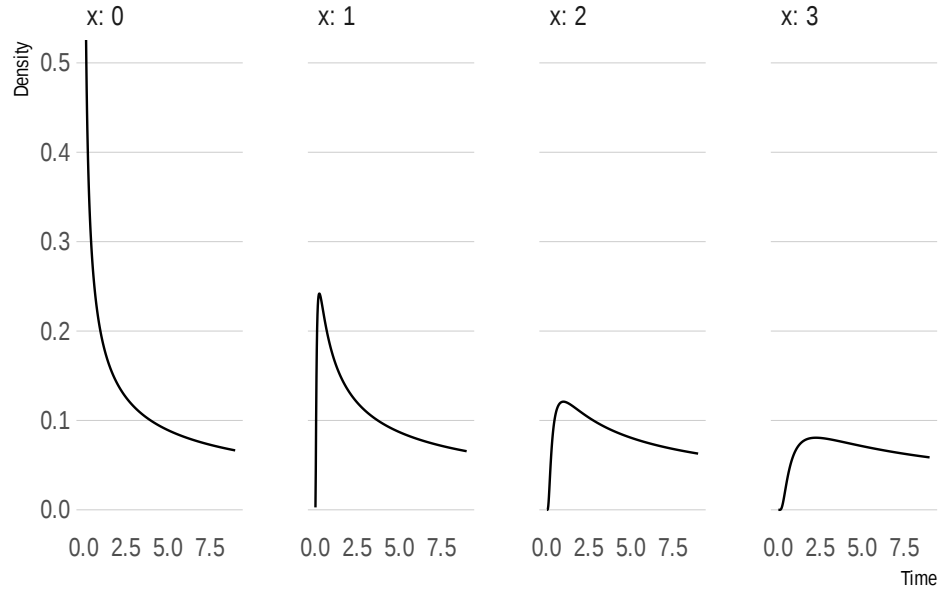


Figure 1: Evolution of the probability density over time at 4 different distances from the release point.

Working the Brownian motion with discrete time and space leads to a Random Walk, in which the walker's position after a large number of independent steps is distributed according to a Normal distribution of total variance $\sigma^2 = \varepsilon^2 \, t/\delta_t$, where $\varepsilon$ is the step size in space, $t$ is the total time elapsed and $\delta_t$ is the time step.

Thus, the diffusivity of the process in $\mathbb{R}^2$ is given by

$$D = \frac{\varepsilon^2}{4\delta_t}. \tag{3}$$

## 4.2 Survival model

We assume that the probability that an individual survives any given day is a constant that we call **Probability of Daily Survival** (PDS) $\pi_s$. As such, the probability of surviving up to day $t$ after release is

$$\pi_s^t = \exp\{t \log \pi_s\} = \exp\{-\beta t\}, \tag{4}$$

with $\beta = -\log \pi_s > 0$ and $t > 0$.

This formulation allows to work in continuous time, not only in discrete days.

From a survival perspective, let $T_0$ the time-to-natural-death in the absence of traps. Equation (4) gives the survival function $S(t) = P(T_0 > t)$ which decreases exponentially at an unknown constant rate. This corresponds to a constant hazard rate equal to:

$$h_0(t) = -S'(t)/S(t) = \beta.$$

Thus, while we don't observe any deaths directly, we observe captures in traps, which work as censoring for this time-to-event process. Indeed, if individuals were captured at some point, they must have been alive at that point.

However, we don't observe **all** individuals alive at some point (only the captured fraction). Meaning that we can't analyse this survival process independently from the capture process. The standard approach was to consider the fraction of captured individuals as a constant proportion of the total. But it is unclear whether this is a sensible or valid hypothesis.

### 4.3 Probability of capture

The dispersal process is only partially observed through the capture of a fraction of the individuals in a set of traps located at fixed positions in the area.

We assume that, in the absence of prior death or capture in the same or other traps, the probability of capture of a released individual is proportional to the density $\rho$ at the location $x_i$ of the trap $i$ at day $t$ and the survival probability.

> Why did I state the proportionality to the survival probability? Furthermore, shouldn't we take into account a range of influence of the trap, rather than only the density at the specific point? Or is this a acceptable approximation?

Let $T \in \{0, 1, \dots, I\}$ a categorical variable describing the fate of an individual, which can be either $T = 0$ if the individual is not captured during the experiment or $T = i$ if the individual is captured in trap $i$.

Let $X_i > 0 \,|\, T = i$ be the capture time of an individual, **conditional to capture on trap** $i$

$$\tilde{\pi}_i(t) = P(X_i = t \,|\, X_i \geq t, T = i) \propto \pi_s^t \, \rho(x_i, t) \tag{5}$$
$$= \alpha_i t^{-1/2} \exp\big\{-\beta t - \gamma_i/t\big\}, \quad t > 0 \tag{6}$$

with $\alpha_i = k_i/\sqrt{\pi 8 D} > 0$, $\beta = -\log \pi_s > 0$ and $\gamma_i = \|\mathbf{x}\|^2/8D > 0$.

The normalising constant $k_i$ ensures that the unconditional distribution sums to 1.

> Here I was thinking in terms of daily probabilities in order to avoid integrating continous times. But it may be better to think in terms of hazards.

Again, this is conditionally to the absence of prior capture in the same or other traps. The quantity and location of other traps, as well as the time elapsed, would necessarily influence this conditional capture probability.

The **unconditional** probability is then

$$\pi_i(t) = P(X_i = t) = P(X_i = t \mid X_i \geq t, T = i)(1 - P(\text{prior capture})) \tag{7}$$
$$= \tilde{\pi}_i(t)(1 - P(\text{prior capture})) \tag{8}$$
$$= \tag{9}$$

Essentially, **I think that this problem can be formulated in terms of a survival analysis with competing risks**.

Since the capture of an individual on different days are mutually exclusive events, the cumulative probability of capture $\pi_i^c(t) = \sum_{t=0}^{T} \pi_i(t)$ is the probability of capture at any time since release and up to time $t$.

$$\pi_i^c(t) = \alpha_i \sum_{x=0}^{t} x^{-1/2} \exp\left\{-\beta x - \gamma_i/x\right\}, \tag{10}$$

with $\alpha_i = k_i/\sqrt{\pi 8 D} > 0$ and $\gamma_i = \|\mathbf{x}\|^2/8D > 0$.

Note that in continuous time we would get an integral, which has a analytical indefinite result as:

$$\int x^{-1/2} \exp(-bx - c/x)\, \mathrm{d}x =$$
$$\frac{\sqrt{\pi} e^{-2\sqrt{b}\sqrt{c}}\left(-\operatorname{erf}(\frac{\sqrt{c}-\sqrt{b}x}{\sqrt{x}}) + e^{4\sqrt{b}\sqrt{c}}(\operatorname{erf}(\frac{\sqrt{b}x+\sqrt{c}}{\sqrt{x}}) - 1) + 1\right)}{2\sqrt{b}} + \text{constant}, \tag{11}$$

with $b > 0, c > 0$, where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2}\, \mathrm{d}x$, $x > 0$ is the *error function*[3].

---

[3]https://en.wikipedia.org/wiki/Error_function

This yields the following analytical formula for the cumulative capture rate for a trap $i$ in the absence of other captures:

$$\pi_i^c(t) = \frac{\alpha_i}{2}\sqrt{\frac{\pi}{\beta}}e^{-2\sqrt{\beta\gamma_i}}\left[-\operatorname{erf}(\frac{\sqrt{\gamma_i}-\sqrt{\beta}t}{\sqrt{t}}) + e^{4\sqrt{\beta\gamma_i}}\left(\operatorname{erf}(\frac{\sqrt{\gamma_i}+\sqrt{\beta}t}{\sqrt{t}})-1\right)+1\right],$$
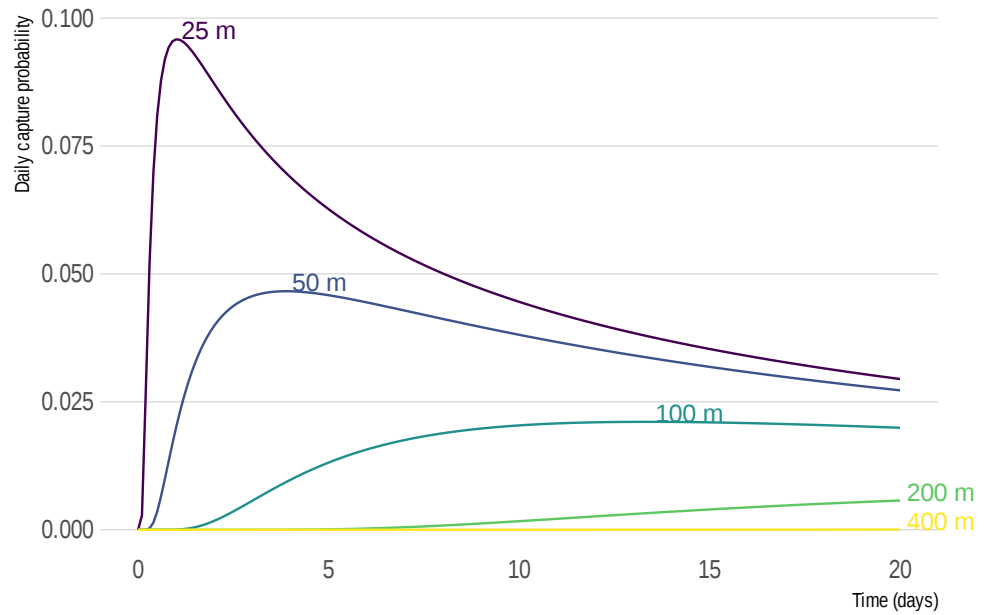
$$(12)$$



Figure 2: Daily probability of capture for each trap using the simulated parameters and uniform trap attractiveness. The scale parameter $k$ is arbitrary.

Figure 2 shows that each trap has a different capture rate with a maximum at a different moment in time, depending on the distance from the release point. The number, location and dates of captures will hopefully be informative enough to identify the diffusivity, survival and attractiveness parameters.

## 4.4 Observation model

We observe the numbers $Y$ of captured individuals in traps $i = 1, \ldots, I$ installed at survey days $j = 1, \ldots, J_i$. The survey periods for trap $i$ span from day $t_{ij}^0$ to $t_{ij}^1$.

Assuming that $N$ individuals were released at day $t = 0$, and that each of them has a cumulative capture probability on trap $i$ at survey $j$ conditional to the absence

of other captures described by (12), the number of captures $Y_{ij}$ follows a Binomial distribution with parameters $N$ and $\tilde{p}_{ij} = \pi_i^c(t_{ij}^1) - \pi_i^c(t_{ij}^0)$.

However, each of the $N$ individuals can either end up in only one of the surveys $i, j$ or not being captured in the course of its life with probability $p_0$. Thus, the outcome vector $(N - \sum Y_{ij}, Y_{ij})$ follows a Multinomial distribution with $N$ trials, $K = 1 + \sum_{i=1}^{I} J_i$ mutually exclusive events and probabilities $p_k$ where,

$$p_0 + \sum_{i=1}^{I} \sum_{j=1}^{J_i} p_{ij} = 1$$

This allows normalising the conditional probabilities $\tilde{p}$. I can do this across traps, but not over time, because time is not exchangeable.

## 5 TODO

- Integrate simulation functionality into `sit`.

- Use the simulator to check the assumption of capturing a constant fraction of the active individuals in a set of traps, in order to estimate survival.

- The realised density of the individuals at a given time progressively deviates from the expected Gaussian density, due to the capture process.
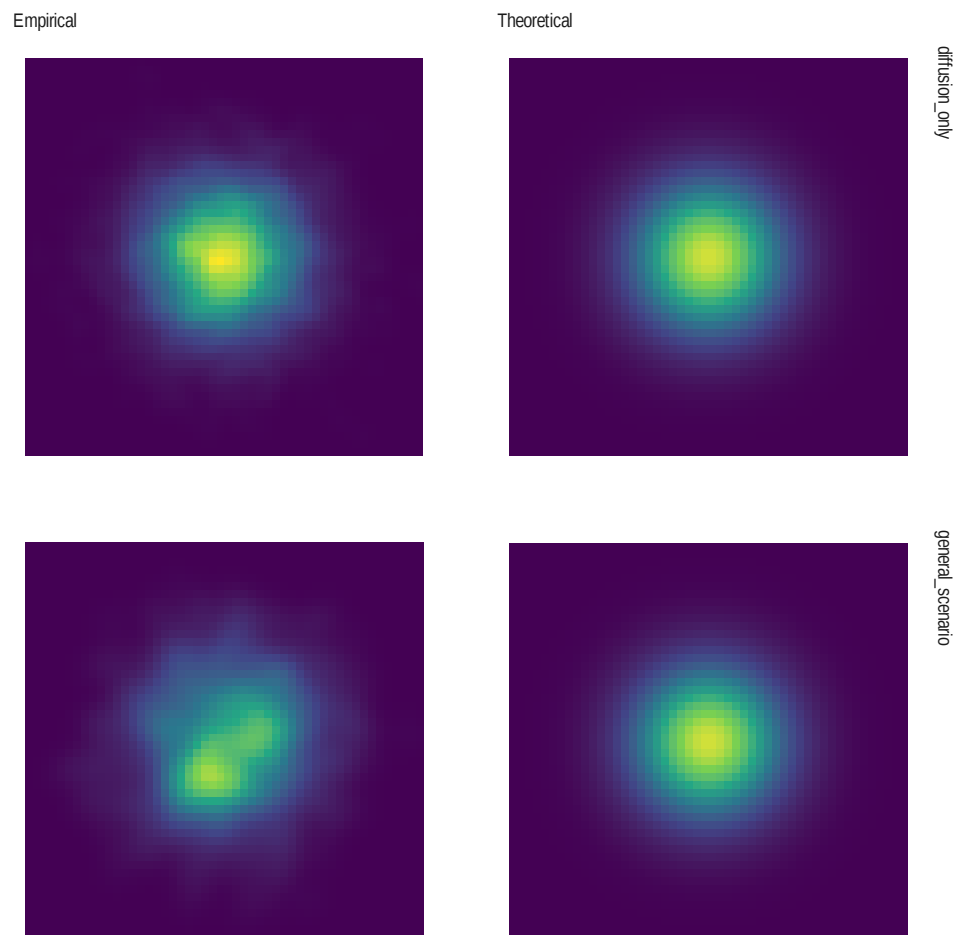


Figure 3: Density of individuals at day 10

To be more precise, we should the capture of individuals affects the diffusion process (the death process as well, but it is independent from the location and

time). Therefore, it needs to be accounted for in the diffusion equation, and the density of individuals derived as the solution to such modified equation.

The diffusion equation actually follows from the continuity equation[4], thus assuming no *sources* or *sinks*. However, traps act as skinks of individuals.

The continuity equation can be generalised to account for sinks[5]. And the corresponding diffusion equation should be solved to find the solution. Perhaps numerically.

# 6 Conclusions

- …

# References

Degenne, P., D. Lo Seen, D. Parigot, R. Forax, A. Tran, A. Ait Lahcen, O. Curé, and R. Jeansoulin. 2009. "Design of a Domain Specific Language for Modelling Processes in Landscapes." *Ecological Modelling* 220 (24): 3527–35. https://doi.org/10.1016/j.ecolmodel.2009.06.018.

Dufourd, Claire, and Yves Dumont. 2012. "Modeling and Simulations of Mosquito Dispersal. The Case of Aedes Albopictus." *BIOMATH* 1 (2). https://doi.org/10.11145/j.biomath.2012.09.262.

Dumont, Y., C. Dufourd, Michail D. Todorov, and Christo I. Christov. 2011. "Spatio-Temporal Modeling of Mosquito Distribution." In, 162–67. Albena, (Bulgaria). https://doi.org/10.1063/1.3659916.

Guidoum, Arsalane Chouaib, and Kamal Boukhetala. 2020. "Performing Parallel Monte Carlo and Moment Equations Methods for Itô and Stratonovich Stochastic Differential Systems: *R* Package **Sim.DiffProc**." *Journal of Statistical Software* 96 (2). https://doi.org/10.18637/jss.v096.i02.

Lequime, Sebastian, Paul Bastide, Simon Dellicour, Philippe Lemey, and Guy Baele. 2020. "Nosoi: A Stochastic Agent-Based Transmission Chain Simulation Framework in R." Preprint. Ecology. https://doi.org/10.1101/2020.03.03.973107.

Petzoldt, Thomas, and Karsten Rinke. 2007. "**Simecol** : An Object-Oriented Framework for Ecological Modeling in *R*." *Journal of Statistical Software* 22 (9). https://doi.org/10.18637/jss.v022.i09.

---

[4]https://en.wikipedia.org/wiki/Diffusion_equation#Derivation
[5]https://en.wikipedia.org/wiki/Continuity_equation#Differential_form

Royle, J. Andrew, Richard B. Chandler, Rahel Sollmann, and Beth Gardner, eds. 2014. *Spatial Capture-Recapture*. Amsterdam: Elsevier.

Wilensky, Uri. 1999. *NetLogo*. Northwestern University. Evanston, IL.: Center for Connected Learning; Computer-Based Modeling. http://ccl.northwestern.edu/netlogo/.