Competition and space structure in an efficient population genetics model of carcinogenesis

Tommaso Tarchi

September 29, 2023

University of Trieste

Introduction

Original paper

Title: Efficient simulation under a population genetics model of carcinogenesis

Authors: Zhu Tianqi, Hu Yucheng, Ma Zhi-Ming, Zhang De-Xing, Li Tiejun, Yang Ziheng

Year: 2011

Journal: Bioinformatics

Volume: 27

Number: 6

Pages: 837-843

Genotypic space

From the paper: ”It is now generally accepted that the rate-limiting step in the process of carcinogenesis is the accumulation of new mutations. [...] Here we develop a hybrid algorithm that can efficiently simulate the evolutionary processes of large populations.”

Genotypic space

From the paper: "It is now generally accepted that the rate-limiting step in the process of carcinogenesis is the accumulation of new mutations. [...] Here we develop a hybrid algorithm that can efficiently simulate the evolutionary processes of large populations."

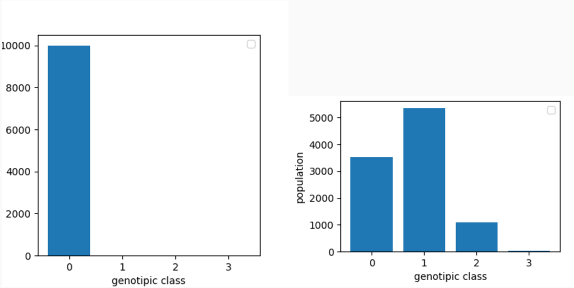


Figure 1: example of population distribution in the genotypic space at $t = 0$ (**left**) and t = waiting time (**right**); $m = 4$.

Fixed-population assumptions

- Fixed population.

Fixed-population assumptions

- Fixed population.
- Fitness landscape: $f_i = 1.01^i$.

Fixed-population assumptions

- Fixed population.
- Fitness landscape: $f_i = 1.01^i$.
- Constant and uniform mutation rate (μ).

Fixed-population assumptions

- Fixed population.
- Fitness landscape: $f_i = 1.01^i$.
- Constant and uniform mutation rate (μ).
- Possible events:
    1. a type j cell is **replaced** by a type j' one, with rate:

$$a_{jj'} = \frac{x_j f_{j'} x_{j'}}{\sum_{l=0}^{m-1} f_l x_l};$$

    2. a type j cell **mutates** into a type $j+1$ one, with rate:

$$a_j = \mu x_j.$$

Dynamic-population assumptions

- Variable population.

Dynamic-population assumptions

- Variable population.
- Fitness landscape: $f_i = 1.01^i$.

Dynamic-population assumptions

- Variable population.
- Fitness landscape: $f_i = 1.01^i$.
- Constant and uniform mutation rate (μ).

Dynamic-population assumptions

- Variable population.
- Fitness landscape: $f_i = 1.01^i$.
- Constant and uniform mutation rate (μ).
- Possible events:
  1. a type j cell is **born**, with rate:
  $$a_j^b = \frac{\tilde{N}(t) f_j x_j}{\sum_{l=0}^{m-1} f_l x_l},$$
  where $\tilde{N}(t)$ is a parameter capable of controlling the population size;
  2. a type j cell **dies**, with rate:
  $$a_j^d = x_j;$$
  3. a type j cell **mutates** into a type $j+1$ one, with rate:
  $$a_j^m = \mu x_j.$$

Exact algorithm

Pseudocode:

1. Compute events rates.
2. Apply Gillespie algorithm to sample time to next event e and next event.
3. Update time to $t' = t + e$ and system's state.
4. If population of the m-th genotypic class is different from zero, set waiting time to t' and stop, otherwise go back to step 1.

Hybrid algorithm

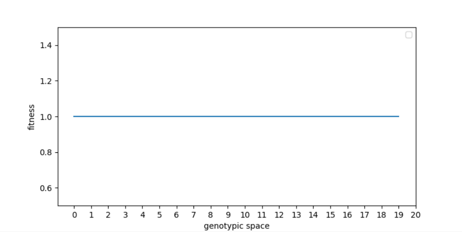
Pseudocode:

1. Compute events rates.
2. Partition events in **critical** and **non-critical** ones.
3. Determine tau-leaping step length τ.
4. Determine (by Gillespie's method) time to next critical event e.
5. If $e < \tau$, update system's state according to Gillespie algorithm, otherwise go directly to next step.
6. Let $h = min\,(e, \tau)$; update system's state applying tau-leaping algorithm over time h.
7. Update time to $t' = t + h$.
8. If population of the m-th genotypic class is different from zero, set waiting time to t' and stop, otherwise go back to step 1.

My humble contribute

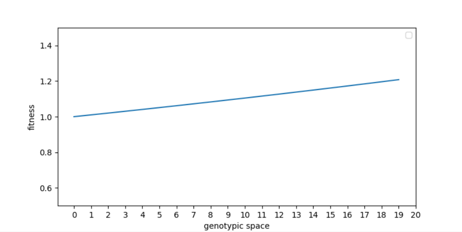
Fitness landscapes

1. Flat: $f_i = 1$



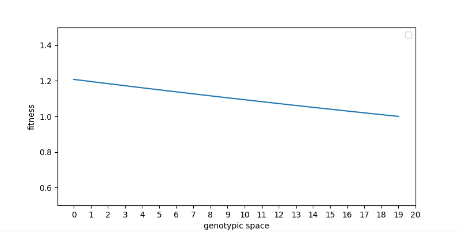
Fitness landscapes

1. Flat: $f_i = 1$
2. Static increasing: $f_i = 1.01^i$



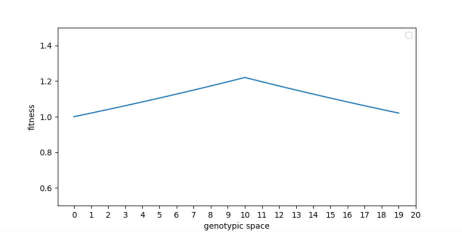
Fitness landscapes

1. Flat: $f_i = 1$
2. Static increasing: $f_i = 1.01^i$
3. Static decreasing: $f_i = 1.01^{m-i-1}$



Fitness landscapes

1. Flat: $f_i = 1$
2. Static increasing: $f_i = 1.01^i$
3. Static decreasing: $f_i = 1.01^{m-i-1}$
4. Static 'mountain': $f_i = 1.01^{m-|m-2i|}$



Modelling competition

Two **assumptions**:

- the closer two organisms in the phenotypic space, the stronger the competition for resources
- the genotype is the only factor determining the phenotype

Modelling competition

Two **assumptions**:

- the closer two organisms in the phenotypic space, the stronger the competition for resources
- the genotype is the only factor determining the phenotype

Result:

The more 'crowded' a portion of the genotypic space, the lower the fitness of the organisms in that region.

Fitness landscapes

1. Flat: $f_i = 1$
2. Static increasing: $f_i = 1.01^i$
3. Static decreasing: $f_i = 1.01^{m-i-1}$
4. Static 'mountain': $f_i = 1.01^{m-|m-2i|}$
5. static decreasing + dynamic:
   $f_i = 1.01^{m-i-1} + 0.1 - 0.01 \sum_{j=0}^{m} \frac{x_j}{N} e^{-|i-j|}$
6. Static 'mountain' + dynamic:
   $f_i = 1.01^{m-|m-2i|} + 0.1 - 0.01 \sum_{j=0}^{m} \frac{x_j}{N} e^{-|i-j|}$

Spatial structure (naive)

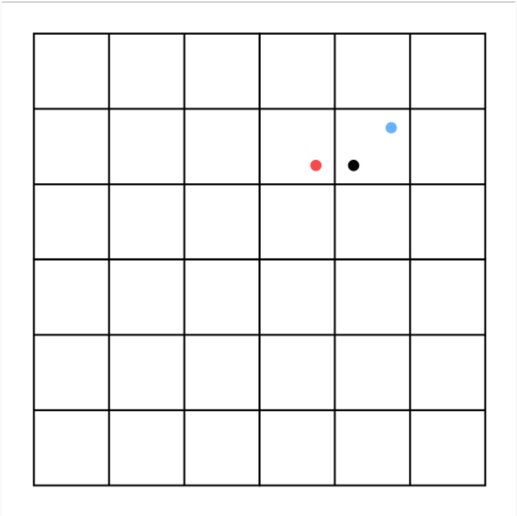


Figure 2: 'world' of the spatial model with resolution=16.

Spatial structure (less naive)

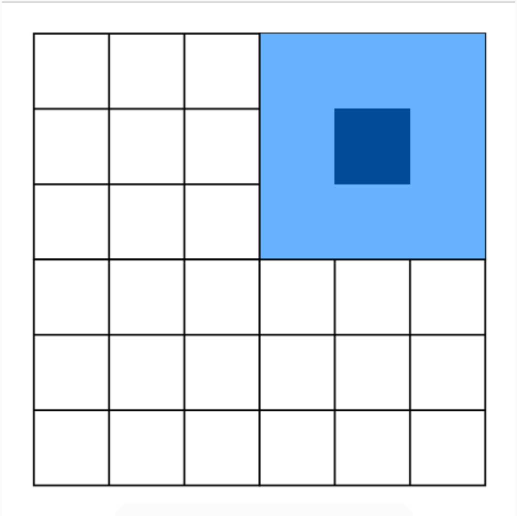


Figure 3: 'world' of the spatial model with resolution=16.

Spatial structure (less naive)

Three possible events:

1. a type j cell is born, with rate:
$$a_j^b = \frac{\tilde{N}(t)}{res} \frac{f_j x_j'}{\sum_{l=0}^{m-1} f_l x_l'};$$
2. a type j cell dies, with rate:
$$a_j^d = \frac{N}{N + \frac{1}{4}\sum_{\{ngb\}} N_k} x_j';$$
3. a type j cell mutates into a type $j+1$ one, with rate:
$$a_j^m = \frac{N}{N + \frac{1}{4}\sum_{\{ngb\}} N_k} \mu x_j'.$$

Where *res* is the resolution, $\#ngb$ is the number of neighbour areas and
$$x_i' = x_i + \frac{1}{4} \sum_{\{ngb\}} x_k$$

Spatial structure (less naive)

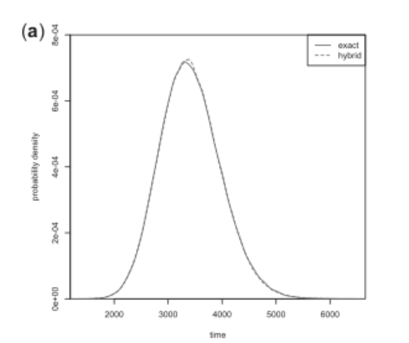
Time to next event was **heuristically** fixed by taking at each iteration:

$$\tilde{e} = \frac{\min\left(\{e\}_{areas}\right) + \max\left(\{e\}_{areas}\right)}{2},$$

where $\{e\}_{areas}$ is the set of times to next critical event computed on each area.

Simulation results

Reproducing results of the original paper



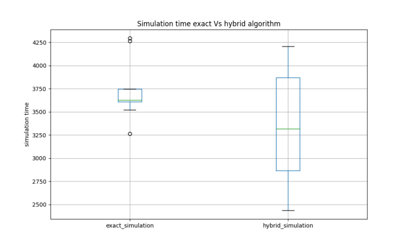


Figure 4: comparison of waiting time obtained with fixed-population exact and hybrid algorithms in original paper (**left**) and our work (**right**); parameters: $N = 10^3$, $m = 20$, $\mu = 10^{-3}$, fitness: static increasing.

Reproducing results of the original paper

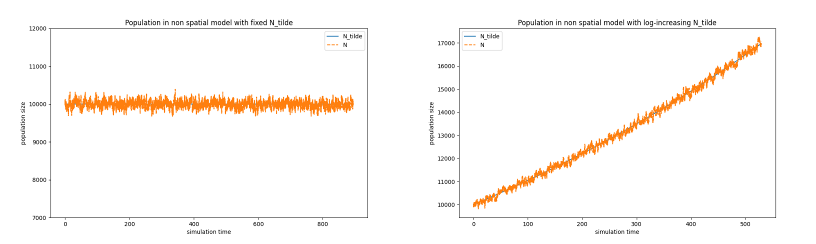


Figure 5: behaviour of N and N_tilde in time in our dynamic-population model with constant (**left**) and increasing (**right**) $\tilde{N}$; parameters: $N = 10^4$, $m = 4$, $\mu = 10^{-4}$, fitness: static increasing.

Changing the fitness landscape

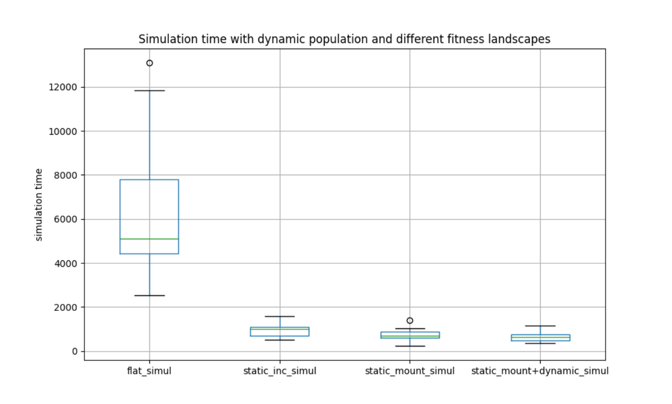


Figure 6: waiting time obtained with fixed-population hybrid model for different fitness landscapes; parameters: $N = 10^4$, $m = 4$, $\mu = 10^{-4}$; **note**: *static decreasing* never converged and *static decreasing + dynamic* almost never converged.

Changing the fitness landscape

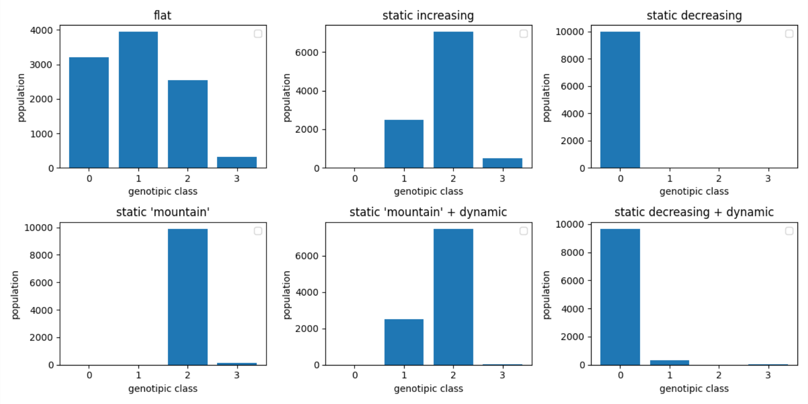


Figure 7: final state obtained with fixed-population hybrid model for different fitness landscapes; parameters: $N = 10^4$, $m = 4$, $\mu = 10^{-4}$.

Introducing space

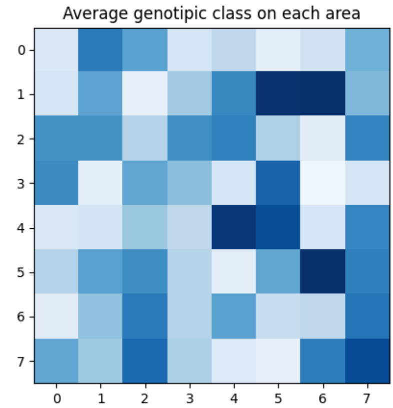


Figure 8: final distribution of the average genotypic class in space obtained with spatial model; parameters: $\tilde{N} = 10^4$ (constant), $m = 4$, resolution $= 64$, $\mu = 10^{-4}$, fitness: static increasing.

Waiting time in spatial model

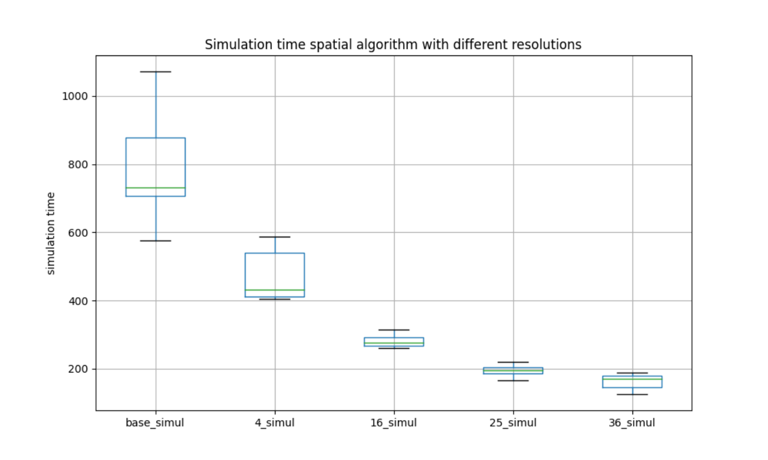


Figure 9: waiting time obtained with spatial model for different resolutions (from 1 to 36); parameters: $\tilde{N} = 10^4$ (constant), $m = 4$, $\mu = 10^{-4}$, fitness: static increasing.

Population in spatial model

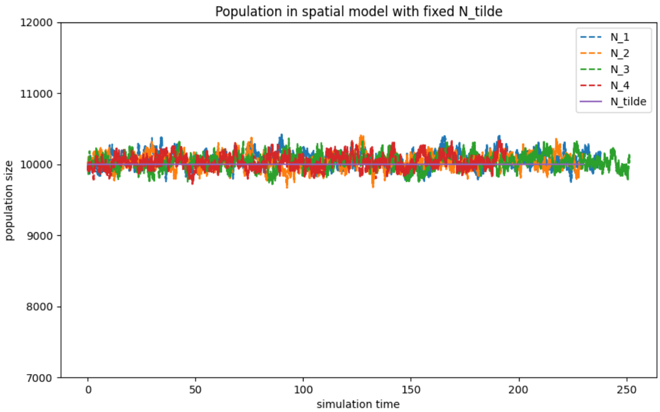


Figure 10: *N*'s behaviour in time for four simulations of the same spatial model; parameters: $\tilde{N} = 10^4$, $m = 4$, $\mu = 10^{-4}$, resolution $= 16$, fitness: static increasing.

An efficiency study

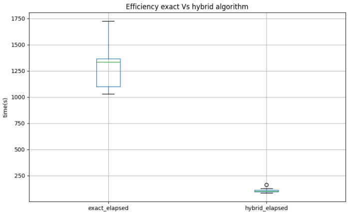


Figure 11: efficiency (in seconds) for fixed-population exact and hybrid models; parameters: $N = 10^3$, $m = 20$, $\mu = 10^{-3}$, fitness: static increasing.

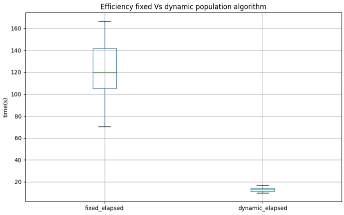


Figure 12: efficiency (in seconds) for fixed-population and dynamic-population hybrid models; parameters: $N\left(\tilde{N}\right) = 10^3$, $m = 20$, $\mu = 10^{-3}$, fitness: static increasing.

An efficiency study

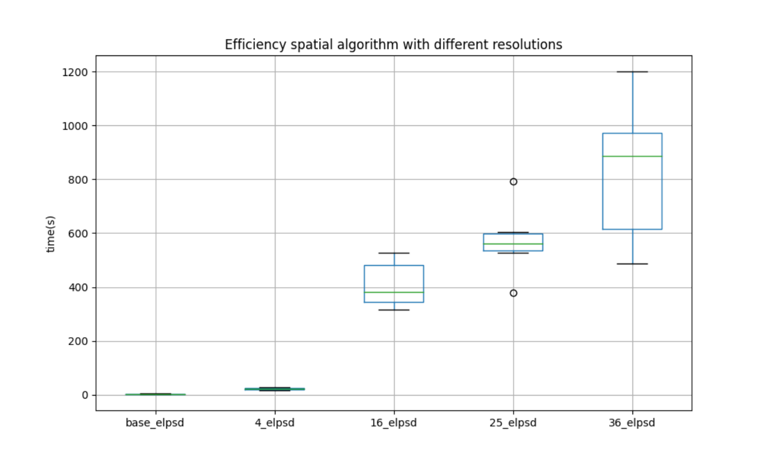


Figure 13: efficiency (in seconds) of spatial model for different resolutions (from 1 to 36); parameters: $\tilde{N} = 10^4$ (constant), $m = 4$, $\mu = 10^{-4}$, fitness: static increasing.

Thank you!