




Peer Community In Ecology

An R package for flexible and fast Stochastic Cellular Automata modeling

Samuel Alizon  based on peer reviews by **Broder Breckling** and 1 anonymous reviewer

Alexandre Génin, Guillaume Dupont, Daniel Valencia, Mauro Zucconi, M. Isidora Ávila-Thieme, Sergio A. Navarrete, Evie A. Wieters (2024) Easy, fast and reproducible Stochastic Cellular Automata with chouca. bioRxiv, ver. 6, peer-reviewed and recommended by Peer Community in Ecology.

<https://doi.org/10.1101/2023.11.08.566206>

Submitted: 11 March 2024, Recommended: 26 August 2024

Cite this recommendation as:

Alizon, S. (2024) An R package for flexible and fast Stochastic Cellular Automata modeling. *Peer Community in Ecology*, 100686. [10.24072/pci.ecology.100686](https://doi.org/10.24072/pci.ecology.100686)

Published: 26 August 2024

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Stochastic Cellular Automata (SCA) are a popular modelling tool because in, spite of their simplicity, they can generate a variety of spatial patterns. This makes them particularly appreciated, for instance, to validate the insights of analytical or semi-analytical spatial models that make simplifying assumptions, e.g. moment equations models. A first limit to SCA are that as soon as details are added to the model, reproducibility issues may occur. Computation speed is also an issue, especially for large populations. The work by Génin *et al.* addresses these two issues through the development of an R package, *chouca*.

The use of the package is designed to be as smooth as possible: users only need to define the type of possible transitions along with their rates, the parameter values, the number of neighbours, and the initial state of the landscape. The main function returns the population dynamics of each state and even the final state of the landscape.

In addition to its flexibility, an asset of *chouca* resides in its use of the Rcpp package, which compiles the model designed by the user in C++. This allows for high computation speed, which can be further boosted by using parallelising options from R.

In their manuscript, the authors use ecological models to illustrate the more advanced possibilities opened by *chouca*, e.g. in terms of graphical interpretation or even to estimate parameter values by computing likelihood functions (the implementation in R does make it very appropriate for statistical inference in general). The package still has some limitations, and, for example, it currently only applied to 2D rectangular grids and it cannot include elaborate movement processes. However, some of these could be addressed in future releases and *chouca* already has the potential to become central for SCA modelling, both for beginners and

expert users, especially in ecology.

References:

Alexandre Génin, Guillaume Dupont, Daniel Valencia, Mauro Zucconi, M. Isidora Ávila-Thieme, Sergio A. Navarrete, Evie A. Wieters (2024) Easy, fast and reproducible Stochastic Cellular Automata with chouca. bioRxiv, ver.6 peer-reviewed and recommended by Peer Community in Ecology
<https://doi.org/10.1101/2023.11.08.566206>

Reviews

Evaluation round #2

DOI or URL of the preprint: <https://doi.org/10.1101/2023.11.08.566206>

Version of the preprint: 5

Authors' reply, 23 August 2024

Many thanks to all parties involved in the review of this manuscript. We included most suggestions provided by B. Breckling below in the manuscript, whom we warmly thank again for dedicating time to our work. Please find below a short point-by-point response to the comments (mostly to acknowledge the changes made in our revision).

Alexandre Génin, on behalf of all co-authors

Review by Broder Breckling, 29 Jul 2024 13:09

2nd Review of the paper

"Easy, fast and reproducible Stochastic Cellular Automata with chouca"

by Alexandre Génin, Guillaume Dupont, Daniel Valencia, Mauro Zucconi, M. Isidora Ávila-Thieme, Sergio Navarrete, and Evie A. Wieters Available at

<https://www.biorxiv.org/content/10.1101/2023.11.08.566206v5.full.pdf>

The authors present a software which makes the application of a particular type of Cellular Automata (CA) easier. It departs from the R (statistics software) environment and allows a subsequent C++ compilation (back-end). Thus, larger grids can be run faster compared to a model that uses an interpreted language like R. The rule specification of models using the described software is limited to an operation of 2d rectangular grids. Cell state transition rules can use probability functions depending on the cells own state, the state of a 4-cell or 8-cell neighbourhood, and depending on the overall grid state. The current implementation would exclude descriptions of particle movement processes (like a random version of Langton's ant). Thus, there remains a considerable development potential to expand the applicability of the approach to a wider range of ecologically relevant interaction types that can be captured in other modelling environments.

The paper is well written and publishable in the present form. It is clear and understandable in particular to those, who are not programmers but who could be interested in an application of the software to model ecological interactions. The revision the authors considered the comments of the previous review in a reasonable way. The following statements explain the reviewer's view and are not necessarily requirements for further update.

Thank you for taking the time for this second round of review. We included most of your suggestions in our revision, as they will help show better how chouca fits within the landscape of models based on SCAs.

Some details of the paper to be looked at:

The abstract does not fully inform about the limitations of the approach. Readers may be disappointed to learn in the text that (random) movement processes cannot be run in the current implementation. This excludes a wide range of stochastic ecological CA applications. It would be reasonable to clarify this already in the abstract.

We now refer to the fact that probabilities are defined at the level of a cell, hinting at the fact that pairs of cells switching state may not be represented ("chouca supports SCA based on rectangular grids where transition probabilities are defined for each cell", l. 41).

The keywords could be more informative by adding "stochastic" to cellular automata and also mention the name of the software.

This has been included in the revision (l.54)

The limitation of the software to rectangular grids has the problem, that for many processes, discretisation artefacts prevail, which cannot be avoided by using a higher grid resolution. For an analysis of the relevance of results compared to empirical observations, it would be reasonable to confirm findings with simulations using an alternative grid topology, e.g. triangular, hexagonal, or even a Voronoi tessellation of a simulated area. This could be a further development perspective.

We now mention the use of non-rectangular grids as possible improvements (l.338).

The paper states in line 101 ff: "Probabilities of transition are assumed to depend only on (i) the proportion of neighbors in each state, captured by the vector $\mathbf{p} = (p_1, \dots, p_n)$, (ii) the proportion of cells in a given state in the whole landscape, $\mathbf{c} = (c_1, \dots, c_n)$, and (iii) a set of constant model parameters. chouca has been primarily designed for modeling the dynamics of sessile organisms over space, which do not move and reproduce through the dispersal of propagules."

Comment: this is the scope of the `_gap-model_` approach, elaborated e.g. by Shugart in the 1980ies to study ecological succession dynamics. The term is widely known in ecology as a model type. It might be useful to add that term (see e.g. Shugart, H. H., & West, D. C. 1981. Long-term dynamics of forest ecosystems: Computer simulation models, which allow for numerous seedlings and the long lives of large trees, predict how forests will respond to different management techniques. *American Scientist*, 69(6), 647-652.)

We now mention the term "gap model" in the corresponding sentence. We added a reference to the Shugart & West publication, which is a good illustration of the approach (l.106). Thank you for the suggestion.

Line 135 ff: The paper explains an example application description to illustrate the supported interaction types. It uses a transition between 3 different cell states. The example is presented very clear and well understandable.

In the part "graphical explorations" (line 227 ff), the authors show another 3-state model (host – parasitized-empty). This is a special case of an excitable media pattern where the according states are usually named excitable, excited, and refractory. The underlying interaction process occurs in self-organisation dynamics in various contexts not only in ecology but also in physiology, physics, and chemistry. To inform readers that this interaction type can be run with the software, it would be useful to include the term "Excitable Media" in the text.

This is now included in the new revision (l. 252).

The paper states in line 269 ff: "Because SCA are defined on grids, a natural application is to compare their output to empirical raster data, such as remote-sensing images, to infer local-scale ecological interactions from landscape-wide spatial patterns"

Comment: such a comparison can be biased because of CA discretisation artefacts. On a rectangular grid with a 4-cell neighbourhood processes in vertical or horizontal direction require less iteration steps than those in diagonal direction towards the same distance. This can influence resulting simulation pattern and thus the

comparison with empirical data (depending on rule-specification, the effect might be less pronounced with an 8-cell neighbourhood). Practical applications require a consideration of discretisation effects. It could be useful to point to this requirement when discussion comparisons with empirical pattern.

We do mention the fact that discretization artefacts may arise from using rectangular grids in the conclusion when mentioning limitations, as this is true of all simulations run with rectangular grids (even when not comparing to empirical data; l. 339).

The paper states in line 315 ff: "Applying such an approach to empirical data would require further testing of model assumptions, for example, to investigate whether facilitation occurs on the recruitment of new plants instead of on the mortality of adult plants – this can be done by simply changing the model definition above." Comment: the authors use the described model to explain, how the search for a parameter set can be done that approximates an observed pattern best. They show how a pattern comparison with systematically varied parameter identifies the size of 2 parameter that were used to generate the target pattern. The demonstration is suitable and understandable, however, from an ecological perspective, the model is questionable due to simplification but illustrates the pattern evaluation procedure well.

Thank you for this comment – it is indeed a simplistic model but its purpose is mostly illustrative as you note. We aim at showing how this approach is made possible in principle, but of course real-world applications likely require a more complex model.

As a positive point it has to be mentioned, that the connection to R can help in various statistical evaluations. The authors state in line 348 ff that the "code used for this work is freely-available at Zenodo under a CC-BY license (Génin, 2024), along with chouca version v0.1.99 used at the time of writing". This is a good condition to work with the software package and an achievement of the authors with regard to the relevance of their contribution.

Thank you for this encouraging comment. We appreciate your feedback on our manuscript.

Decision by **Samuel Alizon** , posted 03 August 2024, validated 03 August 2024

a valuable contribution to cellular automata modelling

The reviewer and I are happy with the changes made and agree that this article can be recommended by PCI Ecology.

The reviewer made a series of suggestions that the authors may want to include, when this (minor) revision decision.

Reviewed by **Broder Breckling**, 29 July 2024

2nd Review of the paper

"Easy, fast and reproducible Stochastic Cellular Automata with chouca"

by Alexandre Génin, Guillaume Dupont, Daniel Valencia, Mauro Zucconi, M. Isidora Ávila-Thieme, Sergio Navarrete, and Evie A. Wieters

Available at

<https://www.biorxiv.org/content/10.1101/2023.11.08.566206v5.full.pdf>

The authors present a software which makes the application of a particular type of Cellular Automata (CA) easier. It departs from the R (statistics software) environment and allows a subsequent C++ compilation (back-end). Thus, larger grids can be run faster compared to a model that uses an interpreted language like R. The rule specification of models using the described software is limited to an operation of 2d rectangular grids.

Cell state transition rules can use probability functions depending on the cells own state, the state of a 4-cell or 8-cell neighbourhood, and depending on the overall grid state. The current implementation would exclude descriptions of particle movement processes (like a random version of Langton's ant). Thus, there remains a considerable development potential to expand the applicability of the approach to a wider range of ecologically relevant interaction types that can be captured in other modelling environments.

The paper is well written and publishable in the present form. It is clear and understandable in particular to those, who are not programmers but who could be interested in an application of the software to model ecological interactions. The revision the authors considered the comments of the previous review in a reasonable way. The following statements explain the reviewer's view and are not necessarily requirements for further update.

Some details of the paper to be looked at:

The abstract does not fully inform about the limitations of the approach. Readers may be disappointed to learn in the text, that (random) movement processes cannot be run in the current implementation. This excludes a wide range of stochastic ecological CA applications. It would be reasonable to clarify this already in the abstract.

The keywords could be more informative by adding "stochastic" to cellular automata and also mention the name of the software.

The limitation of the software to rectangular grids has the problem, that for many processes, discretisation artefacts prevail, which cannot be avoided by using a higher grid resolution. For an analysis of the relevance of results compared to empirical observations, it would be reasonable to confirm findings with simulations using an alternative grid topology, e.g. triangular, hexagonal, or even a Voronoi tessellation of a simulated area. This could be a further development perspective.

The paper states in line 101 ff: "Probabilities of transition are assumed to depend only on (i) the proportion of neighbors in each state, captured by the vector $\mathbf{p} = (p_1, \dots, p_n)$, (ii) the proportion of cells in a given state in the whole landscape, $\mathbf{c} = (c_1, \dots, c_n)$, and (iii) a set of constant model parameters. chouca has been primarily designed for modeling the dynamics of sessile organisms over space, which do not move and reproduce through the dispersal of propagules."

Comment: this is the scope of the `_gap-model_` approach, elaborated e.g. by Shugart in the 1980ies to study ecological succession dynamics. The term is widely known in ecology as a model type. It might be useful to add that term (see e.g. Shugart, H. H., & West, D. C. 1981. Long-term dynamics of forest ecosystems: Computer simulation models, which allow for numerous seedlings and the long lives of large trees, predict how forests will respond to different management techniques. *American Scientist*, 69(6), 647-652.)

Line 135 ff: The paper explains an example application description to illustrate the supported interaction types. It uses a transition between 3 different cell states. The example is presented very clear and well understandable.

In the part "graphical explorations" (line 227 ff), the authors show another 3-state model (host – parasitized-empty). This is a special case of an excitable media pattern where the according states are usually named excitable, excited, and refractory. The underlying interaction process occurs in self-organisation dynamics in various contexts not only in ecology but also in physiology, physics, and chemistry. To inform readers that this interaction type can be run with the software, it would be useful to include the term "Excitable Media" in the text.

The paper states in line 269 ff: "Because SCA are defined on grids, a natural application is to compare their output to empirical raster data, such as remote-sensing images, to infer local-scale ecological interactions from landscape-wide spatial patterns"

Comment: such a comparison can be biased because of CA discretisation artefacts. On a rectangular grid with a 4-cell neighbourhood processes in vertical or horizontal direction require less iteration steps than those in diagonal direction towards the same distance. This can influence resulting simulation pattern and thus the

comparison with empirical data (depending on rule-specification, the effect might be less pronounced with an 8-cell neighbourhood). Practical applications require a consideration of discretisation effects. It could be useful to point to this requirement when discussion comparisons with empirical pattern.

The paper states in line 315 ff: "Applying such an approach to empirical data would require further testing of model assumptions, for example, to investigate whether facilitation occurs on the recruitment of new plants instead of on the mortality of adult plants – this can be done by simply changing the model definition above." Comment: the authors use the described model to explain, how the search for a parameter set can be done that approximates an observed pattern best. They show how a pattern comparison with systematically varied parameter identifies the size of 2 parameter that were used to generate the target pattern. The demonstration is suitable and understandable, however, from an ecological perspective, the model is questionable due to simplification but illustrates the pattern evaluation procedure well.

As a positive point it has to be mentioned, that the connection to R can help in various statistical evaluations. The authors state in line 348 ff that the "code used for this work is freely-available at Zenodo under a CC-BY license (Génin, 2024), along with chouca version v0.1.99 used at the time of writing". This is a good condition to work with the software package and an achievement of the authors with regard to the relevance of their contribution.

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/2023.11.08.566206>

Version of the preprint: 3

Authors' reply, 12 June 2024

Dear editors - thank you for handling our manuscript. Please see our response letter attached to this form.

[Download author's reply](#)

Decision by **Samuel Alizon** , posted 13 May 2024, validated 13 May 2024

a useful software package to better position in the current literature

As you will see in the attached reviews, the two reviewers and myself all agree about the usefulness of your software package for the community, but we noticed a few issues that would need to be addressed in a revision before envisaging a formal recommendation. First, as pointed out by the first reviewer, the width of the cellular automata literature could be better acknowledged and, therefore, so could be the limitations of the package. Second, as pointed out by the second reviewer, further discussing existing packages on other platforms, such as Python, would help better position your package. Finally, to follow-up on a comment from reviewer #1 about biological applications, I would like to add that the example of lizard scales is also one of the most inspiring ones I have seen of late (Manukyan et al 2017 Nature, DOI: 10.1038/nature22031).

Minor suggestions

- l.28,29,30: repetition of "scales"
- l.29: "Markovian" might be a bit technical word for the first sentence of a ecology abstract.
- l.35: mention reproducibility?
- l.35: delete "more importantly"?
- l.84: "computationally" instead of "compute"?
- l.86: mention moment equations rather than just "pair approximation"? (see also the review by Lion 2016, JTB 10.1016/j.jtbi.2015.10.014)

I.98: I would start a new section here because this seems to be more about "The Model" rather than the "Introduction".

I.123: Perhaps mention Rcpp at this stage?

I.124: This could be rephrased a bit into a classical "Results" section or "Illustration"

I.195: Write "This can be improved by" instead of "improved on by"?

I.213: Schneider et al (2016) does not seem to be in the bibliography.

I.217: The previous section already seemed to be an illustration example. Perhaps merge the two could help?

[Download the review](#)

Reviewed by **Broder Breckling**, 13 May 2024

Génin et al present a framework to implement a kind of simplistic cellular automata with probabilistic state transitions. They use illustrative ecological examples to show the functionality. Because they apply R as a computational framework, it can be expected that it makes the access for users easier compared to genuine programming. However, the approach is syntactically rather limited and excludes important ecological interaction types, e.g. it operates on 2d rectangular grids only and excludes direct cell-to-cell relation that would allow 1:1 mutual state change of cells. Besides specific application cases, its benefit could be in the introduction to spatial modelling in a didactical context. For research applications, in most practical cases a higher flexibility and complexity in rule-specification might be required.

In the following, comments to particular statements in the text are added. Because the paper does not provide line numbers, the commented sentences are copied below.

Specific comments

The paper states

"Stochastic cellular automata (SCA) are models that describe spatial dynamics using a grid of cells that switch between discrete states over time, depending only on the current state (Markovian processes)."

Comment

This is a rather narrow specification which excludes an application for many practical requirements. If for the characterisation of a cell's state continuous variables are excluded, many relevant processes cannot be modelled.g. the concentration of a substance, continuous variation in pressure, temperature, etc. at a particular grid position. This is in particular relevant for ecological applications.

The paper states

"de novo implementation of SCA for each specific system and application represents a major barrier for many practitioners."

Comment

It is questionable if this is actually a barrier. Even without object-oriented programming, it is not such a challenge to develop a CA. For a rectangular CA with a von Neumann neighbourhood, two nested FOR-loops together with conventional calculation would do as a main component to set the layout for an iteration cycle.

The paper states

"we built chouca, an R package that translates intuitive SCA model definitions into compiled code,"

Comment

this may be a major advantage compared to interpreted languages. The authors present a figure (Fig. 3) which compares runtimes of different grid size and indicate a considerable efficiency.

The paper states

"ad hoc implementations found in the literature,"

Comment

here, a reference would be useful to clarify which implementations the authors have in mind

The paper states

"Conway's game of life,"

Comment

a reference would be nice

The paper states

“The probabilities of a cell switching from one state to another is assumed to depend on model parameters, the global state of the system (the proportion of cells in each state), and the local neighborhood of the focal cell.”

Comment

Frequently, the change of the cell's states in CAs are considered to depend on the state of the cell itself and its neighbours. The specification here, expands that and offers to operate on two levels of neighbourhood – a local and a global one. This is possible, however, the evaluation of the entire grid state to specify a new cell's state is usually not required as a CA standard. Nevertheless, under particular condition, this could be an interesting option.

The paper states

“In many cases, the explicit numerical simulation must be run, which is often done on small grids to reduce computation time.”

Comment

What do the authors understand as a “small grid”? Since computational potentials expand with technical developments, this statement is relative and tends to lose relevance. Is the grid size stated below (p.7: “a typical 128 x 128 grid (Figure 3)”) a small one?

The paper states

“This opens the possibility of errors in code and often makes it difficult to reproduce model simulations.”

Comment

this is a widespread issue throughout ecological modelling. It may be a useful contribution of this approach, if for this specific segment in modelling, result reproduction may become easier.

The paper states

The R package chouca works with 2-dimensional rectangular grids of cells (a “landscape”).

Comment

In an upcoming version, the authors should expand it to 3d grids, then allowing to represent also waterbodies or processes referring to atmospheric layers.

A further step then could be to facilitate the use of triangular or hexagonal grids. To have an alternative to rectangular grids would allow for tests of the extent of discretisation artefacts – which are quite relevant for rectangular grids with regard to ecological processes.

The paper states

“It is important to note that this excludes cellular automata in which an intermediate distance of interaction is considered (e.g. through a dispersion kernel; Muthukrishnan et al., 2016), or those in which a preferential direction exists (e.g. modeling water redistribution on a slope; Mayor et al., 2013). Other types of SCA not fitting these constraints are those in which two cells swap their respective state, e.g. when modelling the movements of a predator in a landscape (Pascual et al., 2002)” ... “In practice, this functional form is flexible enough to approximate the probabilities of transition of many ecological models”

Comment

this excludes a very large extend (if not the majority) of potential ecological CA applications. If not even the change of a state's cell with its neighbour could be represented, then the whole class of pattern generating processes like Diffusion Limited Aggregation as well as a majority of trophic interaction modelling would be excluded.

The paper states

“Arid systems provide a good illustration of this approach: in those systems, plants often facilitate each other, which results in their aggregation into patches, and has important consequences for the resilience of those systems to changes in aridity”

Comment

even more often, in arid systems plants limit each other through competition for water access and give rise to gaps in vegetation cover. Facilitation is a process that helps in the survival of seedlings and is usually only a part of the relevant interactions. To describe plant dynamics under arid conditions might be more convincing by using "inhibition" rather than "facilitation"...