

Evolving Optimal Spatial Allocation Policies for Complex and Uncertain Environments

Marta Vallejo^(✉), David W. Corne, and Verena Rieser

School of Mathematical and Computer Sciences,
Heriot-Watt University, Edinburgh, UK
mv59@hw.ac.uk

Abstract. Urban green spaces play a crucial role in the creation of healthy environments in densely populated areas. Agent-based systems are commonly used to model processes such as green-space allocation. In some cases, these systems delegate their spatial assignment to optimisation techniques to find optimal solutions. However, the computational time complexity and the uncertainty linked with long-term plans limit their use. In this paper we explore an approach that makes use of a statistical model which emulates the agent-based system's behaviour based on a limited number of prior simulations to inform a Genetic Algorithm.

The approach is tested on a urban growth simulation, in which the overall goal is to find policies that maximise the inhabitants' satisfaction. We find that the model-driven approximation is effective at leading the evolutionary algorithm towards optimal policies.

Keywords: Agent-based model · Genetic algorithm · Statistical model · Optimisation · Uncertainty · Green space planning

1 Introduction

The main purpose of urban planning is to improve the community's quality of life by creating a better social, economical and physical environment. One of the most urgent research issues within this broad field is the study of mechanisms that can mitigate the ecological degradation that is linked with modern urban expansion. One possible strategy is reserving a collection of selected areas to transform them into recreational parks.

However, this process is not as simple as choosing arbitrarily a random number of stands, the time planning and geographic distribution of these spaces needs careful consideration to ensure the quality and quantity of environmental services provided to the surrounding community [2].

This paper is a revised and extended version of a previous publication [1] reported in the Proceedings of the 5th International Conference on Agents and Artificial Intelligence. The key additions cover: Sects. 3 and 4 improvement of the complexity of non-urban cells prices and its inclusion as a new source of uncertainty. In Sect. 5, a new heuristic is studied to enrich the comparison phase. Finally in Sect. 6 more experimental results and comparative evaluations are performed.

There is much active research in designing long-term feasible public open space plans, whereby researchers interested in urban planning and sustainability have investigated a range of agent-based systems and similar mechanisms to explore the consequences of different strategies [3–5].

One of the most common interests in such work is the study of the dynamics involved in urban growth, which is linked with the relative distribution of urbanised, industrial and green spaces along with their impact on quality-of-life issues, and how these factors depend on the broad strategies in place for land-use [6].

However, the computational time complexity of simulations and the many and varied sources of uncertainty can limit the use of these systems. The aim of this paper is to address this situation by wrapping optimisation over the agent-based simulation process, but use a statistical model of the agent-based simulation in place of the real knowledge. This requires a limited number of prior simulations of the agent-based urban growth system in order to allow the use of an evolutionary algorithm to optimise urban growth policies.

Note that similar simulation-based approximations for optimisation are also used in other fields, such as user simulations for spoken dialogue systems [7], emulators for managing uncertainty in complex models, such as climate models (MUCM) or to reduce the computational time required to run the optimisation procedure in combinatorial problems [8].

The approach is tested on a typical urban growth simulation, in which the overall goal is to find policies that maximise the 'satisfaction' of the residents by the protection of an optimal subset of green spaces. The computational results are compared and evaluated with those gathered from several simple heuristics.

The remainder of the paper is organised as follows. Section 2 focusses on various introductory and preliminary details, covering the urban planning problem, the role of agent based simulation, and evolutionary algorithms. Section 3 then provides a detailed account of the models, assumptions and processes we employ in our experiments. Section 4 is devoted to the sources of uncertainty that are handled by our new statistical genetic algorithm approach. Computational experiments are specified in Sect. 5, and the results are presented and discussed in Sect. 6. Section 7 then draws some conclusions and we discuss further research.

2 Problem Definition

Open green urban areas play an important role in maintaining a healthy urban environment. Among all their favourable effects, their crucial impact in the economy, quality of life and in the local climate of the cities [9, 10] can be highlighted. However their distribution and location should be carefully studied by developing an adequate, long-term planning strategy. The fact that makes this task particularly difficult is the fact that the urban expansion is a complex process where their effects can appear at many time-scales.

2.1 Urban Open Space Planning

There is a lack of agreement on how to implement and implant a given planning process and which measures should be selected. On this regard, the most remarkable points to discuss are:

- Goal settings: how to select adequate planning criteria.
- Deciding the most suitable size for the open space according to the current and expected necessities of the population.
- Accessibility & location.
- The design of the potential activities according to different age, cultures and ethnic groups.

The present work follows a *demand approach* where the planning process is based on attributes of the specific target population. These attributes are:

- Size of the urban population.
- Subjective personal preferences.
- Residential distribution.

2.2 Problem Formulation and Techniques

The problem domain of the present paper can be included within the field of stochastic control theory. The developed model represents a paradigm of allocation of resources within a sequential decision-making simulator.

Generally speaking, a sequential planning problem can be defined as follows: an environment which can be described as a state-space set S and an action set A where S and A are both finite. Each state $s \in S$ is dependent on the previous state of the system and the action $a \in A$ taken. The transition function δ controls how actions modify the state of its environment.

$$s_{t+1} = \delta(s_t, a) \quad (1)$$

We define a policy Π such that the mechanism in charge of selecting the next action is based on the current perception of the environment. This perception can be total or partial:

$$\begin{aligned} \Pi : S &\rightarrow A \\ \Pi(s_t) &= a_t \end{aligned} \quad (2)$$

In turn, the action a influences as well its environment provoking the change of the current state. The process starts in the state s_0 and by means of the sequential application of the policy Π , further actions are chosen.

2.3 Cellular Automata and Agent-Based Modelling

The present study is based on the results collected from a basic urban growth model where topological layout of the city is represented by a Cellular Automata (CA). CA was proposed in the late 1940s by John von Neumann and Stanislaw

Ulam for discrete space-time representation of problems which obey their local physics [11]. It is based on the assumption that by means of local interactions, the model is capable of representing complex phenomena. The dynamics of the CA are generated by a set of transition functions which define how cells can evolve from one state to another.

The inhabitants who populate the city are modelled with the use of an Agent-Based Model (ABM) approach. ABM has been used to understand the interconnections, interdependences and feedbacks created among a set of heterogeneous individual entities in order to fulfil their goals.

ABM along with CA taking the role of representing land-use change dynamics have been applied broadly in the field of urban development. Mentionable is their use to simulate allocation decisions [12, 13] or in residential selection within a non-stationary housing market [14, 15]. Reference [16] applies these tools to analyse how prices affect urban agent behaviour. Finally, [17] studied the role of transportation in the evolution of an urban region.

2.4 Genetic Algorithm

Genetic Algorithm (GA) [18] can be defined as an heuristic that mimics the behaviour of natural selection postulated by the English naturalist Charles Darwin in the 19th Century [19]. This search strategy is based on the assumption that nature evolves by the course of new generations preserving the species more suited to their environment. The tools defined by a GA to improve the population over time are the use of mechanisms like reproduction, mutation, crossover and selection.

Here we use GA to optimise an allocation of resources problem, concretely the placement of green spaces over a urban area such that some objectives are met. GA has been successfully used to solve complex spatial problems [20, 21]. However, its performance in uncertain environments has been questioned [22, 23] due to the fact that a simple GA has insufficient data to deal directly with uncertainty. This weakness is the main reason why a GA, under this kind of scenarios, should be defined carefully and provided with the support of external tools in order to overcome these difficulties.

There exist different attempts and techniques that can be applied to GA to provide it with this extra functionality. In [24] a Genetic-Algorithm-Aided Stochastic Optimisation Model is applied to cope with the uncertainty related to the study of air quality in urban areas. In contrast to probabilistic approaches [25] resorts to anti-optimisation techniques (local search) to overcome the uncertainty generated by the ageing factor presented in many engineering problems. Following the same approach [26] successfully applies a variant constrained multi-objective GA in a simulated topology and shape optimisation problem under uncertainty.

3 Model Description

The selected ABM-CA framework is used to represent a basic urban growth model with a monocentric spatial structure based on the traditional Alonso's urban economic model [27]. The strategy of this model to explain the modern urbanisation process is based on the maximisation of a utility function. Urban pattern formation is the consequence of individual urban residence preferences which achieve an economic competitive equilibrium between housing and commuting costs.

The physical layout of the city is configured by a 2-dimensional lattice of 50×50 cells. Each cell corresponds to a physical portion of the city and it can be populated by more than one agent (a family unit). The evolution of the city is ruled by an internal schedule with a determined time-horizon of finite duration. The dynamics of agents and cells allow the model to evolve between a set of predefined one-directional states at each time step.

The types of the cell presented in the grid can be broadly divided into two main groups: urbanised and non-urbanised cells.

3.1 Urban Cells

Urban cells represent cells that have been transformed from native ecosystems into either impermeable surfaces or green areas formed normally by non-native species [28].

In the model, when cells receive the permission to be urbanised, which figuratively means that dwellings are constructed, they can allocate population that is represented by agents. Agents decide their residence location by searching a trade-off between their personal preferences and their economical restrictions. This search involves the interaction among different parameters of the model and assumes global knowledge of the current offer. The decision is represented by the maximisation of the following utility function:

$$\begin{aligned} \max U = (w, z, x, p : w > 0, z > 0, x \geq 0, p > 0) \\ \text{such that: } w - z - kx + p = 0 \end{aligned} \quad (3)$$

where x represents the distance from the household to the Central Business District (CBD) that is located in the centre of the lattice, w is the wage received monthly. This quantity is defined by a uniform random process and does not change throughout the life time of the agent, z is the price of the residential good and k is the constant marginal community cost. Finally, p represents the agent's preference for houses located close to green areas which implies his acceptance to pay more for this kind of houses. This parameter is an extension of the economic competitive equilibrium described by [27]. Following this utility function agents populate the urban cells of the grid.

Prices of the Urban Cells. They represent the amount of money that agents have to pay regularly as a rental cost. Its value varies with the time and is dependant on the following factors:

- **The Demand.** The demand is defined according to the number of agents living in a given cell.
 - The demand for certain preferred locations increases their price.
 - The drop in population of a cell decreases its price.
 - If one cell does not receive any new neighbour during a determined period of time, its value is reduced.
- **Proximity to Green Areas.** This factor affects positively the final price of the dwellings in 10%.

3.2 Non-urban Cells

Non-urban cells are cells that have not undergone a urbanisation transformation.

Biological Value. At the beginning of the simulation the model assigns a stochastic value called *BioCellValue* to the set of non-urbanised cells. This parameter represents the ecological value of this parcel of land and is generated by a uniformly random process $\mathcal{U}(0, 1)$. Apart from its initial value, the cells are influenced by its neighbourhood:

```

BioNeighbourValue(c)=0;
for each cell n in neighbourhood(c){
    if(BioCellValue(n)>= 0.7)
        BioNeighbourValue(c) += 0.01;
    if(BioCellValue(n)<= 0.3)
        BioNeighbourValue(c) -= 0.01;
}
BioValue(c) = BioCellValue(c) + BioNeighbourValue(c);

```

The final *BioValue* is used to identify different land-types of the model. If it is bigger than 0.7 this cell is classified as a *forest* cell, otherwise it is considered *agricultural*. The belonging to each category is dynamic over the time. Furthermore the model experiences a continuous bio-degradation which provokes changes from *forest* to *agricultural* state due to the urban expansion.

Governments can adopt a wide range of interventionist mechanisms to restrict the ownership over the land and control its use, acting as a response to social requirements over gardens and parks to provide a set of services based on the proximity to potential users. Among these measures the local authority can assume the proper ownership of the land like in the case of Stockholm city [29] and assign them partially or totally the function of urban green spaces.

Based on that premise the model delegates the responsibility of selecting the best non-urbanised stands to a new special agent called *Municipality*. This agent does not interact with the rest of agents, but his main goal consists of managing the purchase and protection of green areas within the city by means of a monetary income received periodically called *budget*.

Purchase of Land. As it was stated previously, the location of green areas is a crucial factor in its future use because the kind of services that a park can provide is linked with the concept of proximity. Reference [30] states that the distance to a green area influences the frequency of use and the activities that can be undertaken. According to this criterion, green areas can be classified into the following groups:

- Access within a short walk (less than 300 m).
- Access within a long walk (from 300 to 600 m).
- Access with help of any means of transport (larger than 600 m).

The same study concludes that people do not generally use a green area if it is located beyond a threshold of 300–400 m. In the model the location selection is performed sequentially in each time-step and is limited by the *budget* and the configuration of the lattice in this precise moment. Once the purchase is concluded, the state of the cell is changed to *protected* and the future construction of urban facilities within it is forbidden.

This selection process can be formulated as follows: if C is defined as the finite set of cells included into the lattice, A the subset of rural cells that can be considered as a candidate cell to be purchased, P the subset of cells that are protected and U the urbanised cells such as $\{A, P, U\} \subset C$ and $A \cap P \cap U = \emptyset$, then the selection of a candidate cell in time t can be defined as:

$$\begin{aligned} & \forall \text{ cell } c \in C \\ & \text{if } \text{price}(c)_t < \text{budget}_t \wedge c_t \notin \{P, U\} \\ & \implies c_t \in A \end{aligned} \quad (4)$$

Once the candidate set is defined, the purchasing and protection phase can be formalised as:

$$\begin{aligned} & \forall \text{ cell } a \in A \\ & \underset{\text{satisfaction}}{\text{max}} \delta(a)_t \\ & \implies a_t \in P \wedge a_t \notin A \wedge \text{update}(\text{budget}_t) \end{aligned} \quad (5)$$

The function δ represents the metric that measures the level of satisfaction of the population in terms of the distance to green areas. See formula 9. Every subset of selected cells has associated a level of satisfaction of the population allocated within the boundaries of the city. The model should select the configuration of green areas which achieves the highest possible level of satisfaction according to the restrictions of the system during the considered period of time.

Prices of Non-urban Cells. It is calculated based on a simplified version of the formula 9 developed in [31]. The price is defined in terms of the current urbanised prices and the distance of the cell to the boundary of the city as:

$$\begin{aligned} & \forall \text{ cell } c \in A \\ & \text{price}(c) = \text{priceBase}(c) * (1 - e^{-\alpha[Z-Z^*(t)]})_+ \\ & \text{urbanPrice} * (e^{-\alpha[Z-Z^*(t)]}) \end{aligned} \quad (6)$$

Where *priceBase* corresponds to Table 1 based on prices of rural land in UK [32]. $\alpha = 0.2$ is the rate of change in price as the distance from CBC increases. Z is the current position of the cell and Z^* the boundary of the city. $Z - Z^*$ represents the distance from the cell to the limit of the city where $Z > Z^*$. Finally *urbanPrice* is the current price of the most recent developed cell multiplied by its population density.

Table 1. Prices per cell.

Type of cell	Area \times price	Final price
1 cell forest	51.8 ha \times £6,600 = £854,700	
1 cell agriculture	51.8 ha \times £3,000 = £388,500	

The plot of the averaged values of the prices during the simulation is depicted in Fig. 1. It is noticeable the critical growth in prices that occurs at the end of the simulation. This is due to the fact that the number of available cells is scarce and their demand increment drastically its value.

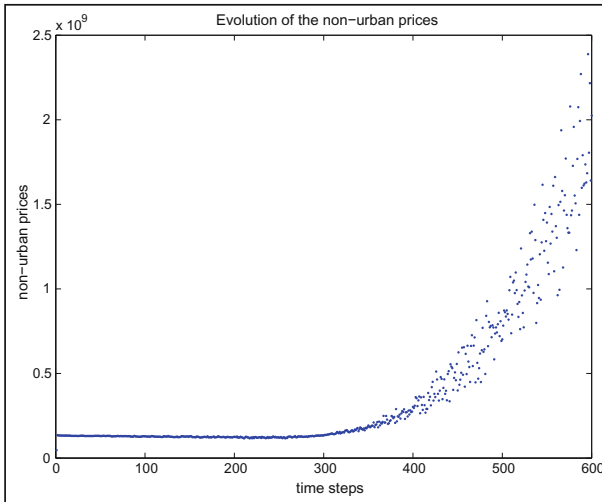


Fig. 1. Evolution of the prices of the non-urban cells throughout the simulation.

4 Sources of Uncertainty

In the present model uncertainty can emerge from a wide variety of sources. Apart from the fact that the implementation of long-term plans always implies to be able to cope with unpredicted future scenarios, the complexity resultant

from the multiple interactions occurred between the elements represented in the model makes their management even more challenging.

Some factors which actively contribute to the increment of the level of uncertainty are mentioned in the list below:

4.1 Urban Property Prices and Green Areas

In the developed model, the selection of green spaces exerts a direct influence on the prices of the surrounding urban cells. [33,34] analyse this tendency reporting a significant increment in the prices of residences located close to urban parks. This aspect is included in the model as the agents' desire to live close to these areas and is represented by the agent's acceptance to pay more for these specific locations. The inclusion of this personal desire provokes a significant growth in the demand of these areas and subsequently in the price and affects the urban spatial spread of the city.

4.2 Ecological Degradation Process and Non-urban Property Prices

From the point of view of the non-urbanised cells, one of the two main parameters which involves a high level of uncertainty is the relationship created between the non-urban price dynamics and the cells' ecological value. Due to the fact that this ecological value also influences its neighbourhood, a significant change in a specific area of the lattice spreads in all directions. The *bioValues* are steady until the distance to the city measured in Manhattan distance is less or equal to 3, otherwise it is applied the following update:

$$\forall \text{ non-urban cell } a \in A$$

$$bioValue(a) = \begin{cases} -0.1 & \text{if } \delta(a, city) = 1 \\ -0.05 & \text{if } \delta(a, city) = 2 \\ -0.01 & \text{if } \delta(a, city) = 3 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Where δ is the distance from the cell to the boundaries of the city. The application of this formula produces an ecological degradation process with the growth of the city. This dynamic influences the *priceBase* of the non-urbanised cells that are closely located to the city and hence, the purchasing process of protected areas that it is restricted to our current *budget*.

4.3 Urbanisation Process

The underlying process of urbanisation is in nature partially random and mainly determined by two factors:

- **The rules of transition of the cells:** based on preselected probabilities.
- **The demand level:** controls the transformation of peri-urban into new urban cells.

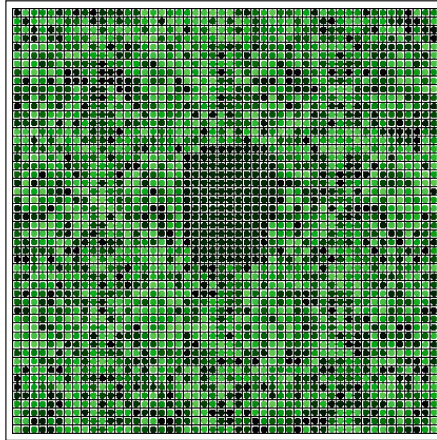


Fig. 2. Environmental values and the effect of the urbanisation process. The range of colours from green to black depicts the ecological values of the cell. Notice that in the centre where the city is located, the black eco-values represent the biological degradation or the metropolitan area (Color figure online).

The knowledge of the urbanisation process is crucial because the set of candidate cells to be protected are restricted to the non-urbanised ones and hence we need to be aware of the complete state of the cells in each time step in order to select and protect non-urban cells.

4.4 Flows of Population

Another significant characteristic of the model is that the city is a non-closed-system. This means that there is an external income flow of new population coming from migration as well as new offspring resulted from the current settled population. The dynamics of these flows are not fixed and predictable and they play a relevant role in the final population distribution within the city. However the density of each future neighbourhood cannot be totally predicted in advance even if there exists a general preference to live close to the city centre in line with the Alonso's model.

Consequently we are not able to know the percentage of population directly affected by a determined location of a new green area and hence, the final satisfaction achieved by a determined configuration of green spaces.

5 Case Study: Allocation of Green Areas

There are three parallel optimisation tasks which form the current case study. The more complex and main objective of the present paper is the GA flow, in contrast to the other two approaches, the random and the best non-optimised strategies that have been developed as a comparison tool.

5.1 Set-Up Phase

The point of departure of the following workflow consists of the definition of the initial model configuration. The model uses as a point of departure two common parameters which share with all the components of the optimisation framework. These parameters are:

- **The Budget.** A stochastic budget assigned to the *municipality* in each sequence of time is decided in advance and it is shared as well by the entire GA population.
- **The Ecological Scenario.** The initial ecological configuration of the lattice is defined by the initial random generation of ecological values. See Fig. 2.

5.2 Configuration of the GA

Chromosome Encoding. The GA evolves through time a population of individuals of size 20 which are chosen randomly from the set of candidate solutions. An individual is encoded as a sequential selections of cells grouped in a predefined number of time steps (Fig. 3). Each of these selections represents a gene and can contain $\{0, n\}$ protected cells chosen by the *Municipality* in one time-step. The superior limit n is bounded by the maximum *budget* available for this time step. Linked with each subset of cells, the remain budget is stored that can be calculated as:

$$remainBudget_t = budget_t - \sum_{i=0}^n price(c_t) + remainBudget_{t-1} \quad (8)$$

Scheme Selection. There exist many selection schemes for GA, among them the present model uses tournament selection (TS) [35]. TS is a robust and simple to code selection mechanism for GA based on the idea of holding a tournament between a group of competitors randomly selected among the population.

Mutation Process. Mutation is a tool used to maintain the diversity among the population of individuals. The mutation process alters one or more values of the genes inherited from the parent. In the present case a mutation consists of changing the set of cells selected to be protected in a slot of time of the simulation. Additionally this implies the update of the associated budget. A successful mutation should meet some constraints:

- A cell cannot be selected twice or more for the same individual.
- The remain budget should be always positive. Debts are not permitted.
- Cells cannot have the state of urbanised when protecting them.

The *budget* can arise potential problems during the mutation process. Due to the fact that non-urbanised cells prices evolve with the time along with the aggregate nature of the budget, a single modification in the selected cells of a gene can influence substantially the amount of money that needs to be used in future purchases. The mutation process should not modify the rest of genes.

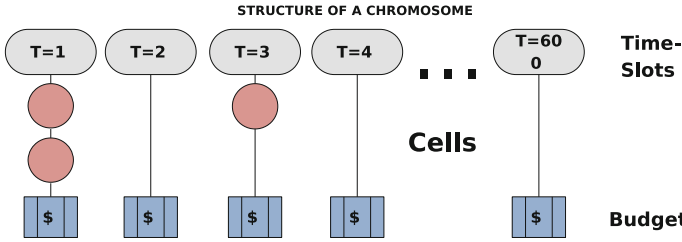


Fig. 3. Three-layers chromosome encoding.

Fitness Function. To measure the contribution of a determined green area to a solution, different kind of metrics can be used according to which aspects want to be emphasized. In the current model the quality of a solution or satisfaction represents the accumulative satisfaction achieved by each person settled on the city with respect to the topological distribution of green areas. Following this approach the fitness function can be defined as follows:

If A is the set of agents of the city, P is the subset of protected cells and C is the set of cells defined in the grid such as $P \subset C$, then for a given time t :

$$\begin{aligned}
 &\forall \text{ agent } a \in A \wedge \forall \text{ protected cell } c \in P \\
 s(a) &= \begin{cases} 3 & \text{if } \delta(a, \text{green}) = \min_{\text{distance}}(a, c) = 1 \\ 2 & \text{if } \delta(a, \text{green}) = \min_{\text{distance}}(a, c) = 2 \\ 1 & \text{if } \delta(a, \text{green}) = \min_{\text{distance}}(a, c) = 3 \\ 0 & \text{otherwise} \end{cases} \quad (9) \\
 \Theta &= \sum_{i=0}^n s(a_i)
 \end{aligned}$$

δ is defined as the function that calculates the distance from the location of a given agent a to the closest green area in the grid using Manhattan distance. Besides we define s as the function which retrieves the individual satisfaction achieved by a given agent a . Finally, Θ represents the total satisfaction achieved by the population of size n in the lattice in time step t .

This fitness function is, in turn, linked directly with the spatial spread of the city and the population density of each stand. However, to be able to use a fitness function in the GA, it is necessary to know the location of the entire population in each time step.

5.3 GA Workflow

Collected Data. In this phase the knowledge that the GA cannot infer a priori is gathered from a urban growth version of the simulation. The collected data includes the following elements:

- **The Topological Development of the City.** Due to the fact that only non-urbanised cells can be candidate to be protected it should be gathered when each cell is more likely to be urbanised.
- **The Population Evolution (Number and Location).** The simulation collects statistical data about the amount of agents living in the city and their precise location in the grid in every time step. This density distribution is necessary to calculate the fitness function that is used to measure the quality of an individual solution.
- **The Non-urban Prices Dynamics.** Due to the fact that budget should be always positive and prices can change with the time, it is necessary to know which prices correspond with which non-urban cells throughout the simulation. In this case the mean of the prices in multiple simulations is annotated.

GA Optimisation. Once the data is gathered, the optimisation can be carried out. In this phase the GA population is generated and evolved using TS for 5000 iterations, assuring their convergence. For each new generation, the possible candidate cells should satisfy the constraints described in the mutation procedure to cope with the restrictions derived from the management of an uncertain future.

Test Component. Once the optimisation phase has been concluded and the final individual solution with the highest fitness is selected as the final solution, the test phase is carried out. The test component uses the output data from the GA phase to check the viability of the protected cells analysing how the statistical model compromises the model validity.

These simulations run in a modified version of the model where the green spaces are selected deterministically meanwhile the rest of factors and interactions maintain its complex and unpredictable behaviour. The main purpose of this test step consists of:

- Measuring the real satisfaction of the population.
- Detecting inconsistencies and incompatibilities of the cells selected by the GA. The inconsistencies are linked directly with the quality of the statistical data. With the appropriate data the GA can infer more accurately the situation of the lattice and better overcome the constraints, reducing the number of inconsistencies.

5.4 Other Heuristics

Random. Random approach is a heuristic that allows 10 attempts to randomly select a cell to be purchased in each time-step. The cell selected should have the state of non-urban and the current available budget should cover the price of the cell. The first cell that meets these conditions is protected.

Best Non-planning Option. This strategy is a heuristic which tries to give the best service in the current moment without performing any further plan. The approach buys the terrains located in any of the areas that are adjoining the new

urban development. Concretely, the algorithm retrieves information about the last cell urbanised in the lattice and searches in the subset of cells which form its neighbourhood, the ones that are not urbanised yet. If its price is lower than the current budget, the cell is protected.

6 Computational Results

The results presented in the paper were calculated as averaged over 20 repeated optimisations, all of them in compliance with the assumptions and restrictions commented in previous sections.

Table 2. Satisfaction grouped in 50 time-steps over the three approaches analysed.

	Random	Best	GA
50	97.94	14.77	208.01
100	411.84	160.99	918.39
150	1407.58	729.74	3055.70
200	4651.23	2867.48	9716.36
250	15342.94	10623.94	29537.54
300	47089.13	35887.04	79909.52
350	89843.31	70444.26	130117.76
400	112007.76	91359.08	146638.63
450	119140.86	103118.51	148372.77
500	118023.17	109580.54	143162.94
550	113881.15	113274.05	135731.20
600	108935.25	114972.24	127935.58

Table 2 summarises the results achieved by the three analysed heuristics. The results measure the average of the satisfaction achieved by the population during the 600 ticks of the simulation grouped in 50 steps. The first column shows the random satisfaction, the second represents the best non-optimise heuristic and the third column includes the results for the GA-optimised satisfaction. Graphically, the same data is depicted in Fig. 4. From these results we can state that GA outperforms the random and the best non-optimised heuristic throughout the simulation.

6.1 Comments About the Random Approach

Due to its stochastic selecting mechanism, the random solution spreads more homogeneously and scattered their protected cells. It does not take advantage of the reduced prices at the beginning of the simulation but, in turns, when prices

are too high to be able to purchase any new land (see peak in non-urban prices in Fig. 1), the random approach achieves to give service to the outskirts of the city in contrast to the other two approaches that are more sensitive to the significant increment in prices occurred at the end of the simulation. One limitation of the stochastic random approach is that the efficiency achieved depends strongly on the extension of the land analysed.

6.2 Comments About the Best Non-optimise Heuristic

This heuristic achieves acceptable results when the non-urban prices are low and the growth speed of the city is slow. However as the approach always tries to buy the most expensive non-urban cells that are the ones located closer to the boundaries of the city, when the city grows and the demand for land increases its price, the available budget is not enough to afford new purchases and the heuristic neglects to give proper services to the outskirts of the city. Due to that this heuristic shows the worst results at the end of the simulation. The total amount of cells protected are smaller and concentrated most of them around the city centre.

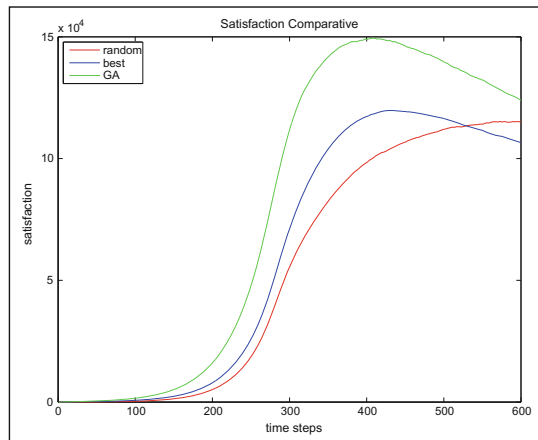


Fig. 4. Representation of the satisfaction achieved by the three heuristics developed.

6.3 Comments About the GA Approach

The GA overcomes both heuristics during the entire simulation. However, the results are closer to the best non-optimised heuristic at the beginning of the simulation and to the random approach at the end. The decline in the satisfaction at the end of the simulation is due to the fact that the algorithm is not able to buy new stands with the current budget and from the 400 time-steps the optimisation is poor. This negative effect could be avoided if the amount of budget assigned in the experiments would have been enough to buy homogeneously new green areas during the entire duration of the simulation.

6.4 Model Validity

The test component checks the validity of a given GA solution using an independent simulation. Moreover it gathers some data to provide information about the quality of the GA solution: the amount of urban inconsistencies and the satisfaction achieved by the solution. A urban inconsistency can be defined by the attempt to protect a cell that is already urbanised.

When an inconsistency occurs, the candidate cell to be protected is rejected and no reward is added to the final satisfaction. Its budget, in turns, is stored for future purchases. As a conclusion, for every inconsistency found in the GA solution, the algorithm reduces its final quality. The behaviour of the inconsistencies Fig. 5 shows that it is not necessary to gather a huge amount of data in order to achieve consistent results. It is noticeable that the model does not achieve a non-inconsistency state even if the number of simulations where the data is gathered increases. This is a consequence that the future cannot be completely predicted.

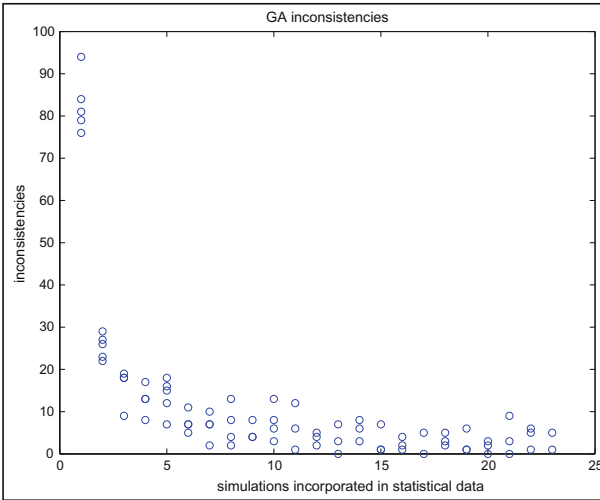


Fig. 5. Urban inconsistencies found in the test of the GA solutions.

7 Conclusions and Future Work

This paper reports results from a proof-of-concept study, which show that statistical model emulator can be used for policy optimisation. In particular, we show how we can capture and represent uncertainty in ABM using data from simulated runs and find optimal urban planning policies with the use of GA.

The strategy is tested in a monocentric urban model where the main objective of the experiment is to distribute a set of green protected areas throughout the

lattice with the goal of achieving the maximum satisfaction from the inhabitants of the city. An individual is considered to be 'satisfied' if a green area is placed close enough to the location of his residence.

The main observation that we draw is that the appropriate prior use of non-optimised simulations was effective in guiding the GA to achieve successful outcomes. The specific approach we took is potentially applicable to a wide range of applications which concern sequential decision making and require time-consuming simulations to evaluate decisions. The results on our case study suggest there is considerable promise in our approach. The ability to successfully address a wider range of optimisation problems of this kind could lead to a new generation of tools for use in urban planning. However, in the meantime, various aspects of the approach need further investigation. Among them are three main directions:

Evaluation of Statistical Simulation-Based Approaches for ABM Optimisation. Related research fields, such as optimisation of natural conversational strategies in human-machine dialogue, make use of similar simulation techniques to approximate real-world behaviour. In the case of spoken dialogue systems, user simulations are built from small data set of real user interactions [7]. In future work, we want to explore how evaluation techniques for user simulations can be applied to estimate the quality and policy impacts of our ABM simulations.

Improving GA to Include Uncertainty for Sequential Decision Making Problems. In the previous experiments we have used a variant of genetic algorithms which does not explicitly encode uncertainty endured by the model environment. In future work, we plan to investigate advanced evolutionary algorithms, such as X Classifier Systems [36] for sequential decision tasks, which explore similarities between evolutionary approaches and Reinforcement Learning.

Improve the Complexity of the Urban Model. In particular, we plan to increment the complexity of our current metric including factors like size of the urban park and quality. We will also develop a new ecological metric based on preserving the ecosystems and conduct experiments to compare the trade-off between our current metric and the new one.

References

1. Vallejo, M., Corne, D.W., Rieser, V.: Evolving urbanisation policies - using a statistical model to accelerate optimisation over agent-based simulations. In: ICAART 2013–5th Conference on Agents and Artificial Intelligence, pp. 1–11 (2013)
2. Forsyth, A., Mussachio, L.: Designing Small Parks: A Manual for Addressing Social and Ecological Concerns. Wiley, New York (2005)
3. Parker, D.C., Hoffmann, M.J., Deadman, P., Parker, D.C., Manson, S.M., Manson, S.M., Janssen, M.A., Janssen, M.A.: Multi-agent systems for the simulation of land-use and land-cover change: a review. *Ann. Assoc. Am. Geogr.* **93**, 314–337 (2003)

4. Sasaki, Y., Box, P.: Agent-based verification of von thünen's location theory. *J. Artif. Soc. Soc. Simul.*, **6** (2003)
5. Sanders, L., Pumain, D., Mathian, H., Gurin-Pace, F., Bura, S.: Simpop: a multi-agent system for the study of urbanism. *Environ. Plann. B Plann. Des.* **24**, 287–305 (1997)
6. Robinson, D., Murray-Rust, D., Rieser, V., Milicic, V., Rounsevell, M.: Modelling the impacts of land system dynamics on human well-being: using an agent-based approach to cope with data limitations in koper, slovenia. *Comput. Environ. Urban Syst. (Special Issue: Geoinformatics)* **2010**(36), 164–176 (2012)
7. Rieser, V., Lemon, O.: Reinforcement Learning for Adaptive Dialogue Systems: A Data-driven Methodology for Dialogue Management and Natural Language Generation. *Theory and Applications of Natural Language Processing*. Springer, Heidelberg (2011)
8. Vallejo, M., Vargas, P., Corne, D.: A fast approximative approach for the vehicle routing problem. In: 2012 12th UK Workshop on Computational Intelligence (UKCI), pp. 1–8 (2012)
9. Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R., Paruelo, J., Raskin, R., Sutton, P., van den Belt, M.: The value of the world's ecosystem services and natural capital. *Ecol. Econ.* **25**, 3–15 (1998)
10. Nowak, D., McPherson, E.: Quantifying the impact of trees: the chicago urban forest climate project. *Unasylva* **44**, 39–44 (1993)
11. Newmann, J.V.: *Theory of Self-Reproducing Automata*. University of Illinois Press, Urbana (1966). (edited by Burks, A.W.)
12. Otter, H.S., van der Veen, A., de Vriend, H.J.: Abloom: location behaviour, spatial patterns, and agent-based modelling. *J. Artif. Soc. Soc. Simul.* **4**, 1147–1152 (2001)
13. Brown, D.G., Robinson, D.T.: Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecol. Soc.* **11**, 46 (2006)
14. Devisch, O., Timmermans, H., Arentze, T., Borgers, A.: An agent-based model of residential choice dynamics in nonstationary housing markets. *Environ. Plann. A* **41**, 1997–2013 (2009)
15. Parker, D.C., Filatova, T.: A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Comput. Environ. Urban Syst.* **32**, 454–463 (2008). (GeoComputation: Modeling with spatial agents)
16. Filatova, T., Parker, D., van der Veen, A.: Agent-based urban land markets: agents' pricing behavior, land prices and urban land use change. *J. Artif. Soc. Soc. Simul.* **12**, 3 (2009)
17. Miller, E.J., Hunt, J.D., Abraham, J.E., Salvini, P.A.: Microsimulating urban systems. *Comput. Environ. Urban Syst.* **28**, 9–44 (2004)
18. Holland, J.: *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor (1975)
19. Darwin, C.: *On the Origin of Species by Means of Natural Selections: or the Preservation of Favoured Races in the Struggle for Life*. Murray, London (1861)
20. Pukkala, T., Kurttila, M.: Examining the performance of six heuristic optimisation techniques in different forest planning problems. *Silva Fennica* **39**, 67–80 (2005)
21. Lu, F., Eriksson, L.O.: Formation of harvest units with genetic algorithms. *For. Ecol. Manag.* **130**, 57–67 (2000)
22. Wu, J., Zheng, C., Chien, C.C., Zheng, L.: A comparative study of monte carlo simple genetic algorithm and noisy genetic algorithm for cost-effective sampling network design under uncertainty. *Adv. Water Resour.* **29**, 899–911 (2006)

23. Rieser, V., Robinson, D.T., Murray-Rust, D., Rounsevell, M.: A comparison of genetic algorithms and reinforcement learning for optimising sustainable forest management. In: *GeoComputation* (2011)
24. Qin, X., Huang, G., Liu, L.: A genetic-algorithm-aided stochastic optimization model for regional air quality management under uncertainty. *J. Air Waste Manag. Assoc.* **60**, 63–71 (2010)
25. Wang, N., Yang, Y.: Target geometry matching problem for hybrid genetic algorithm used to design structures subjected to uncertainty. In: *IEEE Congress on Evolutionary Computation, CEC '09*, pp. 1644–1651 (2009)
26. Wang, N., Zhang, X., Yang, Y.: A hybrid genetic algorithm for constrained multi-objective optimization under uncertainty and target matching problems. *Appl. Soft Comput.* **13**, 3636–3645 (2013)
27. Alonso, W.: *Location and Land Use*. Publications of the Joint Center for Urban Studies. Harvard University Press, Cambridge (1964)
28. Byomkesh, T., Nakagoshi, N., Dewan, A.: Urbanization and green space dynamics in greater dhaka, bangladesh. *Landscape and Ecological Engineering*, pp. 1–14 (2010). doi:[10.1007/s11355-010-0147-7](https://doi.org/10.1007/s11355-010-0147-7)
29. Passow, S.S.: Land reserves and teamwork in planning stockholm. *J. Am. Inst. Plan.* **36**, 179–188 (1970)
30. Giles-Corti, B., Broomhall, M.H., Knuiaman, M., Collins, C., Douglas, K., Ng, K., Lange, A., Donovan, R.J.: Increasing walking: how important is distance to, attractiveness, and size of public open space? *Am. J. Prev. Med.* **28**, 169–176 (2005). (Active Living Research)
31. Plantinga, A.J., Lubowski, R.N., Stavins, R.N.: The effects of potential land development on agricultural land prices. *J. Urban Econ.* **52**, 561–581 (2002)
32. Riley, C.: Comments on mills & evans. In: *Proceedings of seminar on Land Use Regulation*, Lincoln Institute for Land Policy, Cambridge Mass (2002)
33. Tyrväinen, L., Miettinen, A.: Property prices and urban forest amenities. *J. Environ. Econ. Manage.* **39**, 205–223 (2000)
34. Thorsnes, P.: The value of a suburban forest preserve: estimates from sales of vacant residential building lots. *Land Econ.* **78**, 426–441 (2002)
35. Goldberg, D.E.: A note on boltzmann tournament selection for genetic algorithms and population-oriented simulated annealing. *Complex Syst.* **4**, 445–460 (1990)
36. Wilson, S.W.: Classifier fitness based on accuracy. *Evol. Comput.* **3**, 149–175 (1995)