

# A Reinforcement-Learning Application to Biodiversity Conservation in Costa-Rican Forest

<sup>1</sup>Sabbadin, R., <sup>2</sup>D. Spring, and <sup>1</sup>E. Bergonnier

<sup>1</sup>INRA de Toulouse, Unite de Biométrie et Intelligence Artificielle  
BP 52627 - 31326 CASTANET-TOLOSAN cedex - FRANCE

E-Mail: sabbadin@toulouse.inra.fr

<sup>2</sup>Australian Centre for Biodiversity,

School of Biological Sciences, PO Box 18,

Monash University, Clayton, Victoria 3800 - AUSTRALIA

E-Mail: Daniel.Spring@sci.monash.edu.au

**Keywords:** Reserve site selection, Spatial controlled processes, Dynamic programming, Reinforcement learning

## EXTENDED ABSTRACT

Creating networks of protected nature reserves is the primary means of reducing biodiversity loss. The principle focus of the reserve design literature is on determining which sites to reserve to maximise the number of species conserved. A good reservation policy is one that conserves as many species as possible. Until recently, most site selection models have been static: they assumed that sites threatened by development would be reserved immediately after the optimal plan is determined. This rarely occurs in practice, often because there are insufficient funds available when the optimal plan is determined, or because not all sites are immediately available to be reserved. But once postponed, the reservation decisions might never be implemented if the targeted sites become developed first.

Development often takes the form of a contagion process in which the sites most likely to become developed are near sites that have already been developed. This pattern is evident, for example, in Rondonia in Brazil, where the proximity of a site to the nearest deforested area is a strong determinant of the site's probability of being cleared. The strong dependence of site development risk on the site's proximity to cleared forest has also been found in more heavily developed countries such as Costa Rica. Importantly, this "spreading contagion" form of deforestation is found in much of the Brazilian Amazon and other "frontier forests", which are the last remaining large tracts of ecologically intact forest. If such forests are to be conserved in efficient way, it is essential that interdependencies in site development risks be considered. Graph-theoretic models are thus important for addressing ecological management problems in which connectivity among sites and their dynamic properties need to be considered.

Recently, several dynamic reserve site selection models of increasing complexity based on stochastic

dynamic programming (SDP) have been proposed, allowing to model very large spatially explicit problems. Approximate heuristic solution methods and more recently simulation-based Reinforcement Learning (RL), have been developed. Here, we present the first application of this RL method to a (very) large realistic example. Namely, we used available data concerning Costa-Rica forest cover type and changes between 1997 and 2000 and required forest types and minimum areas for the persistence of 90 species of birds and mammals living there, in order to build a large SDP model. The complete model is huge since more than 19000 sites are modelled and the available yearly funding allows to conserve hundreds of hectares each year. Thus even RL method cannot solve such problems.

To overcome the induced complexity, we propose to split the country in different zones containing a few hundreds threatened sites each and in which one site could be reserved each year. In this paper we illustrate our approach on one of these zones, containing 309 sites and 75 species present, 51 of each can be preserved. We now plan to deal with the full scale problem. Indeed, rather than considering completely independent subproblems of tractable sizes we will consider their dependencies and model the problem of efficiently allocating funds for reservation between zones each year.

## 1 INTRODUCTION

The primary means of reducing biodiversity loss is to create networks of conservation reserves. In most cases, the establishment of a reserve network is a gradual, accretive process, comprising a sequence of land acquisitions through time. One of the reasons for this is that not all sites are available for purchase at the same time (Meir et al. (2004)); another reason is that funding for site acquisitions at any given time is insufficient to acquire all sites (Costello and Polasky (2004)). Conservation organisations that build reserve networks over an extended period of time must contend with the risk that sites will be developed before they can be reserved (Costello and Polasky (2004); Meir et al. (2004)).

Recently, dynamic models based on stochastic dynamic programming (SDP) have been used to solve reserve selection problems of this type (Meir et al. (2004); Costello and Polasky (2004); Sabbadin et al. (2007)). In these models, unreserved sites are irreversibly developed each year with a given probability, but only a limited number of sites can be reserved each year, because of budgetary or site-availability constraints. The problem is to design a dynamic reservation policy that results in the maximum expected number of species conserved at the end of the problem horizon. These models are used to investigate the importance of the timing of selections in habitat conservation programs. Most models consider a random development pattern, in which development probability for each site is independent of the development status of neighbouring sites. However, it is more likely that development will occur as a “contagion” process, beginning when new roads make regions economically and logistically accessible (Laurance et al. (2004)).

Sabbadin et al. (2007) propose to consider development as a contact process over a graph, and propose site selection algorithms based on *reinforcement learning* techniques (Bertsekas and Tsitsiklis (1996)). The purpose of this paper is to illustrate the site selection method proposed by Sabbadin et al. (2007) with a case study based on deforestation and birds species requirements data concerning Costa-Rica forests. In a first step, we recall the site selection model of Sabbadin et al. (2007) and the RL-based site selection technique and then we will describe the Costa-Rica case study.

## 2 DYNAMIC RESERVE SELECTION MODEL

The dynamic reserve selection model described in (Sabbadin et al. (2007)) is inspired from the one of Costello and Polasky (2004), but here space is represented explicitly, by a connectivity graph over

sites. Indeed, the considered area where a reserve is to be built is represented by a network (or graph)  $\mathcal{G}$  with  $J$  vertices (the sites which can be selected). A symmetric connectivity matrix  $G(J \times J)$  specifies a neighbourhood relation:  $G(j, j') = 1$  if sites  $j$  and  $j'$  are neighbours, and  $G(j, j') = 0$  if not. Then, we define the neighbourhood of a site  $j$  as  $N(j) = \{j' \in 1 \dots J, G(j, j') = 1\}$ .

At a given time period  $t$ , any site  $j$  can be in one of the three following states : *developed*, *reserved* or *unreserved*. Thus, the state  $S_t$  of sites can be unambiguously described by the means of two of the three vectors  $D_t, R_t, U_t$  where  $D_t(j) = 1$  means that site  $j$  is developed, and  $D_t(j) = 0$  means that it is not.  $R_t$  and  $U_t$  model whether sites are reserved or unreserved. It is clear that for any site  $j$ , exactly one of  $D_t(j), R_t(j), U_t(j)$  equals one, and the two others equal zero. Thus we define  $S_t = (D_t, R_t)$ .

Now, in the site selection model, two distinct elements have to be considered: i) the state dynamics (under a site selection policy) and ii) the impact of this dynamics on the persistence of the species we wish to protect.

Let us first describe the state dynamics model. The state of sites will evolve over time under the influence of two types of factors: controlled and uncontrolled factors.

- Controlled factor. At any time period, it is possible to select one unreserved site for reservation, thus changing its state from *unreserved* to *reserved*.
- Uncontrolled factor. At any time period  $t$ , it is assumed that any unreserved site  $j$  which is not selected for reservation can become developed at the end of the period with a known probability  $p_j$ .

At every time step  $t$ , an action is chosen, consisting in selecting an unreserved site  $a_t \in \{1 \dots J\}$  for reservation ( $U_t(a_t) = 1$ ). Then,  $R_{t+1}(a_t) = 1$  and  $U_{t+1}(a_t) = 0$  (the site becomes reserved). Development occurs concurrently and randomly on non-selected unreserved sites: All unreserved sites at time  $t$  which are not being selected for reservation have a probability to become developed. Here we assume that development is a contagion process : the probability that an undeveloped site  $j$  becomes developed depends on the development status of the *neighbours* of  $j$ ,  $N(j)$ . This is the first use of the neighbourhood relation. More precisely, we assume that we have a model of the development probabilities of any undeveloped site  $j$  given the state of the sites in its neighbourhood:  $p_j(S_{t+1}(j) = D | S_t(j) = U, S_t(N(j)), a_t)$  where  $S_t(N(j)) = \{S_t(j'), j' \in N(j)\}$ .

At this point, it should be noticed that development

and reservation are assumed to be irreversible : a *developed* or *reserved* site remains in this state forever. Thus, the development / reservation process will always end in an *absorbing state* in which no unreserved site persists and the absorbing state will be reached in a number of time steps bounded by the number of sites since at each time step one undeveloped site becomes irreversibly reserved.

Now, let us describe how we model the impact of the state of sites on species existence in a given region. We consider that  $I$  species can live in the region. For each species, only part of the sites form a suitable environment for persistence (food, habitat, etc.). Thus, another matrix,  $M(J \times I)$  specifies which sites are suitable for which species:  $M_{ji}$  equals 1 if site  $j$  is suitable for species  $i$ , and 0 if not.

It is assumed that a species  $i$  can exist in site  $j$  if and only if site  $j$  is not developed (i.e. is in state *reserved* or *unreserved*). In addition, it is often assumed that species need a minimum connected area of suitable environment to persist. Therefore, in our model we consider explicitly the area of the sites ( $A(j)$  is the area of site  $j$ ) and for each species  $i$ , a *minimum required area*  $MRA(i)$ . Species  $i$  will be definitely *protected* when a connected subnetwork of reserved sites of total area at least  $MRA(i)$  is reserved. Note that in our model, the same connectivity matrix  $G$  over sites is used for modelling both the development contact process and the site connectivity relation. However, this is not required in order to apply our model.

The objective of a reserve selection problem is of course to minimise species losses, or equivalently to maximise the number of species present in reserved sites when the process has reached an absorbing state (all sites are either reserved or developed). Our goal is to find a *policy*  $\pi$  assigning to any possible state  $S_t$  a site to reserve. Such a policy should be defined so as to maximise the expected value of the number of species reserved when an absorbing state of the process is reached. From now on, we give up the subscript  $t$  in the notations of the state and action variables for sake of simplicity since i) the process is assumed to be stationary (transitions and rewards do not depend on time) and ii) it can be shown (see Puterman (1994)) that optimal policies are in this case stationary. We finally define a *reward function*  $r(S, a)$  as the number of additional species which are protected when site  $a$  is reserved in state  $S$ .

Now, let us consider a *trajectory*  $\tau$ , that is an alternate sequence of states and actions, starting in an arbitrary state  $S^0$  and ending in an absorbing state  $S^k$  :  $\tau = (S^0, a^0, S^1, a^1, \dots, S^{k-1}, a^{k-1}, S^k)$ . We define the *value*  $V(\tau)$  of such a trajectory  $\tau$  as the number of species eventually protected at the end of

the trajectory. Thus,  $V(\tau)$  is exactly the number of species protected in state  $S^k$ . The following equality can be easily shown, which will be used in the dynamic programming solution method for the reserve selection problem :

$$\forall \tau = (S^0, a^0, S^1, a^1, \dots, S^{k-1}, a^{k-1}, S^k),$$

$$V(\tau) = \sum_{i=0}^{k-1} r(S^i, a^i).$$

A fixed policy  $\pi$  does not define a single trajectory  $\tau$  when applied in a start state  $S$ , but rather a probability distribution over a set of possible trajectories. The value  $V_\pi(S)$  of this policy is defined as the expected number of new species which can be protected by applying  $\pi$ , from start state  $S$  :

$$V_\pi(S) = E[V(\tau)|S, \pi]$$

where  $E[V(\tau)|S, \pi]$  is defined over the set of possible trajectories generated by policy  $\pi$  applied in initial state  $S$ .

There exists, in theory, a policy  $\pi^*$  maximising  $V_\pi$  for all states. However, it is a difficult task to compute it when the number of sites is too high. It soon becomes even impossible to express such a policy in tabular form. Therefore, in the next Section we show how to compute an approximately optimal policy using a *parameterised reinforcement learning algorithm*.

### 3 PARAMETERISED REINFORCEMENT LEARNING SOLUTION METHOD

Reinforcement learning is a set of simulation-based methods which allow for the solution of large-scale Markov Decision Problems (Bertsekas and Tsitsiklis (1996)), such as the reserve site selection problem we are interested in.

When solving a problem expressed as before, the optimal policy  $\pi^*$  can be computed from  $V_{\pi^*}$  as:

$$\pi^*(S) = \arg \max_a \left\{ r(S, a) + \sum_{S'} p(S'|S, a) \cdot V_{\pi^*}(S') \right\} \quad (1)$$

In the RL approach, the optimal value function  $V_{\pi^*}$  is approximated by a (linear) parameterised value function  $V_{\varepsilon^*}$  which is computed through repeated simulations of trajectories. Then, a policy  $\pi_{\varepsilon^*}$ , is computed from  $V_{\varepsilon^*}$  using Eq. (1) where  $V_{\pi^*}$  is replaced with  $V_{\varepsilon^*}$ .  $\pi_{\varepsilon^*}$  approaches the optimal policy  $\pi^*$ . One further problem is that computing  $\pi_{\varepsilon^*}$  from Eq. (1) is itself a difficult task, since the number of sites  $S'$  that can be immediately reached from  $S$  can be huge. So, that computation is often performed by sampling over the possible  $S'$ , using the transition probability  $p$ .

In the following we describe how a parameterised linear approximation of the optimal value function is computed, and how to compute *online* the next site to reserve in the current state of the region.

For very large SDP problems, such as the reserve selection problem, when the number of sites is large, it is not convenient to compute the exact optimal value function  $V_{\pi^*}$  in tabular form. It may be more reasonable to look for an approximate, parameterised, value function  $V_{\varepsilon^*}$ , which can be expressed much more concisely than  $V_{\pi^*}$  itself. A linear approximation of  $V_{\pi^*}$  is often used, searched for in the set of parameterised value functions of the form

$$V_{\varepsilon}(S) = \varepsilon(1)\psi_1(S) + \dots + \varepsilon(k)\psi_k(S).$$

The  $\varepsilon(i), i \in \{1, \dots, k\}$  are parameters and the  $\psi_i$  are arbitrarily given real-valued functions called *features* (Bertsekas and Tsitsiklis (1996)).

Then, the objective of feature-based reinforcement learning algorithms is to compute a parameters vector  $\varepsilon^*$  such that  $V_{\varepsilon^*}$  is a reasonable approximation of  $V_{\pi^*}$ . The general way is to use a simulation of the controlled process in order to compute a sequence of parameters vectors  $(\varepsilon_n)$ , in the form

$$\varepsilon_{n+1} = \varepsilon_n + \Delta(S_n, a_n, S_{n+1}, r(S_n, a_n))$$

where  $\Delta(S_n, a_n, S_{n+1}, r_n(S_n, a_n))$  is a correction factor computed from the output of the current simulation trial. The most common implementation of the above principle is the *gradient descent* method, where updates take the form:

$$\varepsilon_{n+1} = \varepsilon_n + \alpha_n(R_n - V_{\varepsilon_n}(S_n))\nabla_{\varepsilon_n} V_{\varepsilon_n}(S_n).$$

where  $R_n$  is a direct estimation of the value of  $V_{\pi^*}$  drawn from the current trial and past experience. A simple such estimation consists in using

$$R_n = \max_a r(S_n, a) + V_{\varepsilon_n}(S_{n+1}).$$

Now, one simple case of *features* is of particular interest (Tsitsiklis and Van Roy (1996)). This is the one in which the  $\psi_n$  take their values in the set  $\{0, 1\}$ . In this case, the above equation simplifies into

$$\forall i \in 1, \dots, k \varepsilon_{n+1}(i) = \varepsilon_n(i) + \alpha_n(\max_a \{r(S_n, a) + V_{\varepsilon_n}(S_{n+1})\} - V_{\varepsilon_n}(S_n))\psi_i(S_n).$$

This will be the parameters update function which we will use to solve the reserve selection problem.

Now, let us describe the approximation structure in the reserve selection problem. In this problem, we choose to use the following  $J$  features:  $\psi_i(S) = 1$  if site  $i$  is not developed, and  $\psi_i(S) = 0$  if it is developed, for  $i = 1 \dots J$ . Thus,  $\psi_i(S) = \psi_i(S(i))$  only depends

on the state of site  $i$ , and not on the global state of the problem. In addition, concerning the learning rate parameter  $\alpha_n$ , which should simply decrease to 0 as  $n$  grows, we choose to define it as  $\alpha_n = \frac{1}{n(a)}$ , i.e. the number of time site  $a$  has been reserved so far, during the learning phase of the algorithm.

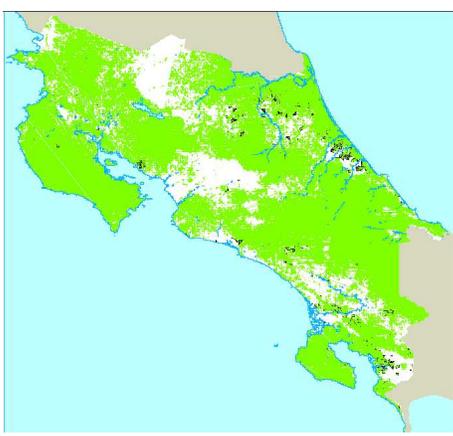
## 4 A RESERVE SELECTION PROBLEM IN COSTA RICA

In the remaining of the paper we show an implementation of the RL approach described above to a reserve design case study concerning Costa Rican forests. This case study is built from two sets of data. The first is a GIS data set used for building a deforestation model over the Costa Rican forest, and the second comprises the life traits of tenths of birds species living in this forest. We next describe the GIS data set which was available and show how it was used to build the contagion deforestation probabilities used in our RL model.

### 4.1 Building the deforestation probabilities

Development often takes the form of a contagion process in which the sites most likely to become developed are near sites that have already been developed. This pattern is evident, for example, in Rondonia in Brazil, where the proximity of a site to the nearest deforested area is a strong determinant of the site's probability of being cleared (Arellano-Neri and Frohn, 2001; F.J.B. and J.P., 2007). The strong dependence of site development risk on the site's proximity to cleared forest has also been found in more heavily developed countries such as Costa Rica (Robalino and Pfaff, 2005). Importantly, this "spreading contagion" form of deforestation is found in much of the Brazilian Amazon (Laurance et al., 2004) and other "frontier forests", which are the last remaining large tracts of ecologically intact forest. If such forests are to be conserved in efficient fashion, it is essential that interdependencies in site development risks be considered. Green et al. (2005) identified graph-theoretic models as an important method for addressing ecological management problems in which connectivity among sites and their dynamic properties need to be considered. We use a GIS data set comprising 4 layers (Robalino and Pfaff (2005)), that describes the changes of forest cover in Costa Rica between years 1997 and 2000, to estimate the probabilities involved in the contagion process:

- 1) Forested area 1997 : 15856 polygons showing forested areas.
- 2) Area deforested between 1997 and 2000 : 3261 polygons.
- 3) Reserved areas : 32 (large) polygons.
- 4) Region types : 12 different types (mountain-wet, mountain-dry, etc.).



**Figure 1.** Deforested area between 1997 and 2000.

Figure 1 shows forested areas in 2000 (in green), as well as sites deforested between 1997 and 2000. By merging the four layers we were able to draw a map comprising 19165 polygons, each representing one site. For each site, in addition to its geometry, we know its status (R, D or U), the type of region it is in (mountain, lowland, wet, dry...) and its area. Note that the sites are of very different sizes, varying from a few hectares for many of them, to more than 100000 ha for a few of them.

We assumed in our model that the deforestation probabilities for a given undeveloped site were functions of site-specific parameters and of its neighbouring environment. More precisely, in order to measure the effect of neighbour sites, we draw a connectivity graph between sites. For each site we draw a disc, which centre is the centre of gravity of the polygon representing the site and which surface is 1.5 that of the polygon. Two sites are neighbours if their associated disks intercept. Using this method, the average number of neighbours for a given site is 29.5. However, large sites can have up to 3142 neighbours, while half of the sites have at most 18 neighbours. From this graph and other GIS data we derive a logistic model of the deforestation probability of a given (U)ndeveloped site  $i$ :

$$\text{logit}(p_i) = \alpha_1 A_i + \alpha_2 DB_i + \alpha_3 N_i + \beta(RT(i)) \quad (2)$$

$A_i$  is the area (in ha) of site  $i$ .  $DB_i$  is the distance of site  $i$  to the closest “big site”. A site is considered “big” if its size exceeds 1000 ha (there are 158 such sites in the map). The big sites were found to have a very low probability to become developed, and their proximity was shown to influence a site’s development probability.  $N_i$  is a variable measuring the potential impact of the developed neighbours of site  $i$  on its probability of development:

$$N_i = \sum_{j \in N(i), S_j = D} \log_{10} A_j.$$

$N_i$  is thus a function of the state of the neighbours of  $i$ . It represents the contribution of the *contagion effect* to the development probability of an undeveloped site. Finally,  $\beta$  is a vector of 12 parameters, one for each region type ( $RT(i)$  is the region type of site  $i$ ).

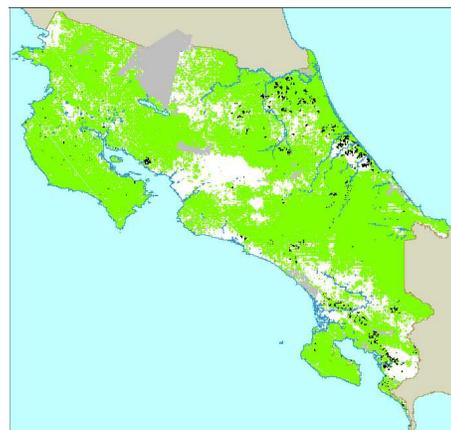
The parameters  $\alpha_1, \alpha_2$  and the additional parameter vector  $\beta$  were adjusted by logistic regression by comparing the sites which were undeveloped in 1997 and became developed in 2000 with the ones which remained undeveloped. Note that in our study we consider a time step of 3 years between each reserve selection decision is implemented.

The results of the logistic regression are summarised in Table 1. The significance of the parameters was measured by the range of the development probabilities, marginalised over the different factors. The third and fourth columns in Table 1 give those ranges, for the observed values of the parameters. These probabilities show that all parameters have an influence on development probabilities, and should be taken into account.

**Table 1.** Deforestation probabilities estimation.

Variable	Coef.	$P_{min}$	$P_{max}$
$N_i$	$-3.9 \times 10^{-4}$	0.00	0.07
$DB_i$	$2.12 \times 10^{-5}$	0.07	0.16
$\beta(RT(i))$	0.2 to 14.8	0.00	0.18
$A_i$	$-2.5 \times 10^{-6}$	0.00	0.007
Overall		0.00	0.22

Figure 2 shows a map obtained after simulation of the deforestation process, given the above described model, after a ten steps run (30 years). Over that period, 8380 sites, totalling 65000 ha (0.35% of the forested area) were developed. This is in line with the current estimates of the deforestation rates in the literature (2000 to 3000 ha/y).



**Figure 2.** Deforested area after a simulated 30 years scenario.

## 4.2 RL reward model

In the Costa Rica reserve selection problem, expert knowledge on birds species requirements was available, summarised in an excel file indicating for each species the total (connected) area which has to be reserved in suitable regions for guaranteeing its persistence. In addition, for each species is indicated which types of regions are suitable. Table 2 shows an extract of the file.

**Table 2.** Area required (in ha) and suitable area type : PMR = pre-mountain, L = lowland, R = rainforest, W = wetland.

Name	min area	requirement
Orange-billed Sparrow	19	PMR
Brown-billed Scythebill	21	PMW or PMR
Olive Tanager	19	PMR
Azure-hooded Jay	17	LMR

From this table, we computed for each species and after each simulation step, the area of the largest connected network of suitable reserved sites remaining (using the same connectivity function as for the contagion process). When this area meets the requirement (column one, in hectares), the species is considered as protected and a reward of 1 is received. Thus, after a given simulated trajectory, the total reward equals the number of birds species protected in the corresponding reserve network.

## 4.3 RL solution for the reserve selection problem

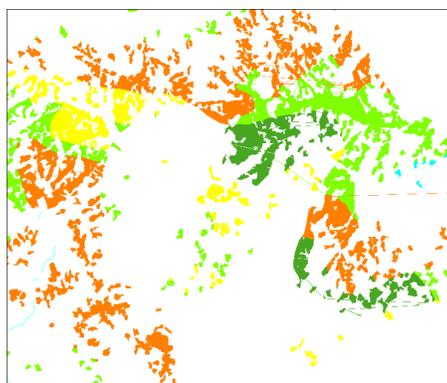
Our objective was to run the RL algorithm described in the previous section to the whole Costa Rica problem. However, it became quickly obvious that the method was not applicable to problems comprising more than 1000 sites. Furthermore, our RL algorithm assumes that one site is reserved at a time, which is not realistic in the Costa Rica problem in which enough money is available for reserving a dozen of mean size sites during each time step. For these reasons, we chose to split the problem by considering several small (with a few hundred mean size sites) regions, and looking for independent policies within each region. A further advantage of using this *divide and conquer* strategy is that it favours conservation of birds species all over the Costa Rica territory, and not only locally, making persistence more robust.

We next show the results obtained within one such small region comprising 309 sites (Figure 3 shows where the region is located and Figure 4 shows the detail of the considered region). In this region, 85 sites are not suitable for any bird species. Overall, 75 species could live in that region, however the



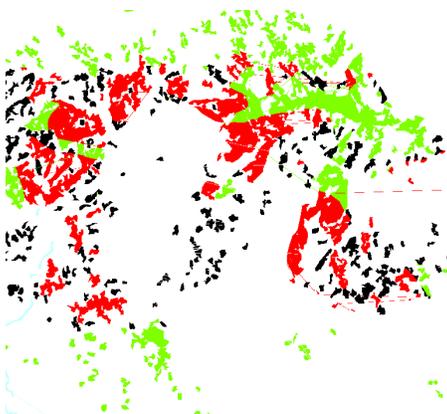
**Figure 3.** Area selected for RL illustration.

minimum area requirements are met for only 54 species. Thus, any site selection policy can at best protect 54 species. The RL algorithm converged to a



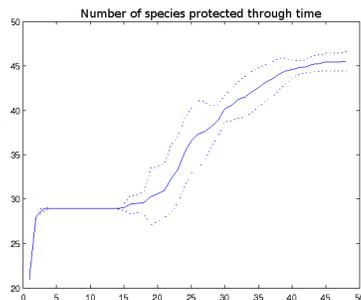
**Figure 4.** Detail of the selected area. Colours represent region types.

vector of values  $\epsilon^*$  within a few hours.  $\epsilon^*$  indirectly represents a site selection policy. We then evaluated this policy by simulation. The result of one simulation run (50 3-years steps) is represented in Figure 5.



**Figure 5.** One 150 years simulation run. Black: deforested, red: reserved, green: unreserved.

We also evaluated our algorithm by plotting the number of species protected over time, averaged over ten runs (Figure 6), together with the mean squared error. Note the very specific form of this curve. In the first steps of the simulation, sites are selected greedily with respect to the number of species which can be immediately protected. These species which have a low minimum area requirement are protected first as they would be, using a greedy policy. Then, after 3 or 4 steps, there only remain species needing a large area to become protected. Several sites have to be selected before any additional species is protected. This corresponds to the plateau in Figure 6, during which the number of protected species does not change. In the mean time, sites may become (randomly) developed, and the effect of this random development will only be seen later (steps  $\geq 15$ ) when the total reserved area is sufficient to protect new species. It is only at that time that the curves for the various runs diverge. Note that for a greedy strategy the initial rise of the curve would certainly be similar, however the plateau would be longer and the last part of the curve would lie below the one represented here.



**Figure 6.** Number of species protected over time.

## 5 CONCLUDING REMARKS

We have presented here an illustration of the RL method proposed by Sabbadin et al. (2007) for solving reserve site selection problems. Our point was mainly to assess the efficiency of the method and its scalability. In a first step we designed an automated method, based on GIS data exploitation for building a model of the development dynamics. We then collected from the literature some data about species requirements, which we used to build an immediate reward function. The scalability of the method was then assessed and it turned out that we could solve efficiently problems comprising several hundred sites and for which one site only could be selected at a time.

In order to solve really large scale problems (thousands of sites, dozens being reserved each time step), which is our objective, we have two

options at hand: i) either grouping sites together, thus considering more *coarse grained* problems, with fewer sites and fewer site selection options and ii) adopt a *divide and conquer* strategy, by considering separately in parallel several subregions between which funding is shared once and for all and independent site selection policies are computed. We are currently exploring both options, but maybe in the future a *hierarchical approach*, common in Artificial Intelligence problems would be the best: solving a simplified coarse grained problem for deciding on global funding allocation between subregions and then solving in parallel several corresponding subproblems, using the approach exposed here.

## 8 REFERENCES

- Arellano-Neri, O. and R. Frohn (2001). Image-based logistic regression parameters of deforestation in rondonia, brazil. In *Geoscience and Remote Sensing Symposium*, Volume 5, pp. 2236–2237.
- Bertsekas, D. P. and J. N. Tsitsiklis (1996). *Neuro-Dynamic Programming*. Belmont, Massachusetts: Athena Scientific.
- Costello, C. and S. Polasky (2004). Dynamic reserve site selection. *Resource and Energy Economics* 26, 157–174.
- F.J.B., O.-F. and M. J.P. (2007). Thresholds inlandscape structure for three common deforestation patterns in the brazillian amazon. *Ecology Landscape* 7, 1061–1073.
- Green, J., A. Hastings, P. Arzberger, F. Ayala, K. Cottingham, F. Cuddington, F. Davis, J. Dunne, M.-J. Fortin, L. Gerber, and M. Neubert (2005). Complexity in ecology and conservation: Mathematical, statistical and computational challenges. *Bioscience* 55, 501–510.
- Laurance, W., A. Albernaz, P. Fearnside, H. Vasconcelos, and L. Ferreira (2004). Deforestation in amazonia. *Science* 304(5674), 1109–11.
- Meir, E., S. Andelman, and H. Possingham (2004). Does conservation planning matter in a dynamic and uncertain world? *Ecology Letters* 7, 615–622.
- Puterman, M. L. (1994). *Markov Decision Processes*. New York: John Wiley and Sons.
- Robalino, J. and A. Pfaff (2005). Deforestation is contagious evidence of spatial interactions from forest clearing in costa rica. (Unpublished).
- Sabbadin, R., D. Spring, and C. E. Rabier (2007). Dynamic reserve site selection under contagion risk of deforestation. *Ecol. Modelling* 201(1), 75–81.
- Tsitsiklis, J. N. and B. Van Roy (1996). Feature-based methods for large-scale dynamic programming. *Machine Learning* 22, 59–94.