Ecological Modelling 351 (2017) 77–86

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Time to fly: A comparison of marginal value theorem approximations in an agent-based model of foraging waterfowl

Matt L. Miller[a,*], Kevin M. Ringelman[b], John M. Eadie[c], Jeffrey C. Schank[a]

[a] University of California, Psychology Department, One Shields Ave., Davis, CA 95616, United States
[b] Louisiana State University, School of Renewable Natural Resources, LSU Agricultural Center, Baton Rouge, LA 70803, United States
[c] University of California, Wildlife, Fish, and Conservation Biology, One Shields Ave., Davis, CA 95616, United States

ARTICLE INFO

Article history:
Received 17 August 2016
Received in revised form 13 February 2017
Accepted 14 February 2017

Keywords:
Optimal foraging theory
Decision-making
Algorithm
Simulation

ABSTRACT

One of the fundamental decisions foragers face is how long an individual should remain in a given foraging location. Typical approaches to modeling this decision are based on the marginal value theorem. However, direct application of this theory would require omniscience regarding food availability. Even with complete knowledge of the environment, foraging with intraspecific competition requires resolution of simultaneous circular dependencies. In response to these issues in application, a number of approximating algorithms have been proposed, but it remains to be seen whether these algorithms are effective given a large number of foragers with realistic characteristics. We implemented several algorithms approximating marginal value foraging in a large-scale avian foraging model and compared the results. We found that a novel reinforcement-learning algorithm that includes cost of travel is the most effective algorithm that most closely approximates marginal value foraging theory and recreates depletion patterns observed in empirical studies.

1. Introduction

Optimal patch selection describes how foragers should be distributed across heterogeneous landscapes with respect to food items and to each other. Under simplifying assumptions, optimality theory predicts that foragers consuming continually-replenishing resources might select patches according to the ideal free distribution (Fretwell and Lucas, 1969), in which animals distribute themselves in the patches proportionate to the gain rate of resources. However, for many animals, patch selection is dynamic: when resources are depleted in one area, animals must find new patches containing those resources. In 1976, Charnov proposed marginal value theorem (MVT), an analytical solution for determining when to leave a foraging patch that predicts that a forager should depart when the intake rate for that patch falls below the long-term average intake rate across all available patches. Charnov's theoretical result sparked a flurry of empirical and modeling studies that suggested other patch-departure rules including fixed-time, fixed-intake, minimizing, maximizing, and inter-reward intervals (e.g., Cowie, 1977; Hodges, 1981; Iwasa et al., 1981; McNair, 1982; Green, 1984; McNamara and Houston, 1985).

There remains some uncertainty as to how well any animal fits the predictions of MVT and whether any animal could actually gather the information required to implement MVT (Stephens and Krebs, 1987; Stephens et al., 2007). Nonetheless, animals likely use some foraging strategy to be able to thrive in their environment, whether through evolution, learning, development, or some combination of these processes. We may expect such strategies to approximate MVT.

Since MVT has broad application, including agent-based models (ABMs) that explore how individual behaviors scale up to create patterns at larger scales (Grimm et al., 2005; DeAngelis and Mooij, 2005; McLane et al., 2011) and non-biological systems such as robot foraging (Ulam and Balch, 2004), finding the best algorithmic approximation for MVT is of critical importance for applied foraging contexts. In this paper, we explore several avenues by which behavioral and biological researchers can optimally program foraging agents that approximate MVT, using overwintering waterfowl as a test species. In particular, we are interested in which algorithms provide the longest survival times in an environment with depleting resources. We examine two approaches suggested in the robotics literature, subsequently modified to account for

* Corresponding author.
E-mail addresses: hzmiller@ucdavis.edu (M.L. Miller), kringelman@agcenter.lsu.edu (K.M. Ringelman), jmeadie@ucdavis.edu (J.M. Eadie), jcschank@ucdavis.edu (J.C. Schank).

http://dx.doi.org/10.1016/j.ecolmodel.2017.02.013
0304-3800/

Table 1
Abbreviations used in this paper.

Abbreviation	Meaning
ABM	Agent-based model
DT50M	Days to 50% mortality
DTD	Days to deficit
GUD	Giving-up density
IFD	Ideal free distribution
LTFA	Long-term-forward average
MVT	Marginal value theorem
OMVT	Online MVT algorithm
RLMVT	Reinforcement-learning MVT algorithm
XOMVT	Extended OMVT algorithm
XRLMVT	Extended RLMVT algorithm

travel time, and examine how these different approximations to MVT affect energy intake and survival in our modeled system.

1.1. Marginal value theorem

We consider three different implementations of marginal value theorem in this paper: Charnov's (1976) original MVT, or *classical MVT*; an algorithm approximating MVT based on reinforcement learning (Wawerla and Vaughan, 2009), or *reinforcement-learning MVT*; and an algorithm approximating MVT using continuously-updated estimates (Wawerla and Vaughan, 2010), or *online MVT* (abbreviations are summarized in Table 1). In the latter two cases, we consider both the original techniques suggested by Wawerla and Vaughan as well as modified versions of these methods that take into account differential costs of travel between foraging areas, or *extended reinforcement-learning MVT* and *extended online MVT*.

1.2. Classical MVT

In 1976, Charnov proposed a mathematical model for the amount of time a forager should remain in a patch of a given quality. If we consider that patches can be assigned a *patch type* based on patch quality, identified by p, and a forager has a net gain in a patch of type p of $g_p(T_p)$ if it spends T_p amount of time in that patch type, then the marginal rate of intake as length of time in the patch type increases is

$$\frac{\partial g_p\left(T_p\right)}{\partial T_p}. \tag{1}$$

(See Table 5 for notation). Similarly, if we are given the travel time between patches, t', and the cost of travel E_{travel}, we can calculate the net intake rate for each patch type,

$$\frac{g_p\left(T_p\right) - t'E_{travel}}{t' + T_p}. \tag{2}$$

If we know the proportion of each patch type in the environment, π_p, we can then calculate the average net intake rate for the whole environment,

$$\frac{\sum_p \pi_p g_p\left(T_p\right) - t'E_{travel}}{t' + \sum_p \pi_p T_p} \tag{3}$$

Charnov showed that T_p is optimized when the marginal rate was equal to the average net intake rate for the environment. That is, more time spent in the patch would be wasted effort and less time would fail to exploit useful resources.

Calculating this rate requires omniscient knowledge of patch quality across the environment. The gain function g must be sufficiently characterized to calculate its rate of change as time-in-patch increases, and for heterogeneous patches, g must be characterized for all patches in the environment. Even in the case of a single forager it is uncertain that the gain function can be adequately characterized (see, for example, Stephens and Krebs, 1987) because average intake rate depends on the time in patches, and time in patches depends on average intake rate. With multiple foragers and exploitation competition, both marginal and net intake rates change as a function of the number of foragers in the patch; an ideal forager would not only resolve its own circular dependency in MVT, but would also have to solve it for every other forager. Ignoring the fact that animals cannot gather this information in the real world (Bartoń and Hovestadt, 2013), the problem of circularity in applying MVT has suggested that computation even with perfect information is intractable (Wawerla and Vaughan, 2010). These difficulties in applying MVT are not surprising since MVT was derived as a theoretical optimum that behavior might approach in the limit, not as a strategy for decision-making and not under real-world conditions such as resource competition and depletion (Stephens and Krebs, 1987; Wajnberg et al., 2006).

1.3. Reinforcement-learning MVT

Because of these difficulties in developing optimal foraging algorithms that satisfy MVT, algorithms that approximate MVT have been proposed to determine time in patches, often relying on very simple rules (Gibb, 1958; Krebs, 1973; Krebs et al., 1974; McNamara, 1982). One of the more promising approaches was proposed by Wawerla and Vaughan (2009) who estimated the optimal patch departure time by simulating reinforcement learning, which we refer to as *reinforcement learning MVT*, or RLMVT.

RLMVT is based on an approach to simulating reinforcement learning by implementing the *n-armed bandit* using *softmax* to overcome local optima (Sutton and Barto, 1998; see Supplemental material C.1). This algorithm can be used to optimize foraging behavior. Consider that the essential problem in implementing MVT is determining the energy gain rate at which the agent should switch patches (the *switching threshold*). An agent can choose to switch at too high a rate, which will result in spending too little time in any given patch; it can choose to switch at too low a rate, which will result in spending too much time in any patch; or it can choose the optimal rate, which should converge with the predictions of MVT. If we consider each of the switching thresholds as one of the slot machine's n arms and we equate the net gain across all patches using that switching threshold as its reward, we can use the n-armed bandit algorithm to find the optimal rate at which to switch patches under MVT. That is, the expected reward for a given switching threshold (corresponding to an arm) is the average net gain rate experienced in patches for which that threshold was used. The softmax algorithm helps greedy optimization systems like the n-armed bandit from settling into local basins of attraction by making choices that currently appear less than optimal more likely to be selected; it also prevents the expected reward values for less-likely thresholds from becoming outdated by allowing them to be selected occasionally (and thus updated) throughout the simulation. Softmax uses a parameter called *temperature* (τ), in which higher temperatures make less-optimal-appearing choices increasingly likely, going to equal probabilities for all choices regardless of expected reward at infinitely-high temperatures (see Supplemental material C.1 for softmax details).

There are two complications: (1) switching threshold is a continuous variable, while the arms on the slot machine represent discrete values; and (2) the optimal switching threshold may change as food resources are depleted. Wawerla and Vaughan (2009) conceptualized this mapping between MVT and the n-armed bandit and addressed the first complication by generalizing Sutton and Barto's algorithm for a continuous action space (detailed in Supplemental material C.1). The second complication arises as a result of agents foraging within a finite environment: classical MVT

assumes an infinitely large landscape in which individual patch depletion does not change the overall long-term average gain rate (Charnov, 1976). In real environments, depletion over time continuously lowers the average gain rate if there is no replenishment, and thus the optimal switching threshold also decreases. This can be accounted for implicitly by setting τ to a reasonably high level, which increases the chances of testing multiple switching thresholds regardless of expected reward.

Using RLMVT, a foraging agent can determine and track the switching threshold that provides it with the highest long-term gain rate across encountered patches. The agent accomplishes this empirically without categorizing patches, determining average densities across the environment, or making explicit corrections for resource competition or depletion. The agent simply forages in a patch until the gain rate drops to the level that the agent has learned will optimize its overall gain if it switches patches.

1.4. Online MVT

Wawerla and Vaughan (2010) sought to improve their prior reinforcement-learning technique with a continuous estimation algorithm we refer to as *online MVT*, or OMVT. The authors hoped to overcome the requirement for startup initialization (during the first n steps) and inefficiencies of the softmax algorithm (see Supplemental material C.1). To this end, they devised a method that estimates the long-term average gain rate by continuously updating that estimate from ongoing foraging experience (that is, *online* estimation).

Using OMVT, the foraging agent keeps track of the gain rate it experiences over time and the time it takes the agent to switch patches on average. From this information, the agent develops an estimate of the average gain rate across patches in its environment (see Supplemental material C.2 for algorithm details). The agent compares its current gain rate to its estimate of the average gain rate and switches to a new patch when the current rate falls below the estimated average rate.

1.5. Extended MVT methods

The above algorithms above do not include the costs of traveling between patches; indeed, travel cost is explicitly omitted in OMVT (Wawerla and Vaughan, 2010). However, the opportunity cost of travel time is still included by its effect on gain rate. In Wawerla and Vaughan's research domain (addressed to robotics researchers), the energetic cost of traveling is not accounted for as part of the foraging model since those robots were not foraging for energy. However, in domains in which energetic costs must be minimized within a finite time horizon, the cost of travel due to patch-switching decisions may become important, since any strategy that increases how often an agent switches patches will ultimately lead to an increase in number of switches per item retrieved, and thus increased cost for that strategy.

In animal foraging models, one must consider the cost of switching patches to obtain more food; thus, the *net energy gain* that allows for the cost of traveling to a new patch must be used to optimize foraging behavior. Furthermore, unlike robots, most foragers are limited in the amount of foraging they can accomplish in one day due to limits in food intake and daily behavioral patterns such as sleep. Because of these constraints, when resources are plentiful (for example, early in the foraging season), foragers may not ever reach the optimal switching threshold even in a single patch before they are done foraging for the day. When this is combined with central-place foraging behavior, which requires the forager to return to a central location at the end of a foraging bout then travel back to a foraging patch at the beginning of the next bout, the energetic cost of switching patches may assume an even larger role in decision-making, especially as patches become depleted. Since these constraints were not applicable to Wawerla and Vaughan's (2009, 2010) research question, their algorithm was explicitly simplified to exclude consideration of patch-switching cost.

However, as we are considering the behavior of animals like waterfowl, which are central-place foragers with dietary and behavioral limits to foraging behavior, we extend RLMVT and OMVT to include the cost of switching between patches.

The adjustments required are simple. For extended RLMVT (XRLMVT), we simply include the cost of travel to the patch when calculating the achieved gain rate for each patch, μ (see Supplemental material, Eq. C3):

$$\mu = \frac{f - t' E_{travel}}{t' + t}, \tag{4}$$

where f is the amount of food obtained, E_{travel} is the energetic cost of inter-patch travel, t' is the time spent traveling to the patch, and t is the amount of time spent foraging in the patch.

For extended OMVT (XOMVT), we adjust the total sum of previous gains, G, used to calculate the estimated long-term gain rate (see Supplemental material, Eq. C11). In OMVT, when the agent is in patch p,

$$G = \sum_{q=1}^{p-1} g_q\left(T_q\right). \tag{5}$$

where $g_q(T_q)$ is the gain in patch q when a forager spends time T_q in that patch. For XOMVT, this becomes

$$G = \sum_{q=1}^{p-1} g_q\left(T_q\right) - t'_q E_{travel}, \tag{6}$$

where t'_q is the time taken to switch from patch q-1 to patch q.

Note the effects of these small changes on how each model behaves. In the case of RLMVT, the extended model decreases the expected reward of a given switching threshold; thus, high thresholds that lead to more frequent switching will be penalized. In the case of OMVT, the reduction in G, which is in the numerator of the estimated average gain rate, leads to a decrease in that rate; thus, the forager will stay in the patch longer since lower rates of return will be tolerated.

These slight modifications increase the biological validity of the model. We hypothesized that they would also provide better decision-making for our agents, leading to more effective foraging on a biologically-realistic landscape.

2. Methods

To assess the best decision-making algorithm, we operationalized foraging effectiveness in two ways: food acquisition efficiency and survival time. Food acquisition efficiency was assessed by days to deficit (DTD), the number of days elapsed before the average forager expended more energy than it acquired (based on a three-day rolling average to prevent early detection resulting from stochastic behavior). Survival time was assessed by the number of days to 50% mortality (DT50 M), when half the population had died due to starvation. DTD and DT50 M provided an absolute measure of the algorithm's effectiveness. In particular, longer survival times (DT50 M) given the same amount of food in the environment indicate more efficient use of those resources.

Simulation of foraging was conducted using a pre-existing model of waterfowl foraging behavior, SWAMP (Miller et al., 2013). This agent-based model (ABM) was created as a conservation management decision support tool for mixed wetland-agricultural

areas. The version of SWAMP used for this study minimized topological differences in the foraging landscape to regularly arranged squares with varying total food densities on patches of the same type, as required to test MVT behavioral rules. A brief synopsis of the model with flowcharts is provided in Supplementary material A. Key elements of the model are described below.

The ABM described by Miller et al. (2013) was augmented for this study with a component that allowed each forager to make patch-switching decisions during foraging bouts using one of the MVT approximations listed above, either RLMVT, OMVT, XRLMVT, or XOMVT. When this component indicated that the forager should switch patches, the forager randomly picked one of the nearest patches to the current patch and moved to it. A foraging bout encompassed all of the temporally-contiguous foraging behavior for a single day (that is, there was effectively only one bout per day) regardless of the number of patches visited; foraging bouts were delimited by return to the forager's home refuge. Patches that the forager had already visited during the current foraging bout were excluded from consideration for switching until a later bout (that is, the forager had to return to its home refuge before reconsidering a patch that had already been visited).

For both reinforcement learning approximations (RLMVT and XRLMVT), all fixed parameters in the *n*-armed bandit with softmax apparatus were held equal: foragers used 5 bins (n in the equations above) and the softmax temperature was 15.0 (τ). Other parameters not covered in the main text are described in the online supplement (Appendix A).

In addition to the MVT approximations, two alternatives were used as bracketing comparisons. In the *no optimization* condition, foragers stayed in the patch to which they were randomly assigned at the beginning of a foraging bout for the duration of that bout (the No MVT/Random condition). Assignment was proportional to patch area; since patch areas were equal in this model, each patch had an approximately equal number of foragers assigned to it at the beginning of the foraging bout. In the *ideal free distribution* condition, foragers also remained within the same patch throughout a given foraging bout; however, foragers were assigned to patches at the beginning of a bout using optimal matching based on the ideal free distribution (IFD) theory (the No MVT/IFD condition). IFD, first proposed by Fretwell and Lucas (1969), suggests that foragers will distribute themselves to match available resources so that depletion due to foragers will be proportional to food availability. In this way, denser food patches will deplete more rapidly so that patches of all qualities will be effectively exhausted at the same time despite their original differences in food availability (Sutherland, 1996; p. 36–37). Like classical MVT, however, IFD requires that foragers be omniscient in regards to the distribution of food in their environment and the behavior of other foragers, a biologically unreasonable assumption (Pierce and Ollason, 1987). (IFD also has other limitations, e.g., Bernstein et al., 1988 and Bernstein et al., 1991; though these limitations do not affect its usefulness as a comparison technique in these simulations.) Unlike MVT, IFD is computationally tractable and thus serves as a reasonable model with which to compare the various algorithms. We simulate IFD by stochastically distributing foragers proportional to the current food density on patches at the beginning of each foraging bout; that is, in IFD simulations, agent patch selection is optimized with respect to resources and other foragers on a daily basis.

A total of 80,000 agents were modeled on a simulated 10 km by 10 km landscape; the number of agents was based on scaling the target ecosystem (see below) down to the modeled area. Simulated time steps were a maximum of 15 min. Food availability on the landscape was varied systematically, starting at maximum values replicating a well-studied waterfowl foraging ecosystem, California's Central Valley (Central Valley Joint Venture, 2006): forageable landscape at 35% of the landscape, composed of 80% commercial rice field and 20% moist soil wetland habitat. We varied these proportions to produce foraging landscapes of different abundance (see Table 2 for quantitites). These landscape variations were tested to determine if the comparative efficacy of foraging strategies differed in more or less abundant landscapes. Simulations ended when all agents ran out of energy reserves.

Table 2
Foraging acreage scenarios.

Moist Soil Acreage (ha)	Rice Field Acreage (ha)	Total Forageable Acreage (ha)
700	2700	3400
1400	2000	3400
2000	1400	3400
400	1600	2000
800	1200	2000
1200	800	2000

Basic parameters for agent metabolism, foraging efficiency, food energy values, and patch food initial densities were held equal across all simulations at values previously found to be biologically plausible (Miller et al., 2013). We present all model parameters in Appendix A (Supplementary material). Thirty trials were simulated for each combination of landscape scenario, initial patch distribution (random or IFD), and MVT strategy (no patch switching, RLMVT, OMVT, XRLMVT, and XOMVT). IFD was only used with no patch-switching strategy.

The basic premise of MVT is that optimal foragers should switch patches when their gain rate falls below the long-term average gain rate in the environment; that is, the switching threshold should be equal to the long-term average. Determining the long-term average would allow a useful comparison of each algorithm's performance to actual MVT-predicted optimization. However, as noted above, this average rate cannot be determined *a priori* because of the circular causality between the average rate and forager behavior. Furthermore, classic MVT considers the average rate to be a constant because it considers an infinite landscape in which the long-term average rate across the landscape is unaffected by the depletion of individual patches. This clearly is not the case in many ecosystems and was not the case in our simulation. Real-world landscapes may only produce a fixed quantity of resources each season which are exhausted before being replenished. Our simulations model this kind of landscape: as time moves forward, the agents have removed food resources so that patch densities decrease and overall gain rates must therefore be lower than they were earlier in the simulation. Despite our inability to calculate a long-term average gain rate in advance, it is possible to calculate the achieved average gain after a simulation is complete. To account for the change in the overall landscape's gain rate in finite time, instead of using a constant value for the average rate, we calculated an average rate that changed on a daily basis to reflect resource depletion. This rate was the average of daily gain rates only from the beginning of the given day through the end of the simulation. A rate calculated in this way is affected by the foraging strategy in use, but nonetheless does reflect the only true measurement possible of the rate in the environment, in particular given that the environment includes the foraging behavior of its denizens. Because this was the long-term gain rate going forward, we call this the long-term-forward average rate (LTFA rate). Since a single day is short compared to the length of the overall simulation and resource depletion across a single day is therefore relatively small compared to the total amount of food in the landscape, the LTFA rate may be a reasonable approximation of the optimal switching threshold for that day. Therefore, we used the LTFA rate as a metric to compare for the achieved switching threshold to MVT predictions as a further test of the effectiveness of each algorithm.

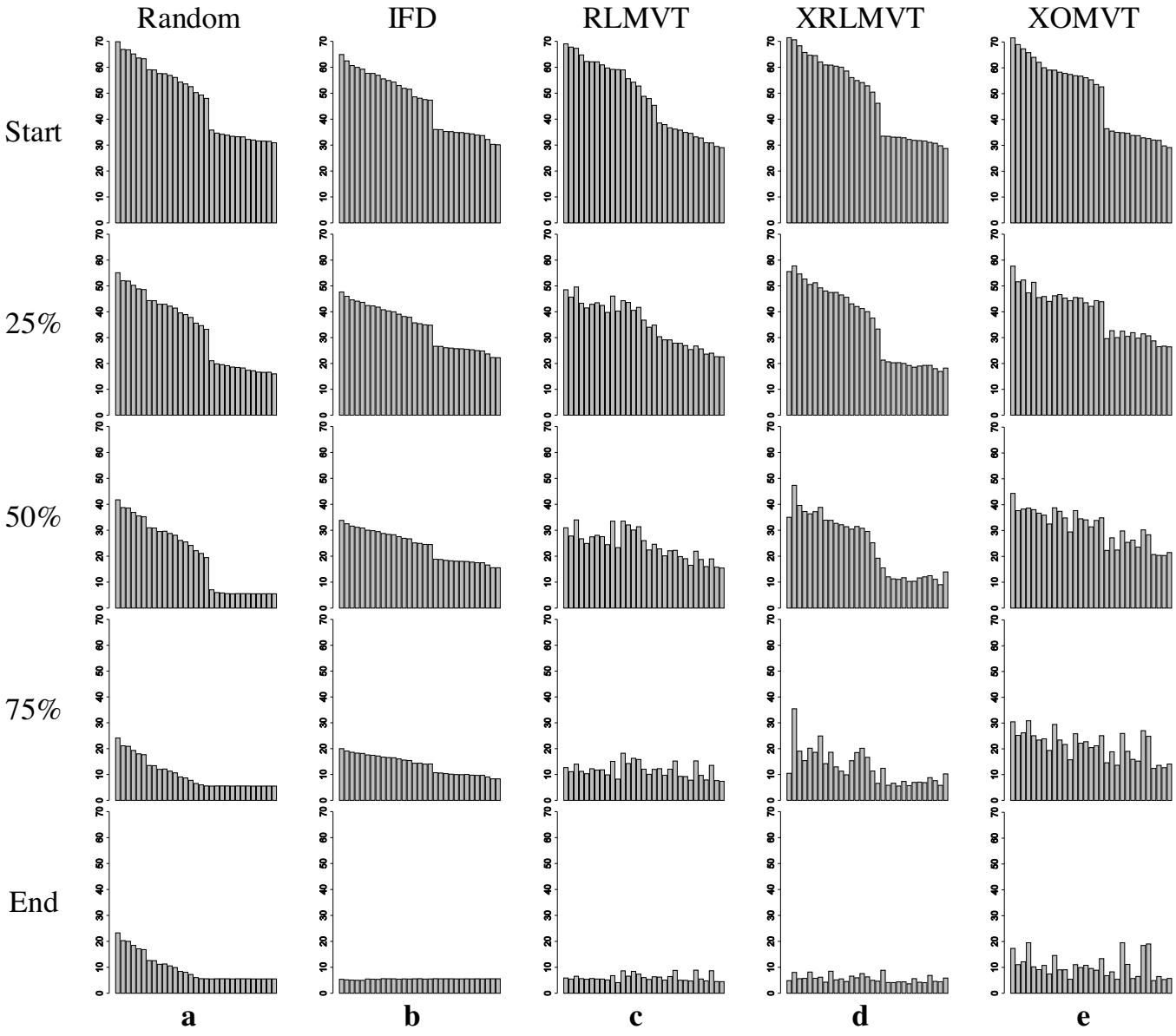

Fig. 1. Patch density distributions over time for typical simulations: (a) random initial selection with no MVT, (b) IFD initial selection with no MVT, (c) random initial selection with RLMVT, (d) random initial selection with XRLMVT, and (e) random initial selection with XOMVT. All simulations shown above are for 1400 ha rice acreage and 2000 ha moist soil acreage. On all graphs, patches are consistently ordered by food density at the beginning of the simulation (that is, patch order on the x-axis remains the same for all five time points graphed).

The ABM used herein collects additional data that the original model (Miller et al., 2013) did not collect. The ABM used for this study calculated the mean and standard deviation of patch density, the number of foragers in each patch, and the correlation between number of foragers in a patch and that patch's density at each time step. This ABM also captured on a daily basis the average gain rate for all foragers, the average gain rate at which foragers switched patches (that is, the current average switching threshold), and the average number of patch switches per day.

All statistical analyses were completed in R (R Core Team, 2013), including the calculation of LTFA.

3. Results

For all analyses, we examined each landscape combination separately. However, we noted that these results fell into two categories. In the first set of cases, all of the results showed very similar patterns, simply scaled to the amount of food in the landscape (for example, a larger number of simulation days for environments with more food). In the second set of cases, there was a continuum of results from the least rich environment to the most rich environment. Thus, for results belonging to the first set, we only present results from the richest environment (1400 ha rice, 2000 ha moist soil), and for results belonging to the second set, we present results from the richest environment (as above) and the poorest environment (1600 ha rice, 400 ha moist soil). Furthermore, OMVT and XOMVT results are very similar with slight scaling differences. Thus, except for the direct measure of the outcomes of interest in which we report all conditions, we report only the XOMVT results since these results were a slight improvement over OMVT. Complete sets of graphs for all conditions and all strategies are presented in the supplementary material, Appendix B.

One of the predictions of IFD is that patches with more resources will be depleted at a faster rate than those with less resources so

that patches will become more homogeneous as they approach depletion; this pattern is observed in empirical observations of foraging environments (Sutherland, 1996; Nolet et al., 2006). Typical results across simulation time are shown in Fig. 1 for random patch selection with no MVT (Fig. 1a), IFD with no MVT (Fig. 1b), random patch selection with RLMVT (Fig. 1c), random patch selection with XRLMVT (Fig. 1d), and random patch selection with XOMVT (Fig. 1e; OMVT is omitted as it performed similarly to XOMVT). Random and IFD both had very smooth progressions to their end states. Random patch selection with no MVT strategy has uniformly decreasing patch density but patches were not of uniform density at the end of the simulation; approximately half the patches were depleted to the forager's giving-up density (GUD; the level below which they would no longer attempt to utilize a patch), but many of the initially dense patches were not completely exploited by the end of the simulation. IFD patch selection showed the expected pattern in which more dense patches were exploited more than less dense patches; all patches reached a uniform level near the GUD. RLMVT recapitulated the IFD results very closely, though with some noisiness in the order in which patches were exploited; this raised the question, examined below, of why RLMVT did not perform as well as IFD. XRLMVT appeared to be performing in much the same way as random patch selection for most of the simulation, with early signs of approaching IFD results at 75% of simulation time; by the end of the simulation, however, XRLMVT also achieved efficient patch exploitation. This time course resulted from the patch-switching algorithm being conservative in switching patches early in the simulation because of the costs of moving and then becoming more willing to switch patches as patch densities became very small. This allowed XRLMVT to obtain similar final results to IFD without incurring switching costs when resources were plentiful. Both XOMVT and OMVT exploited patches in a similar pattern to IFD, but at much later stages of the simulation than IFD and RLMVT (for example, XOMVT's pattern at simulation end resembled RLMVT's at 75%). Both also had somewhat noisier exploitation patterns than the other strategies. This resulted from the large number of patch switches early in the simulation while resources were still plentiful; this depleted forager energy reserves unnecessarily and prevented them from completing efficient utilization of the environment, as we examine next.

To determine energy expenditures for the foraging strategies, we analyzed the average number of patch switches per foraging bout per forager for each, averaged across all 30 trials (Fig. 2). The XOMVT strategy, in particular, had an extremely high number of patch switches. Before day 100, the XOMVT strategy causes foragers to switch more often than the maximum number for RLMVT; before day 30, XOMVT leads to more switches than the maximum number for XRLMVT. Ultimately, the XOMVT strategy peaks at more than four times as many switches per foraging bout as RLMVT and XRLMVT. Furthermore, XRLMVT peaks at only about half the number of switches as RLMVT. This peak also occurs some 50 days later than the RLMVT peak. This supports the idea that foragers using XRLMVT do not switch patches until very late in the simulation, conserving resources needed for survival until efficient patch use becomes necessary. The area under each curve can be thought of as a proxy for time spent switching patches; direct examination of Fig. 2 should make it clear that XRLMVT foragers spend substantially less time switching patches (particularly than XOMVT foragers), and consequently spend substantially less energy switching patches.

To understand the basic MVT processes driving these results, we compared the achieved gain rate, the gain rate at which agents switched patches (the switching threshold), and the LTFA rate; the average number of switches per bout was graphed on the second axis to allow direct examination of its relationship to the foraging metrics (Fig. 3). Recall that our expectation is that the switching

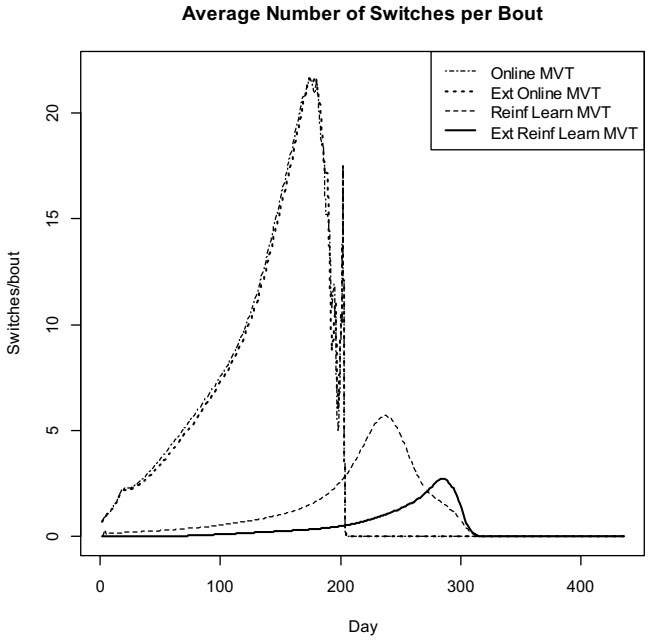

Fig. 2. Average switches per bout per forager for MVT approximation strategies (richest landscape condition).

threshold (solid black line) should approach the LTFA rate (dashed red line) when the forager is behaving in accordance with MVT. It is immediately obvious that in the online MVT strategy (XOMVT) the switching threshold rapidly converges to the current *overall* gain rate (solid red line), well above the LTFA rate (and for most of the first half of the simulation, also greater than even the long-term gain rate at the beginning of the simulation). By comparison, RLMVT does a fairly good job of tracking the LTFA rate, though in the poorest landscape it overestimates the optimum switching threshold for the first half of the simulation. XRLMVT performs most accurately in tracking the LTFA rate, almost exactly matching it in the poorest landscapes. In richer landscapes, XRLMVT even slightly underestimates the LTFA rate, which may be a preferable strategy if at the beginning of the simulation most patches are sufficiently rich to provide for the agent's daily needs. In such cases, it may be unnecessary to switch patches early on, until patch densities begin to fall so low that meeting daily energy requirements becomes uncertain; therefore, foragers should avoid switching patches until later in the simulation.

It is worth observing that the LTFA rates are not equal across the different algorithms. As noted above, this rate is affected by the foraging strategies in use by the environments' foragers. In particular, we draw attention to the fact that the XOMVT algorithm's LTFA rate is highest, the RLMVT LTFA rate is lowest, and the XRLMVT has a LTFA rate intermediate between the two. We examine this result in detail in the discussion, below.

These processes led to the following results in our primary outcomes of interest. As expected, the IFD simulations performed best by both measures of effectiveness. The order of effectiveness for the remaining conditions (again, for both measures, which were highly correlated) from best performance to worst was XRLMVT, RLMVT, random patch assignment, XOMVT, and OMVT (Tables 3 and 4).

We determined each strategy's effect on the outcome variables by examining the differences between conditions. Algorithms' improvements were judged in comparison to the non-IFD, no-MVT results. XRLMVT improved both DTD and DT50 M by the greatest amount among MVT strategies (DTD, 24.32; DT50 M, 18.69), though, as expected, this improvement was not as great as that for IFD (DTD, 44.60; DT50 M, 31.60). RLMVT did not change DTD

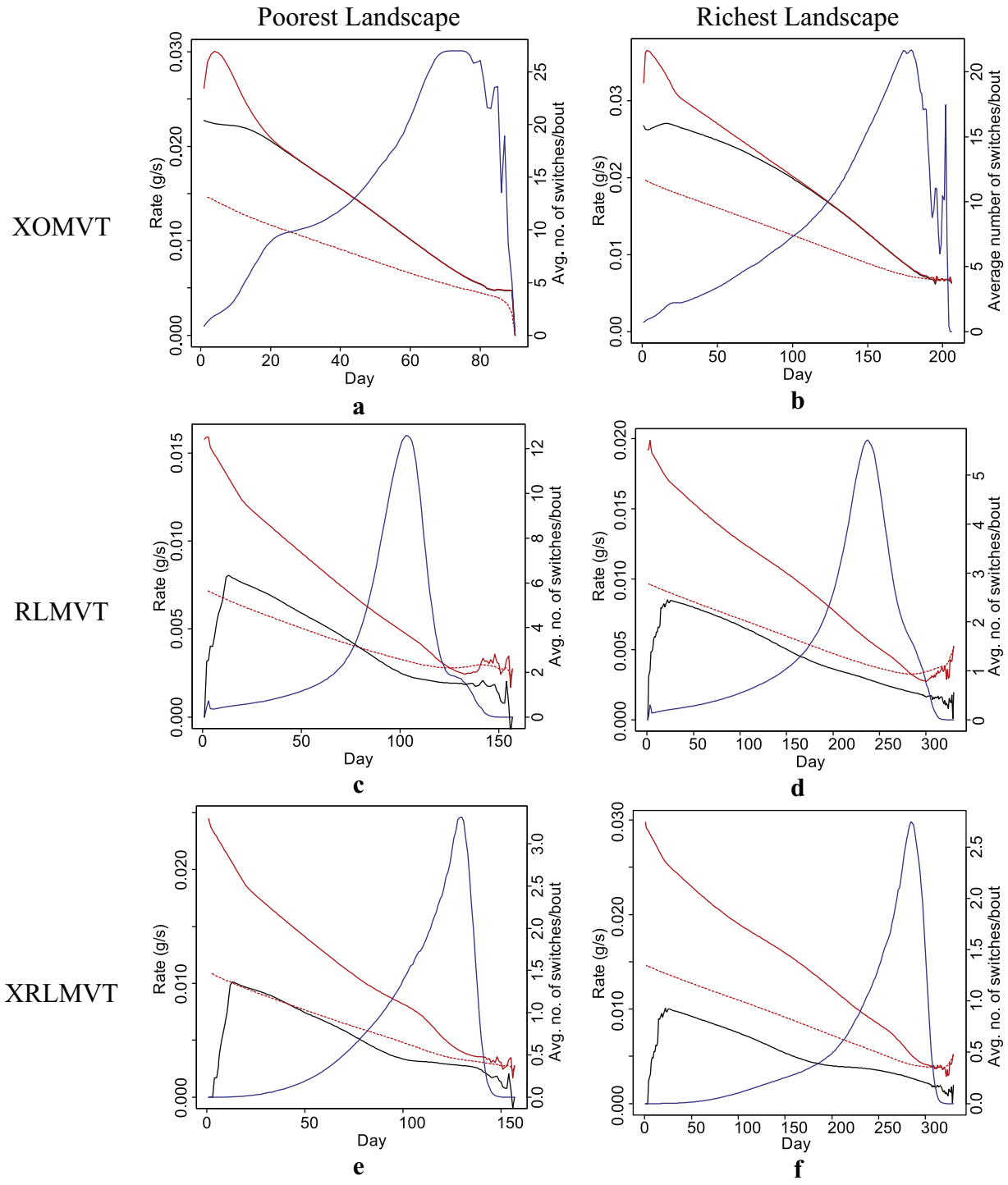

Fig. 3. Overall average gain rate per forager (solid red line), LTFA rate (dashed red line), rate at patch switch (black line), and switches per bout per forager (blue line, on right-hand axis). XOMVT (a) and (b); RLMVT (c) and (d); and XRLMVT (e) and (f) are shown for poorest landscape conditions (a), (c), and (e); and for richest landscape conditions (b), (d), and (f). Note that scales change between graphs.

Table 3
Days to deficit (DTD) for each condition (average across replications, standard errors in parentheses).

	No MVT	OMVT	RLMVT	XOMVT	XRLMVT
IFD	195.9 (4.65)				
Random	151.3 (3.99)	87.5 (2.38)	152.6 (3.52)	88.4 (2.46)	175.7 (4.26)

Table 4
Days to 50% mortality (DT50M) for each condition (average across replications, standard errors in parentheses).

	No MVT	OMVT	RLMVT	XOMVT	XRLMVT
IFD	221.2 (4.62)				
Random	189.6 (3.71)	129.9 (2.63)	187.7 (4.08)	130.8 (2.72)	208.3 (4.31)

Table 5
Notation used in this report and its supplements.

Symbol	Meaning	Symbol	Meaning
a	Action, bin, or "arm" selected for n-armed bandit	R_a	Estimated or expected reward for action, bin, or "arm" a
b	Index variable	R_{a_s}	Estimated or expected reward for action a at time-step s
d_{reward}	Discount rate for reward value rolling average	s	Time step index
d_{travel}	Discount rate for travel time estimator	t	Time spent in patch (typically only the length of one time-step)
$E_{\max}$	Lipid energy storage maximum capacity	t_{handle}	Handling time for food items
E_{travel}	Energy cost of travel	T	Sum total of time in previous patches
f	Amount of food in current patch	T_p	Time spent in patch (or patch type) p
H_{end}	Latest ending time for foraging bout	t'	Travel time between patches
H_{start}	Starting time for foraging bout	t'_p	Travel time from previous patch to patch p
i	Food intake in current step	$\hat{t}'$	Estimated travel time between patches
$I_{\max}$	Maximum daily food intake by mass	$\hat{t}'_s$	Estimated travel time between patches at time-step s
g	Gain function	z	Trial number
G	Sum total of previous gains	z_a	Total number of trials for action, bin, or "arm" a
$g_p(T_p)$	Gain in patch (or patch type) p after spending time T_p in patch	α'	Attack constant for foraging
$g_q(T_q)$	Equivalent to $g_p(T_p)$ when p is used otherwise in equation	δ	Initial width of action bins for reinforcement learning
k	Random-walk size factor for bin optimization	θ	Switching threshold rate
$\hat{M}_p$	Estimate of long-term return in patch (or patch type) p	θ_a	Switching threshold rate represented by action a
n	Number of actions, bins, or "arms" for n-armed bandit	θ_{max}	Expected maximum gain rate
N	Total number of patches	θ_{min}	Expected minimum gain rate
p	Patch or patch type	Θ_a	Bin center for action a (that is, the specific rate represented by a)
$P(a)$	Probability of action a	Θ_{a_s}	Bin center for action a at time-step s
q	Index variable for patch or patch type when p is otherwise used	μ	Current gain rate for specific patch
r	Reward for n-armed bandit	π_p	Proportion of patch (or patch type) p in environment
r_z	Reward for trial z on n-armed bandit	ρ	Patch density
R	Estimated or expected reward	τ	Softmax "temperature"

or DT50 M by a substantial number of days (DTD, 1.28; DT50 M, −1.96); indeed, for both measures, the results are within the 95% confidence intervals of the non-IFD, no-MVT algorithm. Both OMVT and XOMVT resulted in decreases in DTD and DT50 M suggesting once again that the online MVT strategies result in excessive energy expenditure under all conditions.

4. Discussion

Our goal was to explore the effectiveness of several optimal foraging algorithms designed to approximate MVT. In our simulations of biologically-plausible foragers, extended versions of Wawerla and Vaughn's algorithms (2009, 2010) performed better across a range of landscape conditions. However, we also found that for our simulation conditions and agents, reinforcement learning algorithms (RLMVT and XRLMVT) performed far better than continuous estimation, or online, algorithms (OMVT and XOMVT). Indeed, the online algorithms performed worse than simple random patch assignment with no patch switching, while the reinforcement learning algorithms approached the performance of IFD simulations. XRLMVT achieved an effectiveness, compared to IFD simulations, of 89.7% for days to energy deficit and 94.2% for days to 50% mortality.

The factor that drives the differences in foraging effectiveness is the number of patch switches that each algorithm imposes on foraging agents. The online algorithms are aggressive in seeking better patches, and in so doing exhaust food resources and energy reserves very rapidly. The reinforcement learning algorithms impose very low patch-switching rates until late in the simulation, as patches become depleted, and thus conserve food and energy. Online algorithms attempt to derive a switching threshold based on the agents' experiences of gain rates, while reinforcement learning algorithms use achieved energy gains to determine if a threshold is more or less useful. The online methods' divergences from foraging gains, focusing only on meeting rate targets, drive their inefficiency. On the other hand, reinforcement learning algorithms select any good-enough gains that minimizes energy expenditure, resulting in more conservative switching until some patches are too depleted to provide net energy gains.

For systems in which short-term gain rate maximization is more important than length of survival in a depleting landscape, however, our results suggest that the very dynamics that make the online algorithms less effective under this study's assumptions would likely favor the online algorithms. Future research might also be conducted in which the various algorithms compete against one another in the same environment to determine if the reinforcement-learning algorithms' long-term survival and efficient use of resources are more evolutionarily beneficial than the short-term acquisitiveness of the online algorithms, or if the converse is true in spite of the online algorithms' shorter survival times. However, this question was beyond the scope of the research reported herein.

Evidence that long-term efficiency drives the success of reinforcement-learning algorithms in this study can be seen by comparing each algorithm's switching threshold with the overall achieved gain rate and the LTFA rate. The online algorithms' switching threshold converged to the achieved gain rate, which was well above any long-term gain rate average for the environment, and thus not in line with classical MVT. On the other hand, the reinforcement learning algorithms approached the LTFA rate for the simulated environments. Because the reinforcement learning algorithms achieved high degrees of foraging effectiveness, we conclude that the LTFA is an appropriate target threshold for patch switching in finite landscapes with depletion. As we have discussed, the LTFA is definitely affected by the algorithm in use. However, the online algorithms switching thresholds exceeded their LTFA rates despite the fact that those algorithms had the highest LTFA rates; besides being a violation in the strict sense of the basic idea of MVT (switching threshold should match the average gain rate), it could be argued that these gain rates were inflated by the aggressive resource exploitation of that algorithm. More surprisingly, even though XRLMVT was *not* as aggressive as the RLMVT algorithm, it actually achieved higher LTFA rates. Therefore, despite the fact that LTFA is dependent on the foraging strategies being used, thereby being somewhat circular itself, we argue that this study demonstrates that it is the proper target to achieve in depleting finite landscapes.

Further research into the XRLMVT algorithm would yield information regarding specific parameterization of the reinforcement-learning system. In particular, adjustment of the softmax temperature may improve foraging effectiveness. Since lower softmax temperatures favor efficient exploitation of apparent optimal behaviors at the cost of exploration, but higher temperatures favor exploration of changing conditions and alternative maxima at the expense of efficient use of discovered optimal behaviors, the temperature parameter space should be explored to find the optimal tradeoff value for these two competing forces. Alternatively, annealing, in which the temperature is decreased as the simulation proceeds, or reverse-annealing, in which the temperature is increased as the simulation proceeds, might be explored. Adjusting the temperature in response to current forager success may also yield fruitful results: for instance, lowering the temperature to exploit discovered optima when the cost of switching patches is much lower than intake, but allowing the temperature to rise again as the cost of switching patches begins to approach intake levels.

The research reported herein only models exploitative competition. Further research regarding the effects of interference competition on these algorithms can be accomplished by adding interference abilities to the agents. We expect that the additional patch switching imposed by competition will increase cumulative travel costs (Fretwell, 1972), resulting in a decrease in the necessary gain rate to remain in a patch. That would reduce the number of patch switches due to patch quality. Such a result would be in accord with theoretical expectations, since under interference competition it is more desirable for a forager to remain in a patch that it controls. Since, as noted at the end of the description of RLMVT (above) our algorithms do not impose assumptions about how gain rates emerge but allow foragers to simply maximize total gain based on foraging experience, we predict that results with interference competition will be substantially similar to those in this study. This will have to be tested in future research. For our current report, we note that there are many systems in which exploitative competition is more important than interference competition, including the overwintering waterfowl that we used as our example organism. Nevertheless, other factors such as predation risk (Sih, 1984; Sih, 1998), especially from waterfowl hunters (Dooley et al., 2010; Lancaster et al., 2015), and nutritional requirements (Loesch and Kaminski, 1989) may influence patch selection and departure decisions.

Effective, efficient foraging strategies such as XRLMVT may find application in many problems beyond ecological simulations. Specifically, since this algorithm performs optimally by taking the cost of switching patches into account, XRLMVT might be well-suited to applications that are particularly sensitive to movement costs. Models of mate choice, in which switching rate must be calibrated for efficient evaluation of mates (Sullivan, 1994); models of group selection and formation (Stephens et al., 2007); and robot design for remote, difficult-to-access areas such as the deep sea and outer space, in which limited energy resources must be strictly conserved, are areas of potential application.

MVT and the implementations described herein advance our understanding of how animals should optimize sampling of their environment (Dall et al., 2005). The simulation model we developed emphasizes the collection and use of personal information to reduce environmental uncertainty; however, public information on habitat quality can also influence patch selection by animals (Danchin et al., 1998; Danchin et al., 2004; Doligez et al., 2003; Doligez et al., 2004; King and Cowlishaw, 2007), including waterfowl (Pöysä, 2003; Ringelman et al., 2016). Moreover, the degree to which animals learn about their environment and continually adjust their assessments (Bayesian updating) remains an important topic in animal behavior (Oaten, 1977; Green, 1980; Koops and Abrahams, 2003; Valone, 2006; Trimmer et al., 2011). Our model of foraging waterfowl could be extended to model birds flocking (*sensu* Thorn, 2003) to a previously profitable foraging patch (public information and group foraging) or by dynamically adjusting the softmax temperature in response to a changing environment (Bayesian updating).

In this simulation, we have shown that a reinforcement-learning paradigm that accounts for patch-switching costs is more effective than several alternatives and approaches a theoretical upper bound of effectiveness. This strongly suggests that foraging strategies must use adaptive learning. While the empirical question of the strategies that are actually used by foraging animals in patchy environments cannot be answered by simulations alone, simulations can be used to evaluate the relative effectiveness of hypothesized strategies. In conservation applications, it will ultimately be important to understand and predict outcomes based on actual foraging strategies. Our approach allows the investigation and comparison of strategies such as MVT, social learning, and decisions based on interactions with other foragers. In short, our models provide explicit predictions about the pattern of patch depletion and relative foraging efficiency. Future empirical research on forager patch utilization can draw on these results to investigate strategies of adaptive foraging behavior and the interaction of those strategies with conservation management scenarios.

Author contributions

M.M., J.E., and J.S. developed the concept for the research; M.M. and J.S. designed the research, M.M. coded the model, M.M and K.R. analyzed the data, M.M., K.R., J.E., and J.S. wrote the paper.

Funding

This work was supported by funds from Delta Waterfowl, the Dennis G. Raveling Endowed Waterfowl Professorship, the University of California-Davis Provost's Fellowship, and the National Science Foundation Graduate Fellowship grant #DGE1148897.

Conflict of interest

The authors report no conflicts of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolmodel.2017.02.013.

References

Bartoń, K.A., Hovestadt, T., 2013. Prey density, value, and spatial distribution affect the efficiency of area-concentrated search. J. Theor. Biol. 316, 61–69.
Bernstein, C., Kacelnik, A., Krebs, J.R., 1988. Individual decisions and the distribution of predators in a patchy environment. J. Anim. Ecol. 57, 1007–1026.
Bernstein, C., Kacelnik, A., Krebs, J.R., 1991. Individual decisions and the distribution of predators in a patchy environment. II. The influence of travel costs and structure of the environment. J. Anim. Ecol. 60, 205–225.
Central Valley Joint Venture, 2006. Central Valley Joint Venture Implementation Plan: Conserving Bird Habitat. U.S. Fish and Wildlife Service.
Charnov, E.L., 1976. Optimal foraging, the marginal value theorem. Theor. Popul. Biol. 9, 129–136.
Cowie, R.J., 1977. Optimal foraging in great tits. Nature 268, 137–139.
Dall, S.R., Giraldeau, L.-A., Olsson, O., McNamara, J.M., Stephens, D.W., 2005. Information and its use by animals in evolutionary ecology. Trends Ecol. Evol. 20, 187–193.
Danchin, E., Boulinier, T., Massot, M., 1998. Conspecific reproductive success and breeding habitat selection: implications for the study of coloniality. Ecology 79, 2415–2428.
Danchin, E., Giraldeau, L.A., Valone, T.J., Wagner, R.H., 2004. Public information: from nosy neighbors to cultural evolution. Science 305, 487–491.

DeAngelis, D., Mooij, W., 2005. Individual-based modeling of ecological and evolutionary processes. Annu. Rev. Ecol. Evol. Syst. 36, 147–168.
Doligez, B., Cadet, C., Danchin, E., Boulinier, T., 2003. When to use public information for breeding habitat selection? The role of environmental predictability and density dependence. Anim. Behav. 66, 873–988.
Doligez, B., Part, T., Danchin, E., Clobert, T., Gustafsson, L., 2004. Availability and use of public information and conspecific density for settlement decisions in the collared flycatcher. J. Anim. Ecol. 73, 75–87.
Dooley, J.L., Sanders, T.A., Doherty Jr., P.F., 2010. Mallard response to experimental walk-in and shooting disturbance. J. Wildlife Manage. 74, 1815–1824.
Fretwell, S.D., Lucas, H.L., 1969. On territorial behavior and other factors influencing habitat distribution in birds. I. Theoretical development. Acta Biotheor. 19, 16–36.
Fretwell, S.D., 1972. Populations in Seasonal Environments. Princeton University Press.
Gibb, J.A., 1958. Predation by tits and squirrels on the eucosmid *Ernarmonia conicolana*. Anim. Ecol. 27, 375–396.
Green, R.F., 1980. Bayesian birds: a simple example of Oaten's stochastic model of optimal foraging theory. Theor. Popul. Biol. 18, 244–256.
Green, R.F., 1984. Stopping rules for optimal foragers. Am. Nat. 123, 30–43.
Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H., Weiner, J., Wiegand, T., DeAngelis, D.L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310, 987–991.
Hodges, C.M., 1981. Optimal foraging in bumblebees: hunting by expectation. Anim. Behav. 29, 1166–1171.
Iwasa, Y., Higashi, M., Yamamura, N., 1981. Prey distribution as a factor determining the choice of optimal foraging strategy. Am. Nat. 117, 710–723.
King, A.J., Cowlishaw, G., 2007. When to use social information: the advantage of large group size in individual decision making. Biol. Lett. 3, 137–139.
Koops, M.A., Abrahams, M.V., 2003. Integrating the roles of information and competitive ability on the spatial distribution of foragers. Am. Nat. 161, 586–600.
Krebs, J.R., Ryan, J., Charnov, E.L., 1974. Hunting by expectation or optimal foraging: a study of patch use by chickadees. Anim. Behav. 22, 953–964.
Krebs, J.R., 1973. Behavioral aspects of predation. In: Bateson, P.P.G., Klopfer, P.H. (Eds.), Perspectives in Ethology. Plenum Press.
Lancaster, J.D., Davis, J.B., Kaminski, R.M., Afton, A.D., Penny, E.J., 2015. Mallard use of managed public hunting area in Mississippi. J. Southeast. Assoc. Fish Wildl. Agencies 2, 281–287.
Loesch, C.R., Kaminski, R.M., 1989. Winter body-weight patterns of female Mallards fed agricultural seeds. J. Wildlife Manage. 53, 1081–1087.
McLane, A.J., Semeniuk, C., McDermid, G.J., Marceau, D.J., 2011. The role of agent-based models in wildlife ecology and management. Ecol. Mod. 222, 1544–1556.
McNair, J.N., 1982. Optimal giving-up times and the marginal value theorem. Am. Nat. 119, 511–529.
McNamara, J.M., Houston, A.I., 1985. Optimal foraging and learning. J. Theor. Biol. 117, 231–249.
McNamara, J.M., 1982. Optimal patch use in a stochastic environment. Theor. Popul. Biol. 21, 269–288.
Miller, M.L., Ringelman, K.M., Schank, J.C., Eadie, J.M., 2013. SWAMP: an agent-based model for wetland and waterfowl conservation management. Simulation 90, 52–68.
Nolet, B.A., Gyimesi, A., Klaassen, R.H.G., 2006. Prediction of bird-carrying capacity on a staging site: a test of depletion models. J. Anim. Ecol. 75, 1285–1292.
Oaten, A., 1977. Optimal foraging in patches: a case for stochasticity. Theor. Popul. Biol. 12, 263–285.
Pöysä, H., 2003. Parasitic common goldeneye (*Bucephala clangula*) females lay preferentially in safe neighbourhoods. Behav. Ecol. Sociobiol. 54, 30–35.
Pierce, G.J., Ollason, J.G., 1987. Eight reasons why optimal foraging theory is a complete waste of time. Oikos 49, 111–117.
R Core Team, Software, 2013. R: A Language and Environment for Statistical Computing (version 3.0.2). R Foundation for Statistical Computing, Vienna, Austria (Available from) http://www.R-project.org.
Ringelman, K.M., Eadie, J.M., Ackerman, J.T., Sih, A., Loughman, D.L., Yarris, G.S., Oldenburger, S.L., McLandress, M.R., 2016. Spatiotemporal patterns of duck nest density and predation risk: a multi-scale analysis of 18 years and more than 10 000 nests. Oikos, http://dx.doi.org/10.1111/oik.03728.
Sih, A., 1984. The behavioral response race between predator and prey. Am. Nat. 123, 143–150.
Sih, A., 1998. Game theory and predator-prey response races. In: Dugatkin, L.A., Reeve, H.K. (Eds.), Game Theory and Animal Behavior. Oxford University Press.
Stephens, D.W., Krebs, J.R., 1987. Foraging Theory. Princeton University Press.
Stephens, D.W., Brown, J.S., Ydenburg, R.C. (Eds.), 2007. Foraging: Behavior and Ecology. University of Chicago Press.
Sullivan, M.S., 1994. Mate choice as an information gathering process under time constraint: implications for behaviour and signal design. Anim. Behav. 47, 141–151.
Sutherland, W.J., 1996. From Individual Behaviour to Population Ecology. Oxford University Press.
Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press.
Thorn, A., 2003. Considerations in spatially explicit, individual based modeling of waterfowl foraging behaviors. In: Master's Thesis. University of California, Davis, CA.
Trimmer, P.C., Houston, A.I., Marshall, J.A.R., Mendl, M.T., Paul, E.S., McNamara, J.M., 2011. Decision-making under uncertainty: biases and bayesians. Anim. Cogn. 14, 465–476.
Ulam, P., Balch, T., 2004. Using optimal foraging models to evaluate learned robotic foraging behavior. Adapt. Behav. 12, 213–222.
Valone, T.J., 2006. Are animals capable of Bayesian updating? An empirical review. Oikos 112, 252–259.
Wajnberg, E., Bernhard, P., Hamelin, F., Boivin, G., 2006. Optimal patch time allocation for time-limited foragers. Behav. Ecol. Sociobiol. 60, 1–10.
Wawerla, J., Vaughan, R.T., 2009. Robot task switching under diminishing returns. In: IROS 2009: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE.
Wawerla, J., Vaughan, R.T., 2010. Online robot task switching under diminishing returns. In: Artificial Life XII: Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems. MIT Press.