Dear Drs. Xue, Fern, and Sheldon,

We have received all the reviews for the paper you submitted to JAIR, entitled "Scheduling Conservation Designs for Maximum Flexibility via Network Cascade Optimization". The reviewers agree that the paper is well-written and makes an interesting contribution which I believe indicates that paper could *eventually* be published in JAIR. However, they also identify several difficulties, which preclude publication at this stage. Since JAIR's policy is to accept papers that only require minor revisions, we cannot currently accept the paper. However, we would be happy to see a revised paper on this subject--addressing the reviewers' concerns--published in JAIR. Thus, I strongly encourage you to resubmit the paper.

Attached you will find the reviewers' comments and suggestions. Please carefully attend to them.

In my opinion, there are two main issues that need to be addressed.

1. Motivation for including this work in JAIR. Reviewer #1 raises some significant concerns about relevance of and, if I can interpret his/her comments, interest in the submission among the AI audience in the problem and solution techniques. My opinion is not as negative, however, as the reviewer's. I think the topic and content do belong in JAIR, but would like to see some clear justification for members of the broader AI community for whom this may not be clear.

2. The main issue that leads to the decision of "reject with encouragement to re-submit", are the detailed comments of Reviewer #2 considering the true stochastic nature of the problem. I believe his/her arguments are compelling both in pointing out the problem and in the suggested solution of changing the formal definition of the problem. Alternatively, you may want to consider his/her suggestion of the linear time graph-based solution. [If you do pursue this alternative, you may want to look at the paper by

Reich & Lopes in INFORMS Journal on Computing, vol. 23 no. 3 460-469, on stochastic path problems].

In addition, Reviewers #1 and #2 have provided detailed comments about where some clarity is needed and Reviewer #3 provides some nice suggestions for questions you might answer to point toward the generality of your methodological approach.

Please note that papers can be resubmitted to JAIR only once. If you choose to revise the paper, please submit it through our standard web process, and include a cover letter describing the changes and responding to the reviewers' criticisms.

Thank you for submitting your paper to JAIR.

Regards,

Chris Beck
Associate Editor

Review 1

Review of JAIR 3850: Xue, Fern, and Sheldon, "Scheduling Conservation
        Designs for Maximum Flexibility via Network Cascade
        Optimization"

This paper addresses the problem of scheduling the purchase of land
parcels through time so that a given population of organisms spreads
as rapidly as if all the parcels had been purchased right away. This
is an interesting problem and the paper is nicely written. The
authors show how the original stochastic problem can be approximated
by a deterministic one, how the deterministic problem can be thought
of as a novel Steiner tree problem variant, and how that Steiner
problem can be solved via a primal-dual algorithm. The algorithm is

evaluated on both synthetic data and a real case study, against CPLEX, with nice results. I am happy to recommend the paper for publication in JAIR subject to some minor revisions detailed below.

My most serious concern is about relevance. I happen to be interested in computational sustainability in addition to AI and I am glad that techniques relevant for AI, such as decision-making under uncertainty, are also relevant to saving the world (or at least the Red-cockaded Woodpecker). But I do not see what this paper has to do with AI, other than sharing some tools. Certainly no one would argue that every high-quality paper that mentions MDPs belongs in JAIR. JAIR's scope is described as "all areas of AI". Since JAIR has a role as the archival journal of the AI community, and the AI community (via the AAAI conference) has embraced computational sustainability, perhaps one can justify having this paper in JAIR. But I would at least urge the authors to explain in the introduction what, if anything, those interested in AI might glean from this paper, and why they should read it. (And I would urge the editors of JAIR to seriously consider this issue and clarify it on the submission page.)

We have reworked the introduction and summary section to better articulate the aspects of the work that are of interest from a more general AI perspective. The second paragraph of the introduction attempts to articulate the general characteristics of the problem that challenge the existing state-of-the-art in AI planning/scheduling. This paragraph also foreshadows and points the reader to the final section (second paragraph) for a further discussion on this point. There we attempt to articulate how the general schema of our approach is one that should be considered more often for problems with the outlined challenges.

The central problem is formalized as finding a schedule that maximizes flexibility while still achieving optimal reward (= upfront purchase). This is confusing, as it is obviously impossible to achieve optimal reward without upfront purchase because the population spread model has non-zero probability of any parcel becoming a candidate for occupancy at every time step. Later in the paper, it becomes clear that the authors mean that the schedule need only achieve optimal reward with respect to the sampled scenarios. This should be made explicit right at the start of the paper (and certainly by equation 1).

We have since revised the problem formulation to include a (1-\epsilon) approximation factor in both the stochastic and deterministic problem definitions equation (1) and (2). At the end of section 3.2 we now discuss the role of epsilon as a factor that allows one to vary the flexibility-reward trade-off. We also discuss at the end of 3.2 and 3.3 the case of epsilon=0 and how it will only yield flexible solutions in the sampled scenario case

(covered in Section 3.3). The end of Section 3.3 then motivates our overall solution approach that focuses on first solving the epsilon=0 case and the use of early stopping for epsilon > 0.

In the reduction from set cover, the edge sets for the SW-DSG problem are not specified. If the sets only contain the edges going into a single node, then I think the reduction breaks because the same set might be selected using multiple different set nodes and you don't want to pay the cost multiple times. So this needs to be clearer.

The reduction was not broken, but we see that it was easy to miss the definition of the edge sets in the proof, which is key to addressing the above point. In trying to clarify the proof better, we were able to simplify the construction significantly and hence have rewritten the proof. The new construction and proof is quite a bit simpler, since there is now a one-to-one correspondence between the sets of the set cover instance and the edge sets of the SW-DSG instance.

I think it should be explicitly acknowledged that you don't have bounds on the error of the SAA approximation, hence results like Figure 5 are relevant to see how close you get to optimal reward.

Before equation 2 we have added a comment about what can be said using standard concentration inequalities and also mention that in practice one needs to empirically verify the adequacy of the chosen number of scenarios.

I was very concerned that you get better results than CPLEX's optimal solution, and the current explanation seems ambiguous. Did you evaluation the solution yourself and get a different value than CPLEX thought it achieved? Or does CPLEX think that value is optimal? Four significant figures doesn't seem like a lot to ask of a professional package - have you confronted IBM with this result?

CPLEX has the correct evaluation of its solution and it thinks the solution is optimal. According to the CPLEX technical documents, CPLEX allows a small error tolerance and this is likely the reason for the result. We have not looked into this further since it is a relatively small discrepancy. We have updated the footnote with this information.

It seemed to me that equations 1 and 2 should have min instead of argmin, since we want the same element that is passed to the min_c function.

We corrected this.

The paper goes through much notation in order to define the Steiner tree variant and a primal-dual optimization algorithm, but the main claimed benefit of the approach is fast computation of a lower bound.

However, the lower bound is very loose (factor of 2 seems loose to me)
and I can't help but think that some simple local search might be even
more scalable and practical on this problem. I wouldn't want to force
the authors to include an investigation of such an approach before
this work is published, but if there is any additional reason to think
that the proposed approach is somehow advantaged over local search, or
a reason why a factor of two is actually "quite useful" as claimed,
that would be important to add.

Often in the combinatorial optimization literature a factor of 2 guarantee is considered
quite good. This is of course a matter of opinion. At the end of the day, it seems
desirable to be able to reliably provide some lower bound, in order to give assurance that
the algorithm is not stuck in some horrible part of the solution space. The LP is not
scalable enough to reliably provide lower bounds---our method is the only approach we
know of to provide such bounds for large problems.

We agree that it would be interesting to consider local search approaches, especially
starting from the solutions we currently return. Things can only get better, but those
would not help provide or improve the lower bound.

Since the authors spend a lot of effort to prove that any decreasing
function should give the correct optimization objective, it was
confusing to hear that different functions (values for \beta) gave
different results. I assume this is because of some feature of the
primal-dual heuristic but it would be nice to have this clarified.

There can be multiple optimal solutions. Changing the function can change the solution
that is preferred.

The justifications for considering the scheduling problem separately
from choosing the parcels are weak. None of the features of the fully
adaptive problem given on page 2 (eg, large state space) are unique to
the fully adaptive case.

We agree and have expanded this justification in paragraph 7 of the introduction. The
main point is analogous to the distinction between open-loop planning and closed-loop
planning in large stochastic domains. In the former an action sequence is sought, while
in the later a (partial) policy is required, which is a much large solution space to
consider.

Your results are very nice, but they directly show (esp just before
section 6.3) how important the initial design is. It is a shame that
you don't consider that aspect of the problem!

This is one variable that we chose to keep fixed in these experiments, noting that we

minor things:

p2 "strikes an important middle ground" seems like a mixed metaphor to
  me

please make clear that (1) and (2) are intended to be done
simultaneously, and that you are seeking only near-optimal spread

p3: emits -> admits

p4: "in terms of the number of SAA samples" do you mean "for a given
  number of SAA samples"

  I think you can put the citation to the AAAI version in the
  acknowledgments section of the paper (which currently seems to be
  missing).

p7: the the

  so called -> so-called

  the notation seems non-optimal: k, i, j are used in two different
  ways. What about changing k to n and i,j,k to a,b,c?

p9 I suggest reminding the reader about c near eq 2

  "non-standard" what's so non-standard about it? it seemed pretty
  natural to me. maybe you mean "not amenable to MIP encoding"?

We clarified this point.

p10 I think it would be clearer to reverse "t:t \geq \pi(p)" to "t:\pi(p)
   \leq t" in the sums.

   I think a picture would help a lot in proving this proposition -
   it seemed unduly hard to follow. The "We now note that..." was
   not immediately clear to me.
We added some details to the induction equation to improve the presentation.

   space in "\beta \in (0,1)"

   Please remind the read what you mean by Steiner tree problem -
   there are many kinds, and you don't mean the kind where new
   vertices can be introduced.

We now have an informal definition to it in the end of section 4.2 to make it clear.

p11: please make clear in definition of E_{p,t} that u represents all
    incoming edges
Explanation added.

citation for DST NP-completeness
Added.

p12: the distinction you are trying to make between proposition and
    theorem was lost on me.

I suggest explicitly saying *weighted* set cover, and including
    the weights in the definition.
We made the changes.

purchases -> purchased
p13 DW-DSG -> SW-DSG

to encode an SW_DSG problem -> to encode it

While, -> While

an binary -> a binary

all *terminals*

p14 please explain the dual problem much more, since it is central to
    your algorithm.
We attempted to add some additional description throughout the section.

p15 via *only* edges in A

I would also have liked more details in how the algorithm worked,
    and in the proof of theorem 2.
We have tried to extend the section with additional clarification, description, and
intuition.

p18 rewards -> reward

p19 why all the italic around Figure 5?

p20 I suggest combining Figures 7 and 8 to make comparison easier, so
one's eye doesn't have to pass over the large caption. Clarify
in caption that free parcels are not "available" but rather
"already conserved".
The two figures are now combined. We have an explanation of "free parcels" in the 1st
paragraph in 6.1. Initially some parcels are already conserved. They are free (cost is 0)
for the planner and available for RCW to spread to.

It seemed to me from my printout that the red areas (at least at
    t=100) included areas beyond those in green and gray at t=0, which
    I thought was impossible (can only populated patches in conserved
    parcels). If I am seeing the figures wrong, can they be made
    larger and darker (perhaps running them vertically down an entire
    page)? Or else clarify how this can happen!
Thank you very much for looking at the figures so carefully! The figures of the
population spread are not the correct plots. The right ones are shown now.

The word Steiner appears uncapitalized twice in the references.
Corrected.

Would it help connect to other AI literature to mentioning other
determinizations of large stochastic planning problems? (Eg,

FF-Replan, Bent and Van Hentenryck, and friends...)

Review 2

TRACKING NO.:   3850
ARTICLE: Scheduling Conservation Designs for Maximum
               Flexibility via Network Cascade Optimization
AUTHORS: Shan Xue, Alan Fern and Daniel Sheldon
RECOMMENDATION: Reject with resubmission encouraged

General comments:

I generally liked the paper, and I have very little to complain about
the main technical content starting with Section 4. However, I
recommend rejecting the paper (with resubmission encouraged) for two
issues with Section 3:

1) There is a large problem with the motivation for the stochastic
   optimization problem introduced in Section 3.2. The way that the
   problem is defined, it is not truly stochastic since it essentially
   ignores the probabilities of the problem except for the distinction
   between the three cases of
   - probability 0
   - probability 1
   - everything else
   I explain why this is the case under "Why it is not truly a
   stochastic optimization problem" below.

Thanks for pointing this issue out. Your comments below are correct and we have
revised the problem formulation accordingly. You will see that both equation 1 and 2,
the stochastic and deterministic problems are now defined with respect to a reward
approximation factor epsilon. We have a discussion about this in the paper and note the
point you raise about trivial solutions when there is no approximation for the stochastic
problem. We believe that the revised formulation more accurately captures what is going
on.

   If the probabilities are essentially ignored by the formalization,
   then it is not an adequate model of the underlying application
   problem. From an environmental policy perspective, it does not make
   sense to have a problem definition that treats
   colonization/survival probabilities of some very small epsilon and

colonization/survival probabilities of 1 - epsilon equally.

Of course the presented algorithm *does* take the probabilities into account, but this is only an artifact of the imperfect approximation of the stochastic problem by sampling. The (from an application perspective) wrong problem is solved with an approximation which happens to produce useful results due to the nature of the approximation. This is fortunate, but should not be published in this form, since the formal problem definition must make sense in its own right.

Rather than having the right thing happen accidentally as an artifact of the algorithm used, the probabilities must have a direct impact on the problem statement for the probabilistic model to make sense. The high-level reason why in the current paper they do not is because of the insistence on a policy that has *exactly* the same reward as the upfront schedule. This means that every possible scenario, no matter how unlikely, must be covered to avoid losing optimality of the reward.

The most sensible way (for me) to put the probabilities back into the problem is to avoid insisting on optimal reward, and instead focusing on the tradeoff between reward and solution cost. Section 6.3 already goes into this direction, and I think that a revised paper should define the SOP as the problem considered in this section or some variation of it.

2) If we are willing to accept that the SOP as defined in the paper is actually what we want to solve, then (at least under some mild and realistic assumptions on the probabilities in the problem definition) the methodology used in the paper of approximating with a deterministic problem and providing an approximate solution for this deterministic problem is not a good idea. The SOP can be solved optimally in very short time with a direct method. More on this under "Direct solution of the stochastic problem" below.

Hence, in a revision of the paper, the basic problem definition would need to be changed to include some kind of balance between the cost and reward of the conservation policy, e.g. along the lines of what is done in Section 6.3. Even though this requires a rather fundamental change to the motivation and writeup of the paper, I think this still a "reject with resubmission encouraged" rather than an an outright

"reject" because I hope that most or all of the basic algorithmic pieces can be reused for the modified problem.

Why it is not truly a stochastic optimization problem:

For patches i and j, let p_{ij} be the probability that patch j is occupied at time t+1 given that patch i is occupied at time t.

Given these probabilities, the horizon H and the set of initially occupied patches, we can define the set of all possible scenario graphs (as in the paper). Moreover, we can associate a probability distribution with scenario graphs, measuring the probability $p(G)$ with which any given scenario graph G arises from the sampling process.

The expected reward of the upfront schedule is clearly:

sum_{G: scenario graph} $p(G)$ * [number of patches i
    for which G has a path from the root to the terminal i]

Let S be the set of scenario graphs with nonzero probability. A policy achieves the same expected reward as the upfront policy iff for each scenario graph G in S and each patch i s.t. G has a path from the root to the terminal i, at least one such path from the root to terminal i in G is bought by the schedule. (On the one hand, if the policy satisfies this criterion, it is clearly as good as the upfront policy. On the other hand, if it does not, there exists a scenario with nonzero probability on which it is strictly worse than the upfront policy, and hence its expected value is worse.)

Note that this characterization of the optimal policies does *not* depend on the precise probabilities of the scenario graphs: it only depends on S, i.e., on the question whether or not the probability of a given scenario graph is 0. Hence, every problem transformation that modifies the probabilities p_{ij} without changing the set S will leave the set of solutions invariant. Any transformation that changes p_{ij} with $0 < p_{ij} < 1$ to another value in the open range $(0, 1)$ has this property.

In conclusion, the exact probabilities p_{ij} don't matter; it only matters whether they are 0, 1, or in between.

Direct solution of the stochastic problem:

Consider the special case of the problem where $p_{ij} < 1$ for all i, j. I will call this the "mortality assumption".

The mortality assumption seems realistic, since there is no guarantee in nature that an individual survives or procreates in any given year. (I tried to check if this assumption holds in the Sheldon et al. dataset used in the experiments, but it is not clear: Eq. (5) in the Sheldon et al. paper may assign $p(i, j) = 1$ in cases where there is exactly one neighboring territory within the foraging radius. One cannot see from the paper whether or not such a case exists.)

If the mortality assumption holds, then there are no edges that are present in *all* scenario graphs (except for the edges leaving r). Consequently, for every possible (present in *some* scenario graph) path X from r to a terminal, there exists a scenario graph with nonzero probability where the *only* edges in the graph are the edges in X. Therefore, an optimal policy must include all edges in path X to cover this scenario graph.

It follows that a policy is optimal iff it includes all edges on all paths from the root to some terminal node in all scenario graphs. This set can be computed by producing the scenario graph where all edges with nonzero probability are present, performing a forward sweep from r (to eliminate edges unreachable from r) and a backward sweep from the terminals (to eliminate edges that are not relevant). The computation can be performed in linear time in the size of the scenario graph and leads to an optimal solution for the stochastic problem.

(Without the mortality assumption, the problem becomes harder because "certain edges" and "uncertain edges" need to be treated differently. The resulting problem is another Steiner tree variant. Even in this case, I think that solving the problem directly is algorithmically preferable to solving the determinization, since the problems are similar in style, but the determinization is larger due to the need to generate multiple scenarios.)

Minor points:

Here we comment on the more significant points and have adjusted the text according to most of the comments below. Thanks for the careful proof reading.

1. The ordering $<_c$ introduced in Section 3.2 is the standard concept of (weak) dominance, and the minimum elements are exactly the Pareto-optimal curves. Because of this, Proposition 1 is a very basic result on Pareto-optimality and weighted sums and need not be

proved again. If you really want to provide a proof, it can be much shorter since the result is so simple. (Compare to your proof of Theorem 2, which is much shorter even though it is much more complicated.)

We agree this is a straightforward result. We include the proof from first principles for completeness.

2. The definition of f at the beginning of Section 4.1 says that it is a function from {0, ..., H}, but the proof of Prop. 1 relies on the value $f(H + 1)$ in Eq. 3 and elsewhere ($f(t + 1)$ for $t = H$).

   I think the cleanest solution is to mention briefly in the proof of the proposition that for the purposes of the proof $f(H + 1)$ denotes an arbitrary value smaller that $f(H)$.
We now define f as a function on {0, …, H+1}.

3. In Figure 3:

   In Eq. (2), the two sums should not range over all vertices "h"/"j", but over "h: (i, h) in \mathbb{E}"/"j: (j, i) in \mathbb{E}".

   (It's possible, but wasteful to allow non-edges here, but then (4) and (5) would need to be adjusted to make sure the corresponding variables are set to 0.)   In Eq. (6), "u^k_k - u^k_r" should be "(u^k_k - u^k_r)" (parentheses needed).

Page 13, terminal -> terminals

4. In Algorithm 1, in line 5 the restriction "s \notin solution" is unnecessary since this is already implied by definition of C(k) and Cut(k). I think the algorithm would read nicer with the explicit mention of the implied condition removed (but I won't insist).

In line 6, it should be mentioned how ties are broken. Uniformly randomly? Arbitrarily but not randomly? Preferring the lowest index?
Any of these can be used and hence we don't specify in the code. In our implementation we use lowest index first.

5. In the proof of Theorem 2, the notation "\Delta^*" is not introduced, I think. Even though it is clear that it stands for "\Delta(s^*, k)", this should be said explicitly. (Very nice proof, by the way!)

6. In Section 6.1, the following information would be useful to have
   as reference points for the data in Table 1:

   - the cost of the upfront design (you say that it has a budget
     constraint of $320M, but it's not clear how close the design got
     to this maximum allowed amount)

   - the cost of the design that buys everything at the *latest*
     possible time (t = H). This is a trivial lower bound on the
     optimal cost, and it would be useful as a reference point for the
     PD bound. Maybe I'm missing something, but if we can assume that
     the upfront cost is 320 (million dollars), then this trivial
     lower bound should equal 320 * beta^H. With beta = 0.96, this
     gives the following bounds for different H values (compare the LP
     and PD bounds from Table 1):

   H = 15 => bound = 173.5
   H = 20 => bound = 141.4
   H = 25 => bound = 115.3
   H = 30 => bound = 94.0
   H = 35 => bound = 76.7
   H = 40 => bound = 62.5

   This would mean that the PD lower bound is worse than the trivial
   bound in all cases. However, looking at the costs of the actually
   computed solutions, I see that these numbers cannot be correct. I
   guess the reason is that the upfront cost is *not* $320M. All the
   more reason to mention it.

   Or maybe the discrepancy exists because the Steiner tree
   algorithm doesn't buy certain parcels *at all*? The way the
   algorithm is written up, there seems to be nothing in it that
   guarantees that every parcel is eventually bought: depending on
   how the sampling process plays out, it is possible that there is
   some parcel that is completely useless (i.e., contains no
   reachable patch and contains no edges useful for reaching patches
   in other parcels), and then it seems to me that the Steiner tree
   algorithm would never purchase this parcel. How is this handled,
   considering that the problem as defined requires that every
   parcel is eventually bought (at time H at the latest)? (Clearly
   the problem could, and in my opinion should, be extended to allow

not buying a parcel at all if it turns out not to be useful. But
the problem isn't currently defined that way, unless I overlooked
something.)

The reason is as you said that the algorithm doesn't buy certain parcels if the purchasing is not helpful. If you look at the cost curves in Figure 6, you would find than when H is smaller, the total cost at H is far less than $320M. The definition of the problem is to minimize the surrogate total cost while maintaining the same reward. There is no constraint requiring purchasing all the parcels in the end.

7. In the discussion of the lower bounds in Table 1, you write: "We
see that the lower-bound produced by the LP is significantly
tighter [...]. However, the LP is unable to be solved for the
largest problem, while our approach still yields a lower bound."

If the LP cannot be solved within the given resource limits, cannot
a lower bound still be computed? My understanding is that LP
solvers are capable of incrementally producing bounds in an anytime
fashion.

This may depend on the specific LP solver. In CPLEX, we did not get such information.

8. At the end of section 6.3, you write "First we notice that the
average reward achieved by the early-stopping schedules is within
the specified error tolerance". This is not always true: consider
the data point for H = 80 and epsilon = 0.2 in Figure 13.

We change the sentence as '… is almost always within the specified error tolerance…', as it is true in all the data points except H = 60, epsilon=0.05 and H=80, epsilon=0.2.

Formatting, grammar, typos etc.:

generally: "i.e. " (in LaTeX source)

=> "i.e.\ " (backslash + space) in LaTeX source

(Without this, spacing will be wrong. Where LaTeX thinks that a
sentence ends, it will add more space than between words. It thinks
that a sentence ends after periods, with the exception of periods that
follow isolated *capitalized* letters. One, but not the only, example
of "i.e. " instead of "i.e.\ " is on page 2.)

ditto for "e.g.", "et al.", "i.i.d." and other abbreviations
p. 2, "at anytime"

=> "at any time"

p. 3, "emits"

=> "admits"

p. 3 and elsewhere (e.g. p. 17), "Red-Cockaded Woodpecker"

=> "red-cockaded woodpecker"

p. 3, "(e.g. see the review article by Williams, Revelle, and Levin (2005))"

=> See http://www.jair.org/authorinstrs.html, point 1., 6th bullet point. (These rules apply to all academic writing, not just for JAIR.)

p. 3, "the work of Sheldon et al.'s (2010)"

=> "the work of Sheldon et al.\ (2010)"

Figure 1, "purchased nodes, and"

=> should have no comma

p. 8, "Note that, the"

=> should have no comma

p. 9, "build from"

=> "built from"

p. 9 and elsewhere, "Set Weighted Steiner..."

=> "set-weighted" should be hyphenated

p. 10, "Plug this expression ... and rearranging yields"

=> "Plugging..." (unless this is intended as an imperative, but this would be an odd way to address the reader since such a style is not used anywhere else in the paper).

p. 10, "for which X...(i, H) = 1 and hence contribute"

=> grammar issue; e.g. write "and which hence contribute" instead, or better, rephrase.

p. 12, "are purchases as sets"

=> "are purchased as sets"

p. 13, "While, solving"

=> should have no comma

p. 13, "lines 2-4"

=> "lines 2--4" (range dash); similarly later "6-9" => "6--9"

p. 16, "If so, then the edge set can be eliminated."

The "so" does not have a grammatical referent here.

p. 17, "the standard errors are negligible"

=> These are not "standard *errors*" since they are not deviations around a "correct" mean. "Standard deviations" would be better. Or write "variance is negligible".
Yes, we mean standard deviations and revised.
Table 1: In JAIR's style, I think the caption belongs below the table (as in AAAI style), not above it (as in Springer style). There isn't enough vertical space between caption and table currently because the JAIR style doesn't really support captions above the table.

p. 18, "the lower-bound"

=> "the lower bound" (since this is a noun phrase)

p. 18, "is unable to be solved"

=> "cannot be solved" or similar (it's not the LP that is unable but rather the LP solution algorithm)

Figure 5, "horizon H"

=> "horizon H"

Figures 7 and 8 run far into the margins.

p. 22 and elsewhere, "Early-Stopping"/"early-stopping"

=> "Early Stopping"/"early stopping"

In Section 6.3, the figures use the "inverted three" form of epsilon
(\varepsilon in LaTeX), while the text uses the "lunate" form
(\epsilon in LaTeX). Either is fine, but they should be consistent.

References:

"steiner" => "Steiner" (in many references)

At least three references to AAAI papers are wrong; the conference is
not called "Twenty-Fifth/Twenty-sixth Conference on Artificial
Intelligence", but rather "Twenty-Fifth/Twenty-Sixth *AAAI* Conference
on Artificial Intelligence".

When referring to the same or similar conferences, please use a
similar style. Compare Crowley et al./Golovin et al./Kumar et al. on
the one hand to Sheldon et al./Xue et al. on the other hand.

Charikar et al.:

"dircted" => "directed"

"on Discrete Algorithm" => "on Discrete Algorithms"

Shapiro:

"carlo" => "Carlo"

Review 3

Reviewer comments on the paper Scheduling Converservation Designs for Maximum
Flexibility via Network Cascade Optimization

Summary: This works considers the problem of scheduling the purchase of parcels
of over a time horizon to achieve maximum population spread (while trying to
delay purchases of each parcel as much as possible) as a stochastic programming
problem. A metapopulation model is used to generate scenarios and the Sample
Average Approximation is used to generate a deterministic approximation to the
stochastic problem. The deterministic problem is equivalent to what the authors
call the set-weighted Steiner graph problem for which a primal-dual
approximation algorithm is developed that apparently performs well in
computational experiments.

Comments: The paper is well written and the major ideas and results are
developed in a very clear and effective manner. The major contribution is in
the construction of the deterministic problem from the SAA approach i.e. the

set-weighted Steiner tree equivalent (a generalization of the Steiner tree problem developed in this paper) for which a primal-dual approximation is amenable. It is not obvious from the description of the problem that this is a potential approach to attack the problem. The computational experiments are for the most part convincing (see below). Stochastic integer programming problems are notoriously difficult to solve and such creative approaches (mixing good modeling and use of approximation algorithms) exhibited in this paper are I believe valuable. In conclusion, I recommend the paper for acceptance in JAIR.

Minor comments:
Computational experiments: it would be nice to get the size of MIPs (# variables and constraints) to get an understanding of what CPLEX is dealing with.
From the MIP formulation, generally #variable = #parcel * #horizon + #patch * #horzion * #scenario = 443*H+2500*H*N.

#constraints is harder to estimate. It partly depends on the #edges in the network. Presumably with so many # patches, it is huge.

We now mention this in the experimental section.

Would the title be more accurate to mention that this is conservation of a single species?
It would be more accurate, but we have stuck with the current title since it is already quite long.

The primal-dual schema can be seen to be a type of greedy algorithm for other types of combinatorial optimization problems e.g. the weighted set covering/packing? Can you interpret the primal-dual schema more naturally as a greedy method for the problem in this paper?
What else can be modeled as a set-weighted Steiner tree problem?

We have revised some of the description in that section and attempt to paint the "greedy heuristic" intuition. This is a reasonable way to view the algorithm, noting that it is a principled way of arriving at a heuristic, which also comes with a lower bound on performance.

We now also give another example, from networking, of a problem that would be well modeled by the set-weighted Steiner graph problem. This is at the end of Section 4.2.

--

J. Christopher Beck, PhD, LEL
Associate Professor, Industrial Engineering

Associate Chair, Research

Department of Mechanical & Industrial Engineering
Faculty of Applied Science & Engineering | University of Toronto
5 King's College Rd., Toronto, Ontario M5S 3G8
Office: 40 St. George Street, BA8126
jcb@mie.utoronto.ca | tidel.mie.utoronto.ca
Tel 416.946.8854 | Fax 416.978.7753