May 16, 2024

Adam Scott
Environment Advisor - Wildlife
Burgundy Diamond Mines Limited
900-606 Fourth Street SW
Calgary, Alberta T2P 1T1

Dear Mr. Scott:

Review of the Barren-ground Caribou Movement and Habitat Selection Analyses from Telemetry Data Report

The Department of Environment and Climate Change, Government of the Northwest Territories (GNWT-ECC) has reviewed Burgundy Diamond Mine's (Burgundy) *Barren-ground Caribou Movement and Habitat Selection Analyses from Telemetry Data* report (telemetry report) for the Ekati Diamond Mine, prepared by Rettie et al. Burgundy circulated the report on February 24, 2024, and GNWT-ECC posted the report to the Online Review System (ORS) for public review and comment on March 01, 2024. On April 11, 2024, Burgundy held a virtual workshop to present the telemetry report to reviewers and to provide more information about the study objectives, methods, and results. The main objective of the telemetry report was to provide a detailed analysis of radio-collar location data to examine the responses of caribou to the Ekati Diamond Mine and mine roads after accounting for the distribution of waterbodies, eskers, landcover categories (mostly vegetation types), and insect abundance.

GNWT-ECC appreciates the opportunity to review the first draft of the telemetry report and understands that Burgundy will respond to and address the review comments in a final draft of the report to be provided later in 2024.

GNWT-ECC staff (James Hodson and Mélanie Routh) reviewed the telemetry report and contracted Dr. Brian J. Smith from Wildland Resources to provide a detailed review of the analytical methods used in the report. Detailed comments and recommendations are provided in Topic/Comment/Recommendation format submitted using the ORS Excel comment template. However, a more detailed review by Dr. Smith is attached in pdf format to this letter and should be read in its entirety to have a thorough understanding of the context and rationale for his comments and recommendations. His review also includes several specific minor comments that GNWT-ECC encourages Burgundy to consider when revising the final draft of the telemetry report.

GNWT-ECC's overarching comment with respect to the first draft of the telemetry report is that the methods and analyses appeared to be exploratory/descriptive in nature and were unnecessarily complex and convoluted. This detracted from addressing the main question of interest, which is whether movements of barren-ground caribou are affected by proximity to mine infrastructure and mine roads. GNWT-ECC believes the main focus of the report should be on testing hypotheses related to this question. GNWT-ECC recommends that Burgundy re-analyze the data using one overarching integrated step-selection analysis model (iSSA) that considers all of the covariates of interest, and which directly includes interactions between distance to mine and distance to roads with turning angles and step lengths, and that inferences then be drawn from the direction and statistical significance of the coefficients in the model. The same overarching model structure should be applied to each season and sex. A small number of other a priori defined models could be included to test alternate hypotheses. Additionally, Burgundy could consider including an interaction between distance to mine and distance to roads, to evaluate whether the influence of proximity to one type of feature is affected by the proximity to the other (e.g., are roads more strongly avoided when caribou are closer to the mine footprint?).

GNWT-ECC acknowledges the time and effort that Burgundy has put into the telemetry report thus far and believes that the comments and recommendations made by different parties on the first draft of the report will ultimately lead to a more informative and useful product.

Should you have any technical questions please do not hesitate to contact Ms. Mélanie Routh, Wildlife Biologist (Cumulative Effects), at Melanie_Routh@gov.nt.ca or (867) 767-9237, extension 53228.

Sincerely,

Ms. Mélanie Routh
Wildlife Biologist (Cumulative Effects)
Environment and Climate Change

Attachment – Dr. Brian J. Smith's review of the telemetry report.

c. Ms. Heather Sayine-Crawford
Director, Wildlife Management Division
Environment and Climate Change

Mr. James Hodson
Regional Biologist, North Slave Region
Environment and Climate Change

Review of "Barren-ground caribou movement analyses from telemetry data."

Reviewed by:

Brian J. Smith, PhD
Movement Data Specialist
Utah State University
brian.smith@usu.edu

Overview

This document presents my comments on the report of Rettie *et al.* (2024) regarding caribou movements near the Ekati Diamond Mine:

> Rettie, W.J., R.S. Rempel, and L.M. Ainsworth. 2024. Barren-ground caribou movement analyses from telemetry data. Ekati Diamond Mine Wildlife Effects Monitoring. Report prepared by Paragon Wildlife Research and Analysis Ltd., Winnipeg, MB for Arctic Canadian Diamond Company Ltd.

My comments on this report are generally negative. I have some serious concerns about the misunderstanding of the models used here as well as the overall analysis. As a result of these concerns, I have low confidence in the robustness of the results presented and their subsequent interpretation to meet the objectives stated in the report. I have organized my main concerns into sections that deal with (1) misunderstood models, (2) covariate screening, (3) model validation, (4) movement characterization, and (5) pseudoreplication. I explain my concerns in depth in their respective sections below, but here I will briefly summarize.

My first section of comments deals with misunderstood models, specifically misunderstanding the distinctions between habitat selection functions (HSFs), step selection functions (SSFs, sensu Fortin *et al.* 2005), and integrated step selection functions (iSSFs, sensu Avgar *et al.* 2016). HSFs are synonymous with resource selection functions (RSFs, sensu Manly *et al.* 2002), but they are not a component of (i)SSFs as was implied by Rettie *et al.* (2024) in this report. Furthermore, SSFs are not iSSFs without a movement model; rather, they are more like an iSSF where the movement model is *assumed* rather than *estimated*, resulting in biased habitat selection parameters. Rettie *et al.* (2024) used SSFs in place of iSSFs for the stated purpose of being able to predict without a movement simulation. This is a false distinction and an incorrect use of these models. Furthermore, these misunderstandings in the models propagated to improper use of a fixed habitat kernel in Phase II models and incorrect assertions about having properly modeled movements using an iSSF in the section on movement characterization (Section 3.10).

My second section of comments deals with the effects of covariate screening on the resulting inference. The workflow itself has many layers of covariate screening/model selection prior to inference. Inference follows null hypothesis significance testing (NHST), i.e., Rettie *et al.* used p-values to determine whether an effect was statistically significant. Despite widespread appreciation that any data-driven variable screening process will bias p-values (e.g., Freedman 1983), Rettie et al. used 3 screening steps in Phase I and a fourth screening step in Phase II prior to drawing inference. Phase III repeated the steps of Phase I and Phase II, but on finer scale (1-h) data, so the same concerns apply. Each of the variables presented in Table 2-3 (landcover, oestrid, and mosquito variables) are more prone to be spurious correlations than their p-values reflect. However, the fourth screening method used only in Phase II (and the latter half of Phase III) was the Bayesian Information Criterion (BIC), which is more conservative than Akaike Information Criterion (AIC); i.e., it tends to favor models with fewer variables. Rettie *et al.* favored AIC in Phase I, but they favored BIC in Phase II when considering the mine-related variables (presented in Table 2-4). They did not provide a detailed rationale for why they should treat the mine-related variables differently and more conservatively than the other variables.

My third section of comments deals with model validation. iSSFs are typically fit with conditional logistic regression; however, the true underlying model that they assume is not actually conditional logistic regression. Rather, by randomly sampling available steps, the analyst is actually using numerical integration to estimate an integral in the likelihood of the iSSF (Michelot *et al.* 2024). Fitted iSSFs are often evaluated/validated using approaches developed for binary (1/0) data, but care must be taken to understand how these metrics generalize to the iSSF. In this report, Rettie *et al.* improperly used statistics applicable only to binary data to evaluate model performance. Their results showed poor performance in an absolute sense, but they still claimed positive results by asserting similarly poor performance of the fitted models in-sample and out-of-sample. My opinion is that these metrics are an unreliable test of model validity.

My fourth section of comments deals with movement characterization. The movement characterization section of the report is meant to directly address the concerns raised in the report by Poole *et al.* (2021) that concluded the Ekati mine affects caribou movements. Due to the misunderstanding of the distinction between SSF and iSSF, and due to the particular parameterization of what Rettie et al. termed an “iSSF”, those models could not provide inference on changes in movement patterns due to mine infrastructure.

Despite this, Rettie et al. argued that their “iSSFs” were “considered statistically more appropriate and stronger analyses for these data” (p. 30) compared to the methods of Poole *et al.* (2021). I agree that an iSSF is an ideal model for addressing these questions, but Rettie *et al.* did not parameterize their models properly to achieve that goal. Instead, they provided simple linear regressions of movement metrics (step length and turn angle) that do not account for the habitat selection process.

My fifth section of comments deals with pseudoreplication. Data used in this study were collected as repeated measures of individual caribou, but the inferences drawn were at the population level. Appropriate analyses of these data should have used mixed effects models to account for repeated measures on individuals, which are widely available for iSSFs.

Taken together, these comments demonstrate why I have low confidence in the robustness of the results presented and their subsequent interpretation to meet the objectives stated in the report. I find it hard to say what effect the issues I have raised might have had on the conclusions of the report. The report doesn’t explicitly state strong conclusions about effects. The deeper problem is that the analysis doesn’t contain key results that stem from a robust test. It seems to me that the analysis suffers from too many flaws to support any conclusions about effects of the mine on caribou movements.

Clarifying Objectives

Before I address my specific comments, I think it is useful to recapitulate the stated objectives of this report through the lens of the desired outcome. The objectives presented by Rettie *et al.* in this report, taken verbatim from Section 1.1, were (indented and italicized text are direct quotes from p. 2):

> *The broad questions addressed in this report are:*
>
> > *1. Are there effects of the Ekati Diamond Mine on fine-scale barren-ground caribou behaviour?*
> > *2. What are the effects and what are their causes?*
> > *3. At what scale do the effects occur?*
> > *4. Are effects specific to different seasons or sexes? and*
> > *5. What is the magnitude of the effects?*

And later:

> *Additionally, the analyses in this report test the season-specific effect of exposure of caribou to the area within 30 km of the Ekati and Diavik mines on:*
>
> > *6. Total distance moved within the season; and*
> > *7. Delay in arrival time on the next seasonal range.*

Further detail given in the objectives section (Section 1.1) indicated that other objectives were to "provide quantitative analyses for direct comparison with summary information in Poole *et al.* (2021)."

Clarifying the objective of an analysis is a key first step in developing an analysis workflow; i.e., clarifying whether the goal of the analysis exploration, inference, or prediction (Tredennick *et al.* 2021). Given the stated objectives of Rettie *et al.*, and given a pre-existing hypothesis (from Poole *et al.* 2021 regarding the effects of the mine on movement), I posit that the objective of this report was to draw *inference* on whether or not the mine has an effect on movement or habitat selection of caribou. If this is true, the goal of an analyst ought to be to construct a model (or a small set of models) to test the important hypotheses. Often, there should be a clear, *a priori* link between the model structure and the specific hypotheses themselves. From that well-constructed model (or small model set), the analyst can determine how much evidence favors one particular hypothesis over another.

Contrast that approach with the approach of a descriptive study. In a descriptive study, there may be many variables possibly of interest, relationships may be poorly understood, and the analyst might be seeking potential patterns to formulate specific hypotheses (Tredennick *et al.* 2021). While it would be entirely appropriate to have a large number of candidate variables (possibly at multiple spatial scales) and to use model selection to identify the strongest patterns, such an approach would *not* lend itself to inference. There is a rich body of literature on this topic, both within ecology and wildlife biology and in many other applied fields.

The introduction and objectives of the report by Rettie *et al.* seem to indicate (but do not directly state) that inference was the primary goal in this study. However, the analysis workflow more closely resembles that of a *descriptive* study, calling into question the validity of any *inference* drawn from the analyses.

Misunderstood Models

In some literature, "habitat selection analysis" is used as a catch-all to refer to any approach that measures habitat selection (i.e., habitat use disproportionate to habitat availability) (Fieberg *et al.* 2021); that is the way Rettie *et al.* use the term HSA and the way I will use it going forward. Unfortunately, the term "habitat selection function" (HSF) has not, in parallel, been used as a catch-all to refer to any parameterized function that returns habitat selection values. Rather, it has been used as an alternative term for a resource selection function (RSF) to emphasize that “habitat” can be comprised of resources (increase an animal’s fitness), risks (decrease an animal's fitness), and conditions (increase fitness only within a certain range) (Fieberg *et al.* 2021). RSFs are an HSA that have been in use since the 1990s (Boyce & McDonald 1999; Manly *et al.* 2002) and remain very popular today; Rettie *et al.* did not use RSFs (although they refer to HSFs throughout) in this report. RSFs do not assume any movement model, they are agnostic to time, and they thus assume used habitat locations are completely independent. This is a poor assumption for modern telemetry datasets (GPS or similar) that have high fix rates with a strong signal of autocorrelation, but it is conducive to projecting the fitted function in geographic space (Signer *et al.* 2017) to make a map of relative selection strength (Avgar *et al.* 2017).

Step selection functions were developed to deal with the autocorrelation in GPS (or similar) telemetry datasets by defining available habitat at the step level (Fortin *et al.* 2005). Step selection functions are comprised of two components: a movement-independent habitat selection kernel (i.e., a function) and a habitat-selection-independent movement kernel (Fieberg *et al.* 2021; Michelot *et al.* 2024; Signer *et al.* 2024). The realized movements of the tracked animal reflect both of these processes (habitat selection and movement), but statistically, we can parameterize them as two independent kernels. The original SSF presented by Fortin *et al.* (2005) assumed the movement kernel was known; i.e., it was not estimated statistically (Michelot *et al.* 2024). Forester *et al.* (2009) pointed out that this induced a bias in the estimation of the habitat selection parameters and that bias could be alleviated by accounting for the movement process in the estimation of the model (recently reiterated by Michelot *et al.* 2024). Avgar *et al.* (2016) formalized this idea by showing how parametric movement kernels could be estimated with standard regression techniques, thus fully estimating both the habitat selection and movement kernels in a single model. Avgar *et al.* (2016) termed this analysis integrated step selection analysis (iSSA) and the resulting function an integrated step selection function (iSSF). Contrast that terminology with the use of the terms HSA and HSF, which are not parallel, as I discussed above. Rettie *et al.* largely covered this history (see Section 2.8, p. 18) with these relevant citations, but also seemed to introduce important misconceptions along the way.

First, they refer to the movement-free habitat selection kernel of the iSSF as an HSF – this is not consistent with any of the literature they cited. Although it may seem like splitting hairs to differentiate the HSF from the “movement-free habitat selection kernel,” they are not the same. The key distinction is that the latter is conditional on the habitat-selection-free movement kernel, and prediction without that movement kernel is undefined, just as prediction from a multiple linear regression is undefined without all the predictor variables. Importantly, it also incorrectly implies that the movement-free habitat selection kernel can be used as an HSF would be used to produce a map of relative selection strength. In fact, this projection of the movement-free habitat selection kernel is shown in Fig. 2 and Figs. 3-6 through 3-26. This is incorrect, as shown by Signer *et al.* (2017). Rettie *et al.* noted that these maps did not account for the movement process in Section 2.8.5 (p. 23) and correctly cited Signer *et al.* (2017) stating that predicting selection from iSSAs requires (typically) movement simulations. However, they incorrectly stated that these "analytical processes … are not advanced in their development." A general approach for these simulations has been available in the R package 'amt' for multiple years now, and a peer reviewed publication detailing the approach in 'amt' is now published (Signer *et al.* 2024). It is true that simulating from a fitted iSSA prior to this typically required some custom coding, but multiple papers have shown this (including Signer *et al.* 2017) and reviewed the overall approach (Potts & Börger 2022). The literature is clear that it is incorrect to project the movement-free habitat selection kernel in space to make a map.

Second, after (incorrectly) explaining that the process for movement simulations is not developed, they presented a non-sequitur, "Consequently, SSA was chosen for Phase I analyses, and movement covariates for turning angle and step length were excluded from candidate models." This is not a logical consequence of failing to simulate from a fitted iSSF. As I described above, an ordinary SSF (as presented by Fortin *et al.* 2005) *still* contains a movement kernel, but it is *assumed known* rather than estimated statistically. The consequence of that, as I described above, is that the estimation of the habitat selection coefficients is biased (Forester *et al.* 2009). This does not alleviate the previously stated problem, but rather it creates a new one. The movement-free habitat selection kernel is now likely estimated with some unknown amount of bias, but it still cannot be naively projected in space (reviewed by Potts & Börger 2022).

Third, after completing Phase I (including multiple rounds of covariate screening, next section) and arriving at a selected Phase I model, Rettie et al. used this model to create a "covariate" for the Phase II models, which they termed "RSFrisk" (Table 2-4), reiterating the misconception that the movement-free habitat selection kernel is an RSF. They used "RSFrisk" as the only habitat covariate in their new "iSSA Base model" (Table 2-4), inexplicably assuming that the previously estimated habitat selection coefficients must remain fixed in their new model. This is an unjustified assumption. They included log of step length and cosine of turn angle in this base model, which is fairly standard in an iSSA, as I will explain below. From the base model, they created a set of 9 additional models including a single covariate related to the mine infrastructure, modeled as a parabola (including a squared term) and interacting with the fixed "RSFrisk" variable.

Fourth, notwithstanding the fact that the "RSFrisk" variable was an inappropriate treatment of habitat, these 9 additional models only test the hypothesis that the mine infrastructure could change the movement-free habitat selection kernel by modulating selection for the "RSFrisk" variable. Based on the conclusions of Poole *et al.* (2021), Rettie *et al.* expressed an interest in testing whether or not the mine and/or mine infrastructure affected caribou movement (objectives 1, 6, and 7). None of these models were structured to test the effect of mine infrastructure on the selection-free movement kernel, which would have required interactions between cosine of turn angle, log of step length, step length, and the mine variables.

To explain why, I will briefly explain how iSSA estimates the selection-free movement kernel. The key difference between ordinary SSA and iSSA is that, in iSSA, the (randomly generated) available steps are sampled from a parametric step length distribution (e.g., the gamma distribution, which Rettie et al. used) and a parametric turn angle distribution (e.g., the von Mises distribution, which Rettie et al. used). The parameters of the gamma and von Mises distribution are typically estimated by fitting them to the observed step length (gamma) and turn angle (von Mises) distributions. I will refer to these fitted distributions, which are used to create the available steps, as the "tentative distributions". Then, movement terms are included in the model formula to *update* the parameters of the tentative distributions. The coefficients for the movement terms are used in particular formulas to arithmetically update the tentative parameters to the estimated parameters of the selection-free movement kernel (Avgar *et al.* 2016; Fieberg *et al.* 2021). Which movement terms are included in the model formula depends on which distributions are chosen (see Fieberg *et al.* 2021, Appendix C). To update the gamma distribution, the analyst should include step length and log of step length, and to update the von Mises distribution, the analyst should include cosine of turn angle. These updated distributions are the focus of inference, and the coefficients themselves should not be interpreted in the same way coefficients from the movement-free habitat selection kernel are interpreted.

Rettie *et al.* included log of step length (but not step length) and cosine of turn angle in their "iSSA Base model"; however, after BIC model selection (discussed in the next section), the movement parameters were often removed. For those models where the movement parameters were not removed, those models assumed a constant selection-free movement kernel. Rather than reporting the updated selection-free movement kernels, Rettie *et al.* presented only incorrect interpretations of the movement parameters in their report (Section 3.7, p. 44). Had Rettie *et al.* included interactions between the mine variables and the movement parameters, they would have been able to make direct inference on the effects of the mine on caribou movement (see Fieberg *et al.* 2021, Appendix B, for examples). Unfortunately, they did not include these interactions, despite their assertion that their models characterized movements (see my section on Movement Characterization for more discussion). In my opinion, leaving these parameters out of the model directly contradicts some of the stated objectives of this report.

Covariate Screening

There are well known statistical issues that arise when combining data-driven covariate screening procedures with NHST (typically p-value based inference; e.g., Freedman 1983). In that sense, the analysis framework presented by Rettie et al. is particularly problematic. Here, I present a brief summary of their analysis workflow meant to highlight the number of filtering steps. I denote those steps that represent covariate screening with an asterisk:

- Phase I (8-hour data):
  - * Boosted regression tree; keep only covariates with relative influence > 1
  - * “StepAIC” GLMs; filter models using deviance ratio and AIC score
  - Fit conditional logistic regression; * AIC used to choose single best model.
  - Evaluate best model
- Phase II (8-hour data):
  - Start with w(x) predicted from Phase I model
  - Add movement covariates (log of step length, cosine of turn angle) and mine-related covariates
  - *Use BIC to select the best model
- Phase III:
  - Start with data reduction to handle unequal sampling between individuals
  - Repeat steps from Phase I (***) and Phase II (*)

There are four screening steps, and any one of them should be expected to increase the probability of a Type-I error above that estimated in the p-value calculations. Rettie et al. used three screening steps in Phase I and a fourth screening step in Phase II prior to drawing inference. Phase III repeated the steps of Phase I and Phase II, but on finer scale (1-h) data, so my concerns about Phase I and Phase II also apply to Phase III. I cannot quantify how much each screening step made the problem progressively worse – it is possible that using one screening step would have the same effect as using four screening steps if they were using comparable criteria. However, that seems unlikely. Each of the multiple screening methods has the potential to create a bias toward spurious correlations (Type-I errors). This is true for the Phase I variables presented in Table 2-3 (landcover, oestrid, and mosquito variables). Conversely, the fourth screening method used only in Phase II (and the latter half of Phase III) was the Bayesian Information Criterion (BIC), which is more conservative than Akaike Information Criterion (AIC); i.e., it tends to favor models with fewer variables. Rettie et al. favored AIC in Phase I (for landcover, oestrid, and mosquito variables), but favored BIC in Phase II when considering the mine-related variables (presented in Table 2-4). They did not provide a detailed rationale for why they should treat the mine-related variables differently and more conservatively than the other variables.

Given my assumption that this is a study geared toward inference, the model building process should have focused on testing a small number of a priori hypotheses, while controlling for other important drivers of habitat selection. There is a large literature on caribou habitat selection throughout their range that could have been used to guide the model building process, plus the key addition of mine-related variables to *test* (i.e., draw inference about) the effect of the mine on caribou behavior. Instead, the report of Rettie *et al.* largely resembles a descriptive analysis consisting of data dredging and covariate screening/model selection across multiple spatial scales.

Model Validation

Model evaluation and validation is a critical step in any analysis. Especially when overfitting or spurious correlations are a concern, validating with out-of-sample data can alleviate many concerns. Rettie *et al.* presented model validation metrics based on out-of-sample data; however, their approach was focused narrowly on evaluating binary responses, which is not always appropriate for iSSFs.

iSSFs are typically fit with conditional logistic regression; however, the true underlying model that they assume is not actually conditional logistic regression. Rather, by randomly sampling available steps, the analyst is actually using numerical integration to estimate an integral in the likelihood of the iSSF (Michelot *et al.* 2024). Fitted iSSFs are often evaluated/validated using approaches developed for binary data, but care must be taken to understand how these metrics generalize to the iSSF. In this report, Rettie *et al.* improperly used statistics applicable only to binary data to evaluate model performance. Their results showed poor performance in an absolute sense, but they still claimed positive results by asserting similarly poor performance of the fitted models in-sample and out-of-sample. My opinion is that these metrics are an unreliable test of model validity.

Rettie et al. state that "overall performance of the top SSA models was assessed using Rho^2_{adj}." They provide no further description of the method, the software used, or a citation, so I cannot be sure exactly how this statistic was calculated and how it compares with a standard R^2 or pseudo-R^2 statistic. Nevertheless, they stated that it ranged from 0 to 1, with values of 1 indicating better performance. Furthermore, they used four other metrics that depend on the user defining a "cut-point": percent correct classification (PCC), sensitivity, specificity, and Kappa. No further detail is given in the report about how the cut-point was chosen or exactly how it was used. However, in evaluating a binary classifier, the model predictions are continuous probabilities between 0 and 1, but the observed data is either 0 or 1. A cut-point is typically used to classify smaller probabilities as 0s and larger probabilities as 1s for the purposes of calculating PCC, sensitivity, and specificity. I am not familiar with "Kappa", reported by Rettie *et al.* to be "a measure of the agreement between predicted and true values", so I will not comment further on it.

Choosing a particular cut-point is a subjective decision (again, Rettie *et al.* do not describe how they chose each different cut-point for each model). To avoid this, many analysts choose a statistic called area under the curve (AUC), referring to the receiver-operator characteristic curve, which is a plot of sensitivity vs. 1 minus specificity, for many values (ranging between 0 to 1) of the cut-point. The AUC statistic is designed for binary classifiers, but it has an intuitive interpretation for RSFs: it is the probability that a randomly chosen used point has a higher RSF score than a randomly chosen available point. AUC is not appropriate for iSSFs because those models are stratified; the generalization of AUC to the stratified case is called concordance and is widely used to evaluate SSFs. The measures of PCC, sensitivity, and specificity reported by Rettie et al. are conditional on the cut-points, which I would assume were chosen to maximize at least one of those quantities. If my assumption is correct, it suggests that this evaluation tends to be too optimistic. Furthermore, the 0s in this conditional logistic regression (the available steps) are not data (they are sampled by the analyst) but are rather used to numerically estimate an integral. Thus, specificity – in particular – does not evaluate data at all, and PCC and sensitivity are still not motivated from first principles under this model.

Despite being generally inappropriate validation metrics, and despite my assumption that they tend to be generally optimistic, the performance metrics also tend to be low (Appendices: Tables C-1 – C18; Tables E-1 – E-14). Rho^2_{adj} values for the 8-h models were as low as ~ 0.01 and as high as ~ 0.08 (Appendices: Tables C-1 – C-18). Rho^2_{adj} values for the 1-h models were as low as ~ 0.02 and as high as ~ 0.15 (Appendices: Tables E-1 – E-14). It's unclear if these values are "bad" because of the lack of detail on how they are calculated; sometimes pseudo-R^2 for the true generating model under a GLM (with overdispersion) can be very low. There's just not enough information to tell.

Furthermore, in discussing the low Rho^2_{adj} values, Rettie *et al.* stated, "This limited explanation of behaviour is a consequence of analyses that appropriately restricts available habitat to a plausible set of locations based on the movement time interval and the movement patterns of individual animals," (p. 80). I do not think this statement can be substantiated. I think it's more a consequence of the metric used to assess model fit, combined with relatively poor fit. It is not generally true that goodness of fit statistics for an iSSF should be low. They continue, "Little change in relative selective value arises when observed short distance movement steps are matched with random short distance movement steps originating at the same location – available and used locations have similar attributes and selection at this scale is limited." (p. 80). This statement is absolutely not true. It is entirely possible in theory to have – and there are many published papers that show – strong selection strengths over short movements.

For the other metrics, Rettie et al. stated that, "PCC is viewed as the most important measure" (p. 43) and "model validation depends on the performance of the model for the test data closely matching the performance of the train data," (p. 45). I disagree that this is a reasonable criterion for model validation. A very poorly fit model will perform poorly on the in-sample and the out-of-sample data; that the in-sample performance is also poor does not give me confidence in the model. In conclusion, I disagree with the choice of metrics for model evaluation, but nonetheless, the metrics that were presented suggested to me poor overall performance.

Movement Characterization

Rettie et al. devoted a section of the report to movement characterization. This section is meant to directly address the concerns raised in the report by Poole et al. (2021) that concluded the Ekati mine affects caribou movements. Due to the misunderstanding of the distinction between SSF and iSSF, and due to the particular parameterization of what Rettie et al. termed an "iSSF", those models could not provide inference on changes in movement patterns due to mine infrastructure. Despite this, Rettie et al. argued that their "iSSFs" were "considered statistically more appropriate and stronger analyses for these data" (p. 30) compared to the methods of Poole et al. (2021). I agree that an iSSF is an ideal model for addressing these questions, but Rettie et al. did not parameterize their models properly to achieve that goal. In fact, the models they fit are not iSSFs in many cases.

Rettie et al. did provide simple linear regressions of movement metrics (step length and turn angle). Rettie et al. correctly noted that, "The simple relationship of caribou movements to the proximity of mine infrastructure is confounded by habitat selection and the spatial distribution of natural environmental features." However, they then *incorrectly* asserted that, "These relationships were explicitly addressed through iSSAs as described in Section 2.8." (Section 2.9, p. 29). The structure of their models did not, in any way, address the effect of habitat selection on movement (as I covered in detail in the section, "Misunderstood Models"). To accomplish this, they would have needed to include interactions between their movement variables (log of step length and cosine of turn angle) and the mine variables, which they did not. This was a missed opportunity to draw inference directly related to their stated objectives (objectives 1, 6, and 7) while controlling for the effects of habitat selection that they acknowledged confound the observed movement statistics. This makes direct analysis of step lengths and turn angles tricky, but I would generally accept this as a descriptive exercise. Drawing inference from these descriptive analyses is unreliable, notwithstanding my objections about pseudoreplication.

Pseudoreplication

Data used in this study were collected as repeated measures of individual caribou. The ecological and wildlife literature are full of examples of consistent individual variation in habitat selection (Leclerc *et al.* 2016) and cautions about the importance of accounting for individual-level autocorrelation using mixed effects models (Gillies *et al.* 2006; Muff *et al.* 2020). There are other approaches to accounting for individual variation in population estimates, such as a two-step modeling procedure (Craiu *et al.* 2011). Rettie *et al.* seemed to have ignored this problem. They did not address it at all in Phase I and Phase II. In Phase III, they used data reduction (Section 2.8.8.1) to balance the sample by thinning any individual dataset with more than the median locations (p. 27). Data reduction reduces statistical power, increasing the probability of a Type-II error. More simply, many methods exist that would allow an analyst to fit an iSSF with random slopes for balanced population-level inference without sacrificing information (Craiu *et al.* 2011; Klappstein *et al.* 2024; Muff *et al.* 2020). Additionally, the ordinary linear regressions that Rettie et al. presented to characterize movement metrics did not account for individual identity. Using a standard linear mixed model in this context would be absolutely expected for valid inference.

Minor Comments

These comments are minor relative to my previous comments. In my opinion, none of the following are terribly problematic on their own, but taken together and combined with my major concerns above, they reflect on an unreliable analysis.

- Section 2.4.2, p. 9: the seasonal UD analyses seem mostly unrelated to the objectives. However, it seems like Rettie *et al.* did very little to address the autocorrelation in the data. While they did subset the data to one location per animal per day, I would still expect autocorrelation to be strong at the daily scale in animals that have such large ranges over the year. A better estimate of the UD could have been obtained by fitting and AKDE home range that accounts for the autocorrelation in the data (Fleming *et al.* 2015; Silva *et al.* 2021). This would have required no data thinning.

- Section 2.5.3, p. 12: I don't like the data-driven process to identify the spatial resolution of covariates. Yes, it is common in the literature, but it is inherently descriptive, as I have mentioned elsewhere. Identifying the scale at which caribou select a particular land cover type was not the objective here, and this was an unnecessary step when the focus should have been on the effect of the mine variables.
- Section 2.8.1.1, p. 19: the authors stated, "locations with prior steps of 0 m length were removed." Steps of exactly 0 length are typically impossible given the precision of GPS collars is no better than a few meters; instead, they typically arise because of something like duplicate fixes (by the collar) or data duplication during the data management process. However, this is typically handled during data cleaning and prior to creating steps. That these likely errors persisted through the process of creating steps might suggest that the data cleaning protocol was not thorough. The data cleaning process indicated that, "Data were then screened to remove duplicate locations…" (p. 8), but this doesn't seem to have been thorough.
- Section 2.8.1.1, p. 19: five available steps per used step is low. The available steps are used for numerical estimation of an integral (Michelot *et al.* 2024), and few available steps make for very coarse estimation. A small number of available steps also leads to low resolution in model evaluation (e.g., using an appropriate statistic like concordance).
- Section 2.8.3, p. 20: there's no description of the rationale for using boosted regression trees at any point. This is not a standard approach for fitting iSSFs, and while it should generally work with any binary classifier, the true model of interest is not binary.
- Section 2.8.3, p. 21: "In preliminary analyses, collinearity was initially detected, with some VIF values approaching 20, but when proportional cover by water derived from landcover layers was replaced with proportion waterbody area derived from CanVec Series - Hydrographic Features (Natural Resources Canada 2019), collinearity (VIF) was greatly reduced." This is vague and concerning. So is the later treatment of VIF < 5 as a non-problem and the non-reporting of which variables showed high collinearity with the other variables in the model.
- Section 2.8.5, p. 25: "By definition, the top ranked model from the SSA for each sex by season are the Step Selection Functions (SSFs)." By what definition? There is no general consensus that the top-ranked model from an SSA is called an SSF.

- Section 2.8.6, p. 27: “By definition, the top ranked models from the iSSA for each sex by season are the integrated Step Selection Functions (iSSFs).” Same comment as above. This is not convention or definition.
- Table 2-5: Why not combine multiple mine-related variables in a single model? And why not consider interactions between variables? E.g., a reasonable hypothesis is that caribou avoid mine roads more strongly when they are close to the mine itself. That calls for an interaction between the mine and mine roads covariates.
- Table 3-2, Table 3-3, p. 33: I don’t get a sense of how many individuals were removed or the effect of removing individuals on the resulting inference.
- Table 3-9, p. 40: I would have liked to have seen all the VIFs.
- Section 3.6: reporting of quadratic effects (e.g., p. 42) is poor. They are presented as an independent result from the linear effect for the same term. In fact, using a linear and quadratic term actually parameterizes a parabola, rather than a line. These two terms represent a basis expansion, meaning that they are not independent and shouldn’t be interpreted as independent effects. The predictor is still a single dimension, but there are two parameters governing that predictor’s relationship with the response variable.
- Section 3.8, p. 46: “Phase 2 iSSFs provide a better fit to data… than the Phase 1 SSFs…” This is somewhat obvious. By taking Phase I models and *offering* a model selection routine a chance to add covariates, the fit can only improve.
- Section 3.8, p. 47: “Though not part of formal analyses…” Is this stating that figures 3-6 – 3-19 are not part of the analysis? This is the majority of the results figures.
- Section 4, p. 79: Rettie *et al.* stated that the covariates for herd, season, sex, and year are “best addressed with separate sets of models.” However, they note that this would result in 252 potential model sets, “an unmanageable number.” I agree that this is far too many model sets. This further highlights the descriptive and data-dredging tendencies of this analysis. There are pros and cons to representing these factors as separate models vs. interactions in the main model, and those pros and cons are not mentioned anywhere.

Literature Cited

Avgar, T., Lele, S.R., Keim, J.L. & Boyce, M.S. (2017). Relative Selection Strength: Quantifying effect size in habitat- and step-selection inference. *Ecology and Evolution*, 7, 5322–5330.

Avgar, T., Potts, J.R., Lewis, M.A. & Boyce, M.S. (2016). Integrated step selection analysis: bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, 7, 619–630.

Boyce, M.S. & McDonald, L.L. (1999). Relating populations to habitats using resource selection functions. *Trends in Ecology & Evolution*, 14, 268–272.

Craiu, R.V., Duchesne, T., Fortin, D. & Baillargeon, S. (2011). Conditional Logistic Regression With Longitudinal Follow-up and Individual-Level Random Coefficients: A Stable and Efficient Two-Step Estimation Method. *Journal of Computational and Graphical Statistics*, 20, 767–784.

Fieberg, J., Signer, J., Smith, B. & Avgar, T. (2021). A "How to" guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, 90, 1027–1043.

Fleming, C.H., Fagan, W.F., Mueller, T., Olson, K.A., Leimgruber, P. & Calabrese, J.M. (2015). Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator. *Ecology*, 96, 1182–1188.

Forester, J.D., Im, H.K. & Rathouz, P.J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, 90, 3554–3565.

Fortin, D., Beyer, H.L., Boyce, M.S., Smith, D.W., Duchesne, T. & Mao, J.S. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in Yellowstone National Park. *Ecology*, 86, 1320–1330.

Freedman, D.A. (1983). A Note on Screening Regression Equations. *The American Statistician*, 37, 152–155.

Gillies, C.S., Hebblewhite, M., Nielsen, S.E., Krawchuk, M.A., Aldridge, C.L., Frair, J.L., *et al.* (2006). Application of random effects to the study of resource selection by animals. *Journal of Animal Ecology*, 75, 887–898.

Klappstein, N., Michelot, T., Fieberg, J., Pedersen, E., Field, C. & Flemming, J.M. (2024). Step selection analysis with non-linear and random effects in mgcv.

Leclerc, M., Vander Wal, E., Zedrosser, A., Swenson, J.E., Kindberg, J. & Pelletier, F. (2016). Quantifying consistent individual differences in habitat selection. *Oecologia*, 180, 697–705.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. & Erickson, W.P. (2002). *Resource Selection by Animals*. Springer Netherlands, Dordrecht, The Netherlands.

Michelot, T., Klappstein, N.J., Potts, J.R. & Fieberg, J. (2024). Understanding step selection analysis through numerical integration. *Methods Ecol Evol*, 15, 24–35.

Muff, S., Signer, J. & Fieberg, J.R. (2020). Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation. *Journal of Animal Ecology*, 89, 80–92.

Poole, K.G., Gunn, A. & Pelchat, G. (2021). *Influence of the Ekati Diamond Mine on migratory tundra caribou movements*.

Potts, J.R. & Börger, L. (2022). How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. *Journal of Animal Ecology*, 1365-2656.13832.

Signer, J., Fieberg, J., Reineking, B., Schlägel, U., Smith, B., Balkenhol, N., *et al.* (2024). Simulating animal space use from fitted integrated Step-Selection Functions (iSSF). *Methods Ecol Evol*, 15, 43–50.

Signer, J., Fieberg, J.R. & Avgar, T. (2017). Estimating utilization distributions from fitted step-selection functions. *Ecosphere*, 8, e01771.

Silva, I., Fleming, C.H., Noonan, M.J., Alston, J., Folta, C., Fagan, W.F., *et al.* (2021). Autocorrelation-informed home range estimation: a review and practical guide. *Methods in Ecology and Evolution*, 2021, 1–11.

Tredennick, A.T., Hooker, G., Ellner, S.P. & Adler, P.B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, 102.