Check for updates

WILDLIFE BIOLOGY

Research article

Evaluating the accuracy of wolverine identification from photographs of snow tracks by expert observers in North America

David E. Moskowitz[1], Anna C. Machowicz[2], Matthew A. Scrafford[3], Philippe Naveau[4], Jeffrey N. Rose[5] and Stephanie Williams[1]

[1]Cascades Wolverine Project, Twisp, WA, USA
[2]Home Range Wildlife Research, Winthrop, WA, USA
[3]Wildlife Conservation Society Canada, Thunder Bay, ON, Canada
[4]Laboratoire des Sciences du Climat et de l'Environnement, ESTIMR, CNRS-CEA-UVSQ, Gif-sur-Yvette, France
[5]University of Utah, Salt Lake City, UT, USA

Correspondence: David E. Moskowitz (david@davidmoskowitz.net)

Wildlife Biology
2025: e01466
doi: 10.1002/wlb3.01466
Subject Editor:
Jean-Francois Robitaille
Editor-in-Chief: Ilse Storch
Accepted 16 May 2025

North American wolverines *Gulo gulo* are a species of conservation concern across much of their range. The remote and rugged terrain they occupy has led to the development of various remote detection methods. While visual identification of wolverine tracks has a long history of use, it has recently been considered less reliable than other remote identification methods. We evaluated the ability of 29 observers to identify wolverine from photographs of snow tracks. Each observer reviewed 99 observations: 48 of wolverine and 51 of species whose tracks are confusable for wolverine. The identification of each observation used was independently verified by visual observation of the animal that made the tracks or genetic samples collected from the tracks that confirmed field identification. We compared the performance of observers based on their demographic characteristics and the details present in each observation. Observers demonstrated a low false-positive error rate, cumulatively misidentifying observations of other species as definitively wolverine less than 1% of the time. The only clear demographic predictor of an observer's skill was their level of tracking certification via Cybertracker Conservation. Higher certification levels produced more positive detections of wolverine than did low or no certification. Observations with clear details in individual tracks produced the highest rate of positive detections. Depending on the level of certainty expected, observers detected 75–88% of wolverine observations that contained clear details in the tracks and clear track patterns, with a false-positive error rate of 1.1–3.5%. While the error rate did not increase, positive detections occurred less in observations without morphological details in individual tracks or clear track patterns. Observers called these 'unknown' more often. Our results indicate that observers can consistently distinguish wolverine from other species from photographs of snow tracks and refute the assumption that visual track identification is inherently unreliable.

Keywords: citizen science, community science, tracks, *Gulo gulo*, observer reliability, non-invasive research methods, wolverine

© 2025 The Author(s). Wildlife Biology published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

NORDIC SOCIETY OIKOS
www.wildlifebiology.org

Introduction

Wolverines *Gulo gulo* are a low-density circumboreal species of conservation concern across much of their North American distribution. They are currently listed as threatened in the contiguous United States under the federal Endangered Species Act, and the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) lists the wolverine as Special Concern. The species occupies remote rugged terrain that is not easily surveyed. Due to the difficulty of access to this terrain, various methods for remote detection have been employed to aid in the research and monitoring of wolverines including remote cameras, genetic sampling and track identification.

For professional research endeavors, track identification from photographs can provide a cost-effective, low-tech, and easily replicable method for monitoring wolverine. Modern smartphones are commonly carried by members of the public and can provide high-quality georeferenced photographs, making them exceptional snow track data collection tools. Track evidence is simple for the general public to both observe and understand. If subsequent expert review of photographic observations yielded reliable identification results, it would substantiate this particular form of community science and add a demonstrably reliable method to collect data on wolverine and other species. Additionally, with the impacts of winter recreation on wolverine becoming a point of conservation concern (Krebs et al. 2007, Switalski 2016, Heinemeyer et al. 2019, Barrueto et al. 2022, Machowicz et al. 2022), this method of track identification could provide a valuable tool to ascertain the overlap between winter recreation and wolverine habitat use, and to engage the public in research and conservation of the species.

Cascades Wolverine Project (CWP) runs a community science program in the Cascade Mountains of Washington State where wolverine populations appear to be undergoing a range expansion into habitat they were extirpated from early in the 20th century (Aubry et al. 2012, McKelvey et al. 2014). Wolverine habitat in these mountains is primarily roadless with a large amount of elevational relief and dangerous winter field conditions that makes access for studying the species logistically difficult for professional survey efforts (Aubry et al. 2012). CWP solicits photographic observations of putative wolverine tracks from backcountry travelers who recreate in wolverine habitat (primarily backcountry skiers, snowshoers, and snowmobilers in winter; hikers and mountaineers in summer). Photographs are submitted via an online portal and are subsequently independently reviewed by two experts to determine whether the tracks in the photos were made by a wolverine.

In Washington State, between 2016 and 2020 there were a total of 115 verifiable wolverine detections (photographs of the individual or genetic samples) documented by all formal research and monitoring efforts in the state (range 17–29 per year, mean 23, Long and MacKay 2024). Between 2020 and 2024, CWP categorized 46 track observations from the public as 'definitively wolverine' and another 43 as 'likely wolverine' (total 89, range 9–27 per year, mean 17.8, CWP unpubl.). Inclusion of these track observations in various conservation-oriented analyses could substantially increase our understanding of the current distribution of the species in Washington State.

Wildlife tracking is an extension of naturalist skills that have been employed around the world for millennia to study and manage wildlife (Liebenberg 1990, Stander 1997). While the identification of tracks has a long history of use in research and monitoring of wolverine and other sympatric mammal species (Zielinski and Kucera 1995, Matyushkin 2000, Wright and Ernst 2004, Basille et al. 2009, Gardner et al. 2010, Poley et al. 2018), there has been a recent tendency to assume that track identification is an unreliable method for identification of North American wildlife (McKelvey et al. 2006, Ulizio et al. 2006, Franklin et al. 2019). Studies of the reliability of track identification are sparse and yield widely varying results depending on the species that was identified and the observers who were tested (Stander et al. 1997, Evans et al. 2009, Zielinski and Schlexer 2009). We found no studies that have investigated wolverine tracks specifically.

Methods that evaluate observer reliability also vary widely. Stander et al. (1997) visually observed animals making tracks, after which study participants were tested on identification and interpretation of the tracks. Evans et al. (2009) used certified experts to evaluate the interpretation of track and sign evidence by study participants and performed an analysis of an existing dataset collected by the participants to assess the relationship between specific observers and the likelihood of detection of the target species. Zielinski and Schlexer (2009) measured the inter-observer consistency in identification of tracks left on track plates where the definitive identification of the species that made the track was not assumed. Several research efforts assessed the viability of genetic sample collection of wolverines and Canada lynx *Lynx canadensis*, and published how often putative tracks of a particular species yielded genetic samples of that species with an assumption that the genetic identification was absolute (Davison et al. 2002, Squires et al. 2004, McKelvey et al. 2006, Ulizio et al. 2006). We found no published research that tested the reliability of identification of tracks based on photographic evidence.

To assess the validity of the identification method used by CWP, we tested the reliability of observers to identify wolverine tracks in snow from photographs. We investigated and discussed the role of two different types of explanatory variables on the performance of observers: 1) the experience and demographics of observers, and 2) the amount and type of detail in the tracks presented in each observation. We defined reliability by two criteria: 1) the rate of false-positive errors in identification of wolverine (misidentification of tracks made by a species other than a wolverine as a wolverine), and 2) the rate of correct identification of tracks made by wolverine. Given a set of photographic observations that reflects the range of data submitted by the public to CWP's community science project, we predicted: 1) observers would be able to distinguish wolverine from other species via photographs of snow tracks with a false-positive error rate of $< 5\%$ (a standard metric for accuracy in a dataset), 2) observers with

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yakama Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Figure 1. Examples of photographs of wolverine tracks in snow that were the basis of this study.

higher levels of experience and training in track identification would produce fewer errors and correctly identify more wolverine than those with little to no experience or training, 3) observations containing more detail in the tracks and track patterns would produce higher detection rates than observations with limited detail, and 4) the tracks of fisher *Pekania pennanti* and Canada lynx would be the most challenging species to distinguish from wolverine (Fig. 1).

Study area

Wolverines occupy arctic, boreal, and mountainous temperate regions of North America with persistent spring snowpack (Copeland et al. 2010). We collected photographic observations of snow tracks from both wolverine and other species commonly misidentified as wolverine from various locations in North American wolverine range through collaboration with other researchers. Observations came from Wildlife Conservation Society Canada's wolverine radio telemetry study in Red Lake, Ontario, Canada; Swan Valley Connections' Rare Carnivore Monitoring program in Condon, Montana; North Cascades wolverine telemetry project in Winthrop, Washington; and Washington Predator–Prey Project in Winthrop, Washington, as well as incidental observations made by biologists in Washington, Minnesota, Ontario, and Alberta. Observations were made between 2006 and 2021. Observers tested in this study were located in Alaska, California, Colorado, Minnesota, Montana, Texas, Washington, Wyoming, Utah, Alberta, British Columbia, New Brunswick and Ontario.

Methods

Photographic observations

We compiled a database of 99 photographic observations of snow tracks of wolverine (n = 48) and other North American species whose tracks have been confused with wolverine (n = 51). An 'observation' was defined as a set of photographs of the same species and from the same event presented to observers as a single test question (Fig. 2). Each observation had between 1 and 8 photos in total. Eighty-nine of the observations included a recognizable scale or object (e.g. hand, trekking pole, notebook) placed in the photo adjacent to the tracks and the exact measurement of the object was provided to observers.

Observations were designed to mimic the type of data submitted by the public to CWP in terms of species diversity, number of photographs per observation, and quality of the tracks and track patterns portrayed in each observation. We determined which species' tracks could be confused with a wolverine by reviewing the actual species in CWP data that were believed to be wolverine by a member of the public upon submission. Amending a method from Halfpenny et al. (1995), each observation was given a 'photographic snow track quality' (PSTQ) rating to distinguish observations with a high degree of detail in foot morphology and track pattern from those with limited details (Table 1). We distributed observations used in the test between each level of PSTQ to evaluate how the quality of observations affected observer performance (Table 2).

Photographs used in observations were of tracks in snow where the identification of the species that made the tracks was independently verified by one of the following methods: direct observation of the animal (n = 82), remote camera photographs of the animal that made the tracks (n = 1), genetic samples (hair or scat) collected from tracks that match the field identification of the tracks (n = 15), or a combination of multiple methods (n = 1). Direct observation includes instances where the photographer saw an animal and then photographed the tracks, followed tracks until they observed the animal, or photographed tracks of an animal that was caught or released from a live-trap. Along with wolverine

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yahaana Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Figure 2. Example of an observation presented to observers. Photographs are of a single observation of wolverine tracks from release of a live-captured wolverine obtained by researchers in Red Lake, Ontario, Canada, in winter 2021. The observation was assigned a unique number (observation 5 in this case), and each photo within the observation was given a sub-number for direct reference (5-1, 5-2, etc.). Measurements for included scale were also given, in addition to clarifying notes regarding which tracks observers should evaluate.

tracks, we used photographs of snow tracks of fisher, river otter *Lontra canadensis*, Pacific marten *Martes caurina*, Canada lynx, mountain lion *Puma concolor*, wolf *Canis lupus*, domestic dog *Canis lupus familiaris*, coyote *Canis latrans*, red fox (*ulpes vulpes*, black bear *Ursus americanus*, and hoary marmot *Marmota caligata* (Table 2). We used a larger number of observations of Canada lynx (n = 21) and fisher (n = 12) tracks than other species as these were the two species predicted by the authors and colleagues we asked to be the most difficult to distinguish from wolverine based on past field experience.

Observer demographics

We selected observers based on two types of expertise: 1) those with a high degree of wildlife track identification skill, and 2) academically trained scientists with field experience working with wolverine. Within these categories, some observers included both traits. We collected the following demographic information from observers: highest level of scientific academic education, highest level of wildlife tracking certification, self-identified expertise level with wildlife tracking (amateur, defined as 'little or no experience' with wildlife track and sign; professional, defined as 'regularly

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yahaana Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1. Categories for classifying photographic snow track quality (PSTQ) of observations.

PSTQ 1 (n = 29)	Individual track details clear and track pattern clear (highest quality observation). For this category multiple clear tracks must be visible with all of the structures of the animal's foot plainly visible in the track. The trail also must include a clear presentation of the repeating pattern of tracks. Under these circumstances, with a recognizable scale, definitive identification should be possible
PSTQ 2 (n = 17)	Individual track details clear but track pattern obscure. Details for individual tracks as in category 1 but the observation is lacking photographs that show the pattern of footfalls clearly. With clear details in individual tracks and a recognizable scale, identification of a wolverine should be possible
PSTQ 3 (n = 28)	Track details obscure but track pattern clear. A common presentation for tracks in loose snow or older tracks that have been snowed in where the track pattern presents clearly but there is little (overall size and shape possibly) or no (totally snowed in) detail in the individual tracks. If the track pattern is one that is distinctive to mustelids and the overall size of the tracks and track pattern is discernible, identification of a wolverine may be possible. If the track pattern is one shared by numerous species and/or there is no definitive scale for reference, identification as a wolverine is likely not possible
PSTQ 4 (n = 25)	Individual track details obscure and track pattern obscure. Without clear detail in tracks or a plainly visible track pattern, definitive identification of tracks is likely not possible

apply wildlife tracking techniques for work (i.e. biologist, educator, naturalist, etc)'; and expert, defined as 'professional with specific expertise in wildlife tracking (i.e. published literature, Cybertracker specialist certification, providing technical training, etc.)'), approximate number of times they had observed putative wolverine tracks in the field, and past experience with systematic review of photographs of tracks (Table 3). Our sample included 8 female observers and 21 male observers.

The primary objective criteria we used for establishing wildlife tracking skill was certification through Cybertracker Certification (https://cybertracker.org/tracking/tracker-certification/), an international system that awards certification based on performance through a rigorous field test. Our observers included those with no certification (they had not been evaluated by this system), Track and Sign Level 2, Level 3, and Specialist certifications. A Level 2 certification is awarded to those who demonstrate an accuracy level of 80–90% of track and sign identification of a wide variety of difficulties. A Level 3 certification demonstrates an accuracy of 90% or higher for a wide variety of difficulties. To earn the Specialist certification, observers must demonstrate near 100% accuracy in their interpretation of very difficult wildlife track and sign in the field (Liebenberg et al. 2018).

Test design

Due to the limited number of experts to test based on our definition, all of the observers were known colleagues of at least one of the authors of this paper. Each observer was assigned a random personal identification number (PIN) by the test administrator that identified their test answers while allowing for a blind analysis of the resulting data.

We recruited as many observers as possible from within the two primary demographics designated as experts to make as large a sample size as possible. However, both academically trained wolverine researchers and highly certified wildlife trackers represented a relatively small group of observers in North America, producing a final sample size of 29 observers.

Each observer was provided with the same training on how to complete the test in the form of an instructional video and the opportunity to ask questions of the test administrator. The instructional video included training on how to use a ruled scale to determine the actual size of tracks in observations compared to the scale provided in the observations. Additionally, written instructions and reference material suggestions in the form of published materials on wolverine tracks were provided. Observers were informed that they could use any published materials, personal field notes, and photographs to assist them, but they were not permitted to discuss the observations with anyone else. We did not provide training on how to identify wolverine tracks specifically. Observers were not informed of the species included in the test or in what numbers they may occur.

Using an online testing platform (Qualtrics, www.qualtrics.com), each observer reviewed the same 99 observations in the same order. We randomly ordered the observations both by species and by PSTQ. Observers were also given a PDF of all the observations to reference previous observations without exiting the online testing platform. There was no time limit to complete the test and observers could exit and return to the test at a later time if they became fatigued. If an observer had seen an observation prior to the test (e.g. the observer also submitted the original observation) they were instructed to skip that observation. Observers were instructed to categorize each observation into one of four categories: 1) definitively wolverine; 2) likely wolverine; 3) identity unknown; or 4) definitively not wolverine (Table 4).

Analysis

We looked at the cumulative performance of all observers to produce an overall false-positive error and true-positive detection rate. Due to small sample sizes within demographic groups, we used descriptive statistics (mean) with confidence intervals to compare relative reliability of various observer groups and the influence of PSTQ and species on reliability.

To compare the relative effectiveness of various demographic groups we used the mean value of the error and detection rate by individual observer and then averaged these means across every observer within the demographic group along with a confidence interval. To evaluate the effect of PSTQ on how observations were categorized by observers, we took the mean value of how often an individual observer categorized an observation as identifiable versus unknown within each PSTQ level and then averaged those to produce

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yahaana Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Summary of observations used in the test by species, photographic snow track quality (PSTQ) rating and number of photos per observation. Overall percentages are given for the total of non-wolverine observations, total of wolverine observations, and total of all observations. Empty cells represent categories with zero observations.

	PSTQ rating (n)				No. of photos used (n)				
Species (n)	1	2	3	4	1	2	3	4	≥5
American marten (1)				1			1		
Black bear (1)			1			1			
Bobcat (1)		1			1				
Dom. dog (1)		1					1		
Fisher (13)	3		5	5		1	5	3	4
Red fox (2)			1	1			1	1	
Hoary marmot (2)	1	1				1	1		
Lynx (21)	4	1	8	8		2	10	5	4
Mountain lion (3)	2	1					1	2	
River otter (2)	1	1						2	
Wolf (4)	1	3					2	2	
Total non-wolverine observations (51)	23.5% (12)	17.6% (9)	29.4% (15)	29.4% (15)	2.0% (1)	9.8% (5)	43.1% (22)	29.4% (15)	15.7% (8)
Wolverine (48)	35.4% (17)	16.7% (8)	29.2% (14)	20.8% (10)	4.2% (2)	16.7% (8)	31.3% (15)	27.1% (13)	22.9% (11)
Total all observations (99)	**29.3% (29)**	**17.2% (17)**	**29.3% (29)**	**25.3% (25)**	**3.0% (3)**	**13.1% (13)**	**37.4% (37)**	**28.3% (28)**	**19.2% (19)**

Table 3. Overall demographic characteristics of the 29 observers at the time of testing (November 2021–January 2022).

Demographic	No. of observers
Academic degree	
No scientific degree	2
Bachelor of Science	11
Master of Science	12
PhD	4
CyberTracker certification level	
No certification	9
Level 2	3
Level 3	10
Specialist	7
Self-identified expertise level	
Amateur	5
Professional	15
Expert	9
No. of times observed putative wolverine tracks in the field	
Zero observations	10
< 10 observations	11
10–20 observations	5
> 20 observations	3
Past experience reviewing photos of tracks	
Casually	21
Systematically	8

a final mean for each PSTQ. We produced these numbers based on observations categorized as 'definitively wolverine' and by grouping all observations categorized as 'definitively wolverine' and 'likely wolverine'. To test whether the number of photos per observation affected observer outcomes we used a two-sided Kendall's tau rank correlation test evaluated at $\alpha = 0.05$ in R (R Core Team 2023, www.r-project.org).

Results

False-positive errors and true-positive detections

Overall, observers made very few false-positive errors of wolverine when they categorized an observation as 'definitively wolverine'. Twenty-one of the 29 observers made zero false-positive errors. Cumulatively, observers reviewed a total of 2851 observations (29 observers reviewing 92–99 observations each), 1378 of which were wolverine and 1473 of which were other species. Out of the 1473 opportunities to make false-positive errors only 12 were made, an average mean error rate of 0.9% (confidence interval ± 0.6%). When we combined categories of both 'definitively wolverine' and 'likely wolverine', 8 of 29 observers made zero false-positive errors and the averaged mean false-positive error rate for individual observers increased to 3.5% (± 1.3%).

Observers correctly identified tracks as 'definitively wolverine' with an averaged mean of 42.6% (± 5.3%) across all of the observations they reviewed. When we combined observations categorized as 'definitively wolverine' and 'likely wolverine' this increased to 65.6% (± 3.9%). Observers categorized 25.2% (± 4.0) of all observations as 'unknown' when observations did not provide enough detail to make a definitive identification of any sort by the observer.

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yukoma Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4. Categories for classifying the identification of observations.

Category 1: Definitively wolverine	Tracks are the correct size and overall shape. If there is limited detail in the tracks, the track pattern is one of several that is distinctive for this species for the size of the tracks and trail. When tracks show details of the animal's foot, all of the characteristics that are observable are typical for wolverine (e.g. shape, size and distribution of toes and palm pad, length and orientation of the claws). The scale used to identify the size of the tracks is unambiguous. If the track pattern is ambiguous, such as a direct register walk or trot pattern, the individual tracks must present sufficient distinctive characteristics so as to rule out any other species. Identification must include a minimum of two observable characteristics that preclude any other species
Category 2: Likely wolverine	All details that are observable in the tracks, track pattern, and context fit for wolverine, but there is not enough evidence to definitively rule out one or more other species that could make similar-appearing tracks. Photographs that have clear details that fit for wolverine but are lacking an unambiguous scale fall into this category. Definitive identification is not possible
Category 3: Identity unknown	Not enough information in the reported tracks to definitively identify the tracks as either a wolverine or not a wolverine. While it is possible that the tracks could be made by a wolverine, there is not enough evidence to suggest this is more likely than one or more other species
Category 4: Definitively not a wolverine	These observations have enough details to definitively rule out wolverine. The tracks may be identifiable as another species. At a minimum there are characteristics in the tracks or track pattern that eliminate the possibility that the tracks were made by a wolverine

However, observations with clear morphological details and a recognizable track pattern produced true-positive identifications defined as 'definitively wolverine' 74.5% (± 1.1%) of the time. Combining 'definitively' and 'likely wolverine', true-positive detection rates for these observations increased to 88.0% (± 4.8%). As predicted, observations without clear details in footprints and no recognizable track pattern produced a greater number of 'unknown' categorizations than observations with these details.

Role of observer demographics

Overall, there was a high degree of consistency across different demographic groups for rates of correct identifications and errors. The mean false-positive error rates between varying levels of education, the number of times observers had observed putative wolverine tracks in the field, or their previous experience with systematically reviewing track photos all fell within the confidence intervals for one another (Fig. 3). Across the different ways that we grouped observers by demographic characteristics, those with a Specialist certification (n = 7) were the only demographic group that made zero false-positive errors.

Observers with Level 3 and Specialist certifications in wildlife tracking made more true-positive detections than those with no certification or Level 2 certification (Fig. 3). The range of the confidence interval between these two groups did not overlap when the observer categorized an observation as 'definitively wolverine' and when we combined both 'definitively wolverine' and 'likely wolverine' categories.

Species confused for wolverine

Of the 12 false-positive errors that were made in total, six were fisher, five were lynx, and a single false-positive error came from a river otter observation. Lynx comprised 41.2% of the non-wolverine tracks in the test (n = 21) and 41.7% of the total false-positive errors in the study. Observers confused lynx tracks for wolverine 0.8% (± 0.8) of the time. Fisher comprised 25.5% of the non-wolverine tracks in the test (n = 13) and 50.0% of the total false-positive errors. Cumulatively, observers confused fisher tracks for wolverine 1.6% (± 1.4) of the time. There were only two occasions where a single observation led to more than one false-positive error across all 29 observers (two false-positive errors for a single observation of lynx tracks and four false-positive errors for a single observation of fisher tracks).

Influence of PSTQ and number of photographs per observation

The amount and type of detail in the tracks appeared to have an influence on correct identification rates and false-positive errors, as well as on the ability of observers to categorize the observations at all (Fig. 4). Observations that included both clear details of individual tracks and recognizable track patterns (PSTQ 1) produced the highest rate of true-positive detections (74.5% ± 6.6), followed by observations that included details in individual tracks but lacked a discernible track pattern (PSTQ 2, 40.2% ± 10.3). Observations without details in tracks or a recognizable track pattern (PSTQ 4) produced the highest rate of unknown identification (53.5% ± 7.9) as well as the lowest rate of true-positive detections (3.8% ± 2.9). We found no correlation between the total number of photos within a single observation and true-positive (tau = 0.5, p = 0.11) or false-positive rates (tau = 0.32, p = 0.29).

Discussion

We found that a wide variety of observers can distinguish wolverine from other species with similar tracks that overlap the distribution of wolverine in North America from photographs of tracks in snow. Most importantly, observers were able to avoid misidentifications of wolverine with a high degree of accuracy. The ability to avoid false-positive errors is a foundational element of reliability of a dataset (McKelvey et al. 2008). For a species of conservation concern such as wolverine, it is of critical importance to determine the presence of the species where such identification is used, to direct limited funds and resources towards research and conservation efforts (McKelvey et al. 2008).

Like other remote detection methods, it was not possible for observers to definitively positively identify every set of

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yakama Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

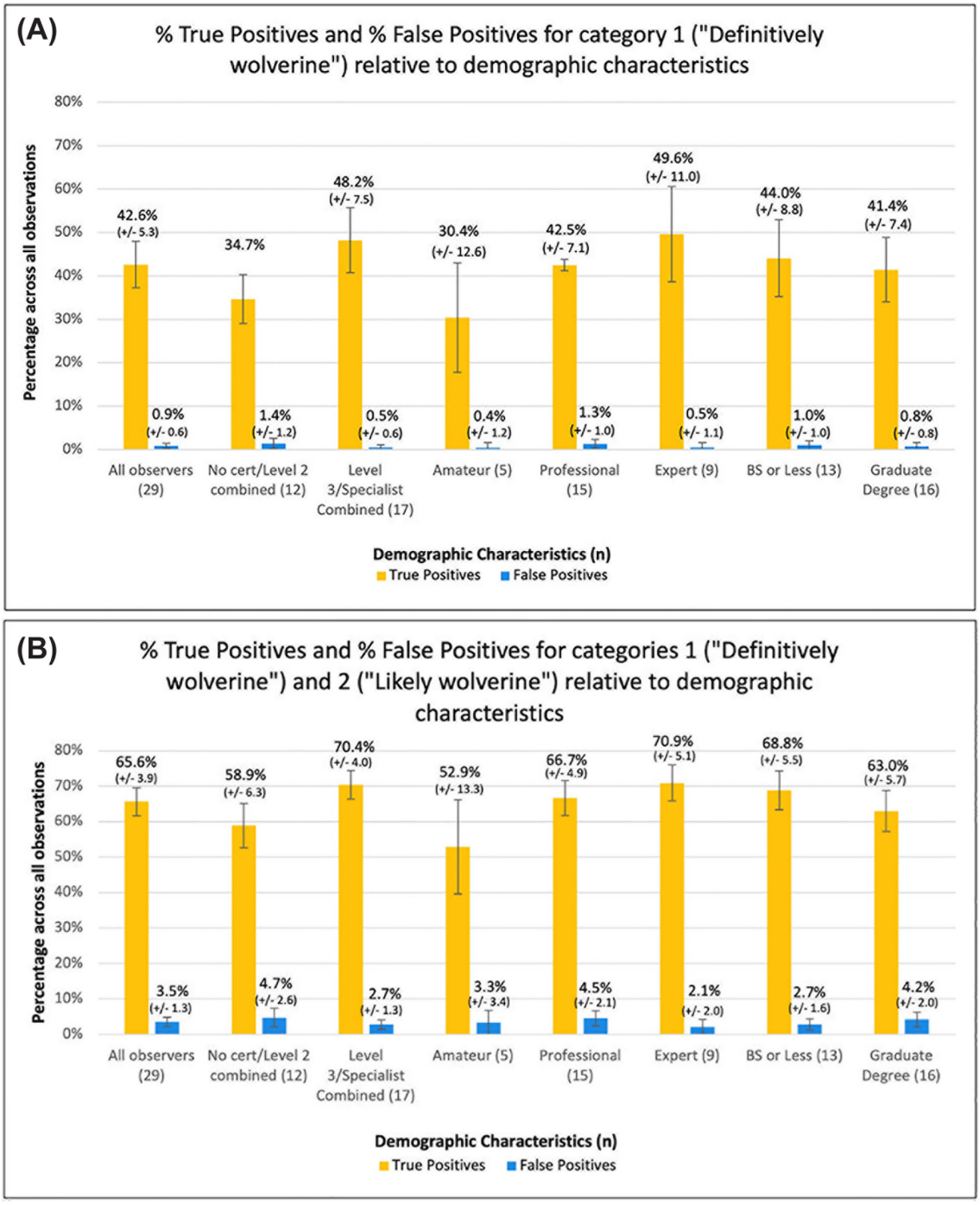


Figure 3. Performance based on observer demographics at the time of testing (November 2021–January 2022). True-positive detection rate and false-positive error rates were calculated for both (A) exclusively category 1 ('definitively wolverine') and (B) when category 1 was lumped with category 2 ('likely wolverine').

wolverine tracks shown to them (Gooliaff and Hodges 2018, Potter et al. 2019, Zett et al. 2022). Rates of true-positive detection were comparable with reported detection rates from remotely collected genetic samples (McKelvey et al. 2006, Ulizio et al. 2006). These results suggest that the use of track identification can produce datasets with low false-positive error rates and allow for use of low-skilled field observers and members of the public (community scientists) to make initial observations while experts can confirm identification out of the field, thereby greatly increasing the survey effort, distribution, and longevity of a study.

Observer demographics

While all demographic groups were able to avoid false-positive errors to a high degree, specific expertise in wildlife tracking led to higher detection rates. These findings were

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yahaana Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

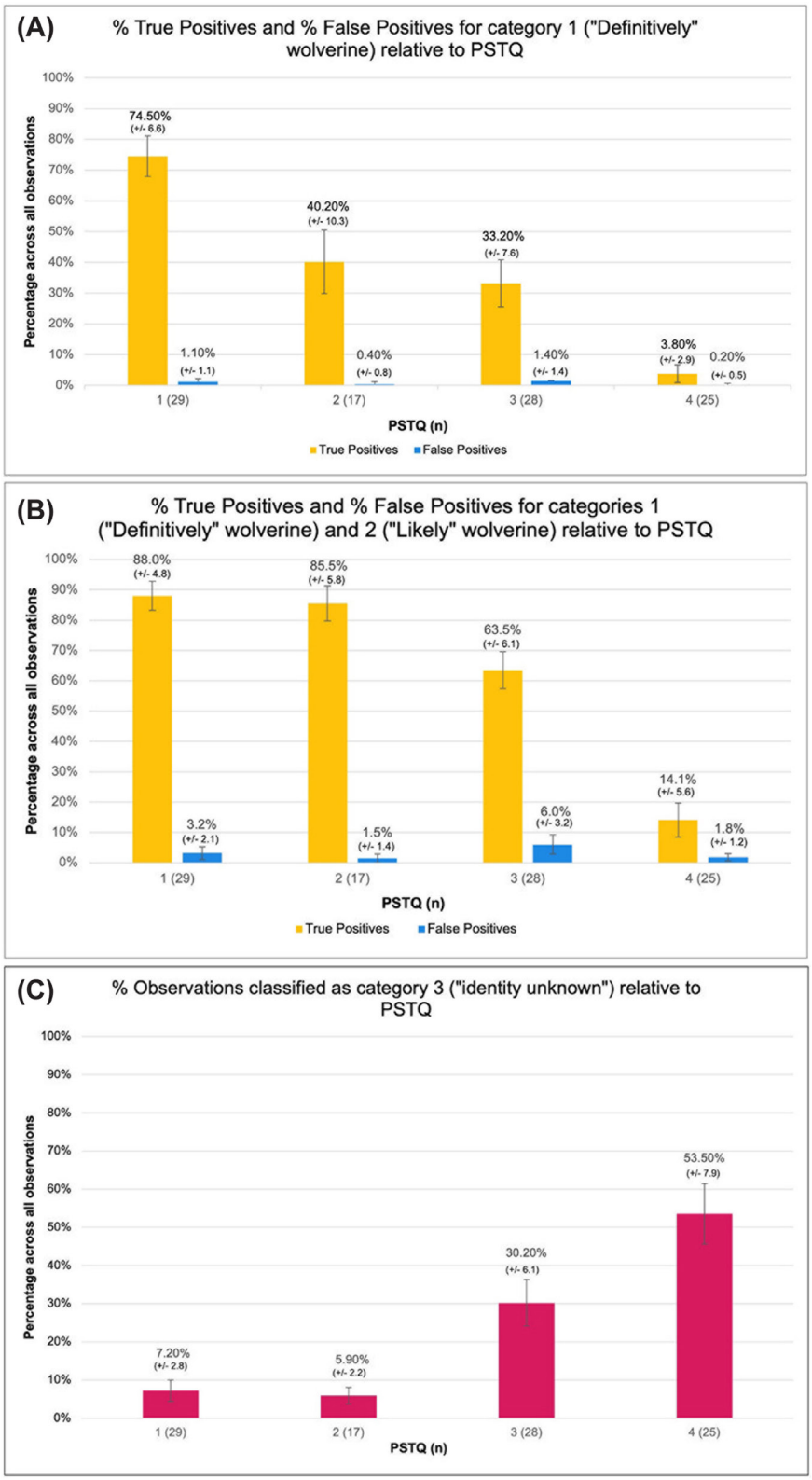


Figure 4. Influence of photographic snow track quality (PSTQ) on true-positive and false-positive rates for all observers. True-positive detection and false-positive error rates were calculated for both (A) exclusively category 1 ('definitively wolverine') and (B) when category 1 was lumped with category 2 ('likely wolverine'). (C) represents the overall rate of category 3 ('identity unknown').

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yakama Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

consistent with previous research that showed highly skilled wildlife trackers in southern Africa with no academic training produced highly reliable results (Stander et al. 1997), and specific training in wildlife tracking was helpful for biologists charged with track identification (Evans et al. 2009).

Predictably, lumping observations categorized as 'definitively wolverine' and 'likely wolverine' yielded both an increase in the total true-positive detections as well as an increase in the number of false-positive errors. This finding suggests that observers can evaluate their own reliability in their stated level of confidence. Additionally, observations with a low PSTQ were more likely to lead to an 'unknown' identification, which also suggested that observers were able to discriminate between photographs that yielded reliable results and those without adequate information for identification.

Avoiding false-positive errors

While the overall number and rate of false-positive errors in this study is low, the factors that produced them are informative and lead to guidelines for avoiding similar mistakes for projects reviewing track observations in the future.

Lynx and fisher were predicted to be the most difficult species to distinguish from wolverine. In certain snow conditions, lynx have tracks that look distinctively different from wolverine but are similar in size. However, when no detail is present in individual tracks besides their overall shape, their tracks can look quite similar. Wolverine and lynx have some track patterns that appear similar (e.g. when walking or trotting) and others that are distinctive from each other (e.g. loping patterns used by wolverine versus the bound used by lynx for high-speed travel). The rate of false-positive errors for lynx tracks was the same as the overall error rate, suggesting lynx are no more difficult to distinguish from wolverine than other species, except in as much as their tracks are often found sympatrically with wolverine.

Fisher leave tracks that look very similar to, but are distinctively smaller than, wolverine. The two closely related species can create all of the same track patterns. Fifty percent of all false-positive errors in the study were from tracks made by fisher. Test observations included photos of both smaller female and larger male fishers (those closest in size to a wolverine). However, all false-positive errors came from fisher observations where the tracks are too small for any of the published size ranges given to observers for wolverine tracks. Therefore, to ensure fisher tracks do not become false-positive errors, researchers may require all definitive identifications of wolverine to objectively show tracks that fall within the expected size range for the species.

CWP uses two expert reviewers to independently categorize each observation received. When there is a consensus in category between reviewers, the categorization of the observation is set. In instances where there is disagreement between reviewers, the most conservative classification is used to avoid false-positive errors (e.g. 'likely wolverine' rather than 'definitively wolverine' or 'unknown' rather than 'likely wolverine'). A more liberal approach is reasonable where the research goal is a larger set of likely detections but the ability to avoid false-positive errors absolutely is less of a concern. Alternatively, the demonstrated error and detection rates found here can be used as the basis for accounting for uncertainty in models built on data collected with this method.

Influence of PSTQ

Observations that included clear details in individual tracks (with or without track patterns associated with them) produced the highest rate of true-positive detections of wolverine. This suggests that photographs with clear track details (PSTQ 1 and 2) will provide the basis for the most accurate dataset.

Observations with limited track detail or a discernable track pattern (PSTQ 4) did not produce an increase in false-positive errors but did increase the likelihood that the observation would be categorized as 'unknown'. These sorts of observations are common in community science submissions (CWP unpubl.). This suggests that their inclusion in an analysis will not substantially increase the number of accurate detections, nor will they introduce a greater error rate.

Advantages and disadvantages of photographic track identification

Given that snow track evidence can be reliably identified by qualified observers, this method has certain advantages. Field observers do not need a high degree of training to identify potential tracks of target species to make observations in the field, which lends itself to community science efforts and inclusion of unskilled field personnel in collecting observations. This method requires no specialized equipment or storage, transport, and analysis of fragile specimens, making it highly cost effective. In appropriate field conditions, track identification is a more consistent way to detect carnivores than remote cameras (Pirie et al. 2016). The tracking method in our study can be easily integrated into recreational pursuits of the general public, while observations can be reviewed in bulk at a later date as opposed to one at a time in the field. In comparison with genetic sampling, the interpretation of evidence is more straightforward with less room for contamination, and a more transparent and repeatable external review of the original evidence.

Conversely, our data suggested that a certain number of wolverine observations will not be categorized as 'definitively wolverine', which will affect how these data can be used, with both the skill of the observer and the quality of the information presented in photographs affecting the positive detection rate. Analysis of these data can take this fact into account in the interpretation of their results. We suggest research projects that employ our method for collecting track photographs in the field provide clear guidelines for field personnel to obtain the highest-quality dataset possible.

We chose a narrow study design that mimics the types of data a community science project receives from the public. To make broader inferences about the accuracy of track identification from photographs of a variety of wildlife species, a study design could be considered in which observers are not

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yakama Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

instructed to look for a specific species and the number of observations is more equally distributed across species.

Given that wildlife tracking is a component of traditional ecological knowledge from around the world, it is not surprising that a wide range of experienced contemporary observers are able to perform this skill in a modern context with a high degree of accuracy. We found the published assumption that track identification produces unreliable data (McKelvey et al. 2006, Ulizio et al. 2006, Franklin et al. 2019) appears to be false in this application.

Research implications

Photographs of wolverine snow tracks can provide reliable, definitive evidence of the presence of wolverine when identification is made by the appropriate observer or set of observers. Wildlife tracking skill is independent of scientific training, and studies using track evidence should ensure that observers interpreting observations have the appropriate expertise. While quality genetic samples can provide much more information about an individual animal than visual observation of tracks alone, the growing tendency to rely on genetic evidence over track evidence for simple species identification perhaps reflects the lack of confidence in this method by those designing surveys, rather than an intrinsic lack of reliability of the method. Snow track evidence combined with community science data collection methods can substantially increase the power of monitoring efforts, with numerous other benefits including community engagement, transparency in research methods, long-term monitoring, and exploring the relationship between wolverine land use and recreational human land use. Replication of this track identification method for other species besides wolverine could expand the use of this inexpensive and accessible monitoring method for a variety of other wide-ranging species of conservation concern.

Acknowledgements – Special thanks to the individuals and organizations that submitted observations and those that participated as observers in our test: Swan Valley Connections (13 final observations), Wildlife Conservation Society Canada (55 final observations), J. Rohrer (6 final observations), Cascades Carnivore Project, C. Raley (2 final observations), R. Klafki (2 final observations), K. Shelton (1 final observation), G. Spence (1 final observation), and B. McConnell, C.M. Vanbianchi, G. Davidson, A. Lieberg, P. Taylor, T. Glass, B.V. Woodruff, S. Engel, J. Litke, R. Long, M. Barrueto, P. McKay, C. McFarland, J. Evans, S. Raymond, A. Martin, M. Kynoch, K. Aubry, M. Boudreau, M. Mayernick, M. Reynerson, J. Poppele, R. Rafferty, and S. Fitkin. We are grateful to the thoughtful review of our test design and observations provided by J. M. Elbroch (6 final observations). Thank you to S. Raymond and B. Windell for their assistance with revisions to this manuscript. Thanks to members of the Washington State Wolverine Research Working Group for your questions and support in this endeavor.

Funding – This research and manuscript preparation was funded by the Cascades Wolverine Project thanks to grants from Patagonia and the Charlotte Martin Foundation, as well as numerous individual donations in support of CWP.

Conflict of interest – As of the date of this publication, the lead author of this paper is on the Board of Directors for Tracker Certification North America, the entity responsible for overseeing Cybertracker Conservation certifications in North America. He also works as an independent contractor implementing field evaluations for this system.

Permits – This study was approved by the Institutional Review Board of the University of Utah for research involving human subjects (study no. 00146972).

Author contributions

David E. Moskowitz: Conceptualization (lead); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (lead); Project administration (lead); Resources (equal); Writing – original draft (equal); Writing – review and editing (equal). **Anna C. Machowicz**: Conceptualization (supporting); Data curation (lead); Formal analysis (equal); Investigation (equal); Methodology (supporting); Project administration (supporting); Resources (equal); Visualization (lead); Writing – original draft (equal); Writing – review and editing (equal). **Matthew A. Scrafford**: Conceptualization (supporting); Formal analysis (supporting); Funding acquisition (supporting); Investigation (supporting); Methodology (supporting); Resources (equal); Writing – review and editing (supporting). **Philippe Naveau**: Formal analysis (supporting); Methodology (supporting); Writing – review and editing (supporting). **Jeffrey N. Rose**: Methodology (supporting); Resources (supporting); Writing – review and editing (supporting). **Stephanie Williams**: Conceptualization (supporting); Funding acquisition (equal); Investigation (supporting); Methodology (supporting); Writing – review and editing (supporting).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/wlb3.01466.

Data availability statement

There are no additional data for this paper.

References

Aubry, K. B., Rohrer, J., Raley, C. M., Weir, R. D. and Fitkin, S. 2012. Wolverine distribution and ecology in the North Cascades ecosystem: 2012 annual report. – USDA Forest Service, Pacific Northwest Research Station.

Barrueto, M., Forshner, A., Whittington, J., Clevenger, A. P. and Musiani, M. 2022. Protection status, human disturbance, snow cover and trapping drive density of a declining wolverine population in the Canadian Rocky Mountains. – Sci. Rep. 12: 17412.

Basille, M., Herfindal, I., Santin-Janin, H., Linnell, J. D. C., Odden, J., Andersen, R., Arild Høgda, K. A. and Gaillard, J. M. 2009. What shapes Eurasian lynx distribution in human dominated landscapes: selecting prey or avoiding people? – Ecography 32: 683–691.

Copeland, J. P., McKelvey, K. S., Aubry, K. B., Landa, A., Persson, J., Inman, R. M., Krebs, J., Lofroth, E., Golden, H., Squires,

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yakama Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

J. R., Magoun, A., Schwartz, M. K., Wilmot, J., Copeland, C. L., Yates, R. E., Kojola, I. and May, R. 2010. The bioclimatic envelope of the wolverine (*Gulo gulo*): do climatic constraints limit its geographic distribution? – Can. J. Zool. 88: 233–246.
CyberTracker2022. Tracker certification. – https://cybertracker.org/tracking/tracker-certification/.
Davison, A., Birks, J. D. S., Brookes, R. C., Braithwaite, T. C. and Messenger, J. E. 2002. On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. – J. Zool. 257: 141–143.
Evans, J. W., Evans, C. A., Packard, J. M., Calkins, G. and Elbroch, M. 2009. Determining observer reliability in counts of river otter tracks. – J. Wildl. Manage. 73: 426–432.
Franklin, T. W., McKelvey, K. S., Golding, J. D., Mason, D. H., Dysthe, J. C., Pilgrim, K. L., Squires, J. R., Aubry, K. B., Long, R. A., Greaves, S. E., Raley, C. M., Jackson, S., MacKay, P., Lisbon, J., Sauder, J. D., Pruss, M. T., Heffington, D. and Schwartz, M. K. 2019. Using environmental DNA methods to improve winter surveys for rare carnivores: DNA from snow and improved noninvasive techniques. – Biol. Conserv. 229: 50–58.
Gardner, C. L., Lawler, J. P., Hoef, J. M. V., Magoun, A. J. and Kellie, K. A. 2010. Coarse-scale distribution surveys and occurrence probability modeling for wolverine in interior Alaska. – J. Wildl. Manage. 74: 1894–1903.
Gooliaff, T. J. and Hodges, K. E. 2018. Measuring agreement among experts in classifying camera images of similar species. – Ecol. Evol. 8: 11009–11021.
Grigione, M. M., Burman, P., Bleich, V. C. and Pierce, B. M. 1999. Identifying individual mountain lions *Felis concolor* by their tracks: refinement of an innovative technique. – Biol. Conserv. 88: 25–32.
Halfpenny, J. C., Thompson, R. W., Morse, S. C., Holden, T. and Rezendes, P. 1995. Snowtracking. – In: Zielinski, W. J. and Kucera, T. E. (eds), American marten, fisher, lynx and wolverine: survey methods for their detection. Technical report. Pacific Southwest Research Station. Forest Service, US Department of Agriculture, pp. 91–119.
Heinemeyer, K., Squires, J., Hebblewhite, M., O'Keefe, J. J., Holbrook, J. D. and Copeland, J. 2019. Wolverines in winter: indirect habitat loss and functional responses to backcountry recreation. – Ecosphere 10: e02611.
Krebs, J., Lofroth, E. C. and Parfitt, I. 2007. Multiscale habitat use by wolverines in British Columbia, Canada. – J. Wildl. Manage. 71: 2180–2192.
Liebenberg, L. 1990. The art of tracking the origin of science. – David Philip Publishers.
Liebenberg, L., Elbroch, M., Louw, A., Pinto, J. and Masia, W. 2018. CyberTracker Tracker Certification. – Standards report from the CyberTracker Certification Standards Committee.
Long, R. A. and MacKay, P. Database managers 2024. Washington Wolverine Research and Monitoring Group: database of verifiable wolverine detections. – Woodland Park Zoo.
Machowicz, A., Vanbianchi, C. and Windell, R. 2022. Recreation and wildlife in Washington: considerations for conservation. A report on current knowledge. Technical report. – Home range Wildlife Research, Conservation Northwest.
Matyushkin, E. N. 2000. Tracks and tracking techniques in studies of large carnivorous mammals. – Zool. Zh. 79: 425–429.
McKelvey, K. S., Aubry, K. B. and Schwartz, M. K. 2008. Using anecdotal occurrence data for rare or elusive species: the illusion of reality and a call for evidentiary standards. – BioScience 58: 549–555.
McKelvey, K. S., Von Kienast, J. V., Aubry, K. B., Koehler, G. M., Maletzke, B. T., Squires, J. R., Lindquist, E. L., Loch, S. and Schwartz, M. K. 2006. DNA analysis of hair and scat collected along snow tracks to document the presence of Canada lynx. – Wildl. Soc. B. 34: 451–455.
McKelvey, K. S., Aubry, K. B., Anderson, N. J., Clevenger, A. P., Copeland, J. P., Heinemeyer, K. S., Inman, R. M., Squires, J. R., Waller, J. S., Pilgrim, K. L. and Schwartz, M. K. 2014. Recovery of wolverines in the western United States: recent extirpation and recolonization or range retraction and expansion? – J. Wildl. Manage. 78: 325–334.
Pirie, T. J., Thomas, R. L. and Fellowes, M. D. E. 2016. Limitations to recording larger mammalian predators in savannah using camera traps and spoor. – Wildl. Biol. 22: 13–21.
Poley, L. G., Magoun, A. J., Robards, M. D. and Klimstra, R. L. 2018. Distribution and occupancy of wolverines on tundra, northwestern Alaska. – J. Wildl. Manage. 82: 991–1002.
Potter, L. C., Brady, C. J. and Murphy, B. P. 2019. Accuracy of identifications of mammal species from camera trap images: a northern Australian case study. – Austral Ecol. 44: 473–483.
R Core Team 2023. R: a language and environment for statistical computing. – R Foundation for Statistical Computing, Vienna.
Squires, J. R., McKelvey, K. S. and Ruggiero, L. F. 2004. A snowtracking protocol used to delineate local lynx, *Lynx canadensis*, distributions. – Can. Field Nat. 118: 583–589.
Stander, P. E., Ghau, I. I., Tsisaba, D. O., and Oma, I. I., VI. 1997. Tracking and the interpretation of spoor: a scientifically sound method in ecology. – J. Zool. 242: 329–341.
Switalski, A. 2016. Snowmobile best management practices for Forest Service travel planning: a comprehensive literature review and recommendations for management. – J. Conserv. Plan. 12: 1–7.
Ulizio, T. J., Squires, J. R., Pletscher, D. H., Schwartz, M. K., Claar, J. J. and Ruggiero, L. F. 2006. The efficacy of obtaining genetic-based identifications from putative wolverine snow tracks. – Wildl. Soc. B 34: 1326–1332.
Wright, J. D. and Ernst, J. 2004. Effects of mid-winter snow depth on stand selection by wolverines, *Gulo gulo luscus*, in the boreal forest. – Can. Field Nat. 118: 56–60.
Zett, T., Stratford, K. J. and Weise, F. J. 2022. Inter-observer variance and agreement of wildlife information extracted from camera trap images. – Biodivers. Conserv. 31: 3019–3037.
Zielinski, W. J. and Kucera, T. E. (eds) 1995. American marten, fisher, lynx and wolverine: survey methods for their detection. Technical report. – Pacific Southwest Research Station. Forest Service, US Department of Agriculture.
Zielinski, W. J. and Schlexer, F. V. 2009. Inter-observer variation in identifying mammals from their tracks at enclosed track plate stations. – Northwest. Sci. 83: 299–307.

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01466 by Yakama Nation Fisheries, Wiley Online Library on [10/06/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License