An evaluation of the scientific quality of Finnish wolf population monitoring

Henrik Andrén[1], Mikko Mönkkönen[2] & Otso Ovaskainen[3]

[1]Professor in Wildlife Ecology, Grimsö Wildlife Research Station, Department of Ecology, Swedish University of Agricultural Sciences, SE-730 91 Riddarhyttan, Sweden

[2] Professor in Applied Ecology, Department of Biological and Environmental Sciences, P.O. Box 35, FI-40014 University of Jyväskylä

[3]Professor in Mathematical Ecology. Department of Biosciences, P.O. Box 65, FI-00014, University of Helsinki, Finland

Summary

Acquiring reliable estimates of the number of wolves in Finland is important because wolf is an endangered and charismatic large predator but also a game animal. To maintain a viable wolf population and to manage the population in a sustainable way requires accurate estimates of the number of territories with pack (reproducing part of the population), number of territorial pairs (potential future territories with packs), and lone wolves. The main goal of this evaluation is to assess the scientific quality of the Finnish wolf population monitoring program, including both data collection processes and analysis methods, and to provide insights for improvements also from the cost-efficiency point of view. The evaluation is based on material and information provided by Luke personnel, both as written documents and via personal communication, and it also utilizes experiences from other European wolf population monitoring schemes. The strength of the Finnish wolf population monitoring program is that it is a low cost scheme that widely involves citizens in data collection. We however identified two major weaknesses of the monitoring scheme that make its results potentially unreliable and difficult to communicate to the public. The first weakness is the sparsity of the underlying data, and the fact that the largest part of the data consists of only partially validated point observations, rather than e.g. tracks followed for several kilometers, or DNA samples. The second weakness is that the analytical methods used to derive the population size estimates from the data are not well documented, and they contain a number of subjective choices which make a critical evaluation of their accuracy difficult. We conclude that the documentation and development of the data analysis pipeline should be given a priority, as that will enable both the derivation of more accurate population size estimates as well as an improved communication between Luke and the other stakeholders. Increasing substantially the data quality would require adopting elements from monitoring schemes used elsewhere (e.g. Scandinavia), which is possible only if additional resources become available.

1. Background

Wolf is simultaneously an endangered and hunted species that causes damage and creates both negative and positive emotions. Therefore, its management involves potentially conflicting objectives and political pressures. Consequently, wolf population monitoring in Finland has two main objectives.

The first objective of wolf population monitoring is to provide information of the current population status of the species in Finland for conservation purposes. Wolf is listed in the EU Habitat Directive, and accordingly, EU countries have to report the population status of the species every 6th year to the EU commission. Wolf is also red-listed in Finland (IUCN category Endangered), and assessing the threat status requires knowledge about population size and trend.

The second objective of wolf population monitoring is to provide information of the current population status of the species for game management purposes. Wolf is a game animal for which the Ministry of Agriculture and Forestry sets annual hunting quota. Setting the quota according to sustainable management principles requires information of wolf population size. According to Wolf Management Plan (2015), management is based on the number of viable wolf territories. Therefore, the wolf monitoring in

Finland focuses on wolf territories (wolf packs), as the wolf management is territory-based. The aim is to maintain a wolf pack in its area, but to reduce the pack size. Number of wolf pairs is needed to estimate the total wolf population size and to forecast the wolf population development. Wolf pairs often become wolf packs in the next year.

To achieve both of these two objectives, the wolf monitoring program aims at providing information about the number of wolves living in packs, as well as the numbers of territorial pairs and single wolves. Monitoring wolves is however a difficult task, because wolves are elusive, they occur in low densities, and they are able to disperse hundreds of kilometers within few weeks. In contrast, what facilitates their monitoring is their territorial behavior and social structure, where packs or pairs can remain in the same areas over several years.

An ideal population monitoring scheme is cost-efficient, i.e. able to provide information that is adequate and accurate enough for reaching the objectives with minimum cost. In population monitoring, a sufficiently high quality of the data (amount, truthfulness and accuracy) is a basic requirement, because no analysis methods, however elegant, can remedy poor quality data. But improving the data quality comes in with a cost, leading to the questions of whether or not the current data are sufficient, and whether better quality data could be acquired without increasing the cost. An equally important issue is the reliability of the analysis methods used to derive a population size estimate from the data. Given all the uncertainties related to wolf monitoring data, deriving unbiased and reliable estimates is not an easy task.

Because of potentially conflicting goals of wolf population management, it is not only important to derive accurate estimates, but also to communicate the methods and results of the monitoring scheme in a transparent way. This is important for public acceptance of the management decisions, as well as for mutual trust between Luke and the other interest groups.

2. Goals and methods of this evaluation

The main objective of this evaluation is to provide a general picture of the focus and scientific quality of the Finnish wolf population monitoring program. Scientific quality depends on the quality of the data and on the quality of the methods used to interpret the data. By the quality of data we refer to the amount of data as well as its relevance, truthfulness, reliability and accuracy. By the quality of the methods used to handle and analyze the data we refer to their scientific rigor, potential to yield unbiased and reliable estimates, as well as their transparency and easiness to communicate both the results and the methods to the stakeholders. The goal of this evaluation is to scrutinize all these aspects, and point out the strengths, weaknesses, threats and opportunities for future development of the monitoring scheme. This evaluation report also considers the cost-efficiency of the current wolf population monitoring scheme, and provides suggestions how the efficiency can potentially be improved.

Wolf population monitoring programs have been established in other European countries such as in Sweden and Norway (Scandinavian wolf; Liberg et al. 2012), in France (Alpine and eastern Pyrennees populations; Duchamp et al. 2012) and in Italy (Alpine and Apennine populations; Galaverni et al. 2016). This evaluation also aims to provide pertinent insights from these programs for the development of the Finnish monitoring scheme.

The evaluation process was launched in a workshop organized by Luke May 18th 2016. In this meeting, representatives of Luke defined the goals of the evaluation, and provided information about the wolf population monitoring scheme to the evaluation group both as written documents as well as oral presentations. The evaluation group then identified the need for obtaining additional information about the analytical methods used to identify and delineate territories and to estimate the sizes of the wolf packs. Such additional information was provided to the evaluation group on June 29th 2016 (Samuli Heikkinen, personal communication). In addition to the provided material, scientific literature as well as material related to wolf monitoring programs from other countries (Sweden, Norway, France, Italy) was utilized.

The evaluation work was conducted during summer 2016, and it culminated into a workshop held in August 17-19 2016, during which the main part of this evaluation report was written. The Luke representatives checked the report for identifying possible factual errors, after which the evaluation group produced the final version.

3. Description of monitoring methods

3.1. Wolf monitoring in Finland

Field work. The wolf monitoring in Finland is mainly based on wolf observations from the public (mainly hunters) from the period from 1st September to 28th February each winter. Local large carnivore contact persons (ca. 2000 persons) within game management units (290 units) check and register some of these wolf observations. The number of wolf observations have increased over the last 5 years and in winter 2012/2013 2312 wolf observations were registered in TASSU by local large carnivore contact persons, whereas in winter 2015/2016 the corresponding number is 6551 wolf observations. For each registered observation, the number of wolves and the geographical coordinates are uploaded in the database TASSU, as well as the information on whether the observation has been checked in the field. On average, the local large carnivore contact persons check about 50 % of all wolf observations registered in TASSU in the field. In winter 2012/2013 most of the registered wolf observations were of single wolves (n = 1465). Number of observation of wolf pairs was 461 and number of observation of wolf packs (three or more wolves) was 386. Also in winter 2015/2016 most observations were of single wolves (n = 3885), 1410 observations were from pairs of wolves and 1256 observation were from three or more wolves.

Luke has five field technicians working with wolf monitoring in the field. They snow-track wolves and make notes about territory marking, estimate pack size etc. In some areas wolf scats are collected for DNA analyses that are used to improve the interpretation of the number of wolf territories in an area. Some wolves are also marked with GPS-collars within the wolf research projects. These GPS-collared wolves are very useful for interpreting the number of wolf territories in an area.

Data sets. The main data set for estimating the number of wolf territories (wolf packs) comes from the database TASSU. The subset of observation of wolf packs is the basis for estimating number of wolf territories in an area. In winter 2012/2013 there were 386 observations of wolf packs and the estimated number of wolf territories was 17 in that same winter, i.e. on average 22 observations of wolf packs per wolf territory. Similar numbers for winter 2015/2016 were 1256 observations of wolf pack and 26 estimated wolf territories, with an average of 48 observation of wolf pack per wolf territory. However, some wolf territories were defined based on only five observations. Furthermore, several observations of wolf packs were from outside the areas that were identified as wolf territories. In addition, pairs of wolves are also estimated using the subset of observation of wolf pairs outside defined wolf territories (wolf packs).

Additional data from DNA analyses and GPS-collared wolves are used to improve the interpretation of the number of wolf territories in some areas.

***Interpretation/ analysis*.** The number of wolf territories in an area is estimated by visually delineating observations of wolf packs (three or more wolves) into territories. Observations of wolf packs close to one another are grouped into the same wolf territory, whereas observations of wolf packs further away from one another are assumed to be from different wolf territories. The area of an average wolf territory (1200 km^2) is used to distinguish between observations close to one another and observation further apart. Data of number of wolves from the observations are also used to group or separate observations into wolf territories (Samuli Heikkinen, personal communication).

When available, data from GPS-collared wolves are also used the delineate wolf territories. GPS-data gives a very good description of the spatial extent of a wolf territory. Results from DNA-analyses are also used to group observation into the same territory or to separate observations between different territories. In addition, observations of wolf pairs outside defined wolf territories (wolf packs) are grouped into the same pair or separated into different pairs in a similar way as wolf territories (wolf packs) are delineated.

The total size of the Finnish wolf population is estimated by adding together all wolves found in packs (range is the minimum and maximum number of wolves observed in a territory) and pairs, and by assuming that 15 % of the total population are single wolves. For wolves belonging to transboundary packs and pairs, only 50 % are counted to the Finnish wolf population. Finally, a small number of wolves is added to the total number to account for uncertainty.

3.2. Wolf monitoring in Sweden/Norway and France.

Sweden and Norway have a shared wolf population and have a joint monitoring system and common reports (Wikenros et al 2014). The wolf monitoring in Scandinavia combines three methods; snow tracking, DNA-analysis and radio-telemetry (GPS-collared wolves), but the focus is on snow-tracking and continuous analyses of DNA-samples during the monitoring period. Around 100 field technicians are employed full time or part-time to find and follow wolf tracks and to collect DNA-samples during the monitoring season (Oct. 1 – Feb. 28). The objective of the DNA-analysis is to identify all territorial wolves at the individual level each season. Between 10-20 wolves are equipped with GPS-collars each year. All monitoring data are recorded in a common database (Rovdata) and complied in an annual report (see e.g. Wabakken et al. 2016).

In France, the total wolf population size is estimated using a multi-event Capture-Recapture model, based on systematic collections of scats and other samples containing DNA and subsequent DNA-analyses, which enables analysis at the individual level. At the "reproductive unit" scale, the wolf howling sessions provide locations of home-sites on which pup scats can be easily collected for genotyping (Duchamp et al. 2012).

3.3. Wolf monitoring costs

The Finnish wolf monitoring cost in 2016 was 257 500 €, which corresponds to about 1000 to 1500 € per wolf. This consists of salaries (researchers 30 000 €, field personnel 80 000 €), DNA analyses (40 000 €), helicopter costs (54 000 €), GPS-collars (44 000 €) and travelling expenses (9 500 €).

The total monitoring costs for wolves in Sweden and Norway together was estimated to about 1 500 000 € in 2011 and the estimated wolf population was about 300 individuals. The recent DNA program is designed and financed to process 400 samples per year for monitoring (about 200 000 € per year). The total monitoring cost corresponds to about 5000 € per wolf (Liberg et al. 2012).

4. Strengths, weaknesses and threats of the Finnish wolf monitoring scheme

4.1. Strengths

The basis of the monitoring scheme are observations collected by layperson, and therefore it involves local people (e.g. hunters and conservationists) living together with wolves. This is a strength of this monitoring scheme, as it allows citizens to participate in the data collection and also gives them some ownership of the results.

Another strength is that, compared to e.g. Sweden, Norway or France, the Finnish monitoring system is a low cost scheme. The quality control of the observations is done by local contact persons (petoyhdyshenkilöt) who check and validate wolf (and other large carnivore) observations and enter the approved observations into TASSU-database. This is a crucially important part of the monitoring scheme because field surveys have shown that up to 50% of observations originally identified as wolf were incorrect (Siira et al. 2009). Local contact persons are volunteers, who get no compensation for costs. Local contact persons are able to check approximately 50% of the wolf observations by their own field visits. By and large, the collection and validation of wolf observations in the current monitoring scheme is a low cost way to accumulate data.

Monitoring provides long-term population data. These data are potentially of great value in scientific work, and conversely, monitoring can greatly benefit from methodological development in basic science. Therefore, tight linkages between persons in charge of designing and organizing the monitoring scheme and the wider scientific community are important. We consider as a strength of the current monitoring scheme that the persons in charge have high quality scientific training, and that they are active in collaborating with scientists outside Luke, as evidenced by their numerous high quality scientific publications.

4.2. Weaknesses

The population size estimates derived from the data and methods described above are likely to involve substantial uncertainty due to three sources of error, which we describe in some detail in this section. In addition to introducing bias, these three sources of error also make it difficult to quantify the confidence limits of the estimates.

Error source 1: not all data are reliable. Approximately half of the field observations in the TASSU database are not checked in the field by large carnivore contact persons, and thus it remains questionable whether all of them actually represent wolf tracks. Results from a pilot study (Siira et al. 2009) suggest that even 50% of wolf tracks reported by volunteers (mainly hunters) belong to other large carnivores (mainly lynx) or other animals (e.g. fox or dog) rather than wolf. While the checked observations are given more weight in the analyses than the non-checked ones, the presence of a large fraction of potentially invalid data is likely to bias the estimates. A further problem is that the misidentified observations are not likely to be a random sample of all observations in the database, as the reliability of track identification varies among volunteers and depends on the training provided to them. Thus, even a large number of observations from the same area may not reliably indicate the presence of wolf.

Error source 2: it is difficult to identify and delineate territories. Probably the most important source of error is that the data acquired for monitoring purposes do not enable a reliable identification and delineation of wolf territories. In addition to the issues related to the truthfulness of the observations (see error source 1), key issues relate to the nature of the data, the limited extent of the data, and the method used to identify and delineate the territories from the data. Identifying a territory is based on a cluster of observations in space. With limited extent of the data, it is possible that some existing territories go totally unnoticed, because there simply are not enough observations to create a cluster.

When identified, the delineation of a territory is based on visual inspection of point patterns of observations on a map. As only a small fraction of the observations involves DNA samples, it is usually not possible to separate individuals from these data. Thus, the delineation of territories contains inevitable uncertainty due to the fact that it is not easy to decide which observations should be grouped to represent a single territory. For example, two aggregates of point observations separated by forty kilometers from each other may represent the same territory or two different territories. This problem is especially severe in areas of high wolf density, as in such areas the neighboring territories are often next to each other. At the same time, it is exactly the high density areas that contribute disproportionally to the total estimate of the national population size. Further, some observations may concern dispersing wolves that are not part of an established a territory.

Comparing the manually conducted territory delineations against independent validation data based on GPS collared wolves (both provided for the purpose of this evaluation by Samuli Heikkinen as a personal communication) gave the following results: out of the ten territories that were known to exist based on GPS data, only five were identified and delineated correctly. Two of the territories were missed entirely, out of which one was however mostly in the Russian side. One territory was misinterpreted to contain two territories, and for two cases the delineated territories covered only a marginal part of the actual territories. This example demonstrates that, without additional data, delineating territories is very difficult and consequently rather unreliable. A further problem is that the method used to delineate territories is not clearly described in any publication, and it contains a large number of subjective choices. The lack of transparency in the analysis method also makes it difficult to communicate the results to the stakeholders; this lack of transparency may be one source of local people's distrust of the results and of Luke's personnel.

Error source 3: estimation of pack sizes is difficult. The estimation of the number of individuals per territory is difficult. This is because the main data source for the number of individuals per pack is the number of individuals inferred from tracks on snow. As the dynamics of the wolf packs involve continuous splitting and merging (Liberg et al. 2012), only part of the pack is likely to be present in any given observation stored in the TASSU database. Indeed, the TASSU data contain high variation in the number of individuals from repeated observations belonging to the same territory. Further, estimating the number of individuals from snow tracks is not always straightforward, with an estimated error rate of ca. 15% (Siira et al. 2009).

4.3. Threats

The accumulation of wolf observations to TASSU-database depends on the efforts of voluntary people. This dependence is a threat as the monitoring may collapse if people find their contribution not rewarding. Therefore, the current scheme would benefit from creating better incentives for citizens to participate in making observations and informing local contact persons about them. Similarly, a low cost way to improve the amount of data available for population monitoring would be to improve incentives for local contact persons so that a larger proportion of original observation can be validated.

With climate change, the snow-free period gets longer and in southern Finland totally snow free winters are likely to become more frequent. This will likely dramatically decrease the number of wolf observations and make tracking much more difficult. Therefore, the present monitoring method, critically dependent on snow tracks, will likely run into problems in the future. Consequently, alternative methods that are not so dependent on snow tracks should be emphasized.

5. Recommendations on how to improve

5.1. The path from data to population estimate should be made transparent

As described above, converting the acquired observation data into estimated numbers of territories and individuals involves a number of choices, which have a critical influence on the accuracy and reliability of the estimates. A major weakness with the current practices is that these choices are not transparently described, and many of them are made more or less subjectively by the researcher conducting the analysis. This leads to unreproducible results that lack scientific rigor, and that are hard to communicate both to the stakeholders as well as to the general public. Therefore, it is critical to document the analysis method so that in principle anyone (with appropriated training) could conduct the analyses or replicate their results, as well as to understand the assumptions made. A difficulty here is that the data stored in the TASSU and other databases are not the only data used, but that they are supplemented e.g. by phone calls and other inquiries to the game management associations.

A potential solution would be to separate the analyses into two phases, of which the first one would consist on inference that can be derived solely based on systematic data (TASSU, DNA-data, data from GPS-collared wolves), and the second one on other information that is used to support or correct the inference derived from systematic data. In this way, the analyses related to the first phase could be made transparent, and much of the analysis procedure could even be automated, thus increasing repeatability and decreasing cost. For example, human-based clustering could be replaced by computer-based clustering, which can be tuned to account for reliabilities of individual observations, distance thresholds based on knowledge of territory sizes, as well as the other criteria used in the manual clustering. Such a systematization and automatization of the data analysis pipeline would also make it possible to evaluate its accuracy against validation data, and thus to optimize the choices made so that they would likely yield on average unbiased estimates, as well as to give confidence limits to the estimates.

In the second phase, additional information would be brought to the interpretation of the data through expert opinion. The separation of the two phases would make it possible to identify and communicate which results are based on systematic analyses of systematic data, and which ones on more subjective expert opinion. The second phase could potentially also include participatory processes, where different stakeholders can provide their interpretations and insights of the more subjective issues to support analysis. Such participatory processes may motivate people to accumulate more data, and potentially also help creating mutual trust among different people and interest groups.

5.2. Instead of reporting point observations, tracks should be followed and scats collected

Scandinavian experience shows that at least 3 km of following the tracks and repeated visits to all territories are needed, both to delineate territories as well as to establish the correct group size. Tracking also enables the collection of scat and urine samples for DNA-analyses, and thus brings additional resolution to both territory delineation and the estimation of pack sizes within territories. The critical is how to achieve such data; we thus discuss the need and options for incentives below.

5.3. Incentives for volunteers to collect and validate observations

The current monitoring scheme critically depends on the participation of motivated volunteers, and may collapse if people find their contribution not rewarding. Identifying and delineating territories, a core procedure in population estimation, would greatly improve if more validated observations were available. Therefore, Luke should invest in creating incentives for volunteers to accumulate more data. We think the best option would be to adopt open science approach and create participatory procedures in data analysis and interpretation. One possibility is to create incentives to individual volunteers who deliver data. Another possibility is to create incentives to local communities, e.g. linking hunting quota to the quality of data by allowing hunting only in territories with reliable information. The third possibility is to move towards the Scandinavian system, where acquiring such data belongs to the duties of governmental organizations and thus is made by paid professionals.

5.4. Data from GPS–collared wolves should continue to be utilized

While we consider that data from GPS-collared wolves should not be the basis of monitoring, it should be continued to use to support monitoring when acquired for research purposes. First, GPS data acquired from the present year are valuable for differentiating nearby territories. Second, both past and present GPS data are valuable in developing and validating the analyses methods. We thus recommend using a larger proportion of the monitoring resources in collecting and analyzing more DNA-samples, but at the same time take the advantage of wolves that are GPS-marked for the purposes of scientific projects.

5.5. Acquiring reliable estimates requires more effort and thus more resources

Much more effort is spent in the wolf monitoring programs in Norway, Sweden and France than in that of Finland, whether effort is measured by time or money, and whether it is calibrated against the area covered or the size of the wolf population. We are confident that without additional resources to the Finnish wolf monitoring scheme, it is very difficult to substantially improve the quality of the data from which the population estimates are derived, and thus the accuracy of the estimates themselves.

5.6. Improved communication

One problem in the wolf management system in Finland is the distrust by citizens towards the results of the monitoring results. Some interest groups often argue that territory number and pack size estimates are underestimates of the true population size, while others may argue that they are overestimates. This is at least partly a communication problem, which may stem from the fact that the monitoring scheme is not transparent. When the process of transforming the observations into the number of territories, pairs and lone wolves, and finally to total population size is kept veiled, and the results are presented as a matter of truth, people may find the estimates unreliable. Also because of this, we recommend describing the analysis methods clearly and transparently, and also pointing out the potential sources of bias and error. We acknowledge the difficulty of communicating the uncertainty of scientific work to laypersons. One potentially operational way to start creating mutual trust would be to open the analytical process and communicate the monitoring results as flow of interpretations, assumptions and decisions one has to make in order to finally come up with estimates of the number of territories, pairs, loners and total population size.

6. Conclusions

The current monitoring scheme results in potentially rather imprecise population estimates, the accuracy of which is very difficult to estimate. This makes the results of the monitoring scheme also difficult to communicate, particularly when the analytical process is not clearly described anywhere. We identified several ways to improve the monitoring scheme. However, we also recognized that without additional resources it is very difficult to substantially improve the quality of the data from which the population estimates are derived, and thus the accuracy of the estimates themselves. We recommend developing incentives for volunteers participating in the data collection, and potentially including participatory processes concerning the data analysis.

Literature cited

Duchamp et al. 2011. Wolf monitoring in France: a dual frame process to survey time- and space-related changes in the population. Hystrix, the Italian Journal of Mammalogy23(1):14-28.

Galaverni, M., Caniglia, R., Fabbri, E. et al. 2016. One, no one, or one hundred thousand: how many wolves are there currently in Italy? Mamm Res 61: 13–24. doi:10.1007/s13364-015-0247-8

Liberg et al. 2012. Monitoring of wolves in Scandinavia. Hystrix, the Italian Journal of Mammalogy 23 (1): 29–34.

Siira, A., Keränen, J. & Kojola, I. 2009. Lumijälkilaskenta suurpetokantojen seurantamenetelmänä: kokemuksia Kainuun 2008 pilottihankkeesta. Riista- ja kalatalous – Selvityksiä 9/2009.
27 s.

Wabakken et al. 2016. Inventering av varg vintern 2015-2016. - Beståndsstatus för stora rovdjur i Skandinavien 1 - 2016. 49 sidor.

Wikenros et al. 2014. Förslag till samordning av inventering av varg i Norge och Sverige. - NINA Rapport 993, 83 sidor.