Sampling_draft1

`r Sys.Date()`

This manuscript (permalink) was automatically generated from DrMattG/TechBk1@76e51c3 on June 23, 2023.

Authors

- **James Martin**
  0000-0001-8937-2262 · quaildoc · gamebirddoc
  Warnell School of Forestry and Natural Resources, University of Georgia, 180 E. Green Street, Athens GA

- **Matthew J. Grainger**
  0000-0001-8426-6495 · DrMattG · Ed_pheasant
  Terrestrial Biodiversity, Norwegian Institute for Nature Research - NINA, Postbox 5685 Torgarden, 7485 Trondheim, Norway

- **Victoria Nolan**
  0000-0002-6069-963X · victorianolan8 · victoria_nolan1
  Warnell School of Forestry and Natural Resources, University of Georgia, 180 E. Green Street, Athens GA

- **William B. Lewis**
  0000-0003-4978-6027 · wblewis7
  Warnell School of Forestry and Natural Resources, University of Georgia, 180 E. Green Street, Athens GA

- **Erlend B. Nilsen**
  0000-0002-5119-8331 · ErlendNilsen · eb_nilsen
  Terrestrial Biodiversity, Norwegian Institute for Nature Research - NINA, Postbox 5685 Torgarden, 7485 Trondheim, Norway

- **Sprih Harsh**
  0000-0002-6462-7668 · spharsh
  Warnell School of Forestry and Natural Resources, University of Georgia, 180 E. Green Street, Athens GA

- **Chloé R. Nater**
  0000-0002-7975-0108 · ChloeRN · chloe_nater
  Terrestrial Biodiversity, Norwegian Institute for Nature Research - NINA, Postbox 5685 Torgarden, 7485 Trondheim, Norway

✉ — Correspondence possible via GitHub Issues

Introduction

What is wildlife science and management?

Seeking truth

Uncertainty

Environmental variation

Linguistic uncertainty

Partial observability

Partial controllability

Structural uncertainty

Opening the Gates of Management and Science

Openness to diversity of knowledge

Wildlife research and management are closely entwined fields with a large number of different stakeholders. Not only that, but the stakeholders typically represent a highly diverse group of actors with complimentary knowledge and expertise relevant to the wildlife systems and/or management challenges at hand [drescher2013]. This diversity of knowledge can be illustrated through the – admittedly very vague – concept of "expert knowledge".
In the eyes of the general public it may be the **researchers** that come to mind first when thinking about "experts" [dommett2019; funk2019]. Researchers collect and analyze data in order to provide a (quantitative) knowledge base for decision-making. Modern science is becoming increasingly more collaborative and typically involves teams of researchers with different and complimentary skills and backgrounds [**carpenter2006?**,**goring2004?**,**cheruvelil2014?**]. At the same time, there is a growing awareness in the scientific community of the need for interdisciplinary work that also makes use of what is often termed "expert knowledge" in scientific work [gosselin2018; [**rubert2021?**]].
When used by a researcher, the term "expert" often refers to any non-researcher who possesses valuable knowledge (in the context of a study/problem). In wildlife research, this typically encompasses managers and citizen scientists, and sometimes also includes indigenous people. **Wildlife managers**, occasionally also referred to as "end users", "decision makers", or "policy makers", are aware of the practical contexts framing wildlife research and often

have a good grasp of both logistical and financial implications of different types of monitoring and management strategies and interventions [gosselin2018]. They may also possess a wealth of (qualitative) knowledge about focal species and ecosystems, as well as about motivations and interests of various other stakeholders in a system [drescher2013; gosselin2018].

Some **citizen scientists** (sometimes also called "community scientists") may possess similar knowledge, at least about the former, based on experiences from their extensive time spent observing, enjoying, and collecting data on species [**miller2012?**,**mckinley2017?**]. They may also have a good understanding of potential issues and challenges with data collection in the field [**miller2012?**]. Last but not least, citizen scientists represent a portion of the general public, and may contribute to debates and decision-making with viewpoints, priorities, and values that are central to a large number of people [[**mckinley2017?**]; binley2021].

Indigenous people have been living as part of and interacting with the natural environment since long before scientific research as we know it today was conducted [berkes2017]. As a consequence, indigenous communities hold intricate knowledge about their lands and waters which is tightly linked to their culture [jessen2022] and which often covers time spans necessary for understanding long-term ecological changes [e.g. savo2016]. Indigenous knowledge is, in many ways, complementary to ecological science and the involvement of indigenous people can lead to more impactful, effective, and just research and management [ban2018].

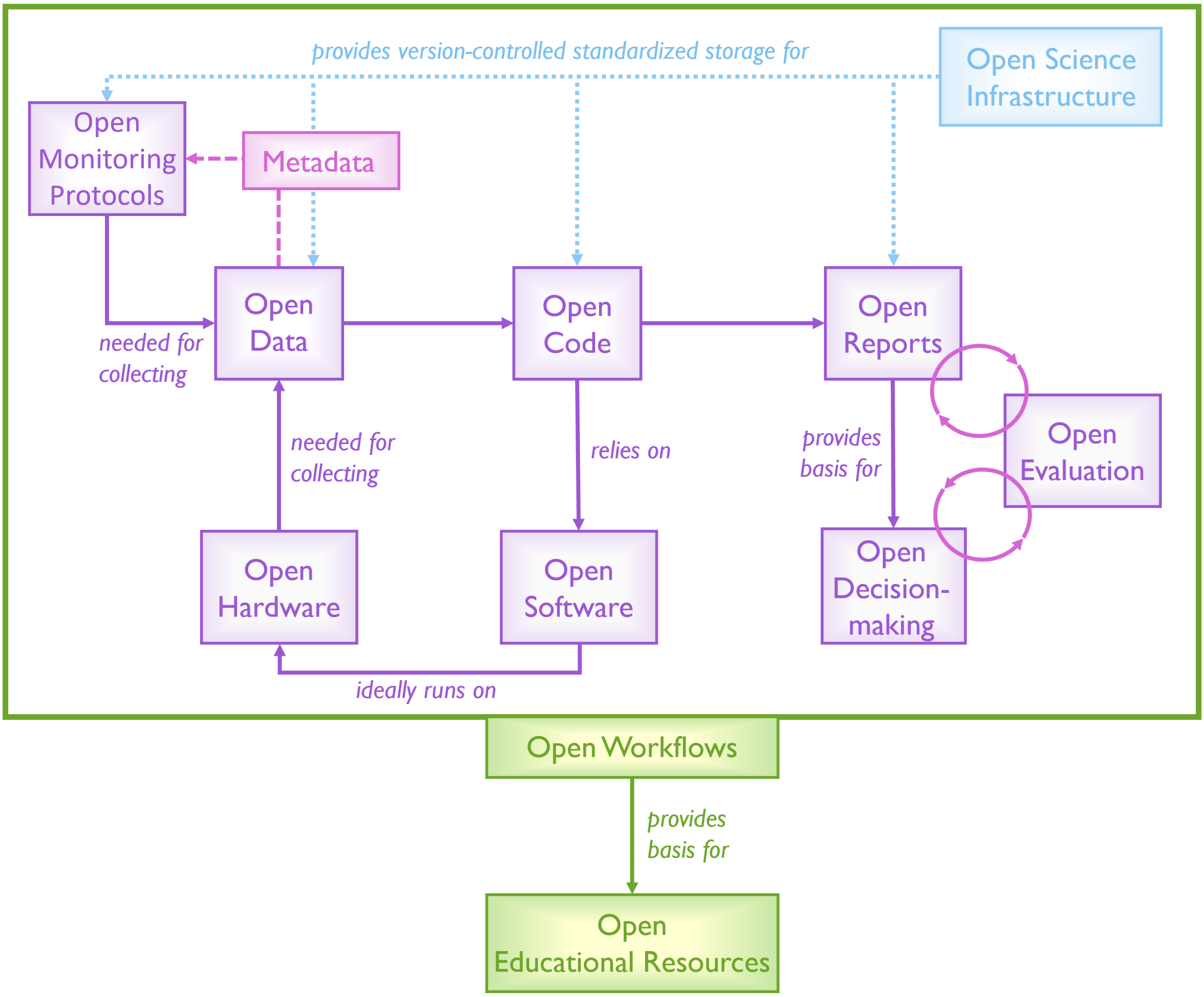
Notably, individuals may also belong to several of the categories introduced above. Wildlife managers, for example, may be actively doing research (or have done so in the past) and may also be collecting citizen science data in their free time. Indigenous people may hold central roles in conservation decision making. Some researchers may have indigenous backgrounds, etc. If anything, that only further enhances the diversity of knowledge surrounding wildlife research and management.

Engagement of diverse stakeholders

Harnessing and making optimal use of the diversity of available knowledge surrounding wildlife and wildlife management requires broad involvement of different stakeholder groups [e.g. **rubert2021?**]. Collaboration is key, and this necessitates that the entire process – from data collection to policy implementation – is inclusive. Four principles form the basis of inclusivity in this context: openness, accessbility, transparency, and reproducibility. These principles are also the corner stones of open science [**hampton2015?**,**powers2019?**,**vincente2020?**]. **Open** processes are clearly visible and offer opportunities for anyone to engage with them and contribute. **Accessible** processes are set up and documented in a way that allows everyone to obtain information relevant to them and be able to understand both the information itself and the way it was generated. **Transparent** processes feature complete documentation of every step of the workflow (this includes data collection, analysis, and presentation but also resulting decision-making) and provide all materials that informed and are produced by them. Finally, the mark of **reproducible** processes is that they are provided as complete workflows and based on tools that everyone can access, meaning that anyone (possessing relevant knowledge) can re-run the workflow and arrive at the same conclusions. Robust decision-making based on diverse knowledge thus requires workflows that are open, accessible, transparent, and reproducible and the next section provides and overview over the components of such workflows (Figure X).

Open and accessible workflows

When considering research workflows, we often only think of the steps connecting data and results. However, complete workflows in wildife research and management begin with planning and implementation of monitoring and end with reporting, decision making, and evaluation. Knowledge is built incrementally, and workflows should therefore also be thought of as cycles where newly gained information and results from evaluation feed back into earlier steps (e.g. sensu adaptive management, [**williams2011?**]; [**nichols2015?**]).



Schematic overview of the components of open and accessible workflows for wildlife research and decision-making.

?? gives an overview over the main components of open and accessible workflows.

It starts with data collection and hence with **monitoring protocols**, which should be openly accessible and documented in a way that they can be understood and – in theory – re-produced/re-implemented by independent parties.
The outcome of monitoring is recorded **raw data**, which should be made available publicly if possible and ideally adhere to FAIR (*F*indabilty, *A*ccessibility, *I*nteroperability, *R*eusability) and/or CARE (*C*ollective benefit, *A*uthority to control, *R*esponsibility, and *E*thics) data principles [wilkinson2016?,carroll2020?,carroll2021?]. Data needs to be supplemented with appropriate and standardized **metadata** to communicate both the structure and content of the data itself, and how it has been generated (i.e. how it is linked to the monitoring protocol).
The process to get from raw data to results typically consists of several steps, e.g. data cleaning, data wrangling/reformatting, data analysis, visualization of results, etc. These steps are implemented using some sort of **code** (manual steps, such as data editing/reformatting in Excel, are best avoided). All code should be well documented, reproducible, and version controlled [see cooper2017? for a guide] and made available via an appropriate repository such as GitHub, GitLab, etc. Formal code review, as is common in e.g. software development, is also a great tool to enhance reproducibility and overall quality of code [ivimey2023?].
Under ideal circumstances, both the **software** used for running code (analyses) and the **hardware** on which it is run should be open. For software, this means that the underlying code is open-source and that the program is – ideally – free to use. Defining openness for hardware is trickier, not least because the term spans a large variety of tech ranging from simple field loggers to sophisticated super-computers. These are typically not "publicly available", but what is crucial in the context of open and transparent workflows is that it is clearly stated what hardware was used and – if applicable – how one may get access to it.
The results from data analysis, their interpretation, and potentially recommendations resulting thereof are presented in **reports**. Whether these take the shape of institutional written reports, scientific articles, oral presentations for different interest groups or any combination of them: they hold important information that should be accessible to anyone interested.
While a majority of written reports nowadays are open access and dissemination presentations often are public events, the same cannot be said of most **decision making** processes. It is not rare for decisions in wildlife management to be made by a small number of individuals behind closed doors. That in itself does not need to be problematic, and may, in many cases, be the most efficient approach to reaching a conclusion and deciding on action. What is important for openness of and trust in wildlife management, however, is that there is an accessible record documenting the decision-making process. That way, anyone interested may gain insights into how and based on what evidence and factors decisions were made.
This is also crucial for the final step of the workflow: **evaluation**. Thoroughly evaluating the different steps of the workflow, from the design of the monitoring to the implementation of management actions and policy, is crucial for improving outcomes and maximizing value for money in the long run.
Finally, it is important to be aware that open scientific and management practices require appropriate **infrastructure**. This includes – but is not limited

to – databases and repositories for storing and sharing data, code, and documents. Relevant infrastructure may be institutional or global, as long as it is accessible, endorses standardized formats, and provides both permanent identifiers and version control.

Wildlife research and management need to be open and accessible to harness the diversity of knowledge available, which in term is crucial to tackle current and future challenges. Open and accessible workflows, from planning of monitoring to evaluation of decisions and actions (Figure X), are key to achieving that. Changing practices towards that goal will happen both through adaptation by current researchers and managers, and by teaching the next generation. Successes and failures in setting up and operating with open workflows provide examples that can be used in **education** and help equipping tomorrow's researchers and managers with the skills they need for successful wildlife management and conservation in a rapidly changing world.

Basics of Management/Decision Science

Value of Information

Evidence

PrOACT

Management Strategy Evaluation

Adaptive Resource Management

Causation and Inference

Asking the right questions in the right way

Estimation questions

Hypothesis driven research

Exploratory research

Causation and correlation

Sufficient causation

Necessary causation

Manipulative Experiments

Observational Studies

Directed Acyclic Graphs (DAGs)

Confounding variables

Mediator and moderating variables

Basics of Robust Experimental Design

Repetition

Replication

Randomization

Controls

Blocking

Response variables (i.e., performance measures in a decision context)

`{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)`

Introduction

The ecological question often revolves around measuring relevant biological variables (survival, density, etc.) on items (like populations, species, subspecies, etc.) of interest in the system under study. In these questions, researchers conduct studies to obtain accurate estimates of characteristics that are important for management and conservation decisions. A vast amount of literature on the subject of **sampling or finite population sampling** (Cochran 1963) exists to help researchers in planning these studies.

The set of all possible individuals from which processes and patterns are to be deduced is called the **target population**. Rarely in any ecological studies, a census of the population is possible. Even in the case of rare and critically endangered species (less in number) or the case of immobile organisms like plants, collecting data from the whole population can be impossible. Thus, in most cases, researchers study a **subset of the population (sample)** and use collected information to draw inferences about the target population. The sample therefore can be described as a group of individuals who participate in research and represent the whole population. In more scientific terms, a sample is a subset of a population randomly selected based on some probabilistic design. In ecological studies, we often hear the term **sampling frame (or study area)**. To efficiently design these studies, it is necessary to understand what this term means. It is a finite set of all individuals that could be measured, and we can use different sampling schemes to obtain items from this frame. The sampling frame usually coincides with the target population, but reasons like accessibility, logistics, budget, etc. can make it differ otherwise. From this sampling frame, we draw a subset or sample of individuals, the **sampling units**, and the items to be measured for different biological variables. Sampling units should be distinct and easy to define (Box 1, Fig. 1). Target population does not always mean the number of animals. It can also be described in terms of geographic area, in which case, the sampling unit would be grids or township or county depending on the research question.

Sampling is a critical part of both descriptive and experimental studies. All field studies require appropriate sampling designs to reduce variation among observations in the study. The choice of sampling method will depend on the objectives of the study, the distribution, and characteristics of the population being sampled among many other factors.

Box 1

Definitions

Population (target population): Collection of all individuals of one species inhabiting an area at a given time, about which some parameters of interest are to be estimated. In wildlife studies, this could be all the individuals of a species or subspecies in a habitat.

Sampling frame: The list of the members of individuals that we randomly select from a target population.

Sample unit: A unique collection of elements (e.g., plots or organisms) on which sample data are collected.

Sample size: The number of samples collected to answer measurements regarding the population of interest.

The target population of a wildlife study could include a broad array of entities and it is important to be specific in defining and identifying it long before the study beings.

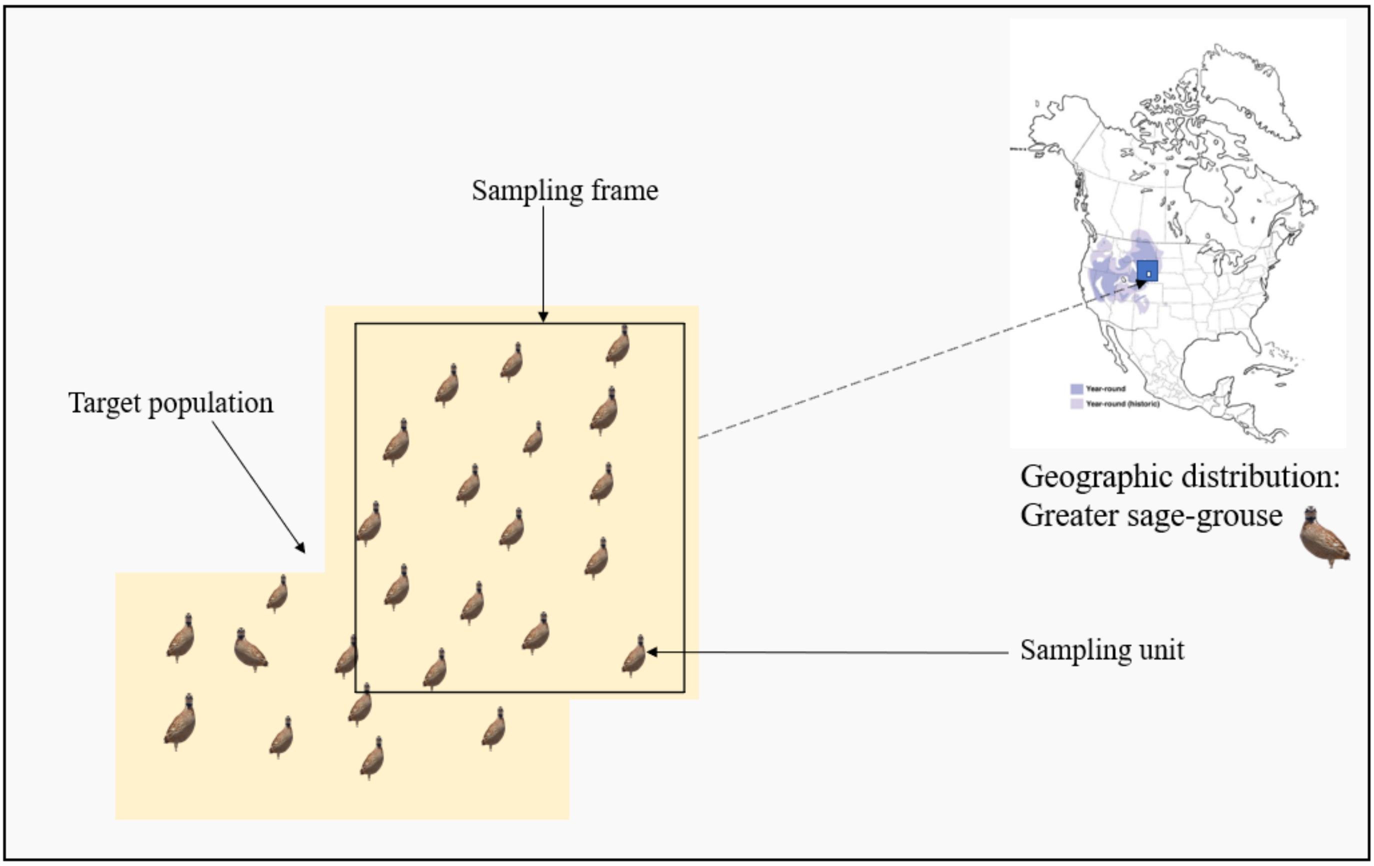


Fig. 1 A target population of greater sage-grouse (*Centrocercus urophasianus*, hereafter grouse) in one county in Wyoming, USA. For reasons like accessibility and permission, the whole county area could not be used for conducting a study. So, the sampling frame represents the area from which sampling units could be collected.

Probability and non-probability sampling methods

For a sample to correctly represent a population, we need to properly identify the target population, followed by identifying the sampling frame, sampling units, and sampling technique. We should also be cautious about resource availability-manpower, logistics, time, etc. So, the next question here is how to draw a sample. There are mainly two types of sampling methods: probability and non-probability sampling. These designs differ in terms of the quality of parameter estimates (Box 2).

Probability sampling: In the probability sampling scheme, every unit from the sampling frame has a non-zero probability of selection. It, therefore, leads to unbiased estimates of the mean and variance for the variable of interest. Therefore, any method aimed at generalizing results drawn by a sample to the whole population of interest must be based on probability sampling.

Non-probability sampling: In non-probability sampling, researchers select samples based on their convenience. In other words, the researchers purposively choose particular units for constituting a sample. Social researchers often use this sampling design to select households or families to conduct their surveys. A few common techniques are **convenience sampling** and **judgment sampling**. In convenience sampling, samples are chosen based on an arbitrary selection procedure. It is often justified based on accessibility and availability of resources like time, budget, etc. In wildlife studies, this sampling scheme is often used for conducting roadside bird surveys, surveys for identifying mammal tracks near roads, etc. Since the location of the target species decides the sampling frame and number of samples, the results from these studies are often highly biased and far from accurate. In judgment sampling, sampling frame, and samples are chosen based on expert knowledge of the system. One common example from wildlife studies is selecting and classifying **study area (sampling frame)** into low-quality and high-quality based on expert knowledge about the area.

There are pros and cons for each of these sampling designs. Probability sampling helps in reducing sample bias and therefore provides an accurate representation of the population. Non-probability sampling is useful when we still need some preliminary data within time and budget constraints.

Box 2

Quality of a parameter estimates can be assessed by its **accuracy** (Fig 2). Accuracy is further defined by how precise and unbiased the data is and refers to the small size of deviations of the estimator from the true population value. **Precision** depicts variation in population and size of the sample (Cochran 1963, Krebs 1999, Zar 1999). Indicators of the precision of an estimator are **standard errors** and **confidence intervals**. Another measure of quality is **bias** which describes how far the average value of estimator is from the true population value. An unbiased estimator centers on the true value for the population.

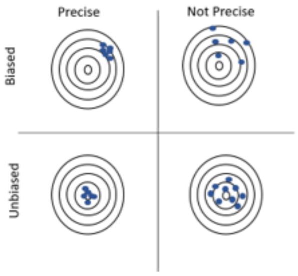


Fig. 2 Accuracy in target shooting in terms of bias and precision.

Let's take an example to illustrate this point. Suppose we were interested in estimating the density of the raptor population in an area. There could be different ways of carrying out this study. One approach might be to divide the study area into a number of grids of equal size and randomly draw a sample of grids to conduct a point-count survey for raptors (Fig. 3a). Within each sampled unit, there will be 3 point-count locations separated by enough distance so that there is no overlap in raptor population between these points. Each of these locations would be surveyed for 15 minutes, in an attempt to count all visible raptors, present there. Each of these surveys will be repeated to understand any temporal variation in the population. To obtain a density estimate, we would divide the total number of raptors in a grid by the size of that grid. As we can see from the graph (Fig. 3b), there exists little variation from one unit to the next. The density estimates of the whole study area would be then the mean value from this sample.

box2a

Box 2 continued.

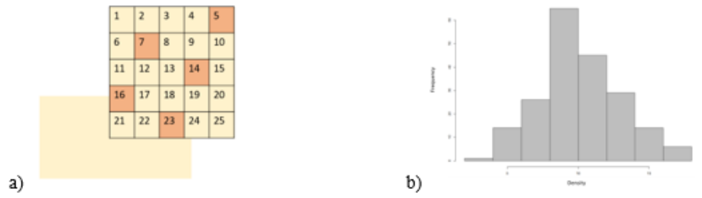


Fig. 3 Hypothetical example of conducting a study based on a a) probabilistic design for estimating raptor counts in a study area and b) density estimates from the study.

Another approach would be conducting a roadside raptor survey. Under this method, we will select roads (Fig. 4a) on the basis of convenience, and accessibility, and will count raptors while driving along these roads at <30km/hr. At the end of each survey, we will count the total number of raptors observed. To obtain a density estimate, the total number of raptors obtained after all the surveys were done will be divided by the total area.

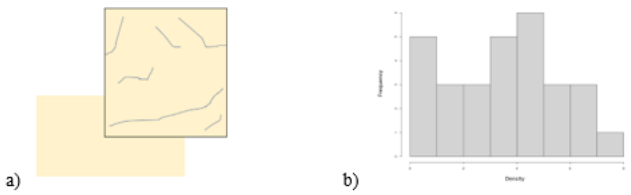


Fig. 4 Hypothetical example of using a) convenience sampling for conducting a study for estimating raptor counts in a study area and b) density estimates from the study.

We see a high variation in results from the roadside raptor survey (Fig 4b). The mean estimate is less precise and not as reliable as the estimate from the earlier study based on a probabilistic design. There was no account for spatial and temporal attributes and therefore, the result is highly biased. We will therefore use estimates from the former study as density estimates for our study area. We strive for accuracy in our estimates by choosing the approach with the least bias and most precision, by applying a robust sampling design, and by obtaining a sufficiently large sample size to provide precise estimates.

box2acont

Sampling designs under the probability sampling method

Although simple random sampling is the most basic technique for sample selection, there are others that are often used in wildlife ecology studies. There are pros and cons associated with each of these designs and they can also be combined to provide a larger set of options for study designs.

Simple random: Simple random sampling is the process of selecting n units from a population of N units such that every unit has an equal probability of inclusion in the sample. Simple random sampling requires that each sample unit be selected independently of all other units. This method should be used if the area of interest is homogeneous with respect to the elements and characteristics of interest. A simple random sample may be obtained by following basic steps: 1) the study area must be completely covered by non-overlapping sampling units, 2) the population of sampling units is assumed to be finite, 3) sampling units can be located and the measurement of the characteristic of interest on the unit is possible. Also, the error in measuring should be small compared to the differences in the attribute from unit to unit, and 4) sample units are sampled without replacement. For example, suppose a farmer is interested in evaluating the health of his cattle, and each of them is ear tagged. So, under simple random sampling, he can generate some random number using a calculator or excel sheet and select those tagged cattle for health assessment.

Systematic: Systematic sampling is the process where sampling units are selected at regular intervals. Under this sampling technique, the sampling frame will be partitioned into n number of primary units and then the selection of units will occur in a systematic fashion based on a random start. This design is easier to execute than simple random sampling. In the above example of assessing cattle health, systematic sampling will be easier to implement in case the cattle are not ear-tagged. Then under systematic design, the farmer can choose every nth cow while they are entering the barn (assuming they follow a queue). Systematic sampling is commonly used to sample vegetation characteristics. For example, determining vegetation characteristics every 10 meters along a line transect in a plot is a classical example of systematic sampling. Systematic sampling has also been criticized in cases when the arrangement of units may follow some pattern in the response variable. For example, let’s say we were interested in the number of people using public areas for birding. We decided to establish a check station and take a count using a systematic sample of days during the study period. This could give us a biased result if every sampled day fell on a work week, then the estimates obtained would be very different from estimates obtained from the weekend count.

Stratified: In wildlife studies, populations tend to be aggregated or clustered, thus sample units closer to each other will be more likely to be similar. For this reason, systematic sampling tends to overestimate the variance of parameter estimates. A uniform grid of points or parallel lines may not encounter rare units. To increase the likelihood of capturing some of these rare units, scientists may stratify the sample such that all units of each distinct type are joined together into strata and simple random samples are drawn from each stratum. Stratified sampling is, therefore, generally used when the population from which the sample is to be drawn does not belong to one homogeneous group and there is a high variation within the population. If, however, the population belongs to a heterogeneous group, the estimates based on earlier sampling designs will be imprecise. If we have prior information associated with the heterogeneity in the population, we can use designs like stratified sampling to select samples which will increase estimates precision. This sampling technique divides the whole population into different mutually exclusive groups (strata) according to some characteristics such as the habitat they inhabit, gender, etc. Ideally, the strata should be homogeneous with respect to the variable of interest (like density, abundance, etc.). This process requires more effort than random sampling but is generally more accurate in terms of representing the population. There is also a limit to the number of strata into which a population can be subdivided. The stratified sampling method is common in wildlife studies, as it helps estimate and contrast parameters among strata. The formal procedure of stratified sampling follows a few steps: 1) specify strata, which must be mutually exclusive, 2) classify all sampling units into their stratum, and 3) draw a simple random sample from each stratum.

Cluster: A probabilistic sampling scheme in which each sampling unit is a cluster of items such as the group of animals. Cluster sampling is generally used in cases when there are predefined groups within the population. These groups can be based on demographics, habitats, geography, etc. The sampling process starts by dividing the population into small groups known as clusters followed by random selection of these clusters to create a sample. This approach has wide applications in wildlife study as many birds and mammals occur in groups during or all parts of the year. Cluster sampling is useful when the cost or time to travel from one sample unit to the next is too high. Cluster sampling can also be performed in stages. Single-stage cluster sampling happens when all the elements of the chosen clusters are included in the sample. Two-stage cluster sampling is when in contrast to single-stage cluster sampling only some units are observed. The formal procedure of cluster sampling follows a few steps: 1) specify appropriate clusters and make a list of all clusters, 2) draw a simple random sample of clusters, and 3) measure all elements of interest in each selected cluster.

Adaptive sampling: In various studies, numerous sampling designs are combined under an adaptive sampling framework. In this technique, we start with an initial probabilistic sample of units and add more units in some pre-defined neighborhood or pre-defined condition to this sample (Thompson and Seber 1996, Williams et al. 2002, Thompson 2003). This process continues until no sampled units satisfy the specified condition. Adaptive sampling offers biologists a way to augment the probability sample with samples from other units without losing the benefits of the original probabilistic design. Rules for the selection of additional samples are established based on some characteristic of the variable of interest (e.g., presence/ absence, age, sex, and height).

Case study: Scientists are worried that ongoing human-induced landscape changes have threatened a grouse population in eastern Wyoming, USA. They decided to conduct a study with two main objectives. They primarily want to understand the impact of this dynamic landscape on the survival of grouse and on lek numbers. The study will be done in Carbon County for a period of 3 years during which they plan to capture 150 grouse in total (both male and female; Fig 5a). For carrying out this study, they divided the study area into 25 equal-sized grids (Fig. 5b). To estimate the survival of grouse, we need to capture individuals and track them for the required time period or till they are alive, whichever comes first. We need to devise a sampling mechanism to select grids from where these grouse can be captured.

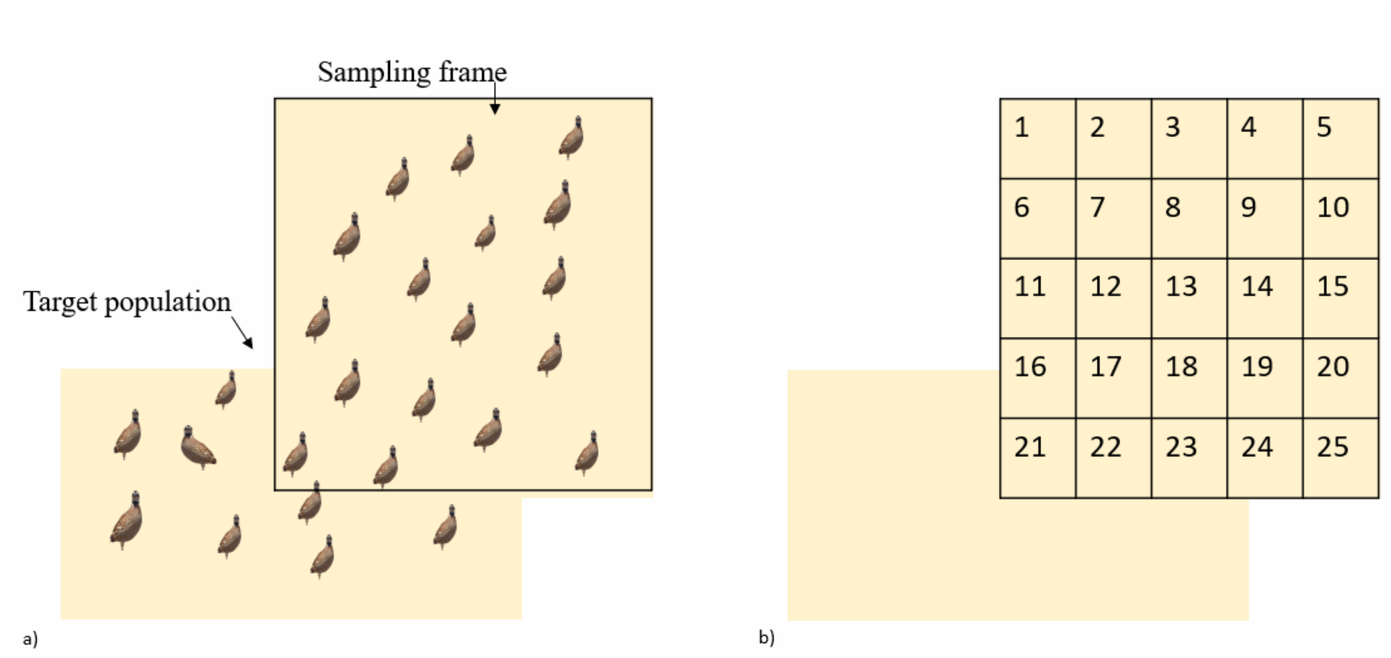


Fig5ab

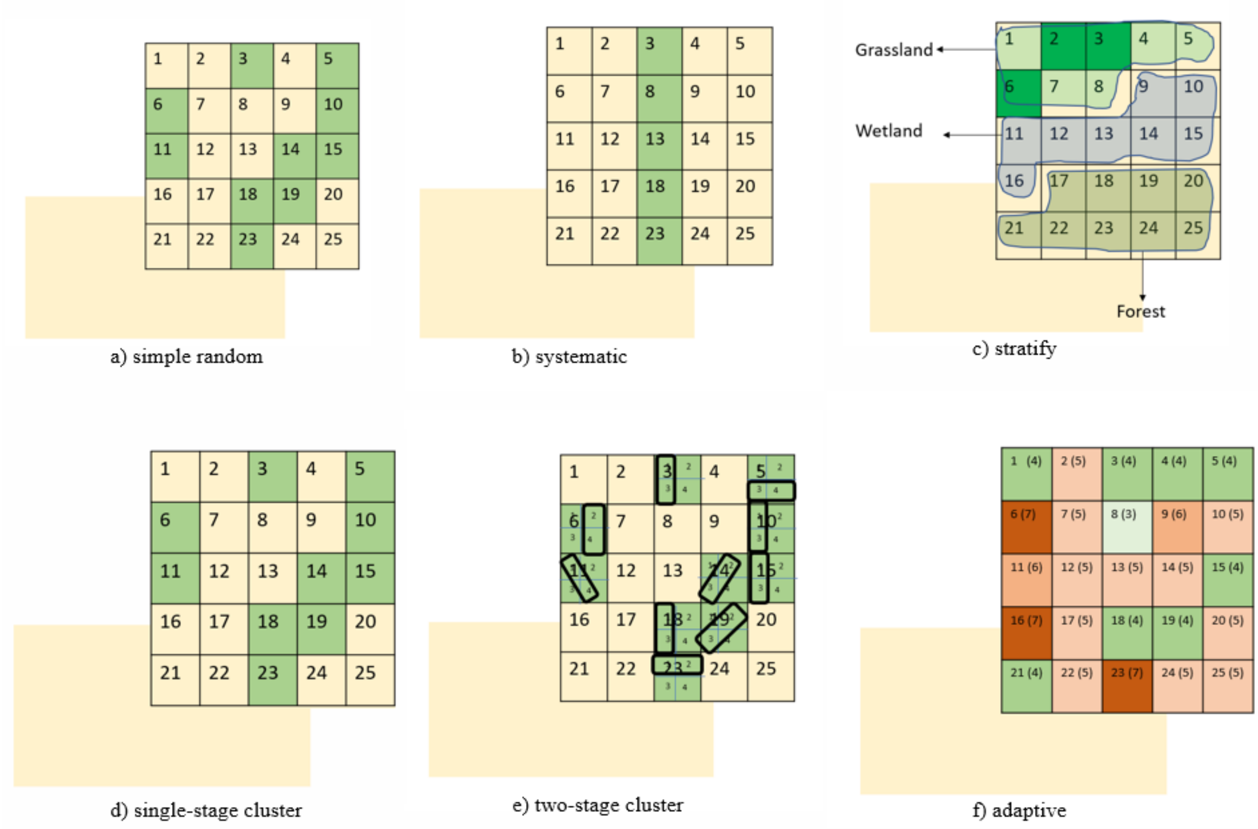
Fig. 5 a) Carbon County in Wyoming, USA for conducting a 3-year study on a grouse population in the sampling frame, b) Study area divided into 25 equal-sized grids for carrying out the study.

In this case study, we will show you how researchers can sample grids using different types of probabilistic sampling schemes. Under random sampling, each of the grids will have an equal probability of getting sampled (Fig. 6a). In a systematic sampling framework, researchers can pick grids at regular intervals. So let's say, they decided to select every 5th grid starting with the 3rd grid. So, 3, 8, 13, 18, and 23 will be their plots from which the grouse will be then captured (Fig. 6b).

Suppose researchers identified three different dominant land cover types (grassland, wetland, and forest), so the random selection of grids can follow a stratified framework. Under this framework, researchers will divide the whole study area into these different strata and then randomly select grids from each of these strata to capture grouse (Fig. 6c).

Lek sites are known to be spatially close to each other. So, we would expect that if a lek is inhabiting one primary sample plot (grid), there are other leks in the neighborhood. So here researchers can use cluster sampling to calculate lek numbers. Under single-stage cluster sampling, they would start by randomly selecting primary sample plots across the study area, then within those primary plots, conduct surveys to calculate all lek numbers within the cluster of all four secondary plots (Fig. 6d). In our example, grids can act as primary plots. In each primary sample plot, there are 4 secondary sample plots (numbered 1 to 4 in some order), so under two-stage cluster sampling, we will randomly select any 2 from these 4 secondary sample plots (Fig. 6e). We will then conduct surveys to calculate lek numbers in 2 of these secondary sample plots.

Under adaptive cluster sampling, we will start with the grids which have the highest number of leks and then will sample the next grids with a similar number of grouse (Fig. 6f). See Box 3 for R codes for each of these sampling designs.



all

Fig. 6 Examples of sampling design a) simple random, b) systematic, c) stratify, d) single-stage, e) two-stage, and f) adaptive (numbers in brackets shows the number of leks in each grid) for selecting grids from an equal-sized gridded system in Carbon County, USA for studying grouse population.

Box 3

```
#to make it reproducible
set.seed(123)

#input all 48 grids
grids<-c(1:25)

#select any 10 grids
sample(x=grids, size=10)
```

a) simple

```
#make it reproducible
set.seed(123)

#create strata
df<- data.frame(id=c(1:25),
              habitat=rep(c("grass","wetland", "forest"),c(8,8,9)))
#select grass strata
grass<-subset(df,habitat=="grass")

#randomly sample from grass strata
grass.sample<-grass[sample(nrow(grass),3),]
```

b) systematic

```
#load required libraries
require(ggplot2)
require(plyr)

#make it reproducible
set.seed(123)

# create reasonable size grid
df<-expand.grid(Lat=seq(5000,9000,1000),
              Long=seq(3000,7000,1000),
              num=0)

# fill grid with known values
df$num=c(4,5,7,5,5,7,5,4,4,5,6,5,5,5,4,7,5,3,6,5,4,5,4,4,4)
df$id=1:nrow(df)

# graph color and text controls
showText<-TRUE
txtSize<-3
heatmap.low<-"lightgreen"
heatmap.high<-"orangered"

#create a heat map
ggplot(df) +
  geom_raster(aes(Lat,Long,fill=num)) + theme_bw() +
  scale_fill_gradient(low=heatmap.low, high=heatmap.high) +
  geom_text(aes(Lat,Long,label=num,fontface="bold"),
        color="black",
        size=5.5)
```

```
#to make it reproducible
set.seed(123)

#input all 48 grids
grids<-c(1:25)

#define a function for systematic sampling
getgrid<-function(N,n){
  k<-ceiling(N/n)
  r<-sample(1:k,1)
  seq(r,r+k*(n-1),k)}

grid_systematic<-as.data.frame(grids)
#call function to draw samples
systematic<-grid_sys[getgrid(nrow(grid_systematic),5),]
```

c) stratify

```
#make it reproducible
set.seed(123)

#Single-stage cluster sampling
#create grids and sample 10 grids
grids<-c(1:25)
primarySample<-sample(grids,10, replace=FALSE)

#Two-stage cluster sampling
#arrange sample in an order and assign id to each sample
primary<-sort(primarySample)
id<-c(1:length(primary))
df<-rbind(id,primary)

# randomly select 2 sub-samples from 4 secondary units
sapply(df[1,], function(g) sample(1:4, 2, replace=FALSE))
```

d) cluster

e) adaptive

codes1

Sampling methodology

In wildlife studies, there are a few commonly used sampling methodologies.

Plots: Plots are widely used to study habitat characteristics, vegetation characteristics, counting animal numbers, etc. Plots' shape can vary from circular to square and represents a geographically defined target population. Wildlife tends to be distributed nonrandomly across the landscape in correspondence to the distribution of their habitat. Their distributions are further impacted by intraspecific and interspecific interactions. Given that distributions and abundance vary, plots should vary in shape and size depending on the studied species. Numerous factors influence plot size, including the biology of the species, their spatial distribution, study objectives, logistical considerations, and cost constraints. For example, larger species with large home ranges require larger plots to include adequate numbers. A 3,500-ha plot might include only 10% of the home range of a grizzly bear (*Ursus arctos horribilis*), the same area could include the entire home ranges of multiple white-footed mice (*Peromyscus leucopus*). Krebs (1999) listed three main approaches to determining optimal plot shape and size for a study: 1) plot size should have the highest precision for a specific study area, 3) plot size which is most accurate and efficient to answer the question of interest, and 3) plot size which is logistically easy to construct and use.

Points: In point sampling, a set of points is established throughout the population, and measurements are taken from each point. A common example is a point-count survey for birds where the distance to each heard or seen bird species of interest is measured from a particular point. Selection of sample points can follow any sampling design, as long as points are spaced apart enough that overlapping of the population between points is a bare minimum.

Transects: In line transects a line or series of lines is randomly or systematically located in the study area. Objects are recorded on either side of the line according to some rule of inclusion. The observer traverses each line, recording the perpendicular distance from the line to each detected animal. These distances are used to estimate the effective width of the area sampled by the transect. Transects can be established using any sampling design as long as each of them is treated as an independent observation and are non-overlapping.

References

Cochran, W. G. 1963. Sampling Techniques, 2nd Ed., New York: John Wiley and Sons, Inc.

Krebs, C. J. 1999. Ecological methodology. Second edition. Benjamin Cummins, Menlo Park, California, USA.

Thompson, S. K. 2003. Editorial: Special issue on adaptive sampling. Environmental and Ecological Statistics 10:507

Thompson, S. K., and G. A. F. Seber. 1994. Detectability in conventional and adaptive sampling. Biometrics 50:713-724

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic, San Diego, California, USA

Zar, J. H. 1999. Biostatistical analysis. 4th edition. Prentice/ Hall, Upper Saddle River, New Jersey.

The Publication and Peer Review Process

Determining coauthorship

Choosing the best outlet

Who should review your manuscript?

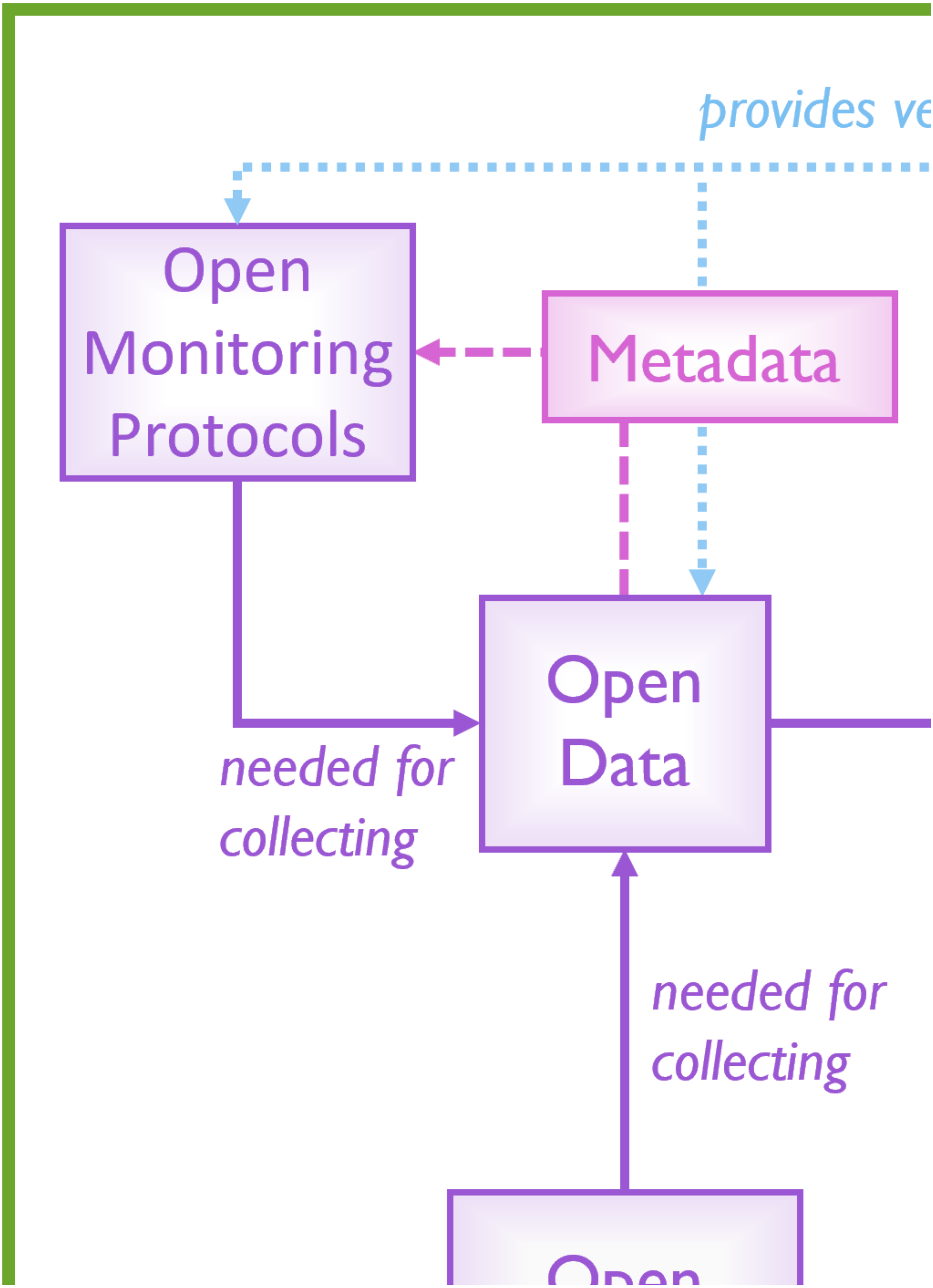
How to constructively provide feedback as a reviewer?

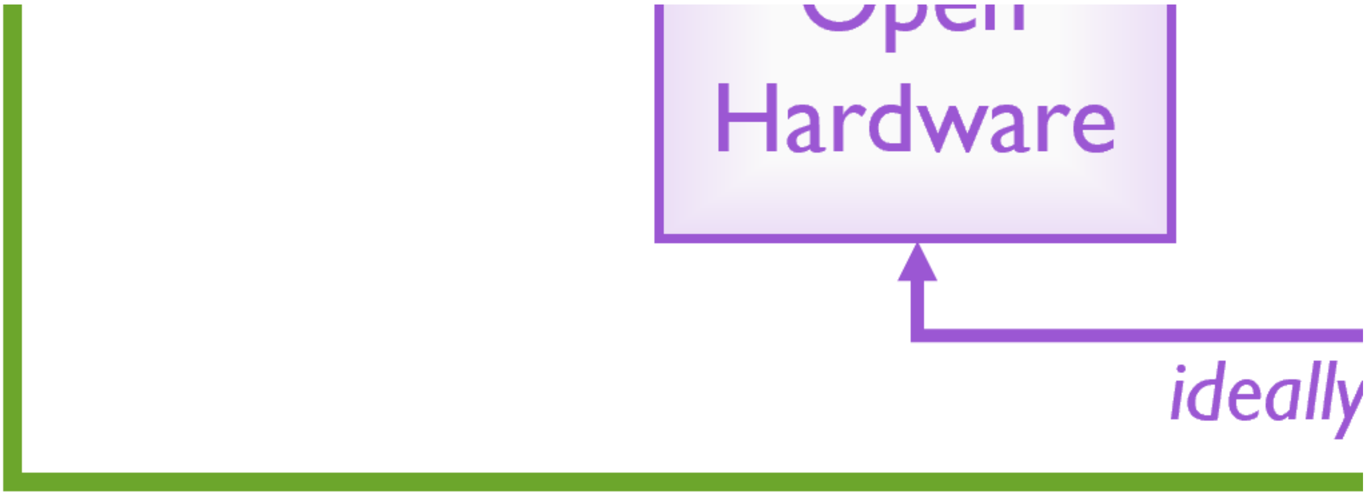
Responding to reviews

Marketing your manuscript after acceptance

Figures

![


provides ve
Open Monitoring Protocols
Metadata
Open Data
needed for collecting
needed for collecting
Open



Tables

References