Evaluating and validating abundance monitoring methods in the absence of populations of known size: review and application to a passive tracking index

Lee R. Allen
Robert Wicks Pest Animal Research Centre, Lee.Allen@daff.qld.gov.au

Richard M. Engeman
National Wildlife Research Centre, richard.m.engeman@aphis.usda.gov

Follow this and additional works at: http://digitalcommons.unl.edu/icwdm_usdanwrc

Part of the Life Sciences Commons

http://digitalcommons.unl.edu/icwdm_usdanwrc/1628

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Animal and Plant Health Inspection Service at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in USDA National Wildlife Research Center - Staff Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Evaluating and validating abundance monitoring methods in the absence of populations of known size: review and application to a passive tracking index

Lee R. Allen & Richard M. Engeman

Abstract Rarely is it possible to obtain absolute numbers in free-ranging populations and although various direct and indirect methods are used to estimate abundance, few are validated against populations of known size. In this paper, we apply grounding, calibration and verification methods, used to validate mathematical models, to methods of estimating relative abundance. To illustrate how this might be done, we consider and evaluate the widely applied passive tracking index (PTI) methodology. Using published data, we examine the rationality of PTI methodology, how conceptually animal activity and abundance are related and how alternative methods are subject to similar biases or produce similar abundance estimates and trends. We then attune the method against populations representing a range of densities likely to be encountered in the field. Finally, we compare PTI trends against a prediction that adjacent populations of the same species will have similar abundance values and trends in activity. We show that while PTI abundance estimates are subject to environmental and behavioural stochasticity peculiar to each species, the PTI method and associated variance estimate showed high probability of detection, high precision of abundance values and, generally, low variability between surveys, and suggest that the PTI method applied using this procedure and for these species provides a sensitive and credible index of abundance. This same or similar validation approach can and should be applied to alternative relative abundance methods in order to demonstrate their credibility and justify their use.

Keywords Activity index · Animal surveys · Model sensitivity · Predators · Trends · Validation

Introduction

Free ranging mammals are at best difficult to survey, and some, such as carnivores, especially try to avoid detection, making them particularly challenging to monitor. Some are also in relatively low numbers in the landscape and/or they occupy terrain and vegetation that makes them difficult to observe (Engeman and Witmer 2000; Witmer 2005; Engeman 2005). Various methods of surveying presence/absence, abundance or density are used to assess free-ranging species (see review in Lancia et al. 1994) with the method chosen being a function of the objectives of the study, the resources available and the size, abundance and behaviour of the target species (Pollock 1995; Engeman 2005; Engeman et al. 2013). Where direct observations or density estimates are not required or possible, indirect counts or indices of relative abundance are employed from which population trends may be monitored (Caughley 1977; Lancia et al. 1994; Pollock 1995; Evangelista et al. 2009). The literature contains numerous examples of where indices are applied to monitor trends in abundance (for examples, see Beier and Cunningham 1996; Blaum et al. 2008; Edwards et al. 2002; Engeman et al. 2000), their response to management actions (Engeman et al. 2000; Eldridge et al. 2000) and sometimes their relationship to other species (for examples, see Catling and Burt 1995; Letnic et al. 2009). The greatest value of indirect abundance measures is in monitoring population trends at the same location or from different, but similar locations at the same time (Lancia et al. 1994; Pollock 1995; Engeman 2005). Knowing or estimating
the exact population size or having a relative abundance estimate is maximally informative when taken in context with comparative values to assess trends, changes or differences (Engeman 2005; Krebs 2008). Caution and, at times, contempt are directed at indirect measures of abundance because they are inappropriately used (for examples, see Allen et al. 2011) and not validated against known populations (Anderson 2001); yet, even direct counts and density estimates are not validated this way because of the difficulty of comparing the method-derived estimate to a known population size (Clark et al. 2003; Engeman 2003; Pollock 1995; but see Curtis et al. 2009). The primary driving force for applying indexing procedures is that they are practical or feasible methods for tracking animal populations. Therefore, validation approaches to indexing procedures is of inherent importance for research and management purposes.

Models to accommodate incomplete counts of true abundance are available but have strict assumptions that are often violated (Anderson 2001; Engeman 2005). Issues of immigration, emigration, births, deaths, various forms of heterogeneity in catchability, loss of identification marks and detection frustrate the valid estimation of absolute numbers of animals in their natural environment within a defined area, with the modern calculation methods to overcome these obstacles often requiring more data than being feasible or possible to obtain (Caughley 1977).

Relative abundance methods have two fundamental assumptions—the population index is at least correlated to true population size (Engeman 2005) and preferably directly proportional to population size with proportionality (relatively) constant (Caughley 1977; Pollock 1995; Witmer 2005). Validating a relative abundance method against these assumptions is impractical because it also requires a known population size (Pollock 1995). Therefore, strict protocols for data collection and standardised approaches for assessing abundance are recommended to minimise violation of the assumptions (see Caughley 1977; Pollock 1995; Engeman 2005; Krebs 2008), although adherence to these principles does not necessarily guarantee credibility of subsequent results. So, here, we address the question “How can indirect abundance methods be evaluated and validated in the absence of a population of known size?”

**Evaluation criteria and framework**

Relative abundance methods are often compared and evaluated in terms of sensitivity, although this term is variously used to describe different attributes of the method. Following the definition of Henke and Knowlton (1995), sensitivity in this paper incorporates three key attributes: the ability of the method to detect the target species at low and high densities (i.e. probability of detection, Gompper 2002), the repeatability or precision of relative abundance values generated by the method (Caughley and Sinclair 1994; Engeman 2005) and the sampling effort required to detect changes in abundance (Catling and Burt 1995). Ideally, a relative abundance method should reliably detect species at low levels of abundance, produce values with negligible variance and be capable of measuring changes in abundance with minimal sampling effort and expertise.

The construction of models to replicate natural and complex systems is a practice widely accepted and reported (Knepell and Arangno 1993; Kleijnen 1995). Validation establishes an initial degree for and/or elevates a model’s credibility (Janová 2012; Law and Kelton 2000) but does not certify a model as being valid or invalid (Carley 1996; Muller and Muller 2003). Models may be used to forecast outcomes of proposed management actions and gain better understanding of how complex systems work and react. Likewise, relative abundance methods are also used to establish qualitative or quantitative insights about species and management effects (Harris and Rayner 1986; Macdonald and Rushton 2003). Broadly, there are three steps in model validation that are applicable to validating relative abundance methods: grounding, calibration and verification.

**Grounding, calibration and verification**

In modelling terms, grounding is an element of model validation that establishes the rationality of the methodology and its potential for predictive accuracy. Grounding may examine the fundamental assumptions, whether they capture the key elements of a particular group or process, whether the model produces expected behaviours consistent with known fact and whether other researchers have made similar assumptions. Calibrating is the process of adjusting a model to fit real-world data. One or more datasets containing known and detailed information are required. Ideally, these should span the range of parameters the model represents. Calibrating relative abundance methods to known populations has been precluded by Engeman (2005), firstly, because of the impracticality of testing the method on a known population (see above); secondly, because the result is only relevant to that specific location and time, and thirdly, the method becomes a population density estimator fraught with the same degree of reliance on analytical assumptions about the relationship between the index and population as has been the bane of the well-known density estimators. The next best alternative to using known populations is to test the method on free-ranging species that represent the range of densities likely to be encountered, ordinal calibration (for example, low, average and high). This evaluation will indicate whether the method is capable of detecting species at low densities and if it can differentiate between density levels. To verify a model, the model’s predictions are compared graphically or statistically.
with real data (Kleijnen 1995) to demonstrate that the model produces the pattern, value or distribution found in real systems. While grounding focuses on the internal assumptions and application of relative abundance methods, verification focuses on the outcomes when using the relative abundance methodology. For example, does the method detect and respond to changes in seasonal conditions, breeding cycles or management actions in a manner consistent with expectations?

Application to a passive tracking index

To illustrate how the validation concepts might be applied in practice, we consider the widely applied passive tracking index (PTI) method (e.g. Engeman and Allen 2000). Briefly, the PTI is one of many indexing methods encompassed by the general indexing paradigm described by Engeman (2005). The tracking plot procedures can take a variety of forms (e.g. multiple 1-m swaths across travel ways or fewer than 1.6 or even 5-km-road-width sections of unformed road). The specific data we consider here were originally described by Allen et al. (1996), and used passive tracking stations (or sand plots) placed at 1-km intervals along unformed dirt roads or sandy creek beds to detect the movements of multiple ground-dwelling and ground-foraging species, identified from their spoor and counted at 24-h intervals for several consecutive days. The index is calculated as the mean of daily means for the number of track intrusions by each species per tracking station per day (e.g. Engeman and Allen 2000).

Grounding: rationality of the PTI approach

We examine the PTI methodology and how the data collected (tracking rate activity) relates to actual abundance, whether the method produces trends and patterns consistent with established knowledge for that species and whether alternative abundance estimation methods have similar or identical assumptions.

Most methods either directly or inadvertently measure a combination of animal abundance and their activity. For example, observations might be detected as spoor on tracking stations or captured as an image (i.e. camera trap) by counting animals as they move about (i.e. spotlight or direct counts) or by capturing them as they move about (i.e. pitfall traps and catch-per-unit-effort approaches). Although measured differently, they are all a function of, and susceptible to, changes in animal activity independent of absolute abundance. The PTI method is passive and does not require any response other than the target animal’s normal daily movements. Thus, PTI calculations are not affected by bait or trap shyness (or prereness) or presence of the observer, although there can be some (individual canid) exceptions (e.g. alpha coyotes avoiding camera traps, Sequin et al. 2003).

Population size is demonstrated to have the largest influence on animal activity detected on tracking plots, although behavioural changes (associated with annual cycles of reproduction, dispersal, etc.), meteorological activity (periods of warm, cool, windy or wet weather proceeding or following fronts) and ephemeral events (like a thunderstorm or passing clouds) can have lesser yet measurable impacts on animal activity (Bider 1968). This environmental and behavioural stochasticity or process error (McCallum 2000) contributes to variability in population estimates for all abundance methods, including density estimation methods. Observational error, inaccuracies in identifying spoor or estimating the abundance value (McCallum 2000), may also add to estimate variability in PTI calculations. By surveying activity over several consecutive days, the PTI methodology smooths the effects of ephemeral and meteorological events that may last hours or days. Therefore, in terms of its approach, the PTI method monitors a key attribute of abundance (animal activity), and although this attribute may be susceptible to influences other than change in absolute numbers by counting several consecutive days and by measuring the precision of PTI values (Engeman 2005), the method compensates and accounts for this variability. Conceptually, the PTI is not dissimilar to other relative abundance estimation methods.

Greater confidence in the reasonableness of a method can be achieved through independent comparisons with alternative techniques to ensure that performance is in line with expectations. For our PTI example, a variety of studies have simultaneously compared spoor detected on tracking stations to alternative methods of estimating abundance. Reviewing a subset of these, we have summarized in Table 1 various independent Australian studies where the PTI method (or similar indexing approaches using tracking stations) has been compared with alternative abundance estimation methods on terrestrial mammal species. In each case, as summarized in Table 1, the PTI results were in line and compared favourably with the other methods, although we note that the PTI demonstrated superior sensitivity for detecting animals and illustrating population changes.

Calibration: using artificial manipulation of population levels and comparison with alternative techniques

PTI values were calculated from 50 tracking stations in two replicate areas separated by a buffer (>10 km) from the nearest tracking station for (up to) 17 consecutive days. After calculating the initial PTI, the dingo/wild dog (Canis lupus dingo and hybrids of Canis lupus familiaris) population was exposed to 50 sodium fluoroacetate (1080) poison baits for 24 h and, subsequent to a recalculation of the PTI, 100 poison baits for 48 h. By calculating the initial PTI and sequentially
reducing the population and repeating the PTI calculations, this experiment evaluated the PTI’s ability to detect wild dogs at high and low densities and to detect incremental reductions in the population. Concurrently, two alternative relative abundance methods—the scent station visitation index (described in Roughton and Sweeny 1982 for use with coyotes Canis latrans) and buried meat index (used on wild dogs and described by Fleming 1996)—provided a comparison between alternative abundance methods. Daily PTI counts and alternative method calculations were plotted against time and statistically compared using Pearson’s correlation. Figure 1 shows the performance of the PTI and two alternative abundance estimation methods (scent station visitation index and buried meat index) at tracking wild dog abundance with each reduction in population size in two replicate areas. Pearson’s correlation coefficients among the three indices were each greater than 0.94. Table 1 presents a summary of studies where activity of species detected as spoor on tracking stations was compared to alternative methods of estimating relative abundance.

Table 1: Australian studies where activity of species detected as spoor on tracking stations was compared to alternative methods of estimating relative abundance

<table>
<thead>
<tr>
<th>Source</th>
<th>Methods</th>
<th>Procedure</th>
<th>Species</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mason (1995)</td>
<td>Compared PTI to walked line transect and aerial surveys at three separate sites with large differences in density</td>
<td>PTI procedure compared first 2-day and last 2-day counts of 50×1 m-wide tracking stations across roads, 1 km apart, monitored for 4 days</td>
<td>Macropods</td>
<td>PTI had high repeatability (r²=0.94, p&lt;0.0001) and precision (21 % coefficient of variation). At low density/low visibility locations, PTI had higher probability of detection compared to the other abundance methods but lost sensitivity at high macropod densities when individual tracks became indistinguishable</td>
</tr>
<tr>
<td>Catling et al. (1997)</td>
<td>Comparison of spotlight counts, track counts and various trapping methods to detect ground-dwelling and arboreal species and to assess their abundance and distribution</td>
<td>284×1 m-wide tracking stations across roads, 0.2 km apart monitored for 3 days</td>
<td>Medium to large ground-dwelling animals</td>
<td>Tracking stations provided comprehensive data on distribution and abundance of wild dogs, foxes, bandicoots, wallabies, possums and cats, but various species of the same genus (e.g. small wallabies) could not be identified separately. Compared to spotlighting and trapping methods, tracking stations had the highest probability of detection for medium to large ground-dwelling species</td>
</tr>
<tr>
<td>Mahon et al. (1998)</td>
<td>Compared spotlight counts and activity of species detected on tracking stations located on roads to stations located randomly</td>
<td>4 independent sites, 8 surveys, 28 randomly located tracking stations 0.75 m×40 m. Road counts on 25 3 m×5 m plots located 0.5 km apart on 3 surveys</td>
<td>Wild dogs, foxes and cats</td>
<td>Road counts and random plots had high correlation for foxes (r=0.95, p&lt;0.01), but not cats (r=0.36, p&gt;0.05). Tracking stations on roads or randomly located plots had higher probability of detection of cats, foxes and wild dogs than did spotlighting</td>
</tr>
<tr>
<td>Edwards et al. (2000)</td>
<td>Compared track plot activity to spotlight counts</td>
<td>Four 10 km×1.8 m tracking stations monitored over 3 days for 9 surveys for over 2 years</td>
<td>Wild dogs and cats</td>
<td>Minimum number of feral cats and wild dogs detected on tracking plots was 5–55 and 8–50 times higher (cats and wild dogs, respectively) with greater precision than for spotlight counts. Probability of detecting cats and wild dogs was low for spotlight counts with low precision (68–244 and 137–265 % coefficient of variation, respectively)</td>
</tr>
<tr>
<td>Eldridge et al. (2002)</td>
<td>Compared track counts on roads and spotlight counts in paired (bailed and unbaited) treatment areas</td>
<td>Three paired replicate sites, three 10 km×1.8 m tracking stations separated by a 5 km buffer, monitored for 3 days in each treatment area. Seven surveys over 3 years</td>
<td>Wild dogs, foxes, cats, rabbits, kangaroos, large reptiles and birds</td>
<td>High precision and repeatability between surveys, significant reduction of wild dogs in baited treatments (mean of 21 %, p&lt;0.05). High correlation in activity trends between treatment areas and between species. Generally lower probability of detection and precision with alternative survey methods although good correlation between fox activity and spotlight counts (p=0.003)</td>
</tr>
</tbody>
</table>
than 0.85, indicating that all three indices had a high level of agreement in tracking daily changes in wild dog activity (full details in Allen et al. 1996). The PTI method, however, had substantially higher proportion of positive readings than the other two indices (i.e. higher probability of detection) and followed the expected decline in numbers (i.e. able to detect small changes in abundance).

Similar comparisons between the PTI values and visual counts (mean number of macropods sighted along a 50-km transect driven for four to six consecutive days) over 4 years and two replicate sites (Allen 2005) were evaluated for large macropod species (*Macropus* spp.). Values and mean daily counts were compared between sites and methods. Figure 2 shows the trends in macropod PTI values and road counts in two replicate areas. The PTI had superior sensitivity for detecting macropods at low densities and in habitats with poor visibility where counts failed to generate an abundance estimate with any level of precision. Correlation analysis between visual counts and PTI values of macropod abundance show non-zero correlation between the methods in only one of two replicate areas ($r=0.45$, degree of freedom ($df$)=18, $F=4.3$, $p=0.05$, versus $r=0.32$, $df=18$, $F=2.0$, $p=0.17$). Comparison of macropod abundance between the two areas (Fig. 2) shows correlation using PTI values ($r=0.81$, $df=18$, $F=32.3$, $p<0.001$), but visual counts correlated at times yet failed to show any overall consistency ($r=0.28$, $df=18$, $F=1.4$, $p=0.25$).

Verification: comparison of PTI trends in adjacent populations

To test the assumption that nearby populations of the same species will respond to seasonal conditions similarly and have synchronous breeding cycles, we compared PTI trends for feral cats, representing a homeotherm, and reptiles (mostly *Varanus* spp.), representing an ectotherm, in two, independent yet nearby populations. PTI values were calculated (as for wild dogs above), from two replicate areas, each of approximately 400 km² separated by a buffer of >10 km (from Allen 2005). PTI values for both species in replicate areas
were plotted against time and correlations between the two areas calculated.

The PTI values for feral cats and reptiles in the two adjoining areas are illustrated in Fig. 3. Although the PTI trends show contrasting patterns of activity between feral cats and reptiles, within species, there are almost identical patterns of activity in adjacent populations. Precision of the PTI values (represented by the error bars which are miniscule at this scale), strong correlation between adjoining areas (correlation coefficients for feral cats \( r = 0.665, t_{17} = 3.67, p = 0.002 \); reptiles \( r = 0.973, t_{17} = 12.16, p < 0.0001 \)) and annual patterns (such as winter dormancy for reptiles) consistent with expected behaviour of these species, demonstrate the PTI method as sensitive and precise.

**Discussion**

The studies used in our demonstration show that the PTI method is a sensitive and reliable method for surveying a variety of terrestrial animals, meaning the PTI showed greater probability of detection at a range of densities (Fig. 1, Table 1); had high precision in abundance values (Fig. 3, Engeman 2005), generally low variability between surveys (Figs. 2b and 3); and was able to detect incremental reductions in abundance (illustrated in Fig. 1). Although the PTI is generally superior to those methods to which PTI has been compared, the PTI lost sensitivity at high macropod densities in the study of Mason (1995) when the activity of multiple macropods muddled the spoor on tracking stations, making individual track counts uncertain.

Populations are predicted to respond to a mixture of top down and bottom-up factors, depending on their relative abundance, and where they are positioned in the food chain. Seasonal conditions (rainfall and temperature) and their effect on plant biomass (providing food resources as well as refuge from predators) form the primary driver for bottom-up factors that produce cascading effects through ascending trophic levels. Thus, seasonal conditions play an integral part on when and by how much populations would expand and contract over time. Understandably, there will be species differences in the lag time and the magnitude of population responses related to each species breeding ecology (gestation length, litter size, age of sexual maturity) and what trophic level they belong to. For example, rodents have short gestation times (21–24 days), reach sexual maturity around 35 days and can breed continuously (Watts and Aslin 1981) and, therefore, will have short eruptive responses to major rainfall events (as reported by Newsome and Corbett 1975) compared to large herbivores or predators because the latter species have longer gestation.
times, produce fewer offspring each year, reach sexual maturity at an older age and occupy higher trophic levels. PTI values increased approximately ninefold for rodent-sized mammals, sixfold for macropods and cats and twofold for wild dogs between drought and flush years (Allen 2005), consistent with this pattern. Similarly, cat, reptile (Fig. 3) and macropod (Fig. 2b) PTI values measured in adjacent areas showed near-identical activity trends consistent with the notion that abundance values and population trends should be similar for identical species living in similar habitats nearby.

The ultimate objective of validation is to establish confidence and credibility (“beyond a leap of faith”, Caughley 1977) to ensure that a method is useful in the sense that it correctly traces changes in abundance and provides accurate information about the species being monitored (Macal 2005). Without validation of the underlying methods, quality research might be rejected or devalued (see for example Witmer 2005; citing the study of Thompson and Fleming 1994 as having applied an abundance method without attempting to validate it).

Detected activity does not flawlessly trace abundance; there are environmental and behavioural stochasticity in activity with some species more susceptible to variability than others. For example, wild dog and reptile activities show large within-year variability due to changes in behaviour with seasonal peaks in autumn, corresponding to mating season and troughs in spring and corresponding to pup rearing (wild dogs, see Fig. 1 in Allen et al. 2011, and for summer activity peaks and winter dormancy of reptiles, see Fig. 3). These peculiarities do not render the PTI method invalid because the precision of the estimates and the annual long-term abundance trends (measured at consistent times of year) reveal valuable information about these species. Examining trends across years using surveys from the same time(s) of year each year is an essential component for valid study design, regardless of the monitoring method (Engeman 2005; Engeman et al. 2013). Besides, independent studies like that of Thomson (1992) on wild dogs show that these peak activity times are associated with increased vocalisation and territory maintenance during mating. Hence, activity trends are real but must be used appropriately to represent abundance; otherwise, invalid comparisons between seasons and species can occur (see critical review of published studies in Allen et al. 2011, 2013).

One might argue that population reductions imposed on wild dog populations by baiting programmes might alter their spatial movement behaviour, thereby influencing PTI values and pre-post baiting comparisons. Fortunately, a recent study of GPS-collared dingoes examined whether such changes in movement behaviour take place before and after baiting. Dingoes surviving the baiting were found not to alter their movement behaviour, including the usage of roads and tracks, demonstrating the validity of using the PTI to assess population reductions from baiting programmes (Allen 2014).

The PTI method is easy-to-use with low-labour inputs yet produces reliable results and is therefore well-suited for resource managers and researchers alike. Uncharacteristic of
alternative abundance methodologies, the PTI method does not require direct contact or interaction with animals, no conditioning of animals that might bias results, no handling of animals, no animal welfare implications and no explanation to stakeholders why animals need to be captured, marked or released again (i.e. in the case of pest species).

In this paper, we have considered the principles of validating mathematical models and applied them to a field method for estimating relative abundance, an application where testing on known populations is impractical. This approach provides a basis to evaluate the sensitivity of the PTI method and compare its responses to accepted abundance estimation methods, known data or established patterns of behaviour. In the examples cited or illustrated, the rationality of PTI methodology was examined; how conceptually animal activity and abundance are related and how various methods make similar assumptions are subject to similar biases or produce similar abundance estimates and trends. The method was then calibrated, not against known populations, but against populations representing a range of densities likely to be encountered in the field. Finally, we compared PTI results against a prediction that adjacent populations of the same species will have similar abundance values and activity trends. This same or similar validation approach can and should be applied to alternative relative abundance methods in order to demonstrate their credibility and justify their use.

References

Macal CM (2005) Model verification and validation. In: Threat anticipation: social science methods and models. The University of Chicago and Argonne National Laboratory, Chicago, IL
Mason S (1995) Activity index as a means of evaluating macropod abundance. Department of Zoology, University of Queensland, St Lucia