



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

How ideas from ecological capture-recapture models may inform multiple systems estimation analyses

Citation for published version:

Worthington, H, McCrea, R, King, R & Vincent, K 2021, 'How ideas from ecological capture-recapture models may inform multiple systems estimation analyses', *Crime and Delinquency*, vol. 67, pp. 2278-2294. <https://doi.org/10.1177/0011128720974319>

Digital Object Identifier (DOI):

[10.1177/0011128720974319](https://doi.org/10.1177/0011128720974319)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Crime and Delinquency

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



How ideas from ecological capture-recapture models may inform multiple systems estimation analyses

Hannah Worthington^{*†}, Rachel McCrea[‡], Ruth King[§], Kyle Vincent[¶]

Abstract

Abundance estimation, for both human and animal populations, informs policy decisions and population management. Capture-recapture and multiple sources data share a common structure; the population can be partially enumerated and individuals are identifiable. Consequently, the analytical methods were developed simultaneously. However, whilst ecological models have been developed to describe highly complex, biologically realistic scenarios, for example modelling population changes through time and combining different forms of data, multiple systems estimation has changed comparatively less so. In this paper we provide a brief description of the historical development of ecological and epidemiological capture-recapture and discuss the associated underlying differences that have led to model divergence. We identify three key areas where ecological modelling methods may inform and improve multiple systems estimation.

Key words: Behavioural effects; integrated modelling; multi-state modelling; temporal data.

^{*}Lecturer in Statistics, University of St Andrews, St Andrews, UK

[†]hw233@st-andrews.ac.uk

[‡]Professor of Statistics, University of Kent, Canterbury, UK

[§]Thomas Bayes' Chair of Statistics, University of Edinburgh, Edinburgh, UK

[¶]Independent Researcher, Ottawa, Canada

1 Introduction

Population assessment, management and policy decision making rely on the robust and precise estimation of the total population size of the target population of interest. Stigmatised, threatened, cryptic or hidden populations are particularly difficult to assess due to their hard-to-reach nature. Whilst a complete census of a population is typically too expensive and impractical to undertake, observing part of the population, a partial enumeration, may be feasible. In an ecological setting, capture-recapture methods are often applied where individuals are observed through time on different capture occasions. For human populations, multiple systems estimation (MSE) is often performed using data from a number of different sources. Sources will vary depending on the target population of interest and a key concept of MSE is the ability to identify individuals across the sources. Typically these sources correspond to different data lists and will be dependent on the population under study. For example, data lists may include; hospital admissions, police records and needle-exchange programmes (for injector related populations); border forces and records of non-governmental organisations (for human trafficking related populations); and humanitarian organisation records and death registries/exhumations (for war casualties). See [Bird and King \(2018\)](#) for further discussion and examples. Further, the data lists record individuals that are then uniquely identifiable using a combination of individual identifiers, such as name, date of birth, address, passport number, community health index (CHI) number (in the UK) or social security number (in the US). The underlying ideas in the data collection, for both capture-recapture and MSE, are the same (noting when or where individuals have been observed) and as a result the methods initially shared a simultaneous development. However, whilst ecological models have, for example, developed to incorporate complex structures for more realistic modelling of changes to the population through time, multiple systems estimation has continued to consider the population as a closed system, that the population is unchanging in its size during the data collection period.

MSE as a method for estimating the size of human populations has a long history. The earliest known application is generally attributed to Graunt in the 1660s for estimating the

population of London, with Laplace applying a similar technique to estimate the population of France in the 1780s (Goudie and Goudie, 2007). Modern applications of MSE include for example, estimating the number of people who inject drugs (King *et al.*, 2009, 2013, 2014), modern day slaves (Sharifi Far *et al.*, 2020a; Silverman, 2014, 2020), homeless populations (Coumans *et al.*, 2017) and the prevalence of human trafficking (Cruyff *et al.*, 2017). However, many challenges still remain for MSE and its ability to provide robust estimates of population sizes. For example, Cruyff *et al.* (2020) demonstrate the importance of model selection on population estimates and the impact of the typically sparse data sets which arise; while Sharifi Far *et al.* (2020a) consider the robustness of the estimates when lists are omitted or combined. Reliable prevalence estimates are important to not only assess the extent of these hidden populations that lead to many societal problems, in addition to the impact on the individuals themselves, but also to be able to detect trends and/or assess policy impact. Advances in ecological capture-recapture methods have led to not only increased precision of population estimates, but also more intricate-level details being identified, including for example, parameters that were previously inestimable from traditional capture-recapture data. By considering some of these statistical advances within the ecological capture-recapture literature, we wish to apply similar rationale to MSE to provide improved prevalence estimates that can better inform policy.

1.1 Brief historical perspective

Capture-recapture methods, motivated by applications in ecology, started to gain traction towards the end of the 1900s and into the 20th century. In particular, they were developed to estimate the size of animal populations using data from two capture occasions (Lincoln, 1930; Petersen, 1896), leading to what is typically referred to as the Lincoln-Petersen estimator. We note that the early approaches used by both Graunt and Laplace are direct applications of this technique. This was followed by the more general K -sample methods (Schnabel, 1938), and further developed by Darroch (1958). However, many of the assumptions of these early models were often unrealistic, including for example, homo-

generality of capture for all individuals and independence of the capture probabilities across the K -samples. To address many of these issues, the 1970s saw a divergence between the models developed for ecological capture-recapture applications and MSE for human populations to account for the different nature of the samples. In particular, within ecology, the population is typically sampled repeatedly through time at a series of capture occasions, while for epidemiological data the samples are collated across different data lists. Thus, importantly, for the ecological capture-recapture setting, there is a temporal ordering of the samples; while for the epidemiological MSE setting the ordering of the data lists are arbitrary and exchangeable. To account for the dependence of sources in MSE applications [Fienberg \(1972\)](#) proposed the set of log-linear models that permitted interactions between different data lists, so that, for example, being observed by one list makes it more/less likely to be observed by another list. These log-linear models provide the foundation for MSE. Conversely, in the temporal capture-recapture setting [Otis *et al.* \(1978\)](#) described models that described three types of dependence: time-dependence (the probability of observing individuals varying by capture occasion); behaviour (individuals captured for the first time behaving differently to those previously captured); and heterogeneity (that the probability of capture differs by individual) ([Pollock \(1974\)](#) see also a review by [Seber \(1986\)](#)). Further model developments have typically developed separately dependent on whether there is a temporal ordering of the samples or not, but in many cases often applying similar statistical ideas. These include, for example (with first citation corresponding to MSE; second citation to capture-recapture), Bayesian model-averaging ([King and Brooks, 2001, 2008](#)), incorporating unobserved individual heterogeneity ([Goodman, 1974](#); [Pledger, 2000](#)), incorporating covariate information ([Huggins, 1989](#); [King *et al.*, 2005](#)), and allowing for non-target members of the population being observed ([Overstall *et al.*, 2014](#); [Pradel *et al.*, 1997](#)). For further information on many of these issues see for example, [King and McCrea \(2019\)](#) and [Worthington *et al.* \(2019a\)](#) for the ecological case; and [Bird and King \(2018\)](#) and [Böhning *et al.* \(2018\)](#) for MSE, including discussion of application areas. However, there have been many further advances within ecological capture-recapture not reflected within the MSE framework, which we

explore in Section 2 with the aim of seeding more methodological developments within MSE.

1.2 MSE and capture-recapture synergies

The idea underlying MSE and capture-recapture is that if the population can be sampled repeatedly, either through time (typical for ecological data) or through different sources (typical for epidemiological data), then the information on when and/or where each observed individual was seen can be used to estimate the probability an individual is not seen. Hence, it is possible to estimate the number of missed individuals and the total population size. The number of unique observed individuals across all of the sources or occasions typically provides only a lower estimate for the total population size; there may be a substantial proportion of the total population not observed by any of the sources or on any occasion.

Data for MSE and capture-recapture can be expressed in the same format; through the recording of *encounter histories*. An example history might be,

$$0 \ 1 \ 1 \ 0 \ 1$$

indicating that this particular individual was observed by sources 2, 3 and 5 but missed by sources 1 and 4 (if considering sources of data), or that they were observed on occasion 2, 3 and 5 but missed on occasions 1 and 4 if considering sampling through time on different occasions. In general, suppose the total population size is given by N , of which n individuals are observed by at least one source or on at least one occasion. If there are K sources/occasions, labelled $k = 1, \dots, K$, then the encounter history for each individual in the population $i = 1, \dots, N$ is represented by $\mathbf{x}_i = \{x_{ij} : i = 1, \dots, N, j = 1, \dots, K\}$. The histories for observed individuals are combined to form an encounter history matrix \mathbf{X} where each row corresponds to an observed individual. In addition to the observed individuals there will be $N - n$ individuals with an all-zero encounter history.

Note that the encounter histories in the above form within the MSE setting simply record whether an individual is seen, or not, by each source within a given time period.

Information on whether individuals have been seen multiple times by a source (repeat sightings) and the order in which an individual was seen by different sources is not included. In general, time information specific to each individual is not retained within the data and does not feature in current models for MSE. To record and release such information may lead to confidentiality issues where individuals could potentially be identified due to their highly unique observation data. We discuss potential options for avoiding these confidentiality issues in Section 2.2.

Methods for both MSE and capture-recapture are generally based on two different statistical distributions: a Poisson model and a multinomial model. [Chao *et al.* \(2001\)](#) provides an excellent overview of the two modelling approaches. In addition to estimating the total population size N , the multinomial model estimates probabilities associated with each possible encounter history. The multinomial likelihood for N and $\mathbf{p} = \{p_{ij} : i = 1, \dots, N, j = 1, \dots, K\}$ given the encounter history matrix (and an all-zero history for missed individuals) is of the form,

$$L(N, \mathbf{p}; \mathbf{X}) \propto \frac{N!}{(N - n)!} \prod_{i=1}^N \prod_{j=1}^K \{p_{ij}^{x_{ij}} (1 - p_{ij})^{1 - x_{ij}}\}.$$

[Fienberg \(1972\)](#) and [Cormack \(1979\)](#) defined a Poisson random variable associated with each observed encounter history. Since a set of independent Poisson random variables leads to a multinomial distribution when conditioned on their sum, [Sandland and Cormack \(1984\)](#) showed that both modelling approaches lead to the same maximum likelihood estimates for the parameter of interest, the total population size N . However, the standard errors for the two modelling approaches differ - see [Cormack and Jupp \(1991\)](#). The equivalence of the approaches in the Bayesian framework, given particular prior specifications on the intercept of the Poisson approach, or total population size in the multinomial specification is explored by [Forster \(2010\)](#). Generally the individual level encounter histories are used for the multinomial approach, whilst summarised contingency tables are used for the Poisson approach. [Bird and King \(2018\)](#) provide an extensive review of the contingency table based approaches, while [King and McCrea \(2019\)](#) provide a perspective building on the multinomial basis.

Both MSE and the models described above for ecological capture-recapture assume that the population is closed. This assumes there are no arrivals or departures from the population during the period over which the data are collected, equivalently that the individuals that form the population being sampled is unchanging. Under highly restrictive conditions, for example very short sampling periods, this assumption may be justifiable, but for many populations under study this is highly unlikely. Data for MSE is often aggregated by year, or perhaps longer, and so the definition of the total population size can be unclear. Assuming closure implies that all individuals were available for the whole sampling duration. Perhaps a more realistic count would be those individuals that were part of the population of interest *at some point* during the sampling period. This latter suggestion requires the possibility that individuals can enter and leave the population at any time. Capture-recapture models commonly work within this open population framework, for example, modelling survival or retention of individuals and explicitly modelling arrivals into the population (Cormack, 1964; Jolly, 1965; King, 2014; McCrea and Morgan, 2014; Newman *et al.*, 2014; Pledger *et al.*, 2009; Schwarz and Arnason, 1996; Seber, 1965; Worthington *et al.*, 2019b).

1.3 Outline of paper

In Section 2 we explore three developments from ecological capture-recapture models that may be used to inform and improve the estimation of population size through MSE. In Section 3 we discuss whether there are elements of MSE that could benefit capture-recapture methods, in particular the combining of different dependent sources of data and consider future developments in both areas.

2 Ecological advances for potential application to MSE

In this section we describe three developments from ecological capture-recapture models and discuss their synergies with MSE. In particular, we discuss: the assumptions relating to interactions between different sources (or capture occasions); individual heterogeneity

and the closure assumption, particularly when data are collected over an extended period of time; and the combining of different forms of data within a single coherent analysis.

2.1 Temporal and behavioural effects

Within ecological studies the capture occasions have a natural temporal order. This is in contrast to the analogous sources used within epidemiological MSE where the sources themselves have no natural order (the encounter histories would change if the sources were reordered). For individuals recorded by multiple sources the temporal information is not available from the contingency table. The presence or absence of the temporal component (for ecological and epidemiological studies, respectively) has a direct impact on the modelling of the data and associated interpretation of the model parameters. However, there remains some commonality and interesting comparisons, motivating further useful avenues of research.

For ecological capture-recapture studies, the model is typically parameterised in terms of the (direct) probabilities of observing an animal on a given capture occasion conditional on its capture history to date (Borchers *et al.*, 2002; McCrea and Morgan, 2014). These time-dependent capture probabilities are combined to form the associated probability of each observed encounter history (equivalently the probabilities associated with each cell of the contingency table). For example, for encounter history $\mathbf{x}_i = \{x_{i1}, \dots, x_{iK}\}$, we have,

$$\mathbb{P}(\mathbf{x}_i) = \mathbb{P}(x_{i1}) \prod_{k=2}^K \mathbb{P}(x_{ik} \mid x_{i1}, \dots, x_{i(k-1)}).$$

In general, even when the study design is specified to minimise the variability of capture across capture occasions, the capture probabilities may still vary by occasion. This may be due to changing weather conditions, or changing behaviour of the individuals over time due to breeding behaviour etc. In this case of time-dependent capture probabilities, assuming that the capture probabilities are common to all individuals so that there is no additional individual heterogeneity to consider and that the capture probabilities across capture occasions are independent, we obtain the time-dependent model denoted by M_t

(Otis *et al.*, 1978). Traditionally, given the cell probabilities it is natural to specify the model within the multinomial framework, with associated model parameters corresponding to the probabilities of being observed at each capture occasion and total population size. This model is equivalent to the independent model for MSE, where the capture of an individual by a particular source does not affect their probability of capture by any of the remaining sources. The Poisson formulation of the independent model instead specifies the mean of the cell probabilities in log-linear form with only the intercept and main effect terms present (Chao *et al.*, 2001). The probabilities for the separate capture occasions can be expressed as a function of the log-linear main effect terms (the exact relationship depends on the particular constraints specified on the terms to achieve uniqueness).

In practice, it is often the case that the capture probabilities are not independent across the different occasions. In particular, we may have behavioural effects where the capture of an individual influences its future capture probabilities (Otis *et al.*, 1978). This is typically referred to as behavioural effects which may correspond to either: (i) a “trap happy” response, where the future recapture probability of an individual is increased following its initial capture (this may occur for example if food is provided to captured individuals); or (ii) a “trap shy” effect, where the future capture probability of an animal is decreased following its initial capture (for example, the trapping and tagging of an animal may be an unpleasant and stressful experience, as a result the individual may identify and avoid future traps). The simplest behavioural model is denoted M_b ; and in the presence of both time and behavioural dependence, M_{tb} . There are numerous types of behavioural effects dependent on the biological setting and known characteristics of the animal. For example, the behavioural response may be a permanent response; or individuals may have a memory of the trap that decreases over time. For such behavioural models the temporal structure of the capture occasions is critically important, as the capture probability of an individual at time k now depends on its previous history. Equivalently, the capture probability of an individual at a given time depends on whether the capture is an *initial* capture or a *recapture* along with a possible model for the dependence on time since previous capture.

We initially consider the behavioural response such that the capture of an individual influences all future capture occasions. In other words, an individual initially captured on occasion k has an increased/decreased capture probability at all future times $\tau = k + 1, \dots, K$ (corresponding to a trap happy/shy response, respectively). This ecological behavioural effect model is in some ways conceptually similar to the log-linear MSE model with all two-way interactions present. Similar patterns hold for alternative behavioural response models. For instance, if the capture of an individual influences *only* the next capture occasion (an individual captured on occasion k has an increased/decreased capture probability at time $k + 1$), then the capture probability for occasion $\tau = k + 2, \dots, K$ is independent of whether or not an individual is captured on occasion k . This is similar to the log-linear model where there are two-way interactions between only “consecutive” sources (where consecutive here relates to the given ordering of the sources listed rather than a chronological/temporal order).

However, there is a fundamental difference between the ecological behavioural models and the two-way interaction log-linear models with important knock-on effects and interpretations. In particular, the behavioural response in the ecological models is a “forward” or “directional” interaction only - for example, the probability of being observed at time $k + 1$ is a function of whether or not an individual is observed at time k ; the probability of being observed at time k is not a function of whether or not it is observed at time $k + 1$. However, for MSE log-linear models, the two-way interaction between sources is symmetrical - if source A affects source B ; then source B affects source A . For a positive interaction, being observed by source A increases the probability of being observed by source B ; and being observed by source B increases the probability of being observed by source A . Similarly, for a negative interaction but the probability of observation by the other source is decreased.

The comparison of log-linear models with the ecological behavioural models raises some interesting perspectives. In many cases, an individual observed by one source may be referred onwards to another source(s). For example, non-governmental organisations may pass on details of individuals to police who then also identify the same individuals

when investigated further; however police may not refer individuals to non-governmental organisations. Such a process describes a directional interaction. Standard log-linear models are unable to formally model such a process (all interactions are symmetric as there is no temporal or referral information); and not incorporating these mechanisms can lead to poor performance (Jones *et al.*, 2014). The ecological capture-recapture models thus potentially motivate the inclusion of temporal information within multiple-source data, thus permitting the development of models with directional interactions for MSE.

2.2 Open population models

The models discussed in the previous sections assume that the population being estimated is closed. The estimate of the total population is therefore a “snapshot” estimate assuming that individuals did not leave the population (due to death, migration or no longer being a member of the target population) nor did new individuals join the population (birth, migration or becoming a member of the target population). Whilst policy makers appear to prefer “snapshot” estimates, the estimation of the population size through time may be more informative by identifying changes occurring within the population.

For example, suppose a contingency table summarises the data collected over a 2-year period by multiple sources. The traditional MSE estimate for N would be interpreted as there being N individuals in the population throughout the two-year period. This singular number can give no indication of increases or decreases in the population over this time. There may have been N unique individuals in the population over the course of the two years, but they may not have all been present concurrently; the number in the population at any one time may not have been as high as N individuals, or the size of the population at the end of the two years may be much smaller/larger than at the start of the period. If MSE is being undertaken to better manage resources, for example health care, then perhaps tracking changes to the population through time could be more informative and beneficial.

Many of the standard open population capture-recapture models, in addition to estimating capture probabilities, estimate apparent survival or retention probabilities. These

parameters express the probability that an individual currently in the population on occasion k is still present in the population on occasion $k + 1$ (Cormack, 1964; Jolly, 1965; Schwarz and Arnason, 1996; Seber, 1965). Stopover models explicitly model the arrival of individuals into the population (Pledger *et al.*, 2009; Worthington *et al.*, 2019b). The reason that arrivals and retention probabilities must be modelled is due to the unknown state of an individual before their first capture and after their final capture. An individual may be present in the population for some time before being initially captured, similarly, an individual may still be in the population but not captured after they are seen for the final time. This unknown state of the individual, whether they are in the population or not, can be modelled as a “hidden” state (or partially observed state since uncertainty only arises over some periods of time). If these hidden states can be established for every individual in the population, including those never seen, then the size of the population at different points in time can be estimated. Multi-state stopover models (Worthington *et al.*, 2019b) offer a further extension to include capture heterogeneity. In addition to states tracking whether an individual is present in the population they can also refer to observable states (e.g. breeding status, location). These observable states may have very different capture probabilities and the time of transition between states may again be unknown (since states are only known when an individual is captured).

If similar multi-state models were to be applied in an MSE setting, then time information would be required. The progression of states from not in the population, to joining the population, to leaving the population, occur in a natural order; it is simply the timing of the transitions that is uncertain. The extra information that would be required could however lead to more informative investigation of the population. For instance, if the arrival and departure time of individuals can be estimated, then the amount of time individuals spend in the population can also be estimated and time spent in the population could inform the probability of capture by a source. If the states refer to the sources that have captured an individual then the transitions between states could model resighting at a source that has already recorded the individual or capture by a further source. This could open up possibilities to identify the expected time gap between sources and potential referrals between

sources.

The data required for time-dependent modelling in an MSE setting may be difficult to obtain. To model transitions between sources it is possible that very large datasets would be required in order to obtain a sufficient number of observations of the different orderings of sources - a problem that would increase significantly with the number of sources used. The largest issue will be in protecting individual identities. By simply retaining the sources that have observed an individual there is a reasonable degree of anonymity. Unless there are very small cell counts individuals will not be identifiable. However, if highly specific covariate information were collected, such as the time of observation by a source, then there is the potential for individuals to be identified. This may be mitigated by instead assigning an arbitrary “time 0” and recording the time gap between observations. The models described here operate in discrete time, and so further anonymity may be achieved by careful selection of the discretisation, though again large datasets may be required in order to have several individuals identified in any one discrete time period.

2.3 Integrated modelling

Integrated population modelling in ecology refers to the combined analysis of multiple data types. The concept was first proposed in [Besbeas *et al.* \(2002\)](#), where ring-recovery data modelled using a product multinomial likelihood was analysed in conjunction with population counts (or population index data) described using a state-space model. This was the first time two disparate modelling approaches were unified into a single analysis. The global model describing both types of data simulatenously requires the assumption of independence of the data as the global likelihood function is formed as the product of component likelihoods. Although some concern is raised regarding the validity of this assumption it has been found that violation of this model assumption does not result in appreciable bias in the estimators (see for example [Abadi *et al.* \(2010\)](#) and [Besbeas *et al.* \(2009\)](#)). One of the benefits of analysing disparate data sets simultaneously is that you can obtain improved precision of some parameters. This is particularly noticeable in the case of multi-state models where estimates of transition probabilities are often associated

with large uncertainty, and the addition of state-specific population counts modelled using the state-space framework improves the precision (McCrea *et al.*, 2010).

It is also the case that it is not possible to estimate certain parameters from a single source of data due to parameter redundancy (Cole and McCrea, 2016; Sharifi Far *et al.*, 2021b; Vincent *et al.*, 2021). For example if just census counts are available you cannot separate the estimation of fecundity or productivity and first year survival. Therefore, by analysing the census data in conjunction with another source of data such as ring-recovery data which contains information on survival, it is then possible to estimate both survival and fecundity.

Within the MSE models there are two parameters: the unknown population size and the capture probabilities. Therefore if other sources of data might provide additional information on capture probability this will result in better estimates of both parameters (due to the correlation of the parameters improvement in precision of capture probability will result in improved precision of N which is the primary parameter of interest).

3 Conclusions and Further Directions

In this paper we have discussed similarities between ecological capture-recapture studies and epidemiological MSE; and focused on three key areas in which capture-recapture methods may inform and improve MSE analyses. Whilst sharing a similar model structure both capture-recapture and MSE can generally be criticised for the assumptions they make. Broadly speaking, MSE ignores temporal information and assumes a static population size whilst capture-recapture ignores potential dependence of observations.

Capture-recapture methods including temporal effects or open population models offer an opportunity for more realistic modelling of the population being counted through time. The incorporation of time into MSE analyses would require a different data structure to the summarised contingency tables that are currently typical. Individual specific information would need to be retained and the issues surrounding the protection of identity would need careful consideration. The benefits of the increased understanding of the

dynamics of the population however may be significant.

An advantage of MSE analyses, that may be relevant to capture-recapture, is the dependence between the sources of data. This aspect is readily accounted for within the model using two-way, or higher, interactions. The interpretation of these interaction terms is not as readily understood in the case of multiple sampling occasions through time. However, surveying of animals can take different forms of which capture-recapture data is only one. It can be advantageous to include multiple forms of information within a single analysis, fitting the model within a single framework. Integrated modelling includes an assumption of independence between the sources, but an approach where this can be relaxed may be preferable and MSE could inform this approach. A potential application is to the analysis of migration data. If there are multiple sites that a species may attend, the choice of site, or the sites an individual is seen at may be influenced by the combinations of sites themselves. Including dependence between the sources, in this case sites, may allow for instance the modelling of related increases/decreases in sightings between sites (similar sites influencing each other for example).

Many capture-recapture studies are repeated annually with data then available across multiple years. Models exist that do not only consider a single-year of data but instead operate on two time scales; a primary level scale operating across several years, and a secondary scale operating within a single year. Many MSE analyses involve data collected over several years, aggregated by year. There may be scope for these capture-recapture models to be applied to MSE data, where individuals could be tracked through years as well as across sources. Capture-recapture data, in addition to time, may also include spatial information on the location a capture occurred. Links between the non-independence of the sources in MSE with the spatial density of a species might be an interesting avenue for further consideration.

There is clearly potential for the two academic communities from ecological statistics and MSE to collaborate to maximise the potential of the information contained in respective data sets.

References

- Abadi, F., Gimenez, O., Arlettaz, R. and Schaub, M. (2010) An assessment of integrated population models: bias, accuracy, and violation of the assumption of independence. *Ecology* **91**, 7–14.
- Besbeas, P., Freeman, S. N., Morgan, B. J. T. and Catchpole, E. A. (2002) Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* **58**, 540–547.
- Besbeas, P., Borysiewicz, R. S. and Morgan, B. J. T. (2009) Completing the ecological jigsaw. In *Modeling demographic processes in marked populations.*, ed. DL Thomson, EG Cooch, and MJ Conroy, 513—539. *Environmental and Ecological Statistics*, Volume 3. Springer-Verlag, New York, New York, USA.
- Bird, S. M. and King, R. (2018) Multiple systems estimation (or capture-recapture estimation) to inform public policy. *Annual Review of Statistics and Its Application* **5**, 95–118.
- Böhning, D., and van der Heijden, P. G. M. and Bunge, J. (2018) *Capture-Recapture Methods for the Social and Medical Sciences*. CRC Press, Boca Raton.
- Borchers, D. L., Buckland, S. T. and Zucchini, W. (2002) *Estimating Animal Abundance: Closed Populations*. Springer-Verlag.
- Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. and Chao, D. Y. (2001) The applications of capture-recapture models to epidemiological data. *Statistics in Medicine* **20**, 3123—3157.
- Cole, D. and McCrea, R. S. (2016) Parameter redundancy in discrete state-space and integrated models. *Biometrical Journal* **58**, 1071–1090.
- Cormack, R. M. (1964) Estimates of survival from the sighting of marked animals. *Biometrika* **51**, 429–438.

- Cormack, R. M. (1979) Models for capture-recapture. In *Sampling Biological Populations*, ed. RM Cormack, GP Patil and DS Robson, 217–255. Fairland: International Co-operative Publishing House.
- Cormack, R. M. and Jupp, P. E. (1991) Inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika* **78**, 911–916.
- Coumans, A. M., Cruyff, M., Van der Heijden, P. G. M., Wolf, J. and Schmeets, H. (2017) Estimating homelessness in the Netherlands using a capture-recapture approach. *Social Indicators Research* **130**, 189–212.
- Cruyff, M., J. van Dijk, and P. G. M. van der Heijden (2017) The challenge of counting victims of human trafficking: Not on the record: A multiple systems estimation of the numbers of human trafficking victims in the Netherlands in 2010-2015 by year, age, gender, and type of exploitation. *Chance* **30**, 41–49.
- Cruyff, M., Overstall, A., Papathomas, M. and McCrea, R.S. (2020) Multiple System Estimation of Victims of Human Trafficking: Model Assessment and Selection. *Crime and Delinquency*. Revision submitted.
- Darroch, J. N. (1958) The multiple recapture census. I. Estimation of a closed population. *Biometrika* **45**, 343–359.
- Fienberg, S. E. (1972) The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika* **59**, 591–603.
- Forster, J. J. (2010) Bayesian inference for Poisson and multinomial log-linear models. *Statistical Methodology* **7**, 210–224.
- Goodman, L. G. (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.
- Goudie, I. B. J. and Goudie, M. (2007) Who captures the marks for the Petersen estimator? *Journal of the Royal Statistical Society: Series, A* **170**, 825–839.

- Huggins, R. M. (1989) On the statistical analysis of capture experiments. *Biometrika*, **76**, 133–140.
- Jolly, G. M. (1965) Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* **52**, 225–247.
- Jones, H. E., Hickman, M., Welton, N. J., De Angelis, D., Harris, R. J. and Ades, A. E. (2014) Recapture or Precapture? Fallibility of standard capture-recapture methods in the presence of referrals between Sources. *American Journal of Epidemiology* **179**, 1383–1393.
- King, R. and Brooks, S. P. (2001) On the Bayesian analysis of population size. *Biometrika*, **88** 317–336.
- King, R., Bird, S. M., Brooks, S. P., Hay, G. and Hutchinson, S. (2005) Sex, age-group, and region influence both injectors' propensity to be listed on capture-recapture data sources and their drugs-related mortality: Prior information and uncertainty in behavioural capture-recapture methods. *American Journal of Epidemiology* **162**, 694–703.
- King, R. and Brooks, S. P. (2008) Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics* **64**, 816–824.
- King, R., Bird, S. M., Hay, G. and Hutchinson, S. J. (2009) Estimating current injectors in Scotland and their drug-related death rate by sex, region and age-group via Bayesian capture-recapture methods. *Statistical Methods in Medical Research* **18**, 341—359.
- King, R., Bird, S. M., Overstall, A., Hay, G. and Hutchinson, S. (2013) Injecting drug users in Scotland, 2006: Number, demography, and opiate-related death-rates. *Addiction Research and Theory* **21**, 235–246.
- King, R., Bird, S. M., Overstall, A., Hay, G. and Hutchinson, S. J. (2014) Estimating prevalence of injecting drug users and associated heroin-related death-rates in England

- by using regional data and incorporating prior information. *Journal of the Royal Statistical Society: Series A* **177**, 209–236.
- King, R. (2014) Statistical Ecology. *Annual Review of Statistics and its Application*, **1**, 401-426.
- King, R. and McCrea, R. S. (2019) Capture-Recapture Methods and Models: Estimating Population Size. In *Handbook of Statistics* Volume 40, 33–83.
- Lincoln, F. C. (1930) Calculating waterfowl abundance on the basis of banding returns. Circular No. 118, U.S. Department of Agriculture, Washington, DC.
- McCrea, R. S., Morgan, B. J. T., Gimenez, O., Besbeas, P., Lebreton, J. and Bregnballe, T. (2010) Multi-site integrated population modelling. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 539–561.
- McCrea, R. S. and Morgan, B. J. T. (2014) *Analysis of Capture-Recapture Data*. Boca Raton, FL: CRC Press/Chapman & Hall.
- Newman, K. B., Buckland, S. T., Morgan, B. J. T., King, R., Borchers, D. L., Cole, D. J., Besbeas, P. T., Gimenez, O. and Thomas, L. (2014) *Modelling Population Dynamics: Model Formulation, Fitting and Assessment using State-space Methods*. Springer.
- Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62**, 3–135.
- Overstall, A., King, R., Bird, S. M., Hutchinson, S. H. and Hay, G. (2014) Incomplete contingency tables with censored cells with application to estimating the number of people who inject drugs in Scotland. *Statistics in Medicine* **33**, 1564–1579.
- Petersen, C. G. J. (1896) The yearly immigration of young plaice into the Limfjord from the German Sea. *Report of the Danish Biological Station* **6**, 5–8.
- Pledger, S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**, 434–42.

- Pledger, S., Efford, M., Pollock, K. H., Collazo, J. A. and Lyons, J. E. (2009) Stopover duration analysis with departure probability dependent on unknown time since arrival. *Environmental and Ecological Statistics* (Edited by D. L. Thomson, E. G. Cooch and M. J. Conroy) **3**, 349—363.
- Pradel, R., Hines, J. E., Lebreton, J.-D. and Nichols, J. D. (1997) Capture-recapture survival models taking account of transients. *Biometrics* **53**, 60–72.
- Pollock, K. H. (1974) The assumption of equal catchability of animals in tag-recapture experiments. PhD thesis, Cornell University, Ithaca, New York.
- Sandland, R. L. and Cormack, R. M. (1984) Statistical inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika* **71**, 27–33.
- Schnabel, Z. E. (1938) The estimation of the total fish population of a lake. *American Mathematical Monthly* **45**, 348–352.
- Schwarz, C. J. and Arnason, A. N. (1996) A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics* **52**, 860–873.
- Seber, G. A. F. (1965) A note on the multiple-recapture census. *Biometrics* **52**, 249–259.
- Seber, G. A. F. (1986) A review of estimating animal abundance. *Biometrics* **42**, 267–292.
- Sharifi Far, S., King, R., Bird, S. M., Overstall, A., Worthington, H. and Jewell, N. P. (2020a) Multiple systems estimation for modern slavery: robustness of list omission and combination. *Crime and Delinquency* – in press.
- Sharifi Far, S., Papathomas, M. and King, R. (2021b) Parameter redundancy and the existence of the maximum likelihood estimates in log-linear models. *Statistica Sinica* – in press.
- Silverman, B. (2014) Modern slavery: an application of multiple systems estimation. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/386841/Modern_Slavery_an_application_of_MSE_revised.pdf.

- Silverman, B. (2020) Model fitting in multiple systems analysis for the quantification of modern slavery: classical and Bayesian approaches. *Journal of the Royal Statistical Society: Series A* **183**, 691–736.
- Vincent, K., and Sharifi Far, S. and Papatomas, M. (2020) Common methodological challenges encountered with multiple systems estimation studies. *Crime and Delinquency* – conditionally accepted.
- Worthington, H., McCrea, R. S., King, R. and Griffiths, R. A. (2019a) Estimation of population size when capture probability depends on individual states. *Journal of Agricultural, Biological and Environmental Statistics* **24**, 154—172.
- Worthington, H., McCrea, R. S., King, R. and Griffiths, R. A. (2019b) Estimation abundance from multiple sampling capture-recapture data via a multi-state multi-period stopover model. *The Annals of Applied Statistics*, **13**, 2043–64.

Author Biographies

Hannah Worthington is a Lecturer in Statistics in the School of Mathematics and Statistics at the University of St Andrews. Her research interests include hidden Markov models applied to problems in ecology, capture-recapture data, incorporating individual heterogeneity and multi-state modelling.

Email: hw233@st-andrews.ac.uk; Tel: 44-1334-461806

Rachel McCrea is a Professor of Statistics at the School of Mathematics, Statistics and Actuarial Science at the University of Kent. She is also Director of the National Centre for Statistical Ecology. Her research interests include integrated population modelling, multi-state models, capture-recapture data and removal and re-introduction modelling.

Email: r.s.mccrea@kent.ac.uk; Tel: 44-1227-824760

Ruth King is the Thomas Bayes' Chair of Statistics in the School of Mathematics at the University of Edinburgh. Her research interests include statistical modelling, capture-

recapture data, multiple systems estimation and missing data applied to problems in ecology, epidemiology and healthcare.

Email: ruth.king@ed.ac.uk; Tel: 44-131-6505947.

Kyle Vincent is an independent consultant whose research focuses on developing innovative strategies that can be used to efficiently study hard-to-reach populations. Much of his work focuses on using network sampling strategies and mark-recapture procedures, which is motivated by challenges encountered when studying populations such as those comprised of human trafficking victims, drug users, or commercial sex workers.

Email: kyle.shane.vincent@gmail.com; Tel: 613-218-0880.