Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7,** 147–156 doi: 10.1111/2041-210X.12519

DEMOGRAPHY BEYOND THE POPULATION

Inverse estimation of integral projection model parameters using time series of population-level data

Edgar J. González[1]*, Carlos Martorell[2] and Benjamin M. Bolker[1,3]

[1]*Department of Mathematics and Statistics, McMaster University, 1280 Main St W, L8S4L8 Hamilton, Ontario, Canada;* [2]*Departamento de Ecología y Recursos Naturales, Facultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, 04510 México D.F., Mexico; and* [3]*Department of Biology, McMaster University, 1280 Main St W, L8S4L8 Hamilton, Ontario, Canada*

Summary

1. Integral projection models (IPMs) allow us to describe quantitatively the dynamics of a population structured by a continuous variable. They rely on information gathered at the individual level by recording survival, reproduction and changes in some structuring variable over time. This requires the ability to track individuals over the course of their entire life cycle. When this is not feasible, we would like to use alternative information to infer a population's dynamics. Time series of population-level data are an option.

2. An inverse modelling approach allows inferring the vital rates of a population when only population-level data, in the form of a time series of the size of a population and the distribution of its individuals along a structuring variable, are available. The approach also allows incorporating estimates obtained through individual-level data. Here, we explore how inverse modelling performs with simulated data and a relatively simple demographic model. We explore scenarios of data availability in terms of time-series length, per-year sample size and availability of independent vital-rate estimates. We also test model performance in a real system using a 15-year long data set from a chamaephyte plant, *Cryptantha flava*.

3. We show that an inverse model can provide accurate reconstructions of the vital rates in a scenario where no individual-level information is available. Better results can be obtained if independent estimates on any vital rate are provided, as was the case for *C. flava* where high interannual variation is present. Parameter estimation becomes more difficult with shorter time series, but per-year sample size can be greatly reduced without significantly affecting parameter accuracy.

4. Inverse modelling of IPMs allow for the estimation of unobserved vital rates, which is important for systems where any or all of the vital rates are hard to quantify. It also helps to determine whether a forward IPM is capturing the population dynamics: if the inverse version produces incorrect reconstructions of the vital rates, the forward IPM can be considered as inadequately describing the system.

Key-words: integrated population model, inverse modelling, inverse problem, model diagnosis, population dynamics, population size, population structure, vital rates

Introduction

Quantifying the demographic behaviour of a species is a key element for understanding its present status and to explore possible scenarios of conservation or management. However, this quantification has traditionally been difficult because it requires tracking individuals over time. This may prove unfeasible because of methodological or ethical reasons or because tracking may impact an organism's adequacy (May 2004; Brooks, McCoy & Bolker 2013). In these cases, ecologists require methods that allow them to infer the demographic behaviour of a population using alternative sources of information that are feasible to gather (Silvertown, Franco & Menges 1996).

In a closed population, its demographic behaviour is determined by the interplay between two vital rates: the recruitment of newborns and the survival of existing individuals. Because individuals differ in their traits (e.g. size, sex, genetic configuration, health status) or in the environment they experience (e.g. temperature, predation intensity, disturbance), these rates differ among individuals (Moran, Hartig & Bell 2015). By increasing variation, we say that these variables *structure* the population, and thus that the vital rates are structured (Tuljapurkar & Caswell 1997; Merow *et al.* 2014). When individuals change their traits or the environment they experience

*Correspondence author. E-mail: edgarjgonzalez@ymail.com

This article forms part of the British Ecological Society journals' Demography Beyond the Population special feature http://wileyonlinelibrary.com/BES_demography

changes over time, new rates become relevant (e.g. individual growth rate when the structuring variable is size). To understand the effect on the vital rates of these variables, we need to measure them at the individual level, which, as stated above, can be difficult to do. However, knowing how vital rates differ among individuals is essential for the development of conservation and management strategies targeted towards those individuals and vital rates that more strongly affect the overall state of a population (Crouse, Crowder & Caswell 1987).

Integral projection models (IPMs) allow studying the dynamics of populations structured by continuous variables (Rees, Childs & Ellner 2014). In an IPM, data gathered by tracking the life cycle of the individuals are used to fit a model to each vital rate. The set of models is then assembled into a single function, called the *kernel*, which fully describes the expected changes in the population from one time step to the next. Starting from an initial population structure, iterating this function produces a time series of expected population structures and sizes. Traditionally, the models used to describe the vital rates have depended on the particular system under study as well as on the type of data available, but the general approach has been to fit some kind of linear model (generalized, additive, mixed, etc.; but *cf*. Rees, Childs & Ellner 2014) to the data on each vital rate separately (Metcalf *et al*. 2013; Merow *et al*. 2014). Using linear models (and their extensions) allows for a relatively small number of parameters to mathematically describe each vital rate, greatly reducing model complexity in contrast to projection models developed for populations structured by discrete variables (i.e. matrix projection models; Easterling, Ellner & Dixon 2000).

The relative simplicity of the demographic models underlying IPMs also allows us to consider *inverse modelling* scenarios, that is cases where only population-level data are available in the form of a time series of population structures and sizes, and we want to infer the value of the parameters associated with each vital rate. They also allow us to explore scenarios where estimates on a subset of the vital rates are available and we would like to infer estimates for unobserved ones by combining the known parameter estimates with population-level data.

In a general sense, solving an inverse problem involves taking the outcome of a series of processes and attempting to estimate these processes. As suggested by the name, every inverse problem relates to a forward one. In our case, the forward problem is essentially an IPM, relating the individual-level vital rates mechanistically to the population sizes and structures observed over time (Fig. 1, left panel).

Versions of this inverse problem have been previously proposed. Fournier, Hampton & Sibert (1998) used age-structured matrix projection models to show that an inverse approach works in the context of fisheries. Wood (1994) applied this approach to infer birth and mortality rates in populations where (continuous) age structures the population. Ghosh, Gelfand & Clark (2012a) used an IPM to reconstruct the vital rates using population structures as input to a similar problem. González & Martorell (2013) attempted an even harder problem: to reconstruct structured vital rates that change over time. These studies have shown that under different assumptions, using inverse modelling is a viable approach to estimate vital rates when only population-level data are available.

In this article, we use inverse modelling to infer some or all of the vital rates associated with a structured population. To do so, we assume that population-level data are available in the form of a time series of population sizes and structures (Fig. 1, right panel). We use simulated data to evaluate how well this approach works for the case where the underlying vital rates as described by the IPM do not change through time, but the population structure has not reached its asymptotic (stationary) state. We show that the length of the time ser-

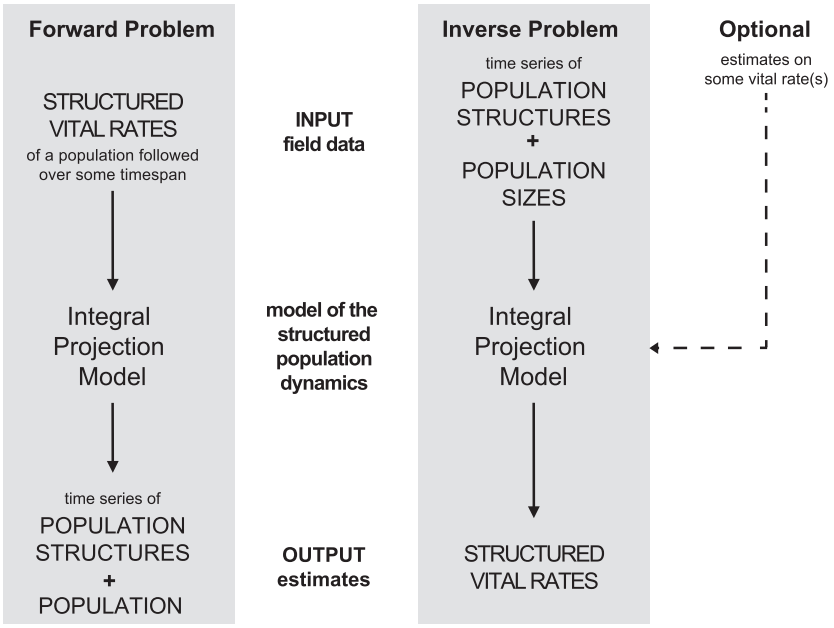

Fig. 1. Comparison of forward (left) and inverse (right) estimations of integral projection model parameters.

ies is more important than per-year individual measurements for reliably estimating model parameters. Finally, we apply the method to data on a real species, the chamaephyte plant *Cryptantha flava*. Due to the high interannual variation, an IPM with constant-through-time vital rates did not adequately capture the dynamics of the population. Therefore, additional information on individual growth rate and among-year variation in survival probability was required to reconstruct the remaining vital rates. We further discuss the implications of inverse modelling versus the more traditional forward approach (Fig. 1, left panel).

Materials and methods

INTEGRAL PROJECTION MODEL

Easterling, Ellner & Dixon (2000) presented the first IPM. In their model, a single variable, size (x), structures a closed population and the vital rates are assumed to remain constant over time. Under these circumstances, the following equation determines how the population structure (n) changes over time as the vital rates (s, g, f_1 and f_2) modify it:

$$n_{t+1}(y) = \int [s(x) \cdot g(y,x) + f_1(x) \cdot f_2(y,x)] \cdot n_t(x)dx, \qquad \text{eqn 1}$$

where the vital rates are $s(x)$, the survival probability of extant x-size individuals; $g(y,x)$, the growth function, or, more precisely, the probability an individual has of changing from size x to y from one time unit to the next; $f_1(x)$, the number of newborns produced by an x-size individual each time unit, and $f_2(y,x)$, the size distribution of newborns produced by x-size individuals.

As eqn 1 shows, an IPM relates the vital rates with a time series of population structures (n_1, n_2,...), and a time series of population sizes, given by the integration of each population structure over the observed size range ($N_1 = \int n_1(x)dx$, $N_2 = \int n_2(x)dx$,...). Data on the vital rates has been traditionally used as input to model the time series. Here, we use data on the latter as input, thus posing an inverse problem (Fig. 1).

There are many alternative functions to describe how vital rates interact to produce the size structure (Merow *et al.* 2014; examples in this Feature). As a case study, we use an IPM with the following structure, close to Easterling, Ellner & Dixon's (2000) original model:

$$s(x) = \text{logistic}(\beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2), \qquad \text{eqn 2}$$

$$g(y,x) = \text{Normal}(\mu = \beta_3 + \beta_4 \cdot x, \sigma = \exp(\beta_5 \cdot \ln(10)),$$

$$f_1(x) = \exp(\beta_6 + \beta_7 \cdot x),$$

$$f_2(y) = \text{Normal}(\mu = \beta_8, \sigma = \exp(\beta_9 \cdot \ln(10))).$$

The vital rates are thus determined by the value of 10 parameters: β_0,..., β_9 (see Appendix S1, Supporting information for details). The inverse problem consists of estimating these parameters using as data a time series of observed population sizes and structures.

We also explored simpler scenarios where estimates on some vital rate (s) were available (i.e. where the optional part in Fig. 1 is included). In many cases, population ecologists have estimates on just some vital rates and would like to infer those that are missing. Therefore, we used as input to the model the estimates associated with observed vital rates along with the time series of population sizes and structures. We explored all possible scenarios where one, two or three vital rates are unknown.

PARAMETER ESTIMATION

To assess whether any given set of values for the 10 parameters can reproduce the observed time series, we substituted these in eqn 2 and calculated the time series of population structures through the iteration of eqn 1, and the time series of population sizes by integrating the structures ($N_i = \int n_i(x)dx$). We used as initial population structure, $n_0(x)$, the first observed size structure (see Appendix S1, Supporting information for details).

The observed and estimated time series can then be compared through a likelihood function. To do this, we need to assign error distributions to the estimated time series. For each population structure in the time series, we assumed that a multinomial distribution describes the probability of observing n individuals of size x. For each point in time i, the probabilities are the estimated population structure n_i. For each population size, we assumed that a Poisson distribution describes the probability of observing N individuals. Again, for each time i, this probability is the estimated population size N_i. Therefore, we had two log-likelihood functions: one that measures the goodness-of-fit between the estimated and observed population structures, l_n, and another that measures the goodness-of-fit between the estimated and observed population sizes, l_N. The overall fit was then given by:

$$l = l_N + l_n. \qquad \text{eqn 3}$$

Estimating all of the parameters simultaneously, rather than independently for each vital rate as in the forward IPM case is a rather challenging computational problem. To find the parameter values producing the maximum likelihood, we resorted to the sequential use of two algorithms: DiffeRential Evolution Adaptive Metropolis (DREAM; Vrugt *et al.* 2009) and Automatic Differentiation Model Builder (ADMB; Fournier *et al.* 2012). We required the use of two algorithms because, under a likelihood approach, we need to find the mode of the likelihood surface. An optimization algorithm (in our case, ADMB) is a good option. However, from preliminary analyses (not shown here), we identified the existence of multiple local maxima and regions where the surface is flat (i.e. the gradient is virtually zero). These facts hindered the use of ADMB alone since it cannot deal with these issues. Therefore, we needed to first identify an area, both unimodal and positive definite everywhere (i.e. with a positive gradient) where ADMB could be implemented. An MCMC algorithm, developed within a Bayesian framework (DREAM), allows for the estimation of the probability distribution of the parameters and thus to identify such area.

The model was coded in C++ and integrated into the R environment (R Development Core Team 2015) using Rcpp (Eddelbuettel *et al.* 2011). We used the `dream` package (Guillaume & Andrews 2012) to run the MCMC algorithm and the solution with the largest likelihood provided by this package was then used as the starting point to ADMB through the `R2admb` package (Bolker, Skaug & Laake 2012). See Appendix S2, Supporting information for the R code used to run the inverse model with a given set of data and parameter intervals.

SIMULATIONS

We simulated a population subject to structured vital rates that remain constant over time. Starting with known values for the parameter esti-

mates (β_i's in eqn 2), we obtained their associated population dynamics (Fig. 2a, b, d, e, dark lines). We chose these values so that the population did not reach its stable state over 100 years (we are using 'year' just as shorthand for the arbitrary time units in the model). Then, we simulated a population that, starting with 10 000 individuals, followed the dynamics given by the vital rates over 100 years (Fig. 2c). We used as the starting population structure the one obtained after iterating eqn 1 once from a uniform distribution of sizes. The time series of population sizes and structures produced by this population was used as data for the inverse model (Fig. 2c, f; see Appendix S1, Supporting information for details). To evaluate model accuracy, we calculated confidence bands for the mean vital rates using the variance–covariance matrix provided by ADMB.

We ran all possible scenarios where one, two, three or all of the vital rates are unknown. We did this by fixing the corresponding parameters at the known values. We also explored the sensitivity of parameter estimation to both the number of individuals measured each year (50%, 90% and 99% reduction in initial sample size), and the number of years for which data were gathered (50, 20, 10 and 5 years).

REAL DATA

To exemplify the inverse model with real data, we used the one provided by Salguero-Gómez *et al.* (2014). This is a 15-year-long annual survey of a population of the chamaephyte *Cryptantha flava* L. in Uintah County, Utah, USA (40° 35′ 42·63″ N, 109°25′ 55·92″ W, 1790 m a.s.l.). Although the data base includes information on an experiment of artificial manipulation of precipitation, we only used the control plots (Fig. 3a, c, e, g, in grey).

As suggested by Ghosh, Gelfand & Clark (2012a), we kept the model as simple as possible to avoid problems of identifiability by eliminating the third parameter of the survival function. Thus, the model used has the structure:

$$s(x) = \text{logistic}(\beta_0 + \beta_1 \cdot x), \qquad \text{eqn 4}$$

$$g(y, x) = \text{Normal}(\mu = \beta_2 + \beta_3 \cdot x, \sigma = \exp(\beta_4 \cdot \ln(10))),$$

$$f_1(x) = \exp(\beta_5 + \beta_6 \cdot x),$$

$$f_2(y) = \text{Negative binomial}(\theta = \exp(\beta_7 \cdot \ln(10)), p = \text{logistic}(\beta_8)).$$

Because individual size is measured as a discrete variable (the number of rosettes in an individual), we changed the distribution of the newborns from Gaussian to negative binomial.

Results

The inverse procedure was able to reconstruct unobserved vital rates using only population-level data in a simulated scenario. When additional information on the rates was supplied (which is known exactly in this case), better estimates were obtained, especially when information on individual variation in growth was supplied. Also, we show that the model works with very small per-year sample sizes, but that short time series make the model non-identifiable. Finally, in the case of *C. flava*, the high between-

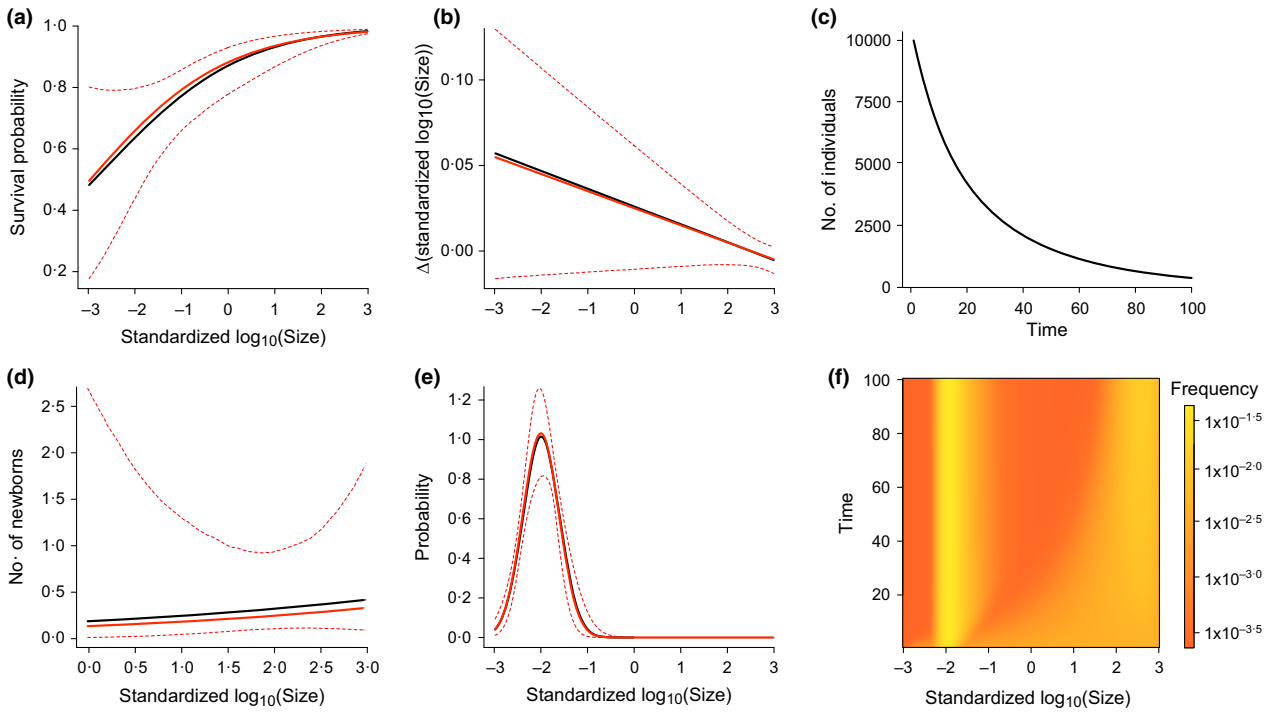

Fig. 2. Initial and reconstructed vital rates (a, b, d, e) and simulated population-level data (c, f). The population followed constant structured vital rates through time (black lines; a: survival, b: growth, d: fecundity, e: newborn size distribution), which produced, over a 100 years, a time series of population sizes (c) and structures (f). The reconstructed structured vital rates when using only population-level data are in solid red (Wald confidence bands in dashed). Reproductive minimum size was set at 0.

year variation in survival probability required the inclusion of individual growth estimates to provide sensible reconstructions for the remaining vital rates.

SIMULATED DATA

Using the sequential (DREAM + ADMB) optimization approach with the simulated data, the model was able to estimate the parameter values used to generate them (Fig. 2, black vs. red lines). However, as the confidence bands show (Fig. 2; dashed red lines), some parameters are easier to estimate than others (Appendix S4, Supporting information). Narrow confidence intervals were obtained for the parameters associated with the distribution of newborn sizes (β_8 and β_9), the mean of the distribution of adult sizes (β_3 and β_4), and the intercept and linear term of the survival function (β_1 and β_2). Fecundity was harder to estimate, with wide confidence intervals for the slope parameter (β_7). The same problem was applied to the quadratic term in the survival function (β_2). This difficulty in estimation is due to a problem of parameter identifiability that stems from the type of data we are providing to the model: aggregate information.

SCENARIOS WITH SOME ESTIMATE ON THE VITAL RATES

Additional data on the vital rates increased or decreased parameter accuracy depending on the parameters been estimated (Fig. 4a; Appendix S3, Supporting information); however, identifiability problems remained present. As expected, the more information was provided on the vital rates, the quicker the MCMC chains in the differential evolution algorithm covered the likelihood surface (results not shown here). This is also reflected in the reduced width of the HPD confidence intervals as more information is provided (Appendix S3, Supporting information). However, the problem of identifiability of the quadratic term in the survival function (β_2) is evident from the distance to the true values, because this presents its highest values when survival is estimated (Fig. 4a, all points where s is unknown).

SCENARIOS OF REDUCED INFORMATION ON THE TIME SERIES

Reducing the years over which the population is followed had a larger impact on accurate parameter estimation than per-year sample size in the scenario of no information on the vital rates (Fig. 4b). A 100-year time series, with 100 individual measurements at the initial time, was enough to correctly estimate the vital rates. On the other hand, a 5-year time series did not provide enough information on the vital rates to estimate them, even when the population started having a large population size (10 000 individuals) and the starting point of the ADMB optimization algorithm was the vector of parameter values from which the data were simulated. In these cases, the model reached the upper bound on β_6, making large individuals unrealistically fecund.

A 10-year time series pushed the limits of our approach (Fig. 4b). When size data were available for a large population (10 000 individuals), the model correctly reconstructed the vital rates. As expected, as sample size reduced, distance to the parameters used to generate the data increased, but only after a large reduction. This is also reflected in the widening of the confidence intervals for the parameter estimates under the different scenarios (Appendix S4, Supporting information).

REAL DATA

Using the approach to reconstruct the vital rates of the *C. flava* population proved challenging. Using only population-level data, the algorithm incorrectly reconstructed the shape of the survival function, as the confidence intervals associated with the observed survival did not include the reconstructed one (results not shown here). Under this scenario, survival of the newborns was overestimated and fecundity was underestimated.

As reported by Lucas, Forseth & Casper (2008), there is high interannual variation in the population dynamics of this species. Therefore, an IPM whose kernel remains constant through time was a poor approximation of the system. This suggested that random effects for interannual variation were required. This approach was explored using ADMB-RE, the version of ADMB that allows including random effects, but a problem of parameter identifiability emerged (i.e. flat regions were attained even when the algorithm starting point was the set of parameter values estimated using the individual-level data). Correct reconstructions of the vital rates were obtained only after assuming known individual growth rate and interannual variation on survival (Fig. 3a, c, d), with large residual outliers obtained for the newborn size distribution (Fig. 3f). Important to note though is that, by construction, the model correctly estimated the population-level data (e.g. population size: Fig. 3g).

Discussion

Integral projection models are widely applied (e.g. the examples presented in this Special Feature). However, they rely on high-quality individual-level data that are costly both in time and money. Here, we used population-level information, which aggregates the individual-level data, to reconstruct the vital rates that produced them. In both cases, an underlying model describes the relation between the individual- and the population-level data: an IPM. As our results show, both with simulated and real data, reconstructing the vital rates from summary information is a difficult endeavour and depends largely on having an IPM that correctly describes the population dynamics.

With simulated data, the model performed well, providing good estimates for the underlying vital rates. This is because the underlying IPM correctly (by construction) described the population dynamics. This demonstrates that vital rates can be reconstructed from population-level data.

We also showed that when estimates for some vital rate(s) are provided, parameter estimation is improved on the missing

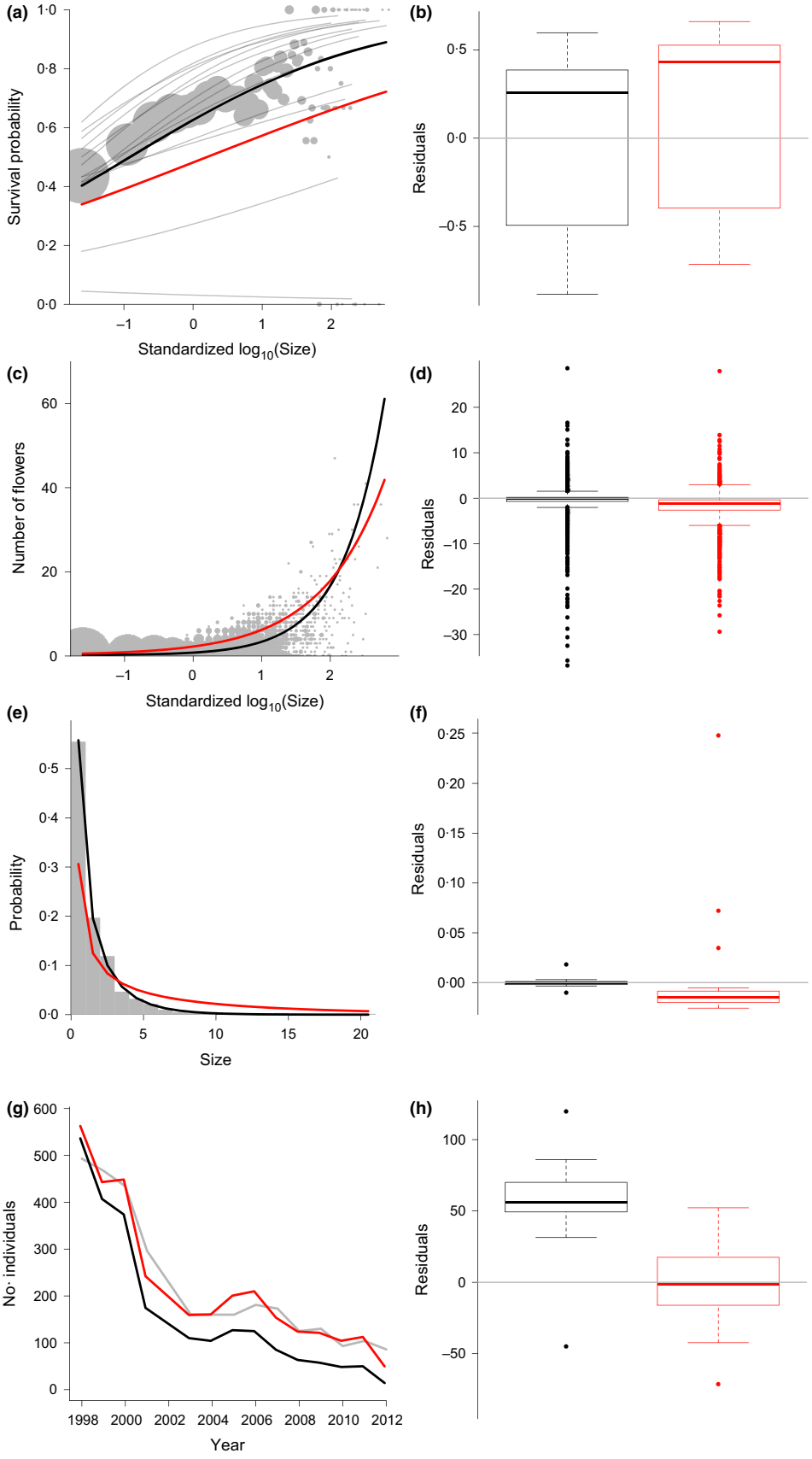

Fig. 3. Vital rates (a, c, e) and population sizes (g) estimated using individual- (black) or population-level (red) data on *Cryptantha flava* (grey). The vital rates were obtained through the inverse model approach assuming known individual growth and between-year survival variation (grey lines in a). Corresponding residuals are in b, d, f and h.

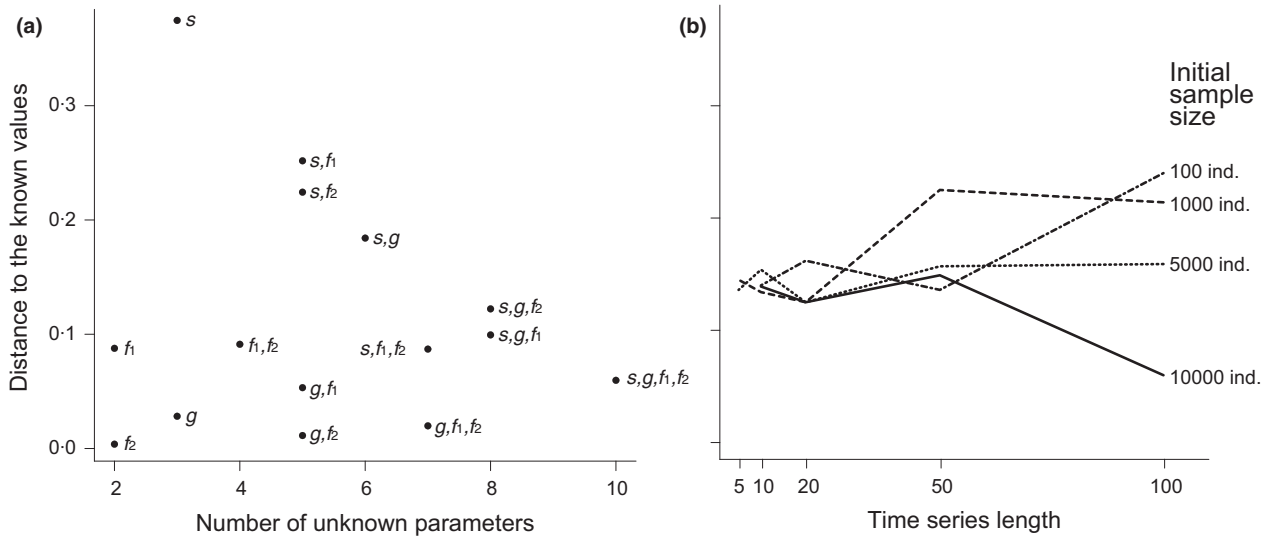

Fig. 4. Procedure performance under scenarios of vital-rate estimates availability (a), time-series length and sample size (b) in terms of the distance between the parameter values used to generate the simulated data and those estimated. In a, all scenarios included population-level data but different number of parameters were unknown (with the associated unknown vital rates indicated). In b, the number of individuals next to each line corresponds to the sample size at time 0 (i.e. Fig. 2c, time = 0). See Appendix S1, in the Supporting information for details on the distance used.

vital rates. Additionally, by changing the length of the time series and the per-year sample size, we found that time-series length is more important than the number of organisms measured each year for accurate parameter estimation (See & Holmes 2015). The method performs well with limited per-year sample size, an attribute probably inherited from the fact that IPMs perform well with limited data (Ramula, Rees & Buckley 2009).

However, how much a time series changes over time is crucial. In our case, a 5-year time series was too short to correctly estimate parameters. This was probably due to the manner in which the parameter values were selected: we chose them so that the initial population structure would not stabilize after 100 years. Such population would experience very small changes from 1 year to the next (Fig. 2f), and thus a 5-year time series would not be informative enough. In such cases, the optimization process is only minimizing sampling error, which leads to incorrect reconstructions as no variation due to differences in the vital rates among sizes is being provided (González & Martorell 2013). Therefore, it would be expected that when the length of the time series is short, only those populations exhibiting large differences in the vital rates between individuals with disparate traits will be accurately reconstructed by the model (Doak, Gross & Morris 2005; See & Holmes 2015).

When real data were used, an IPM with constant underlying vital rates constituted a poor approximation to the system. Good estimates were only obtained when providing estimates on between-year variation in survival and on individual growth. This makes sense because variation in survival was the vital rates displaying the largest between-year variation in its estimates.

High interannual variation in the vital rates pointed toward the inclusion of random effects (Lucas, Forseth & Casper 2008). However, including random effects on parameters is not directly estimated, but only through a nonlinear model (as in our case) is a challenging task (Wakefield 2008). ADMB-RE was considered, but its performance was not adequate as flat regions were reached irrespective of IPM structure. This suggested a problem of non-identifiability. However, model structure could also be the problem. Ghosh, Gelfand & Clark (2012a) suggest the use of simple model structures and that was the reason for our decision to keep model structure simple; certainly, a balance between simplicity and identifiability might had been attained using a model with a few more parameters.

From the simulated and real data, our results suggest that for populations that do not display directional changes in the population structure over a short time span or display highly stochastic changes, inverse modelling will not provide good reconstructions of the vital rates. Under those scenarios, model parameters become non-identifiable, that is multiple combinations of parameter values produce the same population-level data, and the data do not provide enough information to tell those combinations apart. In such cases, additional data (or sensible estimates) on some vital rate will most likely be needed.

AVOIDING UNREALISTIC RECONSTRUCTIONS

When no information is provided to the inverse procedure, two possible scenarios are always lurking: given a population structure, individuals could be all pre-existing or all newborns. In terms of the vital rates, the first scenario would result from near-one survival probability for certain individuals coupled with near-zero fecundity in the entire population. This scenario entails immortality and, therefore, makes no biological sense. The second scenario would result from near-zero survival

probability for pre-existing individuals and high fecundity rates for at least some individuals. This scenario would only make sense if the species were known (or believed) to be an annual.

By imposing bounds (as done by González & Martorell 2013) or priors on the parameters (or combinations of them), these scenarios can be avoided. However, we did not impose informative priors in our implementation of the inverse procedure, as we wanted to evaluate its ability to correctly reconstruct the vital rates in the absence of any information. Certainly, a Bayesian approach will be more appropriate if the researcher possesses such information (Stuart 2010) and it will simplify model analysis, as the maximum likelihood is no longer relevant.

POPULATION-LEVEL MODELS

Fitting observed population structures to an IPM structure has been seen as a way to use population-level information to infer population-level vital rates. The idea behind this point of view is that individual-level information gathered through the follow-up of the life cycle of individuals in a population does not necessarily capture the processes that occur at the population level. To remedy this, population structures have been seen as a better source of information from which to infer the vital rates (Ghosh, Gelfand & Clark 2012a; Gelfand, Ghosh & Clark 2013). We recognize that this approach converges to a similar statistical method to the one we use to infer the vital rates.

However, conclusions drawn from such approach state that vital rates obtained through a population-level model do not necessarily have to match those obtained from individual-level data (Ghosh, Gelfand & Clark 2012b). As we discuss in the next section, we believe that if the IPM correctly represents a population's dynamics, then such match should occur.

INVERSE IPM AS A MODEL DIAGNOSTICS OF A FORWARD IPM

So far, IPMs have been considered as adequately describing population dynamics if the population sizes they project correctly match with the observed ones. This is a diagnosis of the population-level output. However, this procedure does not serve as a means to evaluate the mechanistic description of the population dynamics, that is it does not evaluate the kernel structure, only its output.

To evaluate the kernel structure of an IPM, the vital rates reconstructed by inverse modelling using population-level data should match those obtained through the use of individual-level data, that is residuals should look randomly placed around the line of zero deviation (e.g. Fig. 3b, d, f). This is because an IPM is only a mechanistic model connecting the individual patterns with the population ones. Therefore, if a mismatch exists between the vital rates obtained using forward and inverse modelling, it should be considered as a problem associated with the particular IPM structure used to describe the system.

If we are to provide models that correctly describe a system at both population and individual levels, it is important to have models that use data from both levels. This calls for an integrated approach.

TOWARDS INTEGRATED IPMS

The fact that inverse modelling performed poorly with real data is counterbalanced by the fact that, by construction, the inverse model matches the observed population sizes (Fig. 3g, h). A recurrent problem with traditional MPMs and IPMs is the lack of fit between the observed and estimated population sizes (Williams *et al.* 2011; Jäkäläniemi, Postila & Tuomi 2013; Merow *et al.* 2014). Inverse modelling, coupled with its forward version, represents an opportunity to explore models whose estimates will fit data at both individual and population levels.

This has already been done in the case of animal species whose populations are structured by discrete variables; they are known as integrated population models (Fournier & Archibald 1982; Schaub *et al.* 2007; Schaub & Abadi 2011; Maunder & Punt 2013). Although Schaub & Abadi (2011) clearly pointed towards the application of integrated population modelling to IPMs, no such application has, to our knowledge, been presented yet. However, the idea of using multiple sources of information to inform the estimation of vital rates is not new (Lee 1974, 1985; Wrigley & Schofield 1981; Oeppen 1993; Raftery, Givens & Zeh 1995), and the use of multilevel information is constant in multilevel regression (Gelman & Hill 2007). Nonetheless, within an integrated population-modelling framework, we consider that the scenarios we explored, where estimates on some vital rate are provided, can be seen as a first attempt to perform integrated IPMs.

Estimates provided by an integrated IPM are no substitute to sound modelling practices. IPMs have been used under two assumptions: (i) the model correctly describes population dynamics and (ii) its estimates are accurately estimated. As we discussed above, inverse modelling can help to diagnose model structure. An integrated IPM would address the latter; it provides better estimates because it uses a piece of information that has not been used to inform parameter estimation: population-level data.

Even more, an integrated IPM, by fitting all vital rates at once (as does our approach), allows for the inclusion of aspects that have not been possible to capture so far in traditional forward IPMs: correlations between vital rates (Doak, Kareiva & Klepetka 1994). Given that in traditional IPMs, the data associated with each vital rate are used separately to obtain models that describe the mean vital rate as a function of the structuring variable, the kernel cannot be informed of such correlations. Therefore, an integrated IPM would also allow estimating such correlations and their change along the structuring variable. This will further improve the explanatory power of IPMs.

Acknowledgements

EJG thanks Steve Walker and Guillaume Blanchet for the fruitful discussions on the many statistical elements associated to this problem; Florian Hartig, Sean McMahon and an anonymous reviewer for their valuable comments on improving a previous version of the manuscript and the algorithm used to perform parameter estimation; and Roberto Salguero-Gómez for providing the *C. flava* data base. Finally, we thank the organizers of this Special Feature for inviting our contribution. Funding for this research was provided to EJG by the Mexican Consejo Nacional de Ciencia y Tecnología (CONACYT) and McMaster University through a postdoctoral grant and to BMB through an NSERC Discovery grant.

Data accessibility

R scripts uploaded as online supporting information.

References

Bolker, B., Skaug, H. & Laake, J. (2012) *R2admb: ADMB to R interface functions*. R package version 0.7.9. http://CRAN.R-project.org/package = R2admb

Brooks, M.E., McCoy, M.W. & Bolker, B.M. (2013) A method for detecting positive growth autocorrelation without marking individuals. *PLoS ONE*, **8**, e76389.

Crouse, D.T., Crowder, L.B. & Caswell, H. (1987) A stage-based population model for loggerhead sea turtles and implications for conservation. *Ecology*, **68**, 1412–1423.

Doak, D.F., Gross, K. & Morris, W.F. (2005) Understanding and predicting the effects of sparse data on demographic analyses. *Ecology*, **86**, 1154–1163.

Doak, D.F., Kareiva, P. & Klepetka, B. (1994) Modeling population viability for the desert tortoise in the western Mojave Desert. *Ecological Applications*, **4**, 446–460.

Easterling, M.R., Ellner, S.P. & Dixon, P.M. (2000) Size-specific sensitivity: applying a new structured population model. *Ecology*, **81**, 694–708.

Eddelbuettel, D., François, R., Allaire, J., Chambers, J., Bates, D. & Ushey, K. (2011) Rcpp: seamless R and C++ integration. *Journal of Statistical Software*, **40**, 1–18.

Fournier, D.A. & Archibald, C.P. (1982) A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences*, **39**, 1195–1207.

Fournier, D.A., Hampton, J. & Sibert, J.R. (1998) MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 2105–2116.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. & Sibert, J. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, **27**, 233–249.

Gelfand, A.E., Ghosh, S. & Clark, J.S. (2013) Scaling integral projection models for analyzing size demography. *Statistical Science*, **28**, 641–658.

Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.

Ghosh, S., Gelfand, A.E. & Clark, J.S. (2012a) Inference for size demography from point pattern data using integral projection models. *Journal of Agricultural, Biological, and Environmental Statistics*, **17**, 678–681.

Ghosh, S., Gelfand, A.E. & Clark, J.S. (2012b) Rejoinder: inference for size demography from point pattern data using integral projection models. *Journal of Agricultural, Biological, and Environmental Statistics*, **17**, 693–699.

González, E.J. & Martorell, C. (2013) Reconstructing shifts in vital rates driven by long-term environmental change: a new demographic method based on readily available data. *Ecology and Evolution*, **3**, 2273–2284.

Guillaume, J. & Andrews, F. (2012) dream: DiffeRential Evolution Adaptive Metropolis. R package version 0.4-2.

Jäkäläniemi, A., Postila, H. & Tuomi, J. (2013) Accuracy of short-term demographic data in projecting long-term fate of populations. *Conservation Biology*, **27**, 552–559.

Lee, R.D. (1974) Estimating series of vital rates and age structures from Baptisms and burials: a new technique, with applications to pre-industrial England. *Population Studies: A Journal of Demography*, **28**, 495–512.

Lee, R.D. (1985) Inverse projection and back projection: a critical appraisal, and comparative results for England, 1539 to 1871. *Population Studies*, **39**, 233–248.

Lucas, R.W., Forseth, I.N. & Casper, B.B. (2008) Using rainout shelters to evaluate climate change effects on the demography of *Cryptantha flava*. *Journal of Ecology*, **96**, 514–522.

Maunder, M.N. & Punt, A.E. (2013) A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, **142**, 61–74.

May, R.M. (2004) Ecology: ethics and amphibians. *Nature*, **431**, 403.

Merow, C., Dahlgren, J.P., Metcalf, C.J.E., Childs, D.Z., Evans, M.E.K., Jongejans, E. *et al.* (2014) Advancing population ecology with integral projection models: a practical guide. *Methods in Ecology and Evolution*, **5**, 99–110.

Metcalf, C.J.E., McMahon, S.M., Salguero-Gómez, R. & Jongejans, E. (2013) IPMpack: an R package for integral projection models. *Methods in Ecology and Evolution*, **4**, 195–200.

Moran, E.V., Hartig, F. & Bell, D.M. (2015) Intraspecific trait variation across scales: implications for understanding global change responses. *Global Change Biology*, **22**, 137–150.

Oeppen, J. (1993) Back projection and inverse projection: members of a wider class of constrained projection models. *Population Studies*, **47**, 245–267.

R Development Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org [accessed 15 April 2015]

Raftery, A.E., Givens, G.H. & Zeh, J.E. (1995) Inference from a deterministic population dynamics model for bowhead whales. *Journal of the American Statistical Association*, **90**, 402–416.

Ramula, S., Rees, M. & Buckley, Y.M. (2009) Integral projection models perform better for small demographic data sets than matrix population models: a case study of two perennial herbs. *Journal of Applied Ecology*, **46**, 1048–1053.

Rees, M., Childs, D.Z. & Ellner, S.P. (2014) Building integral projection models: a user's guide. *Journal of Animal Ecology*, **83**, 528–545.

Robinson, W. (2009) Ecological correlations and the behavior of individuals. *International Journal of Epidemiology*, **38**, 337–341.

Salguero-Gómez, R., Kempenich, H., Forseth, I.N. & Casper, B.B. (2014) Long-term individual-level population dynamics of a native desert chamaephyte. *Ecology*, **95**, 577.

Schaub, M. & Abadi, F. (2011) Integrated population models: a novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology*, **152**, 227–237.

Schaub, M., Gimenez, O., Sierro, A. & Arlettaz, R. (2007) Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology*, **21**, 945–955.

See, K.E. & Holmes, E.E. (2015) Reducing bias and improving precision in species extinction forecasts. *Ecological Applications*, **25**, 1157–1165.

Silvertown, J., Franco, M. & Menges, E. (1996) Interpretation of elasticity matrices as an aid to the management of plant populations for conservation. *Conservation Biology*, **10**, 591–597.

Stuart, A.M. (2010) Inverse problems: a Bayesian perspective. *Acta Numerica*, **19**, 451–559.

Tuljapurkar, S. & Caswell, H. (1997) *Structured-population Models in Marine, Terrestrial, and Freshwater Systems*. Vol. 18. Chapman & Hall, New York.

Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M. & Higdon, D. (2009) Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, **10**, 273–290.

Wakefield, J. (2008) Multi-level modelling, the ecologic fallacy, and hybrid study designs. *International Journal of Epidemiology*, **38**, 330–336.

Williams, J.L., Ellis, M.M., Bricker, M.C., Brodie, J.F. & Parsons, E.W. (2011) Distance to stable stage distribution in plant populations and implications for near-term population projections. *Journal of Ecology*, **99**, 1171–1178.

Wood, S.N. (1994) Obtaining birth and mortality patterns from structured population trajectories. *Ecological Monographs*, **64**, 23–44.

Wrigley, E.A. & Schofield, R.S. (1981) *The Population History of England, 1541–1871: A Reconstruction*. Harvard University Press, Cambridge, MA.

Received 9 September 2015; accepted 23 November 2015
Handling Editor: Sean McMahon

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Full description of the model and simulation procedures.

Appendix S2. Implementation of the inverse procedure (invIPM.R) given a time series of population structures and sizes (invipmADMB.dat) and a set of parameter intervals (intervals.csv).

Appendix S3. Estimates, high posterior density intervals and distances for the model parameters under the 15 scenarios of data availability on the vital rates, and the nine scenarios of data availability on the time-series length and the per-year sample size.

Appendix S4. Results associated with the differential evolution of 30 MCMC chains used to estimate initial parameter values for the gradient-based algorithm of the inverse model in the case where only simulated population-level data are available.