**Title:** Inferring structured vital rates from a time series of population sizes and structures: an inverse problem

**Running title:** Inferring structured vital rates

**Word count:**

**Authors:** Edgar J. González\*,1, Carlos Martorell2, Benjamin M. Bolker1,3

**Addresses:**

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

3: Department of Biology, McMaster University, 1280 Main St W, L8S4L8, Hamilton, Ontario, Canada

**Corresponding author:**

***Name*:** Edgar J. González

***Address*:**Department of Mathematics and Statistics, McMaster University, 1280 Main St W, L8S4L8, Hamilton, Ontario, Canada

***E-mail*:**gonzaleze@math.mcmaster.ca

**Abstract**

1. Traditional demographic methods have relied on tracking marked individuals in a population over time to estimate its vital rates. In a structured population these vital rates change as a function of a variable, such as size, that explains most of the among-individual demographic variation. Projection models integrate these structured vital rates, generating a time series of population sizes and structures.

2. Here, we present an inverse problem: inferring the structured vital rates from the time series. We simulated different population dynamics to show that this approach works for the case where the vital rates are structured by a continuous variable, remain constant through time, and the population is not in its stable state. We also explore cases where partial information on the vital rates is available, and apply the method to data from a plant system.

3. We show that this approach can provide accurate reconstructions of the vital rates. The more information on the vital rates is provided the easier it is to find a correct reconstruction.

4. Given the type of input that the method uses, it can be a useful tool for those systems where tracking has not been performed, or is costly or impractical.

**Key-words**

inverse model, population dynamics, population distribution, population density, integral projection model, integrated population model, demography

**Introduction**

Population ecology is a time and money demanding discipline. The information required for the understanding of the dynamics of a population over time, particularly in long-lived organisms, requires the follow up of individuals for a sensible timespan to capture data on the birth, survival, and reproduction rates of its members. Depending on the species and system under study, some of these rates will be harder to measure than others, and cases can exist where no information on any rate is available.

Additionally, individuals differ because they display different traits (e.g., allele, size, weight, sex) or experience different environments (e.g, less/more predation, light intensity, rainfall, disturbance, etc.). By increasing variation among its individuals, we say that these variables *structure* the population, and thus that vital rates are structured (Tuljapurkar & Caswell, 1997; Merow et al. 2014). Although important to understanding the dynamics of a population, accounting for the effect of these variables generally increases costs.

Integrated population models (traditionally called IPMs, but here IPopMs) have been proposed as a single statistical framework to study the dynamics of populations. In IPopMs, data from different sources can be used together and in this way better inform unknown estimates on demographic parameters (Raftery, Givens & Zeh, 1995; Besbeas et al. 2002; Brooks et al. 2004; Schaub & Abadi, 2011; Kéry & Schaub 2012; Chandler & Clark, 2014). Therefore, this framework allows estimating partial information on vital rates to infer those rates that are harder to estimate.

Previous work with IPopMs has focused on scenarios where only partial information on the vital rates is available, usually working with vertebrate species (e.g. Schaub et al. 2007; Kéry & Schaub 2012; Maunder & Punt 2013; but cf. McMahon & Parker 2014). In addition to research explicitly stated as IPopMs, many authors have tried to infer unobserved vital rates from observed ones (REFS, McMahon & Parker 2013). This method can be traced back to inverse/backward projection models in human populations where the aim is to infer mortality and fertility rates, along with population age structure, from annual birth and death records (Lee 1974, 1985; Wrigley & Schofield 1981; Oeppen 1993).

When at least one of the variables that structure a population is continuous, an integral projection model (traditionally called IPM, but here IProjM) is the standard tool to analyse how structured vital rates interact. In an IProjM the user provides information on the vital rates, fits a function to each of these rates and obtains a time series of population structures describing the change of the population over time. Traditionally, the functional forms used to describe these rates have depended on the particular system under study as well as on the type of data available, but the general trend has been to fit some kind of linear model (generalized, additive, mixed, etc.) to the data on each vital rate separately (Metcalf et al. 2013; Merow et al. 2014). Using linear models allows for a relatively small number of parameters to describe mathematically each vital rate, greatly reducing parameter estimation problems when compared with matrix projection models (Easterling, Ellner & Dixon, 2000), while allowing traditional statistical hypothesis testing (Coulson 2012), and model selection to be performed at the level of each vital rate (Metcalf et al. 2013; Merow et al. 2014).

The reduced complexity of IProjMs, coupled with the statistical framework provided by IPopMs, allows for the study of scenarios where very few data are available on some vital rates, up to the point where none is available other than the outcome of their interplay: population size and structure. This last scenario is an inverse problem.

In a general sense, an inverse problem involves taking the results of a series of processes and attempting to estimate these processes. Inverse problems have appeared in many areas of research including astronomy, astrophysics, remote sensing and hydrology (Ambartsumian 1980; Natterer 1986; Brown 1995; Carrera et al. 2005). As can be deduced from their name, every inverse problem must strictly relate to a direct problem. In our case, the direct problem would be to relate the vital rates in a mechanistic way, so that the outcome of such relation are the population sizes and structures observed over time. This is what an IProjM does. Using matrix projection models, Fournier, Hampton & Sibert (1998) showed that a inverse problem approach works in the context of fisheries species, while Ghosh, Gelfand & Clark (2012) have used IProjMs, using population structures as input to a similar problem.

One general change of inverse problems is that, because they are less constrained than traditional statistical modelling approaches, they can easily incorporate large amounts of biological complexity, possibly outrunning the capability of the data to identify parameters. As the capabilities of estimation using complex ecological models has exploded in the past few decades, ecologists have had to confront issues of parameter identifiability and associated estimation problems such as the existence of multiple optima (qualitatively different sets of parameter estimates that each fit the data better than any parameters in their respective neighbourhoods). Despite the repeated cautions from sophisticated modelers on these topics, there are still few concrete, worked examples showing general ecologists how they can diagnose and mitigate such problems (e.g. by simplifying or constraining models) when they arise.

In this paper, we explore the inverse problem of inferring some or all of the structured vital rates associated to a population for which a time series of population sizes and structures is available. We use simulated data to show that how this approach works for the case where an IProjM kernel is constant through time, i.e. the population dynamics remains constant, but the population structure has not reached its asymptotic state. As would be expected with any statistical problem, we show that the less information on the vital rates is available, the harder it is to find the correct reconstruction under a likelihood approach. Also, the less information you have (in terms of sample size and length of the time series), the harder it is to provide reliable reconstructions. Finally, we discuss the problems we've been face with when in trying to identify accurate estimates for the parameters of this challenging inverse problem.

**Materials and methods**

***Model***

Easterling, Ellner & Dixon (2000) presented the first version of an IProjM. It this paper, a single variable structures the population, size (*x*), and the vital rates are assumed to remain constant over time. When the continuous variable *x* that structures a population is size, the following equation relates the population vital rates with the population structure (*n*) at any point in time:

*nt*+1(*y*) = ∫[*s*(*x*)⋅*g*(*y*,*x*) + *f*1(*x*)⋅*f*2(*y*,*x*)]⋅*nt*(*x*)*dx,* eqn. 1

where *nt*(*x*) is the proportion of individuals of size *x* at time *t*; *nt*+1(*y*), the proportion of individuals of size *y* at time *t*+1; and the vital rates are: *s*(*x*), the survival probability of extant *x-*size individuals; *g*(*y*,*x*), the probability an individual has of changing from size *x* to *y* from one unit time 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 IProjM relates the vital rates with a time series of population structures (*n*1, *n*2, ...), and the time series of population sizes given by the integration of each population structure over the observed size range (*N*1 *=* ∫ *n*1(*x*)*dx*, *N*2 *=* ∫ *n*2(*x*)*dx*, ...). The traditional approach has been to use as input to the model the former. Here, we use the latter as input, thus posing an inverse problem.

Many alternative structures to describe which and how vital rates interact to produce the size structure exist. As a case study, we used an IProjM with the following structure, close to Easterling, Ellner & Dixon (2000) original model:

*s*(*x*) = logit(*β*0 + *β*1⋅*x + β*2⋅*x*2), eqns. 2

*g*(*y*,*x*) ~ Normal(*μ* = *β*3 + *β*4⋅*x*, *σ* = exp(*β*5)),

*f*1(*x*) = exp(*β*6 + *β*7⋅*x*),

*f*2(*y*) ~ Normal(*μ* = *β*8, *σ* = exp(*β*9)).

Note that the vital rates are determined by the value of 10 low-level parameters (*β*0, ..., *β*9), and thus that the values of these parameters are sufficient to model the vital rates. Thus the inverse problem consists of estimating these parameters using as data a time series of observed population sizes and structures.

Additionally, we explored simpler scenarios where information on some vital rates is available. In many cases, population ecologists have limited information on the vital rates and they would like to infer the vital rates that are missing. Therefore, we used as input to the model those observed vital rates and a time series of population sizes and structures. We explored all possible scenarios where one, two or three vital rates are unknown, with the inverse problem being the limiting scenario where no information on the vital rates is available.

***Parameter estimation***

To assess whether any given set of low-level IProjM parameter values is able to reproduce the observed time series, we substituted these in eqns 2. We then calculated the time series of population structures through the iteration of equation 1, and the time series of population sizes by integrating the structures. We used as initial population structure, *n*0(*x*), the first observed size structure.

The observed and estimated time series were compared through a composite log-likelihood function. Since we have two log-likelihood functions (see Appendix 1 for details): one that measures the goodness of fit between the estimated and observed population structures, *ln*, and another that measures the goodness of fit between the estimated and observed population sizes, *lN*, we measured the overall goodness of fit through the composite log-likelihood function:

*l* = *w*⋅*ln* + (1 – *w*)⋅*lN*, eqn. 3

where *w* is a weighting factor of the relative importance of fitting the former vs. the latter. Since we did not want to favor one fit over the other, we set *w* = *T*/(*T +* Σ*Ni*). By doing this, we account for the fact that *T* observed population sizes relate to Σ*Ni* measured individuals.

We used a gradient approach for the maximization of *l*. This approach used two optimization algorithms, one heuristic and one gradient-based. The heuristic part consisted of a generalized simulated annealing algorithm (GenSA; Xiang & Gong 2000) to coarsely explore the likelihood function within a large hypercube of possible parameter values. This large hypercube was selected to incorporate as many conceivable population dynamics as possible, becoming in practical terms an unbounded problem. Based on previous work, we knew that the likelihood function presented regions where the log-likelihood could took infinite values (see Fig. X), and thus a gradient-based algorithm would not work. However, we wanted to keep the advantage of such an approach, since GenSA has the caveat that it is time consuming to explore a local basin. Therefore, once a local maximum was identified, we used a gradient-based algorithm (ADMB; Fournier et al. 2012) to further improve parameter estimation within a small basin.

This hybrid approach highly increases the probability that the global maximum is found, although it may take a lot of time to find it (e.g., starting at a random point in this large hypercube, it took up to 6.3×107 evaluations of the model, ~15 h, to find a known global maximum). To reach the correct solution within a reasonable time frame (< 1 h), we ran the model in parallel with 100 starting points (selected with a Latin hypercube sampler) in this large hypercube and restricted the GenSA part of the optimization algorithm to run up to 0.75 h.

The model was coded in C++ and integrated into the R environment (R Core Team 2014) using Rcpp (Eddelbuettel et al. 2011). We used the GenSA package (Xiang et al. 2013) to run the heuristic part of the optimization procedure and the solution provided by this package was then used as input to ADMB through the R2admb package (Bolker, Skaug & Laake 2012). We used the lhs package (Carnell 2012) to obtain the starting points.

***Simulations***

We simulated a population subject to structured vital rates constant over time. Starting with known values for the parameter estimates (*β*i's in eqns. 2; Fig. 1A), we simulated a population that, starting with 10,000 individuals, followed the dynamics given by the modelled vital rates (eqns. 2) over 100 years. We used as the starting population structure the one obtained after one iteration of eqn. 1 of a uniform distribution. We chose the population dynamics so that the population did not reach its stable state over the timespan under consideration. The time series of population sizes and structures produced each year by this population was used as input to the model (Fig. 1B). We ran the 15 possible scenarios where three, two, one or all of the vital rates are unknown, by having fixed the corresponding parameters at the known values. We also explored the impact on parameter estimation of reducing the number of individuals measured on each year, as well as the number of years for which data were gathered. To see how this affected parameter precision we estimated confidence intervals through likelihood profiling (*lprof* option in ADMB; Normal approximation).

***Real data***

***TO BE DONE IF TIME ALLOWS...***

**Results**

***TO BE WRITTEN...***

**It's easy to fit 1, 2, it gets harder for 3 and its very hard to fit 4 at once. The main problem is due to the fact that F1 and F2 are highly correlated (see PCA plot, traces in appendix and likelihood slices in appendix). Variance parameters are also hard to estimate.**

**Discussion**

As an IPopM, this model makes a number of assumptions that should be fulfilled to ensure an accurate and unbiased estimation of parameters (Schaub & Abadi, 2011).

We assume that a static IProjM is a correct model to describe the dynamics of the population for which data are available. This is usually not the case, as the environment is hardly constant through time. However, exploring the feasibility of reconstructing the vital rates using a simple IProjM, as well as establishing which vital rates are harder to estimate, serves as a starting point to explore more complex IProjM structures.

***Simultaneous model fitting***

In a traditional IProjM, vital rates are estimated by separately fitting generalized linear models (GLM) to individual data (Merrow et al. 2014). Our inverse model can be seen as estimating all vital rates at once using a single dataset. This has the advantage of providing a single measure of goodness-of-fit for the entire IProjM.

***Model optimization***

Efficient nonlinear optimization is a challenging task. Even with a limited number of parameters, finding the values that maximize the model likelihood is difficult. A gradient approach (e.g., automatic differentiation) is usually a good alternative when the likelihood function is differentiable everywhere and the Hessian is positive defined everywhere, i.e., when the likelihood function is smooth and has a slope ≠ 0 everywhere. In such cases, the approach is able to find a global maximum in a relative short time frame.

On the other hand, a heuristic approach, such as simulated annealing (Kirkpatrick et al. 1983; Bohachevsky et al. 1986; Suman & Kumar 2006), is more suited for problems where the likelihood function is not differentiable everywhere. The caveat of such approach is that it is time consuming. Therefore, the model is usually run with several starting points, taking as solution the one with the maximum likelihood; however, this does not guarantee that the global solution is found, and large time frames may be involved (Maunder & Punt, 2013).

With our particular problem (10 parameters in the version of the model presented here, but 18 in a more complex one), the likelihood function presents combinations of parameter values that produce infinite values for the likelihood, rendering, in practical terms, the function not differentiable for those values. Also, our likelihood function presents flat regions (slope = 0), making the Hessian non-positive defined. Therefore, the chances of finding a global maximum with a gradient approach were very low (1000 starting points may be required so that a single one reaches a known global maximum). This called for a different approach.

Our hybrid algorithm benefits from the advantages associated to each approach. By using GenSA to explore fast, but coarsely, a large area of possible parameter values for a reasonable time, we avoid those regions where a gradient approach is impractical. Once a smooth and non-zero-slope region is reached, ADMB quickly finds the local maximum. Running in parallel the model with 100 starting points for an hour greatly increased the changes of finding the global optimum.

Obviously, other heuristic optimization algorithms can be used in conjunction with a gradient algorithm. Better ones will probably be required for more complex IProjMs, for which the global maximum will be even harder to find. With a previous version of the model, particle swarm optimization was used with some level of success (González 2008), though the particular platform used at the time (Matlab) limited computational speed.

***Integrated integral projection models***

Although Schaub & Abadi (2011) clearly pointed towards the application of IPopMs to IProjMs, such work has, to our knowledge, not been performed yet. However, the idea of using multiple sources of information to inform the estimation of vital rates is not new. Partial information is a constant in population ecology, and IPopMs allows for the use of all possible sources of data into the estimation of the vital rates, which are the usually the harder part of any population study.

Here we explored the scenarios where one, two, or three vital rates were unknown, showing how our method performs with every possible scenario of restricted information on these rates. Results show that the estimation of one or two vital rates is straight forward, since a few starting points (and in most cases only one) will be enough to find the correct parameter values associated to the missing vital rates.

When three vital rates are unknown...

Finally, the limiting case, where no information on the vital rates is available, is, as discussed in the previous section, a challenging task.

***Population-level models***

Fitting observed population structures to an IProjM structure has been seen as a strategy to use population-level information to infer population-level vital rates. The idea behind this approach 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 where to infer the vital rates (Ghosh, Gelfand & Clark 2012; Gelfand, Ghosh & Clark 2013). This approach, presented by Ghosh, Gelfand & Clark (2012), converges to a similar statistical method to infer the vital rates.

Under such population-level conception, we believe that population size is also another piece of valuable information that would be useful to incorporate into the estimation of the parameters of an IProjM. The use of population size is also supported by similar work done with matrix projection models (Fournier, Hampton & Sibert 1998).

In our model, we incorporated both sources of information through a weighted likelihood... Looking at the population size time series can also be useful to identify density dependence, and to account for it in the IProjM.

***Limitations of the method***

As with all statistical models, confidence in the estimated parameter values depends on data availability. The smaller the data set, the more difficult will be to provide accurate reconstructions of the vital rates. Data availability relates to both the number of years for which population size and structure were recorded, as well as on the per-year number of individuals for which the structuring variable was measured. Here we show that the first seems to be more important for accurate reconstructions than the latter. This...

The structure of the IProjM kernel is also important. If the functions that constitute the kernel do not conform to how actually the population behaves, the reconstructed vital rates will not reflect how the population dynamics really is. This is a problem shared by all statistical models: if the model that one is proposing does not reflect the actual structure of the system, low confidence can be put on the conclusions we derive from it. By using different equations to describe the relation between the vital rates and the structuring variable, and performing traditional model selection/averaging could potentially overcome this problem. However, given the nonlinearity of the problem, the reconstructed vital rates produced by each model should be evaluated visually to confirm that the models under consideration are not associated to grossly dissimilar reconstructions.

***Future research***

IProjMs have mushroomed into a panoply of applications (e.g. the many examples in this special issue). However, following enough individuals over a sensible timespan is a demanding activity in both monetary and human resources. Our aim here is to provide a method that both makes use of all the available information on the population, much in tune with what IPopM seeks, taking this approach to its limit: no information on the vital rates.

In some cases, the global maximum found by the optimization algorithm may not be biologically realistic. Visual examination of the solutions to discard unrealistic ones is always a reasonable approach (González & Martorell 2013), but we believe that constraining the parameter space to only realistic solutions is a more promising alternative. This can be seen, in an optimization framework, as maximizing a multiobjective problem (Suman & Kumar 2005) and, in a Bayesian framework, as imposing weakly informative priors. This approach has already been explored using a gradient-based approach but such priors create flat areas, which in turn make the exploration of the likelihood surface difficult. Hopefully, the hybrid optimization presented here will not suffer from this problem. The question still remains how to do this without introducing considerable bias in parameter estimation, but it makes sense to use a Bayesian framework, as it has proved its usefulness with other inverse problems (Calvetti, Kaipio, Somersalo 2014).

**Acknowledgements**

**Data accessibility**

**References**

Ambartsumian, V.A. (1980). On some trends in the development of astrophysics. *Annual Review of Astronomy and Astrophysics*, **18**, 1-13.

Besbeas, P., Freeman, S.N., Morgan, B.J. & Catchpole, E.A. (2002). Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, **58**, 540–547.

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

Brown, J.C. (1995). Overview of topical issue on inverse problems in astronomy. *Inverse Problems*, **11**, 635.

Calvetti, D., Kaipio, J.P. & Somersalo, E. (2014). Inverse problems in the Bayesian framework. **30**, 110301.

Carnell, R. (2012). lhs: Latin Hypercube Samples. R package version 0.10. http://CRAN.R-project.org/package=lhs

Carrera, J., Alcolea, A.S., Medina, A., Hidalgo, J. & Slooten, L.J. (2005). Inverse problem in hydrogeology. *Hydrogeology Journal*, **13**, 206–222.

Coulson, T. (2012). Integral projections models, their construction and use in posing hypotheses in ecology. *Oikos*, **121**, 1337–1350.

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., 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.

Ghosh, S., Gelfand, A.E. & Clark, J.S. (2012). Inference for Size Demography From Point Pattern Data Using Integral Projection Models. *Journal of Agricultural, Biological, and Environmental Statistics*, **17**, 678–681.

González, E.J. (2008). *Modelación de la Dinámica Poblacional de Cactáceas Endémicas del Valle de Tehuacán a lo largo de un Gradiente de Disturbio Antropogénico Crónico.* Master's Thesis. Universidad Nacional Autónoma de México, México, D.F.

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.

Heisey, D.M., Osnas, E.E., Cross, P.C., Joly, D.O., Langenberg, J.A. & Miller, M.W. (2010). Linking process to pattern: estimating spatiotemporal dynamics of a wildlife epidemic from cross-sectional data. *Ecological Monographs*, **80**, 221–240.

Kéry, M. & Schaub, M. (2012). *Bayesian population analysis using WinBUGS: a hierarchical perspective*. Academic Press, Waltham, MA.

Lavine, M. (2010). Living dangerously with big fancy models. *Ecology*, **91**, 3487–3414.

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.

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

McMahon, S.M. & Parker, G.G. (2014). A general model of intra-annual tree growth using dendrometer bands. *Ecology and Evolution*, **5**, 243–254.

Merow, C., Dahlgren, J.P., Metcalf, C.J.E., Childs, D.Z., Evans, M.E.K., Jongejans, E., Record, S., Rees, M., Salguero-Gómez, R. & McMahon, S.M. (2013). 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. (2012). IPMpack: an R package for integral projection models (M. Rees, Ed.). *Methods in Ecology and Evolution*, **4**, 195–200.

Natterer, F. (1986). *The mathematics of computerized tomography*. Wiley, Chichester.

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.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmuller, U. & Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**, 1923–1929.

Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B.D., Theis, F.J., Klingmüller, U. & Timmer, J. (2013). Lessons Learned from Quantitative Dynamical Modeling in Systems Biology (E. Hernandez-Lemus, Ed.). *PLoS ONE*, **8**, e74335.

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.

Schaub, M. & Abadi, F. (2010). Integrated population models: a novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology*, **152**, 227–237.

Suman, B. & Kumar, P. (2005). A survey of simulated annealing as a tool for single and multiobjective optimization. *Journal of the Operational Research Society*, **57**, 1143–1160.

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

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

Xiang, Y. & Gong, X.G. (2000). Efficiency of generalized simulated annealing. *Physical Review E*, **62**, 4473–4476.

Xiang, Y., Gubian, S., Suomela, B. & Hoeng, J. (2013). Generalized simulated annealing for global optimization: the GenSA Package. *R Journal*, **5**, 13–28.

**Tables**

Table 1. Performance of the model under different scenarios of availability of information on vital rates. Success rate is the number of starting points, out of a 100, that reached the sought solution within 1 h. Vital rates: *s*, survival; *g*, growth; *f1*, fecundity; and *f*2, newborn sizes.

|  |  |  |
| --- | --- | --- |
| Unknown | Known | Success rate |
| *s* | *g, f*1, *f*2 | 100 |
| *g* | *s, f*1, *f*2 | 100 |
| *f*1 | *s*, *g, f*2 | 100 |
| *f2* | *s*, *g, f*1 | 100 |
| *s*, *g* | *f*1, *f*2 | 67 |
| *s, f*1 | *g, f*2 | 100 |
| *s, f*2 | *g, f*1 | 100 |
| *g, f*1 | *s, f*2 | 84 |
| *g, f*2 | *s, f*1 | 77 |
| *f*1, *f*2 | *s*, *g* | 100 |
| *s*, *g, f*1 | *f2* | 84 |
| *s*, *g, f*2 | *f*1 | 14 |
| *s, f*1, *f*2 | *g* | 13 |
| *g, f*1, *f*2 | *s* | 14 |
| *s, g, f*1, *f*2 | *–* | 5 |

Table 2. Estimates and confidence intervals obtained with the simulated data. Initial: values of the parameters from which the data were generated; Estimate: estimated values of the parameters using the model were all the vital rates were assumed unknown; Confidence interval: obtained through likelihood profiling.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Vital rate | Parameter | Initial | Estimate | Confidence interval |
| survival | *β0* | 1.9 | 1.884 | 1.343, 2.395 |
| *β1* | 0.4632 | 0.690 | 0.176, 1.184 |
| *β2* | 0.0046 | 0.010 | -0.292, 0.299 |
| growth | *β3* | 0.025 | 0.026 | -0.015, 0.065 |
| *β4* | 0.99 | 0.990 | 0.975, 1.003 |
| *β5* | -1.3 | -1.326 | -2.402, -0.293 |
| fecundity | *β6* | -2 | -1.708 | -4.754, 1.218 |
| *β7* | 0.3 | 0.278 | -1.061, 1.563 |
| newborn sizes | *β8* | -2 | -1.988 | -2.079, -1.901 |
| *β9* | -0.95 | -0.937 | -1.175, -0.690 |

Table 3. Performance of the model under different per-year sample sizes and number of years available. Here, all vital rates are estimated simultaneously.

***WORKING ON THIS...***

**Figures**

Fig. 1. Simulated data and reconstructed vital rates. The population followed constant vital rates through time (A), which produced, over a 100 years, a time series of population structures (B) and sizes (C). D: Reconstructed vital rates when no information on them is provided (confidence intervals in grey).

***WORKING ON THIS...***

Fig. 2. Likelihood slices for pairs of parameters.

***WORKING ON THIS...***

**Appendix 1**

Description of the likelihoods associated to the goodness of fit of the estimated time series of population sizes and structures and the observed ones.