Ecological Modelling 222 (2011) 2027–2030

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Short communication

Theory put into practice: An R implementation of the infinite-dimensional model

Anna Kuparinen [a,*], Mats Björklund [b]

[a] Ecological Genetics Research Unit, Department of Biosciences, P.O. Box 65, FI-00014 University of Helsinki, Finland
[b] Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 D, S-75236 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 27 February 2011
Received in revised form 29 March 2011
Accepted 30 March 2011
Available online 27 April 2011

Keywords:
Evolution
Function valued trait
Infinite-dimensional model
Growth trajectory
Phenotypic variation
Selection

ABSTRACT

The infinite dimensional model (IDM) is an approach that has been developed for the analyses of phenotypic variation in function valued traits such as growth trajectories and continuous reaction norms. This model is particularly suited for the analysis of the potential and the constraints for growth to evolve under selection on body size. Despite of its applicability to a broad range of study systems IDM has only been applied in a handful of studies, as it is mathematically demanding for scientists not familiar with quantitative genetics methods. Here, we present a user-friendly R implementation of IDM, demonstrate its performance with growth data on nine-spined stickleback (*Pungitius pungitius*). In addition to rearing experiments, individual based size-at-age trajectories are often measured in wild in mark-recapture studies or estimated retrospectively from scales or bones. Therefore, our R implementation of IDM should be applicable to many studies conducted in wild and in a lab, and be useful by making the methodologically challenging IDM approach more easily accessible also in the fields where quantitative genetics methods are less standardly used.

1. Introduction

Phenotypic variability forms a basis for adaptive evolution by providing material for the selection to act upon (Roff, 1992; Stearns, 1992). Yet the evolutionary responses to selection further depend on the extent to which phenotypic variability reflects underlying genetic variation, and analyses of phenotypic diversity can provide useful insights into the potential and constraints of a population to adapt. This is particularly true in wild, where estimation of additive generic variance and covariance among phenotypic traits is difficult but observations on phenotypic trait values are easier to obtain and, thus, much more abundant.

In natural populations, body size is a trait that is typically under selection as it is closely linked to survival and reproductive success (Roff, 1992; Stearns, 1992). However, evolutionary responses to selection upon body size are constrained by the alternative patterns of growth present in a population (e.g. Kirkpatrick and Lofsvold, 1992; Björklund, 1993). For example, if large individuals tend to be large throughout their lives (i.e. they maintain their rank in size over time), then selection favouring large body size at one age increases body size at any age. If alternative patterns of growth exist (i.e. rank in size may change over time), then selection might either increase or decrease body size at other ages.

To address these questions in the context of growth or, more generally, in any function-valued trait, an approach called infinite-dimension model (IDM) was introduced by Kirkpatrick and Heckman (1989) and Kirkpatrick et al. (1990). This model detects alternative patterns of growth (i.e. shapes of the growth trajectory) present in a population, as well as the amounts of phenotypic variation accounted for by each of the growth patterns by decomposing a covariance matrix of size over a set of ages (Kirkpatrick and Heckman, 1989; Kirkpatrick et al., 1990). The advantages of the IDM are that it does not require assumptions about the analytical shapes of the growth trajectories and it provides direct information about how growth can evolve in a population (Kirkpatrick and Lofsvold, 1992; Meyer and Kirkpatrick, 2005). Specifically, if all phenotypic variation is accounted for by one single growth pattern, then responses to selection can only be seen in the elevation of growth trajectories, but not in the changes of its shape. In contrast, the presence of alternative growth patterns allows these patterns to become more abundant if being favoured by selection (Kirkpatrick and Lofsvold, 1992).

Since its introduction the IDM has been applied in a handful of studies (e.g. Björklund, 1993, 1997; Gilchrist, 1996; Björklund et al., 2003; Kingsolver et al., 2004; Ragland and Carter, 2004; Löhmus et al., 2010). However, also concerns about technical demands of the model have been presented: as Berner and Blackenhorn (2007) said, "Unfortunately, this method is technically demanding, so few people have used it and probably will in the future". While in the field of quantitative genetics the mathematics involved to the IDM estimation are well known, this may not be the case in the fields

* Corresponding author. Tel.: +358 40 731 3120.
E-mail address: anna.kuparinen@helsinki.fi (A. Kuparinen).

0304-3800/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2011.03.041

of life-history and wildlife ecology as we all as among scientists in more applied fields such as fisheries research. Nonetheless, tools to investigate phenotypic variations in growth and how growth may respond to selection are required across a wide range of biological studies. To this end, we present here a user-friendly R implementation of the IDM that performs the mathematical calculations involved to the model automatically, thus allowing one to use IDM without technical challenges. We further demonstrate the performance of the developed R package with an example, and discuss the problems that occasionally arise in the analysis of empirical datasets, and ways to deal with these.

2. R implementation of IDM

The starting point for the IDM approach is a dataset containing individual-based measurements of body sizes over a set of N ages. However, the model does not utilize raw growth trajectories but a phenotypic covariance matrix **P** for sizes at the considered ages ($N \times N$ matrix). This can be calculated with any standard statistical package; for instance in R, a covariance matrix is returned by a function `cov()`. (Note that if an additive generic covariance matrix **G** for body size at the considered ages can be estimated, **P** can be replaced with that. In R, the estimation of **G** matrix can be done using functions built in to `MCMCglmm` package). The steps involved in the IDM calculations are (i) calculation of the covariance matrix **P**, (ii) re-scaling of the ages over which size was measured between −1 and 1, (iii) building up a matrix $\mathbf{\Phi}$ of Legendre polynomials at re-scaled ages, (iv) calculation of the eigenvalues and eigenvectors of $\mathbf{C_P} = (\mathbf{\Phi})^{-1}\mathbf{P}\,(\mathbf{\Phi^T})^{-1}$, and (v) calculation of the N alternative growth trajectories (i.e. growth patterns) based on $\mathbf{\Phi}$ and eigenvectors of $\mathbf{C_P}$ using Eq. (8) in Kirkpatrick et al. (1990; eigenfunctions in their terminology), and scaling them to the norm of unity. For equations and further mathematical details, see Kirkpatrick et al. (1990).

We built the above IDM calculations as formulated by Kirkpatrick et al. (1990) into an R function `IDM()` that requires the phenotypic covariance matrix **P**, and a vector of ages at which size was measured, as input parameters. The function returns eigenvalues and eigenvectors of $\mathbf{C_P}$ as well as growth trajectories and the proportions of phenotypic variance that they each account for. Additionally, we developed a function `IDM.bootCI()` that similarly performs IDM calculations but also estimates 95% confidence intervals (CI) for eigenvalues, trajectories, and the proportions of phenotypic variation accounted for by the trajectories via bootstrapping. Due to limited sampling negative eigenvalues might result. This is not a major problem when it comes to **P**-matrices compared to **G**-matrices (Meyer and Kirkpatrick, 2010). Since eigenvalues tend to be overdispersed (op. cit), in the `IDM.bootCI()` the negative values are forced upwards to be zero, while the dominant eigenvalue is reduced by the same amount, i.e. the eigenvalues are truncated but their mean remains unchanged (i.e. squeezing *sensu* Kirkpatrick and Lofsvold, 1992). If the number of observations is large compared to the number of ages, negative eigenvalues as a result of sampling error are rare (Meyer and Kirkpatrick, 2008). The function `IDM.bootCI()` is parameterized with a complete size-at-age data matrix (i.e. missing values are not allowed), an age vector, and the requested number of bootstrap simulations. The reason for developing the two alternative functions is that `IDM()` can be applied to any arbitrary covariance matrix, which can be convenient, for example, if the covariance matrix must be estimated piecewise due to missing values (see further discussion on this below).

The R functions described above are built into an open-source R package InfDim that can be downloaded from CRAN or obtained from the first author. The package was created using 2.10.1 (R Development Core Team, 2009) and R Tools for Windows (http://www.murdoch-sutherland.com/Rtools/).

3. An example with fish growth data

To illustrate the performance of our R implementation of IDM, we used growth data on 53 female nine-spined stickleback (*Pungitius pungitius*) collected in a common garden experiment. The individual fish were reared at 17 °C, fed *ad libitum*, and their body length was measured at the ages of 17, 45, 80 and 115 days (Fig. 1a; Kuparinen et al., 2011). Phenotypic variation in growth patterns was investigated by parameterizing the `IDM.bootCI()` function with the growth data, and the bootstrap sampling for CIs was repeated 1000 times.

About 89.9% of phenotypic variation (Fig. 1b) in the patterns of stickleback growth was accounted for by the first trajectory that predicted positive correlation between body sizes across the study period (Fig. 1c). However, 7.7% of the phenotypic variation (Fig. 1b) was accounted for by an alternative growth trajectory that describes negative correlation between early and late growth, with the turning point being at about 80 days (Fig. 1d). Variation associated with the third and the fourth trajectory was minor (<2.4%; trajectories not shown). The IDM analysis demonstrates that fish growth in the experiment was largely restricted to two alternative growth patterns; most individuals that were large (small) at some point tended to be large (small) over the entire study period, while a few that were initially large (small) tended to be among the smaller (larger) ones towards the end of the experiment.

The executable R code for the above analyses is provided in the documentation file of the `IDM.bootCI()` function, and the stickleback growth data is build into the InfDim package under the name `fish`. Similarly, an example to use `IDM()` is provided in the function's documentation file. This example utilizes the mouse growth data given in Kirkpatrick et al. (1990), and returns results identical to those reported by Kirkpatrick et al. (1990).

4. Discussion

The advantage of our R implementation of the IDM is that it allows the use of the model without technically demanding calculations – a task that has been found challenging (Berner and Blackenhorn, 2007). In addition to quantitative genetic studies and animal breeding applications (Meyer and Kirkpatrick, 2005), individual based size measurements over a common set of ages (as required by the IDM model) are often collected also in wild in mark-recapture studies (e.g. Björklund, 1997), or estimated retrospectively from annual scale or bone patterns (e.g. Schirripa and Goodyear, 1997; Kuparinen et al., 2009a) in studies monitoring life-histories and temporal trends in those. Therefore, IDM along with its R implementation are widely applicable also to fields where quantitative genetics methods are less standardly used and the R implementation of IDM may provide an opportunity to re-analyse already existing datasets. Moreover, even though here discussed and demonstrated in the context of growth trajectories, IDM can be applied also to other function valued traits, such as, for instance, continuous reaction norms (Kingsolver et al., 2004).

Missing values are a practical problem often present in repeated measures, such as size-at-age trajectories, particularly when data is collected from wild populations. An observation can be lacking due to failure in recapturing an individual, because an individual has left the study area, or because the individual has died. This poses some challenges for the IDM approach, specifically for the estimation of the phenotypic covariance matrix for sizes over a set of ages. In some cases trajectories containing missing values can be simply omitted: for example, if an individual died because of being