# My, how you've grown: a practical guide to modeling size transitions for Integral Projection Model (IPM) applications

Tom E.X. Miller*[a] and Stephen P. Ellner[b]

[a]Department of BioSciences, Rice University,Houston, TX
[b]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York

April 30, 2020

**Authorship statement:** All authors discussed all aspects of the research and contributed to developing methods, analyzing data, and writing and revising the paper.

**Data accessibility statement**: No original data appear in this paper. Should the paper be accepted, all computer scripts supporting the results will be archived in an appropriate public repository such as Dryad or Figshare, with the DOI included at the end of the article.

*Corresponding author. Department of BioSciences, Rice University, Houston, TX 77005-1827. Email: tom.miller@rice.edu Phone: 713-348-4218

## Abstract

## 1 Introduction

Structured demographic models – matrix and integral projection models (MPMs and IPMs) – are powerful tools for data-driven modeling of population dynamics and viability that are widely used in basic and applied settings. In contrast to MPMs for populations with discrete structure (life stage, age class, etc.), IPMs are a relatively recent development (Easterling et al., 2000) best known for their ability to represent populations structured by continuous variables, most commonly size. A second, related innovation[1] of the IPM framework is its emphasis on modeling size-dependent vital rates as sub-models of the projection 'kernel'. This means that ecologists could assemble an IPM using familiar regression tools and leverage their capacity for covariates and hierarchical variance structures (i.e., random effects) to statistically describe growth, survival, reproduction, and other demographic transitions as functions of size, in the context of other factors (age, competition, temperature, etc.) affecting those rates. The relative ease and flexibility of the regression-based approach has facilitated a growing body of IPM literature that examines how covariates of individual demographic fate scale up to affect population dynamics (cite examples) and that partitions demographic heterogeneity associated with space, time, or individual identity (cite examples).

Compared to other vital rates, growth is special. The regression sub-models for survival and reproduction only need to give the expected values of these vital rates as functions of size. Modeling growth for an IPM is about more than the expected value: the full probability distribution of future size, conditioned on previous size, must be defined. Modeling that distribution should be done just as carefully as any other other part of building the model.

Steve's stray bits to be incorporated:

Modeling growth in IPMs is *modeling*, not statistics. Statistical tools and tests inform the process, but do not dictate it. We are not claiming that we know how to find the "right" model for any particular data set or application through some model selection procedure. Our goal is to demonstrate how a simple workflow, using tools that were nonexistent or not readily available when IPMs came into use, makes it straightforward and relatively easy to identify when the default choice (Gaussian with possibly nonconstant variance) is a poor fit to the data, and to then choose and fit a substantially better growth model that is no harder to use in practice.

The default growth model traces back to Easterling et al. (2000). The analyses described there pre-date version 1.0 of R by several years, and they were quickly replaced by better methods such as the linear and generalized linear mixed models regression functions in R, notably the option to fit linear mixed models with nonconstant growth variance. But as Peterson et al. recently noted, despite the growing list of other options and repeated calls in the literature for greater attention

[1] *This 'innovation' is not unique to IPMs but it was rare in MPMs before IPMs.*

to growth modeling, the general state-of-the-art in the literature remains stuck where it was 15 or so years ago, using the default model without pausing to examine critically whether or not it actually provides a good description of the data. Our goal here is to break that logjam, by building on Peterson et al. to provide a practical "new default" that researchers can follow to choose, fit, and implement better descriptions of individual growth for IPMs and any other size-structured population model.

The modeling workflow that we suggest runs as follows:

1. Fit a "pilot" model with a Gaussian distribution of future size conditional on current size and any other covariates, having fitted non-constant variance.
2. Use the fitted variance function to compute standardized residuals. If the Gaussian pilot model is valid, the set of standardized residuals should be Gaussian with constant variance.
3. Use statistical and graphical diagnostics, which we detail below, to identify if and how the standardized residuals deviate from Gaussian: are there skew and kurtosis, and if so, how do those vary across the size distribution? If the standardized residuals are as expected under the pilot model, skip to the last step.
4. Otherwise, identify a distribution with nonzero skew and/or kurtosis that provides a good fit to the standardized residuals. Tools to automate this step are available in R.
5. Refit the growth model using the chosen distribution, with shape (skew and kurtosis) varying according to the patterns seen in the graphical diagnostics. How this can be done depends on features of the model at hand, in particular whether or not it includes random effects.
6. Test the final model through graphical diagnostics comparing simulated growth increments with the data. A good model will generate simulated data that look like the real data.

The assumption of non-constant variance in the pilot model means that it is not necessary to seek a data transformation that stabilizes the growth variance. Transformation may have other benefits (e.g., log transformation avoids spurious negative sizes, or some other transformation may be helpful in modeling survival or fecundity). However, it does entail some model selection to specify the mean-variance relationship of future size. In contrast, we recommend that statistical model selection should be the basis for choosing the form of the final regression model. Rather, its form should be chosen to match the observed properties of the scaled residuals, and at most slightly modified based on final diagnostic tests.

## Acknowledgements

This research was supported by US NSF grants DEB-1933497 (SPE) and ....

## Literature Cited

Easterling, M. R., Ellner, S. P., and Dixon, P. M. (2000). Size-specific sensitivity: applying a new structured population model. *Ecology*, 81(3):694–708.