

# IPMpack: an R package for demographic modeling with Integral Projection Models (v.1.2)

Jessica Metcalf, Sean M. McMahon, Rob Salguero-Gomez, Eelke Jongejans

August 2, 2012

The goal of IPMpack is to provide a suite of demographic tools based on Integral Projection Models (IPMs) to support biologists interested in making projections for populations where demography is strongly linked to a continuous variable, such as size. The package includes functions that can take data, such as size or age, as well as environmental covariates, and build models of growth, survival and fecundity. Functions are defined that then take these statistical models and construct IPMs. IPMpack has tools that compare different functional forms for the underlying statistical models, plotting them and returning model scores, as well as tools for diagnostic tests of the IPM models themselves. There are also methods to build population models for varying environments, use Bayesian methods to sample population parameters, estimate longevity and passage time, sensitivity and elasticity (of either parameters or matrix elements), and much more.

This vignette is intended to introduce the concepts of IPMs as well as the implementation of IPMpack to biologists with a wide range of quantitative skills. This vignette is for IPMpack version 1.3, and so we encourage users to contact the IPMpack team at [IPMpack@gmail.com](mailto:IPMpack@gmail.com) with any feedback or mistakes they find. We also host a blog at R-forge IPMpack Web Site that contains news of updates, new features, and announcements of papers and meetings relevant to IPMs.

# 1 Introduction to Integral Projection Models

An Integral Projection Model (IPM) is a demographic tool that can estimate the dynamics of populations where individuals' fates depend on state variables that are continuous (e.g., weight, diameter at breast height, height, limb length, rosette diameter) or quasi-continuous (e.g., number of leaves, age, number of reproductive structures) and may be a mixture of discrete and continuous variables. IPMs track the distribution of individuals  $n$  across these state variables between census times (e.g., year  $t$  and year  $t + 1$ ) by projecting from models that define the underlying vital rates (e.g., survival, growth, and reproduction) as a function of the (quasi-)continuous state variables. For detailed introductions to IPMs see Easterling et al. (2000), and Ellner & Rees (2006, 2007).

Briefly, an IPM is defined by a kernel  $K$  that represents probabilities of growth between discrete or continuous stages, survival across these stages, and the production of offspring and offspring recruitment. For example, in the simplest case, where the population is structured by a continuous covariate, size, then

$$n(y, t + 1) = \int_L^U K(y, x) n(x, t) dx \quad (1)$$

where  $n(y, t + 1)$  is the distribution across size  $y$  of both established and new individuals in census time  $t + 1$ ,  $n(x, t)$  the distribution across size of individuals in census time  $t$ , and  $L$  and  $U$  the lower and upper size limits modeled in the IPM, respectively.

Multiple functional forms for both demographic processes as well as their error structures can be accommodated with IPMPack. The  $F$  kernel (equation 4) describes per-capita contributions of reproductive individuals to number of new individuals at the next census. Multiple size-dependent or size-independent vital rates can be fitted within the  $F$  kernel, reflecting for example reproductive probability, number of reproductive structures (e.g. flowers in plants, basidia in fungi), number of propagules within reproductive structure (e.g. seeds in inflorescences), and so on. Additionally, a range of constants ( $c_1, c_2, \dots$ ) can be included if there are no data for a stage. All of these will be multiplied to obtain the eventual fertility for individuals of each size. Finally, the  $F$  kernel definition includes a probability density function describing the size of offspring recruiting into the population,  $f_d$ .

From equation 1:

$$n(y, t + 1) = \int_L^U K(y, x) n(x, t) dx = \int_L^U [T(y, x) + F(y, x)] n(x, t) dx, \quad (2)$$

where

$$\int_L^U T(y, x) n(x, t) dx = \int_L^U \text{surv}(x) \text{growth}(y, x) dx, \quad (3)$$

and

$$\int_L^U F(y, x) n(x, t) dx = \int_L^U c_1 c_2 c_3 \dots \text{fec1}(x) \text{fec2}(x) \text{fec3}(x) \dots f_d(y, x) dx \quad (4)$$

After numerically solving these kernels, key ecological and evolutionary quantities such as the population rate of increase  $\lambda$ , the stable population size structure, the net reproductive rate  $R_0$ , and many others can be estimated (see Caswell 2001 for more a comprehensive discussion).

Essentially, the same tools are available for IPMs as for discrete projection matrices (matrix population models), e.g., estimation of population growth rate, sensitivities, elasticities, life table response experiment [LTRE] analyses, passage time calculations, etc. (Caswell 2001, Cochran & Ellner 1992, and others). The main difference between an IPM and a matrix model is that while in discrete projection matrices the number of classes (i.e., number of stages in the life cycle

of the study species) must be defined **a priori**, IPMs impose the discretization of the three-dimensional surface defined by equation 1 in the last step. This produces a typically large matrix (e.g., 100 x 100 cells) that is more robust to biases from matrix dimensionality (Zuidema et al. 2010, Salguero-Gomez & Plotkin 2010) and sample size (Ramula et al. 2009) than classical matrix models.

The goal of IPMpack is to provide a centralized set of quantitative techniques based on IPMs to help ecologists and evolutionary biologists model populations. IPMpack can accommodate multiple vital rates from complex life cycles all grouped into two main sub-kernels:  $T$  and  $F$  (equation 2) <sup>1</sup>.

This vignette will now walk through the steps of a basic IPM analysis. We first describe the kind of data necessary to build an IPM. If a user begins ‘from scratch’, they must input data in a specific format (described below). However it is possible to jump past this step and use IPMpack capabilities on IPMs that were developed outside of IPMpack. That is, if a user wants quick figures, summary statistics, other analyses on an IPM matrix already built, IPMpack can readily accommodate that. However there are some features that, because of the object-oriented coding require some specific structures (and other features that do not). Please refer to the manual files and the rest of this vignette for this information. However a user wants to implement IPMpack, the vignette will begin with data input. We will then walk through how to build and analyse a basic IPM model. More complex models will be introduced later, with options to create unique class objects and methods, as well as run comparative model testing and Bayesian implementations.

## 2 Getting started: setting up the data for IPMpack

For users who prefer to define IPM matrices using their own statistical tools, there is no requirement for the data to be in any particular format, and most of the functions in IPMpack will operate on the matrices directly (e.g., life expectancy, sensitivity of matrix elements, etc.). However, to use IPMpack’s full capacities, the individual-level demographic data must be organized in a **data frame** (a class of object in R [see help file for ‘data.frame’ in base]), where each row represents one observation of an organism in the population at one census time  $t$  with the following potential column names:

- **size**: size of individuals in census time  $t$  \*
- **sizeNext**: size of individuals in census time  $t + 1$  \*
- **surv**: survival of individuals from census time  $t$  to  $t + 1$  (contains: 0 for death or 1 for survival) \*
- **fec1, ...**: as many columns as desired relating size to sexual reproduction. For example, this might be:
  - **fec1**: probability of reproduction (output: 0 for no reproductive or 1 for reproductive)
  - **fec2**: number of reproductive structures (output: 1, 2, 3, ...) when individual is reproductive, that is, when  $\text{fec1} = 1$
  - **fec3**: number of propagules (output: 1, 2, 3, ...) per reproductive structure (e.g. seeds per flower in reproductive plant individual)
  - ...
- **stage**: stage of individuals in census time  $t$ . For rows in the data frame where **size** is not an NA, then this must be the word “continuous”. Where **size** is NA, any variety of named discrete stages may be defined (e.g. “seed bank”). If this column is missing, many procedures

<sup>1</sup>Note than in the seminal paper by Easterling et al. (2000) this kernel was referred to as  $P$ , but here we follow the terminology by Caswell (2001) and call it  $T$  instead). The  $T$  kernel (equation 3) describes **growth** between demographic censuses conditional on individuals’ survival (**surv**).

in IPMpack are designed to simply fill in this column assuming that only “continuous” state variables describe the life cycle of the species, i.e. there are no discrete stages. For running `makeFecObj`, the column must be a factor. If not supplied, the function will generate this column assuming all individuals are “continuous”.

- **stageNext**: stage of individuals in census time  $t + 1$ , in the simplest case, “continuous” or “dead”. As above, this column is not essential for many procedures in IPMpack. For running `makeFecObj`, the column must be a factor. If not supplied, the function will generate this column assuming all individuals that are alive are “continuous”.
- **number**: number of individuals corresponding to each row in the data frame. For all rows corresponding to movement between continuous stages, this value will be 1, but for movement between discrete stages (e.g., from “dormant seeds” to “seeds ready to germinate”) then this number may be  $> 1$ , potentially directly reflecting observed individuals in the data. This information avoids having a data frame with a row for every discrete stage (e.g. seed). As above, many procedures in IPMpack will simply assume that this value is always 1.
- **covariate**: value of a discrete covariate in census time  $t$ , such as light environment at time  $t$ , age at  $t$ , patch at  $t$ , etc.
- **covariateNext**: value of a discrete covariate in census time  $t + 1$ .
- ...any other covariates of interest, named as desired by the user are possible too (e.g., precipitation, habitat, temperature, etc).
- **offspringNext**: if the size contained in `sizeNext` corresponds to the size of an offspring, this column will contain either the value “sexual” or “clonal” (depending on whether sexual or clonal reproduction is being considered). If this column exists, rows that take these two values will be excluded from the growth analyses (functions `makeGrowthObj` and variants thereof, see below).

The \* symbol above indicates the minimum columns in the data frame required to obtain passage time and life expectancy calculations. These values form the  $T$  kernel. If sufficient additional columns are available, a full life-cycle model, containing the  $F$  kernel, can be produced and further analyses are possible. Although `size` and `sizeNext` can be transformed, many of the utility functions assume no transformations in columns in the original data frame pertaining to fertility. Transformations can be formally called in various parts of the package and appropriate  $F$  matrices built that account for these transformations.

### 3 The basics: building an IPM

First, the user must install IPMpack from CRAN using `install.packages("IPMpack")` and then load IPMpack into an R session (`library(IPMpack)`) (see help files for problems with installation or loading).

```
> library("IPMpack")
```

Next, the user must input demographic data. As mentioned above, most functions of IPMpack require a data file with at minimum columns called `size`, `sizeNext`, `surv`, where ‘size’ is size at time  $t$ , ‘sizeNext’ is size one census later, and ‘surv’ is a series of 0s and 1s, indicating if the individual survived or not. In the case of ‘size’ and ‘sizeNext’, data can be transformed (e.g., onto a log scale), if appropriate via functions built into IPMpack. For the purpose of learning how to use IPMpack, the user can either use his/her own data (adjusted to have the appropriate headings, as aforementioned), or generate them with a function built into IPMpack:

```
> dff <- generateData()
```

A quick check indicates that this contains sensible (fictional) information:

```
> head(dff)
```

	size	sizeNext	surv	covariate	covariateNext	fec	stage
1	7.672077	7.1416931	1	1	1	0.0000000	continuous
2	5.384852	NA	0	1	0	0.6731872	continuous
3	NA	1.9162983	NA	0	0	NA	<NA>
4	1.685810	0.8629548	1	1	0	0.0000000	continuous
5	4.658631	NA	0	1	0	0.0000000	continuous
6	5.657233	2.0253533	1	1	0	13.7487432	continuous

```
stageNext
1 continuous
2 continuous
3 continuous
4 continuous
5 continuous
6 continuous
```

for simplicity, no discrete covariates are included in this first example. Figure 1 (p. 6) is produced by the following code:

```
> plot(dff$size, dff$sizeNext, xlab = "Size at t", ylab = "Size at t+1")
```

IPMpack is written in object-oriented code, using *S4* objects. This means that extra object classes are used by IPMpack, with methods assigned to those classes that do particular things to specific objects. An example for those familiar with *R* is the `plot` function. When applied to two vectors, it produces an x-y plot, but when applied to a fitted linear regression, it provides a series of diagnostic plots. In other words, the ‘plot’ method is object-specific and does different things to objects of class ‘numeric’ and objects of class ‘lm’.

IPMpack contains defined classes for growth, survival and fertility objects, and associated methods that allow the user to build IPM objects. In addition, this object-oriented structure in IPMpack uses methods from IPM objects to calculate life expectancy, passage times, and other population estimates of interest. The advantage of object-oriented programming is its flexibility: for example, the same machinery can be applied to suites of underlying regression forms and the user can take advantage of pre-existing highly generalized *R* functions, such as `predict`. The needs of any particular dataset may require different object and method definitions. Towards the end of this vignette we also describe how to define a new class and a new method (e.g., a new growth object for a specific life-history structure, and a new growth method applicable to plotting information from that object).

As an example, let us first define objects built as simple polynomial regressions from the generated data. The source code of `generateData` will confirm that the survival data is built around a polynomial logistic regression relating size at  $t$  to survival from  $t$  to  $t + 1$ , and the growth data is built around a polynomial regression relating size at  $t$  to size at  $t + 1$ . To make growth and survival objects that reflect this, the user must implement:

```
> gr1 <- makeGrowthObj(dataf = dff, Formula = sizeNext~size+size2)
> sv1 <- makeSurvObj(dff, Formula = surv~size+size2)
```

In both these functions, the argument `Formula` contains formulas of the type used in linear or logistic regressions in *R*, built around the possible defined range of transforms of `size` currently available (`size2` which is `size2`, `size3` which is `size3`, and `logsize` which is `log(size)`). Currently further transforms of `size` are not possible. This function can also be used to fit models that include a single discrete covariate (e.g., light environment, age, etc) as long as this exists in the `dataf` in a column named `covariate`. For instance, the user could model the population dynamics according to `size + covariate` or `size + logsize*covariate`, etc. For more complex analyses,

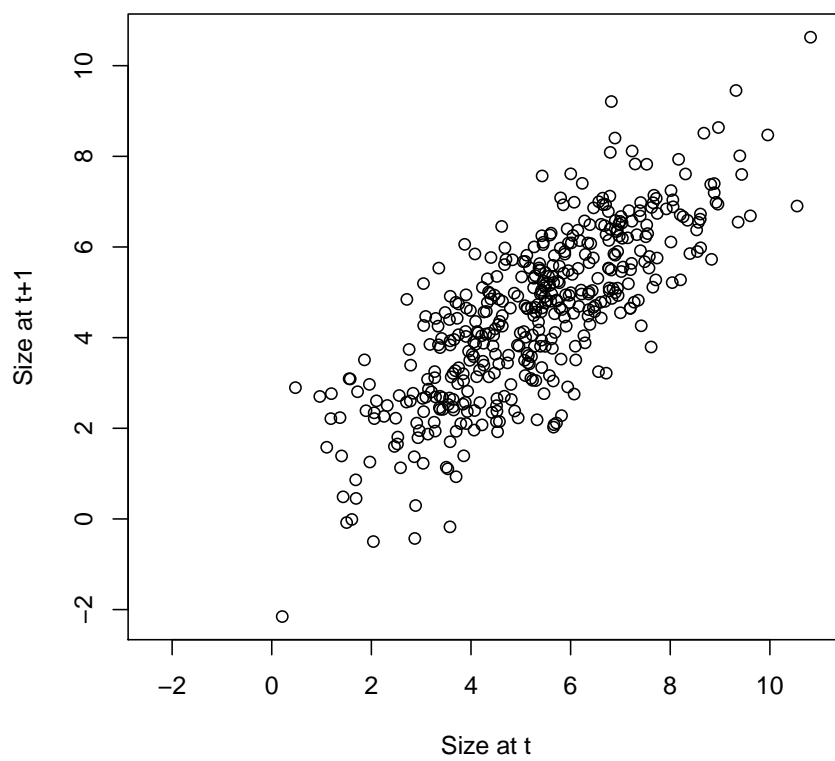


Figure 1: Size at  $t$  and size at  $t+1$

other covariates (time since fire, precipitation, etc) can be fitted as long as they exist within `dataf`. For the growth model, possibilities for the response variable in the `Formula` are: `sizeNext` meaning that the response variable is size at the next census time, or `incr` meaning that the response variable is the size increment that has accrued between the two census times (common among tree demographic studies), and `logincr` meaning that the response variable is the log of the size increment that has accrued between the two census intervals.

Glancing at the source code will confirm that all these functions simply fit a linear regression relating size at  $t+1$  or increment to size at  $t$  and covariates for growth, as for survival. The survival and growth objects created have a slot called 'fit' that holds the regression and a slot `sd` that holds the variance around the regression.

```
> gr1

An object of class "growthObj"
Slot "fit":

Call:
lm(formula = Formula, data = dataf)

Coefficients:
(Intercept)          size          size2
    0.338962     0.823702    -0.005149

Slot "sd":
[1] 1.136372
```

IPMpack contains two functions that allow the user to check these two relationships against the data used for them in order to explore goodness of fit and effect of mesh size, shown in Figure 2 (p. 8).

```
> par(mfrow = c(1, 2), bty = "l", pty = "m")
> p1 <- picGrow(dff, gr1)
> p2 <- picSurv(dff, sv1, ncuts = 30)
```

To build a demographic model describing survival and growth transitions from these objects, the user can use the function `createIPMTmatrix`, i.e.:

```
> Tmatrix <- createIPMTmatrix(nBigMatrix = 50,
+                             minSize = -5, maxSize = 35,
+                             growObj = gr1, survObj = sv1,
+                             correction = "constant")
```

where `nBigMatrix` is the number of bins used, `minSize` and `maxSize` define the limits of the IPM,  $U$  and  $L$  in the equations above. Typically, these ranges should usually extend to beyond the smallest and largest size measurement, but the user might want to exclude outliers). The objects `growObj` and `survObj` define changes in size and survival as defined above.

IPMpack includes a useful function `diagnosticsTmatrix` that provides a series of plots indicative of whether bin choice and size range is adequate. Applying this function as a preliminary step before obtaining demographic and evolutionary output from IPMs can help identify basic problems in the fitting of vital rate functions or in the creation of the IPM matrices before proceeding. Several common problems can be diagnosed with the panelled figure produced by `diagnosticsTmatrix` (Figure 3).

The top-left panel shows the range of the data and the range of the state variable fitted in the matrix. If these are mis-matched, the limits of the data used in building the `Tmatrix` can be adjusted with the `minSize` and `maxSize` arguments in `createIPMTmatrix`. One common problem

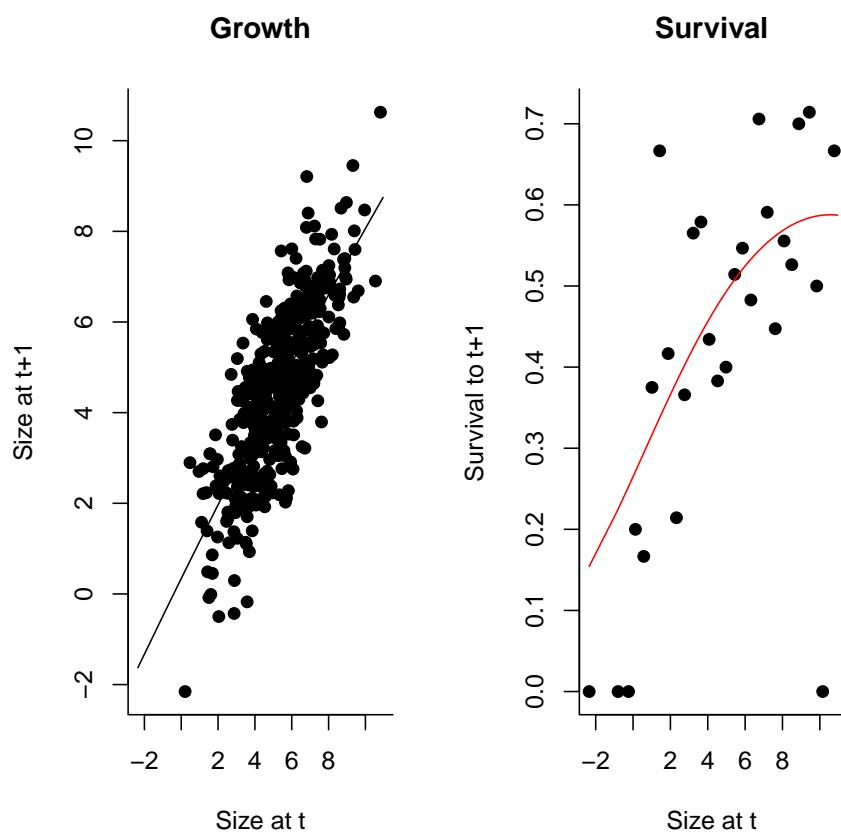


Figure 2: Growth and survival objects



```
[1] "Range of Tmatrix is "  
[1] 3.360023e-247 1.642476e-01
```

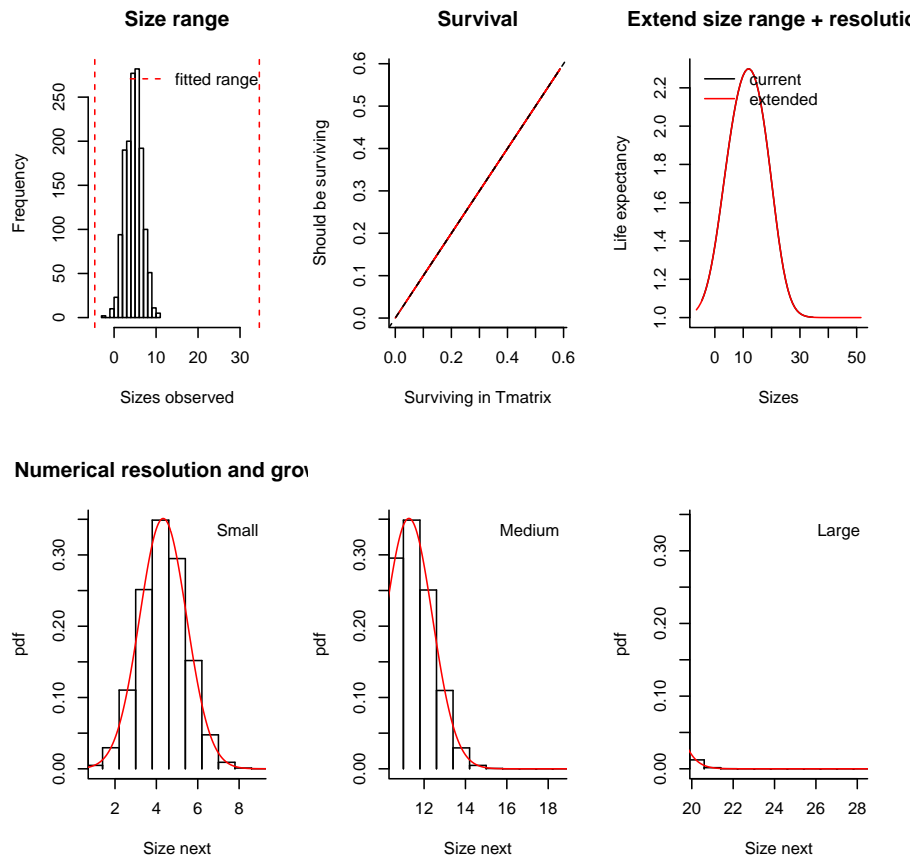


Figure 3: Diagnostic plots for the Tmatrix.

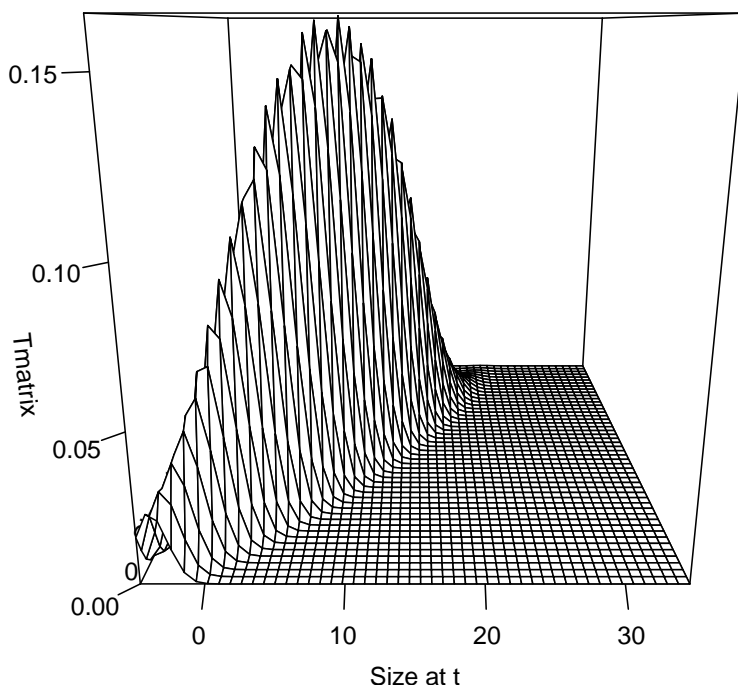


Figure 4: Transition matrix encompassing survival and growth transitions only

can result from the loss of parts of the continuous distribution when binning. This can mean that the sum of the columns of the matrix do not match the fitted survival. The top-middle panel of Figure 3 will show this discrepancy if the red and black lines do not overlay. To correct this, `createIPMTmatrix` has an argument `correction = "constant"` which ensures that the columns sum to the fitted survival by adding a constant. The top-right panel checks whether extending the size range included in the matrix and extending the number of bins does not alter basic predictions from the IPM. The three lower panels show the discretized IPM (histograms) and the theoretical density function. If mismatched, increasing the `ttnBigMatrix` argument may correct this discrepancy.

The `createIPMTmatrix` function builds around methods defined so that it will provide appropriate output whatever the survival and growth objects are (e.g. error structure, covariates). The T matrix contains a matrix defining the transitions, but also other useful slots, e.g., the meshpoints, covariates, grid size. The user can access this information by writing:

```
> slotNames(Tmatrix)

[1] ".Data"          "nDiscrete"      "nEnvClass"      "nBigMatrix"
[5] "meshpoints"     "env.index"      "names.discrete"
```

and finally, the user can plot the Tmatrix using `persp` (Figure 4). Next, with this, the user can obtain the life expectancy, and passage time to a chosen size (here set at the mean) for the range of meshpoints

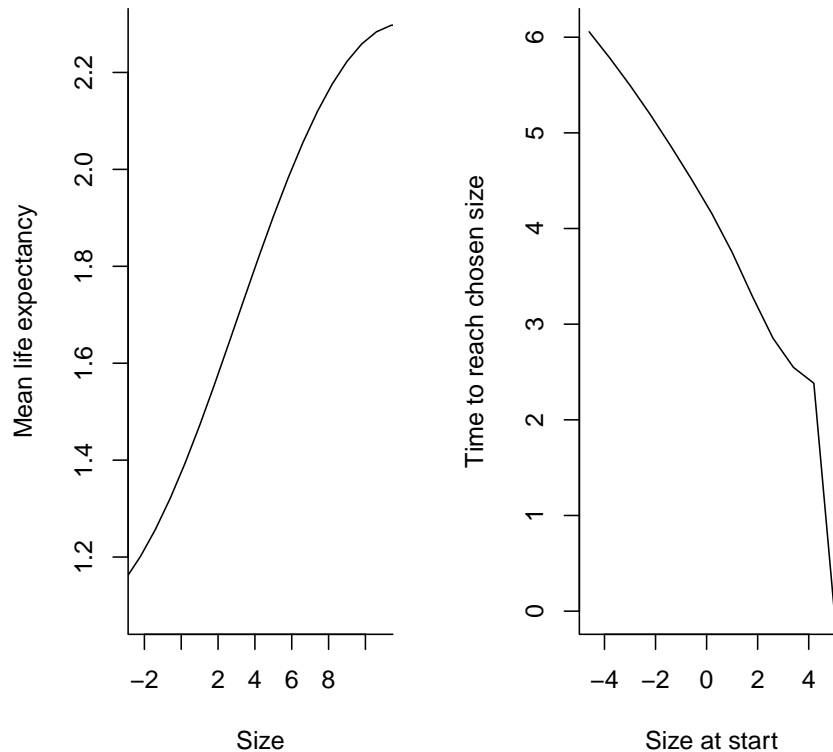


Figure 5: Associated Life Expectancy and Passage Time

```
> LE <- meanLifeExpect(Tmatrix)
> pTime <- passageTime(mean(dff$size, na.rm = TRUE), Tmatrix)
```

and the user can also plot these against `Tmatrix@meshpoints` to examine how life expectancy and passage vary as a function of size (Figure 5 p. 11). The function `runSimpleModel` takes as minimum arguments a data frame and a target size (i.e., here type: `runSimpleModel(dff, chosenSize = 4)`) and runs this analysis to create figures for survival, growth, life expectancy and passage time as shown so far, assuming the simplest possible models of survival and growth (basic linear and logistic regressions, no covariates except for size, etc).

If the user defines a fertility object - which for instance is not always easy with for example trees - IPMpack can also create a transition matrix describing movement between sizes attributable to fertility.

```
> fv1 <- makeFecObj(dff, Formula = fec~size,
+                   Family = "gaussian",
+                   Transform = "log")
> Fmatrix <- createIPMFmatrix(nBigMatrix = 50, minSize = -5,
+                             maxSize = 35,
+                             fecObj = fv1,
+                             correction = "constant")
```

Note that `makeFecObj` will use the relationships defined in `Formula` and the family defined in `Family` with transforms defined in `Transform` in the order supplied. Please note that the fecundity

Formula MUST match the Transform since IPMPack needs to combine the parameters from Formula with the right Transform to appropriately build the F matrix.

The default arguments required to run `makeFecObj` to create a fecundity object from which an F matrix with no discrete stage can be built are `offspringSplitter=data.frame(continuous=1)`, `offspringTypeRates=data.frame(NA)` and `fecByDiscrete=data.frame(NA)`. Additionally, note that if there are values other than “continuous” in the `stage` column of the data-frame named `dff` in the example above, then the function will assume that multiple offspring classes are required, and the result will be an IPM with `nBigMatrix` + the number of offspring classes deduced (which is the number of names in `stage` other than “continuous”). This may lead to a mismatch with the size of the T matrix unless a discrete transition matrix is explicitly being included in the T matrix (see below, incorporating discrete stages).

If the data-frame contains an extra column `offspringNext` that takes the values `sexual`, and that corresponds to rows where both `size` and `sizeNext` are different from NA, the user can define a relationship between maternal size and offspring size through the `makeFecObj` argument `offspringSizeExplanatoryVariables`. The default is to only fit an intercept, equivalent to simply having a mean and variance of offspring size. The function `makeFecObj` also allows users to simply over-write the mean and variance of offspring size with the values of their choice (arguments `meanOffspringSize` and `sdOffspringSize`).

The function `makeClonalObj` operates identically to `makeFecObj` except that offspring are only considered for fitting the distribution of mean and standard deviation of offspring size if the column `offspringNext` takes the values `clonal`. Rows where `offspringNext` takes the values “sexual” or “clonal” are excluded from the survival and growth analyses (`makeGrowthObj` and `makeSurvObj`).

The user can combine the F matrix with (an identically built, i.e., same bin number, size limits and discrete classes) survival-growth transition T matrix to obtain a full Integral Projection Model, and its population growth rate  $\lambda$ , sensitivity, elasticity, etc.

```
> IPM <- Tmatrix + Fmatrix
> eigen(IPM)$value[1]
```

```
[1] 3.291736
```

```
> sensitivity <- sens(IPM)
> elasticity <- elas(IPM)
```

These outputs can be plotted against the meshpoints (Figure 6 p. 13). In addition to perturbation measures from mesh cells, the user can also obtain sensitivity and elasticity of particular parameters that underlie the kernels, e.g., doing:

```
> res <- sensParams(growObj = gr1, survObj = sv1, fecObj = fv1,
+                   nBigMatrix = 50, minSize = -5, maxSize = 15)
> res
```

```
$slam
```

grow (Intercept)	grow size	grow size2
0.08914487	0.19297888	0.45532391
sd growth	surv (Intercept)	surv size
0.02286143	0.23717006	0.50668214
surv size2	reprod 1 (Intercept)	reprod 1 size
1.18025847	2.88966864	6.14291441

```
$elam
```

grow (Intercept)	grow size	grow size2
0.030218218	0.158965040	-0.002344424
sd growth	surv (Intercept)	surv size
0.025920727	-0.242068790	0.132091451
surv size2	reprod 1 (Intercept)	reprod 1 size
-0.014579335	1.743533749	1.385737689

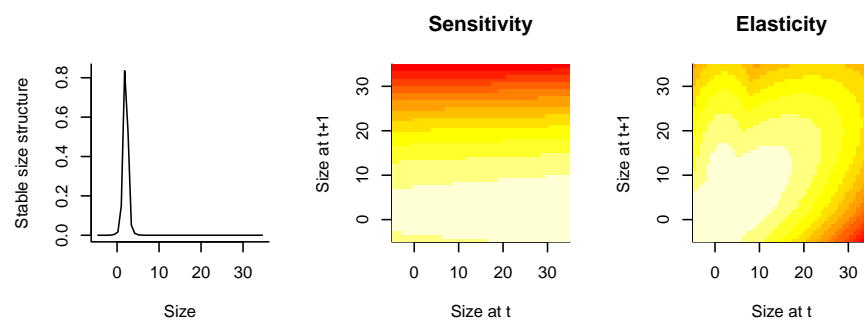


Figure 6: Measures off a full IPM

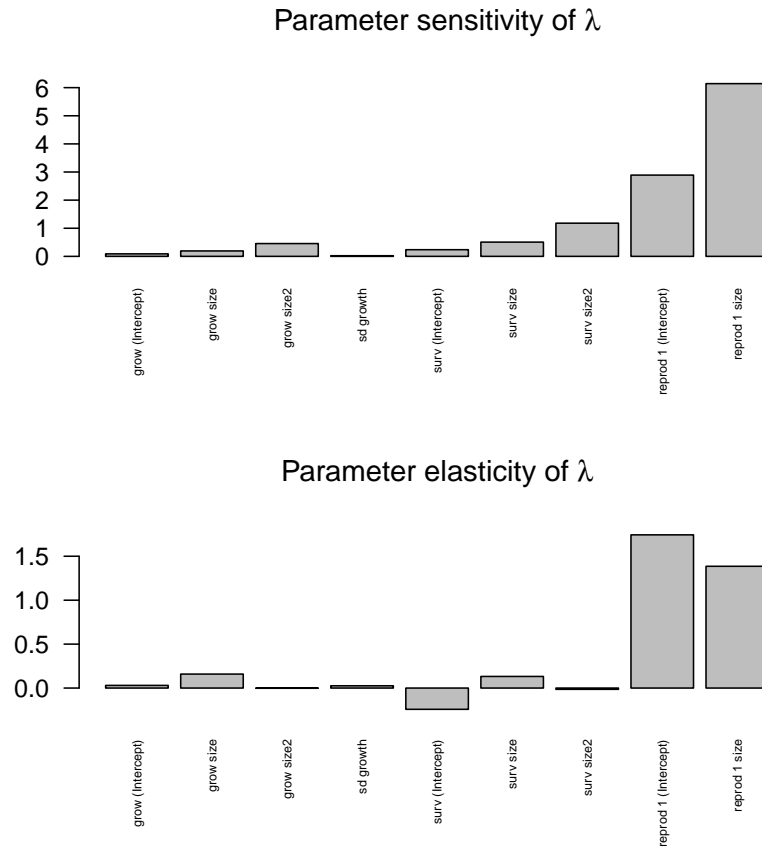


Figure 7: Sensitivity and elasticity of parameter values

and this output can be plotted out (Figure 7 p. 14) using

```
> par(mfrow = c(2, 1), bty = "l", pty = "m")
> barplot(res$slam, main = expression("Parameter sensitivity of "*lambda),
+         las = 2, cex.names = 0.5)
> barplot(res$elam, main = expression("Parameter elasticity of "*lambda),
+         las = 2, cex.names = 0.5)
```

## 4 Incorporating discrete stages

Populations are often structured by both discrete and a continuous stages. For example, many plant populations may persist for many years in a seedbank as well as having size-determined fates after they germinate. IPMPack can incorporate this variability for complex life cycles (Ellner & Rees 2006). To illustrate using discrete stages in an IPM, we generate data that includes both discrete and continuous life-history stages:

```
> dff <- generateDataDiscrete()
```

A quick check indicates that these data contain several types of stage classification (and not just "continuous" as seen up till now):

```
> table(dff$stage)
```

continuous	dormant	seedAge1	seedOld
950	50	35	32

Given this data structure, the user can make a fertility object that reflects the fact that propagules (seeds in this example) produced in one year may recruit directly into the continuous phase (e.g., seedling), or may end up in a discrete stage (e.g., seed bank). The `makeFecObj` (and similar functions) has an argument called `offspringSplitter` that allows the user to define these paths:

```
> fv1 <- makeFecObj(dataf = dff, Transform = "log",
+                   offspringSplitter = data.frame(continuous = 0.2,
+                   dormant = 0, seedAge1 = 0.8, seedOld = 0),
+                   fecByDiscrete = data.frame(dormant = 0,
+                   seedAge1 = 0, seedOld = 0))
```

In this example, 20% of seeds produced at  $t$  end up in the continuous part of the population structure at  $t+1$  (for example, they might directly recruit as rosettes from one year to the next) and 80% of seeds recruit into the "one year old seeds" stage. Although in this case no individuals are recruited at  $t+1$  into the "dormant" or "old seeds" stages (since these will come from adult plants or the seed bank), they are included because `offspringSplitter` is where IPMpack identifies all the existing discrete stages. The argument `fecByDiscrete` reflects the fact that none of the discrete classes addressed in this example are likely to directly produce offspring (which may not always be the case). The resulting fecundity object can be used with `createIPMFmatrix` in the usual way:

```
> Fmatrix <- createIPMFmatrix(fecObj = fv1, nBigMatrix = 5,
+                             minSize = min(dff$size, na.rm = TRUE),
+                             maxSize = max(dff$size, na.rm = TRUE),
+                             correction = "constant")
```

The user also needs a `Tmatrix` that reflects the same structure. The continuous part of the `T` matrix will be the standard structure:

```
> gr1 <- makeGrowthObj(dataf = dff,
+                      Formula = sizeNext~size)
> sv1 <- makeSurvObj(dff, Formula = surv~size)
```

Movement in and out of discrete stages is defined via an add-on of a transition matrix, that is defined using:

```
> discTrans <- makeDiscreteTrans(dff)
```

which captures survival and transitions between discrete stages and the continuous stage (note that this function will not work unless the data frame `dff` contains appropriate columns `stage` and `stageNext`). The user can then construct the `T` matrix:

```
> Tmatrix <- createIPMTmatrix(nBigMatrix = 5,
+                             growObj = makeGrowthObj(dff),
+                             survObj = makeSurvObj(dff),
+                             discreteTrans = discTrans,
+                             correction = "constant")
```

Note that both the `T` matrix and the `F` matrix in this example have a rather small number of bins just for ease of comparison, and that a higher number is almost certainly advisable. The user can examine both matrices:

```
> print(Tmatrix)
```

```

An object of class "IPMmatrix"
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 3.800000e-01 0.000000e+00 0.000000e+00 9.811801e-03 6.231452e-01
[2,] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
[3,] 0.000000e+00 4.439560e-01 4.323308e-01 0.000000e+00 0.000000e+00
[4,] 6.199776e-01 7.252743e-02 1.127820e-01 4.539650e-01 1.595343e-20
[5,] 2.236014e-05 4.046762e-08 3.182419e-15 1.513117e-25 2.556848e-04
[6,] 2.666217e-25 9.177937e-25 1.148429e-52 5.243367e-100 4.260333e-38
[7,] 1.051088e-60 8.460862e-52 5.300052e-114 1.889016e-224 7.380227e-122
[8,] 1.369952e-111 3.170416e-89 3.128132e-199 0.000000e+00 1.329179e-255
      [,6]      [,7]      [,8]
[1,] 1.040093e-01 2.981780e-04 2.808002e-08
[2,] 0.000000e+00 0.000000e+00 0.000000e+00
[3,] 0.000000e+00 0.000000e+00 0.000000e+00
[4,] 3.740428e-81 4.619418e-165 9.808780e-289
[5,] 2.882524e-24 1.711748e-67 1.747705e-150
[6,] 2.309470e-17 6.594469e-20 3.237495e-62
[7,] 1.923709e-60 2.641239e-22 6.235018e-24
[8,] 1.665919e-153 1.099824e-74 1.248401e-35
Slot "nDiscrete":
[1] 3

Slot "nEnvClass":
[1] 1

Slot "nBigMatrix":
[1] 5

Slot "meshpoints":
[1] 4.1 14.3 24.5 34.7 44.9

Slot "env.index":
[1] 1 1 1 1 1

Slot "names.discrete":
[1] "dormant" "seedAge1" "seedOld"

```

```
> print(Fmatrix)
```

```

An object of class "IPMmatrix"
      [,1] [,2] [,3]      [,4]      [,5]      [,6]      [,7]
[1,] 0 0 0 0.000000e+00 0.000000e+00 0.0000000000 0.0000000000
[2,] 0 0 0 1.418320e-01 2.471971e-01 0.4308362621 0.7508984252
[3,] 0 0 0 0.000000e+00 0.000000e+00 0.0000000000 0.0000000000
[4,] 0 0 0 5.818153e-03 1.014038e-02 0.0176735178 0.0308029241
[5,] 0 0 0 1.721265e-02 2.999969e-02 0.0522860340 0.0911285888
[6,] 0 0 0 1.090479e-02 1.900581e-02 0.0331249623 0.0577330281
[7,] 0 0 0 1.479429e-03 2.578476e-03 0.0044939898 0.0078325112
[8,] 0 0 0 4.298113e-05 7.491121e-05 0.0001305617 0.0002275541
      [,8]
[1,] 0.0000000000
[2,] 1.3087302407
[3,] 0.0000000000
[4,] 0.0536859807

```



```

[5,] 0.1588267280
[6,] 0.1006220778
[7,] 0.0136511729
[8,] 0.0003966009
Slot "nDiscrete":
[1] 3

Slot "nEnvClass":
[1] 1

Slot "nBigMatrix":
[1] 5

Slot "meshpoints":
[1] 0.0274954 2.3942312 4.7609670 7.1277028 9.4944386

Slot "env.index":
[1] 1 1 1 1 1

Slot "names.discrete":
[1] "dormant" "seedAge1" "seedOld"

```

and check for example that the slot `namesDiscrete` is aligned for both of them, and add them together:

```

> print(Tmatrix+Fmatrix)

      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 3.800000e-01 0.000000e+00 0.000000e+00 9.811801e-03 6.231452e-01
[2,] 0.000000e+00 0.000000e+00 0.000000e+00 1.418320e-01 2.471971e-01
[3,] 0.000000e+00 4.439560e-01 4.323308e-01 0.000000e+00 0.000000e+00
[4,] 6.199776e-01 7.252743e-02 1.127820e-01 4.597831e-01 1.014038e-02
[5,] 2.236014e-05 4.046762e-08 3.182419e-15 1.721265e-02 3.025537e-02
[6,] 2.666217e-25 9.177937e-25 1.148429e-52 1.090479e-02 1.900581e-02
[7,] 1.051088e-60 8.460862e-52 5.300052e-114 1.479429e-03 2.578476e-03
[8,] 1.369952e-111 3.170416e-89 3.128132e-199 4.298113e-05 7.491121e-05

      [,6]      [,7]      [,8]
[1,] 0.1040092920 0.0002981780 2.808002e-08
[2,] 0.4308362621 0.7508984252 1.308730e+00
[3,] 0.0000000000 0.0000000000 0.000000e+00
[4,] 0.0176735178 0.0308029241 5.368598e-02
[5,] 0.0522860340 0.0911285888 1.588267e-01
[6,] 0.0331249623 0.0577330281 1.006221e-01
[7,] 0.0044939898 0.0078325112 1.365117e-02
[8,] 0.0001305617 0.0002275541 3.966009e-04

```

The first three rows and columns concern transitions in and out of the discrete stages; the remainder are the usual T and F matrices describing moving across the continuous variables. The usual types of calculations (sensitivity via `sens`, life expectancy via `meanLifeExpect`, etc) can be applied here too.

## 5 Discretely varying environments

One extension of IPMs is to create a compound IPM matrix where, in addition to moving between continuous sizes, individuals move through discrete environments where the discrete environmental

states have an expected sequence, and therefore can be described by a transition matrix of their own (e.g. light environments for tropical trees, as in Metcalf et al. 2009).

To explore this type of dynamic, the user needs either to provide or simulate an environmental variable at  $t$  and the corresponding value at  $t + 1$ . Here, it has been generated as part of the `generateData` function (See above). From this generated data, the user can then create an environmental transition matrix that describes how the environment tends to move between these states from one census time to the next. If the data has been set up as described, there is a function that will do this for the user:

```
> dff <- generateData()
> env1 <- makeEnvObj(dff)
> env1
```

```
An object of class "envMatrix"
      [,1] [,2]
[1,] 0.219697 0.1923077
[2,] 0.780303 0.8076923
Slot "nEnvClass":
[1] 2
```

The user can now use IPMpack to create a survival-growth transition T matrix that encompasses movement across environments, first redefining the survival and growth objects to fit a discrete covariate, by changing the `Formula` argument:

```
> gr1 <- makeGrowthObj(dff, Formula = sizeNext~size+covariate)
> sv1 <- makeSurvObj(dff, Formula = surv~size+covariate)
```

Note that these functions will only work appropriately for a discrete covariate if the value of the covariate at time  $t$  is available as a column in the data frame names `covariate` and the value of the covariate at the next census is available as a column in the data frame called `covariateNext`. IPMpack functions use the presence of a column in the data frame called `covariate` as a cue to renumber values in these two columns to numeric levels between 1 and the observed number of covariate levels to facilitate looping, and changes them into factors. Once this step is implemented, the user can use these functions to create a compound T matrix, using `createCompoundTmatrix`:

```
> Tmatrix <- createCompoundTmatrix(nBigMatrix = 50, minSize = -5,
+                                 maxSize = 35,
+                                 envMatrix = env1, growObj = gr1,
+                                 survObj = sv1,
+                                 correction = "constant")
```

Essentially, the compound T matrix is a large matrix with stacked IPMs corresponding to each environment, modified to reflect movement between environmental states defined by `env1`. Passage time can be calculated using similar function, but now including the environmental matrix as an argument (equivalent life expectancy functions are in development):

```
> pTimes <- stochPassageTime(Tmatrix@meshpoints[15], Tmatrix, env1)
```

The resulting vectors contain the life expectancy and time to reach each size for individuals starting in each different environmental class, concatenated together (i.e. there are `nBigMatrix` values in the LE matrix ranging over the first environment, then `nBigMatrix` values ranging over the second environment, etc). The user can plot these against meshpoints (Figure 8 p. 19), each colour indicates a different starting environment. Similar syntax can be used for passage time (although note that here the function name has changed).

(Figure 8) :

Adding a fertility object to this analysis, the user can also define a full life cycle IPM model for across environments. With such information, obtaining the stochastic population rate of increase

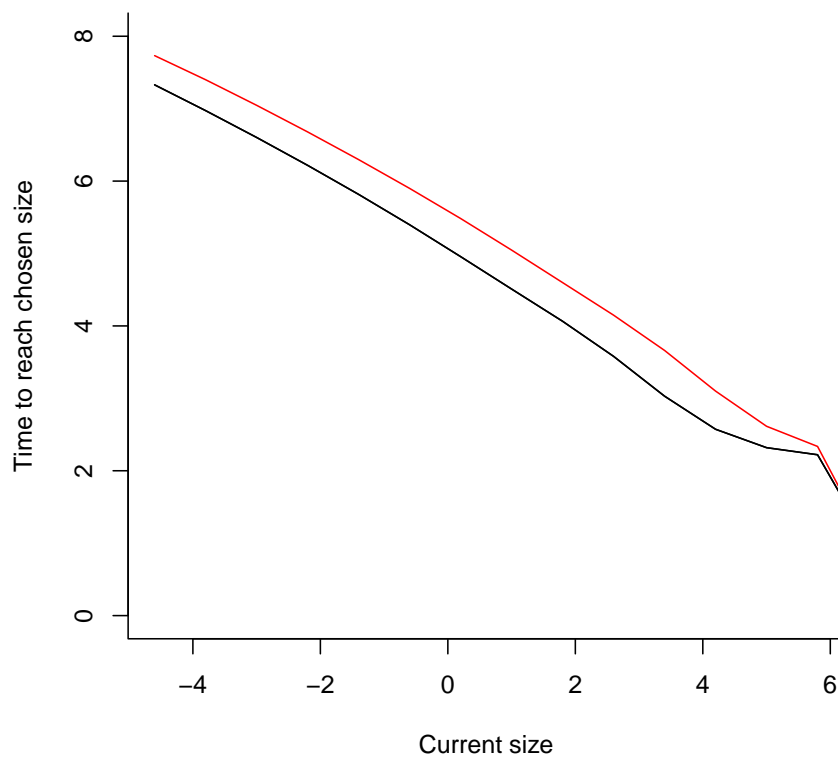


Figure 8: Passage time for a compound IPM; different colours reflect predictions for individuals starting in different environments

$\lambda_s$  in this environment is relatively straight-forward. IPMpack does this by sampling a very large number of environments and corresponding IPMs, and multiplying them together (Childs et al. 2004). At the moment, this is only defined for the case where environments (defined by the discrete covariates) are distributed independently (i.e. the next state does not depend on the previous state). To do this, the user must first define a list of IPMs (each the sum of a matrix of survival-growth transitions, and a matrix of fecundity transitions corresponding to a particular environment):

```
> IPMlist <- makeListIPMs(dataf = dff, nBigMatrix = 25, minSize = -5,
+                           maxSize = 35, explSurv = surv~size+covariate,
+                           explGrow = sizeNext~size+size2+covariate,
+                           explFec = fec~size, Transform="log",correction = "constant")

surv ~ size + covariate
```

Note that in this example IPMpack uses an arbitrary selection of formulas for the various linear and logistic regressions (`explGrow`, `explSurv`, etc). In reality, model selection will need to be used to establish this. Importantly, each environment type requires sufficient data (years of data) to quantify population behaviours in that environment as if it were an IPM in a single environment. Next, the user can estimate  $\lambda_s$  using:

```
> stochGrowthRateSampleList(listIPMmatrix = IPMlist,
+                           nRunIn = 30, tMax = 50)

[1] 1.217271
```

where `nRunIn` defines the number of time steps to discard from the start of the time series in order to remove transient dynamics, and `tMax` is the total number of time steps to run, and should be large enough that increasing it does not substantially change the result (numbers presented here for efficiency are almost certainly not large enough). Note that `stoch.growth.rate` in the package `popbio` will operate in essentially the same way, and provides more information, more efficiently.

## 6 More generally varying environments

An alternative way of inhabiting stochastic environments is to experience continuously changing covariates (rather than moving between discrete states, as the above describes). In this case, rather than building a single megamatrix, desired variables are obtained by multiplying up a suite of matrices and relying on the weak ergodic theorem for convergence (as described for obtaining  $\lambda_s$ , above). IPMpack contains code to do this. The user must first define a new data frame containing several time-varying covariates, and then, build the associated survival, growth and fertility objects (note that the default in IPMpack is to assume that if you have a covariate called "covariate", your aim is to build a compound matrix, so ideally other covariates should have other names):

```
> dff <- generateDataStoch()
> sv1 <- makeSurvObj(dataf = dff,
+                     Formula = surv~size+covariate1+covariate3)
> gr1 <- makeGrowthObj(dataf = dff,
+                       Formula = sizeNext~size+covariate1+covariate2)
> fv1 <- makeFecObj(dataf = dff, fecConstants = data.frame(1.8),
+                   Formula = fec~size, Transform = "log")
```

As before, the user can explore the data:

```
> head(dff)
```

	size	sizeNext	surv	covariate1	covariate2	covariate3	fec
1	6.372907	9.934122	1	1.09804577	-1.876444499	0.2881630	4.356188
2	1.652763	NA	0	-0.09542163	0.695057865	0.4865301	0.000000
3	5.742314	12.691594	1	2.08540290	-0.007290462	1.3259784	5.455125
4	3.396018	7.496433	1	1.10826541	0.805389712	-0.1088741	0.000000
5	NA	-2.110090	1	-0.39795950	2.309424450	-0.9718477	8.409157
6	1.928364	NA	0	-0.13496462	0.114439155	0.2268170	0.000000

	stage	stageNext	number
1	continuous	continuous	1
2	continuous	continuous	1
3	continuous	continuous	1
4	continuous	continuous	1
5	<NA>	continuous	1
6	continuous	continuous	1

and glance at the objects, e.g.,

```
> gr1
```

An object of class "growthObj"

Slot "fit":

Call:

```
lm(formula = Formula, data = dataf)
```

Coefficients:

(Intercept)	size	covariate1	covariate2
1.05611	0.88862	3.00186	0.01198

Slot "sd":

```
[1] 0.2204566
```

To explore demographic projections for this model, the user must decide on a time scale and time length, and define it by a vector called 'tVals'. This time series can be used to reflect patterns in environmental data that repeat. In this example we set tVals to reflect monthly intervals over 4 years, with years as the time scale and built covariates that vary seasonally, i.e., they fluctuates randomly around a sine wave which peaks once a year. From this simulation, the user can generate a matrix containing time as rows and different covariates in columns.

```
> tVals <- seq(1, 4, by = 1/12)
> covTest <- c(1 + 0.5*sin(2*pi*tVals))
> covMatTest <- data.frame(covariate1 = rnorm(length(covTest), covTest, 0.5) - 1,
+                           covariate2 = rnorm(length(covTest), covTest, 0.5) - 1,
+                           covariate3 = rnorm(length(covTest), covTest, 0.5) - 1)
```

Note that if there is no apparent temporal pattern to the data, one could simply generate random normal distributions of the covariates using their observed mean and variance. Other types of temporal patterns (multiannual, etc) are also possible. With this setup, the user can then estimate the stochastic growth rate over these years, using the geometric mean of the population growth rate (Tuljapurkar 1990; Childs et al. 2004), for these particular covariates using:

```
> r <- stochGrowthRateManyCov(covariate = covMatTest, nRunIn = 12*1,
+                             tMax = length(tVals), growthObj = gr1,
+                             survObj = sv1, fecObj = fv1, nBigMatrix = 20,
+                             minSize = 2*min(dff$size, na.rm = TRUE),
```

```

+               maxSize = 1.5*max(dff$size, na.rm = TRUE),
+               nMicrosites = 50, correction = "constant")
> print(r)

[1] -0.001814673

```

Setting `nRunIn = 12*1` in this example is equivalent to discarding the first year of the simulation (likely to contain transients) since the chosen time step is months. Note that in this formula, it was assumed that density-dependence acts on seedling establishment, and that 50 microsites are available for seedling establishment in every time step. Setting `nMicrosites = 0` allows for calculations without density-dependence, and `nMicrosites` can also be a vector, if the number of microsites fluctuates through time. It may also be interesting to have a glance at what has been happening to the population structure over this time-course, and the function `trackPopStructManyCov` allows this. `IPMpack` also contains a dedicated function to depict the results from this, `plotResultsStochStruct`.

## 7 Parameter uncertainty in a constant environment

To illustrate parameter uncertainty in a constant environment, we generate data again, and from these data build a list of survival and growth objects reflecting the parameter posteriors of fitted linear and logistic regression models (taking the simplest case of structure only via a continuous covariate). Note that it is also possible (and simpler) to directly use the variance-covariance of parameters estimates from a maximum-likelihood fit (via `glm` or `lm`) to build a list of growth, survival or fertility objects reflecting parameter uncertainty (see the help file for `getListRegObjects`) but here we illustrate the Bayesian case.

```

> dff <- generateData()
> grlist <- makePostGrowthObjs(dff,
+                             explanatoryVariables = "size",
+                             burnin = 100, nitt = 400)
> svlist <- makePostSurvivalObjs(dff,
+                                explanatoryVariables = "size",
+                                burnin = 100, nitt = 400)

```

This function currently only uses default non-informative priors features. There is a default thinning in `MCMCglmm`, the engine used to derive the posteriors, that reduces simulations by 10. So the number of samples from the posterior used in this simple example `nitt` is rather small (resulting in only  $[\text{nitt} - \text{burnin}] / 10 = 30$ ), and larger numbers are advisable (the default in `MCMCglmm` is 50,000). With output from this, the user can make a list containing multiple T matrices:

```

> TmatrixList <- makeListTmatrix(grlist, svlist, nBigMatrix = 20,
+                                minSize = -5,
+                                maxSize = 35,
+                                correction = "constant")

```

If one of the lists is longer than the other, this function samples the shorter object at random to reach the size of the longer object. Note that in this example the matrix size is rather small just to save time, and larger number of bins are advisable. The function will also construct compound matrices, if an environmental matrix is provided. With this, the user can now obtain posteriors for constant environment models.

```

> res <- getIPMoutput(TmatrixList, targetSize = 5, FmatrixList = NULL)
> names(res)

[1] "LE"          "pTime"       "lambda"      "stableStage"

```

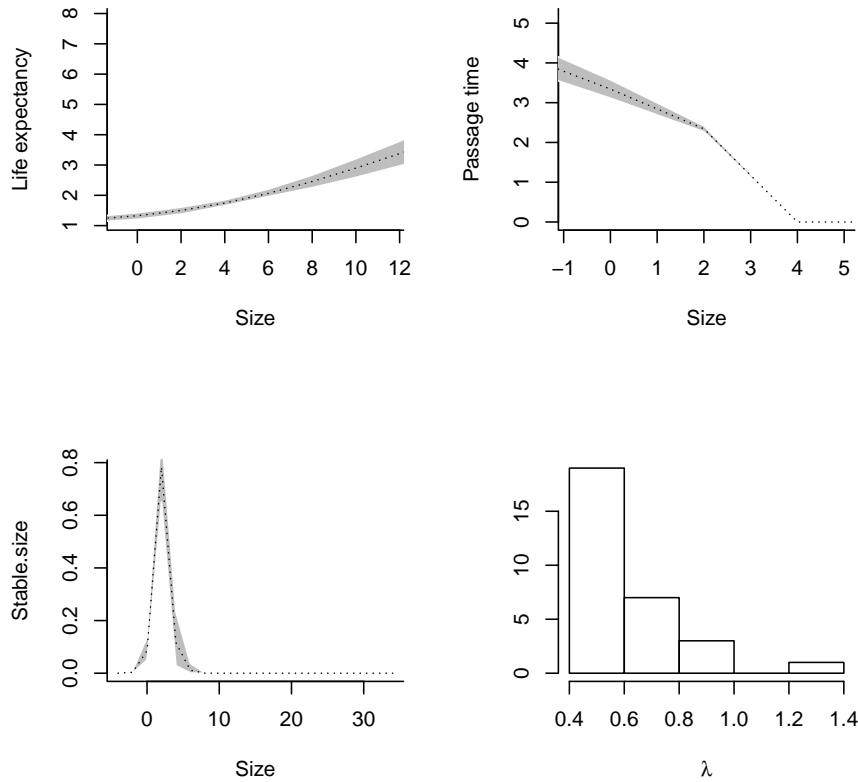


Figure 9: Uncertainty in IPM output

The vector called  $\lambda$  and matrix called `stableSize`, etc, will consist of NAs, unless a list of Fmatrices is also provided, which would allow a complete population projection matrix can be built. `IPMpack` contains a similar function to obtain a list of F matrices, and if such a list is included as the third argument into the function `getIPMOutput` (for which the default is 'NULL'), the function will also return distributions of  $\lambda$ , the stable stage distribution, etc:

```
> fv <- makePostFecObjs(dff, explanatoryVariables = "size+size2", fecConstants=data.frame(0.01),
+                       burnin = 100, nitt = 400, Transform = "log")
[1] 30

> FmatrixList <- makeListFmatrix(fv, nBigMatrix = 20, minSize = -5,
+                               maxSize = 35, cov = FALSE,
+                               correction = "constant")
> res <- getIPMOutput(TmatrixList, targetSize = 5, FmatrixList)
```

Again, larger number of iterations, binsize, etc, are recommended. These results can be visually inspected too (Figure 9 p. 23):

This is a rather slow way of proceeding - a large number of IPMs are being stored in R's memory. A slightly more rapid approach is to use the function `getIPMOutputDirect` that builds an IPM from a sample from the posterior, calculates relevant parameters, then over-writes this with a rebuilt IPM, iterating though the list.

## 8 Building your own objects and methods

What if growth is best reflected by a saturating function, rather than by the linear regression models provided? In this case the user may want to define a new class of growth object. An example of this follows:

```
> setClass("growthObjSaturate", representation(paras = "numeric", sd = "numeric"))
```

Then define the functional form of the mean prediction, with relevant parameters:

```
> fSaturate <- function(size, pars) {  
+   u <- exp(pmin(pars[1] + pars[2] * size, 50))  
+   u <- pars[3] * 1/(1+u)  
+   return(u)  
+ }
```

where the third parameter indicates the asymptotic size. The user can then estimate the parameters by fitting this function to the data using a wrapper function and the R function `optim`.

```
> wrapSaturate <- function(par, dataf) {  
+   pred <- fSaturate(dataf$size, par[1:3])  
+   ss <- sum((pred - dataf$sizeNext)^2, na.rm = TRUE)  
+   return(ss)  
+ }  
> tmp <- optim(c(1, 1, 1), wrapSaturate, dataf = dff, method = "Nelder-Mead")  
> tmp
```

```
$par  
[1] 2.065909 -0.382648 9.162594
```

```
$value  
[1] 506.219
```

```
$counts  
function gradient  
      222      NA
```

```
$convergence  
[1] 0
```

```
$message  
NULL
```

For simplicity, one can assume normally distributed errors:

```
> resids <- fSaturate(dff$size, tmp$par) - dff$sizeNext  
> sdSaturate <- sd(resids, na.rm = TRUE)
```

With these parameters, the user can then define the new growth object:

```
> gr1 <- new("growthObjSaturate")  
> gr1@paras <- tmp$par  
> gr1@sd <- sdSaturate
```

Finally, the user must define a method appropriate for this type of object.



```
> setMethod("growth", c("numeric", "numeric", "numeric", "growthObjSaturate"),
+           function(size, sizeNext, cov, growthObj){
+               mux <- fSaturate(size, growthObj@paras)
+               sigmax <- growthObj@sd
+               u <- dnorm(sizeNext, mux, sigmax, log = F)
+               return(u);
+           })

[1] "growth"
```

By putting `growthObjSaturate` in the signature, R will use this particular method for all objects with this signature. Now, the user can go ahead and use all the other code as previously, without a need for further definitions.

If the user wishes to fit a growth model with, for example, gamma errors, a similar approach can be used, but with ‘`dgamma`’ instead of `dnorm` in the last line of `growth` method, and appropriate slots defined in the object, etc.

## Selected References

- Caswell. 2001. Matrix population models: analysis, construction and interpretation. 2nd ed. Sinauer. Massachusetts, USA.
- Childs, Rees, Rose, Grubb & Ellner. 2004. Evolution of size-dependent flowering in a variable environment: Construction and analysis of a stochastic integral projection model. *Proc. Roy. Soc. Lond. Ser. B.* 271: 471–475.
- Cochran & Ellner. 1995. Simple methods for calculating age-based life history parameters for stage-structured populations. *Ecological Monographs* 62: 345–364.
- Ellner & Rees. 2006. Integral projection models for species with complex life-histories. *American Naturalist* 167: 410–428.
- Metcalf, Horvitz, Tuljapurkar & Clark. 2009. A time to grow and a time to die: a new way to analyze the dynamics of size, light, age and death of tropical trees. *Ecology* 90: 2766–2778.
- Rees & Rose. 2002. Evolution of flowering strategies in *Oenothera glazioviana*: an integral projection model approach. *Proc. Roy. Soc. Lond. Ser. B.* 269: 1509–1515.
- Ramula, Rees & Buckley. 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.
- Salguero-Gomez & Plotkin. 2010. Matrix dimensionality bias demographic inferences: implications for comparative plant demography. *The American Naturalist* 176: 710–722
- Tuljapurkar. 1990. Population Dynamics in Variable Environments. Springer. New York, USA.
- Zuidema, Jongejans, Chien, During & Schieving. 2010. Integral Projection Models for trees: a new parameterization and a validation of model output. *Journal of Ecology* 98: 345–355.