

SPI-IPM Code Manual

Chloé R. Nater

2021-10-04

Contents

About this manual	5
1 Preparing SPI-Birds data for Bayesian analysis	7
1.1 Nest count data	8
1.2 Clutch size data	8
1.3 Fledgling count data	8
1.4 Mark-recapture data	8
1.5 Immigrant count data	8
1.6 Auxiliary data on the sampling process	8
2 IPM Construction	9
2.1 Open population model with 2 age classes	9
2.2 Data likelihoods	11
2.3 Priors and constraints	15
3 Modelling temporal variation	17
3.1 Random year variation	17
3.2 Temporal covariates	17
3.3 Notes on covariate selection	17
4 IPM Implementation	19
4.1 Efficient implementation using NIMBLE	19
4.2 Simulation of initial values	19
4.3 Test runs and full runs: chains, iterations, burn-in, and thinning	19
4.4 Trouble-shooting implementation issues	19

5	Model Assessment	21
5.1	Assessing chain convergence	21
5.2	Plotting data vs. predictions	21
5.3	Comparing estimates from integrated vs. independent analyses	21
5.4	“Reality check” using stochastic simulations	21
5.5	Other approaches	21
6	Visualizing and interpreting direct IPM outputs	23
6.1	Population trajectories	23
6.2	Within-population variation in vital rates	23
6.3	Between-population variation in vital rates	23
6.4	Covariate effects	23
7	Follow-up Analyses	25
7.1	Testing for time-trends	26
7.2	Testing for density-dependence	26
7.3	Investigating cross-population covariation	26
7.4	Quantifying demographic contributions to short term population dynamics	26
7.5	Quantifying demographic contributions to long-term population trends	26
8	Useful extensions and outlook	27
8.1	Adapting the population model for your species/population	28
8.2	Including additional data and informative priors	28
8.3	Building on the multi-population perspective	28

About this manual

Briefly on the need for/value of standardized data and analyses.

Why IPMs are popular and what they are suitable for (Kéry and Schaub, 2011; Plard et al., 2019).

Overview over workflow, code repository & contents of manual (Figure 1).

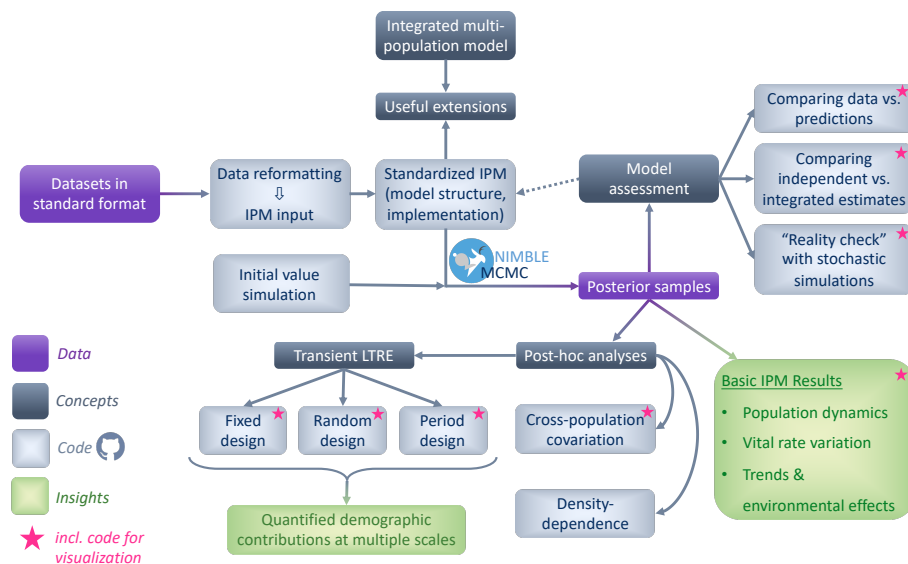


Figure 1: Schematic representation of the SPI-IPM workflow.

In no way complete: great if user's analyses/adaptations become part.

How to cite.

Chapter 1

Preparing SPI-Birds data for Bayesian analysis

Quick recap of SPI-Birds standard format and overview over data types used in IPM (+ how they relate, incl. diagram).

1.1 Nest count data

1.2 Clutch size data

1.2.1 Nest level

1.2.2 Population level

1.3 Fledgling count data

1.3.1 Nest level

1.3.2 Population level

1.4 Mark-recapture data

1.4.1 Individual capture histories

1.4.2 M-array

1.5 Immigrant count data

1.6 Auxiliary data on the sampling process

1.6.1 Nest survey sampling effort

1.6.2 Capture probability proxies

Chapter 2

IPM Construction

2.1 Open population model with 2 age classes

2.1.1 Model description

Population dynamics are represented using a female-based age-structured open population model with a pre-breeding census in spring. At census, females are divided into two age classes: “yearlings” (1-year old birds hatched during the breeding season of the previous year) and “adults” (birds older than one year). The motivation underlying this distinction is that reproductive output often differs for these two age classes in passerine birds. The dynamics of the female segment of the population over the time-interval from census in year t to census in year $t + 1$ can be described with classic matrix notation (Caswell, 2001) as:

$$N_{tot,t+1} = \begin{bmatrix} N_{Y,t+1} \\ N_{A,t+1} \end{bmatrix} = \begin{bmatrix} 0.5F_{Y,t}sJ_t & 0.5F_{A,t}sJ_t \\ sA_t & sA_t \end{bmatrix} \begin{bmatrix} N_{Y,t} \\ N_{A,t} \end{bmatrix} + \begin{bmatrix} Imm_{Y,t+1} \\ Imm_{A,t+1} \end{bmatrix}$$

$N_{tot,t+1}$ represents the total number of yearling and adult females in the population upon arrival in the breeding areas in year t . The total female population size, $N_{tot,t+1}$, is the sum of the numbers of yearling and adult females in the population in year $t + 1$ ($N_{Y,t+1}$ and $N_{A,t+1}$, respectively) and consists of local survivors and recruits from the previous breeding season, as well as immigrant yearling ($Imm_{Y,t+1}$) and adult ($Imm_{A,t+1}$) females.

$F_{a,t}$ represents the expected number of fledglings produced by age class a females during the breeding season in year t and is the product of several vital rates.

First, females in age class a may breed in a nestbox with probability $pB_{a,t}$ upon arrival to breeding areas in year t . Each breeding female may then lay a clutch containing a certain number of eggs (expected number = $CS_{a,t}$), and each egg within the clutch may hatch and survive to fledging. The probability of an egg hatching and surviving to fledging is divided into an age-independent probability of nest success (pNS_t , probability of complete clutch failure = $1 - pNS_t$) and a survival probability of every egg/chick to fledging provided that the nest has not failed entirely ($sN_{a,t}$, with a = age of the mother). Consequently, the expected number of fledglings produced by age class a females in year t is defined as:

$$F_{a,t} = pB_{a,t} \times CS_{a,t} \times pNS_t \times sN_{a,t} \quad (2.1)$$

Fledglings that survive to the next breeding season and remain within the population (probability = sJ_t) contribute to next year's yearling class ($N_{Y,t+1}$). Yearlings and adults that survive to the next breeding season and remain within the population (probability = sA_t) become part of next year's adult age class ($N_{A,t+1}$).

2.1.2 Code implementation including demographic stochasticity

Population process models within IPMs are typically implemented as stochastic models that account for randomness in the outcomes of demographic processes at the individual level ("demographic stochasticity", Caswell, 2001; Kéry and Schaub, 2011). The model described here is no different, meaning that the numbers of breeders, fledglings, and survivors are treated as binomial and Poisson random variables.

Reproduction is modelled via two sets of random variables: a binomial random variable representing the number of breeders in age class a in year t , $B_{a,t}$ and a Poisson random variable representing the number of fledglings produced by breeders of age class a in year t , $Juv_{a,t}$. The implementation in the BUGS language used in the SPI-IPM code (IPMSetup.R, lines 231-241) looks like:

```
for (t in 1:Tmax){
  for(a in 1:A){

    ## 1) Breeding decision
    B[a,t] ~ dbin(pB[a,t], N[a,t])

    ## 2) Offspring production
    Juv[a,t] ~ dpois(B[a,t]*CS[a,t]*pNS[t]*sN[a,t]*0.5)
  }
}
```

In the code, the age indices $a = 1$ and $a = A = 2$ correspond to yearlings and adults, respectively.

Analogous to breeders, the numbers of local survivors – both fledglings surviving their first year and becoming yearlings, and yearlings and adults surviving to the next year – are implemented as binomial random variables (IPMSetup.R, lines 243-255):

```
for (t in 1:(Tmax-1)){

  ## 3) Annual survival of local birds
  # Juveniles -> Yearlings
  localN[1,t+1] ~ dbin(sJ[t], sum(Juv[1:A,t]))
  # Yearlings/Adults -> adults
  localN[2,t+1] ~ dbin(sA[t], sum(N[1:A,t]))

  ## 4) Immigration
  for(a in 1:A){
    N[a,t+1] <- localN[a,t+1] + Imm[a,t+1]
  }
}
```

Immigrant numbers are also treated as outcomes of stochastic processes, and these are detailed in 2.2.5 Immigrant count data likelihood.

2.2 Data likelihoods

IPMs obtain information on the population model's parameters (population sizes and vital rates) from several different data sets. Information in each data set is channeled into model parameters via one or multiple data likelihoods.

The SPI-IPM contains five data modules consisting of a total of eight data likelihoods: nest count data (one likelihood), clutch size data (two likelihoods), fledgling count data (three likelihoods), mark-recapture data (one likelihood), and immigrant count data (one likelihood). The likelihoods contained in each data module are described in detail in the following sub-chapters. The underlying data sets are introduced in Chapter 1 of the manual.

2.2.1 Nest count data likelihood

The population model defines the true size of the female segment of the breeding population in any year t via the year- and age-specific number of breeding females:

$$B_{a,t} \sim \text{Binomial}(N_{a,t}, pB_{a,t}) \quad (2.2)$$

The total size of the female breeding population (the sum of $B_{a,t}$ over all age classes) in year t is expected to correspond closely with the observed number of first clutches laid in any year ($NestCount_t$). The breeding population count thus contains information about both breeding probability ($pB_{a,t}$) and population size ($N_{a,t}$) and its likelihood can be defined as:

$$NestCount_t \sim Poisson((B_{Y,t} + B_{A,t}) \times NS_Data_t) \quad (2.3)$$

Here, the year-specific variable NS_Data_t is a correction factor introduced for dealing with years in which no nest survey data was collected ($NestCount_t = 0$ due to lack of sampling). NS_Data_t is set to 0 for years without data collection, and to 1 in all other years.

The likelihood is written almost identically in the code, the only difference being the use the `sum()` function over age classes 1 to $A = 2$ when specifying the Poisson distribution (this provides more flexibility for extending the model to more than two age classes):

```
for(t in 1:Tmax){
  NestCount[t] ~ dpois(sum(B[1:A,t])*NS_Data[t])
}
```

2.2.2 Clutch size data likelihoods

The counting of incubated eggs provides information about both individual-level clutch size ($CS_{a,t}$) and reproductive output at the population level. Consequently, two separate likelihoods can be specified for within the clutch size data module.

Each individual clutch size observation can be treated as the outcome of a Poisson process with an expected value of $CS_{a,t}$ (where a is the age of the female that laid the clutch, and t the year in which the clutch was laid). In the code, the likelihood is formulated for each clutch size observation x (out of a total of `CS_X` observations) and includes nested indexing of the expectation using data on female age (`CS_FAge`) and year (`CS_year`):

```
for(x in 1:CS_X){
  ClutchSize[x] ~ dpois(CS[CS_FAge[x], CS_year[x]])
}
```

Since both year and female age need to be part of the provided data, the individual-level clutch size likelihood can only be used with complete observations, i.e. clutches for which both number of eggs and age of the mother are known (year should always be known).

For the population-level likelihood, on the other hand, data on clutch sizes can be included irrespective of whether the age of the mother is known. The total number of eggs counted in all nests laid in year t can be described as:

$$EggNoTot_t \sim \text{Poisson}(\text{sum}(B_{Y:A,t} \times CS_{Y:A,t}) \times p_t^{EggNo}) \quad (2.4)$$

Since the product $B_{a,t}CS_{a,t}$ corresponds to all eggs laid in all nests/nestboxes within the study site, another correction factor (p_t^{EggNo}) is needed to account for the fact that eggs may not have been counted in all surveyed nests with breeding activity in each year. p_t^{EggNo} (`EggNoSP[t]` in code) thus contains information of the year-specific proportion of surveyed nests for which the numbers of eggs were counted.

In code, the likelihood for the number of eggs at the population is written as:

```
for(t in 1:Tmax){
  # Expected "true" egg number (by mother age class)
  EggNo.ex[1:A,t] <- B[1:A,t]*CS[1:A,t]

  # Observed egg number (corrected by data availability)
  EggNoTot[t] ~ dpois(sum(EggNo.ex[1:A,t])*EggNoSP[t])
}
```

2.2.3 Fledgling count data likelihoods

Analogous to observations of clutch size, counts of fledglings contain information on reproduction at both individual and population level. Distributions of number of fledglings produced from a clutch are often 0-inflated because incidents of harsh weather, predation, abandonment, etc. may result in complete loss of entire clutches. To account for this, we split data on fledgling numbers and formulated separate likelihoods for the survival of the clutch as a whole (probability of a nest not failing completely, pNS_t) and for each chick subsequently surviving to fledgling (probability $sN_{a,t}$).

Whether or not a clutch succeeded (i.e. at least one chick survived to fledging, `anyFledged`) was coded using 1 (success) and 0 (failure) and modelled as the outcome of a Bernoulli process with a year-dependent success probability pNS_t :

```
for(x in 1:F_X){
  anyFledged[x] ~ dbern(pNS[F_year[x]])
}
```

For the subset of successful clutches, the number of fledglings produced (`NoFledged`) was modelled as a binomial random variable with each egg laid

in the clutch (NoLaid) having a probability of $sN_{a,t}$ (where a = age of the mother) to survive and fledge:

```
for(x in 1:NoF_X){
  NoFledged[x] ~ dbin(sN[NoF_FAge[x], NoF_year[x]], NoLaid[x])
}
```

At the population level, both processes (nest success and survival to fledging conditional on nest success) were combined into a single Poisson likelihood describing the total number of fledglings counted in the population in a given year t as the product of the number of breeding females, clutch size, nest success, and survival to fledging:

$$FledgedTot_t \sim \text{Poisson}(\text{sum}(B_{Y:A,t} \times CS_{Y:A,t} \times pNS_t \times sN_{Y:A,t}) \times p_t^{Fledged}) \quad (2.5)$$

As in the clutch size data module (Chapter 2.2.2), a correction factor quantifying the proportion of nests for which data is available ($p_t^{Fledged}$, FledgedSP[t] in code) is used to account for missing records of fledgling numbers.

In the code, calculation of the expected number of fledglings is based on the expected number of eggs as calculated in the clutch size data module (Chapter 2.2.2):

```
for(t in 1:Tmax){
  # Expected "true" fledgling number (by mother age class)
  Fledged.ex[1:A,t] <- EggNo.ex[1:A,t]*pNS[t]*sN[1:A,t]

  # Observed egg number (corrected by data availability)
  FledgedTot[t] ~ dpois(sum(Fledged.ex[1:A,t])*FledgedSP[t])
}
```

2.2.4 Mark-recapture data likelihood

In the SPI-IPM, capture histories of marked birds are analysed using an age-specific Cormack-Jolly-Seber (CJS) model ((Cormack, 1964; Jolly, 1965; Seber, 1965)), which allows estimation of parameters associated with both annual survival and the (re-)capturing process.

The survival parameters sJ_t and sA_t describe the probability of surviving from one breeding season to the next for fledglings/juveniles and adult females, respectively. Since field based sexing fledglings is not possible in the hatch year, first-year survival in the CJS likelihood was set to $0.5sJ_t$, representing the probability that a fledgling is female and survives (recaptures of adult males were

removed from the capture histories).

The time- and age-dependent recapture parameters $p_{Age,t}^{Recap}$ describe the probability that an age class a individual is recaptured and identified during the breeding season of year t given that it survived the time interval $t - 1$ to t . For many of the bird populations included in SPI-Birds, (re-)captures of birds are carried out at the nestboxes during the breeding season only. Consequently, two conditions need to be met for a (re-)capture besides the individual being alive: 1) the individual breeds in a nestbox (as non-breeders and birds breeding in natural cavities are not captured) and 2) the individual is actually captured and marked/identified while breeding in a nestbox. The recapture probability is therefore the product of the probability of breeding in a nestbox ($pB_{Age,t}$) and the probability of capture and identification given breeding in a nestbox (p_t^{CapB}): $p_{Age,t}^{Recap} = pB_{Age,t} \times p_t^{CapB}$. The parameters $pB_{Age,t}$ and p_t^{CapB} are confounded, and auxiliary data on one of them is required to separately estimate them. Here, raw data on the annual proportions of surveyed nests for which the identity of the breeding female was recorded (as a result of ringing or recapture) is used to approximate p_t^{CapB} , allowing the CJS model to estimate age- and time-specific breeding probabilities ($pB_{Age,t}$).

CJS models in Bayesian frameworks have traditionally been implemented either as latent states models with Bernoulli likelihoods or as multinomial likelihoods for data re-formatted as 'm-arrays' (Gimenez et al., 2007; Kéry and Schaub, 2011). The former typically results in long runtimes and high computational costs, while the improved efficiency of the latter is tied to an implementation that suffers from being neither particularly intuitive nor user-friendly. To avoid these pitfalls, the CJS model in the SPI-IPM is implemented using a marginalized likelihood (that integrates over all latent states) applied to unique (not individual) capture histories only (following Turek et al., 2016):

```
## Likelihood with custom distribution
for (i in 1:n.CH){

  y.sum[i, first.sum[i]:last.sum[i]] ~ dCJS_vv_sum(
    probSurvive = phi.CH[i, first.sum[i]:last.sum[i]],
    probCapture = p.CH[i, first.sum[i]:last.sum[i]],
    len = last.sum[i]-first.sum[i]+1,
    mult = CHs.count[i])
}
```

2.2.5 Immigrant count data likelihood

2.3 Priors and constraints

Chapter 3

Modelling temporal variation

3.1 Random year variation

3.2 Temporal covariates

3.2.1 Continuous variables

3.2.2 Categorical variables

3.2.3 Imputation of missing covariate values

3.3 Notes on covariate selection

Chapter 4

IPM Implementation

4.1 Efficient implementation using NIMBLE

We use the fantastic `nimble` package (de Valpine et al., 2017)!

4.2 Simulation of initial values

4.3 Test runs and full runs: chains, iterations, burn-in, and thinning

4.4 Trouble-shooting implementation issues

Chapter 5

Model Assessment

5.1 Assessing chain convergence

5.2 Plotting data vs. predictions

5.3 Comparing estimates from integrated vs. independent analyses

5.4 “Reality check” using stochastic simulations

5.5 Other approaches

Running for additional years and comparing to non-included data, PPCs, etc.

Chapter 6

Visualizing and interpreting direct IPM outputs

6.1 Population trajectories

6.2 Within-population variation in vital rates

6.2.1 Age-class-specific averages

6.2.2 Year-by-year variation

6.3 Between-population variation in vital rates

6.3.1 Population-specific averages

6.3.2 Year-by-year variation

6.4 Covariate effects

Chapter 7

Follow-up Analyses

7.1 Testing for time-trends

7.2 Testing for density-dependence

7.3 Investigating cross-population covariation

7.4 Quantifying demographic contributions to short term population dynamics

7.4.1 Year-by-year variation in population growth rate (random design LTRE)

7.4.2 Year-to-year differences in population growth rate (fixed design LTRE)

7.5 Quantifying demographic contributions to long-term population trends

7.5.1 Differences in population trajectories between time periods (period design LTRE)

7.5.2 Differences in population trajectories between locations (period design LTRE with time-by-space substitution)

Chapter 8

Useful extensions and outlook

8.1 Adapting the population model for your species/population

8.1.1 Accounting for multiple broods per bird per year

8.1.2 Altering age structure

8.1.3 Individual heterogeneity beyond age: sex, traits, and more

8.2 Including additional data and informative priors

8.2.1 Including partially observed age information

8.2.2 Making the most of auxiliary knowledge about immigrants/dispersers

8.2.3 Letting published values help with estimation when data is sparse

8.3 Building on the multi-population perspective

8.3.1 Joint analysis of data from several populations

8.3.2 Modelling cross-population covariation

8.3.3 Estimating hyper-parameters in large-scale analyses

8.3.4 Unlocking the secrets of dispersal

Bibliography

- Caswell, H. (2001). *Matrix population models: construction, analysis, and interpretation*. Sunderland, Mass.: Sinauer Associates.
- Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, 51(3/4):429–438.
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., and Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with nimble. *Journal of Computational and Graphical Statistics*, 26(2):403–413.
- Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J.-P., and Pradel, R. (2007). State-space modelling of data on marked individuals. *Ecological Modelling*, 206(3-4):431–438.
- Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52(1/2):225–247.
- Kéry, M. and Schaub, M. (2011). *Bayesian population analysis using WinBUGS: a hierarchical perspective*. Academic Press.
- Plard, F., Fay, R., Kéry, M., Cohas, A., and Schaub, M. (2019). Integrated population models: powerful methods to embed individual processes in population dynamics models. *Ecology*, page e02715.
- Seber, G. A. (1965). A note on the multiple-recapture census. *Biometrika*, 52(1/2):249–259.
- Turek, D., de Valpine, P., and Paciorek, C. J. (2016). Efficient markov chain monte carlo sampling for hierarchical hidden markov models. *Environmental and ecological statistics*, 23(4):549–564.