# The multi-state Jolly-Seber model

William Kendall, *USGS Colorado Cooperative Fish & Wildlife Research Unit*

In Chapter 10 the multi-state model was presented, conditioning on the first capture of an individual and then tracking its fate and detections through time to estimate state-specific survival probabilities and probabilities of transition among phenotypic states, locations, sub-populations, disease states, etc; all of this while accounting for state-specific detection probabilities. In Chapter 12, Jolly-Seber models were presented. Under the **POPAN** version of this model (Schwarz & Arnason 1996; Chapter 12.3.2), for each individual first detection probabilities are modeled as well as subsequent detections. Part of modeling that first detection is to also model first entry probabilities, which could represent births, immigration, or some other kind of first arrival. Other parameters estimated are total number of individuals in the population during the study, period-specific abundances, and period-specific numbers of new recruits. In this chapter we present how to implement a combination of these two model structures, a multi-state Jolly-Seber (MSJS) model (similar to Dupuis & Schwarz 2007).

14.1. Sampling protocol and model structure

The idea behind this model is illustrated in the diagram shown at the top of the next page. It represents a system with two states, **A** and **B**, modeled over three time periods. N^* individuals will enter the system over those three time periods. Arrows between states represent survival and state transitions and arrows pointing out of the system represent mortality (or permanent emigration from the system). Individuals from this superpopulation recruit to state **A** or **B** with probability π or $(1 - \pi)$, respectively. Those that will recruit to state **A** do so in sampling periods 1, 2, or 3 with probability $pent_1^{\mathbf{A}}$, $pent_2^{\mathbf{A}}$, and $pent_3^{\mathbf{A}}$, respectively, with a similar pattern for state **B**. For each state the *pent*'s sum to 1.0.

For the survival, transition, and detection processes, in addition to state and time dependence we incorporate age (length of time since the individual has recruited to the system) in addition to time. So individuals in state **A** who have been in the system for (a) sampling periods to date survive with probability $S_t^{\mathbf{A}(a)}$, and given survival, remain in state **A** with probability $\psi_t^{\mathbf{AA}(a)}$ or transition to state **B** with probability $\psi_i^{\mathbf{AB}(a)}$. Individuals in state **B** survive with probability $S_t^{\mathbf{B}(a)}$ and given survival, remain in state **B** with probability $\psi_t^{\mathbf{BB}(a)}$ or transition to state **A** with probability $\psi_t^{\mathbf{BA}(a)}$. Finally, for those in state **A** or **B** during sampling period (t), which have been in the system for (a) sampling periods, the probability of detection is $p_t^{\mathbf{A}(a)}$ or $p_t^{\mathbf{B}(a)}$, respectively. Full age and time structure for a parameter would be overparameterized, but this general structure at least allows for restrictive models that incorporate both effects (e.g., survival as a general function of time period and a linear function of time since arrival).

Rather than having N^* as a parameter in the model (i.e., in the model likelihood), we instead condition on those individuals n^* that are detected at least once during the study, with probability p^*, which is a function of all the model parameters listed above.

There are two principal contexts in which this model could be useful. The first is the case where time periods represent seasons or years, and therefore the interpretations of the parameters are as stated, where S denotes survival, *pent* denotes some kind of recruitment, and N represents annual abundance. The second context is where data are collected across time periods within a season of interest (e.g., at migratory stopover areas, at multiple sea turtle nesting beaches), so that N^* refers to the total number of individuals that used the multiple sites in that season, π is the probability that, e.g., for that season a sea turtle will deposit her first clutch of eggs at beach **A**; $pent_t^{\mathbf{A}}$ refers to the probability an individual that will deposit that first clutch during sampling period (t), $S_t^{\mathbf{A}(a)}$ is the probability that a female that has deposited a clutch at beach **A** during sampling period (t), and has deposited a previous clutches that season, persists in the system to deposit another clutch during sampling period ($t+1$), and $\psi_t^{\mathbf{AB}(a)}$ is the probability that next clutch will be deposited at beach **B**. This scenario is an extension of the within-season sampling and modeling described in the multi-state open robust design model (Chapter 16; Kendall *et al*. 2019).

14.1.1. Data

The data structure for this model is identical to the multi-state model of Chapter 10. For each of K sampling periods detections are denoted by a state-specific code (e.g., **A**, **B**) and non-detections by a '0'. So, example detection histories for the 2-state, 3-period study might be 'A0B', '0BB', etc. The model structure for these two example histories is shown in the table at the top of the next page.

The expression for p^* is based on summing the probabilities of being detected for the first time in each time period. It gets complex quickly with more time periods and benefits from the use of matrix multiplication. Below is the case for just two time periods:

$$p^* = \left[\pi \cdot pent_1^A p_1^A + (1-\pi)pent_1^B p_1^B\right] + \left[\pi\{pent_1^A(1-p_1^A)S_1^A(\psi_1^{AA}p_2^A + \psi_1^{AB}p_2^B) + pent_2^A p_2^A\}\right.$$
$$\left. + (1-\pi)\{pent_1^B(1-p^B)S_1^B(\psi_1^{BA}p_2^A + \psi_1^{BB}p_2^B) + pent_2^B p_2^B\}\right]$$

history	probability expression
A0B	$\left(\pi\left(pent_1^{\mathbf{A}}\right)p_1^{\mathbf{A}(1)}\left[\psi_1^{\mathbf{AA}(1)}\left(1-p_2^{\mathbf{A}(2)}\right)S_2^{\mathbf{A}(2)}\psi_2^{\mathbf{AB}(2)}+\psi_1^{\mathbf{AB}(1)}\left(1-p_2^{\mathbf{B}(2)}\right)S_2^{\mathbf{B}(2)}\psi_2^{\mathbf{BB}(2)}\right]p_3^{\mathbf{B}(3)}\right)/p^*$
0BB	$\left(\left[\pi\left(pent_1^{\mathbf{A}}\right)\left(1-p_1^{\mathbf{A}(1)}\right)S_1^{\mathbf{A}(1)}\psi_1^{\mathbf{AB}(1)}+(1-\pi)pent_1^{\mathbf{B}}\left(1-p_1^{\mathbf{B}(1)}\right)S_1^{\mathbf{B}(1)}\psi_1^{\mathbf{BB}(1)}\right]p_2^{\mathbf{B}(2)}S_2^{\mathbf{B}(2)}\psi_2^{\mathbf{BB}(2)}p_3^{\mathbf{B}(3)}\right.$ $\left.+(1-\pi)pent_2^{\mathbf{B}}p_2^{\mathbf{B}(1)}S_2^{\mathbf{B}(1)}\psi_2^{\mathbf{BB}(1)}p_3^{\mathbf{B}(2)}\right)/p^*$

14.1.2. Estimation

Under this multinomial model we can use maximum likelihood to estimate most parameters of the model. However, as with the Jolly-Seber model, there is confounding of parameters at the beginning and end of the times series. Detection probabilities for time periods 1 and K cannot be estimated without constraints. Estimation of detection probability relies on marked individuals and therefore parameter p_1^s for state s cannot be estimated. In addition, p_K^s is confounded with S_{K-1}^s and ψ_{K-1}^{rs}. In addition, as with the Jolly-Seber model, it's assumed that unmarked and previously marked individuals have the same capture probability.

14.1.3. Derived Parameters

There are a number of derived parameters from **MARK** for the MSJS model, whose variance is developed via the Delta method (Appendix B).

- **Total abundance**

  The first is the total number of individuals that are in the system at some time during the study, N^*. The value is not of interest in many cases, especially for a multi-year study. However, this model can be applied to an open multi-state process within a season, such as sea turtles arriving at and moving among various nesting beaches, where N^* refers to the total number of nesters. Regardless, the derivation of this estimator is

$$\hat{N}^* = \frac{n^*}{\hat{p}^*},$$

  where n^* is the total number of individuals detected during the study.

- **State-specific, period-specific abundance**

  Method 1

  Like total abundance, when all states are observable abundance for each state (s) in sampling period (t) can be derived by inflating the count based on detection probability,

$$\hat{N}_t^s = \frac{n_t^s}{\hat{p}_t^s},$$

  With total abundance for period (t) derived from the sum over all M states

$$\hat{N}_t = \sum_{s=1}^{M} \hat{N}_t^s.$$

Method 2

An alternative approach is to derive period- and state-specific abundance from total abundance and then the dynamics of the system thereafter. The general expression is in terms of matrix multiplication (Kendall *et al.*, in prep) but an example would be (ignoring age-dependence for simplicity):

$$\hat{N}_2^{\mathbf{B}} = \hat{N}^* \left[\hat{\pi} \, \widehat{pent}_1^{\mathbf{A}} \hat{S}_1^{\mathbf{A}} \hat{\psi}_1^{\mathbf{AB}} + (1 - \hat{\pi}) \left\{ \widehat{pent}_1^{\mathbf{B}} \hat{S}_1^{\mathbf{B}} \hat{\psi}_1^{\mathbf{BB}} + \widehat{pent}_2^{\mathbf{B}} \right\} \right].$$

This approach is useful where there are unobservable states, and therefore Method 1 is not an option.

- **Phenology parameters**

Three additional derived parameters are computed which are most meaningful for the case where the data are collected within a given season. These are expected period of first ingress to the system, expected period of last egress from the system, and expected time of residence in the system. To explain the derivation of these parameters we'll use the simple case of three sampling periods and two states **A** and **B**.

Expected Ingress and Egress Periods

Expected period of arrival to the system (*a*) would be an average weighted by the probability of first arrival, in either state:

$$a = 1 \cdot \left(\pi \, pent_1^{\mathbf{A}} + (1 - \pi) \, pent_1^{\mathbf{B}} \right) + 2 \cdot \left(\pi \, pent_2^{\mathbf{A}} + (1 - \pi) \, pent_2^{\mathbf{B}} \right) + 3 \cdot \left(\pi \, pent_3^{\mathbf{A}} + (1 - \pi) \, pent_3^{\mathbf{B}} \right).$$

Expected departure period *g* (meaning, departure immediately after sampling period *g*) involves a more complicated weighted average accounting for arrival and then departure – for example the case of two time periods:

$$\begin{aligned} g = {} & 1 \cdot \left[\pi \, pent_1^{\mathbf{A}} \left(1 - S_1^{\mathbf{A}} \right) + (1 - \pi) \, pent_1^{\mathbf{B}} \left(1 - S_1^{\mathbf{B}} \right) \right] \\ & + 2 \cdot \left[\pi \left\{ pent_1^{\mathbf{A}} S_1^{\mathbf{A}} \left(\psi_1^{\mathbf{AA}} \left(1 - S_2^{A} \right) + \psi_1^{\mathbf{AB}} \left(1 - S_2^{B} \right) \right) + pent_2^{\mathbf{A}} \left(1 - S_2^{A} \right) \right\} \\ & + (1 - \pi) \left\{ pent_1^{\mathbf{B}} S_1^{\mathbf{B}} \left(\psi_1^{\mathbf{BA}} \left(1 - S_2^{\mathbf{A}} \right) + \psi_1^{\mathbf{BB}} \left(1 - S_2^{\mathbf{B}} \right) \right) + pent_2^{\mathbf{B}} \left(1 - S_2^{\mathbf{B}} \right) \right\} \right]. \end{aligned}$$

- **Expected Residence Time**

There are two approaches to estimating the expected number of time periods an individual spends in residence in the system (e.g., in either or both states).

Method 1

The first is easily derived once *a* and *g* are estimated, and is essentially the difference between the two:

$$(g_i + 1) - a_i.$$

Method 2

The second is applicable only where persistence probability $S_t^{s(a)}$ is independent of state and time, and only dependent on the number of periods since first arrival. As in Kendall *et al.* (2019) expected residence time for a three-period study would be:
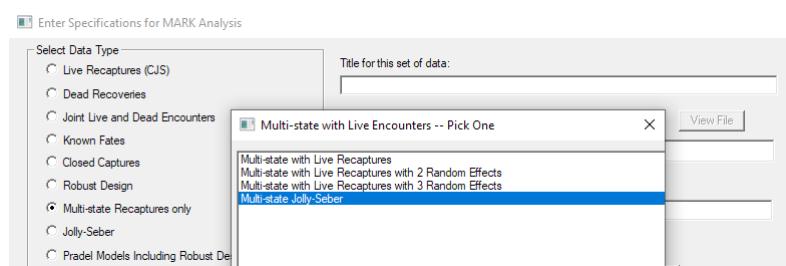
$$\left[1 \times \left(1 - S_i^{\cdot(0)} \right) \right] + \left[2 \times \left(S_i^{\cdot(0)} \right) \left(1 - S_i^{\cdot(1)} \right) \right] + \left[3 \times \left(S_i^{\cdot(0)} \right) \left(S_i^{\cdot(1)} \right) \right].$$

14.2. Implementation of the MSJS model in MARK

To demonstrate implementation of the model we'll use the example of a fish species that exist in two different creeks (**N** and **T**). It's suspected that there is some movement between creeks from year to year, and interest is in annual abundance for each creek, as well was survival and movement. A sample of the data (the complete encounter history data are contained in **fish_stream.inp**) over seven years is

```
0T00000 1;
0T00N00 1;
0T00000 1;
0T0T000 1;
0N00000 1;
0T00000 1;
0T00000 1;
0TT0000 1;
```
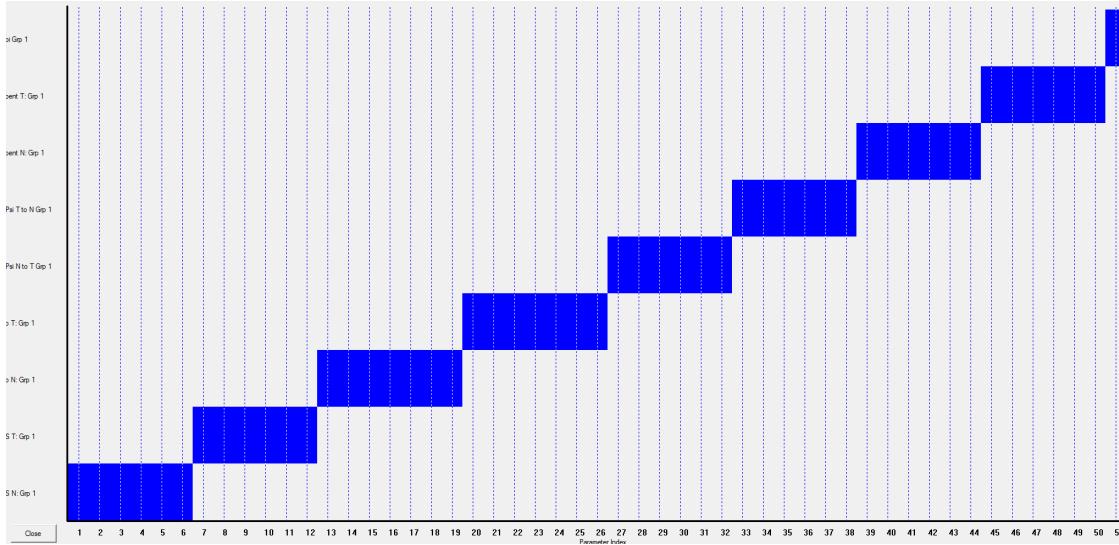
In **MARK**, the MSJS model is listed with the other multi-state models:

The specification screen for the MSJS model is identical to the MS model (chapter 10):
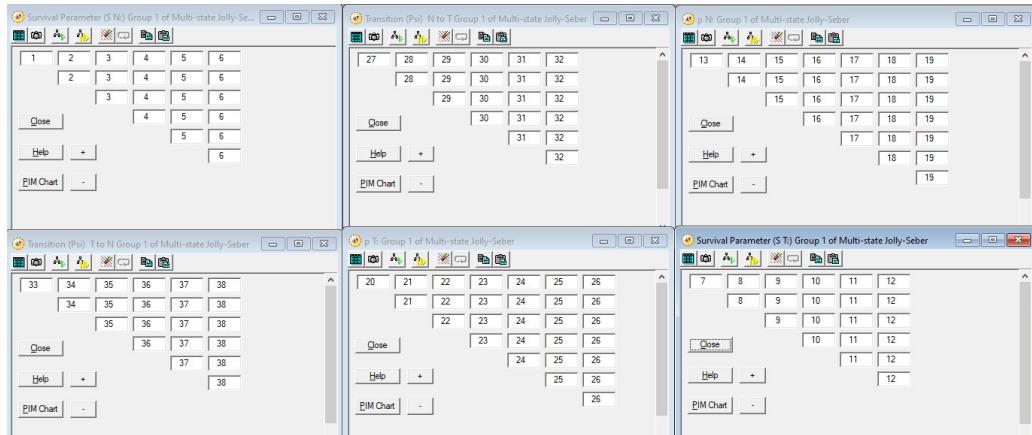
Looking at the PIM chart for this 2-state, 7-period analysis (below), you can see the various PIMs for this model.

Consistent with the parameter description above there are two PIMs, one for each state, for survival (S), detection (p), state transitions (ψ), and recruitment or ingress ($pent$). Notice also from the chart the number of parameters in each PIM, given there are 7 sampling occasions. For p there are 7 columns, one for each occasion. For S and ψ there are 6 columns, one for each time interval. For $pent$ there are also 6 columns, although there are 7 $pent$'s, one for each time period. Because $pent$'s apply to each individual in the superpopulation, which must enter the system some time during the study, the $pent$'s for each state must sum to 1.0. Therefore there are 6 $pent$'s explicitly estimated, with the 7th not listed, computed by subtracting the sum of the estimated parameters from 1. **MARK** permits you to choose either the first or last $pent$ by subtraction. More about that below. The final parameter π is a single parameter in the case of two states (one flip of the coin to determine which state an individual will eventually recruit to). With $s > 2$ states there would be $(s - 1)$ columns in the π PIM, and as in the case of $pent$, one would be derived by subtraction.

You can see below that the PIMs for ψ, p, and S are triangular rather than a single row like the **POPAN** model from Chapter 12 (because the **POPAN** model only permits time variation on parameters).

This is to allow for age in the parameters. Unlike the PIMs of the multi-state model from Chapter 10, the rows of the PIM are not based on the time since first capture, but the time since entering the system. This entry could be due to birth or immigration. For a within-year study for a migratory species, entry could be due to arrival to the study area (e.g., to breed).
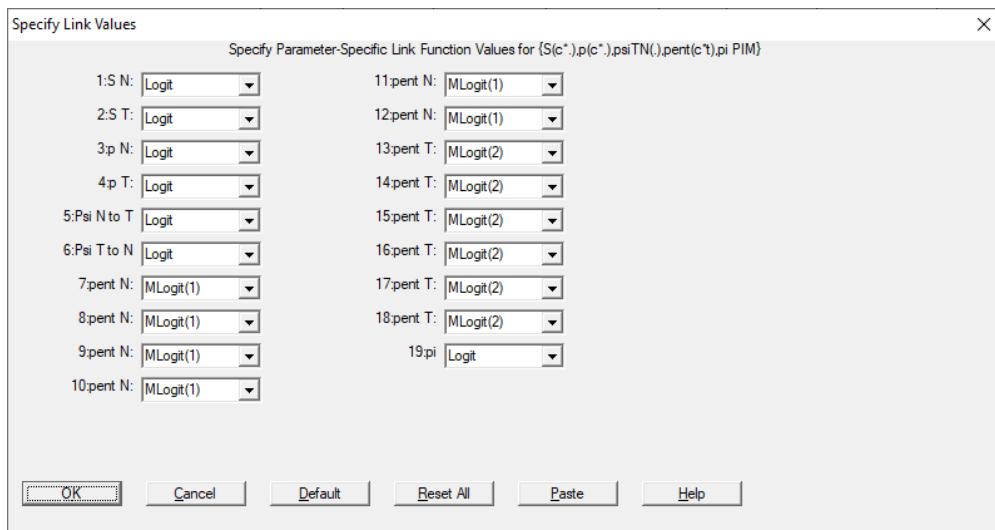
With this model, like with the multi-state model, you can specify which state transition probabilities (which have to add to 1.0) you compute by subtraction of the others from 1. At the top of the screen you would select '**PIM | Change PIM Definitions**', which will lead you to the screen:

You can also designate whether the first or last *pent* for each state is computed by subtraction. After making your choice(s) for which transitions to compute by subtraction, when you click on OK the screen below will appear, and there you choose which *pent* will be computed by subtraction:

Before running a model make sure the link functions for each parameter are appropriate. There are a number of options for most parameters (e.g., Sin, Logit). For those parameters whose values must sum to ≤ 1 you will need to use the '**Mlogit**' link (see Chapters 10, 12, and 16). Below is an example set of link functions for model '**{S(c*.), p(c*.),psiTN(c*.), pent(c*t),pi}**' for the fish example. Since $\sum_{t=1}^{K} pent_t^N = 1.0$ and $\sum_{t=1}^{K} pent_t^T = 1.0$, there are separate Mlogit links for each set.

Running model '**{S(c*.), p(c*.),psiTN(c*.), pent(c*t),pi}**', the real estimates of the model parameters for the fish example are shown at the top of the next page. In this case movement is rare from creek **T** → **N** but apparently does not occur in creek **N**, and approximately 46% of individuals recruit to creek **N**.

```
                            Suckers 091422

      Real Function Parameters of {S(c*.),p(c*.),psiTN(.),pent(c*t),pi PIM}
                                                95% Confidence Interval
         Parameter              Estimate      Standard Error      Lower           Upper
      --------------------------  ---------------  ---------------  ---------------  ---------------
         1:S N:                  0.5384933      0.0197744        0.4995807       0.5769425
         2:S T:                  0.5919121      0.0285067        0.5350840       0.6463845
         3:p N:                  0.1685327      0.0125174        0.1454001       0.1945080
         4:p T:                  0.0730930      0.0087209        0.0577355       0.0921362
         5:Psi N to T            0.7392710E-08  0.5714360E-05   -0.1119275E-04   0.1120754E-04
         6:Psi T to N            0.0085913      0.0039296        0.0034960       0.0209567
         7:pent N:               0.2516169      0.0132614        0.2265247       0.2784877
         8:pent N:               0.2541529      0.0136108        0.2284089       0.2817390
         9:pent N:               0.4207359E-23  0.9942147E-16   -0.1948661E-15   0.1948661E-15
        10:pent N:               0.1008937      0.0108616        0.0815093       0.1242643
        11:pent N:               0.1211070      0.0115416        0.1002439       0.1456094
        12:pent N:               0.2464626      0.0128300        0.2221863       0.2724621
        13:pent T:               0.5300078      0.0224516        0.4858833       0.5736680
        14:pent T:               0.1600354E-22  0.0000000        0.1600354E-22   0.1600354E-22
        15:pent T:               0.0126862      0.0226364        0.3718325E-03   0.3074117
        16:pent T:               0.1303574      0.0180986        0.0987923       0.1701042
        17:pent T:               0.0311927      0.0170593        0.0105365       0.0887134
        18:pent T:               0.1466542      0.0166067        0.1169975       0.1822765
        19:pi                    0.4629684      0.0271680        0.4103377       0.5164375
```

Output for derived parameters is below. The first section is the estimate of N^*, the total number of individuals in the system over 7 years. In this case we don't associate any particular interpretation to this number.

```
                        Estimates of Derived Parameters
      Total Abundance Estimates of {S(c*.),p(c*.),psiTN(.),pent(c*t),pi PIM}
                                            95% Confidence Interval
         Grp.      N*-hat        Standard Error       Lower           Upper
         ----   --------------   --------------   ---------------  ---------------
          1      22382.692        1199.3522        20181.374        24890.971
```

The next section of the derived output contains state-specific annual estimates of abundance using method 1 described above, where $\hat{N}_t^s = n_t^s / \hat{p}_t^s$.

```
      Abundance Estimates of {S(c*.),p(c*.),psiTN(.),pent(c*t),pi PIM}
                                                95% Confidence Interval
      Grp. Str. Occ.    Abundance        Standard Error      Lower           Upper
      ---- ---- ----  --------------   --------------   ---------------  -----------
        1   N    1     267.01066        44.287717        195.73805        371.98267
        1   N    2     2760.2721        236.73340        2341.1885        3273.0162
        1   N    3     4157.7880        273.27642        3664.0090        4738.0883
        1   N    4     2261.0913        204.92974        1897.3608        2703.8370
        1   N    5     2276.8652        188.50091        1942.1674        2683.9680
        1   N    6     2497.0962        207.45364        2128.1210        2944.3988
        1   N    7     3909.9588        312.05931        3354.0522        4581.7793
        1   T    1     1792.2352        263.65711        1350.4740        2394.0270
        1   T    2     7422.5372        808.23490        6011.1782        9196.2520
        1   T    3     4355.7440        617.15611        3317.4798        5758.9987
        1   T    4     2708.5585        347.29814        2114.5003        3485.8964
        1   T    5     3156.3779        409.69251        2457.9262        4076.2783
        1   T    6     2227.1898        299.35621        1718.2058        2901.0564
        1   T    7     3069.7892        405.69437        2378.5686        3981.2314
```

The section of the output that follows contains year-specific estimates of total abundance, summed across states, with precision and confidence interval.

```
Summed Abundance Estimates of {S(c*.),p(c*.),psiTN(.),pent(c*t),pi PIM}
                                              95% Confidence Interval
   Grp. Occ.      Abundance       Standard Error       Lower            Upper
   ---- ----    --------------   --------------   --------------   ------------
     1    1      2059.2459         266.86174         1604.5187        2658.7221
     1    2      10182.809         835.60255         8685.3018        11971.354
     1    3      8513.5320         666.64013         7319.3920        9941.0782
     1    4      4969.6498         399.86870         4250.8568        5823.0305
     1    5      5433.2431         446.57164         4634.9580        6391.4444
     1    6      4724.2860         359.99037         4076.0464        5491.2622
     1    7      6979.7480         505.36337         6067.1253        8053.4654
```

Following this, state- and year-specific estimates of abundance using method 2, where abundance is derived from N^* and state dynamics, are included. In this case this method is not necessary, since fish in each state are captured.

```
Abundance Estimates without TSA of {S(c*.),p(c*.),psiTN(.),pent(c*t),pi PIM}
                                                95% Confidence Interval
  Grp. Str. Occ.      Abundance       Standard Error       Lower            Upper
  ---- ---- ----    --------------   --------------   --------------   ------------
    1    N    1      267.01049         19.831702         231.41773        309.39900
    1    N    2      2765.0420         205.36829         2396.4591        3203.9985
    1    N    3      5019.7973         372.83599         4350.6532        5816.7012
    1    N    4      1323.1853         98.277100         1146.8034        1533.2439
    1    N    5      2343.7588         174.07827         2031.3334        2715.8356
    1    N    6      2314.0910         171.87475         2005.6203        2681.4580
    1    N    7      4082.2938         303.20469         3538.1199        4730.3669
    1    T    1      1792.2375         213.83691         1423.1462        2266.7568
    1    T    2      7032.1381         839.02422         5583.9477        8893.9925
    1    T    3      4979.9577         594.17280         3954.3910        6298.4694
    1    T    4      2585.7473         308.51280         2053.2415        3270.3591
    1    T    5      3488.7066         416.24743         2770.2464        4412.3893
    1    T    6      2011.1368         239.95440         1596.9656        2543.6126
    1    T    7      2859.3713         341.15966         2270.5157        3616.4289
```

The following section contains total abundance by year, based on method 2.

```
Summed Abundance Estimates without TSA of {S(c*.),p(c*.),psiTN(.),pent(c*t),pi PIM}
                                              95% Confidence Interval
   Grp. Occ.     Abundance       Standard Error       Lower            Upper
   ---- ----   --------------   --------------   --------------   ------------
     1    1      2059.2480         214.65324         1683.2955        2528.9713
     1    2      9797.1801         862.76879         8261.7491        11656.372
     1    3      9999.7550         699.83942         8731.5957        11481.729
     1    4      3908.9326         323.30708         3330.4740        4602.0452
     1    5      5832.4654         450.35737         5022.2454        6792.8271
     1    6      4325.2277         294.44171         3790.6770        4947.6005
     1    7      6941.6651         455.26008         6113.1803        7901.7379
```

The final section of output (shown at the top of the next page) includes estimates for expected ingress (arrival for within season model) period, expected egress (departure for within season model) period, and expected residence time in the system for methods 1 and 2 listed above. Again, these parameters could be of interest for a cross-year study, but are more likely to be of interest in examining phenology and duration of study area use for a given season.

```
Expected Ingress, Egress, and Residence Times 1&2 {S(c*.),p(c*.),psiTN(.),pent(c*t),pi PIM}
                                       95% Confidence Interval
 Grp.    Estimate      Standard Error     Lower          Upper
 ----   --------------  --------------  --------------  --------------
    1    3.6450309       0.0628476       3.5218495       3.7682122
    1    4.5600156       0.0535610       4.4550360       4.6649952
    1    1.9149848       0.0637151       1.7901032       2.0398663
    1    2.1383658       0.0843105       1.9731172       2.3036143
```

14.3. Summary

The MSJS model in **MARK** should provide a flexible framework for modeling ingress, survival, state transitions, and abundance in a variety of contexts. This includes populations and metapopulations monitored across years, and well as within-year arrival, use, state transitions, and departure for species that use a study area seasonally.

14.4. References

Dupuis, J. A., and Schwarz, C. J. (2007) A Bayesian approach to the multi-state Jolly-Seber capture-recapture model. *Biometrics*, **63**, 1015-1022.

Kendall, W. L., Stapleton, S., White, G. C., Richardson, J. I., Pearson, K. N., and Mason, P. (2019) A multi-state open robust design: population dynamics, reproductive effort, and phenology of sea turtles from tagging data. *Ecological Monographs*, **89**, e01329.

Schwarz, C. J., and Arnason, A. N. (1996) A general methodology for the analysis of capture-recapture populations in open populations. *Biometrics*, **52**, 860-873.