MIDASPOM: Bayesian Metapopulation Inference Data Analysis using a Stochastic Patch Occupancy Model
Linux User Manual

February 28, 2018

Contents

1. Introduction

MIDASPOM is a Bayesian inference program which enables to estimate parameters and test hypotheses under Stochastic Patch Occupancy Models (SPOMs) using temporal patch occupancy data. This method is implemented in C. MIDASPOM comprises 4 modules:

1. MIDASPOM, which jointly estimates the probability of detection, extinction and colonization parameters
2. MIDASPOM_impute, which estimates the probability of occupancy of each segment and each time point
3. MIDASPOM_dieoff, which estimates the parameter (past local population size) under an increased in situ die-off hypothesis
4. MIDASPOM_loss, which estimates the parameters (past source population size and distance to the linear habitat) under a habitat loss hypothesis
5. MIDASPOM_future, which computes the probability of extinction under different management scenarios

In addition, MIDASPOMA, MIDASPOMA_dieoff and MIDASPOMA_loss correspond to a sparse matrix many-patches approximation versions of modules MIDASPOM, MIDASPOM_dieoff and MIDASPOM_loss.

2. Install

All modules require the MPI (Message Passing Interface) library for parallel computing. To install this library, on Debian, Ubuntu and Mint, type

```
sudo apt-get install mpi
```

on Fedora and Redhat, type

```
sudo yum install mpi
```

Then, to use MIDASPOM, simply decompress the archive in the directory of your choice using the command:

```
unzip MIDASPOM.zip
```

You can then use program MIDASPOM by typing

```
mpirun -np 1 bin/MIDASPOM.out [options]
```

where [options] corresponds to optional switches described in the following section.

Alternatively, you can compile the sources yourself, by typing:

```
make
```

in the MIDASPOM directory. To compile the sources, you will need the atlas library. You can install it on Debian, Ubuntu and Mint by typing

```
sudo apt-get install atlas
```

and on Fedora and Redhat by typing

```
sudo yum install atlas
```

3. The MIDASPOM module

MIDASPOM takes as input temporal habitat occupancy data, and outputs the joint posterior distribution of the extinction and colonization parameters of the SPOM presented in the main text.

3.1. Basic example

The basic command is:

```
mpirun -np nb_CPU MIDASPOM_MPI.out -Y occupancy_file -J survey_file
       -m mean_dispersal -d mean_distance
```

Executable MIDASPOM_MPI uses the mpi library (mpirun command), and several options using switches (see Table 1 for the full list of switches). In its most basic usage, you need to specify the number of CPUs (or threads) on wich to run the program (switch -np), the name of the input files (switches -Y and -J for the occupancy and survey files, respectively), the mean dispersal distance of the species (switch -m), and the distance between consecutive habitat segments (switch -d). For example, to infer the extinction and colonization rates of occupancy data in file Y.txt, for a species which mean dispersal distance is 400m and which occupies a linear habitat with segments 200m long, and using 4 CPUs, one enters:

```
mpirun -np 4 MIDASPOM_MPI -Y Y.txt -J J.txt -m 400 -d 200
```

The input data consists in two space or tabulation-delimited tables. The first table represents temporal occupancy records for segments of a linear habitat. Each row corresponds to a year or another appropriate time unit, and each column corresponds to a segment. The numbers indicates the number of surveys where the segment was detected as occupied. For the previous example, a possible input file *Y.txt* for a linear habitat subdivided into 5 segments surveyed 3 consecutive years would be:

```
0 1 2 1 1
0 0 1 0 1
1 0 1 2 0
```

The second table represents the number of surveys per patch per year. This table has the same dimensions as the occupancy table. For example, a possible input file *J.txt* corresponding to input occupancy file *Y.txt* would be:

```
0 1 2 2 2
0 1 1 2 1
1 1 2 2 0
```

The output is a $s \times u$ table, where n_e and n_c are the number of values of e and c, respectively, on which the posterior is computed. We denote by l and u the lower and upper bounds of the prior on e, and by h and v those of c (defaults are $s = 101$, $l = h = 0$ and $u = v = 1$). The entry of row i and column j corresponds to the log-likelihood of parameters $e = (u - l)\frac{i-1}{n_e-1} + l$ and $c = (v - h)\frac{j-1}{n_c-1} + h$. For example, a possible output file *loglik.txt* for $s = 5$, and default l and u is:

```
-100 -100 -100 -100 -100
-100 -82 -81 -82 -85
-100 -82 -80 -79 -84
-100 -82 -81 -82 -85
-100 -100 -100 -100 -100
```

In that case, the maximum value (-79) is at row $i = 3$ and column $j = 4$, thus the maximum a posteriori estimates of the parameters are $\tilde{e} = (1 - 0) \times \frac{3-1}{5-1} = 0.5$ and $\tilde{c} = (1 - 0) \times \frac{4-1}{5-1} = 0.75$.

The posterior probability can then be computed by computing the exponential of the values, and normalizing the values (see script *plot_posterior.R*).

3.2. Inferring detection probabilities

The probability of detection can be set using the switch -f. When this probability is unknown, we need to run MIDASPOM across a range of values. For example, using a bash script:

```
for i in `seq 0.5 0.05 1`
do
    mpirun MIDASPOM_MPI -m 400 -d 200 -i Y.txt -J J.txt -f $i
done
```

3.3. All command line options

4. The MIDASPOM_impute module

MIDASPOM_impute takes as input the survey and occupancy tables $\mathbf{J}$ and $\mathbf{Y}$ along with estimates of the model parameters $\tilde{p}$, $\tilde{e}$, $\tilde{c}$, and outputs the probability of occupancy of each segment for each surveyed year, and the distribution probability of the number of occupied segments.

Table 1: **Summary of switches for MIDASPOM**

Switch	Model parameter	Default	Description
-m	$1/\alpha$	400 m	mean dispersal distance of the species in meters
-p	ϕ_0	0.5	prior probability that a segment with unknown initial occupancy was actually occupied
-d	d	100 m	mean length of the habitat segments in meters
-i	$\mathbf{Y}$	input.txt	input file for patch occupancy data
-J	$\mathbf{J}$	J.txt	input file for patch occupancy data
-o	-	loglik.txt	output file where posterior distribution is written
-s	-	0.01	Window size for likelihood numerical computation
-l	-	0	lower bound of the prior on e
-u	-	1	upper bound of the prior on e
-h	-	0	lower bound of the prior on c
-v	-	1	upper bound of the prior on c
-f	p	1	probability of detection

4.1. Basic example

```
mpirun -np 4 MIDASPOM_impute_MPI -m 400 -d 200 -i Y.txt -J J.txt -o impute_occupancy.txt
       -q impute_patches.txt -f 0.9 -e 0.4 -c 1.1
```

Executable MIDASPOM_impute_MPI uses the mpi library and some specific switches (see Table 3). At least, you need to specify the number of CPUs (or threads) on wich to run the program (switch -np), the input files Y.txt and J.txt, and the estimated parameters (switches -f, -e, and -c estimated using MIDASPOM_MPI), the mean dispersal distance of the species (switch -m), and the distance between consecutive habitat segments (switch -d). See script *plot_impute.R* for examples on how to plot the results.

4.2. All command line options

Table 2: **Summary of switches for the MIDASPOM_impute module**

Switch	Model parameter	Default	Description
-m	$1/\alpha$	400 m	mean dispersal distance of the species in meters
-p	ϕ_0	0.5	prior probability that a segment with unknown initial occupancy was actually occupied
-d	d	100 m	mean length of the habitat segments in meters
-i	$\mathbf{Y}$	input.txt	input file for patch occupancy data
-J	$\mathbf{J}$	J.txt	input file for patch occupancy data
-o	-	occupancy.txt	output file where imputed occupancies are written
-q	-	patches.txt	output file where the distribution of the number of occupied segments is written
-e	-	1	extinction parameter e
-c	-	1	colonization parameter c
-f	p	1	probability of detection

5. The MIDASPOM_dieoff module

MIDASPOM_dieoff takes as input the probability of occupancy in the first surveyed year, and outputs the likelihood of the population size of a segment before the increased *in situ* die-off, K_D, for the SPOM presented in the main text (hypothesis 1).

5.1. Basic example

```
mpirun -np nb_CPU MIDASPOM_dieoff_MPI -a time_after_event -b time_before_event
       -e extinction_param -c colonization_param -m mean_dispersal -d mean_distance
       -i impute_occupancy
```

Executable MIDASPOM_dieoff_MPI uses the mpi library, the switches from MIDASPOM_MPI, and some specific switches (see Table 3). At least, you need to specify the number of CPUs (or threads) on wich to run the program (switch -np), the time (in years or in the time unit appropriate for the dataset) between the event which increased *in situ* die-off and the timing of the first survey (switch -a) and before the event (switch -b), the extinction and colonization parameters (switches -e and -c; typically estimated using MIDASPOM_MPI), the name of the input probability of occupancy file (switch -i; typically computed by MIDASPOM_impute), the mean dispersal distance of the species (switch -m), and the distance between consecutive habitat segments (switch -d). The input file can contain several time-steps, but only the first one is used for inference.

For example, to infer K_D for an event that occurred 10 years ago from occupancy data in file obs.txt, for a species which mean dispersal distance is 400m, and a habitat with segments of 200m, with extinction and colonization parameters of 0.5, and using 4 CPUs, one enters:

```
mpirun -np 4 MIDASPOM_dieoff_MPI -a 10 -e 0.5 -c 0.5 -m 400 -d 200 -i occupancy.txt
```

The output is a series of s values (default is $s = 151$), and entry i corresponds to the posterior probability of parameter $K_D = 10^{\frac{i-1}{s-1}\left[\log_{10}(u)-\log_{10}(l)\right]+\log_{10}(l)}$, where l and u are the lower and upper bounds of the prior on K_D (defaults are $l = 0.1$ and $u = 100$). For example, a possible output file *lh_dieoff.txt* for $s = 11$ and default l and u is:

```
0.000000 0.000000 0.000000 0.000003 0.154286 0.250559 0.256858 0.258974 0.259911
0.260355 0.260570
```

In that case, the maximum value (0.260570) is in cell $i = 11$, thus the maximum a posteriori estimate is $\tilde{K_D} = 10^{\frac{11-1}{11-1}(2+1)-1} = 100$, and 95% of the maximum value is reached in cell $i = 6$, which corresponds to $K_D = 10^{\frac{6-1}{11-1}(2+1)-1} = 10^{0.5} \approx 3.162278$.

5.2. All command line options

Table 3: **Summary of switches for the MIDASPOM_dieoff module**

Switch	Model parameter	Default	Description
-m	$1/\alpha$	400 m	mean dispersal distance of the species in meters
-p	ϕ_0	0.5	prior probability that a segment with unknown initial occupancy was actually occupied
-d	d	200 m	mean length of the habitat segments in meters
-i	-	input.txt	input file for patch occupancy probability
-o	-	lh_dieoff.txt	output file where likelihood is written
-s	-	151	number of values for numerical likelihood computation
-b	t_b	20	number of years before the disturbance
-a	t_a	-	number of years between the disturbance and the year of the first survey
-e	e	-	extinction parameter
-c	c	-	colonization parameter
-l	-	0.1	lower bound of the prior on K_D
-u	-	100	upper bound of the prior on K_D

6. The MIDASPOM_loss module

MIDASPOM_loss takes as input the probability of occupancy in the first surveyed year, and outputs the joint likelihood of the size of the lost population, K_L, and the distance between the lost population and the first segment of the linear habitat, d_L, for the SPOM presented in the main text (hypothesis 2).

6.1. Basic example

```
mpirun -np nb_CPU MIDASPOM_loss_MPI -a time_after_event -b time_before_event
       -e extinction_param -c colonization_param -m mean_dispersal -d mean_distance
       -i impute_occupancy
```

Executable MIDASPOM_loss_MPI uses the mpi library, the switches from MIDASPOM_MPI, and some specific switches (see Table 4). At least, you need to specify the number of CPUs (or threads) on which to run the program (switch -np), the time (in years or in the time unit appropriate for the dataset) between the habitat loss event and the timing of the first survey

(switch -a), the extinction and colonization parameters (switches -e and -c; typically estimated using MIDASPOM_MPI), the name of the input file (switch -i), the mean dispersal distance of the species (switch -m), and the distance between consecutive habitat segments (switch -d). The input file can contain several time-steps, but only the first one is used for inference.

For example, to infer K_L and d_L for an event that occurred 10 years ago from occupancy data in file obs.txt, for a species which mean dispersal distance is 400m, and a habitat with segments of 200m, with extinction and colonization parameters of 0.5, and using 4 CPUs, one enters:

```
mpirun -np 4 MIDASPOM_loss_MPI -a 10 -e 0.5 -c 0.5 -m 400 -d 200 -i occupancy.txt
```

The output is a $s \times v$ table (default is $s = 151$ and $v = 20$). The entry of row i and column j corresponds to the joint posterior probability of parameters $K_L = 10^{\frac{i-1}{s-1}\left[\log_{10}(u_K)-\log_{10}(l_K)\right]+\log_{10}(l_K)}$ and $d_L = \frac{j-1}{v-1}(u_d - l_d) + l_d$, where l_K and u_K are the lower and upper bounds of the prior on K_L (defaults are $l_K = 0.1$ and $u_K = 100$) and where l_d and u_d are the lower and upper bounds of the prior on d_L (defaults are $l_d = 200$ and $u_d = 4000$). For example, a possible output file *lh_loss.txt* for $s = 11$ and $v = 4$, and default l_K, u_K, l_d, and u_d is:

```
0.027205 0.002034 0.001587 0.001569
0.049097 0.002497 0.001606 0.001570
0.084571 0.003414 0.001644 0.001571
0.133971 0.005226 0.001720 0.001574
0.189908 0.008765 0.001871 0.001581
0.234981 0.015541 0.002171 0.001594
0.256482 0.028018 0.002768 0.001619
0.266358 0.049404 0.003952 0.001669
0.269195 0.081827 0.006280 0.001770
0.269195 0.122334 0.010800 0.001971
0.269195 0.146794 0.019346 0.002370
```

In that case, the maximum value (0.269195) is in cells $(i = 9, j = 1)$, $(i = 10, j = 1)$, and $(i = 11, j = 1)$, thus the maximum a posteriori estimates are $d_L = \frac{1-1}{11-1}(4000 - 200) + 200 = 200$, and K_L between $K_L = 10^{\frac{9-1}{11-1}(2+1)-1} = 10^{1.4} \approx 25.11886$ and $K_L = 10^{\frac{11-1}{11-1}(2+1)-1} = 100$. 95% of the maximum likelihood value is reached in cell $(i = 7, j = 1)$, which corresponds to $d_L = 200$ and $K_D = 10^{\frac{7-1}{11-1}(2+1)-1} = 10^{0.8} \approx 6.309573$.

6.2. All command line options

Table 4: **Summary of switches for the MIDASPOM_loss module**

Switch	Model parameter	Default	Description
-m	$1/\alpha$	400 m	mean dispersal distance of the species in meters
-p	ϕ_0	0.5	prior probability that a segment with unknown initial occupancy was actually occupied
-d	d	200 m	mean length of the habitat segments in meters
-i	-	input.txt	input file for patch occupancy probability
-o	-	lh_loss.txt	output file where likelihood is written
-s	-	151	number of values for numerical likelihood computation of parameter K_L
-v	-	20	number of values for numerical likelihood computation of parameter d_L
-b	t_b	20	number of years before the disturbance
-a	t_a	-	number of years between the disturbance and the year of the first survey
-e	e	-	extinction parameter
-c	c	-	colonization parameter
-l	-	0.1	lower bound of the prior on K_L
-u	-	100	upper bound of the prior on K_L
-L	-	200	lower bound of the prior on d_L
-U	-	4000	upper bound of the prior on d_L

7. The MIDASPOM_future module

MIDASPOM_future takes as input the last survey of habitat occupancy, and outputs the number of simulations for which the population became extinct under several management scenarios, as a function of time since the last survey. The

parameters of the management scenario are the size of the additional source population, K_{source}, the distance between the source and the first segment of the linear habitat, d_{source}, and the size of the populations in each segment relative to present, K_D. The case $K_{source} = 0$ and $K_D = 1$ corresponds to the scenario without any management (i.e., no change in extinction and colonization rates compared to present).

7.1. Basic example

```
mpirun -np nb_CPU MIDASPOM_future_MPI -a duration -m mean_dispersal
       -d mean_distance -i impute_occupancies -q input_param_distribution
```

Executable MIDASPOM_future_MPI uses the mpi library, the switches from MIDASPOM_MPI, and some specific switches (see Table 6). At least, you need to specify the number of CPUs (or threads) on wich to run the program (switch -np), the duration (in years or in the time unit appropriate for the dataset) of the simulations (switch -a), the file containing the posterior distribution of extinction and colonization parameters (switch -q; typically the output of MIDASPOM_MPI processed using script *plot_posterior.R*), the name of the input probability of occupancy file (switch -i), the mean dispersal distance of the species (switch -m), and the distance between consecutive habitat segments (switch -d). The input file can contain several time-steps, but only the first one is used for inference.

For example, to predict the probability of extinction in the next 10 years for $K_{source} = 1$, $d_{source} = 200$, and $K_D = 1$ from occupancy data in file *impute_patches.txt*, for a species which mean dispersal distance is 400m, and a habitat with segments of 200m, with the probability of extinction and colonization parameters from file *posterior.txt*, and using 4 CPUs, one enters:

```
mpirun -np 4 MIDASPOM_future_MPI -a 10 -m 400 -d 100 -i impute_patches.txt
       -q posterior.txt -o pext_future.txt -S 1 -s 200
```

The output is a series of a values (default is $a = 50$). The entry i corresponds to the number of simulations (out of n_{sim}, with $n_{sim} = 10,000$ by default) where all segments were extinct at time i in the future. For example, a possible output file *pext_future.txt* for the previous example is:

```
0 0 0 0 4 8 12 12 32 52
```

In that case, the population survived in at least 1 segment during the next 4 years in all 10,000 simulations. The proportion of extinct populations in 10 years is 52 out of 10,000, which leads to an estimate of the 10-year probability of extinction of 0.0052.

7.2. All command line options

Table 5: **Summary of switches for the MIDASPOM_future module**

Switch	Model parameter	Default	Description
-m	$1/\alpha$	400 m	mean dispersal distance of the species in meters
-p	$Pr(p_{i,0} = 1)$	0.5	prior probability that a segment with unknown initial occupancy was actually occupied
-d	d	200 m	mean length of the habitat segments in meters
-i	-	input.txt	input file for patch occupancy probability
-q	-	posterior.txt	input file with joint distribution of extinction and colonization parameters
-o	-	pext_future.txt	output file where likelihood is written
-n	-	10,000	number of simulations
-a	t	50	duration of the simulations
-S	K_{source}	0	size of the additional source population
-s	d_{source}	200	distance between the additional source population and the first segment of the linear habitat
-D	K_D	0	size of the populations relative to present

8. The MIDASPOMA modules

Modules MIDASPOMA_MPI, MIDASPOMA_dieoff, and MIDASPOMA_loss respectively correspond to the many-patches approximate algorithms 1 and 2. They can be used similarly to their corresponding modules, with an additional switch -n that specifies the number of states that are retained each year (parameter m in algorithms 1 and 2).

In addition, MIDASPOMA_dieoff, and MIDASPOMA_loss can take additional arguments with switches -w and -r. The former outputs the list of states used for the computation. The latter reads a list of states from a file, which can be used to ensure the reproducibility of results.

9. Troubleshooting

This section summarizes possible errors and their potential resolution.

Table 6: **Summary of possible errors**

Error	**Cause**	**Solution**
MPI Error in MPI_Pack_size() (0) Error in NBC_Copy() (0)	not enough threads available	reduce the number after switch *-np*
Output "loglik.txt" only contains NA	a year with total extinction (0 everywhere) is followed by a year with some occupied patches	either add an additional unsampled habitat patch with -1, or change a 0 to -1 if there is uncertainty in one particular patch, or split the input file in 2 and estimate separately the parameters before and after the total extinction