

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

Authors TBD

October 14, 2016

*Land Use and Environmental Planning, Stanford University, USA  
Department of Biology, Stanford University, Stanford, CA 94305-5020, USA*

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Install</b>	<b>2</b>
<b>3</b>	<b>The MIDASPOM module</b>	<b>2</b>
3.1	Basic example . . . . .	3
3.2	All command line options . . . . .	3
<b>4</b>	<b>The MIDASPOM.dieoff module</b>	<b>4</b>
4.1	Basic example . . . . .	4
4.2	All command line options . . . . .	4
<b>5</b>	<b>The MIDASPOM.loss module</b>	<b>4</b>
5.1	Basic example . . . . .	5
5.2	All command line options . . . . .	5
<b>6</b>	<b>The MIDASPOM.future module</b>	<b>5</b>
6.1	Basic example . . . . .	5
6.2	All command line options . . . . .	6
<b>7</b>	<b>Troubleshooting</b>	<b>7</b>

## 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, and exists in single processor and multiprocessor versions. MIDASPOM comprises 4 modules:

1. MIDASPOM, which estimates extinction and colonization parameters
2. MIDASPOM\_dieoff, which estimates the parameter (past local population size) under an increased in situ die-off hypothesis
3. MIDASPOM\_loss, which estimates the parameters (past source population size and distance to the linear habitat) under a habitat loss hypothesis
4. MIDASPOM\_future, which computes the probability of extinction under different management scenarios

## 2. Install

All modules exist in single processor and multiprocessor versions. The multiprocessor versions 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 the single processor version of program MIDASPOM by typing

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

where [options] corresponds to optional switches described in the following section, or run the multiprocessor version by typing

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

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

For the single processor version, the basic command is:

```
MIDASPOM.out -i input_file -m mean_dispersal -d mean_distance
```

For the MPI version, the basic command is:

```
mpirun -np nb_CPU MIDASPOM_MPI.out -i input_file -m mean_dispersal -d mean_distance
```

Executable MIDASPOM\_MPI.out 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 which to run the program (switch -np), 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). For example, to infer the extinction and colonization rates of occupancy data in file obs.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.out -i obs.txt -m 400 -d 200
```

The input data consists in a space or tabulation-delimited table representing temporal occupancy records for segment of a linear habitat. Each row corresponds to a year or another appropriate time unit, and each column corresponds to a segment. A 0 indicates an empty segment, a 1 indicates an occupied segment, a -1 indicates missing data. For the previous example, a possible input file *obs.txt* for a linear habitat subdivided into 5 segments surveyed 3 consecutive years would be:

```
0 1 1 1 1
0 -1 1 0 1
1 0 1 1 0
```

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

```
0.000000 0.000000 0.000000 0.000000 0.000000
0.000000 0.612582 0.757902 0.374970 0.186856
0.000000 1.570907 2.757263 2.273996 1.547265
0.000000 0.844716 2.188557 2.634643 2.234809
0.000000 0.000000 0.000000 0.000000 0.000000
```

In that case, the maximum value (2.757263) is at row  $i = 3$  and column  $j = 3$ , thus the maximum a posteriori estimates of the parameters are  $\tilde{e} = (1 - 0) \times \frac{3-1}{5-1} = 0.5$  and  $\tilde{c} = (1 - 0) \times \frac{3-1}{5-1} = 0.5$ .

### 3.2. All command line options

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	$Pr(p_{i,0} = 1)$	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 data
-o	-	posterior.txt	output file where posterior distribution is written
-s	-	101	Number of values for posterior numerical computation; the posterior is evaluated at $e = l, (u - l) \frac{1}{s-1} + l, (u - l) \frac{2}{s-1} + l, \dots, (u - l) \frac{s-1}{s-1} + l$ (respectively for $c$ ).
-l	-	0	lower bound of the prior
-u	-	1	upper bound of the prior

## 4. The MIDASPOM\_dieoff module

MIDASPOM\_dieoff takes as input the first survey of habitat occupancy, 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).

### 4.1. Basic example

```
mpirun -np nb_CPU MIDASPOM_dieoff_MPI.out -a time_after_event -e extinction_param
      -c colonization_param -m mean_dispersal -d mean_distance -i input_file
```

Executable MIDASPOM\_dieoff\_MPI.out uses the mpi library, the switches from MIDASPOM\_MPI.out, and some specific switches (see Table 2). 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 event which increased *in situ* die-off and the timing of the first survey (switch -a), the extinction and colonization parameters (switches -e and -c; typically estimated using MIDASPOM\_MPI.out), 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  $K_D$  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.out -a 10 -e 0.5 -c 0.5 -m 400 -d 200 -i obs.txt
```

The output is a series of  $s$  values (default is  $s = 151$ ), and entry  $i$  corresponds to the posterior probability of parameter  $K_D = 10^{\frac{i-1}{s-1} [\log_{10}(u) - \log_{10}(l)] + \log_{10}(l)}$ , 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  $K_D = 10^{\frac{6-1}{11-1} (2+1) - 1} = 10^{0.5} \approx 3.162278$ .

### 4.2. All command line options

Table 2: 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	$Pr(p_{i,0} = 1)$	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 data
-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 $K_D$

## 5. The MIDASPOM\_loss module

MIDASPOM\_loss takes as input the first survey of habitat occupancy, 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).

## 5.1. Basic example

```
mpirun -np nb_CPU MIDASPOM_loss_MPI.out -a time_after_event -e extinction_param
-c colonization_param -m mean_dispersal -d mean_distance -i input_file
```

Executable `MIDASPOM_loss_MPI.out` uses the `mpi` library, the switches from `MIDASPOM_MPI.out`, and some specific switches (see Table 3). 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.out`), 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.out -a 10 -e 0.5 -c 0.5 -m 400 -d 200 -i obs.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  $K_L = 10^{\frac{i-1}{s-1} [\log_{10}(u_K) - \log_{10}(l_K)] + \log_{10}(l_K)}$  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  $K_D = 10^{\frac{7-1}{11-1}(2+1)-1} = 10^{0.8} \approx 6.309573$ .

## 5.2. All command line options

## 6. 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).

## 6.1. Basic example

```
mpirun -np nb_CPU MIDASPOM_future_MPI.out -a duration -m mean_dispersal
-d mean_distance -i input_occupancies -q input_param_distribution
```

Executable `MIDASPOM_future_MPI.out` uses the `mpi` library, the switches from `MIDASPOM_MPI.out`, and some specific switches (see Table 5). At least, you need to specify the number of CPUs (or threads) on which 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

Table 3: 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	$Pr(p_{i,0} = 1)$	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 data
-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 $d_L$

MIDASPOM\_MPI.out), the name of the input 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 obs.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.out -a 10 -m 400 -d 100 -i obs.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.

## 6.2. All command line options

Table 4: 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 data
-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

## 7. Troubleshooting

This section summarizes possible errors and their potential resolution. Please report any additional unexplained error to [nalcala@stanford.edu](mailto:nalcala@stanford.edu)

Table 5: 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 <i>-np</i>
Output "posterior.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