#2: Decreasing calculation time

Nadege Belouard*

2024-08-22

Contents

Aim and setup

In case of extremely large species occurrence datasets, it may take a long time to run the analyses. Any number of sectors will provide the accurate results. However, computational time may be decreased by increasing the number of sectors considered. The higher the number of sectors, the larger the invasion radius at which points are compared by pairs in `find_thresholds`, so the fewer distances need to be calculated. However, the lower the number of sectors, the better pre-identification of spatial discontinuities and the more pruned the list of potential jumps, so the faster `find_jumps`. The lowest computational time is therefore obtained by a trade-off between dataset size, invasion radius, and number of sectors.

We demonstrate the effect of the number of sectors on computational time on the SLF dataset.

Load the grid data created in the first vignette

```
grid_data <- read.csv(file.path(here::here(), "exported-data", "grid_data.csv"))
```

Compare calculation times

Run the jumpID functions successively for 16, 40, and 80 sectors and compare computation times.

```
## [1] "Sectors: 16"
## 2024-08-22 14:03:38.20105 Start sector attribution... Sector attribution completed.
## 2024-08-22 14:03:38.249482 Start finding thresholds... Sector 1/16...  2/16...  3/16...  4/16...  5/1
## Threshold analysis done. 4243 potential jumps were found.
## 2024-08-22 14:09:57.273903 Start finding jumps... Year 2014 ... Year 2015 ... Year 2016 ... Year 201
## 2024-08-22 14:10:19.837816 Start finding secondary diffusion... Year 2017 ...Year 2018 ...Year 2019
## [1] "Sectors: 40"
## 2024-08-22 14:11:26.640472 Start sector attribution... Sector attribution completed.
## 2024-08-22 14:11:26.659887 Start finding thresholds... Sector 1/40...  2/40...  3/40...  4/40...  5/4
```

*iEco lab at Temple University, Ecobio lab at the University of Rennes, nadege.belouard@gmail.com

```
## Threshold analysis done. 3747 potential jumps were found.
## 2024-08-22 14:13:20.481658 Start finding jumps... Year 2014 ... Year 2015 ... Year 2016 ... Year 201
## 2024-08-22 14:13:40.582599 Start finding secondary diffusion... Year 2016 ...Year 2017 ...Year 2018
## [1] "Sectors: 80"
## 2024-08-22 14:14:27.939271 Start sector attribution... Sector attribution completed.
## 2024-08-22 14:14:27.958792 Start finding thresholds... Sector 1/80...  2/80...  3/80...  4/80...  5/8
## Warning: no negative survey in the gap identified in sector 23 and year 2022 after 106 km. The spatia
##  24/80...  25/80...  26/80...  27/80...  28/80...  29/80...  30/80...  31/80...  32/80...  33/80... W
## Warning: no negative survey in the gap identified in sector 33 and year 2020 after 113 km. The spatia
##  34/80...  35/80...  36/80...  37/80...  38/80...  39/80...  40/80...  41/80...  42/80...  43/80...
## Threshold analysis done. 5034 potential jumps were found.
## 2024-08-22 14:14:47.273658 Start finding jumps... Year 2014 ... Year 2015 ... Year 2016 ... Year 201
## 2024-08-22 14:15:22.736783 Start finding secondary diffusion... Year 2016 ...Year 2017 ...Year 2018
```

```
##    s potJumps Jumps    Total_time
## 1 16     4243   387 7.807336 mins
## 2 40     3747   387 3.021631 mins
## 3 80     5034   387 2.207451 mins
```

For this dataset, all computational times are decreased by dividing space into 40 sectors instead of 16. Data is not dense enough for dividing space into 80 sectors, as indicated by multiple warning messages from `find_threshold`.

– end of vignette –