Analysis of incubation behaviour using incR

Pablo Capilla-Lasheras (pacapilla@gmail.com)

2017-06-06

1 - Introduction

In this vignette I would like to present and describe a recommended workflow to analyse incubation behaviour using the *incR* R package. This package is currently available through GitHub. It is in active development and the user should not expect to find either perfect code or documentation. However, it is my intention to further develop both if the method results to be appropriate.

My main aim writing this vignette is to introduce the package to potential users in order to provide a piece of code they can use and, at the same time, evaluate the usefulness of the functions that this package contains in a wide suite of circumstances.

1.1 - The package in short

incR contains three types of functions. First, *incRprep* takes a raw time series of temperatures and creates additional useful columns for further analysis. The user could do this by hand but *incRprep* makes it easier, cleaner and quicker. The rest of the functions in the package have been customised to accommodate the data in the form that *incRprep* returns.

The main function in *incR* is *incRscan*. This function implements an algorithm that scans temperature time-series and comes up with suggested positions for the incubating individual, these "positions" being "inside the nest" (coded as 1) or "outside the nests" (coded as 0). Compared to previous methods [1], this algorithm exploits daily variation in nest temperatures and calibrates itself based on daily conditions. *incRscan* analyses temperature variation between pairs of consecutive points, potentially increasing the sensitivity to detect short incubation off-bouts.

At the heart of *incR* lays *incRscan*; however, *incR* aims at being more than only such function. If the job of *incRscan* is performed by other means (or if *incRscan* is further developed and modified), *incR* still provides a suite of functions that allow the extraction of biologically meaningful metrics from a series of observations of the incubating individual. This aims at the development of *incRscan* without affecting the functionality of the package as a whole.

1.2 - Description of the algorithm implemented by ***incRscan***

The idea of developing this approach came up studying blue and great tits in the northern hemisphere in a temperate zone (Scotland). The characteristic nest temperature of incubation in these species varies around 35ºC, therefore, always well above environmental temperature (which don't usually reach 20ºC). When incubating individuals leave the nest, there exists a strong gradient between nest (around 35ºC) and environmental temperatures (somewhere between 5 and 25ºC depending on latitude). Thus, the drop in temperature right after a bird leaves is large. At this point, however, any increase in nest temperature would mean the bird has come back and is heating up the nest. On the other hand, if nest temperature is close to the environmental one, the decrease in temperature after an incubation off-bout would not be very large (the temperature gradient between the nest and the environment is small). In this situation, however, an incubation on-bout would be reflected in a large increase in nest temperature (the gradient between environmental temperature and incubation temperature - circa.35ºC - is large).

The dynamics in nest temperature is exploited by *incRscan*, setting thresholds in temperature variation which characterised changes between incubation on- and off-bouts. The algorithm is thought to work well when there is a marked difference between incubation and environmental temperatures. The utility of the algorithm to analyse data collected in hot environments has not been tested yet. In this scenario the difference between incubation and environmental temperatures would be smaller, making the distinction of incubation on- and off- bouts harder.

Incubation data are sequentially analysed by *incRscan* in a daily basis. I, therefore, explain here the analytical approach summarising one day of incubation (day 1); however, the explanation can be easily extended to understand the analysis of several sequential days. *temp.diff*, *lower.time*, *upper.time*, *sensitivity* and *maxNightVar_accepted* are the arguments that control *incRscan*.

I differentiate two scenarios: i) when nest temperature is higher than environmental temperature by more than *temp.diff* degrees; and, ii) when nest temperature is within *temp.diff* degrees of environmental temperature. Incubation temperature data between *lower.time* and *upper.time*, when incubating individuals are assumed to be in their nests (given the biology of many species, this period can be set at night, when the incubating individual rests on the nest), are chosen to calibrate the algorithm. Within this calibrating window, the minimum difference in temperature between two consecutive data points (eg. minimum (T_i - T_{i-1}), this is equal to the maximum decrease in temperature between pairs of consecutive) is set as a threshold for incubation off-bouts ("maxDrop") in scenario i): assuming that nest temperatures are above environmental values, when the incubating individual is in the nest this value is thought to effectively represent the maximum drop in temperature associated with periods when the incubating individual is in the nest. The threshold for incubation off-bout in scenario ii) is set to $"maxDrop" X sensitivity$ (a default value for *sensitivity* of 0.15 has been found adequate for blue tits in Scotland, however, this value can be set by the user depending on the species of study; see more on this below).

Similarly, "maxIncrease" is defined as the maximum difference in temperature between two consecutive data points (eg. maximum (T_i - T_{i-1}), equal to the maximum increase in temperature between pairs of consecutive), and is set as a threshold for incubation on-bouts ("maxIncrease") under scenario ii). The threshold for incubation on-bouts in scenario i) is set to 0, as any increase in temperature when nest temperature is well above environmental values will indicate the incubating individual has come to the nest.

Once these thresholds are set, temperature differences between successive pairs of data points throughout the day and after *upper.time* are calculated and orderly compared with the value of "maxDrop" and "maxIncrease" following:

T_i - T_{i-1} < maxDrop (= minimum (T_i - T_{i-1})) (1)

T_i - T_{i-1} > maxIncrease (= maximum (T_i - T_{i-1})) (2)

T_i - T_{i-1} being the i*th* temperature record from *i=2* to *i=I* (*I* equal to the last data point, and total number of temperature recordings). Off-bout periods are, then, defined between $T_{i's}$ satisfying (1) and the closest forward situation in which T_j, when $i < j$, satisfies (2) (raise in temperature above what it is expected when the incubating individual is in the nest). On-bout periods start after an off-bout finishes and last until (1) is fulfilled again: when temperature variation between T_i and T_{i-1} indicates the incubating individual has left the nest. To make this calculation conservative and robust against highly variable calibrating periods, whenever $|maxDrop| > maxNightVar\_accepted$ (in degrees C) for a particular studied day, the value of "maxDrop" and "maxIncrease" of the previous day is instead applied. The result of this algorithm is a temporal sequence of 0's and 1's representing on-bout (1) and off-bout (0) periods. Based on this sequence and using other functions of *incR* biological meaningful information can be extracted.

"maxDrop" works well when incubation and environmental temperatures are very different, especially when environmental temperature is much lower than incubation temperature. However, this threshold may not detect off-bouts when nest and environmental temperatures are very similar (see above for further clarification), eg. when after a long incubation off-bout, an off-bout comes after a short on-bout. To circumvent this problem, *incRscan* allows the user to reduce the value of "maxDrop" by a proportion (given by the *sensitivity* argument) when the difference between nest and environmental temperatures is lower than *temp.diff*. The default values for *sensitivity* and *temp.diff* seem to work well for blue and great tits in northern latitudes in

Europe; however, it is highly recommended that the user tries different values and assesses the consistency of the results. Additionally, validated data (for example, using cameras along with temperature loggers to estimate incubation on- and off-bouts) for a subset of incubation days can help to calculate the optimal values of *sensitivity* and *temp.diff* in any circumstances.

2 - Workflow: from raw data to biological metrics

I suggest a workflow to analyse incubation data using *incR*.

2.1 - Getting the package into your computer

The very first step is to download *incR* from GitHub, to do so you will need to install the *devtools* package.

```
install.packages("devtools")  # install "devtools"
```

Then, *incR* can be downloaded:

```
devtools::install_github("PabloCapilla/incR")
```

```
library(incR) # attaching incR
```

2.2 - Read your data and format it using *incRprep*

In this vignette I use example data provided with the *incR* package. These data come from a blue tit nest monitored in Glasgow in 2015 and belong to Barbara Helm and Davide Dominoni, who kindly allowed me to reproduce these data here. Information on the data set can be found in `?incRdataExample`.

```
data(incRdataExample)
df <- incRdataExample
```

```
str(df) # preliminar exploration
```

```
## 'data.frame':    954 obs. of  3 variables:
##  $ DATE    : Factor w/ 954 levels "06/05/2015 21:42",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ valueT  : num  32 32 32.1 32.2 32.2 ...
##  $ env.temp: num  8.76 8.76 8.76 8.76 8.76 ...
```

Now, *incRprep* is applied to the *df* object to prepared the data for further analysis:

```
df.prep <- incRprep (data= df,
                     date.name="DATE",
                     date.format="%d/%m/%Y %H:%M",
                     timezone="GMT",
                     temperature.name="valueT")
```

2.3 - Apply the *incR* algorithm and estimate the position of the inc. individual

The algorithm has been explained above and now it is used on the object returned by *incRprep*. The specified parameters have been checked by comparing their performance against video-calibrated data, yielding very high correspondence (>95%) between the estimated position of the incubating individual using this algorithm and video-camera observations.

```r
inc.example <- incRscan (data=df.prep,            # data frame
                         temp.name = "valueT",
                         lower.time=22,            # beginning of calibrating window
                         upper.time=3,             # end of calibrating window
                         sensitivity=0.15,         # correction when nest Tª is
                                                   # far below maxinc.Temp
                         temp.diff=8,              # diff in Tª from maxinc.Temp that
                                                   # triggers sensitivity change
                         maxNightVariation=2,      # max night variation accepted in
                                                   # calibrating windows
                         env.temp="env.temp")      # column with environmental tempertures
```

```
## Warning in incRscan(data = df.prep, temp.name = "valueT", lower.time =
## 22, : No night reference period for 2015-05-06 - day skipped
```

As you see, you get a warning messages. These warnings are thought to be informative and will tell you something about your data. Most likely, they will mean days without a previous calibrating window (eg. if your data series starts at 11am, then that day won't be assessed as there is no calibrating window available before - as in the example above). Check them but do not assume the function did not work if you see them, probably it did. Further versions of this function will store these messages in a separated data frame, rather than being generated as warnings.

incRscan produces a list with 2 objects. In the first of these, you will find your original data set plus an additional column called *inc.vector*. This variable contains the sequences of on- and off-bouts. In the second object of the list, there is a data frame with information about temperature threshold per day employed to detect on- and off-bouts ("maxDrop" and "maxIncrease). See above for more details.

From now on, I am going to focus on the first object of the *inc.example*. The rest of the functions in *incR* are designed to calculate biological metrics based on the *inc.vector* object returned by *incRscan*. Visualization of the results of *incRscan* is highly recommended and I include here basic code for it.

```r
# selecting the object of interest
df.incubation <- inc.example[[1]]
```

2.3.1 - Plotting the results of *incRscan*

I use Using only one day to increase the resolution of the plot:

```r
# selecting one day of incubation
plot.example <- df.incubation[df.incubation$date=="2015-05-07",]
# plot
plot (valueT ~ index, data=plot.example,
      pch=19,
      main=list("Blue tit incubation 07/05/2015"),
      col=c("red", "black")[plot.example$inc.vector+1],# pch color defined by "inc.vector"
      xaxt="n", las=1, cex.axis=1.5,# no x-axis, horizontal y axis-labels and size
      xlab=list("Time", cex=1.5),
      ylab=list("Temperature (ºC)", cex=1.5))
axis (side=1,                           # x-axis
      at= plot.example$dec.time,        # location of marks
      labels=plot.example$time)         # label for marks
legend ("bottomleft", bty="n",
        pch=c(19,19),
        col=c("red", "black"),
        legend=c("Ind. in", "Ind. out"), cex=1.3)
```

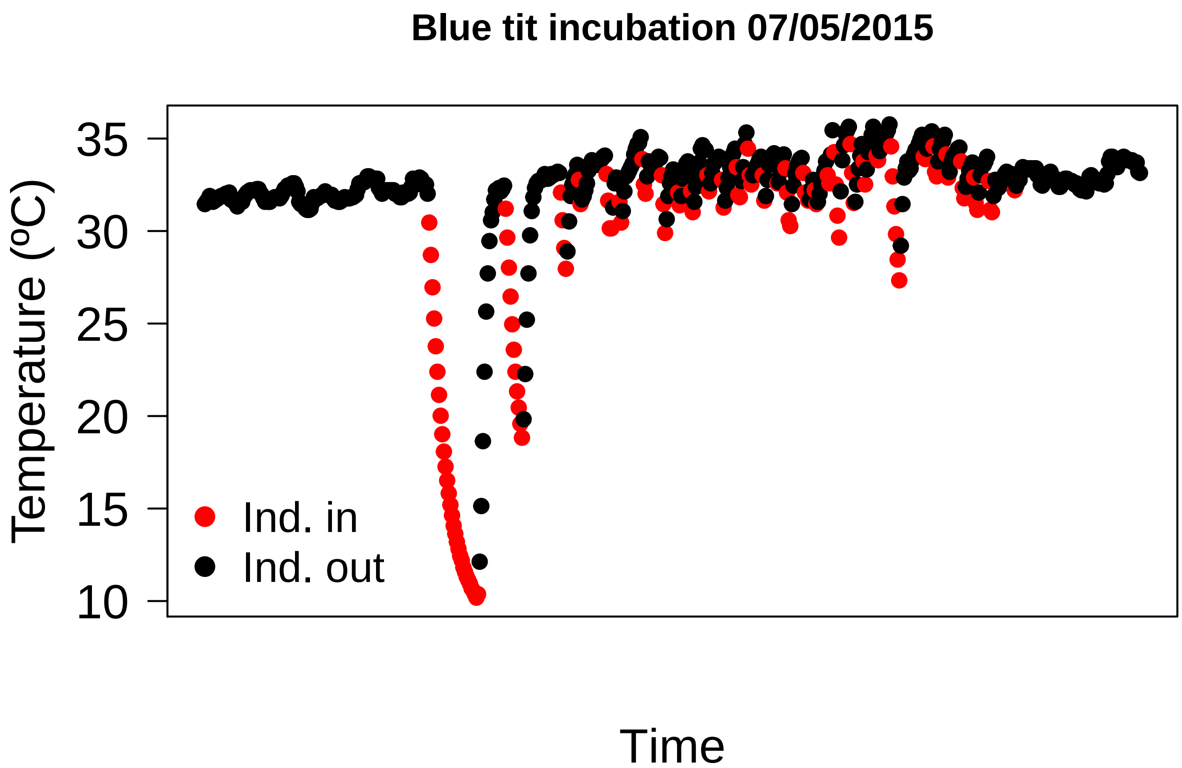

Figure 1. A day of blue tit incubation. Red and black dots represent off- and on-bout periods, respectively.

2.4 - Using incR to extract biologically relevant metrics of incubation

2.4.1 - Daily activity times

First, I calculate onset and offset times of activity using the function *incRactivity*.

```
df.activity <- incRactivity (data=df.incubation,
                             vector.incubation ="inc.vector")
```

Have a look at the new data frame with activity times:

```
df.activity
```

```
##         date    onset   offset
## 1 2015-05-07 5.783333 20.81667
## 2 2015-05-08 5.866667 13.23333
```

It contains onset and offset of activity per day. **WARNING**: this function returns the first off-bout and last on-bout of a day. If the time series is incomplete for one day, onset or offset times should be discarded, as it happens in the example above for the offset time on 2015-05-08. On that day the temperature recording ended around 13:30, and the last on-bout was around 13:15, but it does not biologically represent the offset of activity (see table above). One should check that the days of analysis are completed before concluding on the biology of the results.

You can include activity times in the plot above (the code is not shown) to check the performance of the function:

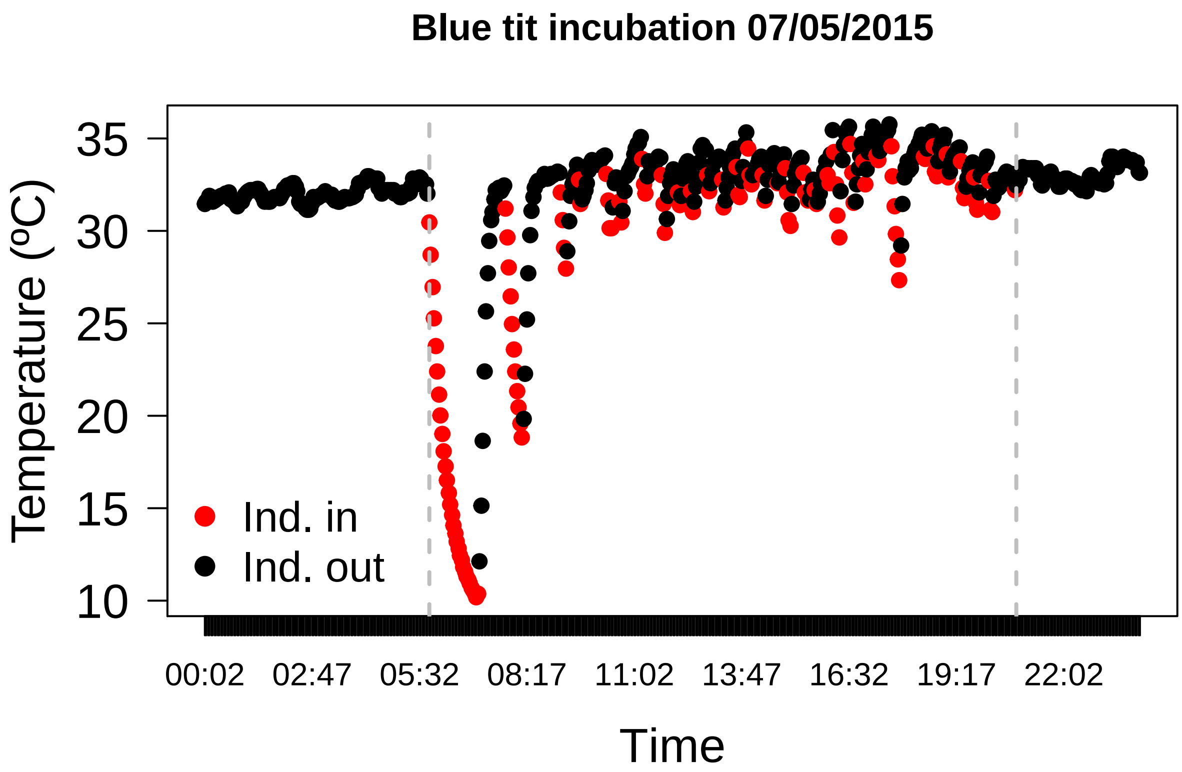

Figure 2. A day of blue tit incubation. Onset and offset estimations are illustrated by grey-dashed lines.

2.4.2 - Daily incubation constancy

The function *incRconstancy* automates the calculation of the daily percentage of time the incubating individual stayed in the nest. The current version of the *incR* package performs this calculation over 24h. Future developments will allow the separation between day/night or other specified time windows. *incRconstancy* assumes that the rate of sampling is constant over the analysed period of time.

```
inc.constancy <- incRconstancy (data=df.incubation,
                                vector.incubation ="inc.vector")
inc.constancy
```

```
##         date  perc.in
## 1 2015-05-07 78.64583
## 2 2015-05-08 67.08075
```

Results are stored in a data frame and for each day of analysis the proportion of time in the nests is shown under the *perc.in* column.

2.4.3 - Number of on- and off-bouts per day

```
# calculation of sampling rate - which is constant for the entire time-series
sr <- df.incubation$dec.time[5]-df.incubation$dec.time[4]
bouts.example <- incRbouts (data=df.incubation,
                            vector.incubation="inc.vector",
                            sampling.rate=sr)
bouts.example
```

```
##         date number.on.bouts number.off.bouts mean.time.on.bout
## 1 2015-05-07               31               32         0.2559140
## 2 2015-05-08                6                7         0.3944444
##   mean.time.off.bout
## 1          0.1281250
## 2          0.5047619
```

A five-column data frame is produced with one day per raw. Date, number on-bouts, number of off-bouts (number of on-bouts + 1 by definition) and mean time duration of on- and off-bouts are displayed in the 5 columns respectively. Mean times are shown in the time units you specify *sampling.rate* - decimal hours in this example.

2.4.4 - Temperature variation for customised time intervals

incRt allows the user to calculate temperature averages and variation between two time windows within 24 hours. It was originally thought to calculate incubation temperatures between day and night. The current version includes three ways of defining "day" and "night". Here I use the onset and offset of activity times calculated by *incRactivity* to define the two day periods, *day* and *night* if one likes. If this option is chosen, one must check that there is a column under the *inc.vector* name giving the position of the incubating individual for each time point:

```
is.null(df.incubation$inc.vector)
```

```
## [1] FALSE
```

```
head (df.incubation$inc.vector, 5)
```

```
## [1] NA  1  1  1  1
```

Check the *incRt* help page to find out more about the functionally of this function. Once that has been checked, the function can be run:

```
incRt (data=df.incubation,
       temp.name = "valueT",
        limits=NULL,
        coor=NULL,
        civil.twilight=FALSE,
        activity.times=TRUE,
        time.zone="GMT")
```

```
##         date day.mean   day.var night.mean night.var
## 1 2015-05-07 30.97323  31.89103   33.69237 0.4758604
## 2 2015-05-08 20.73246 168.97885         NA        NA
```

```
?incRt
```

In this example I calculate temperatures based on activity times estimated by *incRactivity*. Note that there are no night calculations for the 8th of May. Such night is defined by the offset of activity of the 8th and the onset of activity of the 9th. There are no data for the 9th and, therefore, such calculation cannot be made.

---

3 - Final notes

At the beginning of a study it is highly recommended to have incubation validated data (using, for instance, video-recordings). With few days of video and temperature recordings one can calculate the optimal values for `incRscan` and automate the analyses of full data sets with little extra effort.

The current version of the package is expected to be further polished. Please, if you spot any error, bug or spelling problem in the documentation, send me an e-mail to correct it. The code of the package is publicly available, so if you think you can improve it, I would be delighted to get your feedback.

4- References

- Cooper CB, Mills H. 2005. New software for quantifying incubation behavior from time-series recordings. Journal of Field Ornithology 76(4): 352-356.