**STEP-BY-STEP PROCEDURE FOR DATA CURATION**

This document provides a summary of the steps followed to get final and cleaned data from nightjar incubation. Its main goal is to serve as a streamlined procedure for a quick and standardized data processing in future studies with incubation data conducted with red necked nightjar (*Caprimulgus ruficollis*) in Murcia. Field data from nightjar nest monitoring must be collected through a standardized sampling protocol, which will be soon described in a parallel protocol. These data are recorded in a field notebook that includes the same columns and codes as indicated in the *uni\_parent* datasheet, this way allowing for a quick and free-mistake data entering.

1. **Data entering**

When fieldwork has finished, data must be entered in the *Metadata\_nightjar\_incubation* Googlesheet ([here](https://docs.google.com/spreadsheets/d/1h4S3FpTtyno_3E6xr9U77j8tv9ooRnSSAEeVJm1pHIM/edit?gid=1387931867#gid=1387931867)) by following the instructions provided therein (ReadMe sheet). More specifically, the following sheets must be properly full filled with data from fieldwork: *loggers*, *authors*, *species*, *site*, *nests*, *inc\_lay\_esc*, *visits*, *eggs*, *captures* and RFID*.* At this point, it is important to note that some sheets include a column named “MB\_check” which must be completed with “check” if you want to check for potential errors or data incongruences. The *acto\_based* and *use* sheets will be later fulfilled once the actograms have been drawn. All raw data, which are “.msr” (temp), “.csv” (temp) and “.txt” (RFID) must be placed in the corresponding folder (Data) before conducting data checking or analyses. At this point, be careful with the duplicated names of some files. In cases of duplicated files from different names, renamed them to avoid duplications.

1. **First data checking**

After data entering, the nest step is to run the codes named *DB\_backup, DB\_checks* and *DB\_visualise* (in this order). The first code (*DB\_backup*) is aimed to save the Googlesheet as an RData file and to move older files to the freeze folder. Here, it´s better to work directly from the googlesheet (by setting API key). The second code (*DB\_checks*) is aimed to run a detailed checking of the data to look for potential errors or incongruences. This code produces a report (saved in the Reports folder) with the list of all potential errors detected. This report must be carefully checked, and all errors must be corrected (when possible) directly in the source googlesheet. After that, the code *DB\_checks* must be run again to generate an updated report with errors and then confirm that all errors have been properly corrected.

1. **Generating actograms**

Hence, the third code (*DB\_visualise*) must be run to generate actograms. Once generated, it is recommended to take a quite look to each actogram to check everything is congruent with raw data. For example, it is very useful to check it the day of start and day of end in the actograms are the same as in the googlesheet, because sometimes actograms are cut due to any analytical criteria. Also recommended to check if visits are properly depicted and with the congruent (or expected) codes. Importantly, all actograms must be also screened to check if the assignation of each probe to the nest or control site is correct. For example, in the correct way, temperature at nest must be generally lower than control temperature at nighttime, whereas the opposite is expected at daytime. If this is not the case (even, we find the opposite pattern), then probes were swapped, and they need to be switched through the code.

1. **Fulfilling *acto\_based* and *use* sheets**

Once actograms have been generated, then they must be carefully screened to extract the datetime of relevant moments (e.g. predation, nest desertion, hatching, nest desertion, nest leaving by chicks, returning of the chicks to the nest after leaving, returning of the adults do the nest after nest failure, etc) during the incubation process. This information is frequently represented by steep changes in temp. or HR within the actogram. For example, egg-hatching is represented by a steep decrease in HR for the nest-probe. The nest leaving behaviour may be identified as the end of the last incubation bout, thus meaning that adults and chicks definitively leave the nest scrape. All this information must be used to fulfil the *acto\_based* sheet. At this point, it worth to note that information from videos can be particularly useful to confirm the exact datetime of a given moment. Then, both actograms and videos (or metadata from videos) must be used to fulfil the *acto\_based*. After that, then the *use* sheet must be also fulfilled to indicate what type of data (and its usefulness for analyses) are containing in each nest.

1. **Updating actograms**

The *DB\_visualise* code must be run again to get updated actograms including the information from the *acto\_based* and *use* sheets. At this point, it is necessary to change the “extracted” function in the code from FALSE to TRUE (line 26). Then, actograms are again generated. They must be carefully checked again to confirm that the information from *acto\_based* is perfectly depicted and its matches with the info contained in the actogram itself. All these actograms are generated by following the Martin Bulla’s protocol.

1. **Running Hidden Markov models (HMMs)**

To compare the performance of a different data curation procedure, we run also HMMs. These models calculate hidden states (incubation gaps or bouts) as inferred from the raw data, and they provide an accurate estimate of when nightjars are incubating or not (read the document *protocol\_hmm* for detailed information). To run them, we will use all RData files contained in the “to\_extract” subfolder (“Data” folder). Moreover, all Actograms must be also copied to the “Actograms\_all” folder. Hence, the HMM can be run, and they will generate new actograms (hereafter, HMM actograms) but with a different layout as compared to the Bulla’s actograms. For example, HMM actograms do not incorporate a vertical line depicting each hour, neither an indication to twilight or nighttime, neither timing of *acto\_based* events or the daily percentage of nest attendance. However, HMM actograms contain a more accurate delimitation of incubation bouts and gaps (green bars), which facilitates the manually edition process of the actograms. Moreover, variations in temperature and HR are more precisely depicted in the HMM actograms, thus allowing to easily distinguish between real or false incubation gaps.

1. **Post-processing of HMM data (actogram manual edition)**

Once HMM actograms have been obtained, data from HMM can be manually edited by running the code *manual\_postprocessing\_adjust\_NEW* and by applying changes directly into the actograms. This allows to correct mismatches in estimates from HMM regarding to incubation bouts and gaps (green bars). There is a particular protocol describing all the relevant information and codes necessary to proceed with manual actogram edition (*protocol\_manual\_editing*). During the manual edition process, last versions of Bulla’s actograms and HMM actograms must be checked to decide which bouts can be edited or not. For this, the information from videos must be particularly useful (mostly at night) to identify brief incubation gaps (depicted in the actogram has minimal temp variations) or to remove the noise from adult body-shaking behaviour (typically at midday, which could be wrongly interpreted as incubation gaps) at midday. The edition process takes some time because the zoom-in and zoom-out option need 20-40 seconds to update the viewer, so the use of a big enough screen or a powerful server is recommended to reduce time consumption.