MRC | Epidemiology Unit

11.0. Processing Free-Living data through PA Java Software (GPR processing)

The PA Java Software implements the method of cleaning heart rate data described in "*Gaussian Process Robust Regression for Noisy Heart Rate Data*", (Stegle et al, IEEE Trans Biomed Eng, 2008)
This Java code processes the information in two stages:
- Noise classification by Bayesian Clustering
- Gaussian Process Regression to infer latent time series

Essentially, the raw time-series output from the AH is taken in: (Sensor acceleration, heart rate, auxillary data (when available)), along with an initialising set of hyperparameters (defined in the study profile).
A time-series is returned (at a user-specified resolution) relating to: cluster membership (ie. clean/intermediate noise or very noisy), predicted acceleration & heart rate along with the upper and lower confidence intervals.

These "cleaned" files are then fed into STATA for translation to Energy expenditure estimates, generating hourly and summary measures.

11.1. Setting Region of Interest (ROI)

If the files have had "Trim end" time set (as described in section 6.1.1), this will be automatically read and displayed as the "Region of Interest End" by the Java software (see example below).

At present, the java software does not upload the "Trim Start" from the header of the raw text export in the same way as "Trim End" (see example below). Therefore, this has to be done manually in the Java software.
Any files that have been trimmed can be identified in the ReadInfo table of the Actiheart database by looking at the TrimStart Column.

- Double click on file of interest to load into Free-Living GUI (Graphic User Interface).
- If monitor was worn immediately following initialisation of monitor, and trim-end has been selected, this will already have the ROI correctly set (as example below).
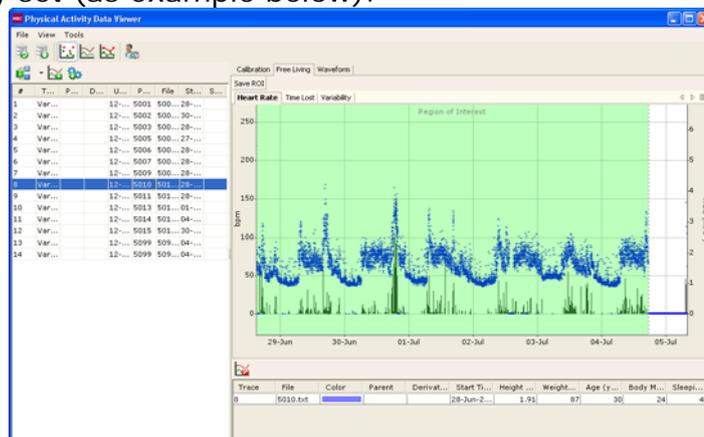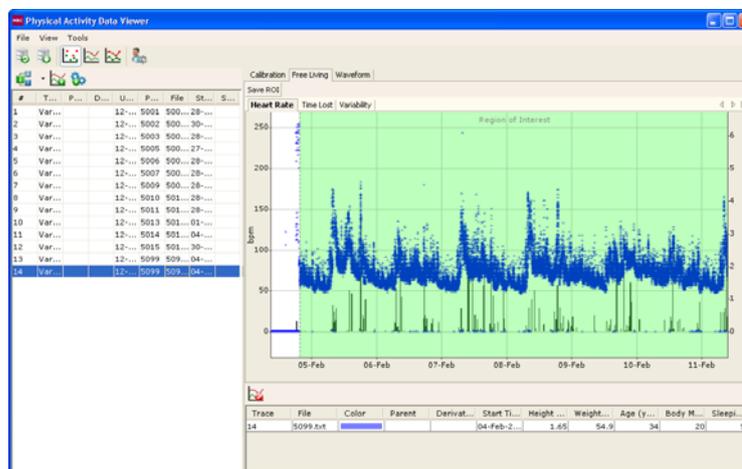
- If monitor was not worn immediately (ie. if monitor posted out for instance), this will need ROI start being set.
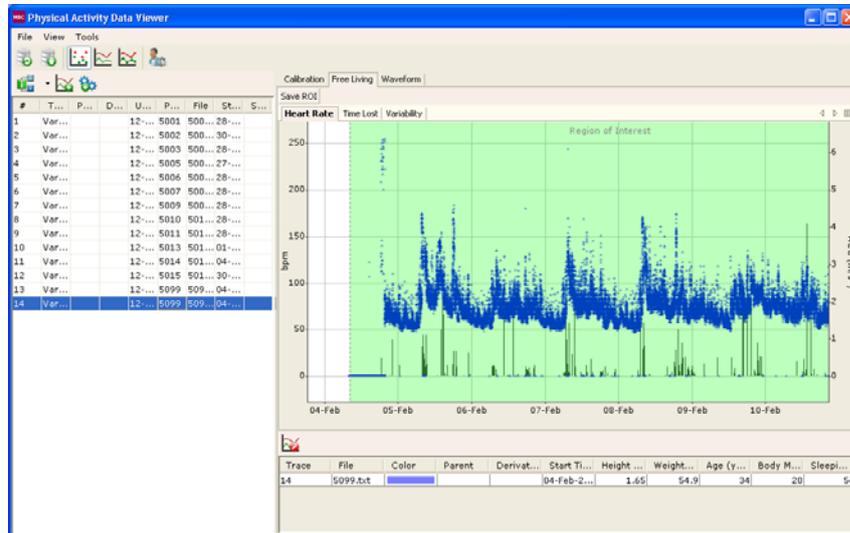
- Click on GUI and zoom in (by rolling ball on mouse) on where monitor has started to be worn (where HR trace begins).
- Right click on the mouse, click "Region of Interest Start"
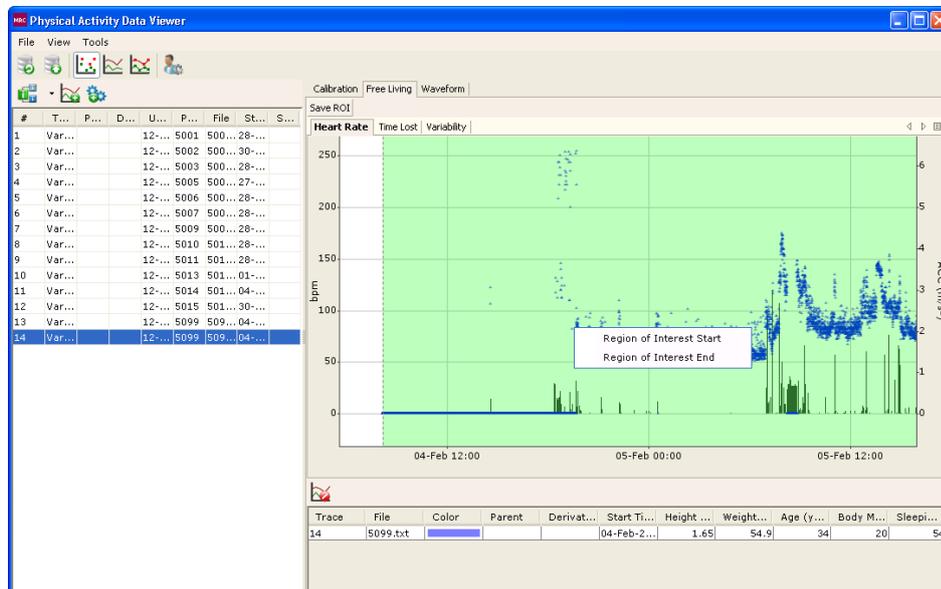
- Click "**Save ROI**" at the top of the Free Living tab.

- **REFRESH the database before queuing files for processing**
- *Note: Removing this data from the ROI means that processing is more efficient as it is not processing additional data before or after the monitor is worn. (In the event that a file has been processed and this Start or End ROI is not set, check that the wear probability parameter has picked this up correctly in the Reviewing STATA plots phase).*
- *Note: If files have been pre-trimmed in commercial software, and was only necessary for a few files, it is possible to view in Read Info table "TrimStart" column to see which files require a pre-trim in Java software.*
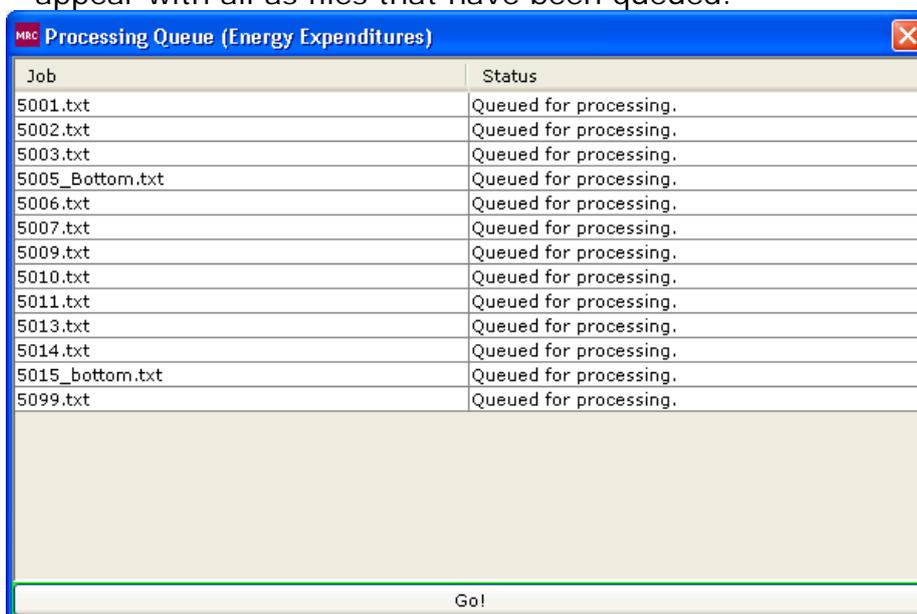
11.2. Selecting files for processing

- Flag red those files that you do not wish to process by clicking on status column relating to the file of interest and selecting the red cross:
- Grouping list by participant helps identify where participant has 2 x free-living files.
- Where a participant has 2 files (for example a re-wear) it may be necessary to only send the better quality trace through.
- Flag as red:
    - Any duplicate files (to avoid duplicated processing)
    - Completely corrupt files
    - Files that appear with no data at all (verify this is expected)

11.3. Queuing files for processing

- Highlight files ready to queue for processing (suggest group by "none" & sort by status. Therefore, if files have been flagged red, these are not selected for processing). Flagging red does not mean if selected will not go through GPR. Red flags must not be selected.
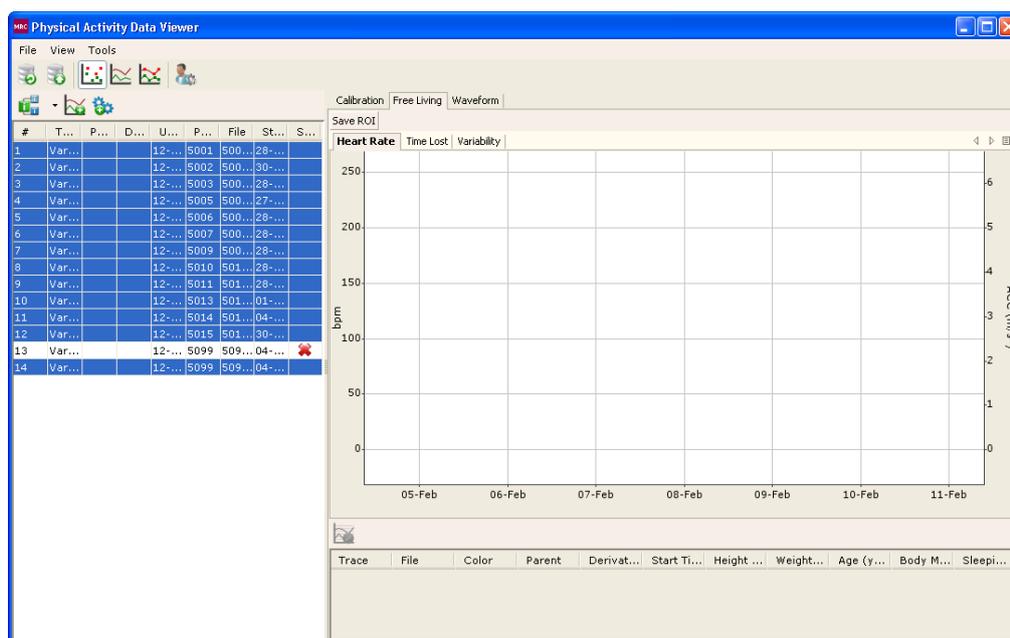
- Click "Queue dataset" icon.

- Once these files have been added to the queue list, the list will appear with all as files that have been queued.

- *NB: It is advisable to make a note of the numbers of the first and last one in the queue as there is no indication once the file has been processed.*
- Example timings for GPR processing: (Dependent on computer processing capability)
    - 10 x 6 day files (60 sec epoch), inferred time resolution of 1.0 takes approximately 55mins
    - 10 x 6 day files (15 sec epoch), inferred time resolution of 0.25 takes approximately 2hrs 40mins

- Once the last file in the list has "Finished processing", scroll up the list to check all have completed successfully.
- The command window of the software will indicate when processing has finished.
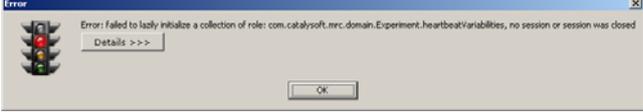

- In the event that errors appear, check the file by viewing in the Java software to see whether the file has any useful information present. If you expect that it should have processed, try re-processing & flag to PA TechTeam
- If further processing needs to be done in the same database, close and reopen the software, and queue the next batch of files (remember to sort in the same way as last time to avoid missing/duplicating files).

11.4. Problem solving

Error	Possible Cause
Error out of memory	Files were still being added to queue list once first file was still processing. If 2nd file processes ok, leave to continue that batch, check export & will probably have to push first file through again. If error persists check job size window is set correctly in .xml
Submatrices index	Usually from files which are completely corrupt or have no data in them. Check through original commercial software to view.
When starting to queue a batch of files	Usually when finished a batch of processing, and starting to queue the next batch without

Error: Failed to lazily initialize a collection of role: com.catalysoft.mrc.domain.Experiment.heartbeatVariabilities, no session or session was closed Details >>> OK	closing and re-opening the software.
Error: Index 1, Size 1	This appears when files have no data in them, but have been processed
File processes the entire file rather than that which was marked as ROI	Following changing an ROI, must click "Refresh" or close and re-open software before queuing.