TAPPED Model Readme File

**Contents of the TAPPED folder**

* **PPE\_Change\_Detection\_Phase\_Estimation\_Model\_v2.R**
  + This is the R script that loads and formats the data and for the TAPPED model to be run.
  + The scripts runs two different JAGS models and then formats the results from both.
* **PPE\_Simple\_Change\_Detection\_Cal\_Pred.txt**
  + This runs the first stage of the TAPPED model defining the change points within the learning curve of a participant.
  + The change detection procedure is taken from this article.
    - Lee, M. D. (2019). A simple and flexible Bayesian method for inferring step changes in cognition. *Behavior research methods*, *51*(2), 948-960.
* **Simple\_Stage\_Estimation\_Model.txt**
  + The stage estimation model takes in the human data from each segment of an individual’s learning curve and then estimates what learning phase that generated that segment of the learning curve the work done by Tenison & Anderson, (2016)
    - Learning phases are estimated based on Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 749.
* **Simple\_Stage\_Estimation\_Model.txt**
  + I also included data from one participant of the data in a spacing study that was analyzed and published in my previous ICCM paper.
* **PPE\_Time\_Functions.R**
  + This R function that will compute the necessary PPE time functions

**How to Run the TAPPED model**

1. Open the PPE\_Change\_Detection\_Phase\_Estimation\_Model\_v2.R and make sure to install that necessary R packages, “dplyr”, “Hmisc”, and R2Jags.

Note: You will also need to install JAGS software to run the Bayesian models. JAGS can be downloaded at  <http://sourceforge.net/projects/mcmc-jags/files/>.

1. Once the data necessary packages and software is uploaded you should be able to read in the data that is in the TAPPED folder and run the software.
   1. The PPE\_Change\_Detection\_Phase\_Estimation\_Model assumes that all of the JAGS scripts and datafile are in the folder together, so I would suggest at least for a first step keeping all of the contents together.

Side Note: the model script assumes that the PPE’s model time functions are already in the data, which is the case in the example data I gave you. In the TAPPED model folder I have included PPE time variable functions, which calculates the necessary time variables needed for the PPE. A broader description of the PPE time variables can be found in following variable

Walsh, M. M., Gluck, K. A., Gunzelmann, G., Jastrzembski, T., & Krusmark, M. (2018). Evaluating the theoretic adequacy and applied potential of computational models of the spacing effect. *Cognitive science*, *42*, 644-691.

1. I’ve gone through and annotated the model file. Overall the model file will run two separate JAGS models. First, the model will determine change points within the participants learning curve. Second, after the change points have been determined the model will estimate the learning phase for each change point. After each of the models have been run all of the information is saved from the models and stored in the output.
2. After the Model is run over the data is outputs the formatted results into a table in the TAPPED folder.
3. Model output
   1. Specifics of the Experimental Set up

"Participant" "Day" "Block" "Pnum"

"Problem" "Base" "Height" "Answer"

"Practice" "Trained" "Seen" "Time"

"Correct" "ComputeTimeout" "InputTimeout" "ComputeDuration"

"InputDuration" "FeedbackDuration" "BlankDuration" "BackspaceCount"

"Input" "Interference4" "InterDist4" "Interference8"

"InterDist8" "TotalSolved" "Phase" "Likelihood1"

"Likelihood2" "Likelihood3" "CountPhase" "AllExpose"

"ExactSpacing" "D1RT" "LIK1" "LIK2"

"LIK3" "CNTPHS" "PHS" "allPHS"

"allCNTPHS" "Experiment\_Instance" "Day\_Trial" "Milli\_sec"

"Cumulative\_Time" "Time\_Stamp\_v2" "ID"

* 1. PPE Specific Variables

"Instances"

"TrialTimeSec" "TrialTimeMin\_diff" "TrialTimeSec\_sumDiff" "TrialTimeSec\_sumDiffneg75"

"T" "N" "c" "s"

"InvLogTime\_diff" "CumAveInvLogTime\_diff" "CumAveInvLogTime\_diff\_lag1"

* 1. PPE Prediction.
     1. PPE\_1, PPE\_2, PPE\_3, PPE\_4, and PPE\_5 are the mean predictions from each of the individual change points.

"PPE\_1" "PPE\_2" "PPE\_3" "PPE\_4" "PPE\_5"

* + 1. H\_HDI\_PPE\_x & L\_HDI\_PPE\_x is the 95% HDI interval for the mean predictions for each of the predictions

"H\_HDI\_PPE\_1" "H\_HDI\_PPE\_2" "H\_HDI\_PPE\_3" "H\_HDI\_PPE\_4"

"H\_HDI\_PPE\_5" "L\_HDI\_PPE\_1" "L\_HDI\_PPE\_2" "L\_HDI\_PPE\_3"

"L\_HDI\_PPE\_4" "L\_HDI\_PPE\_5"

* 1. These are different types of predictions that I explored with the TAPPED model, using the full information from the full posterior distribution, either taking an overall average, weighted average, Weighted Median from all of the detected change points. I never really used these in any of my papers, but they are there if you need them.

"Overall\_Average"

"Weighted\_Average"

"Weighted\_Median" "Weighted\_Mode" "H\_HDI\_Overall\_Average" "H\_HDI\_Weighted\_Average"

"H\_HDI\_Weighted\_Median" "H\_HDI\_Weighted\_Mode" "L\_HDI\_Overall\_Average" "L\_HDI\_Weighted\_Average"

"L\_HDI\_Weighted\_Median" "L\_HDI\_Weighted\_Mode"

* 1. "Cal\_Point", the point up to where the model calibrated
  2. These are the individual change points that were identified by the model

"Tau\_Prime\_1"

"Tau\_Prime\_2"

"Tau\_Prime\_3"

"Tau\_Prime\_4"

* 1. These variables are also a hold over from some explorations that I was doing with the change detection analysis

"w\_round\_mean"

"w\_round\_median" "w\_round\_mode"

* 1. These variables set the weights for each of the change points

"Weight\_Set\_5" "Weight\_Set\_4"

"Weight\_Set\_3" "Weight\_Set\_2" "Weight\_Set\_1"

* 1. Change Point variables

"Stage\_Marker" - This represents each of segments of the learning curves discovered by the change point detection. A 0 means that these were predicted portions of the learning curve.

"Estimated\_Stage" - These are estimated learning phase of each segments.

1. 1 - Declarative Phase
2. 2 – Associative Phase
3. 3 – Procedural Phase
   1. The rest of the columns are hold over variables from some model exploration that I was doing. They will likely not be of interest to you.

"Stage\_1\_Freq" "Stage\_2\_Freq" "Stage\_3\_Freq"

"N\_Stage" "N\_Stage\_L\_HDI" "N\_Stage\_H\_HDI" "D\_Stage"

"D\_Stage\_L\_HDI" "D\_Stage\_H\_HDI" "A\_Stage" "A\_Stage\_L\_HDI"

"A\_Stage\_H\_HDI" "P\_Stage" "P\_Stage\_L\_HDI" "P\_Stage\_H\_HDI"

"All\_Stage" "All\_Stage\_L\_HDI" "All\_Stage\_H\_HDI" "AP\_Stage"

"AP\_Stage\_L\_HDI" "AP\_Stage\_H\_HDI" "DA\_Stage" "DA\_Stage\_L\_HDI"

"DA\_Stage\_H\_HDI"

**Additional Notes and thoughts about generalizing the data to surgical performance data.**

* This model is set up to calibrate to a subset of a participant’s performance during the first day and then predict their performance on Day 3. If you are using the model to only calibrate then you will need to modify the “Num\_Cal” variable.
* To determine the learning phase (i.e., Declarative, Associative, and Procedural) you currently need response time data. If you do not have response time data then you might just want to run the first model only to determine change points within a learning curve.
* The likelihood function for the performance model is a beta distribution, which takes in performance data between .01-.99. You may need to change the likelihood function depending on the characteristics of your performance value (e.g., Bernouli, Binomial, etc). However, modifying the likelihood function should not change any of the other code.