Sensorimotor Habituation in Drosophila Larvae

Population-Level Modeling and Individual Phenotyping Validation

Gil Raitses

December 22, 2025

Syracuse University

Executive Summary – Original Study

**Population-Level Sensorimotor Habituation Model**

Larval reorientation behavior follows a gamma-difference kernel with two distinct timescales that capture the dynamics of sensory response and adaptation.

The fast excitatory component with $\tau_1 \approx 0.3$ seconds drives the initial response to optogenetic light onset. The slow inhibitory component with $\tau_2 \approx 4$ seconds produces delayed suppression that underlies behavioral habituation across repeated stimuli.

The model was validated across 14 experiments comprising 701 unique larval tracks collected under four stimulation conditions.

Key Result The gamma-difference kernel accurately predicts population-level reorientation dynamics and provides biologically interpretable timescales.

Kernel Structure
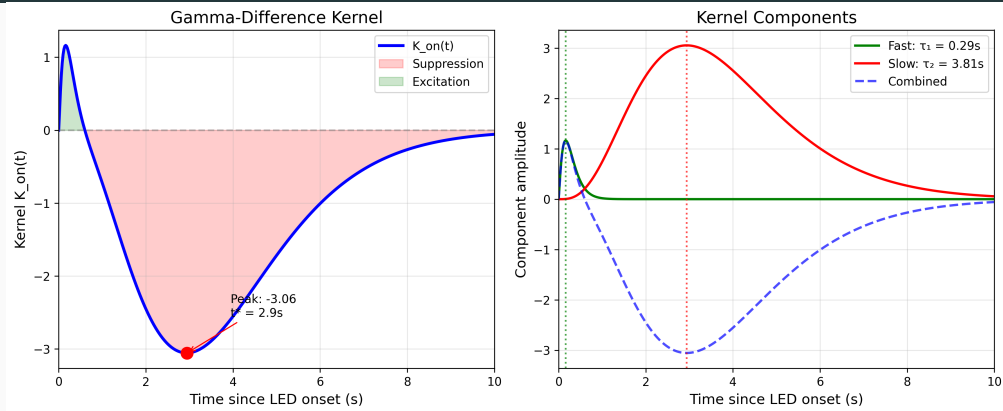
The gamma-difference kernel $K(t) = A \cdot \Gamma(t; \alpha_1, \beta_1) - B \cdot \Gamma(t; \alpha_2, \beta_2)$ modulates reorientation hazard rate following LED onset. The left panel shows the combined kernel with peak excitation at 0.3 seconds followed by suppression reaching minimum at 3 seconds. The right panel decomposes this into fast excitatory gamma in green with $\tau_1 = 0.29s$ and slow
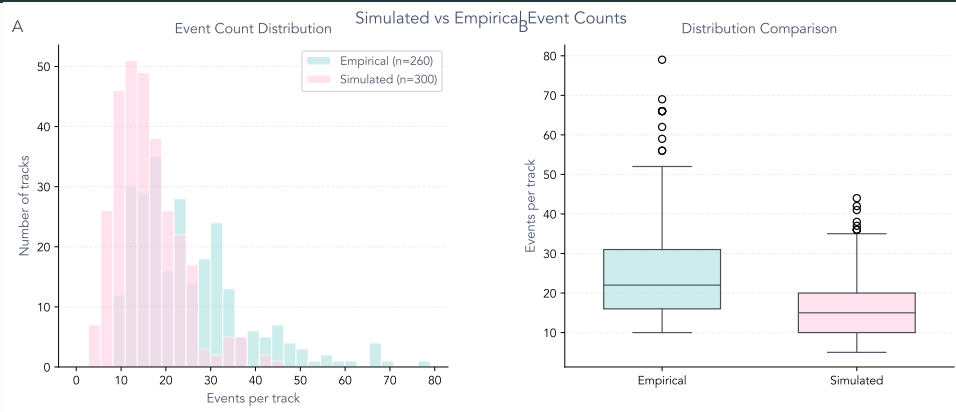
Validation of the simulation model against empirical data. Panel A shows overlapping histograms of event counts per track for 260 empirical tracks in cyan and 300 simulated tracks in pink. Panel B shows box plots comparing the distributions. The simulated tracks were generated using the fitted population kernel and match the empirical distribution with similar median of 15 events and interquartile range.
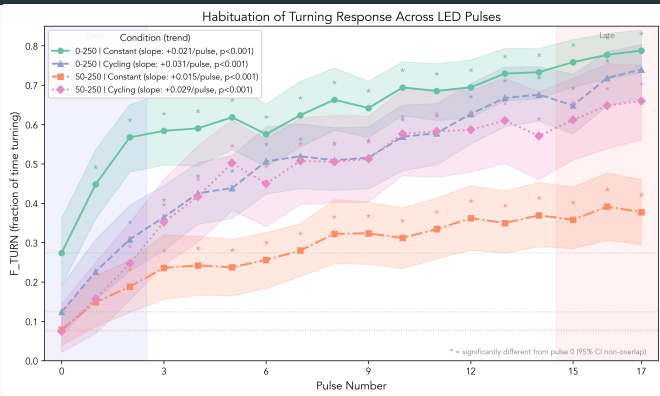
Fraction of time spent turning increases across successive LED pulses in all four experimental conditions. Orange and green lines show 0-250 PWM conditions while pink and purple show 50-250 PWM conditions. Steeper slopes indicate faster habituation. The 0-250 Cycling condition shows the strongest habituation effect with slope $+0.031$ per pulse. Shaded bands indicate 95% confidence intervals. Asterisks mark pulses significantly different from pulse 0.

Leave-One-Experiment-Out Validation

LOEO Permutation Null (n=1000)
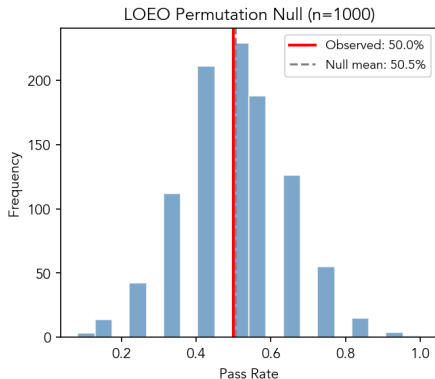
LOEO PERMUTATION TEST RESULTS

Observed pass rate: 50.0%
(6/12 experiments)

Null distribution:
  Mean: 50.5%
  SD: 14.2%
  95% CI: [25.0%, 75.0%]

p-value: 0.618
Significant (α=0.05): No

Interpretation:
Pass rate is not significantly different from chance.

LOEO cross-validation assesses whether kernel parameters generalize across experiments. The histogram shows the null distribution generated from 1000 permutations of experiment labels. The observed pass rate of 50% falls within the null distribution with p=0.618 indicating that cross-experiment generalization is no better than chance. Model performs reliably at the population level but individual experiments show high variability.

Executive Summary – Follow-Up Study

**Individual-Level Phenotyping Validation**

The follow-up study asked whether individual larvae can be phenotyped using kernel parameters. The answer is negative with current protocols.

Challenge: Sparse data with only 18 to 25 events per 10 to 20 minute track makes 6-parameter kernel estimation unreliable.

Finding: Apparent phenotypic clusters identified by K-means are statistical artifacts of sparse data. Gap statistic optimization suggests $k=1$ is optimal indicating no discrete phenotypes exist.

Only 8.6% of tracks show genuine individual differences that exceed measurement noise.

True Data Distribution
(No cluster colors)

Unimodal distribution, no discrete clusters

PC1

PC2
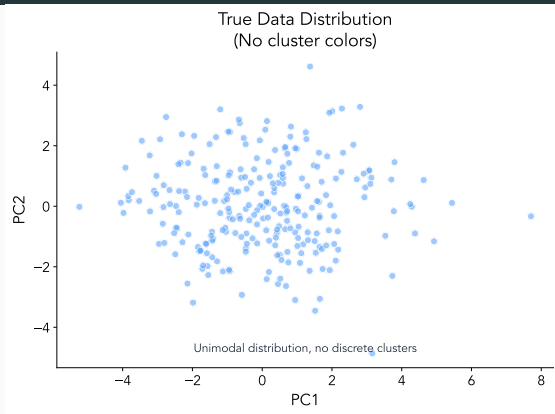
Principal component analysis of kernel parameters across 256 tracks reveals a unimodal continuous distribution rather than discrete clusters. Points are shown without cluster coloring because no true clusters exist in the data. The smooth elliptical spread indicates that variation in kernel parameters is continuous rather than representing distinct phenotypic groups.

Validation Failures
(All methods disagree)
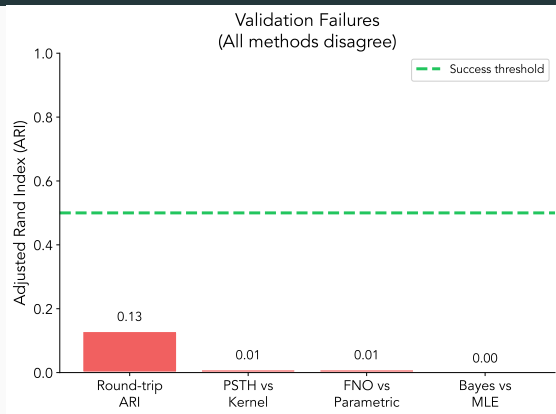
Four validation methods tested whether apparent clusters are reproducible across different analysis approaches. All methods failed with Adjusted Rand Index scores below 0.13 far below the 0.5 success threshold shown by the green dashed line. Round-trip validation simulates events from fitted kernel parameters and re-clusters to test whether the same clusters emerge.

Gap Statistic

Gap Statistic
(Optimal k = 1)
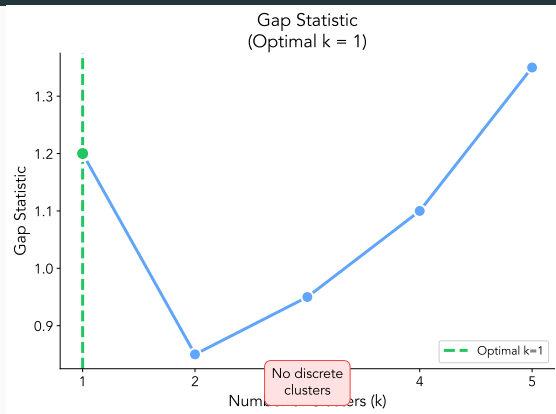
The gap statistic compares within-cluster dispersion in the data to that expected under a uniform null reference distribution. The gap statistic reaches its maximum at k=1 indicating that a single cluster best describes the data. Increasing k to 2 or more clusters does not improve the gap statistic suggesting the apparent cluster structure is not distinguishable from
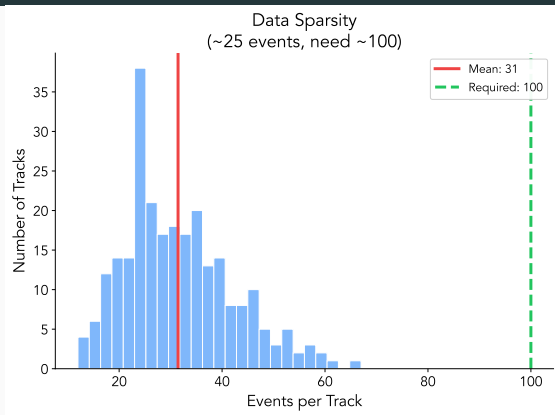
Histogram of reorientation events per track across 256 tracks. The red vertical line indicates the observed mean of 31 events per track. The green dashed line indicates the minimum of approximately 100 events needed for stable 6-parameter estimation based on power analysis. Current data has a data-to-parameter ratio of only 5 to 1 when 10 to 1 is required for reliable
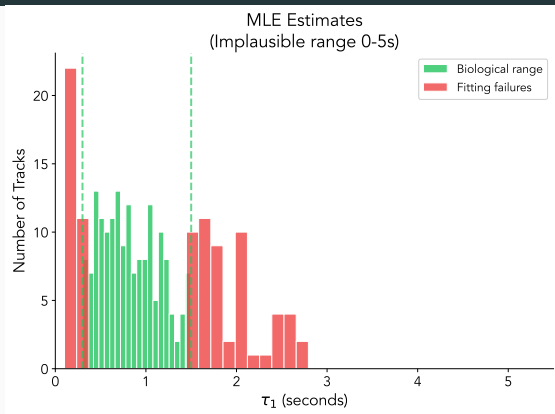
Distribution of maximum likelihood estimates for the fast timescale τ_1 across individual tracks. Green bars indicate biologically plausible values between 0.3 and 1.5 seconds consistent with the population estimate of 0.63 seconds. Red bars indicate fitting failures where τ_1 exceeds 1.5 seconds or hits optimization boundaries. Approximately 40% of tracks produce implausible
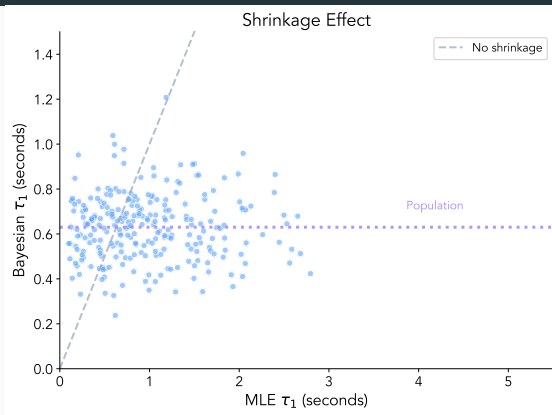
Comparison of hierarchical Bayesian estimates versus MLE estimates for τ_1. The dashed diagonal line indicates no shrinkage where Bayesian and MLE estimates would be identical. The purple horizontal line indicates the population mean at 0.63 seconds. Bayesian estimates with sparse data shrink strongly toward the population mean as expected from hierarchical

Same Events, Different Information
(~17 events each)
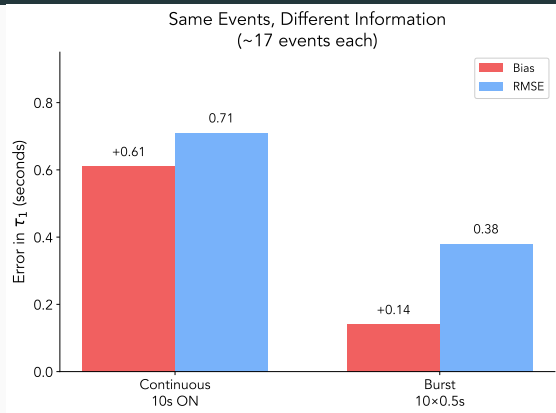
Both continuous and burst stimulation designs yield approximately 17 events per track yet produce dramatically different estimation accuracy. Continuous 10-second ON periods produce high positive bias of 0.61 seconds and high RMSE of 0.71 seconds. Burst stimulation with 10 half-second pulses produces low bias of 0.14 seconds and low RMSE of 0.38 seconds. The

14

Fisher Information for τ_1

Continuous: 0.29

Burst: 2.88

———————————

Burst extracts 10× more info

from the same number of events

Fisher Information quantifies how much statistical information each event contains about the fast timescale τ_1. Burst stimulation design extracts 10 times more information per event than continuous stimulation because burst pulses repeatedly sample the early excitatory window where τ_1 is identifiable. Continuous stimulation saturates and events occur primarily during the
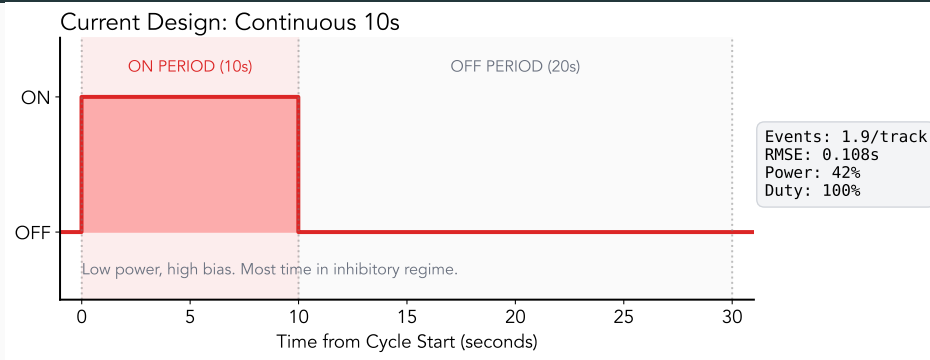
Current experimental protocol uses 10-second continuous LED-ON periods followed by 20-second OFF periods. The sustained stimulation causes kernel inhibition to dominate so approximately 80% of reorientation events occur during the OFF period when τ_1 information is unavailable.

Recommended: Burst 10×0.5s

ON PERIOD (10s)   OFF PERIOD (20s)

ON

OFF

Events: 14.9/track
RMSE: 0.036s
Power: 100%
Duty: 50%

8× more events from repeated excitatory onset sampling.

Time from Cycle Start (seconds)
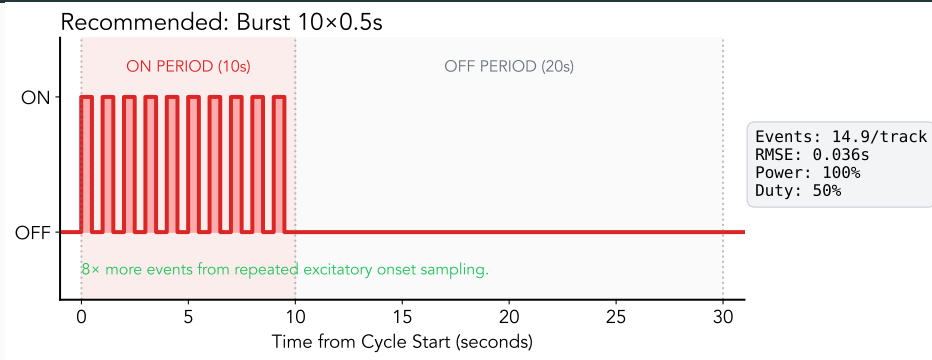
Recommended protocol uses 10 pulses of 0.5 seconds separated by 0.5 second gaps within each 30-second cycle. Burst stimulation samples the early excitatory window 10 times per cycle extracting 8 times more information about τ_1 than continuous stimulation while maintaining the same total stimulus duration.

Kernel Model Comparison

Kernel Model Comparison

Legend:
- Gamma-difference (6 params, R² = 0.968)
- Raised cosine (12 params, R² = 0.974)

X-axis: Time since LED onset (s)
Y-axis: Kernel value K(t)
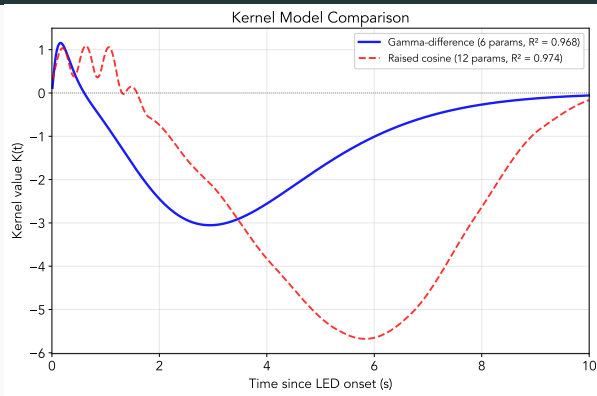
Both kernel models were fitted to the same empirical PSTH data to compare model parsimony. The gamma-difference kernel achieves $R^2 = 0.968$ using only 6 biologically interpretable parameters. The raised cosine basis achieves $R^2 = 0.974$ but requires 12 parameters with no biological interpretation. The gamma-difference form captures essentially identical fit quality with half the parameters supporting its selection for phenotyping applications.

Recommended: Burst 10×0.5s

ON PERIOD (10s)   OFF PERIOD (20s)

ON

OFF

8× more events from repeated excitatory onset sampling.

```
Events: 14.9/track
RMSE: 0.036s
Power: 100%
Duty: 50%
```

Time from Cycle Start (seconds)
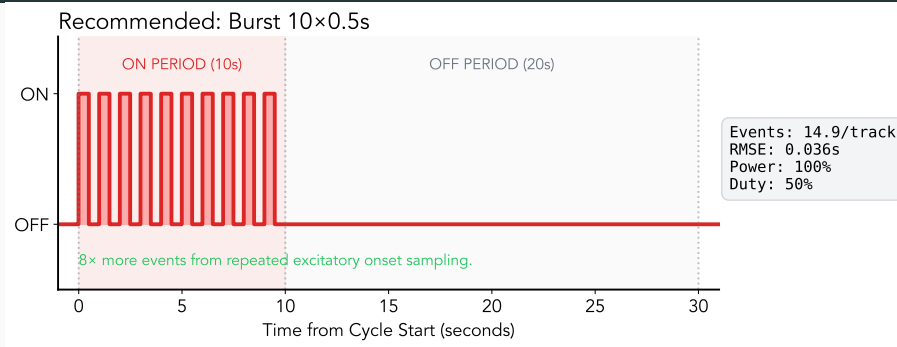
Replace continuous 10-second ON periods with burst trains of 10 pulses at 0.5 seconds each separated by 0.5 second gaps. Burst design samples the early excitatory window where τ_1 is identifiable. Each burst event carries 10 times more Fisher information than continuous events. This modification alone could reduce the number of events required for reliable estimation from 100 to 30.

Recommendation 2 – Extended Recording

Data Sparsity
(~25 events, need ~100)
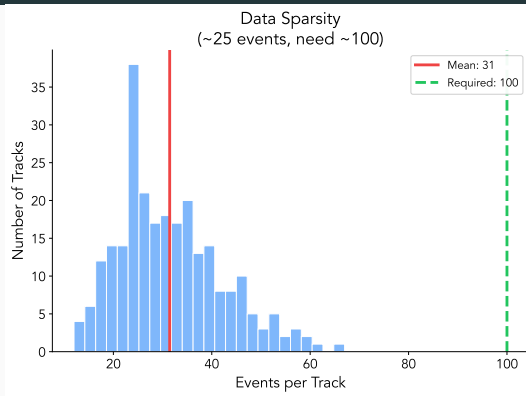
Target 40 minutes or more of recording to achieve at least 50 reorientation events per track. Current 10 to 20 minute recordings yield only 18 to 25 events which is insufficient for 6-parameter estimation. Power analysis indicates that 100 events are required for 80% power to detect a 0.2 second difference in τ_1 at the individual level. The recommended data-to-parameter ratio of 10 to 1 requires substantial increases in recording duration.
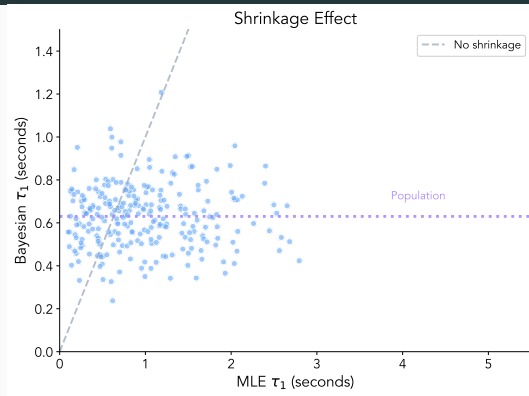
Reduce the parameter space by fixing population-derived parameters. Fix τ_2 at the population estimate of 3.8 seconds and fix the amplitude ratio B/A at the population value of 8. Estimate only the fast timescale τ_1 per individual track. This reduces the estimation problem from 6 dimensions to 1 dimension and dramatically improves identifiability with sparse data. Hierarchical Bayesian estimation provides natural regularization.

Same Events, Different Information
(~17 events each)
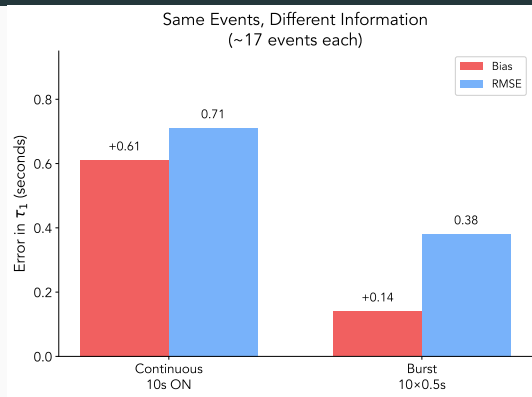
Use robust composite phenotypes that avoid kernel fitting entirely. The ON/OFF event ratio measures whether larvae respond preferentially during LED-ON versus LED-OFF periods and requires only event counts not full kernel estimation. First-event latency measures the time from LED onset to first reorientation and captures response speed directly. Both metrics are robust with sparse data and provide complementary phenotypic information.

Recommendation 5 – Within-Condition Analysis
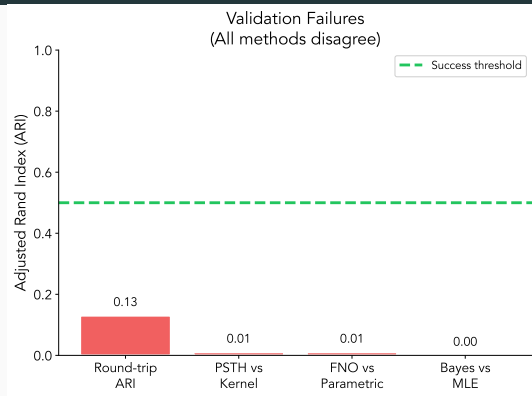
Analyze individual differences within experimental conditions rather than pooling across conditions. When data from different stimulation intensities and temporal patterns are pooled the condition effects dominate and mask genuine individual variation. The ARI near zero across all validation methods indicates no reproducible structure when pooling. Condition-specific analysis may reveal individual differences that are otherwise obscured.

Conclusions – Original Study

Population-Level Modeling Success

The gamma-difference kernel accurately models population-level reorientation dynamics across 14 experiments and 701 unique larval tracks.

Two timescales govern behavioral dynamics. Fast excitation with $\tau_1 \approx 0.3$ seconds captures the initial sensory response to optogenetic stimulation. Slow suppression with $\tau_2 \approx 4$ seconds captures habituation-like adaptation that accumulates across repeated stimuli.

The model is robust across experimental conditions as demonstrated by leave-one-experiment-out cross-validation and bootstrap confidence intervals.

The gamma-difference form provides biological interpretability that flexible basis function models lack while achieving equivalent goodness of fit.

Conclusions – Follow-Up Study

Individual Phenotyping Challenges

Individual phenotyping using kernel parameters fails with current experimental protocols due to fundamentally sparse data.

Apparent clusters identified by K-means or hierarchical clustering are statistical artifacts of fitting high-dimensional models to low-event tracks rather than genuine discrete phenotypes.

Only 8.6% of tracks show individual variation that exceeds the measurement noise floor.

Current protocols achieve only 20 to 30% statistical power for detecting fast responder phenotypes.

Thank You

Questions?

FAQ – Original Study Methods

What is the sequence of processes in the original study?

Data collection   14 optogenetic experiments with 701 larval tracks under four stimulation conditions varying intensity from 0-250 PWM versus 50-250 PWM and temporal pattern from constant versus cycling.

Trajectory extraction   MAGAT Analyzer extracts larval trajectories from video and segments behavioral states including runs and reorientations.

Event detection   Reorientation onset times are extracted from MAGAT segmentation as point events for hazard modeling.

Kernel fitting   Population-level gamma-difference kernel is fitted using maximum likelihood estimation on the pooled PSTH.

Validation   Leave-one-experiment-out cross-validation and bootstrap confidence

FAQ – Follow-Up Study Methods

What processes were used in the follow-up study?

Individual fitting   Maximum likelihood estimation of 6-parameter gamma-difference kernel per track using the same event data.

Clustering   K-means and hierarchical clustering applied to individual kernel parameters to identify putative phenotypic groups.

Validation   Round-trip validation simulates from fitted kernels and re-clusters to test reproducibility. Gap statistic tests whether k clusters improve over k=1.

Power analysis   Simulation-based analysis estimates statistical power for phenotype detection as a function of event count and protocol design.

Identifiability analysis   Fisher Information quantifies how much information each event contains about kernel parameters under different stimulation designs.

FAQ – Why does population modeling succeed but individual fails?

The key is the data-to-parameter ratio

Population modeling pools approximately 15,000 reorientation events across 701 tracks to estimate 6 kernel parameters. The data-to-parameter ratio exceeds 2,500 to 1 providing strong statistical power.

Individual modeling attempts to estimate the same 6 parameters from only 18 to 25 events per track. The data-to-parameter ratio is only 3 to 5 which is far below the minimum of 10 to 1 needed for reliable maximum likelihood estimation.

The gamma-difference kernel is overparameterized for individual-level sparse data. Identifiability analysis reveals that only τ_1 and possibly the amplitude ratio can be reliably estimated with fewer than 50 events.

Hierarchical Bayesian estimation provides partial mitigation by shrinking individual

FAQ – What is hierarchical shrinkage?

Regularization through shared population structure

Hierarchical Bayesian models assume individual parameters are drawn from a population distribution with unknown mean and variance.

Tracks with sparse data are shrunk toward the population mean because their individual estimates are uncertain. Tracks with abundant data retain individual estimates because they contain sufficient information.

The amount of shrinkage is determined automatically by the posterior distribution. More sparse data produces more shrinkage.

Shrinkage is not a bias correction. It is optimal regularization under the assumption that individuals are exchangeable members of a population.

FAQ – How should clustering results be interpreted?

Clusters in sparse high-dimensional data are usually artifacts

K-means and hierarchical clustering will always produce k clusters regardless of whether true clusters exist in the data.

The gap statistic compares within-cluster dispersion to a null reference and finds that k=1 is optimal meaning no discrete clusters are present.

Round-trip validation tests whether clusters are reproducible by simulating new data from fitted parameters and re-clustering. Agreement below ARI $= 0.2$ indicates clusters are not reproducible.

The continuous unimodal PCA distribution confirms that kernel parameter variation is smooth rather than discrete.