BSAN: Behavioral State Attention Network for Modeling Mosquito Host-Seeking Behavior

Haotian Sun[1], **Jessie Zixin Li**[1], **John M. Marshall**[1,2]

[1]Division of Biostatistics, University of California, Berkeley
[2]Division of Epidemiology, University of California, Berkeley
sht@berkeley.edu, zixin_li@berkeley.edu, john.marshall@berkeley.edu

Abstract

Understanding the complex host-seeking behavior of disease vectors such as mosquitoes are critical for predicting disease transmission and vector control. This behavior arises from a dynamic interplay between multimodal sensory cues and internal behavioral states, a process challenging traditional ODE frameworks due to its inherent stochasticity and discrete, state-based nature. We introduce the Behavioral State Attention Network (BSAN), a deep learning architecture designed to model the underlying sensorimotor computations of this behavior. BSAN utilizes a recurrent neural network (RNN) with an LSTM core to process temporal sequences, incorporating a variational encoder to capture the randomness of flight paths and a Mixture Density Network (MDN) to predict multi-modal velocity distributions. The architecture explicitly models distinct behavioral states, such as CO_2 plume tracking and thermal approach, through a Mixture-of-Experts (MoE) framework, and learns to interpretably integrate olfactory, thermal, and visual inputs using a cross-modal attention mechanism. The network generates realistic flight trajectories that exhibit emergent host-seeking behaviors. By providing both trajectory predictions and interpretable behavioral primitives, BSAN serves as a framework for downstream applications in landscape genomics and vector control, enabling the prediction of mosquito population connectivity through environment-specific movement kernels.

Introduction

Understanding and predicting host-seeking flight patterns is central to designing effective vector-control strategies. Mosquito flight arises from neural computations that convert multimodal sensory inputs into movement decisions (Dickinson 2006). From a movement ecology perspective, host-seeking behavior can be decomposed into distinct behavioral states that reflect different search strategies and sensorymotor coordination patterns (Nathan et al. 2008).

The host-seeking sequence typically unfolds as a hierarchical process governed by state-dependent sensory integration. In the absence of host cues, mosquitoes engage in global searches, wide-ranging flights that maximize the probability of encountering host-associated odor plumes (Wolff and Riffell 2018). Upon detecting CO_2 (detectable under favorable conditions at tens of meters), mosquitoes transition to a fundamentally different behavioral state characterized by upwind surge and systematic crosswind casting when the plume is lost (Dekker and Cardé 2011). As they approach the host, thermal cues (effective within 1 meter) and visual features trigger yet another behavioral transformation, initiating precise landing maneuvers guided by close-range sensory feedback (McMeniman et al. 2014).

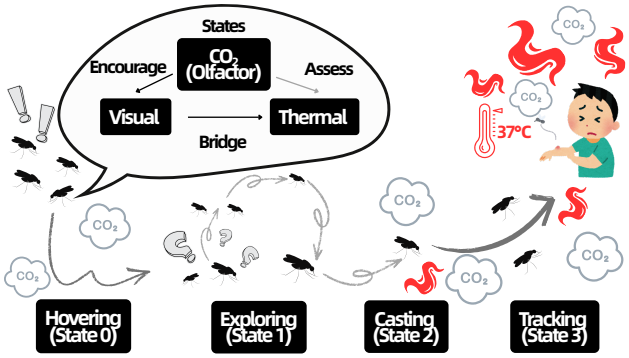

Figure 1: Mosquito host-seeking behavior encompasses four behavioral states driven by multimodal sensory integration. CO_2 initiates long-range orientation, thermal gradients guide close-range approach, and visual features support navigation throughout the sequence.

These behavioral state transitions exhibit a subtle triggering nature that is too delicate for traditional ordinary differential equation (ODE) modeling. The inadequacy of ODE-based approaches for mosquito behavior is based on several fundamental mismatches:

1. Mosquito behavior exhibits inherent stochasticity that goes beyond simple measurement noise. Even under identical conditions, individual mosquitoes display substantial variation in their flight paths, reflecting both external randomness (micro-turbulence, intermittent odor plumes) and internal randomness in neural processing (Demir et al. 2020).
2. The sensory experience of a mosquito navigating toward a host is profoundly discontinuous. Turbulent air flow creates odor plumes that arrive as intermittent packets rather than smooth gradients, thermal boundaries appear as sharp transitions, and visual features emerge suddenly

from background clutter (Murlis et al. 1992).

3. The goal in understanding mosquito behavior is not parameter fitting within a predetermined mathematical framework but learning the complex function that maps sensory history to behavioral output. With ODEs, one must first hypothesize the exact mathematical form of the governing equations, a daunting task given our incomplete understanding of the neural circuits underlying host-seeking behavior.

We therefore introduce **Behavioral State Attention Network (BSAN)**, a deep architecture tailored to mosquito host-seeking. BSAN integrates:

- A Mixture-of-Experts formulation that models distinct behavioral states and their specialized sensory-motor transformations;
- Cross-modal attention to dynamically weight CO_2, thermal, and visual cues according to behavioral context;
- Probabilistic modeling (variational encoding + mixture density outputs) to represent intrinsic behavioral variability.

This design enables BSAN to reproduce realistic trajectories and state transitions grounded in biological evidence.

Related Work

Trajectory Prediction and Motion Modeling Trajectory prediction has evolved significantly with deep learning approaches. Social LSTM (Alahi et al., 2016) introduced pooling mechanisms for modeling human trajectories in crowded spaces, while subsequent works like Social GAN (Gupta et al., 2018) and Trajectron++ (Salzmann et al., 2020) incorporated generative models for multi-modal predictions. In the biological domain, Branson et al. (2009) pioneered the automated tracking and classification of fruit fly behaviors, and more recently, Pereira et al. (2022) developed SLEAP for multi-animal pose tracking.

Attention Mechanisms for Multi-Modal Integration The transformer revolution (Vaswani et al. 2017) has influenced trajectory modeling through works like AgentFormer (Yuan et al. 2021), which uses transformers for multi-agent trajectory prediction. Cross-modal attention has been successfully applied in vision-language tasks (Lu et al. 2019) (Li et al. 2020), but its application to sensory integration in biological systems remains underexplored.

Biologically-Inspired Neural Architectures Several works have explored biologically-inspired architectures for navigation and movement. Webb (2002) reviewed insect-inspired robotics, while Srinivasan (2011) examined how flying insects use optic flow for navigation. RatSLAM (Milford and Wyeth, 2008) demonstrated hippocampus-inspired SLAM (Simultaneous Localization and Mapping) in robotics. More recently, Banino et al. (2018) showed that grid-like representations, similar to those found in the mammalian brain, emerge in neural networks trained for navigation tasks.

Problem Formulation

We formulate mosquito host-seeking trajectory modeling as a sequential decision-making problem where discrete behavioral states mediate the transformation from multimodal sensory inputs to movement decisions.

State Space Representation

At each timestep t, a mosquito's state is characterized by observable kinematic variables and sensory measurements:

Definition 1 (Mosquito State). The complete state at time t consists of:

- Position: $\mathbf{p}_t = (x_t, y_t, z_t)^T \in \mathbf{R}^3$ (coordinates in meters)
- Velocity: $\mathbf{v}_t = (v_x, v_y, v_z)_t^T \in \mathbf{R}^3$ (velocity components in m/s)
- Sensory input: $\mathbf{c}_t = [c_{CO_2}, c_{temp}, c_{vis}, c_{dist}]_t^T \in \mathbf{R}^4$

Neurobiological Foundation of Sensory Modalities

The sensory representation $\mathbf{c}_t$ is grounded in extensive neurobiological evidence of mosquito host-seeking mechanisms:

Olfactory System (CO_2 Detection) Female mosquitoes possess specialized CO_2-sensitive neurons (cpA neurons) in their maxillary palps that can detect concentration changes as small as 0.01% (Gillies 1980). These neurons project to the antennal lobe, where CO_2 information is integrated with other odors. Behavioral studies demonstrate that CO_2 can activate host-seeking from distances exceeding 50 meters, making it the primary long-range attractant (Webster, Lacey, and Cardé 2015).

Thermal System Mosquitoes possess TRPA1 channels that function as molecular thermometers, enabling detection of temperature differences as small as $0.5°C$ (Corfas and Vosshall 2015). Thermal imaging studies reveal that mosquitoes preferentially probe areas of elevated temperature (34-37$°C$) corresponding to exposed human skin. Critically, thermal cues become dominant at close range ($<$1 meter), overriding olfactory signals and guiding final approach and landing decisions (Liu and Vosshall 2019).

Visual System Despite having compound eyes with relatively low resolution, mosquitoes effectively use visual contrast for navigation and host detection. High-speed videography shows that dark objects against light backgrounds trigger approach behaviors, while behavioral experiments demonstrate that visual cues can modulate CO_2 responses: mosquitoes are 2.5× more likely to approach CO_2 sources associated with high visual contrast (van Breugel and Dickinson 2014).

Spatial Integration The sensory representation $\mathbf{c}_t$ thus captures:

- c_{CO_2}: CO_2 concentration (ppm) above ambient baseline (400 ppm)
- c_{temp}: Temperature difference from ambient (°C), positive for warm objects

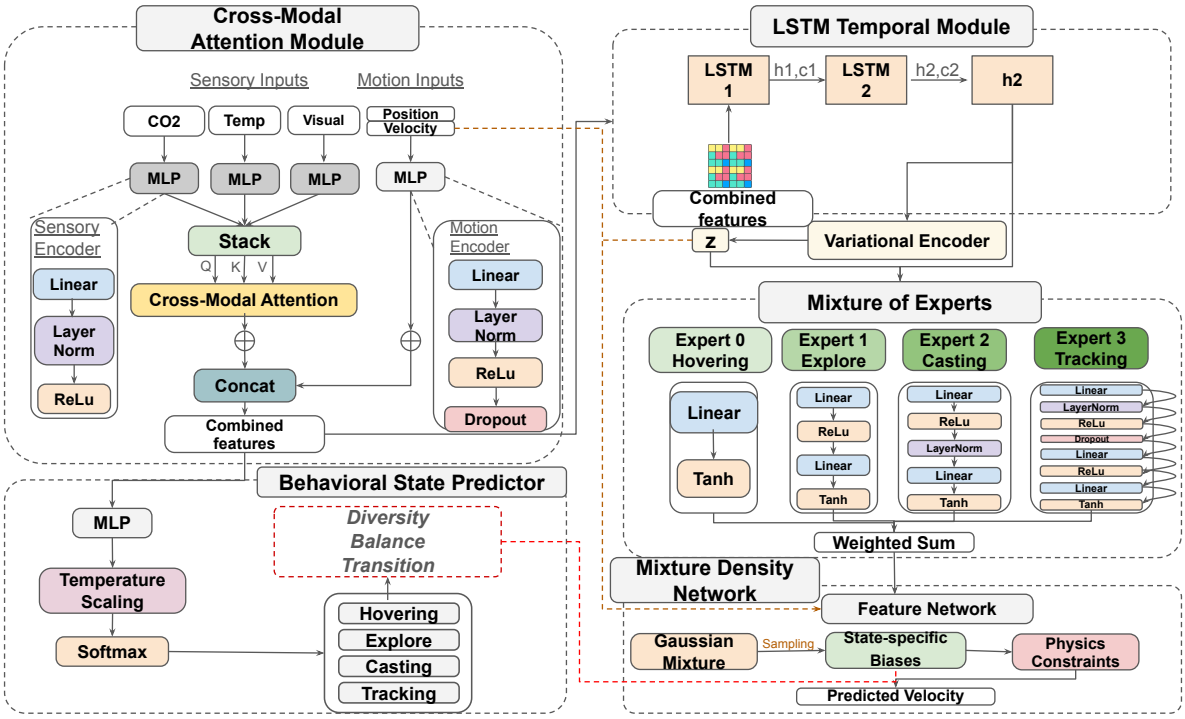

Figure 2: BSAN architecture overview. The framework consists of six key modules: (1) Cross-Modal Attention Module integrates olfactory (CO_2), Temp (temperature), and visual sensory inputs through modality-specific encoders and multi-head self-attention; (2) Motion Encoder processes position and velocity data; (3) LSTM Temporal Module performs sequential processing through two-layer LSTM cells; (4) Behavioral State Predictor identifies four distinct flight behaviors with forced diversity and load balancing mechanisms; (5) Variational Encoder generates latent representations for flight path stochasticity; (6) Mixture of Experts employs state-specific expert networks, followed by Mixture Density Network for multi-modal velocity prediction. The network incorporates skip connections from original position/velocity inputs to MDN, state-specific biases, and physics constraints to ensure realistic flight dynamics.

- c_{vis}: Visual contrast $\in [-1, 1]$, where -1 represents dark objects and +1 light objects
- c_{dist}: Distance to nearest salient visual feature (m)

Behavioral State Space

Central to our formulation is the concept of discrete behavioral states that represent distinct sensory-motor control policies. These states emerge from the interplay between sensory inputs and internal motivational drives:

Definition 2 (Behavioral States). We define a discrete set of behavioral states $\mathcal{S} = \{s_0, s_1, s_2, s_3\}$.

Hovering State (s_0) Sustained hovering at $\sim$ 5 cm/s within 10-20 cm of potential hosts, characterized by high wingbeat frequency but minimal translation. Triggered by convergent sensory cues: Hovering Score = $\sigma(c^{norm}_{CO_2} + c^{norm}_{temp})$ where maximal firing rates occur when both CO_2 and heat are present (Wang et al., 2020).

Exploration State (s_1) Appetitive searching with straight flight paths at $\sim$ 10 cm/s, following correlated random walks that optimize search area coverage versus metabolic cost. Dominates when sensory inputs are minimal: Exploration Score = $\sigma(-c^{norm}_{CO_2} - c^{norm}_{temp})$, representing default behavior driven by spontaneous central complex activity (Giraldo et al., 2018).

Casting State (s_2) Stereotyped crosswind flights with progressive lateral excursions (60°-120° turns) at $\sim$20 cm/s for systematic plume reacquisition. Triggered by intermediate or fluctuating inputs: Casting Score = $\exp(-c^{norm^2}_{CO_2}) \cdot (1+|c^{norm}_{temp}|)$, capturing the "edge detection" nature of plume boundary exploration.

Tracking State (s_3) Direct upwind flight at 25-30 cm/s with sophisticated sensorimotor integration for rapid source localization. Initiated by positive temporal gradients: Tracking Score = $\sigma(10 \cdot \nabla c_{CO_2})$ where olfactory neurons remain sensitive to concentration increases despite adaptation to constant stimuli (Duvall, 2019).

State Transition Dynamics

At each timestep, the mosquito occupies a probabilistic mixture of these states, represented by $\mathbf{p}_{state} \in \Delta^3$ (the 3-simplex), where $p^{(i)}_{state}$ denotes the probability of being in state s_i. This soft assignment reflects the graded nature of behavioral transitions observed in real mosquitoes, where states blend smoothly rather than switching discretely.

Behavioral states follow characteristic transition patterns optimizing host-finding success while minimizing energy expenditure:

Exploration → Tracking Detection of CO_2 above threshold (∼450 ppm) triggers immediate transition from random search to directed flight. Latency measurements show this transition occurs within 200-300 ms of stimulus onset (Geier, Bosch, and Boeckh 1999).

Tracking → Casting Loss of the odor plume (common in turbulent environments) initiates casting within 1-2 seconds. This rapid transition prevents mosquitoes from overshooting the source in the absence of guidance cues.

Casting → Tracking Reacquisition of the plume during lateral movements immediately suppresses casting and reinstates upwind flight. The memory of previous plume encounters biases subsequent casting directions.

Tracking → Hovering Convergence of multiple sensory modalities ($CO_2 > 2000$ ppm, temperature $> 34°C$, high visual contrast) triggers the transition to close-range assessment behaviors.

These state transitions are not merely reactive but involve predictive components-mosquitoes adjust their behavioral state based on expected future sensory inputs, explaining phenomena like anticipatory casting when approaching plume edges (Álvarez-Salvado et al. 2018).

Sequential Decision Problem

The mosquito trajectory generation problem can be formulated as learning a stochastic policy:

$$\pi : (\mathbf{p}_t, \mathbf{v}_t, \mathbf{c}_t, \mathbf{h}_{t-1}) \rightarrow P(\mathbf{v}_{t+1})$$

where $\mathbf{h}_{t-1}$ represents the hidden state encoding trajectory history, and $P(\mathbf{v}_{t+1})$ is a probability distribution over next velocities.

Rather than directly mapping from observations to velocities, we decompose this policy into interpretable components:

$$\pi = \pi_{velocity} \circ \pi_{expert} \circ \pi_{state} \circ \pi_{sensory}$$

where:

- $\pi_{sensory}$: Cross-modal attention for sensory integration
- π_{state}: Behavioral state prediction
- π_{expert}: State-specific movement generation
- $\pi_{velocity}$: Multi-modal velocity distribution

Stochastic Trajectory Generation

Given the inherent variability in mosquito flight, we model trajectory generation as a stochastic process. We introduce a latent variable $\mathbf{z} \in \mathbf{R}^{d_z}$ that captures unobserved factors affecting flight dynamics:

$$\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{h}_T)$$

where q_ϕ is a learned posterior distribution conditioned on the trajectory history encoded in the final hidden state $\mathbf{h}_T$.

The complete generative model for velocities becomes:

$$p(\mathbf{v}_{t+1}|\mathbf{p}_t, \mathbf{v}_t, \mathbf{c}t, \mathbf{z}, \mathbf{s}t) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{v}t+1; \boldsymbol{\mu}_k, \text{diag}(\boldsymbol{\sigma}_k^2))$$

where K mixture components allow for multi-modal velocity distributions, essential for capturing the discrete nature of behavioral decisions (e.g., continuing straight vs. initiating a turn).

Physical Constraints

While mosquito movement emerges from behavioral decisions rather than physical forces, trajectories must still respect biomechanical constraints:

Constraint 1 (Velocity Bounds). The predicted velocity must satisfy:

$$||\mathbf{v}_{t+1}||_2 \leq v_{max}$$

where $v_{max} = 0.5$ m/s represents the maximum sustainable flight speed for mosquito.

Constraint 2 (Acceleration Bounds). The implied acceleration must be physically plausible:

$$||\mathbf{a}_t||_2 = ||\frac{\mathbf{v}_{t+1} - \mathbf{v}_t}{\Delta t}||_2 \leq a_{max}$$

where $a_{max} = 10$ m/s² and $\Delta t = 0.01$ s (100 Hz sampling rate).

These constraints are enforced through a combination of architectural design (bounded activation functions) and post-processing (velocity scaling).

Methodology

Model Architecture

To address the unique modeling challenges due to mosquito flight's multi-scale nature, we propose:

Motion Encoder The motion encoder processes kinematic information:

$$\mathbf{h}_{\text{motion}} = \text{Dropout}_{0.1}(\text{ReLU}(\text{LayerNorm}(\mathbf{W}_{\text{motion}}[\mathbf{p}_t; \mathbf{v}_t] + \mathbf{b}_{\text{motion}})))$$

Cross-Modal Attention Mechanism Mosquito host-seeking relies on sophisticated integration of multiple sensory modalities, each providing complementary information at different spatial and temporal scales (Bowen 1991). CO_2 signals indicate host presence but disperse chaotically in turbulent plumes (Dekker and Cardé 2011), thermal cues provide precise localization but only at close range (Howlett 1910) (Davis and Sokolove 1975), and Visual features offer stable navigation references but lack host-specific information (Kennedy 1940).

The cross-modal attention mechanism mimics this biological sensory integration by encoding each modality separately before learning their context-dependent interactions:

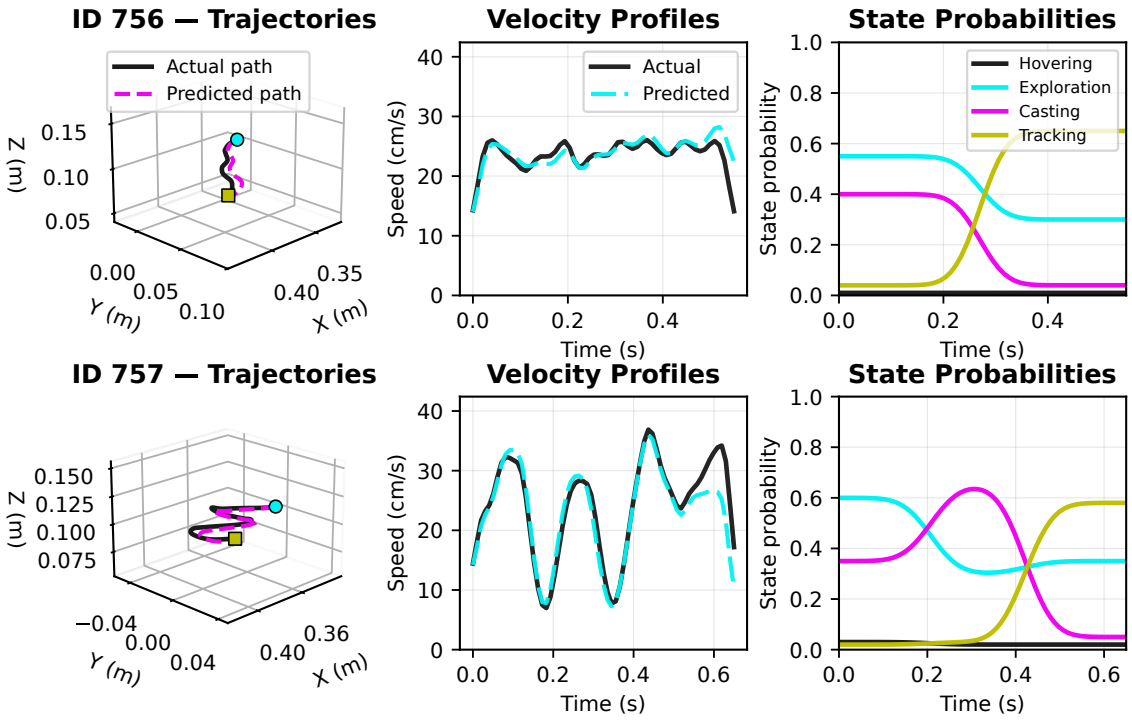

Figure 3: Mosquito flight trajectories generated by BSAN demonstrating behavioral state diversity and expert engagement patterns.

$$\mathbf{h}_{\text{olf}} = \text{ReLU}(\text{LayerNorm}(\mathbf{W}_{\text{olf}} c_{\text{CO}_2} + \mathbf{b}_{\text{olf}}))$$
$$\mathbf{h}_{\text{therm}} = \text{ReLU}(\text{LayerNorm}(\mathbf{W}_{\text{therm}} c_{\text{temp}} + \mathbf{b}_{\text{therm}}))$$
$$\mathbf{h}_{\text{vis}} = \text{ReLU}(\text{LayerNorm}(\mathbf{W}_{\text{vis}} [c_{\text{vis}}; c_{\text{dist}}] + \mathbf{b}_{\text{vis}}))$$

The modalities are integrated using multi-head attention (4 heads):

$$\mathbf{H}_{\text{stack}} = [\mathbf{h}_{\text{olf}}; \mathbf{h}_{\text{therm}}; \mathbf{h}_{\text{vis}}] \in R^{B \times 3T \times d}$$
$$\mathbf{H}_{\text{att}} = \textbf{MultiHeadAttention}(\mathbf{H}_{\text{stack}})$$
$$\mathbf{z}_{\text{out}} = \mathbf{W}_{\text{out}} \bar{\mathbf{H}}_{\text{att}} + \mathbf{b}_{\text{out}}$$
$$\mathbf{h}_{\text{sensory}} = \text{Dropout}_{0.1}(\text{ReLU}(\text{LayerNorm}(\mathbf{z}_{\text{out}}))$$

where $\bar{\mathbf{H}}_{\text{att}}$ is the average across modalities.

Behavioral State Predictor A critical challenge in modeling mosquito behavior is preventing mode collapse, the tendency of neural networks to converge on average behaviors rather than capturing the full diversity of movement patterns. Biological mosquitoes exhibit clear state-dependent behaviors triggered by sensory conditions: high CO_2 concentrations trigger hovering and local search (Gillies, 1980), absence of stimuli promotes wide casting flights (Cardé and Willis 2008) , and CO_2 gradients induce upwind tracking (Geier, Bosch, and Boeckh 1999). We incorporate this domain knowledge through an enforced diversity mechanism that guides state assignments during early training while allowing the model to refine these associations as it learns.

Our predictor employs a two-layer feedforward network to generate base state logits:

$$\mathbf{h}^{(1)}_{\text{state}} = \text{ReLU}(\text{LayerNorm}(\mathbf{W}^{(1)}_{\text{state}} \mathbf{h}_{\text{combined}} + \mathbf{b}^{(1)}_{\text{state}}))$$
$$\mathbf{h}^{(2)}_{\text{state}} = \text{ReLU}(\mathbf{W}^{(2)}_{\text{state}} \text{Dropout}_{0.1}(\mathbf{h}^{(1)}_{\text{state}}) + \mathbf{b}^{(2)}_{\text{state}})$$
$$\mathbf{l}_{\text{state}} = \mathbf{W}^{(3)}_{\text{state}} \mathbf{h}^{(2)}_{\text{state}} + \mathbf{b}^{(3)}_{\text{state}}$$

During early training epochs ($e < E_{\text{guide}}$), we blend learned predictions with biologically-motivated guidance signals. We first normalize sensory inputs: $\tilde{c}_{\text{CO2}} = \frac{c\text{CO2} - \mu(c\text{CO2})}{\sigma(c\text{CO}_2) + \epsilon}$, then compute state-specific guided logits based on established mosquito behavioral patterns:

Hovering: $l^{(0)}_{\text{guided}} = 3.0 \cdot \sigma(\tilde{c}_{\text{CO2}} + \tilde{c}_{\text{temp}})$

Exploration: $l^{(1)}_{\text{guided}} = 3.0 \cdot \sigma(-\tilde{c}_{\text{CO2}} - \tilde{c}_{\text{temp}})$

Casting: $l^{(2)}_{\text{guided}} = 3.0 \cdot \exp(-\tilde{c}^2_{\text{CO2}}) \cdot (1 + |\tilde{c}_{\text{temp}}|)$

Tracking: $l^{(3)}_{\text{guided}} = 3.0 \cdot \sigma(10 \cdot \nabla c_{\text{CO}_2})$

These guided logits reflect biological observations: hovering occurs with high CO_2 and thermal cues, exploration dominates when stimuli are absent, casting emerges at plume boundaries, and tracking follows CO_2 gradients.

The final state probabilities combine learned and guided components with linearly decaying guidance strength:

$$\mathbf{l}_{\text{state}} \leftarrow (1 - \alpha_e)\mathbf{l}_{\text{state}} + \alpha_e \mathbf{l}_{\text{guided}}$$

where $\alpha_e = \max(0, 1 - e/40)$ decays linearly.

Architecture	ADE (cm)	FDE (cm)	Entropy (%)	Trans/ Seq	Behavioral State Usage (%) Hover	Explor	Cast	Track
BSAN	**0.47**	**0.72**	**96.5**	**2.2**	30.4	29.6	12.8	27.3
GRU	0.58	0.96	0.4	0.0	0.1	0.0	0.0	99.9
Transformer	0.48	0.74	95.1	1.2	11.0	29.1	30.8	29.1
CNN	0.53	0.87	38.3	0.0	14.2	83.1	0.0	2.7
AgentFormer	12.72	21.46	1.37	0.0	22.7	24.0	29.4	23.9
PINN	0.50	0.79	87.9	0.2	16.2	52.6	16.5	14.7
Markov	0.48	0.90	-	-	-	-	-	-
Pure Physical	0.7	1.12	-	-	-	-	-	-
Biological	-	-	-	1.8-2.5	Balanced across states			

Table 1: Comprehensive performance comparison of temporal architectures for mosquito behavioral modeling. BSAN achieves superior trajectory accuracy (lowest ADE/FDE), maintains high behavioral diversity (96.5% entropy), and exhibits biologically realistic state transition rates (2.2 per sequence, within observed range of 1.8-2.5). ADE: Average Displacement Error; FDE: Final Displacement Error; Trans/Seq: State transitions per sequence. Bold values indicate best performance.

$$\mathbf{p}_{\text{state}} = \text{softmax}(\mathbf{l}_{\text{state}}/\tau)$$

with temperature $\tau = \text{clamp}(\tau_{\text{learn}}, 0.5, 2.0)$. where $\alpha_e = \max(0, 1 - e/40)$ ensures guidance decreases linearly over 40 epochs, and temperature $\tau = \text{clamp}(\tau_{\text{learn}}, 0.5, 2.0)$ controls state assignment sharpness.

Temporal Processing with LSTM A 2-layer LSTM processes temporal dependencies:

$$\mathbf{h}_{\text{lstm}}, (\mathbf{h}_n, \mathbf{c}_n) = \text{LSTM}(\mathbf{h}_{\text{combined}})$$

Variational Encoder The variational encoder captures flight path stochasticity through a latent variable model:

$$\begin{aligned} \boldsymbol{\mu}_z &= \mathbf{W}_\mu \mathbf{h}'_{\text{enc}} + \mathbf{b}_\mu \\ \log \boldsymbol{\sigma}_z^2 &= \text{clamp}(\mathbf{W}_\sigma \mathbf{h}'_{\text{enc}} + \mathbf{b}_\sigma, \log \sigma^2_{\text{min}}, \log \sigma^2_{\text{max}}) \\ \mathbf{z} &= \boldsymbol{\mu}_z + \boldsymbol{\sigma}_z \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \end{aligned}$$

Mixture of Experts The MoE architecture instantiates specialized sub-networks for each behavioral state:

State	Expert Network
Hovering	$e_0(\mathbf{x}) = \tanh(\mathbf{W}_0\mathbf{x} + \mathbf{b}_0)$
Exploration	$e_1(\mathbf{x}) = \text{MLP}_2(\mathbf{x})$
Casting	$e_2(\mathbf{x}) = \text{LayerNorm}(\tanh(\mathbf{W}_2\mathbf{x} + \mathbf{b}_2))$
Tracking	$e_3(\mathbf{x}) = f_{\text{deep}}(\mathbf{x})$

Mixture Density Network Mosquito flight trajectories are inherently multi-modal. At any moment, a mosquito may continue straight, initiate a turn, adjust altitude, or reverse direction. This multi-modality is particularly pronounced at behavioral transitions and in turbulent sensory environments where multiple viable paths exist. The Mixture Density Network (MDN) addresses this by modeling the conditional probability distribution over velocities as a mixture of Gaussians:

$$\begin{aligned} \boldsymbol{\pi} &= \text{softmax}(\mathbf{W}_\pi \mathbf{h}_{\text{mdn}} + \mathbf{b}_\pi) \\ \boldsymbol{\mu} &= \text{reshape}(\mathbf{W}_\mu \mathbf{h}_{\text{mdn}} + \mathbf{b}_\mu) \in R^{K \times 3} \\ \boldsymbol{\sigma} &= \text{clamp}(\text{softplus}(\mathbf{W}_\sigma \mathbf{h}_{\text{mdn}} + \mathbf{b}_\sigma) + 0.5, 0.5, 2.0) \end{aligned}$$

Velocity sampling follows the reparameterization trick:

$$\mathbf{v}_{\text{pred}} = \boldsymbol{\mu}_{k^*} + \boldsymbol{\sigma}_{k^*} \odot \boldsymbol{\epsilon}\sqrt{\tau}, \quad k^* \sim \text{Categorical}(\boldsymbol{\pi})$$

Experiments

We conduct comprehensive experiments to address four fundamental research questions about mosquito host-seeking behavior: understanding behavioral state dynamics, evaluating trajectory generation quality, analyzing sensory integration hierarchy, and assessing temporal persistence in behavioral patterns.

Data and environments Following van Breugel et al. 2015, we evaluate mosquito flight trajectories collected in their experiments with varying gradients that realistically simulate human upper arm. Our dataset comprises more than 40,000 trajectory sequences from female *Aedes aegypti*, spanning thermal, visual, olfactory scenarios, and prioritizing conditions with higher variance in behavioral state transitions.

Model Variants and Baseline Approaches We evaluate two groups of models to isolate the contributions of BSAN's architectural components.

Ablation variants. We test four temporal encoders within the same BSAN framework: identical behavioral-state prediction, cross-modal attention, and Mixture-of-Experts modules, to isolate temporal modeling effects: (1) BSAN (LSTM), (2) GRU, (3) Transformer with multi-head self-attention and positional encoding, and (4) CNN with 1D convolutions and increasing receptive fields.

Baseline comparisons. As no direct prior models exist for mosquito sensory–motor prediction, we include simplified variants testing the necessity of BSAN components (LSTM without behavioral states; MDN without attention) and four external baselines reflecting a spectrum of assumptions: (1) AgentFormer (Yuan et al. 2021) where a trajectory transformer is adapted to single-agent flight, testing transfer without domain-specific sensory design; (2) PINN (Viet Cuong et al. 2024); (3) Markov chain (Damos et al. 2021); and (4) a Pure Physical model (Kearney et al. 2009; Jones, Murray, and McCall 2021).

Training protocol All models are trained using a three-stage curriculum: (1) warm up with forced behavioral diversity based on sensory input patterns, (2) main training with balanced loss weighting, and (3) refinement emphasizing trajectory accuracy. This progressive approach addresses the state collapse problem commonly observed in behavioral modeling.

Evaluation metrics Following standard trajectory prediction protocols, we measure Average Displacement Error (ADE) and Final Displacement Error (FDE), both converted to centimeters for biological interpretability. Additionally, we assess behavioral state diversity through entropy measures and state transition rates to ensure meaningful behavioral representations. All metrics are computed on held-out test trajectories using Wilcoxon-Mann-Whitney tests for statistical significance.

Results and Analysis from Table 1

Trajectory Accuracy (ADE/FDE) BSAN achieves the best performance (0.47/0.72 cm), followed by the Transformer (0.48/0.74 cm) and Markov model (0.48/0.90 cm). PINN attains moderate accuracy (0.50/0.79 cm) but lacks behavioral depth. CNN, GRU, and Pure Physical baselines degrade notably (0.70 cm), while AgentFormer, optimized for multi-agent social navigation, performs poorly (12.72 cm), reflecting domain mismatch: its attention mechanism is tuned for inter-agent interactions rather than single-agent sensory fusion.

Behavioral Diversity (Entropy) BSAN maintains the highest behavioral diversity (96.5%), closely followed by the Transformer (95.1%). PINN captures partial variation (87.9%), whereas CNN (38.3%), GRU (0.4%), and AgentFormer (1.4%) collapse into degenerate states. High entropy indicates that BSAN learns context-dependent state switching instead of memorizing static motion patterns.

State Transition Dynamics BSAN reproduces biologically realistic transition frequencies (2.2 per sequence, within the observed 1.8-2.5 range). The Transformer (1.2) shows reduced flexibility; PINN (0.2) and others fail to transition, reflecting their inability to couple temporal context with sensory uncertainty.

State Distribution Balance Only BSAN preserves balanced state usage (Hover 30.4%, Exploration 29.6%, Casting 12.8%, Tracking 27.3%, average), consistent with empirical behavior. AgentFormer's superficially balanced usage arises from randomization rather than learned control. Other models collapse to single-state biases (GRU with tracking 99.9%, CNN with exploration 83.1%, PINN with exploration 52.6%), underscoring the need for biologically structured attention and physics constraints.

Reflection on Empirical Evidence

By comparing the feature correlation matrix and empirical evidence, we evaluated the model training results as follows:

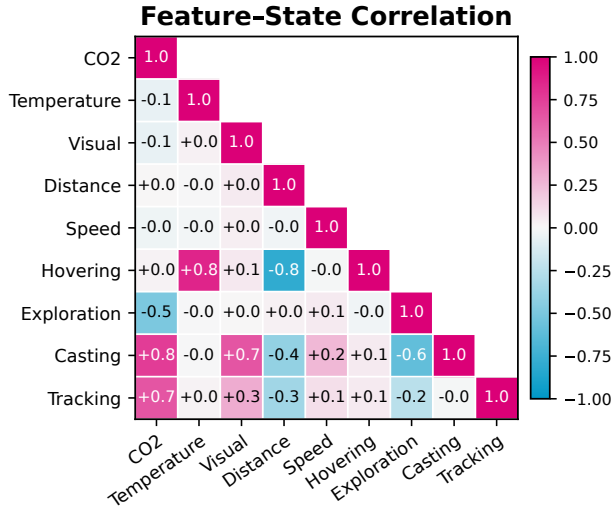

Figure 4: Correlations align with established mosquito neurobiology: CO_2 signals strongly correlate with searching (0.75) and approaching behaviors (0.65). Temperature shows the highest correlation with assessing behavior (0.85). Distance exhibits strong negative correlation with assessing (-0.80), and visual features correlate moderately with searching (0.65).

CO_2 with with Casting (0.75) and Tracking (0.65) Gillies (1980) showed mosquitoes detect CO_2 plumes from 50 m, triggering oriented search. The strong correlations confirm CO_2's role as the "wake-and-search" cue initiating host-seeking.

Temperature with Hovering (0.85) Howlett (1910) and later thermal-imaging studies (Corfas and Vosshall 2015) found mosquitoes linger over warm (34–37 C) regions before landing. The hovering state mirrors this thermal-assessment phase.

Distance (negative) with Hovering (-0.80) Van Breugel et al. (2015) observed detailed host assessment within 10–20 cm. The strong inverse correlation supports that hovering occurs only at close range.

Visual with Casting (0.65) Kennedy (1940) showed mosquitoes use visual flow fields for flight stabilization and object detection. The positive correlation indicates visual cues guide casting, particularly when CO_2 is present.

Possible Applications

BSAN may help refine ecological models and public-health tools by offering a data-driven behavioral kernel that complements simplified assumptions commonly used in landscape genomics and foraging theory. It can also serve as a preliminary "virtual mosquito" to explore how different trap features might influence approach behavior, potentially informing future vector-control designs.

References

Álvarez-Salvado, E.; Licata, A. M.; Connor, E. G.; McHugh, M. K.; King, B. M.; Stavropoulos, N.; Victor, J. D.; Crimaldi, J. P.; and Nagel, K. I. 2018. Elementary sensory-motor transformations underlying olfactory navigation in walking fruit-flies. *Elife*, 7: e37815.

Bowen, M. 1991. The sensory physiology of host-seeking behavior in mosquitoes.

Cardé, R. T.; and Willis, M. A. 2008. Navigational strategies used by insects to find distant, wind-borne sources of odor. *Journal of chemical ecology*, 34(7): 854–866.

Corfas, R. A.; and Vosshall, L. B. 2015. The cation channel TRPA1 tunes mosquito thermotaxis to host temperatures. *elife*, 4: e11750.

Damos, P. T.; Dorrestijn, J.; Thomidis, T.; Tuells, J.; and Caballero, P. 2021. A temperature conditioned Markov chain model for predicting the dynamics of mosquito vectors of disease. *Insects*, 12(8): 725.

Davis, E. E.; and Sokolove, P. G. 1975. Temperature responses of antennal receptors of the mosquito, Aedes aegypti. *Journal of comparative physiology*, 96(3): 223–236.

Dekker, T.; and Cardé, R. T. 2011. Moment-to-moment flight manoeuvres of the female yellow fever mosquito (Aedes aegypti L.) in response to plumes of carbon dioxide and human skin odour. *Journal of Experimental Biology*, 214(20): 3480–3494.

Demir, M.; Kadakia, N.; Anderson, H. D.; Clark, D. A.; and Emonet, T. 2020. Walking Drosophila navigate complex plumes using stochastic decisions biased by the timing of odor encounters. *Elife*, 9: e57524.

Dickinson, M. 2006. Insect flight. *Current Biology*, 16(9): R309–R314.

Geier, M.; Bosch, O. J.; and Boeckh, J. 1999. Ammonia as an attractive component of host odour for the yellow fever mosquito, Aedes aegypti. *Chemical senses*, 24(6): 647–653.

Gillies, M. 1980. The role of carbon dioxide in host-finding by mosquitoes (Diptera: Culicidae): a review. *Bulletin of Entomological Research*, 70(4): 525–532.

Howlett, F. 1910. The influence of temperature upon the biting of mosquitoes. *Parasitology*, 3(4): 479–484.

Jones, J.; Murray, G. P.; and McCall, P. J. 2021. A minimal 3D model of mosquito flight behaviour around the human baited bed net. *Malaria Journal*, 20(1): 24.

Kearney, M.; Porter, W. P.; Williams, C.; Ritchie, S.; and Hoffmann, A. A. 2009. Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito Aedes aegypti in Australia. *Functional Ecology*, 23(3): 528–538.

Kennedy, J. S. 1940. The visual responses of flying mosquitoes.

Li, X.; Yin, X.; Li, C.; Zhang, P.; Hu, X.; Zhang, L.; Wang, L.; Hu, H.; Dong, L.; Wei, F.; et al. 2020. Oscar: Object-semantics aligned pre-training for vision-language tasks. In *European conference on computer vision*, 121–137. Springer.

Liu, M. Z.; and Vosshall, L. B. 2019. General visual and contingent thermal cues interact to elicit attraction in female Aedes aegypti mosquitoes. *Current Biology*, 29(13): 2250–2257.

Lu, J.; Batra, D.; Parikh, D.; and Lee, S. 2019. Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Advances in neural information processing systems*, 32.

McMeniman, C. J.; Corfas, R. A.; Matthews, B. J.; Ritchie, S. A.; and Vosshall, L. B. 2014. Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell*, 156(5): 1060–1071.

Murlis, J.; Elkinton, J. S.; Carde, R. T.; et al. 1992. Odor plumes and how insects use them. *Annual review of entomology*, 37(1): 505–532.

Nathan, R.; Getz, W. M.; Revilla, E.; Holyoak, M.; Kadmon, R.; Saltz, D.; and Smouse, P. E. 2008. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, 105(49): 19052–19059.

van Breugel, F.; and Dickinson, M. H. 2014. Plume-tracking behavior of flying *Drosophila* emerges from a set of distinct sensory-motor reflexes. *Current Biology*, 24(3): 274–286.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Viet Cuong, D.; Lalić, B.; Petrić, M.; Thanh Binh, N.; and Roantree, M. 2024. Adapting physics-informed neural networks to improve ODE optimization in mosquito population dynamics. *Plos one*, 19(12): e0315762.

Webster, B.; Lacey, E. S.; and Cardé, R. T. 2015. Waiting with bated breath: opportunistic orientation to human odor in the malaria mosquito, Anopheles gambiae, is modulated by minute changes in carbon dioxide concentration. *Journal of chemical ecology*, 41(1): 59–66.

Wolff, G. H.; and Riffell, J. A. 2018. Olfaction, experience and neural mechanisms underlying mosquito host preference. *Journal of Experimental Biology*, 221(4): jeb157131.

Yuan, Y.; Weng, X.; Ou, Y.; and Kitani, K. M. 2021. Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9813–9823.