STEPPY: MULTI-SCALE, MULTI-SPECIES WILDLIFE MOVEMENT FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Reliable forecasts of wildlife movement would enable numerous advances in ecology and conservation, but forecasting wildlife movement is a challenging predictive task, requiring models to represent complex, heterogeneous, and multi-scale movement across species and individual animals. Recent empirical studies indicate that even state-of-the-art wildlife movement forecasting methods only modestly outperform simple statistical approaches, even when leveraging complex environmental predictors. Further, existing methods are typically specialized for individual species and for specific predefined forecasting horizons, though research in ecology indicates that movement patterns may be shared across taxa and across different scales of movement. We propose a new movement forecasting model, STEPPY, that produces movement forecasts for any species and any time horizon dynamically at prediction time, leveraging ecologically-inspired architectural designs that enable it to learn useful and shared patterns across species and spatiotemporal scales. STEPPY outperforms existing state-of-the-art forecasting methods significantly, improving probabilistic forecasting by 18% on future trajectories from individuals who appear in the training data and 20% on trajectories for held-out individuals, even without incorporating any explicit environmental predictors. We will make our code, models, and data public upon publication.

1 INTRODUCTION

Wildlife movement is a complex, multi-scale phenomenon that shapes ecosystems around the world (Nathan et al., 2008). Advances in animal tracking technology from the past several decades—such as lightweight GPS hardware to track animals of diverse shapes and sizes across all corners of the globe—have provided scientists and decision makers with unprecedented insight into global-scale wildlife movement patterns and their consequences (Kays et al., 2015; Cagnacci et al., 2010). By analyzing this data we are beginning to understand how and why animals migrate (Aikens et al., 2022; Mouritsen, 2018; Jesmer et al., 2018), how to mitigate human-wildlife conflict (Blondin et al., 2020; Block et al., 2011; Pekarsky et al., 2021), and how anthropogenic activity influences and is influenced by movement (Oliver et al., 2026; Tucker et al., 2023).

The ability to reliably *forecast* future movement of individual animals could revolutionize conservation decision-making, for example by providing early warning signs of changes in migration patterns (Gomez et al., 2025; Forrest et al., 2026b). However, predictive methods are not yet capable of robustly forecasting wildlife movement across species, scales, and ecosystems (Kay et al., 2026). Existing methods, most notably step selection functions (SSFs; Fortin et al. (2005); Avgar et al. (2016)), largely originated as descriptive statistical models designed to make inferences about the impacts of various environmental factors on individual wildlife movement. SSFs were not specifically designed for forecasting, and they struggle to generalize in this context (Kay et al., 2026). Current state-of-the-art predictive movement methods—all deep learning extensions of SSFs (Forrest et al., 2026a; Cífka et al., 2023; Fronville et al., 2026)—only modestly outperform simple statistical approaches for predicting movement, and often generalize worse to new individuals than simply drawing random movements from population-level movement distributions (Kay et al., 2026).

We identify several weaknesses in current predictive methods. **First,** almost all existing methods are Markovian—they predict an animal's next movement conditioned only on its current location and bearing. This means that important behavioral properties such as behavioral persistence and spa-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tial memory are ignored or included simplistically as a derived spatial covariate (Kim et al., 2024), despite their influence on wildlife movement (Fagan et al., 2013; Toledo et al., 2020). **Second,** almost all existing methods are fit and evaluated on a single study of a single species, thus limited to the available data for that species, and do not have the capability to learn inter-species movement similarities and differences. The only exception is MoveFormer (Cífka et al., 2023), which uses a transformer (Vaswani et al., 2017) to predict the next movement conditional on both taxonomy and a sequence of recent movements. This has led to state-of-the-art generalization performance (Kay et al., 2026), yet empirically we find that its contextual reasoning is still limited. **Third,** most methods assume temporal regularity in locations and are designed to forecast at a single time horizon, requiring retraining to forecast at different horizons and limiting shared information across scales.

We introduce **STEPPY, the first multi-scale, multi-species wildlife movement forecasting model.** *Multi-scale:* STEPPY removes the need to train and predict at a fixed time horizon and learns to reason across scales when predicting future movement. *Multi-species:* STEPPY is able to learn movement patterns across species, and predicts better than single-species models. **STEPPY outperforms all prior work by a large margin**, improving probabilistic forecasting by 18% on future points from individuals who appear in the training data and 20% on new individuals who do not, by leveraging several ecologically-inspired architectural components including a novel relative spatiotemporal positional encoder and random Fourier basis spatial decoder. Finally, **STEPPY harnesses long-term contextual movement history** to improve predictions, and we show quantitatively and qualitatively that STEPPY's predictions continue to improve as we increase contextual movement history.

2 BACKGROUND

Wildlife movement data and terminology. Ecologists and conservationists use deployed on-animal sensors (*e.g.* GPS-based radio-telemetry tags (Cagnacci et al., 2010)) to collect information about animal movement. We define an animal movement *study* as a deployment of on-animal sensors to a population of animals. Studies typically focus on a single species, but range widely in the number of devices that are deployed, the sampling rates, and how long they record for. Data is often highly expensive and challenging to collect, but the scale of collected data is undeniable: there are over eleven billion logged animal locations on the animal tracking data repository Movebank (MoveBank, 2025).

Animal-mounted GPS devices typically collect temporally regular sequences of time-stamped locations. We define the basic unit of an animal movement sequence as a *step* (Fortin et al., 2005): a straight-line segment between successive locations in a sequence of observations of an individual animal. More formally, let $(\boldsymbol{\ell}_{t_0}, t_0), \ldots, (\boldsymbol{\ell}_{t_n}, t_n)$ with $t_0 < t_1 < \cdots < t_n$ denote a sequence of observed locations and times. Then, for $i = 1, \ldots, n$, the ith step is the movement from $\boldsymbol{\ell}_{t_{i-1}}$ to $\boldsymbol{\ell}_{t_i}$.

MOVEBENCH. We evaluate the performance of STEPPY on MOVEBENCH (Kay et al., 2026), the first and only large-scale benchmark for probabilistic wildlife movement forecasting. MOVEBENCH comprises 2.5M GPS trajectory data points from over 800 individuals of 110 different species of wildlife. These data are sourced from 147 independent observational studies, each with up to 10 individuals of 1–2 species.

MOVEBENCH formalizes probabilistic wildlife movement forecasting as follows. Given a desired forecast horizon $\tau > 0$, the goal is to predict a distribution over where an individual animal is likely to be τ units of time in the future. Let $\mathbf{L}_t \in \mathbb{R}^2$ denote the animal's location at time t, represented as a random vector in metric Cartesian coordinates, and let $\boldsymbol{\ell}_t$ denote the observed location. Let $c_t^{(\tau)}$ denote any contextual information available at a forecast origin $(\boldsymbol{\ell}_t, t)$, which may contain past locations and other conditioning information and may depend on τ. Then, formally, the task is to model the conditional predictive distribution of $\mathbf{L}_{t+\tau}$,

$$P\Big(\cdot \mid c_t^{(\tau)}, \tau\Big). \tag{1}$$

MOVEBENCH evaluates predictions using the ***energy score*** **(ES)**, a proper scoring rule taking the following form:

$$\mathrm{ES}(P, \boldsymbol{\ell}_{t+\tau}) = \underbrace{\mathbb{E}_{\tilde{\mathbf{L}} \sim P} \big\| \tilde{\mathbf{L}} - \boldsymbol{\ell}_{t+\tau} \big\|_2}_{\text{expected Euclidean error of forecast}} - \frac{1}{2} \underbrace{\mathbb{E}_{\tilde{\mathbf{L}}, \tilde{\mathbf{L}}' \overset{\text{iid}}{\sim} P} \big\| \tilde{\mathbf{L}} - \tilde{\mathbf{L}}' \big\|_2}_{\text{spread of forecast}}.$$

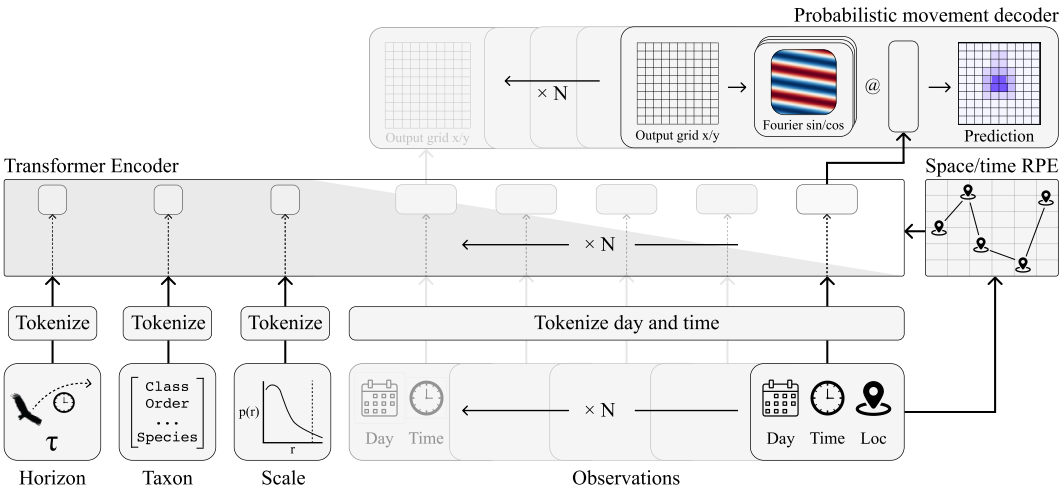


Figure 1: **STEPPY model architecture.** STEPPY generates probabilistic forecasts of where an animal is likely to be τ units of time in the future. **Inputs (bottom):** STEPPY's predictions are conditioned on the forecast horizon τ, the animal's taxonomic information, a population-level spatial scale, and the animal's N recent movements, represented as a sequence of discrete spatiotemporal observations recorded by GPS. **Encoder (middle):** Each input is tokenized and input to a transformer encoder. Observations receive a relative spatiotemporal encoding based on their distance from each other in world coordinates and time, allowing them to attend to other steps based on their spatial and temporal distance from each other. The encoder uses a causal self-attention mask that ensures observations only attend to observations that occur before them. **Decoder (top):** The hidden representation of each observation token is combined with random Fourier basis features of candidate locations to produce a spatial prediction of where the animal will be τ units of time in the future.

Where P is a τ-horizon predictive distribution for the future location $\mathbf{L}_{t+\tau}$ issued at time t (as in Eq. (1)), and $\boldsymbol{\ell}_{t+\tau}$ is the ground truth location at time $t+\tau$. Lower values represent better forecasts. The ES rewards probabilistic forecasts that both accurately predict movement (the first term) and appropriately represent uncertainty (the second term). Like all proper scoring rules, it is minimized in expectation by predicting the true conditional distribution of the future location (Gneiting & Raftery, 2007).

MOVEBENCH proposes evaluation of wildlife movement forecasting at four forecast horizons: $\tau \in \{1 \text{ hour}, 12 \text{ hours}, 1 \text{ day}, 7 \text{ days}\}$. It tests methods at each horizon over two distinct generalization challenges: **(1) a *temporal* test set**, which evaluates how well methods forecast future movement of individuals included in the training set, and **(2) an *individual* test set**, which evaluates how well methods forecast movement of individuals not included in the training set.

3 METHOD

To represent trajectories and generate predictions, STEPPY utilizes an autoregressive transformer architecture (Vaswani et al., 2017) in combination with a novel input representation and spatiotemporal encoding, probabilistic movement decoder, and multi-scale training strategy. See Fig. 1.

Sequence-level inputs. For each sequence, we tokenize: (1) the species that the sequence corresponds to, (2) the desired temporal forecast horizon the model should predict future movement for, and (3) an individual-specific movement-scale representation calculated as a summary statistic from the training data of the individual (if past data from the individual is present in the training data), or an aggregate estimate from the set of individuals of that species in the training data (for held-out individuals). Details of the implementations of each are in Appendix Sec. A.

Observation inputs. The remainder of the inputs to STEPPY represent observations—locations and times at which the animal was observed. Each corresponds to a single recorded GPS point. We use observations in two ways: first, we tokenize the date and time as inputs to the encoder (see Appendix Sec. A for details); second, sptiotemporal relationships between observations enter the model through a novel adaptation of relative positional encoding described below.

Encoder. STEPPY's encoder is a transformer (Vaswani et al., 2017) with causal self-attention on the observations. That is, we create a self-attention mask that enforces that all observations can only attend to observations that occurred before them. All tokens can attend to the tokens for forecast horizon, spatial scale, and taxon. By default, we use a four layer transformer with 12 attention heads per layer, each with 64 channels, and hidden dimension 768, which gave the best trade-off between computation and validation-set performance when trained on MOVEBENCH.

Relative spatiotemporal encoding. We use relative positional encodings Shaw et al. (2018) to provide each observation with context about its spatial and temporal distance to other observations in the sequence. We implement this as a 3D version of rotary positional encodings (RoPE; Su et al. (2024)), where the three dimensions are x, y, and time. Unlike prior two or three-dimensional extensions to RoPE (Heo et al., 2024; Liu et al., 2025), we *compute our relative spatial distances in world coordinates*, *i.e.* in meters in a Cartesian (projected) coordinate system. This allows attention weights within the encoder to depend on actual spatial and temporal distances between observations, enabling observation tokens to learn to attend more or less to each other based on distance in absolute space and time.

Specifically, for observation $(\boldsymbol{\ell}_{t_i}, t_i)$, we define normalized spatiotemporal coordinates $\mathbf{p}_i = (\ell_{t_i,x}/d(\tau),\ \ell_{t_i,y}/d(\tau),\ t_i/\tau)$, where $d > 0$ is a spatial scale and τ is the forecast horizon. For each of the three coordinates, RoPE rotates pairs of query and key features by angles proportional to that coordinate at a set of fixed frequencies. Consequently, the attention score between observations i and j depends on the differences between these rotation angles, and therefore on their normalized relative offset $\mathbf{p}_j - \mathbf{p}_i = ((\ell_{t_j,x} - \ell_{t_i,x})/d(\tau),\ (\ell_{t_j,y} - \ell_{t_i,y})/d(\tau),\ (t_j - t_i)/\tau)$. We use fixed axial RoPE as in Heo et al. (2024) with base frequency 10,000 as in Su et al. (2024), applied separately to x, y, and time and only applied to observation tokens. Our implementation allocates 24 of the 64 channels per layer to each spatial coordinate and 16 to time.

Decoder. We decode from the final hidden representation of each observation token to a spatial probability map of the likelihood that the animal moves to any location "near" the current observation at τ-time-horizon. We concurrently predict a forecast for every observation in the sequence regardless of how many predecessor observations it has, so that the model learns to forecast with different amounts of context. Taking inspiration from MOVEBENCH, we represent this prediction as a discrete 101×101 grid spanning $2d(\tau)$ in both directions as defined in Eq. (2). To produce a spatial prediction, we project each hidden observation representation into a query and take its dot product with learned embeddings of normalized grid coordinates. These embeddings use random Fourier features with Gaussian-sampled frequencies (Rahimi & Recht, 2007; Tancik et al., 2020), allowing predictions to capture spatial structure scross scales. A softmax over the grid converts the resulting scores into probabilities.

Multi-scale training. Our goal is to train STEPPY to make forecasts for any continuous time horizon for any animal. At every training iteration, for a given minibatch size B and sequence length N, we: **(1)** Select a forecast horizon. Empirically, we find that shorter forecast horizons require more training steps to converge. We therefore sample $\tau \sim p(\tau)$, where $p(\tau) \propto \tau^{-1.25}$ for $\tau \in [20\,\text{min}, 7\,\text{days}]$, allocating more training iterations to shorter horizons. We set the sampling exponent to 1.25 based on validation experiments. **(2)** Select B studies for the batch. We sample uniformly across all studies. **(3)** Sample a single trajectory of length $\leq N$ from each of the B studies. We first sample a observation at random from the study, which becomes the final point in our context window (the "current location"), and take the $N - 1$ previous steps as our training sequence, padding if neccesary to build a sequence of length N. In our experiments, $B = 8$ and $N = 512$ by default based on validation performance.

Additional training data outside of MOVEBENCH. MOVEBENCH's preprocessing involves subsampling individuals (5–10 per study), trajectory lengths (maximum of two years of observations per individual), and sampling rates (maximum of one observation per hour). To explore the impact of training data scale on predictive performance, we additionally download all raw data from the 147 publicly-available studies in the benchmark and add every trajectory from every individual to our training set, sampled at a maximum of one observation every 20 minutes. To prevent leakage with the MOVEBENCH test data, we truncate each training trajectory at the earliest time point present in that study's temporal test set, so there is no temporal overlap between training and test data. This process adds an additional 19,093,436 training observations from an additional 3,908 individuals. As STEPPY does not require access to environmental features, adding this data is simple. For baselines

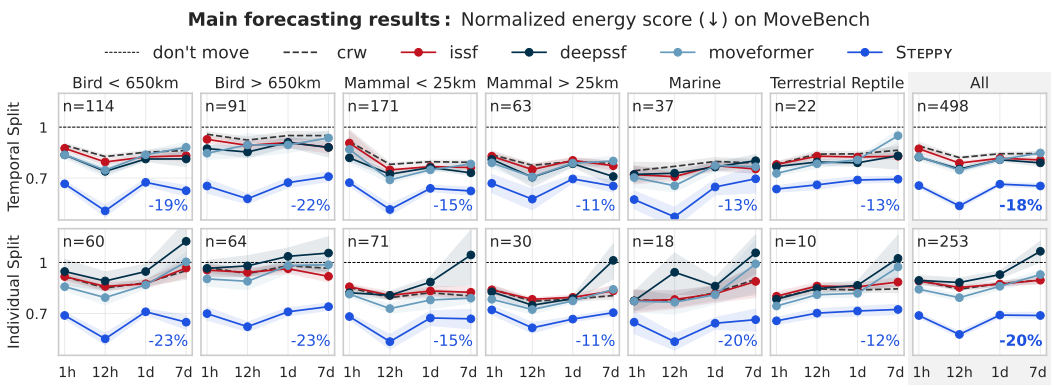


Figure 2: **Probabilistic forecasting results on MoveBench.** Results comprise 110 species aggregated by taxonomy and median per-study maximum movement distance. Lower normalized energy scores are better. **Steppy outperforms existing state-of-the-art wildlife movement forecasting methods significantly**. It is more accurate for every taxonomic grouping and every forecast horizon in MoveBench, and is the only method to consistently outperform the correlated random walk (CRW) baseline in every scenario. On average, Steppy produces 18% better forecasts than the next-best method for held-out future time points of individuals that occur in the training data, and 20% better forecasts for held-out individuals that do not.

which *do* require environmental features, the storage and processing requirements are prohibitive. For fair comparison we other methods—which are trained only on the original MoveBench training set—we report results with and without this additional training data.

4 Experiments

4.1 Main forecasting results

We evaluate Steppy against the following baselines, as proposed by MoveBench: **(1) Don't move**, a simple baseline that always predicts that the animal remains in its current location; **(2) Correlated random walk**, a simple baseline that always predicts the study-level training-set marginal step-length and turn-angle distribution for each individual; and the three best-performing existing methods on MoveBench: **(3) Integreted step selection functions (iSSF, Avgar et al. (2016))**, **(4) deepSSF (Forrest et al., 2026a)**, and **(5) MoveFormer (Cífka et al., 2023)**, each with AlphaEarth (Brown et al., 2025) environmental covariates. As per the original designs of each method, we fit and evaluate (2), (3), and (4) independently per-study, and (5) is the only method trained and evaluated cross-study, as it is the only other method that natively supports multi-species training.

The main probabilistic forecasting results on MoveBench are visualized in Fig. 2. We follow the evaluation protocol in MoveBench: first, energy scores are computed across all individuals in each study at each forecast horizon. They are then normalized by the average movement distance in that study at that time horizon in order to make them comparable across studies and prevent animals with larger spatial ranges from dominating aggregate metrics. Per-study energy scores are aggregated according by taxonomic groups (bird, mammal, marine, and terrestrial reptile), and by the median travel distance of the individuals in the study. Recall that lower energy scores are better.

Steppy outperforms every baseline at every time horizon for every taxonomic/range grouping in both test sets, often significantly, despite using no explicit environmental information at training or evaluation time (unlike many of the baselines). On average, Steppy outperforms the next best model by 18% on the temporal test set and by 20% on the individual test set. The largest improvements over the next-best model are a 22% and 23% reduction in ES for long-distance birds in the temporal and individual test sets, respectively. Steppy is also the only method to consistently outperform the simple correlated random walk baseline on held-out individuals. Compared to the correlated random walk, Steppy's forecasts are 23% better on both the temporal and individual test sets, compared to the next-best method's respective improvements of only 6% and 2%.

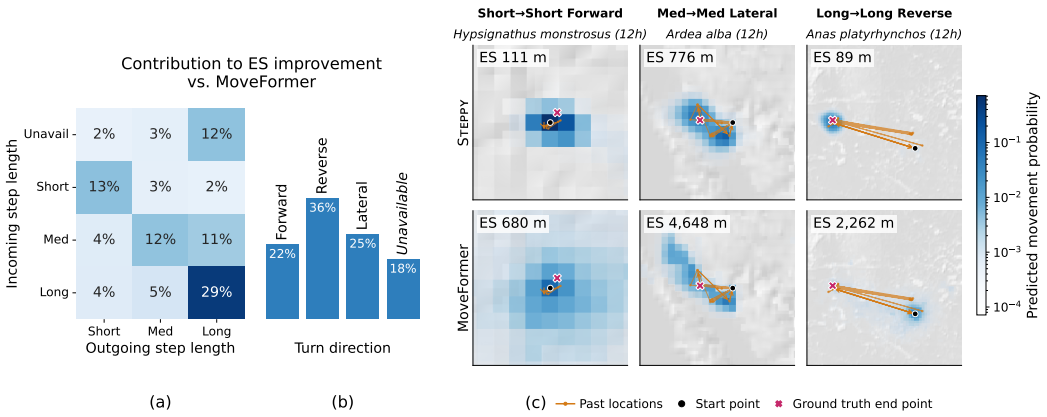


Figure 3: **Performance improvement breakdown.** We stratify test predictions of STEPPY and the next best method, MoveFormer, by **(a)** step length, *i.e.* distance between subsequent observations (short, medium, and long) and **(b)** turn direction, *i.e.* change in bearing between steps (forward, lateral, and reverse), and compute the contribution of each to STEPPY's overall energy score improvement. Predictions where recent step length is maintained (*e.g.*a long step followed by another long step), and where heading is reversed, are the largest contributing categories. In **(c)**, we visualize examples. In the short→short and medium→medium cases, STEPPY makes use of recent context to predict tighter distributions around recent movement locations than MoveFormer, improving its forecast accuracy. In the long→long reverse case, STEPPY uses recent historical examples to predict the animal will return to a recent location it has revisited several times.

In Fig. 3 (a) and (b), we quantitatively break down STEPPY's improvements over the next-best method, MoveFormer, based on *step lengths* and *turn angles* between steps, computed at the four time horizons in MOVEBENCH. More details on how these are computed are in Appendix Sec. B.

When analyzing step lengths (Fig. 3a), we find that the majority of STEPPY's improvements (54%) come from more accurately predicting that an animal maintains its most recent historical step patterns: *i.e.*, a short step preceding a short step or a long step preceding a long step. We also see that 17% of STEPPY's improvement comes from cases where incoming step-length information is unavailable (*e.g.* at the beginning of a sequence), and this is most notable for long outgoing steps. When a discontinuity in a trajectory is observed, MoveFormer truncates the context, while STEPPY includes historical tokens regardless of their temporal spacing, allowing it to incorporate more historical context. When analyzing turn angles (Fig. 3b), we see that 36% of STEPPY's improvement over MoveFormer comes from better predictions of movement reversal, *i.e.*when an animal turns back toward the origin of its most recent step. Forward and lateral movements also contribute.

We show qualitative examples of these improvements in Fig. 3c. In the first two columns, STEPPY makes use of recent short-to-medium length movements to predict that in the next step the animal will remain within a small range that it has been frequenting, while MoveFormer constructs a more diffuse prediction that assigns probability mass outside of the local area. In the rightmost column, a mallard (*Anas platyrhynchos*) displays a pattern in recent history whereby it frequents two different locations repeatedly, reversing its trajectory between each visit. STEPPY takes advantage of this context, predicting confidently that it will return to a commonly-visited location, whereas MoveFormer assigns high probability to the animal remaining in its current location.

4.2 ABLATION STUDIES

We ablate the key components of STEPPY, and report results in Table 1. For each row, we remove one component from STEPPY and report the change in normalized energy score averaged over all studies and timescales for each test set. Recall positive changes indicate worsened performance.

Inputs and history. We see that STEPPY is relatively robust to the set of features used in the observation tokens. Switching our representation (time of day and day of year) for MoveFormer's (absolute latitude, longitude, time, step length and bearing) has a minimal effect, worsening performance on held-out individuals by approximately 1%. Features for time of day improve predictive performance by 2–3%, while day of year features have minimal effect. This indicates that STEPPY has learned some diel movement patterns, but annual trends have either not been learned or are not

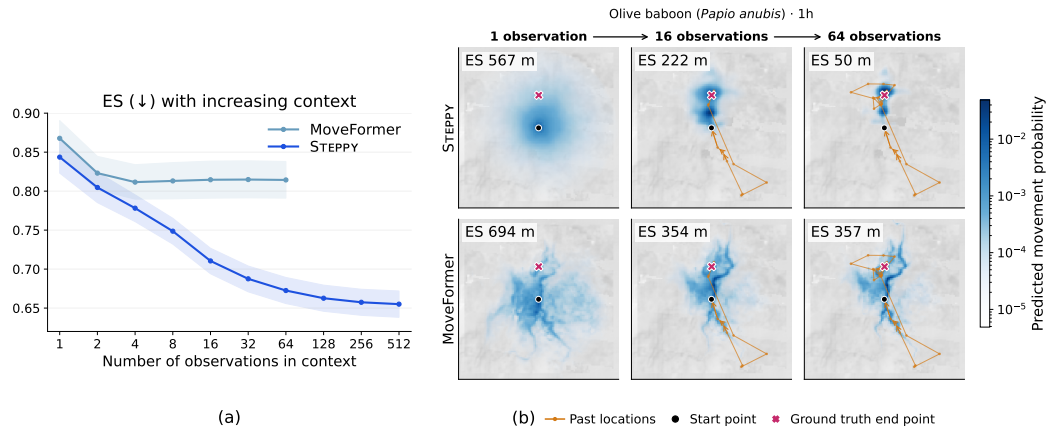


Figure 4: **Effects of longer step context on forecasting accuracy. (a)** We truncate the amount of context available to STEPPY and MoveFormer at test time and evaluate the effects on forecasting accuracy. We see that both methods perform similarly with no historical context. As more context is available, STEPPY's predictions improve considerably, and most of this improvement comes from the first 20% of available context; MoveFormer's improvements with additional context are more modest. **(b)** Qualitative examples of predictions with increasing context. Each panel represents the same test example, with predictions from STEPPY and MoveFormer made with different amounts of context. As STEPPY receives more context from recent movements, its predictions get progressively tighter around locations that are consistent with recent spatial patterns, whereas MoveFormer's predictions remain diffuse within a broader area of recent movement.

useful for this data. Interesting, including *no features whatsoever* in the input tokens—making each observation a simple learned token—only reduces performance by approximately 2%, indicating that most predictive performance gains stem from the spatial reasoning elsewhere in the network.

In contrast, observation history has a significant effect. We compare two history construction choices of MoveFormer to our decision to include all possible observation history in the context window: truncating the history at "burst" boundaries, *i.e.* when there are long temporal gaps between observations, and using "steps" spaced at the forecast horizon rather than all observations at their native sampling rate. Truncating at gaps leads to a 11–13% performance drop, while using steps leads to an approximate 2% performance drop. These results indicate that STEPPY is able to make use of contextual observations despite sampling noise (*e.g.* long gaps between fixes) or mismatch between sampling rates and forecast horizon. Importantly, STEPPY is able to learn useful patterns across scales: performance improves for longer forecast horizons when we include higher-temporal-resolution observations.

Table 1: STEPPY ablations.

Component ablated	ΔES (↓) Temporal	Individual
Observation tokens		
→ No date	+0.2%	+0.1%
→ No time	+2.4%	+2.3%
→ No features	+2.3%	+1.8%
→ Use MoveFormer rep	+0.0%	+1.0%
→ Truncate at gaps	+12.6%	+11.3%
→ Use τ-horizon steps	+2.1%	+1.8%
Taxonomy input	+0.4%	+1.2%
Scale input	+0.4%	−0.3%
Space/time RPE		
→ Space only	+1.9%	+1.4%
→ Time only (XY feat)	+3.9%	+3.0%
→ None (XY + PE)	+4.4%	+3.4%
Decoder		
→ $\Delta x, \Delta y$ regression	+75.0%	+64.6%
→ MoveFormer head	+4.5%	+5.0%
→ Linear layer	+0.3%	+0.1%
Cross-study training	+12.4%	+9.5%
Cross-horizon training	+1.8%	+1.5%
Continuous horizon training	+0.5%	+0.3%
Extra training data	+6.3%	+4.9%

We further explore the use of movement history in Fig. 4 by varying the historical context seen by STEPPY and MoveFormer at test time, from a single step (*i.e.* current location only) up to the full context available. STEPPY's performance improves significantly with more context (see Fig. 4a), and we continue to see improvements up to a 512 context length (our default setting); on the other hand, MoveFormer benefits from some context but plateaus after just 4 historical observations. In Fig. 4b we show two qualitative examples of this on an example trajectory of an olive baboon (*Papio anubis*). The predictions made by both methods with no historical observations are quite similar spatially, reflecting learned general movements of the species. Recall that MoveFormer, unlike STEPPY, also has access to explicit environmental context, which leads to visible environmental artifacts from elevation changes in the landscape, but

its energy score is worse. As more context is added, STEPPY's predictions improve significantly, tightening the forecast to reflect the recurring movement pattern exhibited by the individual.

We find taxonomic inputs modestly improve performance, up to 1.2% on held-out individuals. Including the individual scale token explicitly also has minimal effect on predictive performance, however the model also has access to this scale information via the relative positional encoding.

Space/time relative positional encoding. Removing the relative temporal information from the positional encoding, so that only the day and time features from the observation input are available, reduces performance by 1–2%. Removing the relative spatial information (replacing it with x and y features in the input token) has a larger effect of between 3–4%. Removing the relative positional encoding entirely, instead relying on x and y features in the input tokens and using a standard index-based position encoding, worsens performance up to a 4.4% performance drop on the temporal test.

Decoder. STEPPY predicts a probabilistic spatial map, trained with a negative log-likelihood (NLL) loss. Replacing this objective with a simple regression head that predicts a $\Delta x, \Delta y$ is catastrophic, leading to a 65–75% reduction in performance. Using MoveFormer's step-selection head—which is trained with NLL on subsampled candidate locations rather than the full output grid—leads to approximately 5% worse performance, indicating that learning on the full spatial grid is useful. A simple linear decoder without Fourier features performs on par with our Gaussian Fourier approach, but with significant computational cost—7.84M parameters versus 180k for our proposed decoder.

Other training techniques. Incorporating additional training data beyond MOVEBENCH has a significant effect, improving performance by 5–6%. Training one model across studies—compared to training a single model per study—improves performance even more significantly on both held-out time points and held-out individuals (a 12.4% and 9.5% improvement respectively). This indicates that additional training data, even from different study systems, is helpful for learning general conditional predictive patterns of animal movement. Training across forecast horizons—compared to a single model per horizon—improves performance by 1–2%. Including fully continuous horizons into training (rather than training just at the four time horizons in MOVEBENCH) further improves performance by up to 0.5% at those four horizons. These takeaways are notable—from species to time horizons, a generalist model outperforms specialist models trained for each species or time horizon directly, further indicating STEPPY's ability to share information across species and scales.

5 DISCUSSION AND CONCLUSIONS

STEPPY is the first multi-scale, multi-species model for forecasting wildlife movement, and it improves probabilistic forecasting over existing methods by 18% on future points from known individual animals and 20% on unknown individuals, a major step forward for the field. STEPPY has numerous potential applications for understanding animal movement processes such as behavioral states, spatial memory and cognitive maps, as well as practical applications in biodiversity conservation. We highlight key takeaways and opportunities for future work:

We can accurately predict wildlife movement without access to explicit environmental features. One interesting design decision of STEPPY is the choice to predict wildlife movement *without* explicit environmental context, which is in direct contrast to most prior wildlife movement forecasting methods. A notable outcome of this work is the discovery that *high predictive accuracy can be achieved without explicit access to environmental information*. This result highlights that important movement properties are contained in the structure of the observed movement trajectory, such as behavioral states, spatial memory, and even environmental interactions (such as revisiting productive foraging locations), even if these properties are not labeled explicitly in the data. STEPPY is able to learn and use such implicit patterns to improve predictive performance. An additional benefit is that, by not requiring explicit environmental information, STEPPY is significantly more lightweight to train and deploy both computationally and infrastructurally (Kay et al., 2026), while providing a strong foundation for future extensions that may combine STEPPY's powerful conditional movement predictions with explicit environmental or biological information.

STEPPY learns general *and* species-specific movement patterns. The predictive performance of STEPPY improves when training across taxa, even when explicit taxonomic information is not provided, and improves further with explicit taxonomic inputs. This suggests that pooling data across species provides useful training signal for learning general conditional movement patterns, even

when these patterns differ considerably across species. This does not mean that STEPPY is "taxon-free" without these labels, just as it is not "environment-free" without environmental labels—it may implicitly learn to leverage species-specific patterns in recent movement context. An exciting direction for future work is to explore to what extent these derived patterns reflect ecological phenomena such as phylogenic, morphological, lifestyle, or size-based traits (Beumer et al., 2026).

Opportunities to extend STEPPY. *Adding explicit environmental context.* Although STEPPY produces highly accurate and realistic predictions, some predictive outputs, such as estimating landscape connectivity and testing hypothetical scenarios (Cowan et al., 2025; Hofmann et al., 2024), require a representation of the spatial landscape for animals to respond to. Providing a representation of environmental context may further improve STEPPY's performance and engender a wider range of applications. *Beyond next-step forecasting.* While STEPPY is able to produce forecasts for any time horizon at prediction time, its forecasts can still be considered conditional *next-step* predictions. The ability to additionally predict future paths—*i.e.* multi-step predictions—directly could provide additional ecological insight into the possible routes animals take through landscapes. Extending this to even longer time horizons could enable more accurate predictions of longer-range phenomena such as space use and home ranging (Signer et al., 2024; Potts et al., 2022). *Toward a foundation model for wildlife movement.* We have seen initial evidence that methods like STEPPY can improve performance with additional training data. Further, STEPPY (and the other movement forecasting baselines benchmarked here) operate solely on GPS data and remotely-sensed environmental data; incorporating other animal-borne sensors such as accelerometers may improve predictive performance and enable learning more fine-scale and transferable patterns across scales. The potential of such scaling is not yet well understood, and will required coordinated data collection efforts to address gaps and increase the diversity, scale, and representativeness of public movement data.

6 ADDITIONAL RELATED WORK

Forecasting non-wildlife movement. The most common trajectory data sources in the machine learning literature involve human and vehicle tracking (Zheng et al., 2010; Yuan et al., 2010; O'Connell et al., 2015). Early forecasting methods for vehicles or humans draw on statistical state-space and hidden Markov models, as well as recurrent architectures such as LSTMs for direct sequence forecasting (Langrock et al., 2012; Alahi et al., 2016). Contemporary methods leverage graph neural networks and transformer-based architectures to capture complex patterns in trajectory prediction, including tasks like single-step forecasting, multi-step prediction, and behavioral classification (Shi et al., 2021; 2023; Fujii et al., 2025; Wei et al., 2024; Zhu et al., 2024). However, many of these methods are explicitly map-constrained: they assume movement happens only within fixed structures, like roads, and leverage this structure extensively (Newson & Krumm, 2009; Wang et al., 2022). Animal movement instead operates in continuous space, as well as at larger and more diverse spatiotemporal scales than is captured by existing human and vehicle benchmarks (Kay et al., 2026). Further, animal movement is inherently stochastic, requiring probabilistic forecasts, whereas most human/vehicle methods focus on point predictions Shi et al. (2021; 2023); Pellegrini et al. (2009).

Wildlife movement forecasting. In ecological contexts, *wildlife movement* has instead traditionally been conceptualized through the lens of *habitat selection*, where movement is driven by an animal's assessment of environmental conditions at prospective destinations. This perspective underlies *step-selection functions* (Fortin et al., 2005; Avgar et al., 2016), which implement this conceptual framework as a binary classification task where models must differentiate between observed locations and unobserved alternatives based on environmental features and their movement behavior. Deep learning architectures have begun to gain popularity alongside the statistical models used in traditional step selection methods (Forrest et al., 2026a; Cífka et al., 2023; Fronville et al., 2026).

While not their original intended use, step-selection functions have been applied to forecast wildlife movement (Signer et al., 2024; Rogers et al., 2025; Signer et al., 1771; Potts & Börger, 2023; Hofmann et al., 2023; Cowan et al., 2025; Westaway et al., 2025; Forrest et al., 2025; 2026b). Beyond step selection, a few works directly forecast movement for single studies, including southern elephant seals (Medina Hernández et al., 2025), sika deer (Kazama et al., 2024), and migratory gulls Wijeyakulasuriya et al. (2020). Of these prior works only MoveFormer (Cífka et al., 2023) is capable of multi-species wildlife movement forecasting, and no prior work enables movement forecasts to be generated dynamically at any timescale. STEPPY is the first work to enable both.

AI USE STATEMENT

Tasks with required disclosure: In this work, we used generative AI tools to help implement methods. We have not used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, assist with translation, clean and reformat dataset, support qualitative and thematic data analysis, or interpret results.

Tasks with recommended disclosure: Additionally, we used generative AI tools to create and modify scientific figures, suggest experimental parameters, and create and edit software code.

We have reviewed all AI-assisted work. All LLM-generated code was reviewed for correctness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Code is included in the supplemental material. All data used in our experiments is publicly available with Creative Commons licensing.

REFERENCES

Ellen O Aikens, Iris D Bontekoe, Lara Blumenstiel, Anna Schlicksupp, and Andrea Flack. Viewing animal migration through a social lens. *Trends in Ecology & Evolution*, 37(11):985–996, 2022.

Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 961–971, 2016.

Tal Avgar, Jonathan R Potts, Mark A Lewis, and Mark S Boyce. Integrated step selection analysis: bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, 7(5):619–630, 2016.

Larissa T Beumer, Anne G Hertel, Raphaël Royauté, Marlee A Tucker, Jörg Albrecht, Roxanne S Beltran, Francesca Cagnacci, Sarah C Davidson, Nandintsetseg Dejid, Roland Kays, Andrea Kölzsch, Ashley Lohr, Eike Lena Neuschulz, Kamran Safi, Anne K Scharf, Matthias Schleuning, Martin Wikelski, and Thomas Mueller. MoveTraits-A database for integrating animal behaviour into trait-based ecology. 29:e70297, 2026. ISSN 1461-023X,1461-0248. doi: 10.1111/ele.70297.

Barbara A Block, Ian D Jonsen, Salvador J Jorgensen, Arliss J Winship, Scott A Shaffer, Steven J Bograd, Elliott Lee Hazen, David G Foley, GA Breed, A-L Harrison, et al. Tracking apex marine predator movements in a dynamic ocean. *Nature*, 475(7354):86–90, 2011.

Hannah Blondin, Briana Abrahms, Larry B Crowder, and Elliott L Hazen. Combining high temporal resolution whale distribution and vessel tracking data improves estimates of ship strike risk. *Biological Conservation*, 250:108757, 2020.

Christopher F Brown, Michal R Kazmierski, Valerie J Pasquarella, William J Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, et al. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025.

Francesca Cagnacci, Luigi Boitani, Roger A Powell, and Mark S Boyce. Animal ecology meets gps-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2157–2162, 2010.

Ondřej Cífka, Simon Chamaillé-Jammes, and Antoine Liutkus. Moveformer: a transformer-based model for step-selection animal movement modelling. *bioRxiv*, pp. 2023–03, 2023.

Mitchell A Cowan, Scott W Forrest, Samantha A Setterfield, Judy A Dunlop, Lesley A Gibson, and Dale G Nimmo. The impact of mining on animal movement and landscape connectivity revealed through simulations and scenarios. *Ecological Applications*, 35(7):e70134, 2025.

William F Fagan, Mark A Lewis, Marie Auger-Méthé, Tal Avgar, Simon Benhamou, Greg Breed, Lara LaDage, Ulrike E Schlägel, Wen-Wu Tang, Yannis P Papastamatiou, James Forester, and Thomas Mueller. Spatial memory and animal movement. 16:1316–1329, 2013. ISSN 1461-023X,1461-0248. doi: 10.1111/ele.12165.

Scott W Forrest, Dan Pagendam, Michael Bode, Christopher Drovandi, Jonathan R Potts, Justin Perry, Eric Vanderduys, and Andrew J Hoskins. Predicting fine-scale distributions and emergent spatiotemporal patterns from temporally dynamic step selection simulations. *Ecography*, 2025 (2):e07421, 2025.

Scott W Forrest, Dan Pagendam, Conor Hassan, Jonathan R Potts, Christopher Drovandi, Michael Bode, and Andrew J Hoskins. Predicting animal movement with deepSSF: A deep learning step selection framework. *Methods in Ecology and Evolution*, 17(2):371–391, 2026a.

Scott William Forrest, Mitchell A Cowan, David Denis Hofmann, Natasha Klappstein, Jonathan Potts, Théo Michelot, John Fieberg, Johannes Signer, James Pay, Charlotte R Patterson, Andrew Hoskins, Rachel Mawer, Luca Börger, Michael Bode, and Dale Nimmo. Predictive movement ecology using spatial simulations. *EcoEvoRxiv preprint*, 2026b.

Daniel Fortin, Hawthorne L Beyer, Mark S Boyce, Douglas W Smith, Thierry Duchesne, and Julie S Mao. Wolves influence elk movements: behavior shapes a trophic cascade in yellowstone national park. *Ecology*, 86(5):1320–1330, 2005.

Thibault Fronville, Maximilian Pichler, Johannes Signer, Marius Grabow, Stephanie Kramer-Schadt, Viktoriia Radchuk, and Florian Hartig. Analyzing animal movement using deep learning. *arXiv preprint arXiv:2603.24009*, 2026.

Ryo Fujii, Hideo Saito, and Ryo Hachiuma. Towards predicting any human trajectory in context. *arXiv preprint arXiv:2506.00871*, 2025.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Sara Gomez, Holly M English, Vanesa Bejarano Alegre, Paul G Blackwell, Anna M Bracken, Eloise Bray, Luke C Evans, Jelaine L Gan, W James Grecian, Catherine Gutmann Roberts, et al. Understanding and predicting animal movements and distributions in the anthropocene. *Journal of Animal Ecology*, 94(6):1146–1164, 2025.

Byeongho Heo, Song Park, Dongyoon Han, and Sangdoo Yun. Rotary position embedding for vision transformer. In *European Conference on Computer Vision*, pp. 289–305. Springer, 2024.

David D Hofmann, Gabriele Cozzi, John W McNutt, Arpat Ozgul, and Dominik M Behr. A three-step approach for assessing landscape connectivity via simulated dispersal: African wild dog case study. *Landscape Ecology*, 38(4):981–998, 2023. doi: 10.1007/s10980-023-01602-4.

David D Hofmann, Dominik M Behr, John W McNutt, Arpat Ozgul, and Gabriele Cozzi. Dispersal and connectivity in increasingly extreme climatic conditions. *Global Change Biology*, 30(5): e17299, 2024.

Brett R Jesmer, Jerod A Merkle, Jacob R Goheen, Ellen O Aikens, Jeffrey L Beck, Alyson B Courtemanch, Mark A Hurley, Douglas E McWhirter, Hollie M Miyasaki, Kevin L Monteith, et al. Is ungulate migration culturally transmitted? evidence of social learning from translocated animals. *Science*, 361(6406):1023–1025, 2018.

Justin Kay, Shir Bar, Ellen O. Aikens, Martin Becker, Francesca Cagnacci, Juliet Cohen, Scott W. Forrest, Jessica Kendall-Bar, Madeleine Lucas, Macon Overcast, Meredith S. Palmer, Will Rogers, Nicholas J. Russo, Christian Rutz, Larissa T. Beumer, Michael Brown, Ying-Chi Chan, Sarah C. Davidson, Diego Ellis Soto, Anne G. Hertel, Roland Kays, Benjamin Koger, Guram Mikaberidze, Thomas Mueller, Ruth Oliver, Thorsten Papenbrock, Robert Patchett, Jared A. Stabach, Dane Taylor, Scott W. Yanco, and Sara Beery. Movebench: A benchmark for global-scale wildlife movement forecasting, 2026. URL https://arxiv.org/abs/2609.15780.

Roland Kays, Margaret C Crofoot, Walter Jetz, and Martin Wikelski. Terrestrial animal tracking as an eye on life and planet. *Science*, 348(6240):aaa2478, 2015.

Kentaro Kazama, Katsuhide Fujita, Yushin Shinoda, and Shinsuke Koike. Sika deer trajectory prediction considering environmental factors by timeseries transformer-based architecture. *Expert systems with applications*, 250:123630, 2024.

Dongmin Kim, Peter R Thompson, David W Wolfson, Jerod A Merkle, L G R Oliveira-Santos, James D Forester, Tal Avgar, Mark A Lewis, and John Fieberg. Identifying signals of memory from observations of animal movements. 12:72, 2024. ISSN 2051-3933,2051-3933. doi: 10.1186/s40462-024-00510-9.

Roland Langrock, Ruth King, Jason Matthiopoulos, Len Thomas, Daniel Fortin, and Juan M. Morales. Flexible and practical modeling of animal telemetry data: Hidden markov models and extensions. *Ecology*, 93(11):2336–2342, 2012. doi: 10.1890/11-2241.1.

Zikang Liu, Longteng Guo, Yepeng Tang, Tongtian Yue, Junxian Cai, Kai Ma, Qingbin Liu, Xi Chen, and Jing Liu. Vrope: Rotary position embedding for video large language models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 14471–14483, 2025.

Jorge Medina Hernández, Jorge P Rodríguez, Clive R McMahon, Ana MM Sequeira, and Víctor M Eguíluz. Improving prediction region accuracy in marine animal movement with temporal fusion transformer. *Scientific Reports*, 2025.

Henrik Mouritsen. Long-distance navigation and magnetoreception in migratory animals. *Nature*, 558(7708):50–59, 2018.

MoveBank. Movebank for animal tracking, 2025. URL `https://www.movebank.org/cms/movebank-main`. `https://www.movebank.org/cms/movebank-main`.

Ran Nathan, Wayne M Getz, Eloy Revilla, Marcel Holyoak, Ronen Kadmon, David Saltz, and Peter E Smouse. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, 105(49):19052–19059, 2008.

Paul Newson and John Krumm. Hidden markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2009. doi: 10.1145/1653771.1653818.

Meghan O'Connell, moreiraMatias, and Wendy Kan. Ecml/pkdd 15: Taxi trajectory prediction (i). `https://kaggle.com/competitions/pkdd-15-predict-taxi-service-trajectory-i`, 2015. Kaggle.

Ruth Y Oliver, Scott W Yanco, Diego Ellis-Soto, Brett R Jesmer, Juliet Cohen, Song Gao, Robert Patchett, Tal Avgar, Keith Bildstein, Nicholas W Bakner, et al. Interacting effects of human presence and landscape modification on birds and mammals. *Science*, 392(6800):879–884, 2026.

Sasha Pekarsky, Ingo Schiffner, Yuri Markin, and Ran Nathan. Using movement ecology to evaluate the effectiveness of multiple human-wildlife conflict management practices. *Biological Conservation*, 262:109306, 2021.

Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, pp. 261–268. IEEE, 2009.

Jonathan R Potts and Luca Börger. How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. *Journal of Animal Ecology*, 92(1):16–29, 2023.

Jonathan R Potts, Luca Börger, Bronson K Strickland, and Garrett M Street. Assessing the predictive power of step selection functions: How social and environmental interactions affect animal space use. *Methods in Ecology and Evolution*, 13(8):1805–1818, 2022.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007.

Will Rogers, Scott Yanco, and Walter Jetz. Choices to landscapes: Mechanisms of animal movement scale to landscape patterns of space use. *Ecology Letters*, 28(12):e70279, 2025.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 464–468, 2018.

Liushuai Shi, Le Wang, Chengjiang Long, Sanping Zhou, Mo Zhou, Zhenxing Niu, and Gang Hua. Sgcn: Sparse graph convolution network for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8994–9003, 2021.

Liushuai Shi, Le Wang, Chengjiang Long, Sanping Zhou, Mo Zhou, Zhenxing Niu, and Gang Hua. Trajectory unified transformer for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 20660–20669, 2023.

J Signer, J Fieberg, and T Avgar. Estimating utilization distributions from fitted step-selection functions. ecosphere. 2017; 8: e01771, 1771.

Johannes Signer, John Fieberg, Björn Reineking, Ulrike Schlägel, B Smith, Niko Balkenhol, and Tal Avgar. Simulating animal space use from fitted integrated step-selection functions (issf). *Methods in Ecology and Evolution*, 15(1):43–50, 2024.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

Sivan Toledo, David Shohami, Ingo Schiffner, Emmanuel Lourie, Yotam Orchan, Yoav Bartan, and Ran Nathan. Cognitive map–based navigation in wild bats revealed by a new high-throughput tracking system. *Science*, 369(6500):188–193, 2020.

Marlee A Tucker, Aafke M Schipper, Tempe SF Adams, Nina Attias, Tal Avgar, Natarsha L Babic, Kristin J Barker, Guillaume Bastille-Rousseau, Dominik M Behr, Jerrold L Belant, et al. Behavioral responses of terrestrial mammals to covid-19 lockdowns. *Science*, 380(6649):1059–1064, 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Jiankai Wang, Meng Meng, Longkai Guo, Jiayi Xu, Lucio Dai, Yu Jiang, and Ming Yang. Ltp: Lane-based trajectory prediction for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Tonglong Wei, Yan Lin, Youfang Lin, Shengnan Guo, Jilin Hu, Haitao Yuan, Gao Cong, and Huaiyu Wan. Plmtrajrec: A scalable and generalizable trajectory recovery method with pre-trained language models. *arXiv preprint arXiv:2410.14281*, 2024.

Dylan M Westaway, Dale G Nimmo, Scott W Forrest, and Mitchell A Cowan. Navigating fragmentation: movement ecology and habitat selection of a large, long-lived skink in an agricultural landscape. *Landscape ecology*, 40(12):232, 2025. doi: 10.1007/s10980-025-02253-3.

Dhanushi A Wijeyakulasuriya, Elizabeth W Eisenhauer, Benjamin A Shaby, and Ephraim M Hanks. Machine learning for modeling animal movement. *PloS one*, 15(7):e0235750, 2020.

Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*, pp. 99–108, 2010.

Yu Zheng, Xing Xie, Wei-Ying Ma, et al. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33(2):32–39, 2010.

Yuanshao Zhu, James Jianqiao Yu, Xiangyu Zhao, Xun Zhou, Liang Han, Xuetao Wei, and Yuxuan Liang. Unitraj: Learning a universal trajectory foundation model from billion-scale worldwide traces. *arXiv preprint arXiv:2411.03859*, 2024.

A INPUT TOKENIZATION

A.1 SPATIOTEMPORAL CONDITIONING

Conditioning on spatiotemporal scale. We design STEPPY to be able to make forecasts for any timescale τ dynamically at prediction time. Different forecast horizons depend differently on past observations—for instance, short time-horizon forecasts may pay more attention to recent high-resolution movement, while longer-range forecasts may depend more on coarser-scale trends. We enable STEPPY to extract scale-specific representations by conditioning on two features representing the spatiotemporal scale of the queried forecast: **(1) the queried forecast horizon τ itself**; and **(2) a τ- and individual-specific spatial scale**, which is a summary statistic computed from the training data of the individual we are generating a forecast for. Specifically, we use the same summary statistic used by MOVEBENCH to determine the spatial support of its output maps,

$$d(\tau) = \min(2\,Q_{0.95}(D_\tau),\, Q_{0.99}(D_\tau))\,, \tag{2}$$

where Q_p denotes the empirical p-quantile and D_τ is the collection of valid observed displacement magnitudes (*i.e.* "step lengths") over horizon τ, pooled across training individuals in the study containing the animal of interest. This scale represents a reasonable estimate of the maximum distance the animal is likely to move in τ units of time.

We condition STEPPY on these temporal and spatial scales by prepending its input sequence with one token for each, with features $\log(\tau/1\text{ hour})$ and $\log(d(\tau)/1\text{ meter})$ respectively. Each is projected to the model's hidden dimension (768 by default) and summed with a learned token-type embedding.

A.2 CONDITIONING ON TAXONOMY.

We also design STEPPY to make forecasts for an animal of any species at prediction time, and to learn shared patterns across taxa where it is helpful to do so. To enable this behavior, for each input sequence, we encode the taxonomy of the animal of interest into a single token that is prepended alongside the sequence-level scale tokens described above. Because every individual in our dataset has the same kingdom (Animalia) and phylum (Chordata) we break down each individual's taxonomy into five ranks: class, order, family, genus, and species. We maintain a dictionary of all observed instances of each rank at training time which we use to tokenize each label. We learn a 128-dimensional token for each label at each rank, as well as a learned rank token that is shared across all examples. We then average these vectors over the known ranks and add a learned depth embedding indicating how many ranks are represented. Missing or unseen taxa truncate the lineage at the last known ancestor. We then project these features to the model's hidden dimension and sum them with a learned token-type embedding.

A.3 SEQUENCE TOKENIZATION.

In our design, each observation token contains features for the time of day and day of year that the observation occurred, providing context for learning daily and annual movement patterns. Each token encodes time of day in local mean solar time, h_i, and day of year, d_i, using circular features: $[\sin(2\pi h_i/24), \cos(2\pi h_i/24), \sin(2\pi d_i/325.25), \cos(2\pi d_i/325.25)]$. This representation respects the periodic nature of both variables and avoids artificial discontinuities at the boundaries of the daily and annual cycles. To tokenize, we project these features to the model's hidden dimension and sum them with a learned token-type embedding.

B HOW WE COMPUTE THE COMPARISONS FOR FIGURE 3

To compute STEPPY's improvements relative to MoveFormer over different step lengths and turn angles, we use MOVEBENCH's thinning procedure at each time horizon, resulting in sequences of observations that are approximately τ time units apart. A step length represents the distance between each subsequent pair of locations, and a turn angle represents the change in bearing between subsequent pairs of steps. Contiguous sequences of evenly-spaced observations are maintained where possible; if there is a gap between observations of longer than $3\tau/2$, a new sequence is created. Step lengths and turn angles are only defined within contiguous sequences. To compute the results

in Fig. 3, for every test case where STEPPY produces a prediction with a lower energy score, we record its incoming step length, outgoing step length, and turn direction with respect to its incoming bearing. We then sum the individual energy score improvements within each category to determine their contribution to the overall improvement over MoveFormer.