A framework for pre-processing individual location telemetry data for freshwater fish in a river section

Dominique Lamonica, Hilaire Drouineau, Hervé Capra, Hervé Pella, Anthony Maire

A framework for pre-processing individual location telemetry data for freshwater fish in a river section

Dominique Lamonica*[a,b], Hilaire Drouineau[c], Hervé Capra[a], Hervé Pella[a], Anthony Maire[d]

[a] *INRAE, RiverLy, HYNES (Irstea-EDF R&D), F-69625, Villeurbanne, France*
[b] *Institute of Landscape and Plant Ecology, University of Hohenheim, Stuttgart, Germany*
[c] *INRAE, EABX, HYNES (Irstea-EDF R&D), F-33612, Cestas, France*
[d] *EDF R&D, LNHE (Laboratoire National d'Hydraulique et Environnement), HYNES (Irstea - EDF R&D), 6 quai Watier, 78401 Chatou Cedex, France*

Abstract

Animal movement study often relies on individual tracking. The data scale (in time and space) varies according to the species, the environment where individuals live, or the exogenous processes that drive movement. To explore freshwater fish movement in rivers, fine-scale data are needed. Also, in rivers, recorded telemetry frequently shows missing data and location errors. The irregular time-steps, huge amount of data, environmental complexity (river section) and how fish move in such anisotropic environments undermine the use of statistical frameworks such as state-space models. To deal with these specificities, data pre-treatment can be required. We propose a generic method of telemetry data pre-processing, which can be transposed to other datasets. This framework includes interpolation to handle trajectories at fine time scales and performs data analysis within a state-space model. We combined analyses on observed and simulated data at various interpola-

*Corresponding author. Tel.: +33472208732.
Email address: dominique.lamonica@gmail.com (Anthony Maire)

tion time-steps to choose the one that best preserves the general movement while reducing the total amount of data required. First, we directly compared raw and interpolated data, and the results of parameter inference of a simple state-space model using the interpolated data. The state-space model infers behavioural state based on speed and turning angle between successive locations in animal trajectories. We also included two additional variables computed from raw data: a quantitative indicator of the correspondence between the interpolated trajectory and the raw data, and the variance of turning angles of raw data within the interpolation time-step. We were finally able to determine the most appropriate time-step to obtain locations that were regularly spaced in time and to reduce the amount of data while maintaining the precision of the raw data. Computational time was reduced 12-fold by using a 30-second time-step to interpolate data simulated at 3-second intervals. The inclusion of the two variables derived from raw data compensated for the loss of information in interpolated trajectories and allowed more efficient discrimination between behaviours.

Keywords: Animal location data, Movement model, State-space model, Switching behaviour, Bayesian inference, Parameter estimation

1. Introduction

Movement is a key issue in animal ecology and has been the focus of increasing research, especially in aquatic ecology, in both marine and freshwater environments (Giuggioli and Bartumeus 2010, Lennox et al. 2017, Nathan et al. 2008). Movement influences many processes, at individual, population and community levels: habitat selection (Block et al. 2011, Capra et al. 2017), migra-

tion (Bultel et al. 2014, Drouineau et al. 2017, Tétard et al. 2016, Tétard et al. 2019), trophic dynamics (Lima 2002), spread of disease (Carraro et al. 2017, Jonsen et al. 2001, Pinder et al. 2005) and adaptability to climate change and extreme events (Boucek et al. 2017). The study of animal movement often relies on individual tracking. Recent technical progress has revolutionized such studies: the development of high-frequency tags enables high-frequency data collection at fine spatial resolution (Cagnacci et al. 2010, Hussey et al. 2015, Lennox et al. 2017). Tracking data at high temporal resolution are invaluable for species that are rarely static although staying in the same area. Location per hour, half-day or day provides information on individual movement within an animal's living-range (e.g., within a region, or along a migration route) over a long period of time rather than small quick displacements within or between local habitats (Capra et al. 2018, Donaldson et al. 2014). In contrast, tracking several locations per minute allow displacements to be described at fine scale, enabling precise investigations of suitable migratory conditions, habitat selection or behaviour choices at individual level (e.g. (Capra et al. 2017, Cooke et al. 2004, Tétard et al. 2019)). Telemetry data at a very fine time scale (of a few seconds) are being increasingly collected worldwide (Cooke et al. 2013, Hussey et al. 2015, Lennox et al. 2017). Positions are generally estimated by triangulation, either by satellite or by multiple fixed receivers.

Statistical modelling frameworks have been developed to analyse aquatic telemetry data (Whoriskey et al. 2019) by assessing and correcting position errors (Bergé et al. 2012, Roy et al. 2014), dealing with missing data (Woillez et al. 2016), inferring behaviour (Dorazio and Price 2019, Thiebault et al. 2018,

Vermard et al. 2010), and investigating the influence of the environment on individual movements (Bestley et al. 2013, Drouineau et al. 2017, Patterson et al. 2009). State-space modelling is one of the useful existing statistical frameworks for animal movement analysis (Dorazio and Price 2019, Jonsen et al. 2003, Joo et al. 2013, Patterson et al. 2008).

Most animal movement studies have focused on large terrestrial or marine species, e.g. (Andersen et al. 2017, Bailey et al. 2008, Franke et al. 2006, Hedger et al. 2008, Mcclintock et al. 2012) which move in a wide and open environment, few have dealt with freshwater organisms. Telemetry data for anisotropic, irregular environments, such as large rivers, show some specificities (Cooke et al. 2013). Firstly, triangulation is made difficult by ground irregularities and the presence of vegetation, causing frequent signal loss. Secondly, the precision of triangulation varies in space (Bergé et al. 2012). And thirdly, the anisotropic closed conditions of hydrographic networks are specific limitations on animal movement (Quaglietta and Porto 2019, Sutherland et al. 2015) and complicate the trajectory analysis. Fish navigation is known to be influenced by physical cues such as current fields and physical obstacles (e.g., dams or river banks) (Goodwin et al. 2014) leading to highly orientated navigation which does not fulfil the isotropy assumption usually applied in open environments. The second major issue is the temporal and spatial scales at which fish movement is characterised. Telemetry studies of freshwater fish generally focus on seasonal movement, e.g. (Dorazio and Price 2019, Fraley et al. 2016, Koehn and Nicol 2016, Muhlfeld 2012), and modelling frameworks are rarely applied to analysing individual movements at fine scales. However, such fine-scale data and corresponding analytical methods are critical to study-

ing small-scale foraging movements of fishes within their home range, which (Dingle 1996) called "station keeping" movement, in contrast to migration over large distances.

The combination of environmental limitations, which entail irregularities in signal recording, and fine spatial and temporal scale has three main consequences. Firstly, irregular time-steps have to be handled. Secondly, analysing the huge amount of data generated requires great computer power, which may lead to a trade-off between reducing computing time and not degrading fine-scale data quality. And thirdly, the consideration of entire individual trajectories from the beginning to the end of the recording period is not always necessary; a certain number of trajectories spaced in time can also be used to explore individual behaviour.

To deal with irregular time-steps, an appropriately longer time-step can be chosen, coupled with interpolation, if necessary, to deal with any missing data. Increasing the time-step generally reduces the rate of missing data while also reducing the amount of data to be analysed. However, it impairs overall precision compared to the raw dataset, and may impact ecological interpretations based on these trajectories. In this paper, we propose a generic method of data processing to accurately infer individual behaviours from trajectories at fine temporal and spatial scales. Analysis was performed at different time-steps and compared so as to select the most appropriate one: *i.e.*, the one that preserved the general movement while efficiently discriminating between behaviours (here, slow and fast movements). To this end, we directly compared the raw and interpolated data, and compared the results of the parameters inference of a simple state-space model with the interpolated

data. We did not deal with location error, as our aim was to infer individual behaviour based on observed movement rather than reconstruct the exact individual trajectories. Moreover, usual correction methods are based on an isotropy assumption (*i.e.*, that there is no favoured direction within the space, and animal navigation consequently depends only on behaviour and not on environmental characteristics such as flow fields around physical obstacles), which is not fulfilled here.

To illustrate this approach, we used telemetry data for individual fish collected in the Rhône River at 3-second intervals (Capra et al. 2017). The initial objective of the study was to infer the relationships between fish behaviour and hydraulic conditions through analysis of fish movements. In this context, we used a state-space model to discriminate fish behaviours. For that, raw fish location data (hereafter referred to as "raw data") must be pre-processed. While maintaining the precision of the raw data, gaps between locations must be dealt with so as to obtain regularly time-spaced data.

2. Material and methods

2.1. Case study

Bergé et al. (Bergé et al. 2012) collected telemetry data on freshwater fish in the Rhône River using the HTI (https://www.innovasea.com/fish-tracking/) acoustic fixed telemetry system. Our system includes a set of pre-positioned hydrophones used to detect ultrasounds emitted by acoustic tags (frequency of 307 KHz). Tags signals that allow the identification of the tag, and precise positioning of the tag through a triangulation process

provided that the signal is detected by at least 3 hydrophones (hydrophones were all connected to a single controller and synchronized with UTC time to improve the triangulation). Further details are provided in Berg et al. (Bergé et al. 2012).

Locations of 94 individuals of various species were tracked for 3 months at a time interval of 3 seconds. The dataset suffered from the usual defects affecting tracking data. Firstly, individuals were not systematically located every 3 seconds during the 3 months of the experiment. There were two types of gap in the data: large gaps, in which the individual signal was lost for several minutes to several days, and small gaps where the individual signal was lost for a few 3-second periods (*i.e.*, 3 seconds to a few minutes). A second defect was specific to the triangulation process of the HTI system: some successive locations form artefactual "star-shaped" trajectories (Appendix A, Figure A1). Though this type of star-shaped pattern is specific to the triangulation process, it is more generally one of the types of location error besetting most tracking studies. The fine time-scale data may also incur a specific problem: when an individual is static or moving very slowly, its successive locations can be tracked as being as distant as when it is moving faster, due to triangulation error.

2.2. Proposed data pre-processing

Three pre-processing steps were performed on the raw data. The first was the choice of a time-step p. The second dealt with the large gaps between certain successive locations. For this, we considered using a threshold s equal to twice the time-step p, in order to split the trajectory in two while avoiding interpolation in-between points that were too far apart in time. If the dura-

tion between two successive locations $(x, y)_i$ and $(x, y)_{i+1}$ (x and y denoting the spatial coordinates and i the index of the location) is greater than s, then these two locations are assumed to belong to two distinct independent trajectories. The raw data for a given trajectory are denoted hereafter as the "support of the trajectory". The third step consisted in the linear interpolation of new locations within each trajectory according to the time-step p to deal with missing data within each trajectory. Interpolations were carried out using the *move* R package (Kranstauber and Smolla 2008).

We tested 6 interpolation time-steps (with the associated s threshold): 30, 60, 120, 180, 240, and 300 seconds, denoted by p_k, $k \in [1, 6]$. In order to test the different means of pre-processing the raw data, 16 of the 731 trajectories of individual European catfish (*Silurus glanis*) were selected. European catfish were chosen because this species included well-tracked individuals presenting numerous locations, and thus longer trajectories than for other species.

The complete trajectories can comprise three general patterns: (i) travelling (T), in which the individual moves over a long distance, where start and end locations are distant from one another; (ii) stationary (S), where locations are concentrated within a short perimeter (potentially "star-shaped") with start and end locations nearby; and (iii) a mixed pattern (B, for "bi-type") where a stationary move follows a long move or vice versa. Four trajectories of each for the stationary and travelling types, and 8 trajectories for the mixed type were selected for the test dataset. Two examples of each trajectory type at each time-step are shown in Figure 1.

2.3. Validation criteria

To compare the interpolated trajectories using the different time-steps p_k, several criteria were considered.

2.3.1. Computed variables

Two variables were computed to quantify the consistency between the raw data and each interpolated trajectory. They showed different patterns according to the type of trajectory, and were computed to see how they varied according to time-step. The two additional variables, denoted by U_1 and U_2, concerned speed and turning angles respectively. For a trajectory j associated with time-step p_k, let $(x, y)_{n_j \times p_k}$, $n_j \in [0, N_j]$ (with N_j the total number of locations in the trajectory j) be the n_j^{th} location in the trajectory. Considering the support of the trajectory j, let $i \in \left[0, I_{n_j}\right]$ be an index for raw data locations, denoted by $(x, y)_{i_{n_j}}$, between time $n_j \times p_k$ and $(n_j+1) \times p_k$ of the trajectory locations and $\theta_{i_{n_j}}$ the angle between the locations $(x, y)_{i_{n_j}}$, $(x, y)_{(i+1)_{n_j}}$ and $(x, y)_{(i+2)_{n_j}}$.

For a time-step p_k and a total time $n_j \times p_k$ of the trajectory j, $U_{1_{k,n_j}}$ and $U_{2_{k,n_j}}$, are defined as follows:

$$U_{1_{(k,n_j)}} = \frac{d\left((x, y)_{n_j \times p_k}, (x, y)_{(n_j+1) \times p_k}\right)}{\sum_{i=0}^{I_{n_j}-1} d\left((x, y)_{i_{n_j}}, (x, y)_{i_{n_j+1}}\right)} \tag{1}$$

where $d(X_1, X_2)$ represents the covered distance between locations X_1 and X_2. U_1 is thus the ratio between the covered distance according to the interpolated data and the covered distance according to the raw data, within the interpolation time-step. U_1 represents a quantitative indicator of the

correspondence between the interpolated trajectory and the raw data: the higher U_1, the better the fit.

$$U_{2_{(k,n_j)}} = var(\theta_{i_{n_j}}) \tag{2}$$

where $i \in [0, I_n]$ and $var(X)$ denotes the variance of X. U_2 is thus the variance of turning angles of raw data within the interpolation time-step. If the variance is small, the trajectory is rectilinear within the interpolated time-step and little information is lost. However, a large variance may indicate a star-shaped pattern, especially if associated with a low U_1.

2.3.2. State-space modelling

We developed a state-space model based on (Morales et al. 2004) to discriminate between the different individual behaviours in the 16 selected trajectories. Model parameters were estimated independently for each time-step, so as to assess variations in parameter estimates according to the time-step. Results for each time-step where also compared with and without adding U_1 and U_2, to see whether behaviour discrimination was improved by including these two variables.

Model definition. Model states correspond to the succession of fish behaviours at each time increment, with two possible behaviours: "Resting" (denoted by R), which corresponds to slow or erratic movements, and "Moving" (denoted by M), which corresponds to fast, oriented movements. We assumed constant behaviour switching probabilities between successive time increments. The observation model links the state at time t to corresponding movement variables (*i.e.*, speed between two locations and turning angles between two

moves) (Morales et al. 2004). Low mean speed and high turning angles variance are taken to characterise "Resting" behaviour, whereas high speed and mean turning angles around 0° are taken to characterise "Moving" behaviour. "Resting" behaviour is expected to predominate in "Stationary" trajectories, "Moving" behaviour in "Travelling" trajectories. The model is written as follows:

Transition matrix

$$M_q = \begin{pmatrix} q_{R\to R} & 1 - q_{R\to R} \\ 1 - q_{M\to M} & q_{M\to M} \end{pmatrix} \tag{3}$$

State equation

$$z_t \sim \mathcal{B}(M_q[z_{t-1}]) \tag{4}$$

Observation model

$$\begin{aligned} y_{v_t} &\sim \mathcal{G}(a[z_t], \lambda[z_t]) \\ y_{\phi_t} &\sim \mathcal{WC}(b[z_t], \rho[z_t]) \end{aligned} \tag{5}$$

with $q_{R\to R}$ and $q_{M\to M}$ being the probability of maintaining resting, or moving, behaviour following resting, or moving, behaviour, and z_t being the behaviour at time t (R or M). Concerning the data, y_{v_t} is the observed speed between $t-1$ and t, and y_{ϕ_t} is the observed turning angle between $t-2$ and $t-1$ and $t-1$ and t. Concerning the parameters, $a[z_t]$ and $\lambda[z_t]$ describe the speed for behaviour z at time t and $b[z_t]$, and $\rho[z_t]$ describes the turning angle for behaviour z at time t. $\mathcal{B}$, $\mathcal{G}$ and $\mathcal{WC}$ represent the Bernoulli distribution, the Gamma distribution and the Wrapped Cauchy distribution, respectively. Given that U_1 and U_2 allow integration of information derived from the raw data independently of the interpolation, two models were tested to determine

whether taking into account of U_1 and U_2 improved the model's behaviour discrimination. The first (hereafter, "Model 1") included only variables from interpolated data (observed speed and observed turning angles). In the second (hereafter, "Model 2"), U_1 and U_2 were added to the variables, the observation model becoming as follows:

$$\begin{aligned} y_{v_t} &\sim \mathcal{G}(a[z_t], \lambda[z_t]) \\ y_{\phi_t} &\sim \mathcal{WC}(b[z_t], \rho[z_t]) \\ U_{1_t} &\sim \mathcal{B}eta(\alpha[z_t], \beta[z_t]) \\ U_{2_t} &\sim \mathcal{N}(\mu[z_t], \sigma[z_t]) \end{aligned} \tag{6}$$

with $\alpha[z_t]$ and $\beta[z_t]$ being the parameters describing U_1 for behaviour z at time t, $\mu[z_t]$ and $\sigma[z_t]$ being the parameters describing U_2 for behaviour z at time t. $\mathcal{B}eta$ and $\mathcal{N}$ represent for the Beta distribution and the normal distribution, respectively.

Computation. Bayesian inference was used to fit the model to the data: *i.e.*, the trajectories obtained for each time-step p_k. A single model was fitted for all selected trajectories at once. Prior distributions were defined summarising all available information on each parameter (Appendix B, Table B1). Markov Chain Monte Carlo (MCMC) computations were performed using JAGS software and the *rjags* R package (Plummer 2009, R Core team 2018). A total of 10,000 iterations were performed as a burn-in phase, and inference was based on 20,000 additional iterations for each of the three independent chains (with different initiations). The Gelman and Rubin tests (Gelman and Rubin 1992) were used to check the convergence of the estimation process. The computation times for the different time-steps p_k were compared. The estimated behaviour at each time-step was recorded from the

MCMC iterations. For each location, the credibility of resting and moving behaviours was calculated, as the mean of the behaviours estimated for all the MCMC iterations. We also calculated the mean duration of maintaining each behaviour (*i.e.*, the expected value of the geometric distribution with switching probability, multiplied by the time-step), with the medians and 95% credibility intervals of the posterior distributions for both switching probabilities.

Twelve fits of the models were performed for all 16 selected trajectories: one with Model 1 and one with Model 2, for each of the 6 tested time-steps. We also performed 2 additional fits using all the individual trajectories: one with Model 1 and one with Model 2, for the time-step which appeared to be the most appropriate after checking the various criteria. As convergence might not be reached for a simple model with two behaviours (Morales et al. 2004), we also tested similar models (without and with U_1 and U_2) including a third intermediate behaviour.

Simulated data analysis. A Markov chain of 10,000 behaviours was simulated with fixed transition probabilities ($q_{R\to R} = 0.99$ and $q_{M\to M} = 0.97$) using Model 1. Each point in the chain stands for a theoretical location of the individual and, for each point, a speed and a relative angle was derived from the distributions of speeds and relative angles of each behaviour (with the following fixed parameter values for speed and turning angles distributions: $a_1 = 0.1$, $a_2 = 1.5$, $\lambda_1 = 10.5$, $\lambda_2 = 5$, $b_1 = 3.15$, $b_2 = 0$, $\rho_1 = 0.3$ and $\rho_2 = 0.9$). Then, for each chain point, a theoretical location was calculated. Thus, we obtained a series of 10,000 locations associated with a behaviour. The theoretical time-step of this trajectory was 3 seconds. The

pre-processing described in the Section 2.2 was applied: for the 6 tested time-steps, the simulated trajectory was interpolated, speeds and turning angles were calculated, as well as the two additional variables U_1 and U_2. Then, we estimated the parameters of Model 1 (with speeds and turning angles only) and Model 2 (with speeds, turning angles and the two additional variables U_1 and U_2) for each time-step and for the raw trajectory.

Computation time was recorded and, when the model converged, the goodness of fit between the simulated and inferred behaviours was computed. For the 3-second trajectory (raw simulated data), inferred and simulated behaviours were compared directly. For the interpolated trajectories, first we computed the mean of the simulated behaviours between two locations within the time-step, then the root mean squared error for all the behaviours of the trajectory were calculated, comparing for each location the mean simulated behaviour and the inferred one. To compare results between time-steps, the root mean squared error was divided by the number of locations in the interpolated trajectory.

3. Results

3.1. Influence of the time-step between interpolated locations on behaviour inference

3.1.1. General correspondence of interpolated trajectories and raw data

Superimposition of raw data and interpolated trajectories is (for convenience) shown for 6 typical trajectories out of the 16 studied trajectories in Figure 1. For bi-type 1 and the two stationary trajectories, the interpolated trajectories did not show any major differences in overall movement between

time-steps. However, the duration of the trajectories increased with the time-step, which was expected because of the increasing threshold separating two distinct trajectories, whereas the number of locations decreased. On the contrary, for bi-type 2 and the two travelling trajectories with the most changes in direction, the interpolated trajectories did not show any major differences for time-steps below 60 seconds, but diverged from the raw data for the four longer time-steps (120, 180, 240 and 300 seconds).

3.1.2. Additional variables

The ratio between covered distance from interpolated and raw data, namely U_1, showed a similar pattern for all types of trajectory and for all time-steps (Figure 2a). As expected, U_1 decreased with increasing time-step, due to shorter covered distance in interpolated data. In the 5 time-steps longer than 30 seconds, the difference in U_1 between the three types of trajectories was greater than in the first time-step (30 seconds). Likewise, the variance of turning angles within the interpolation time-step, namely U_2, showed a pattern similar to U_1 (Figure 2b).

3.1.3. Modelling results

Parameter estimation. The MCMC algorithm consistently converged according to Gelman and Rubin diagnostics for each simulation performed with the 16 selected trajectories. For all parameters, posterior distributions were narrower than prior distributions. The narrowness of the posterior distributions suggests that sufficient information was available from the data to accurately estimate the model parameters. Overall, there were progressive differences between parameter estimates with increasing time-step (Figure 3). However,

the posterior distributions of the first two time-steps (30 and 60 seconds) differed strongly from those of the other four time-steps, all of which were quite similar.

The mean of speed distributions of resting behaviour a_R decreased with increasing time-step, because of decreasing estimated travel distance in the interpolated trajectory. Mean turning angle distributions b_R and b_M were constant for all time-steps. The modes of U_1 (the ratio between the covered distance from the interpolated and raw data, within the interpolation time-step) distributions α_R and α_M decreased with increasing time-step, because the longer the time-step, the lower the fit between interpolated trajectory and raw data. Mean U_2 for moving behaviour μ_M increased with increasing time-step, where the turning angles were more acute (*e.g.*, travelling trajectories in Figure 1). Mean U_2 for the resting behaviour μ_R was similar for all time-steps.

Concerning the transition probability of maintaining resting behaviour $q_{R\to R}$, posterior distributions were similar for the four longer time-steps (120 to 300 seconds), with a median value around 0.975. For the second tested time-step (60 seconds), the median of the posterior distribution was lower (around 0.95), and much lower for the shortest time-step of 30 seconds (around 0.88). Concerning the transition probability of maintaining moving behaviour $q_{M\to M}$, posterior distributions were similar for the two longest time-steps of 240 and 300 seconds, with a median value around 0.90, and for the 30, 120 and 180 second time-steps, with a median value around 0.87. For the 60 second time-step, the median value of the posterior distribution was higher (around 0.93).

Inferred behaviours. Behaviours were highly discriminated for all tested time-steps, although discrimination was a bit lower for the shortest time-step, of 30 seconds (Table 1). Using Model 2, inferred behaviours were similar for all time-steps for the two travelling trajectories and the second stationary trajectory (Figure 4). Stationary trajectory 1 showed a higher proportion of resting behaviour for the four longest time-steps (120 to 300 seconds) (Figure 4); for the two shortest time-steps, some movements were associated with moving behaviour, or else were not discriminated (Figure 5). Similarly, the stationary phase at the end of the bi-type 1 trajectory was more associated with moving behaviour (or else not discriminated) for the two shortest time-steps (Figure 5). Concerning the bi-type 2 trajectory, a higher proportion of resting behaviour was inferred for the four longest time-steps (Figure 4), due to the increased number of locations in the stationary phase at the end of the trajectory (Figure 5).

Simulated data analysis. For the three longest time-steps (180, 240 and 360 seconds), the models (Model 1 and Model 2) failed to converge and discriminate between behaviours. For raw data, 99% of the inferred behaviours matched the simulated ones (RMSE of 0.01) (Appendix D Table D1). For the three shortest tested time-steps (30, 60 and 120 seconds), RMSE increased with time-step. For the two shortest time-steps (30 and 60 seconds), RMSE was lower on Model 2 than on Model 1. Part of the trajectory, with true and inferred behaviours for the three shortest tested time-steps (30, 60 and 120 seconds), is shown in Appendix D, Figure D1.

3.2. Behaviour discrimination with the state-space model

3.2.1. Interest of the two additional variables

For stationary trajectories ("S"), U_1 was lower than for travelling trajectories ("T"), and showed average values for bi-type trajectories ("B") (Figure 2 (a)). This was due to the higher rate of star-shaped trajectories when the individual was stationary: in star-shaped trajectories, distance covered was shorter on interpolated than raw data, due to the triangulation artefact. U_2 was lower for travelling than stationary trajectories and intermediate for bi-type trajectories (Figure 2 (b)). The differences shown by U_1 and U_2 between the different types of trajectory provide information for discriminating "resting" versus "moving" behaviour related to "stationary" and "travelling" trajectory types, respectively. Behaviours were slightly better discriminated (*i.e.*, were in most cases inferred as being either resting or moving), with the addition of U_1 and U_2 (Table 1).

Finally, we fitted the model on all the 731 available trajectories, except for those comprising fewer than 5 locations. At first, neither of the two models (Model 1 and Model 2) was able to converge. We therefore added a third behaviour in the models. This third behaviour was an intermediate between the resting and moving behaviours, with mean turning angle close to that of the resting behaviour, and mean speed between the mean speeds of the other two behaviours. With the third behaviour, Model 2 converged whereas Model 1 still failed to converge.

3.2.2. Fit between inferred behaviours and observed trajectories

Overall, with Model 2, inferred behaviours were consistent with observed trajectories: moving behaviours were mostly inferred in travelling trajecto-

ries, and resting behaviours in stationary trajectories (Figure 4). For the bi-type 2 trajectory, the travelling (starting) and stationary (ending) phases mostly corresponded to moving and resting behaviours, respectively, as expected. Some movements in stationary phases were, however, associated with moving behaviour, or with resting behaviour with low certainty, for the two shortest time-steps. Increasing time-step increased mean duration for each behaviour (Table 2), and also increased the difference in duration between the behaviours: the longer the time-step, the longer the resting duration, compared to the moving duration.

4. Discussion

4.1. Influence of time-step on behaviour discrimination and computation time

A wide range of durations between two consecutive locations (*i.e.*, time-steps) are used in animal movement studies: for example from 60 seconds (eels, (Bassett and Montgomery 2011)) to several minutes (caribou, 15 minutes (Andersen et al. 2017)), one hour (sea lions, (Breed et al. 2012)), or several hours (turtles, 6 hours (Bailey et al. 2008), wolves, 12 hours (Franke et al. 2006)). Time-steps should be adapted to the species, its travel mode and the question being addressed. For resident fish moving in small areas, time-steps shorter than several minutes are necessary to explore the different behaviours. We observed progressive but not major differences between variables (*i.e.*, the movement variables mean speed and turning angles (Appendix A, Figure A2 (a) and (b)) and the additional variables U_1 and U_2) used to evaluate the tested interpolation time-steps. As of the second tested time-step

(60 seconds), behaviours were better discriminated. However, with time-steps longer than 120 seconds, interpolation generated trajectories that were very remote from the raw data, considerably degrading location information. This can cause problems: for instance, when the individual often switches behaviour or when spatial data such as environmental parameters are to be included in the analysis. Likewise, the differences in mean duration per behaviour according to time-step showed that switching phases may be overlooked if a long time-step is used and, consequently, information on species ecology can be lost due to unreliable interpolation. Furthermore, model fit to the data (speed and turning angles; data not shown) was better for the two shortest time-steps. This is supported by the analysis of simulated data, which gave satisfactory behaviour predictions for the two shortest time-steps (30 and 60 seconds), poorer prediction for 120 seconds, and no discrimination between behaviours for the three longest time-steps (180, 240, and 360 seconds).

Overall, computation time (Appendix B, Table B2) was reasonable. There were no large differences between the time-steps from 60 to 240 seconds. However, computation time was almost twice as long for the shortest (30 seconds) as for the longest (300 seconds) time-step. For simulated data, computation time was almost 12-fold longer for raw data (3 seconds) than for the shortest time-step (30 seconds) and 2.5-fold longer for the 30-second than the 60-second time-step (Appendix D, Table D2). Such a difference can be critical when analysing all trajectories of all individuals together, even though data generated by interpolation made the model inference feasible, in terms of computation time, for all tested time-steps.

4.2. Model results and behaviour discrimination

The results concerning the movement descriptors (*i.e.*, speed and turning angle distributions) in resting behaviour should be regarded with caution. Most artefactual trajectories that are "star-shaped" are stationary phases which mainly include resting behaviour. Nevertheless, interpolation with all tested time-steps reduced that artefact, enabling discrimination between moving and resting behaviours. The high estimated transition probabilities $q_{R\to R}$ and $q_{M\to M}$ (above 0.80) imply that behaviours presented long duration, and are clearly influenced by our choices of "exemplary" trajectories displaying contrasted movement phases.

In order to check whether behaviours were adequately discriminated according to the tested time-steps, the selected trajectories had to be composed of phases that clearly represent the two main behaviours. As a consequence, some other patterns of movement were firstly discarded. When all trajectories were included (even excluding those with fewer than 5 locations), the simple model with two behaviours did not converge, partly due to the presence of these other movement patterns. In addition, we assumed a constant transition matrix, which is too simplistic since fish behaviour is influenced by environmental factors. This might also prevent the model from converging. Further validation with all data is needed, but a simple model tested on a set of sample trajectories is still useful to determine the pre-processing to be performed on the data.

Additional variables. Focusing on the 16 selected trajectories, including the two variables U_1 and U_2 in the model improved discrimination of behaviours. The simulated data analysis showed that including the additional variables

reduced prediction error for the two shortest time-steps (30 and 60 seconds). Thus, it appears highly profitable to combine these kinds of variable, calculated from raw data, with a short time-step to compensate the deficit in behaviour discrimination. Furthermore, using all trajectories (except for those with fewer than 5 locations) confirmed the benefit of including the additional variables U_1 and U_2 for a given interpolation time-step (here, 60 seconds). Other additional variables could have been considered such as specific indicators related to the known behaviours of the studied species.

4.3. Implications of the interpolation and trajectory cutting processes for behaviour discrimination

Tools from earlier studies of state-space models (*e.g.*, (Johnson et al. 2008, Jonsen et al. 2003, Vermard et al. 2010)) attempted to deal with irregular time-steps, location errors or the reconstruction of entire trajectories. Such tools could have been appropriate to process the present data, but the prime issue was computing time, and increasing the time-step solved this while also dealing with most of the numerous small gaps in location. Combined with the use of the two additional variables U_1 and U_2, the proposed processing reduced computing time, dealt with irregularly time-spaced locations and preserved the information provided by the initial 3-second- time-step of the raw data. For the present study, several smaller unconnected trajectories provided enough information, and we did not seek to determine individual locations and behaviours when the signal had been lost for a long period of time (several hours). To achieve convergence with the 3-behaviour model, the shortest trajectories (fewers than 5 points) were deleted from the dataset. This necessity might be due to a bias in estimating transition probabilities

for these very short trajectories. An initial behaviour (namely z_{init}, see Appendix C, Model code) was mandatory to initiate the Markov chain of behaviours across time: z_{init} was derived from a categorical distribution with the 3 equal probabilities. When short trajectories are numerous, the estimated transition probability from the initial behaviour to the first behaviour of the trajectory has a major weight, although it is only an "artefact" of the modelling procedure.

As short trajectories may bias the estimation of transition probabilities, a more suitable dataset would favour long trajectories (*i.e.*, with the maximum number of points). For this, two possibilities emerged. Firstly, long time-steps could be used, so that the raw dataset is less divided: trajectories are longer in time, and also likely made up of numerous points. However, trajectories shorter in time than twice the chosen time-step were deleted due to interpolation, because at least two movements are needed to calculate speed and turning angle. These trajectories would be preserved with a smaller time-step. Secondly, a short time-step could be preferred, in which case trajectories are denser in points, but the cutting process of the raw data to obtain trajectories excludes more points than with a long time-step, because the s threshold is lower.

The cutting process and the chosen s threshold are thus appear key points in data pre-processing. The present cutting method has two main disadvantages. Firstly, the different time-steps were difficult to compare, as the number of locations differed between the tested time-steps. Secondly, with a small threshold, trajectories are liable to be small, which could lead to overlooking some switches in behaviour. One solution could be to use a higher

threshold, chosen on the basis of the histogram of trajectory durations, combined with a small time-step. But this solution would lead to interpolating data on time ranges for which no observed locations are available, which is not desirable. An auto-correlogram of covered distances in the raw data could also give indications on where to cut the raw data to obtain trajectories. Missing locations could also be considered as missing data, instead of interpolating them. In the present case, this was not possible because using a Wrapped Cauchy distribution for turning angles within JAGS required using observed turning angles as input rather than observed variable (see Annex C Model code and BUGS trick in (Morales et al. 2004), Supplementary Information).

4.4. Outline

The present study developed a generic method of data pre-processing to handle trajectories at fine time scales and infer behaviours based on telemetry data, which could be transposed to other datasets. Data pre-processing is an essential step in trajectory analysis, although rarely highlighted. Increasing time-steps allowed efficient discrimination between behaviours, with locations regularly spaced in time and a smaller amount of data to process. At the same time, the additional variables computed from the raw data compensated for the loss of information in interpolated trajectories resulting from the increased time-step. The time-step should be adapted according to the ecology and habitat preference of the studied species. The raw data cutting process should be explored to optimise trajectory length while preserving small time-steps and correspondence between trajectories and raw data.
We demonstrated the possibility of discriminating behaviours for the whole

dataset of a given individual, using a state-space model. This opens up interesting perspectives. Individual variation in movement or behavioural parameters could be quantified using the trajectories of several individuals, by hierarchical modelling (Jonsen 2016) using a 3-behaviour state-space model including additional variables computed from the raw data. The reduction in computation time is then particularly valuable when all trajectories of several individuals are included. Datasets for several individuals of various species have more sources of variability. Hierarchical modelling is then necessary to deal with individual and species variability. Environmental variability requires explicit modelling of the link between environmental variables, such as hydraulic or thermal parameters, and the transition matrix. Considering that i) behavioural state and habitat selection are linked, and ii) movement and trajectory changes are behavioural state proxy, movement analysis is a way of investigating dynamical selection of habitat (short term, below several minutes) in a highly contrasted and variable (below the hour) environment. Therefore, dynamically favourable habitats (according to dynamic hydraulic conditions) should be mapped, providing quantitative information to evaluate the impact of events such as thermal discharge or dam functioning.

5. Acknowledgements

We thank the Agence de l'Eau Rhône-Méditerranée-Corse, Electricité de France (EDF-DTG and EDF R&D), the European Union/FEDER and the Aquitaine Region for their financial support. This study is part of a partnership research program in hydrobiology (HYNES) between EDF R&D and IRSTEA.

6. Figures titles and captions

6.1. Figure 1

Title. Six trajectories and raw data superimposed, for all the tested time-steps.

Caption. Gray triangles are the raw data and black dots are the interpolated locations (linked by black lines) for each time-step (*i.e.*, the trajectories). These trajectories were chosen because they are characteristic of a stationary behaviour (Stationary 1 and 2), of an active travelling behaviour (Travelling 1 and 2), or of an alternation between these two behaviours (Bi-type 1 and 2).

6.2. Figure 2

Title. Distributions of the additional variables U_1 (a) and U_2 (b), for all the tested time-steps.

Caption. U_1 is the ratio between covered distances according to interpolated data and covered distances according to raw data. U_2 is the variance of turning angles of raw data within the interpolation time-step.

6.3. Figure 3

Title. Posterior distributions of the parameters, for all the tested time-steps, from Model 2.

6.4. Figure 4

Title. Percentages of each inferred behaviour for six trajectories, for all the tested time-steps (computed from Model 2).

Caption. Resting and Moving correspond to locations where resting and moving behaviour respectively was inferred with probability > 0.8 (*i.e.*, for 80% of the MCMC iterations the estimated behaviour was resting/moving). Hesitating corresponds to locations where behaviours were inferred with probability < 0.8.

6.5. Figure 5

Title. Six trajectories with inferred behaviour, for all the tested time-steps (computed from Model 2).

Caption. These trajectories were chosen because they are characteristic of a stationary behaviour (Stationary 1 and 2), of an active travelling behaviour (Travelling 1 and 2), or of an alternation between these two behaviours (Bi-type 1 and 2).

7. Appendix A: Figures

7.1. Figure A1

Title. Example of a star-shaped trajectory.

7.2. Figure A2 (a) and (b)

Title. Distributions of movement variables (speed and turning angles), for all the tested time-steps.

7.3. Figure A3

Title. The 16 selected trajectories with inferred behaviour (red: moving, blue: resting), for the 60-second time-step (computed from Model 2).

8. Appendix B: Table of model parameters

9. Appendix C: JAGS code for Model 2

10. Appendix D: Simulated data analysis

10.1. Tables 1 and 2, Appendix D

10.2. Figure D1

Title. Portion of the simulated trajectory with true (triangles) and predicted (circles) behaviours for the three shortest tested time-steps (30, 60 and 120 seconds, from top to bottom), with Model 1 (left side) and Model 2 (right side). Red stands for moving behaviour, and blue for resting.

11. References

[Andersen et al. 2017] K. H. Andersen, A. Nielsen, U. H. Thygesen, H. H. Hinrichsen, and S. Neuenfeldt. Using the particle filter to geolocate Atlantic cod (Gadus morhua) in the Baltic Sea, with special emphasis on determining uncertainty. *Canadian Journal of Fisheries and Aquatic Sciences*, 64(4):618–627, 2007.

[Bailey et al. 2008] H. Bailey, G. Shillinger, D. Palacios, S. Bograd, J. Spotila, F. Paladino, and B. Block. Identifying and comparing phases of movement by leatherback turtles using state-space models. *Journal of Experimental Marine Biology and Ecology*, 356(1-2):128–135, 2008.

[Bassett and Montgomery 2011] D. Bassett and J. Montgomery. Home range use and movement patterns of the yellow moray eel Gymnothorax prasinus. *Journal of Fish Biology*, 79(2):520–525, 2011.

[Bergé et al. 2012] J. Bergé, H. Capra, H. Pella, T. Steig, M. Ovidio, E. Bultel, and N. Lamouroux. Probability of detection and positioning error of a hydro acoustic telemetry system in a fast-flowing river: Intrinsic and environmental determinants. *Fisheries Research*, 125-126:1–13, 2012.

[Bestley et al. 2013] S. Bestley, I. D. Jonsen, M. A. Hindell, C. Guinet, and J. B. Charrassin. Integrative modelling of animal movement: incorporating in situ habitat and behavioural information for a migratory marine predator. *Proceedings of The Royal Society B: Biological sciences*, 280(1750):2012–2262, 2013.

[Block et al. 2011] B. A. Block, I. D. Jonsen, S. J. Jorgensen, a. J. Winship, S. a. Shaffer, S. J. Bograd, E. L. Hazen, D. G. Foley, G. a. Breed, A.-L. Harrison, J. E. Ganong, A. Swithenbank, M. Castleton, H. Dewar, B. R. Mate, G. L. Shillinger, K. M. Schaefer, S. R. Benson, M. J. Weise, R. W. Henry, and D. P. Costa. Tracking apex marine predator movements in a dynamic ocean. *Nature*, 475(7354):86–90, 2011.

[Boucek et al. 2017] R. E. Boucek, M. R. Heithaus, R. Santos, P. Stevens, and J. S. Rehage. Can animal habitat use patterns influence their vulnerability to extreme climate events? An estuarine sportfish case study. *Global Change Biology*, (April):1–13, 2017.

[Breed et al. 2012] G. A. Breed, D. P. Costa, I. D. Jonsen, P. W. Robinson, and J. Mills-Flemming. State-space methods for more completely capturing behavioral dynamics from animal tracks. *Ecological Modelling*, 235-236:49–58, 2012.

[Bultel et al. 2014] E. Bultel, E. Lasne, A. Acou, J. Guillaudeau, C. Bertier, and E. Feunteun. Migration behaviour of silver eels (Anguilla anguilla) in a large estuary of Western Europe inferred from acoustic telemetry. *Estuarine, Coastal and Shelf Science*, 137(1):23–31, 2014.

[Cagnacci et al. 2010] F. Cagnacci, L. Boitani, R. A. Powell, and M. S. Boyce. Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1550):2157–62, 2010.

[Capra et al. 2017] H. Capra, L. Plichard, J. Bergé, H. Pella, M. Ovidio, E. McNeil, and N. Lamouroux. Fish habitat selection in a large hydropeaking river: Strong individual and temporal variations revealed by telemetry. *Science of the Total Environment*, 578:109–120, 2017.

[Capra et al. 2018] H. Capra H., H. Pella, and M. Ovidio. Individual movements, home ranges and habitat use by native rheophilic cyprinids and non-native catfish in a large regulated rive. *Fisheries Management and Ecology*, 25:136–149, 2018.

[Carraro et al. 2017] L. Carraro, L. Mari, M. Gatto, A. Rinaldo, and E. Bertuzzo. Spread of proliferative kidney disease in fish along stream networks: A spatial metacommunity framework. *Freshwater Biology*, (March), 2017.

[Cooke et al. 2004] S. J. Cooke, C. M. Bunt, and J. F. Schreer. Understanding fish behavior, distribution, and survival in thermal effluents using

667 fixed telemetry arrays: A case study of smallmouth bass in a discharge
668 canal during winter. *Environmental Management*, 33:140–150, 2004.

669 [Cooke et al. 2013] S. J. Cooke, J. D. Midwood, J. D. Thiem, P. Klimley,
670 M. C. Lucas, E. B. Thorstad, J. Eiler, C. Holbrook, and B. C. Ebner.
671 Tracking animals in freshwater with electronic tags: past, present and
672 future. *Animal Biotelemetry*, 1(5):1–19, 2013.

673 [Dingle 1996] H. Dingle. *Migration: the Biology of Life on the Move*. New
674 York: Oxford University Press.

675 [Donaldson et al. 2014] M. R. Donaldson, S. G. Hinch, C. D. Suski, A. T.
676 Fisk, M. R. Heupel, and S. J. Cooke. Making connections in aquatic
677 ecosystems with acoustic telemetry monitoring. *Frontiers in Ecology*
678 *and the Environment*, 12: 565–573, 2014.

679 [Dorazio and Price 2019] R. M. Dorazio and M. Price. State-space models
680 to infer movements and behavior of fish detected in a spatial array of
681 acoustic receivers. *Canadian Journal of Fisheries and Aquatic Sciences*,
682 76(4):543–550, 2019.

683 [Drouineau et al. 2017] H. Drouineau, F. Bau, A. Alric, N. Deligne, P.
684 Gomes, and P. Sagnes. Silver eel downstream migration in fragmented
685 rivers: use of a Bayesian model to track movements triggering and du-
686 ration. *Aquatic Living Resources*, 30:1–9, 2017.

687 [Fraley et al. 2016] K. M. Fraley, J. A. Falke, R. Y., and S. Ivey. Seasonal
688 Movements and Habitat Use of Potamodromous Rainbow Trout Across

689 a Complex Alaska Riverscape. *Transactions of the American Fisheries*
690 *Society*, 145(5):1077–1092, 2016.

691 [Franke et al. 2006] A. Franke, T. Caelli, G. Kuzyk, and R. J. Hudson. Pre-
692 diction of wolf (Canis lupus) kill-sites using hidden Markov models. *Eco-*
693 *logical Modelling*, 197(1-2):237–246, 2006.

694 [Gelman and Rubin 1992] A. Gelman and D. B. Rubin. Inference from iter-
695 ative simulation using multiple sequences. *Statistical science*, 7(4):457–
696 511, 1992.

697 [Giuggioli and Bartumeus 2010] L. Giuggioli and F. Bartumeus. Ani-
698 mal movement, search strategies and behavioural ecology: a cross-
699 disciplinary way forward. *Journal of Animal Ecology*, 79: 906–909, 2010.

700 [Goodwin et al. 2014] R. A. Goodwin, M. Politano, J. W. Garvin, J. M.
701 Nestler, D. Hay, J. J. Anderson, L. J. Weber, E. Dimperio, D. L. Smith,
702 and M. Timko. Fish navigation of large dams emerges from their mod-
703 ulation of flow field experience. *Proceedings of the National Academy of*
704 *Sciences of the United States of America*, 111(14):5277–82, 2014.

705 [Hedger et al. 2008] R. D. Hedger, F. Martin, J. J. Dodson, D. Hatin, F.
706 Caron, and F. G. Whoriskey. The optimized interpolation of fish posi-
707 tions and speeds in an array of fixed acoustic receivers. *ICES Journal*
708 *of Marine Science*, 65(7):1248–1259, 2008.

709 [Hussey et al. 2015] N.E. Hussey, S. T. Kessel, K. Aarestrup, S. J. Cooke,
710 P. D. Cowley, A. T. Fisk, R. G. Harcourt, K. N. Holland, S. J. Iverson,
711 J. F. Kocik, J. E. M. Flemming, and F. G. Whoriskey. Aquatic animal

telemetry: A panoramic window into the underwater world. *Science*, 348(6240):1255642, 2015.

[Johnson et al. 2008] D. S. Johnson, J. M. London, M. A. L., and J. W. Durban. Continuous-Time Correlated Random Walk Model for Animal Telemetry Data. *Ecology*, 89(5):1208–1215, 2008.

[Jonsen et al. 2001] I. D. Jonsen, R. S. Bourchier, and J. Roland. The influence of matrix habitat on Aphthona flea beetle immigration to leafy spurge patches. *Oecologia*, 127(2):287–294, 2001.

[Jonsen et al. 2003] I. D. Jonsen, R. A. Myers, and J. M. Flemming. Meta-analysis of animal movement using state-space models. *Ecology*, 84(11):3055–3063, 2003.

[Jonsen 2016] I. D. Jonsen. Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Sci Rep*, 6:20625, 2016.

[Joo et al. 2013] R. Joo, S. Bertrand, J. Tam, and R. Fablet. Hidden Markov Models: The Best Models for Forager Movements? *PLoS ONE*, 8(8), 2013.

[Koehn and Nicol 2016] J. D. Koehn and S. J. Nicol. Comparative movements of four large fish species in a lowland river. *Journal of Fish Biology*, 88(4):1350–1368, 2016.

[Kranstauber and Smolla 2008] B. Kranstauber and M. Smolla. Visualizing and Analyzing Animal Track Data. Rpackage version 2.1.0, 2008.

[Lima 2002] S. L. Lima. Putting predators back into behavioral predator-prey interactions. *Trends in Ecology & Evolution*, 17(2):70–75, 2002.

[Lennox et al. 2017] R. J. Lennox, K. Aarestrup, S. J. Cooke, P. D. Cowley, Z. D. Deng, A. T. Fisk, R. G. Harcourt, M. Heupel, S. G. Hinch, K. N. Holland, N. E. Hussey, S. J. Iverson, S. T. Kessel, J. F. Kocik, M. C. Lucas, J. M. Flemming, V. M. Nguyen, M. J.W. Stokesbury, S. Vagle, D. L. VanderZwaag, F. G. Whoriskey, and N. Young. Envisioning the Future of Aquatic Animal Tracking: Technology, Science, and Application. *BioScience*, 67:884–896, 2017.

[Mcclintock et al. 2012] B. T. Mcclintock, R. King, J. Matthiopoulos, B. J. Mcconnell, and J. M. Morales. A General Modeling Framework for Animal Movement in Discrete Time Using Multi-State Random Walks. *Ecological Monographs*, 82(3):335–3349, 2012.

[Morales et al. 2004] J. M. Morales, D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell. Extracting More Out of Relocation Data : Building Movement Models As Mixtures of Random Walks. *Ecology*, 85(9):2436–2445, 2004.

[Muhlfeld 2012] C. C. Muhlfeld, J. J. Giersch, and B. Marotz. Seasonal movements of non-native lake trout in a connected lake and river system. *Fisheries Management and Ecology*, 19(3):224–232, 2012.

[Nathan et al. 2008] R. Nathan, W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse. A movement ecology paradigm

for unifying organismal movement research. *Pnas*, 105(49):19052–19059, 2008.

[Patterson et al. 2008] T. A. Patterson, L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. State-space models of individual animal movement. *Trends in Ecology and Evolution*, 23(2):87–94, 2008.

[Patterson et al. 2009] T. A. Patterson, M. Basson, M. V. Bravington, and J. S. Gunn. Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology*, 78(6):1113–1123, 2009.

[Pinder et al. 2005] A. C. Pinder, R. E. Gozlan, and J. R. Britton. Dispersal of the invasive topmouth gudgeon, Pseudorasbora parva in the UK: A vector for an emergent infectious disease. *Fisheries Management and Ecology*, 12(6):411–414, 2005.

[Plummer 2009] M. Plummer. rjags: Bayesian graphical models using mcmc. Rpackage version 1.0.3-12, 2009.

[Quaglietta and Porto 2019] L. Quaglietta and M. Porto. SiMRiv: An R package for mechanistic simulation of individual, spatially-explicit multistate movements in rivers, heterogeneous and homogeneous spaces incorporating landscape bias. *Movement Ecology*, 7(1):1–9, 2019.

[R Core team 2018] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[Roy et al. 2014] R. Roy, J. Beguin, C. Argillier, L. Tissot, F. Smith, S. Smedbol S. and E. De Oliveira. Testing the VEMCO Positioning System: spatial distribution of the probability of location and the positioning error in a reservoir. *Animal Biotelemetry*, 2:1, 2014.

[Sutherland et al. 2015] C. Sutherland, A. K. Fuller, and J. A. Royle. Modelling non-Euclidean movement and landscape connectivity in highly structured ecological networks. *Methods in Ecology and Evolution*, 6(2):169–177, 2015.

[Tétard et al. 2016] S. Tétard, E. Feunteun, E. Bultel, R. Gadais, M. L. Bégout, T. Trancart, and E. Lasne. Poor oxic conditions in a large estuary reduce connectivity from marine to freshwater habitats of a diadromous fish. *Estuarine, Coastal and Shelf Science*, 169:216–226, 2016.

[Tétard et al. 2019] S. Tétard, A. Maire, M. Lemaire, E. De Oliveira, P. Martin, and D. Courret. Behaviour of Atlantic salmon smolts approaching a bypass under light and dark conditions: importance of fish development. *Ecological Engineering*, 131:39–52, 2019.

[Thiebault et al. 2018] A. Thiebault, L. Dubroca, R. Mullers, Y. Tremblay, and P. Pistorius. "m2b" package in R: deriving multiple variables from movement data to predict behavioural states with random forests.. *Methods in Ecology and Evolution*, 9:1548–1555, 2018.

[Vermard et al. 2010] Y. Vermard, E. Rivot, S. Mahévas, P. Marchal, and D. Gascuel. Identifying fishing trip behaviour and estimating fishing

effort from VMS data using Bayesian Hidden Markov Models. *Ecological Modelling*, 221(15):1757–1769, 2010.

[Whoriskey et al. 2019] K. Whoriskey, E. G. Martins, M. Auger-Méthé, L. F.G. Gutowsky, R. J. Lennox, S. J. Cooke, M. Power, and J. Mills Flemming. Current and emerging statistical techniques for aquatic telemetry data: A guide to analysing spatially discrete animal detections. *Methods in Ecology and Evolution*, (February):1–14, 2019.

[Woillez et al. 2016] M. Woillez, R. Fablet, T.-T. Ngo, M. Lalire, P. Lazure, and H. de Pontual. A HMM-based model to geolocate pelagic fish from high-resolution individual temperature and depth histories: European sea bass as a case study. *Ecological Modelling*, 321:10–22, 2016.

Table 1: Percentage of locations where moving or resting behaviour was inferred with probability > 0.8.

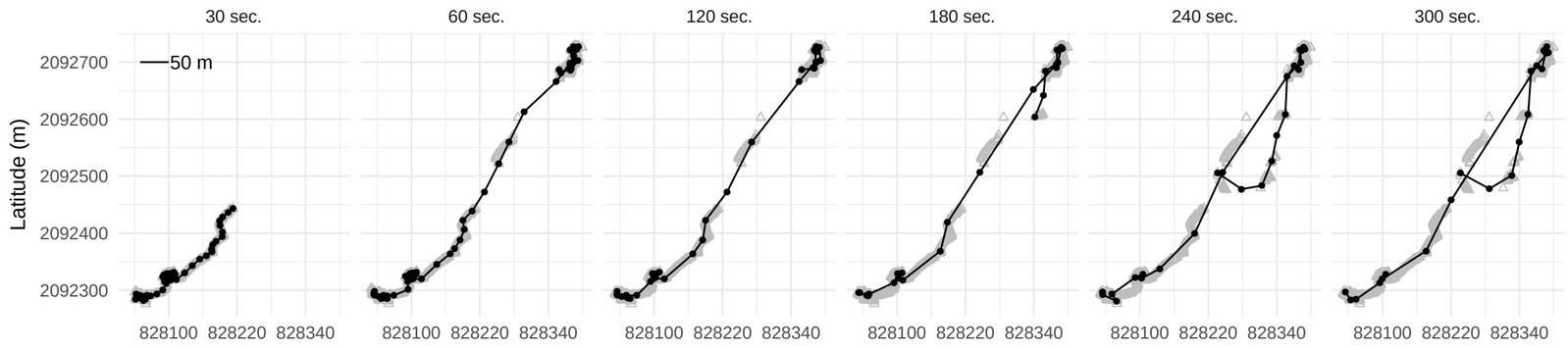
Time-step (sec.)	Model 1, without U_1 and U_2 (%)	Model 2, with U_1 and U_2 (%)
30	83	91
60	92	94
120	96	98
180	92	99
240	92	99
300	91	99

Table 2: Mean duration (minutes) in resting or moving behaviour according to the tested time-steps (computed from Model 2).

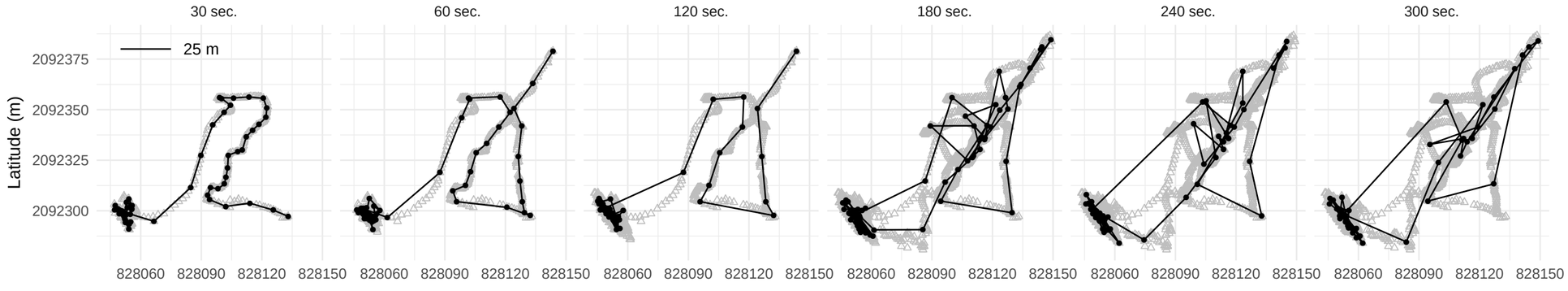
Time-step (sec.)	Resting	Moving
30	4.1 $[3.2, 5.5]$	3.7 $[2.9, 4.9]$
60	21 $[14, 33]$	12 $[8.8, 18]$
120	65 $[43, 106]$	15 $[11, 22]$
180	118 $[74, 208]$	24 $[17, 36]$
240	229 $[125, 499]$	38 $[25, 61]$
300	234 $[129, 486]$	45 $[30, 73]$

The mean duration (median and 95% credibility interval) in resting, or moving, behaviour was computed as the expected value (*i.e.*, the number of independent trials to get the first success) from the geometric distribution of probability equal to $1 - q_{R \to R}$, or $1 - q_{M \to M}$, multiplied by the time-step.

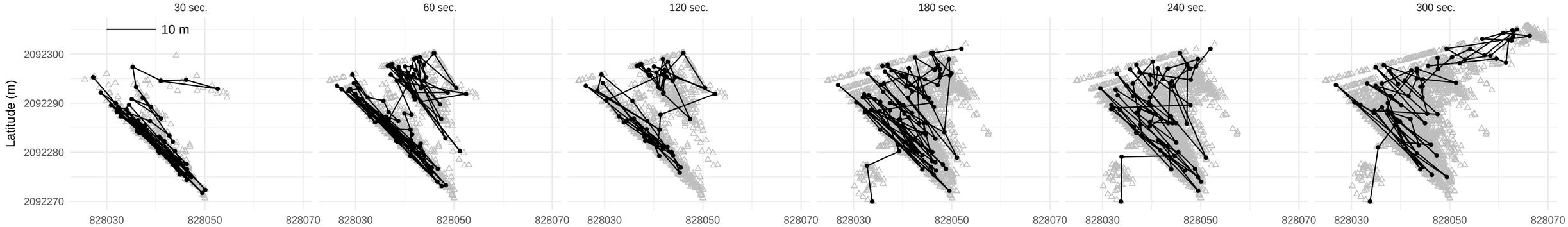
Bi−type 1



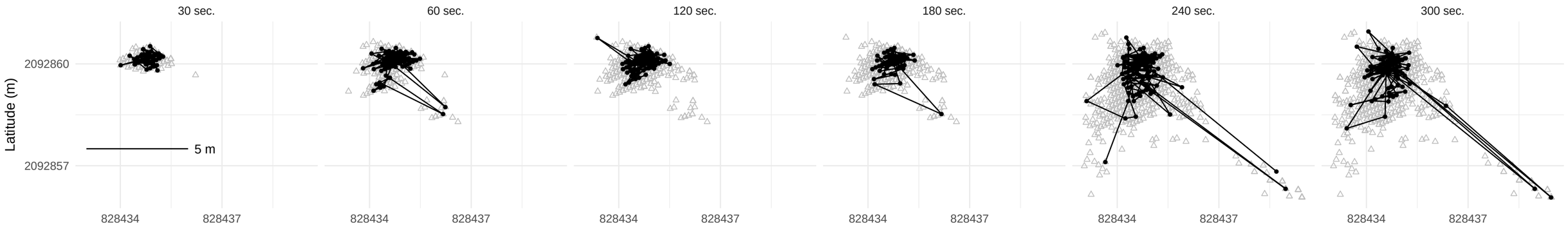
Bi−type 2



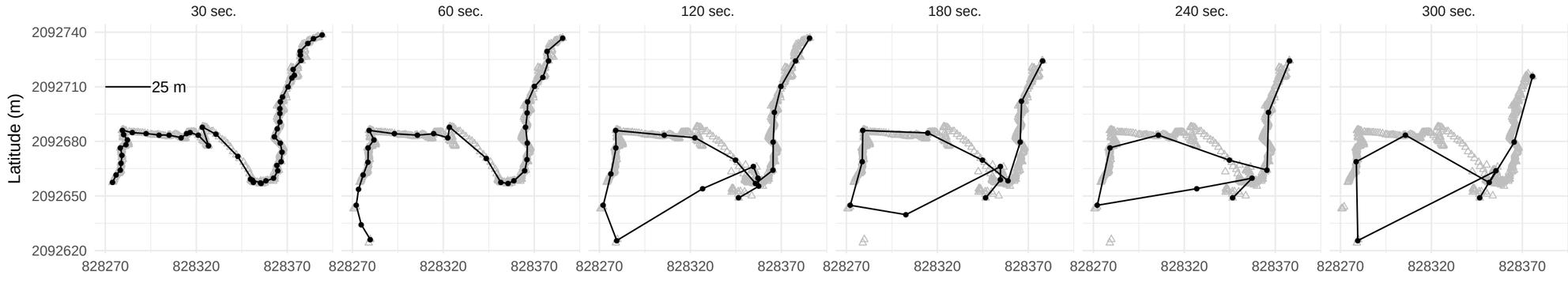
Stationary 1



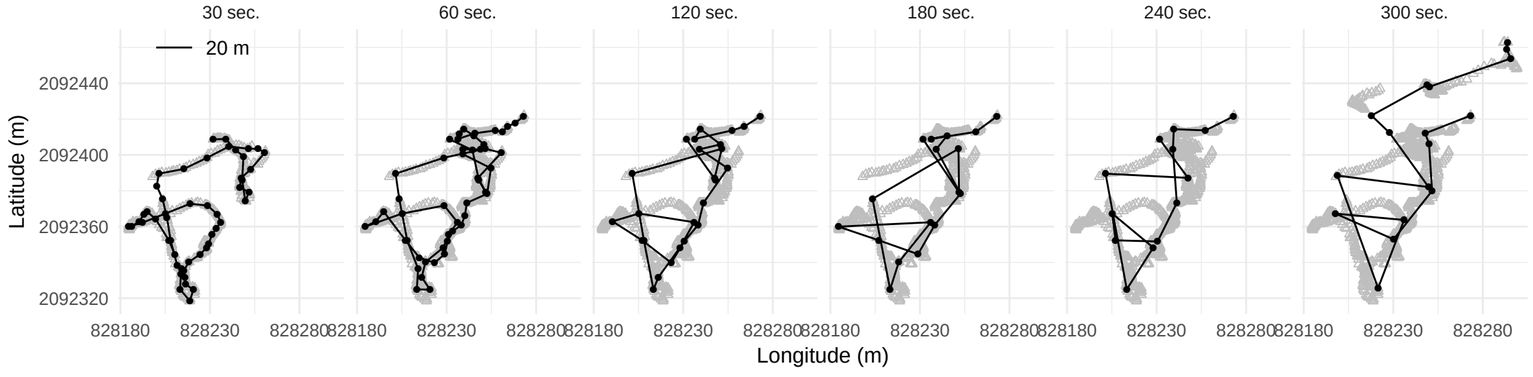
Stationary 2

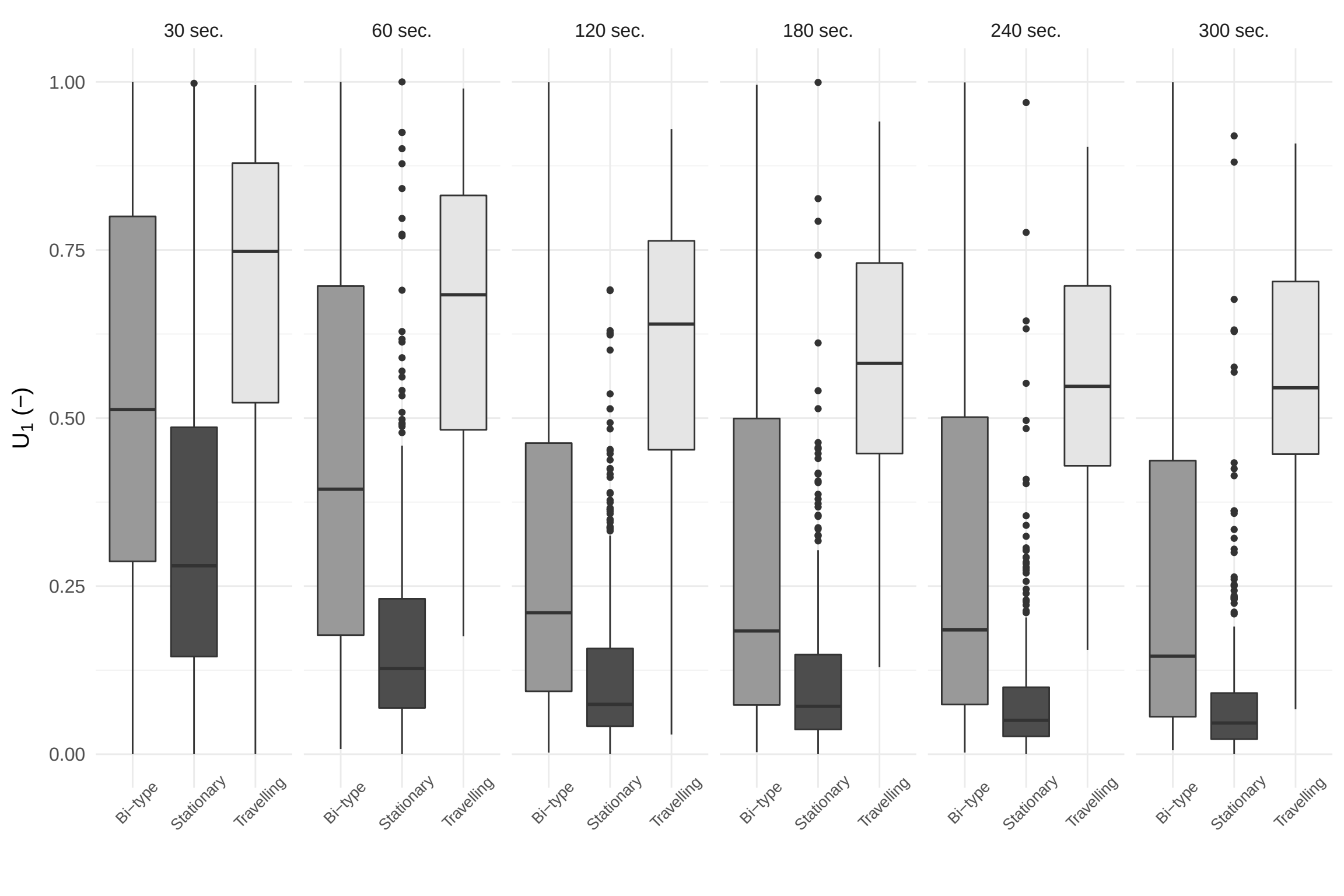


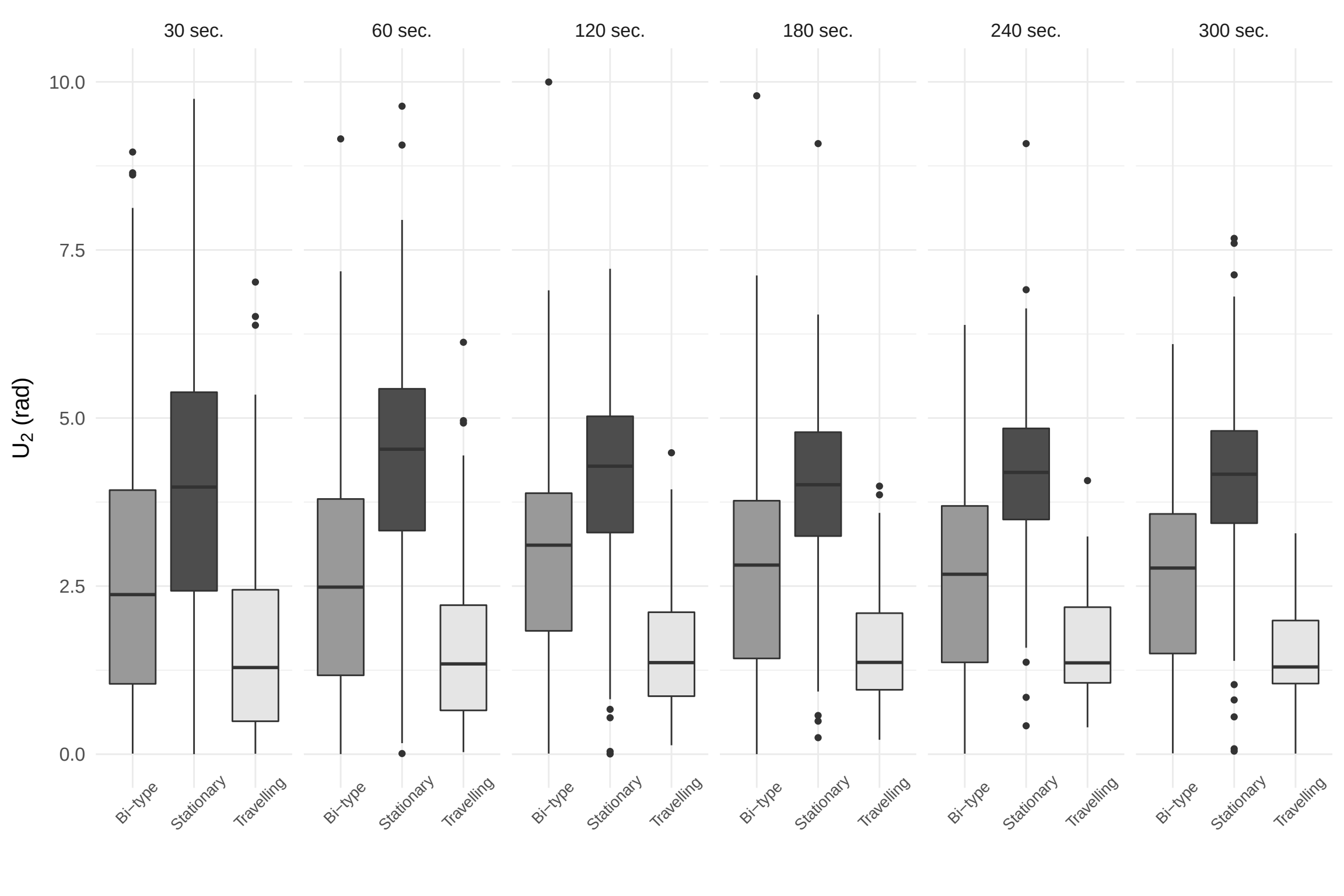
Travelling 1

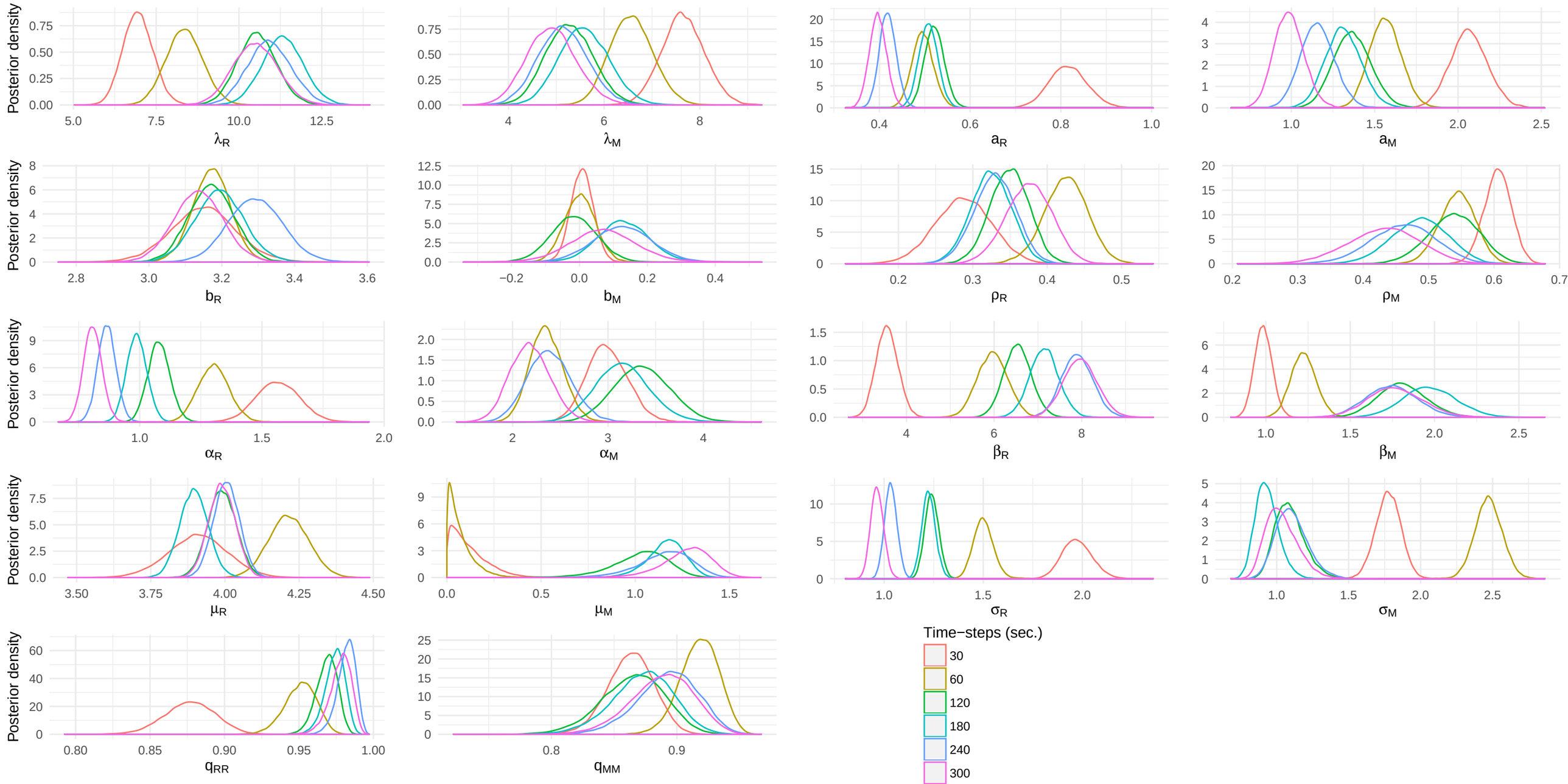


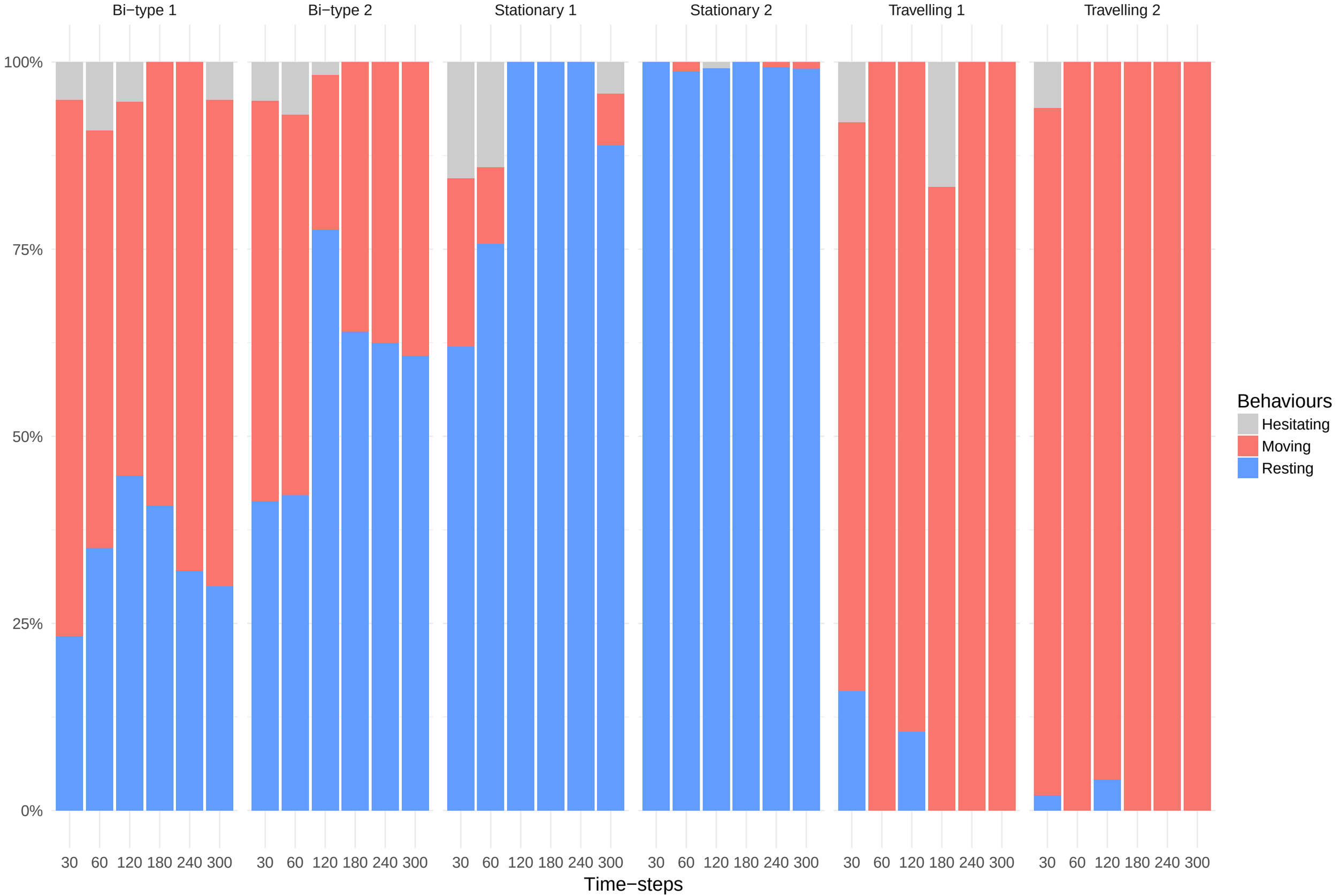
Travelling 2



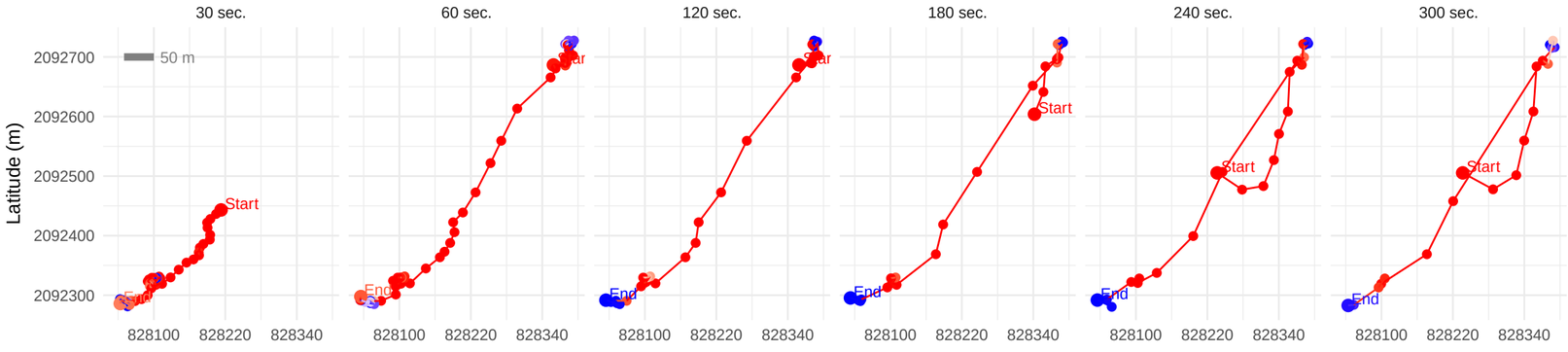

30 sec.
60 sec.
120 sec.
180 sec.
240 sec.
300 sec.
1.00
0.75
0.50
0.25
0.00
U_1 (–)
Bi−type
Stationary
Travelling


30 sec.
60 sec.
120 sec.
180 sec.
240 sec.
300 sec.
10.0
7.5
5.0
2.5
0.0
U_2 (rad)
Bi−type
Stationary
Travelling

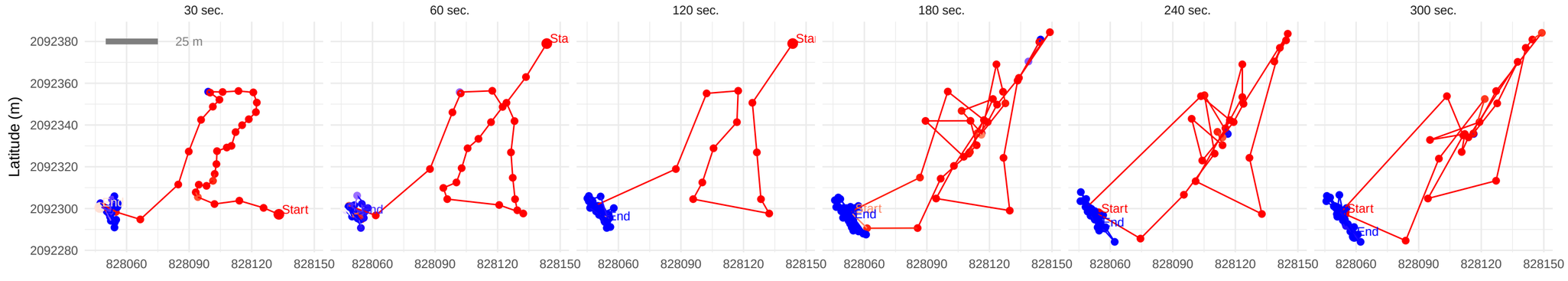

Posterior density
λR
λM
aR
aM
bR
bM
ρR
ρM
αR
αM
βR
βM
μR
μM
σR
σM
qRR
qMM
Time−steps (sec.)
30
60
120
180
240
300


Bi−type 1
Bi−type 2
Stationary 1
Stationary 2
Travelling 1
Travelling 2
100%
75%
50%
25%
0%
30
60
120
180
240
300
Time−steps
Behaviours
Hesitating
Moving
Resting

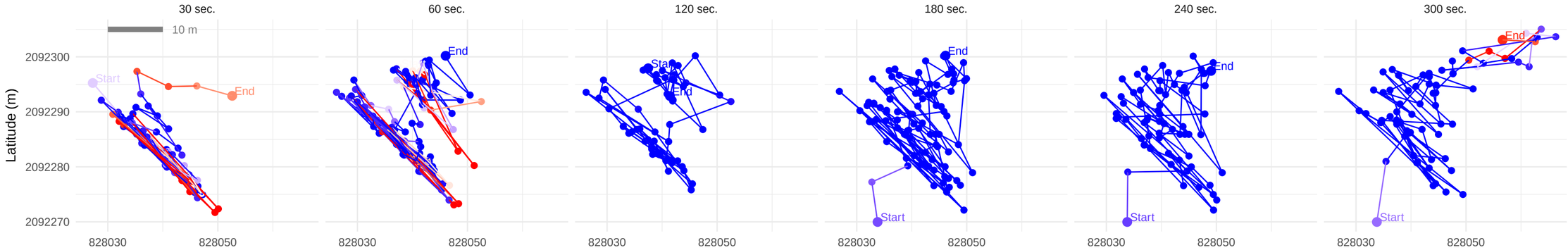
Bi−type 1



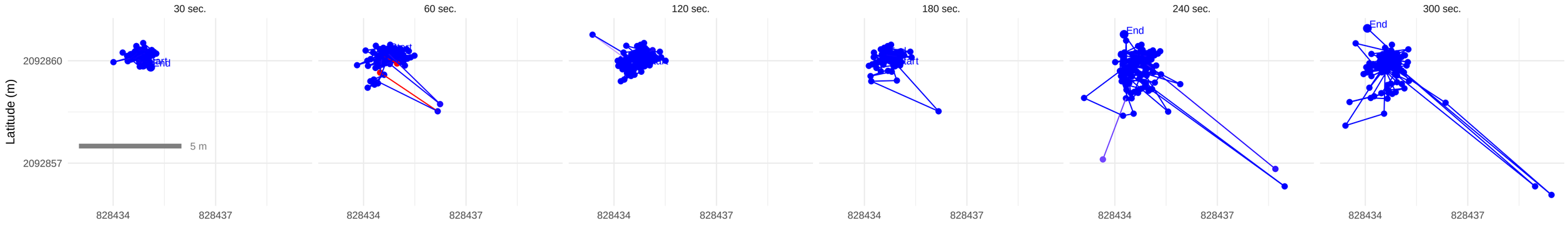
Bi−type 2



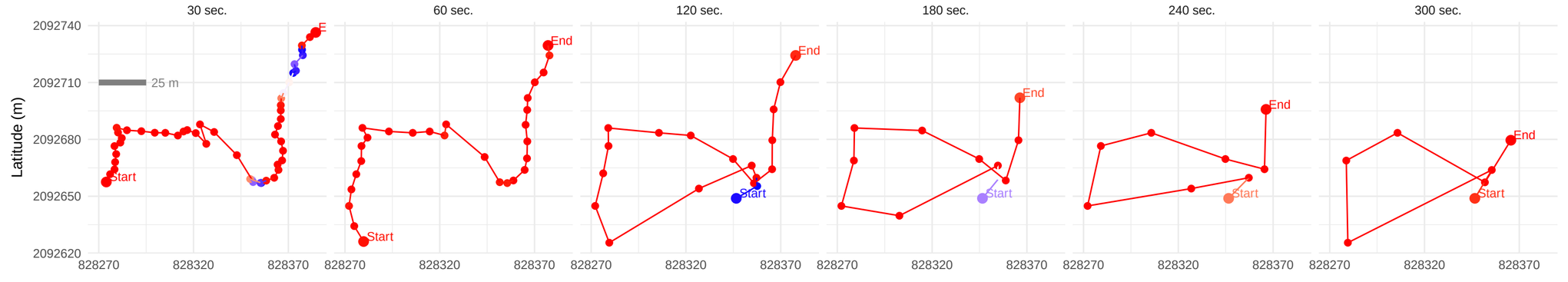
Stationary 1



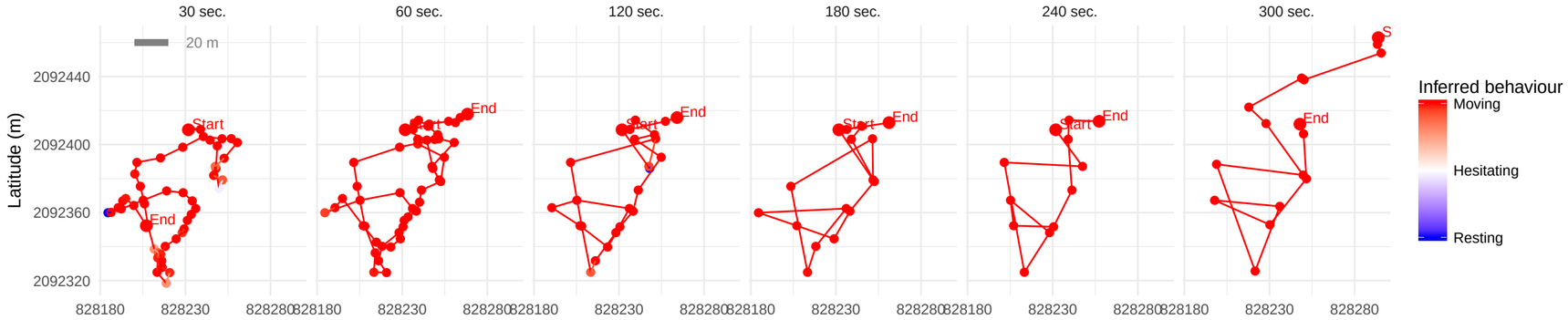
Stationary 2

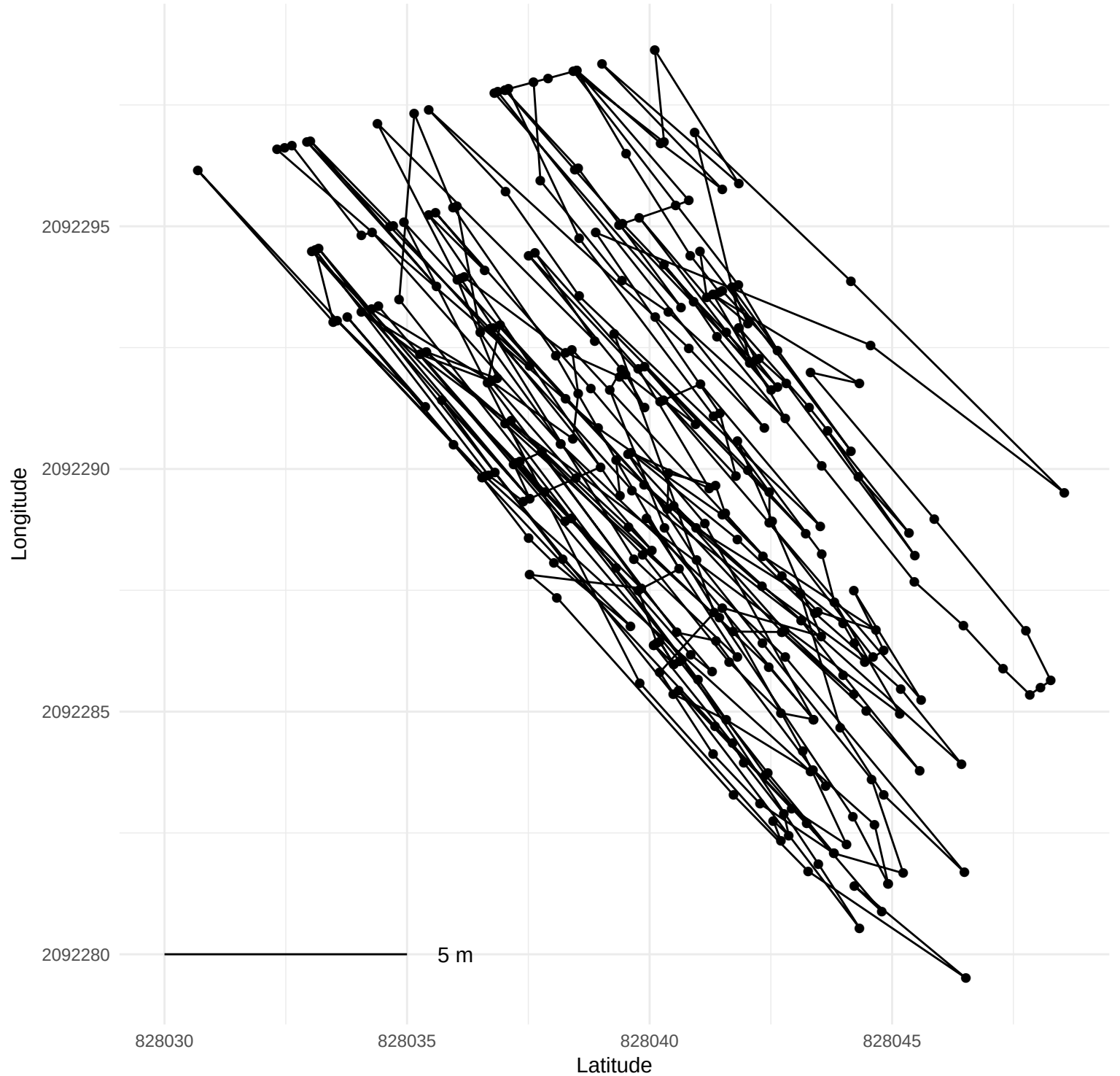


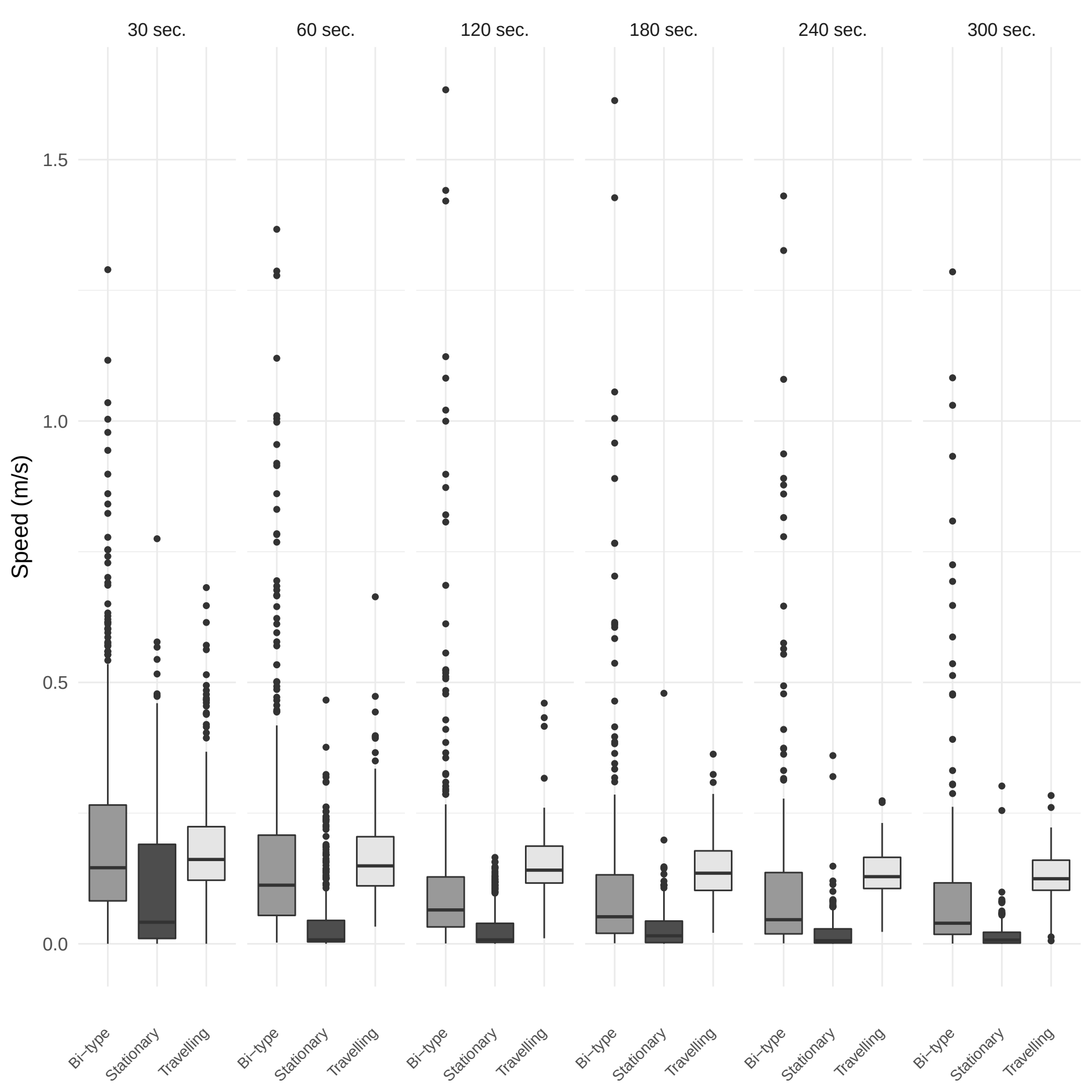
Travelling 1



Travelling 2




2092295
2092290
2092285
2092280
Longitude
5 m
828030
828035
828040
828045
Latitude


30 sec.
60 sec.
120 sec.
180 sec.
240 sec.
300 sec.
Speed (m/s)
1.5
1.0
0.5
0.0
Bi−type
Stationary
Travelling

30 sec.

60 sec.

120 sec.

180 sec.

240 sec.

300 sec.

Bi−type

Stationary

Travelling

330 − 30

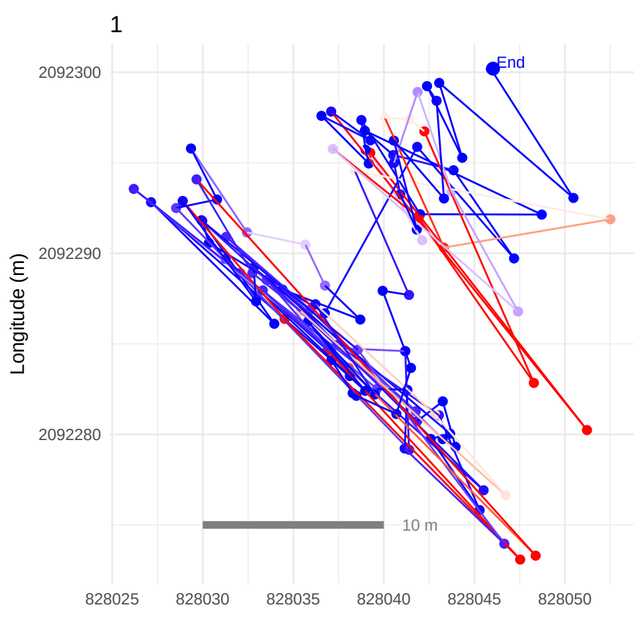
30 − 90

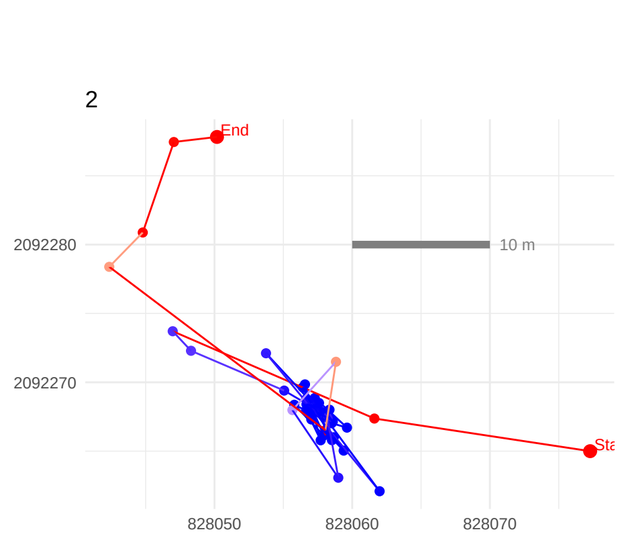
90 − 150

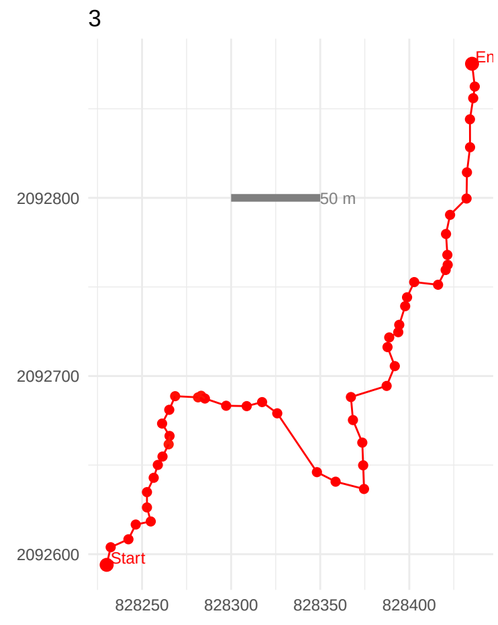
150 − 210

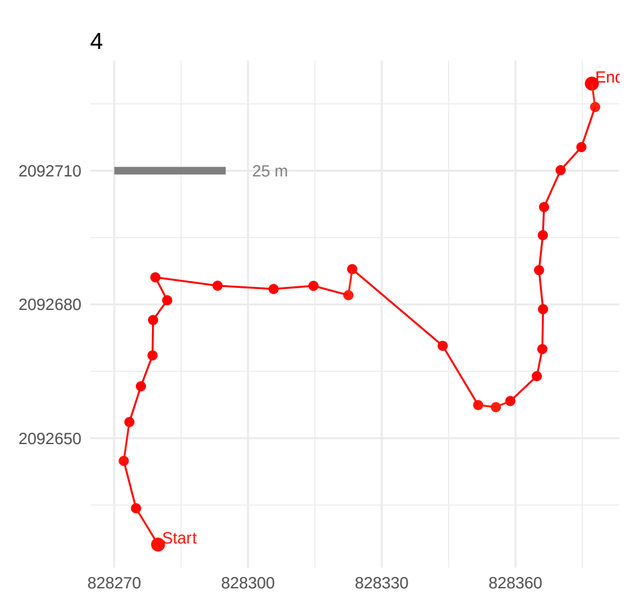
10 − 270

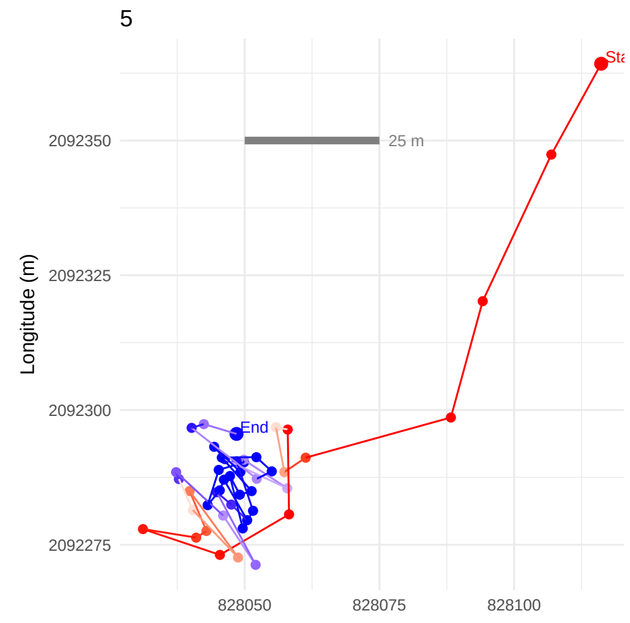
70 − 330

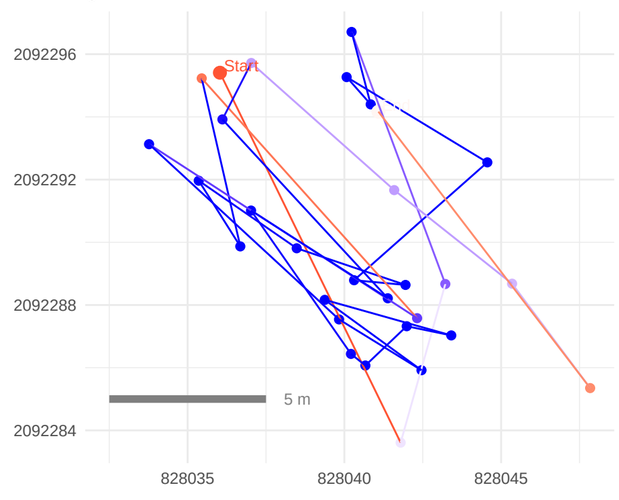

1
End
10 m
Longitude (m)
2092300
2092290
2092280
828025
828030
828035
828040
828045
828050

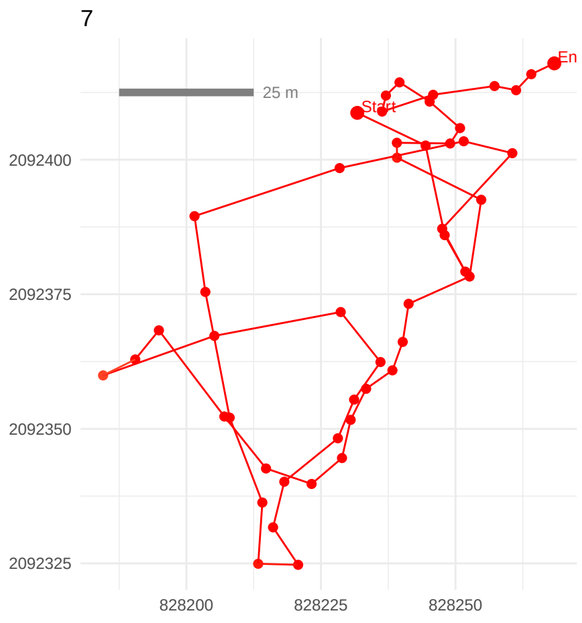

2
End
10 m
2092280
2092270
828050
828060
828070

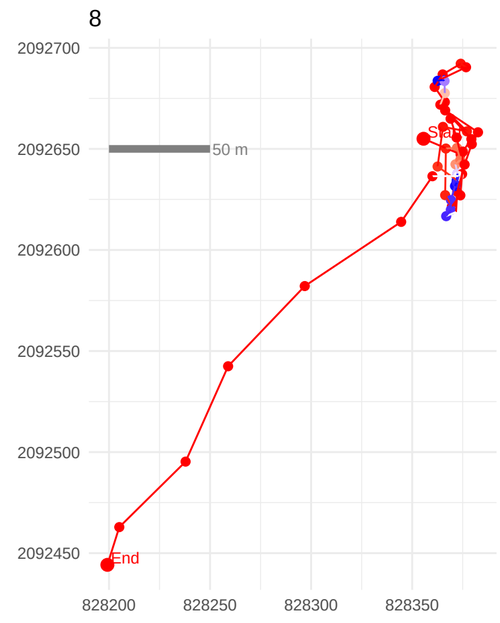

3
50 m
2092800
2092700
2092600
Start
828250
828300
828350
828400

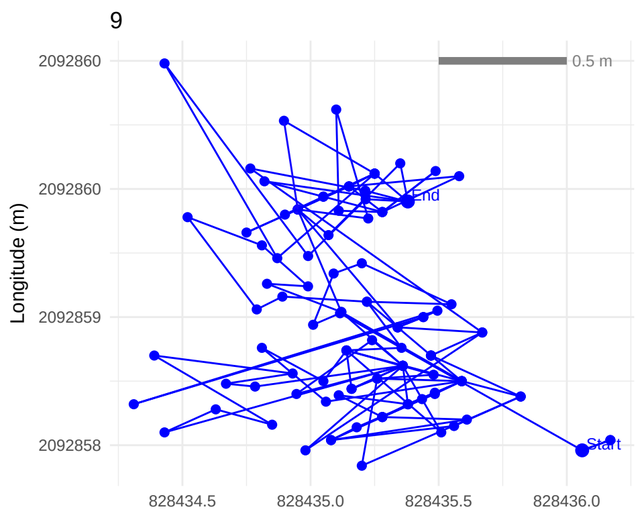

4
25 m
2092710
2092680
2092650
Start
828270
828300
828330
828360

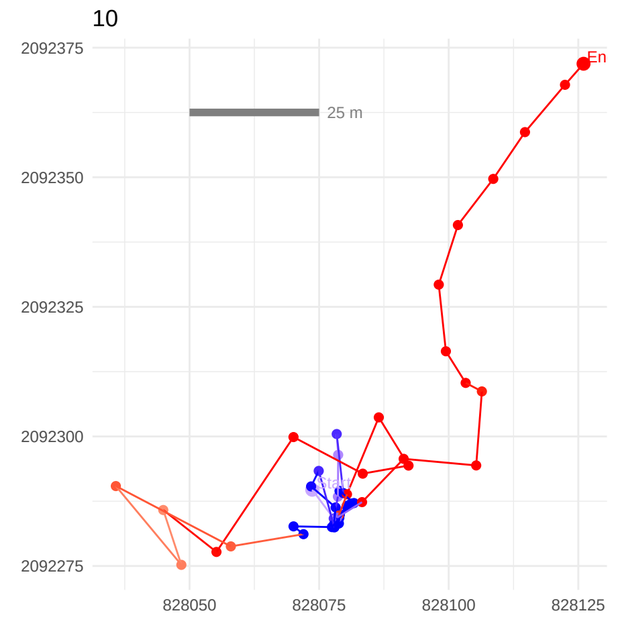

5
25 m
Longitude (m)
2092350
2092325
2092300
2092275
End
828050
828075
828100

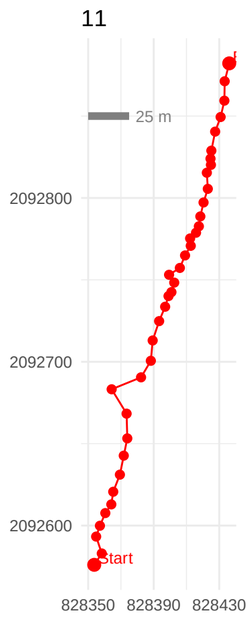

6
Start
5 m
2092296
2092292
2092288
2092284
828035
828040
828045

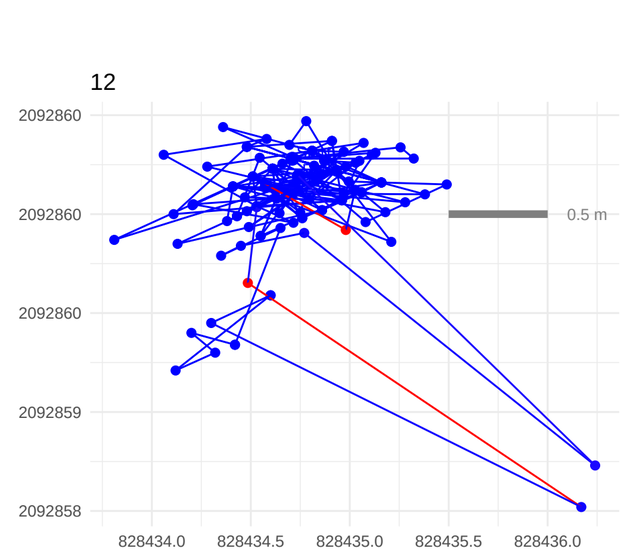

7
25 m
Start
2092400
2092375
2092350
2092325
828200
828225
828250

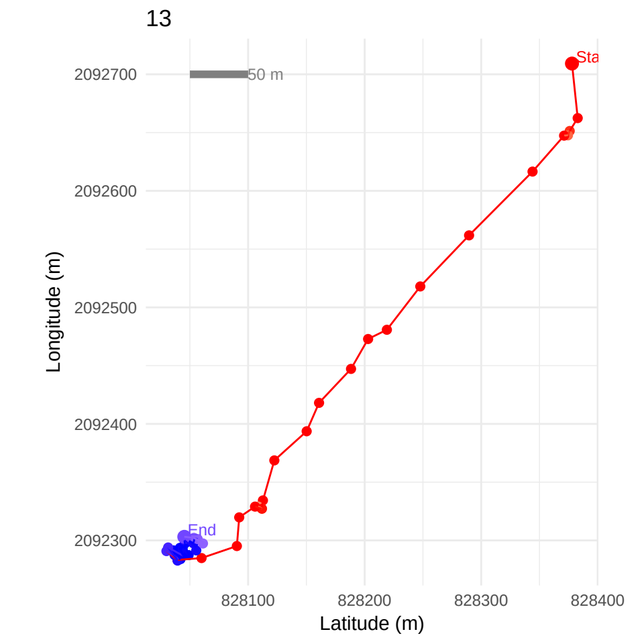

8
50 m
2092700
2092650
2092600
2092550
2092500
2092450
End
828200
828250
828300
828350

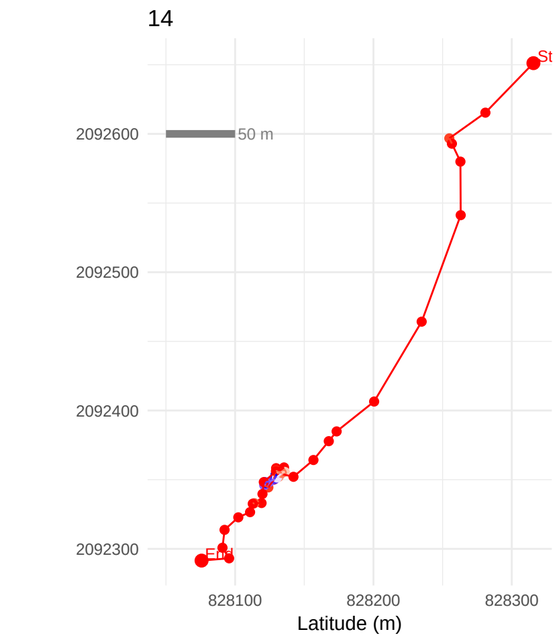

9
0.5 m
Longitude (m)
2092860
2092860
2092859
2092858
End
Start
828434.5
828435.0
828435.5
828436.0

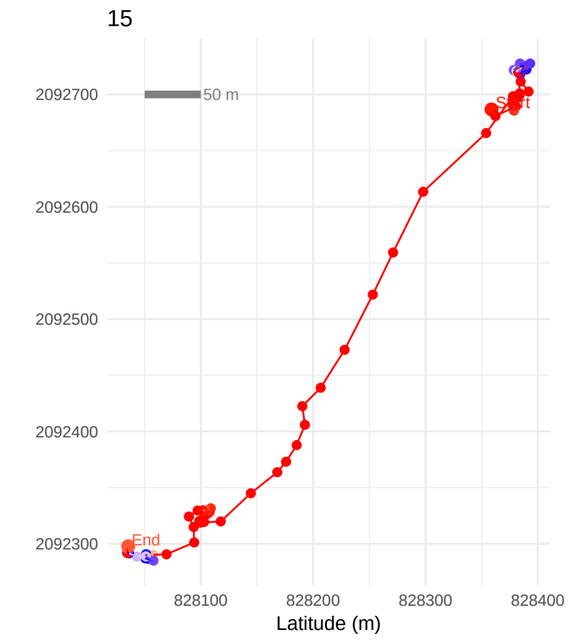

10
25 m
2092375
2092350
2092325
2092300
2092275
828050
828075
828100
828125

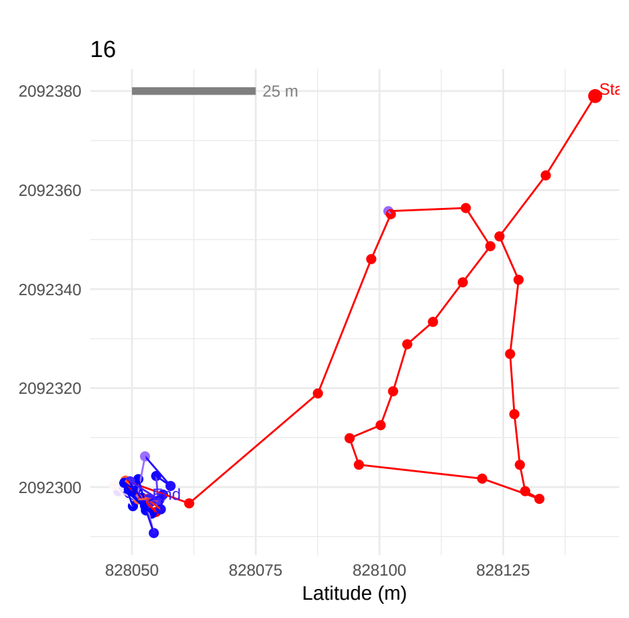

11
25 m
2092800
2092700
2092600
Start
828350
828390
828430


12
0.5 m
2092860
2092860
2092859
2092858
828434.0
828434.5
828435.0
828435.5
828436.0

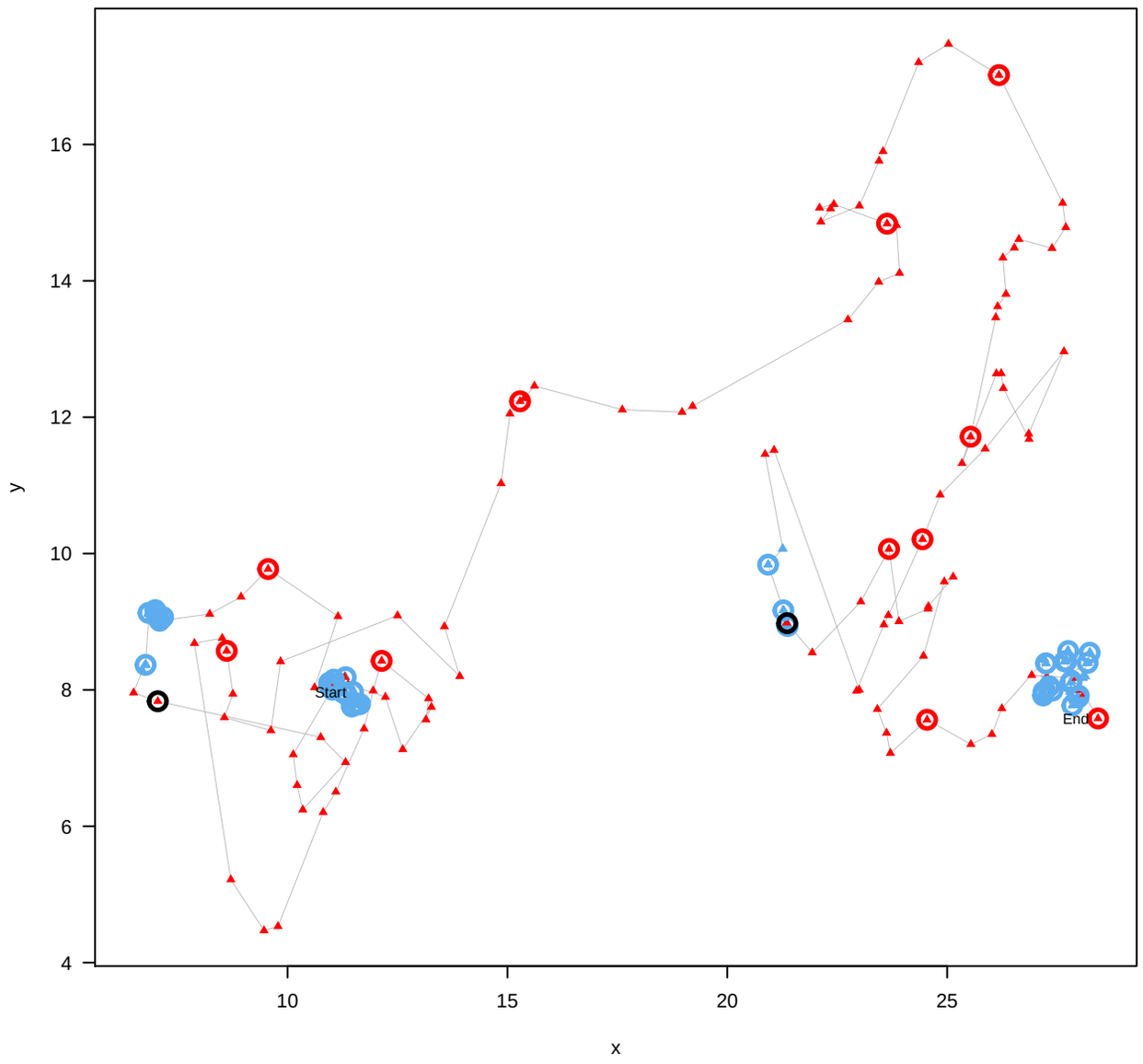

13
50 m
Longitude (m)
2092700
2092600
2092500
2092400
2092300
End
828100
828200
828300
828400
Latitude (m)


14
50 m
2092600
2092500
2092400
2092300
828100
828200
828300
Latitude (m)

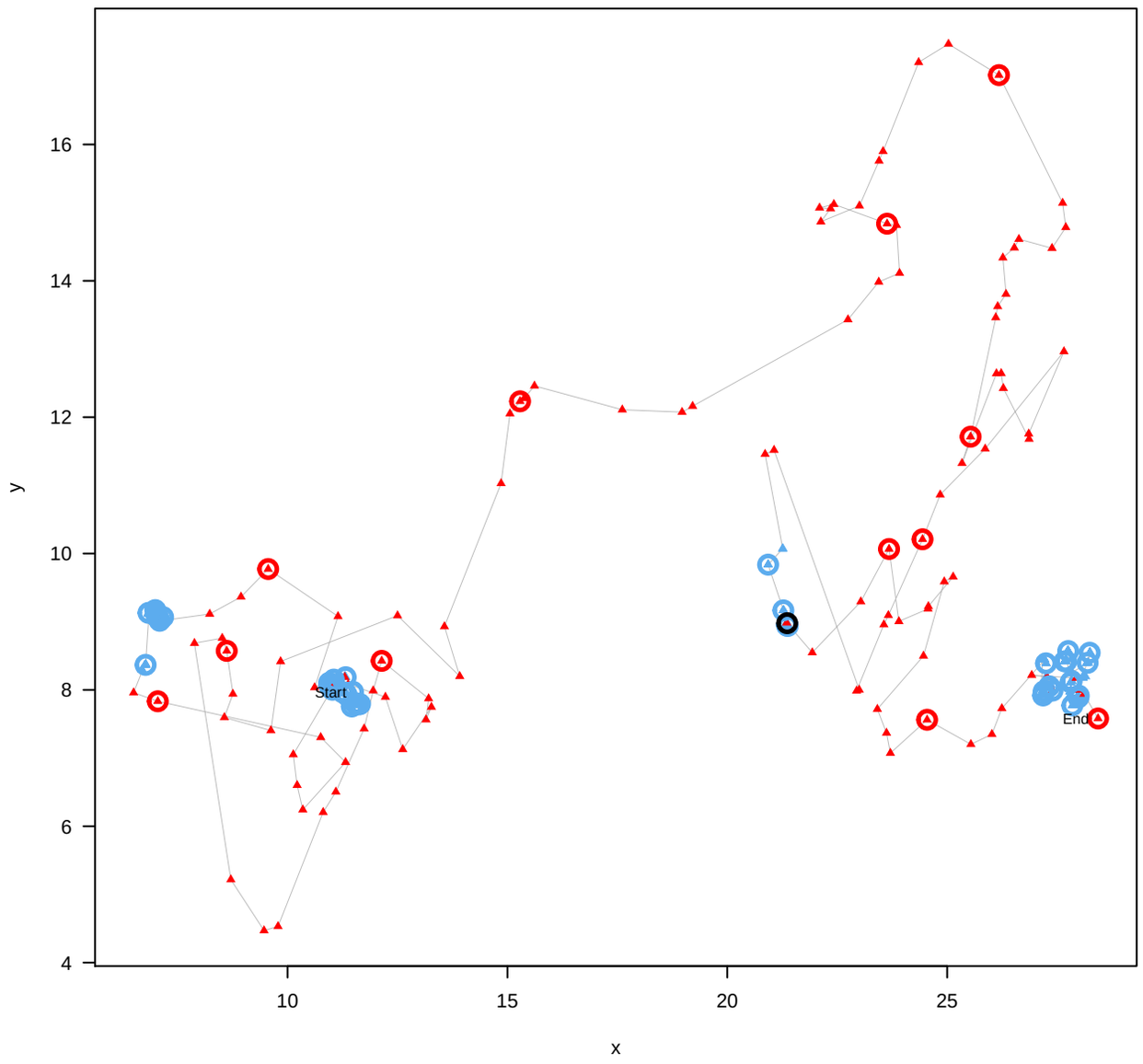

15
50 m
2092700
2092600
2092500
2092400
2092300
End
828100
828200
828300
828400
Latitude (m)


16
25 m
2092380
2092360
2092340
2092320
2092300
828050
828075
828100
828125
Latitude (m)

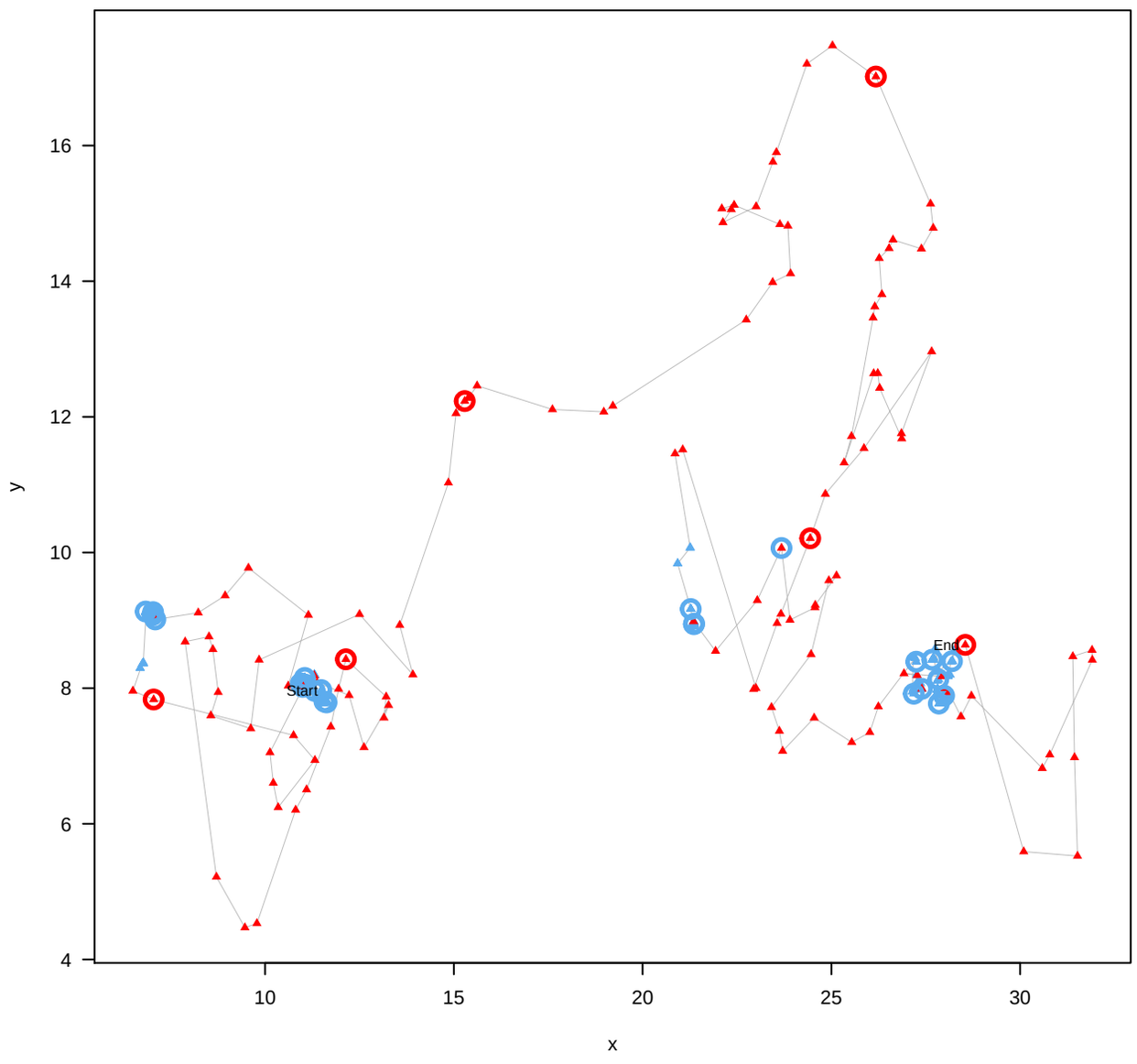
time−step=30s, without U1 and U2



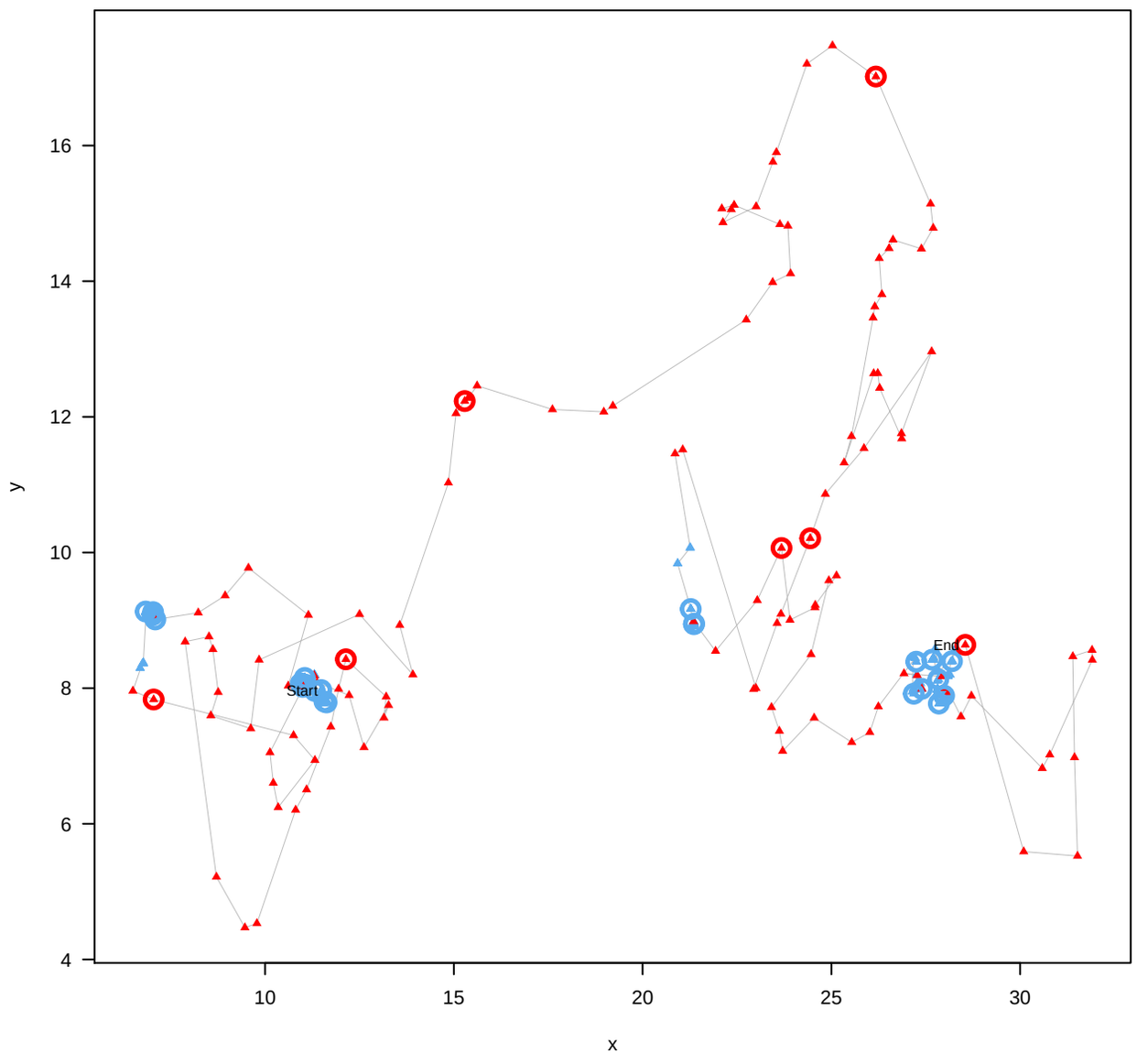
time−step=30s, with U1 and U2



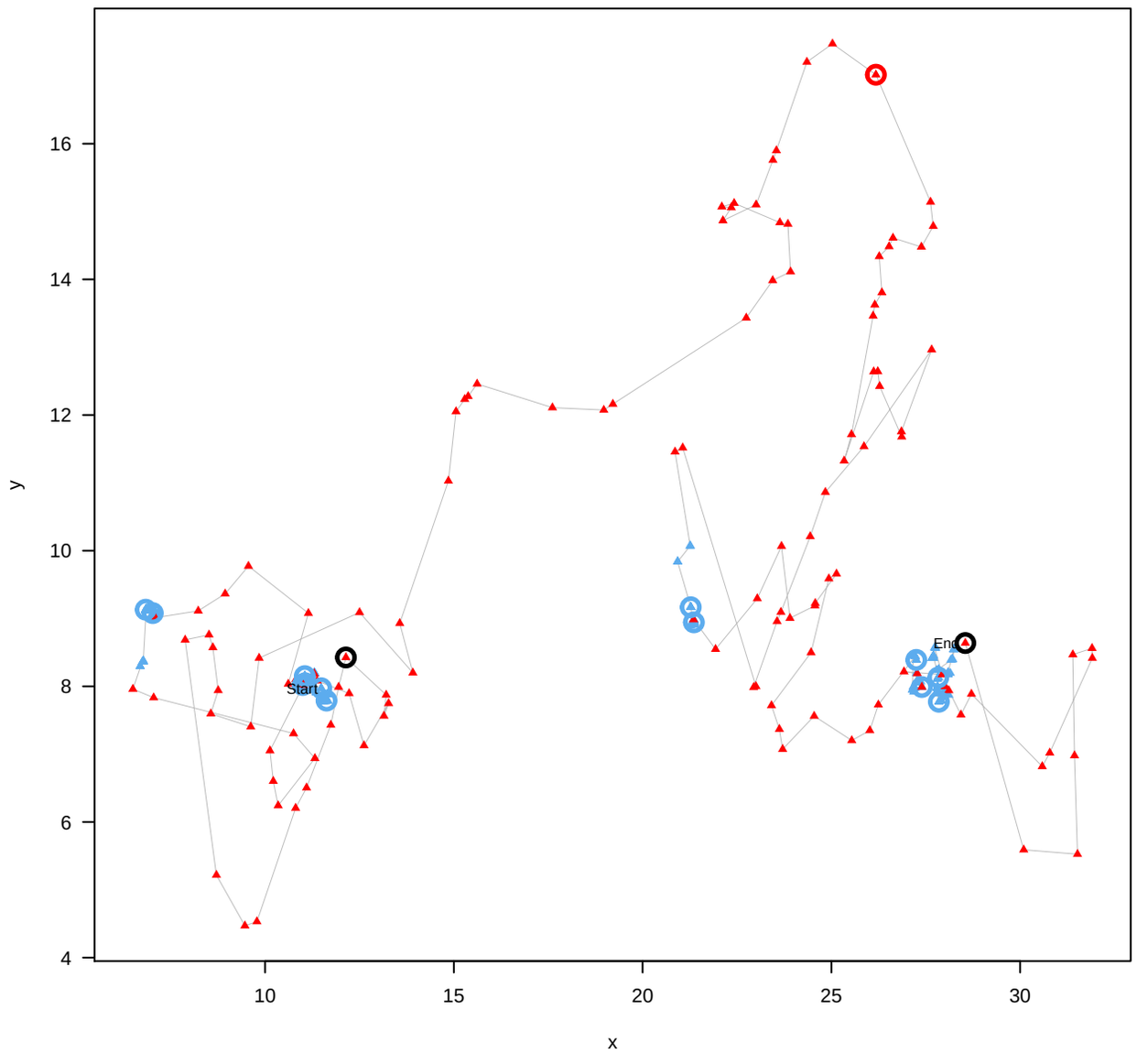
time−step=60s, without U1 and U2



time−step=60s, with U1 and U2



time−step=120s, without U1 and U2



time−step=120s, with U1 and U2