Combining disciplines: dealing with observed and cryptic animal residencies in passive telemetry data by applying econometric decision-making models

Stijn Bruneel[1,2*], Pieterjan Verhelst[2], Jan Reubens[3], Stijn Luca[4], Johan Coeck[5], Tom Moens[2], Peter Goethals[1]

[1]*Department of Animal Sciences and Aquatic Ecology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium,*
[2]*Marine Biology Research Group, Ghent University, Krijgslaan 281, 9000 Ghent, Belgium,*
[3]*Flanders Marine Institute (VLIZ), Wandelaarkaai 7, 8400 Ostend, Belgium,*
[4]*Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Ghent, Belgium,*
[5]*Research Institute for Nature and Forest (INBO), Havenlaan 88, bus 73, 1000 Brussels, Belgium,*
[*]*To whom correspondence should be addressed; E-mail: stijn.bruneel@ugent.be*

Abstract

Migratory species do not necessarily behave migratory continuously. An important aspect of studying migratory species is therefore to distinguish between movement and resident behaviour. Telemetry is a rapidly evolving technique to study animal movement, but the number of data processing techniques to account for resident behaviour remains limited. In this study we describe how models that were initially developed to predict human customer behavior, i.e. two-part and three-part models, provide new insights in the movement of migrating eel by accounting for resident behaviour apparent from telemetry data sets. In econometrics, two-part models take into account that the decision of a customer to purchase an item and the decision of

the customer on the purchase quantity of the concerning product, might be affected by different factors. Similarly, the factors that affect the decision of a fish to migrate or to stay resident might be different from the factors that affect the swimming speed of the fish. Telemetry data of eel movement in the Permanent Belgian Acoustic Receiver Network (PBARN) of the Scheldt Estuary was used. This network with high detection probabilities allowed residencies to be recognized, defined, and introduced as zero values in a movement-residency data set. Two-part models, which consider movement decision, i.e. residency or movement, and movement intensity, i.e. swimming speed, as two different processes or parts of one larger model, outperformed one-part models that do not make that distinction. This underlines the complex migration behaviour eels exhibit. These two-part models in turn were outperformed by three-part models that also accounted for cryptic (i.e. unobserved) residencies. While the one-part model identified the tides and the distance from the most upstream gate as most important for movement, the three-part models identified the tides as most important for the movement decision and the distance from the most upstream gate as most important for the movement intensity. Considering movement decisions, cryptic residencies and movement intensity in modelling efforts increased model performance by 10 %, underlining the importance of acknowledging the potentially complex behaviour animals exhibit.

Keywords: Acoustic telemetry, Fish movement, Residency, Gate, Two-part and three-part model, Eel migration

1. Introduction

Zero values are often encountered in ecological count data where they typically represent absences. However, zeros may have different meanings as they may arise from real absences due to habitat unsuitability, or from false absences due to observer and design errors (Blasco-Moreno et al., 2019). Similarly, in telemetry studies, data-sets can be heavily zero-inflated if moments of non-detection are considered as zeros (Brownscombe et al., 2019; Whoriskey et al., 2019), with the meaningfulness of these zeros being strongly dependent on the network design (Bruneel et al., 2020). Since the objective of many telemetry studies is to describe movement behaviour of animals, zero values could be used as an indication of non-movement or residency. However, accounting for resident behavior, represented as zero values in telemetry data, might require adapted models. Therefore, the aim of this study is to evaluate currently used models and to assess the potential of alternative models to deal with such data.

A good network design is key to defining zero values. In estuarine and riverine acoustic networks with good detection probabilities, receivers may act as gates that tagged animals need to pass to leave a specific area (Kraus et al., 2018; Steckenreuter et al., 2017). Therefore, animals which remain undetected could still be positioned within a zone of the study area, i.e. between two gates, allowing periods of non-detection to be considered as residencies between detections (Bruneel et al., 2020).

However, the specific animal behaviour between detections often remains

entirely unknown, unless some expert-knowledge, such as typical swimming speed and spawning period, is integrated. For example, a fish known to migrate during a certain period would be expected to perform highly unidirectional movement behavior and unexpected travel delays would be an indirect indication of resident behavior between detections. Since these residencies between detections cannot be observed directly, they are referred to as cryptic residencies.

Zero-inflated data sets often require adapted statistical tools. Depending on the nature of zeros, different statistical ecological models have been suggested. If both false and true zeros are likely to be present, zero inflated models are typically used, while hurdle models are used when there are only true zeros (Zuur et al., 2009). More specifically, hurdle models assume that two processes result in two distinct signals, i.e. zero versus not zero, while zero-inflated models assume that both processes can yield zero values. For example, in case detection probabilities are low, individual fish not being detected might actually be present, yielding false zeros in addition to structural zeros. In such a case, zero-inflated models would be most appropriate. Within the field of ecology, zero-inflated and hurdle models are typically used for modelling count data inflated with observed absences (Blasco-Moreno et al., 2019; Joseph et al., 2009; Zuur et al., 2009). Although continuous and proportional ecological data sets are omnipresent, model equivalents for these types of data are not often used. However, for continuous ecological data, such as fish swimming speed, models with a similar approach but different

underlying distribution could be useful (i.e. while for count data, Poisson or negative binomial distributions are typically used, Gaussian or Gamma distributions would be more appropriate for continuous data). In econometric studies for example, the continuous equivalents of hurdle models, known as two-part models, have already been used frequently (Deb and Holmes, 2002; Farewell et al., 2017).

Excess zeros are often considered a nuisance as they typically require more complex models with more parameters to be defined (Warton, 2005). However, explicitly accounting for zero-values may be useful as they may represent a unique signal of an unconsidered process. For example, in econometrics, two-part models have been widely used to study customer behaviour (Neelon et al., 2016; Pohlmeier and Ulrich, 1995). A customer might decide to purchase a product (Will I buy this?), but after that decision he/she would also need to decide on the quantity of the product (How much of it will I buy?). The conditions that drive the customer to purchase may be different from those driving the level of consumption. Hence, accounting for each process separately may be necessary to understand customer behavior.

Similarly, the factors that trigger fish movement, i.e. the movement decision, may be different from the factors determining the distance or speed with which the fish moves, i.e. the movement intensity. Hence, accounting for movement decision and intensity separately may also be necessary to understand fish movement behavior. Therefore, the aim of this study was to assess the added value for predictions and the implications for ecological knowledge

of distinguishing between both processes of fish movement behavior. More specifically, we compared the predictive performance and inferred ecological knowledge of one-part and two-part models describing the movement behaviour of migrating eel (*Anguilla anguilla L.*) in the Scheldt Estuary. In addition, to assess whether a further compartmentalization (e.g. distinction between upstream and downstream movement) would provide added value, different three-part models were constructed and compared with the one-part and two-part models.

Given the increasing data availability and complexity entailed by the exponential increase of possible associations among predictors, machine learning is gaining ground among movement ecologists because of its high predictive performance and alleged ease of use (Joseph et al., 2017; Wang, 2019). However, although machine learning is built on a statistical framework, the outputs of pattern-learning algorithms are often difficult to interpret in the wider context of system functioning (Bzdok et al., 2018). Therefore, in practice, the choice between machine learning and statistical models is typically determined by the purpose, which is either to make predictions or to infer knowledge, respectively. However, since ecologists typically want the best of both worlds, i.e. a model that is interpretable in terms of ecological knowledge while remaining broadly applicable for predictions, statistical models and machine learning should be treated as complementary tools. Therefore, we also compared the interpretability and the predictive performance of statistical models (i.e. one-part and two-part regression models), hybrid models

(i.e. three-part models that combine neural networks with generalized linear regression) and machine learning algorithms (i.e. conditional inference random forests (RF)) for the current telemetry data set.

2. Materials and methods

2.1. Study area

The Schelde Estuary is a well-mixed estuary of 160 km long without transversal man-made migration barriers and characterized by strong currents, high turbidity and a large tidal amplitude up to 6 m (Cornet et al., 2016). The estuary can be divided in two regions (upstream to downstream): the Zeeschelde, which spans 105 km from Ghent to Antwerp (Belgium), and the Westerschelde, which covers the 55 km from Antwerp to the mouth of the estuary at Vlissingen (The Netherlands). The width of the Zeeschelde varies between 50 to 1350 m while the width of the Westerschelde varies between 2000 and 8000 m (Fig. 1). The description of the study area was adopted from Bruneel et al. (2020). This study was limited to the part of the Zeeschelde, because of the relatively low detection probability of the gates in the Westerschelde (see section 2.3).

2.2. Tagging procedure

At the tidal weir in Merelbeke (Ghent), 100 eels were caught and internally tagged with V13 (VEMCO Ltd., Canada) coded acoustic transmitters (Verhelst et al., 2018). After capture, surgery and recovery (Thorstad et al.,

114 2013), fish were released at the nearest receiver. Of the 100 tagged eels, 58
115 migrated. The migration period of these 58 eels was determined (Verhelst
116 et al., 2018) and used for further analysis. A more detailed description of
117 the tagging procedure is provided in Appendix A. The description of the
118 tagging procedure was adopted from Bruneel et al. (2020).

119 *2.3. Acoustic network*

120 Within the framework of the Belgian LifeWatch observatory, a permanent
121 longitudinal network of receivers (VR2W, VEMCO Ltd, Canada) has been
122 deployed since the spring of 2014 in the Schelde Estuary (Reubens et al.,
123 2019a). Currently, the network consists of 25 receivers, deployed from the
124 river bank, which were combined into 18 gates that are on average 4969 m
125 apart (Fig. 1 and Table B.1). At four locations (s15, s16, s17 and s18), a
126 receiver on each side of the estuary was deployed to cover the whole width.
127 The exact detection range for the different receivers in the Zeeschelde was
128 unknown, but ranges between 300 m and 1000 m (Verhelst et al., 2018). Re-
129 sults from the network in the North Sea suggest that it is strongly dependent
130 on current velocity and wave action and will therefore be characterized by
131 a strong spatial and temporal variability (Reubens et al., 2019b). The de-
132 tection probability of the gates was estimated using the conditional nature
133 of fish movement throughout the system (Brownscombe et al., 2019). Since
134 there are no other pathways to the North Sea, tagged fish have to pass the
135 different gates in a well-defined order and detection probability can be defined

as the probability of detecting a tag moving past a specific gate (Melnychuk, 2012; Perry et al., 2012). The detection probabilities of the different gates are given in Table B.1. The description of the acoustic network was adopted from Bruneel et al. (2020).

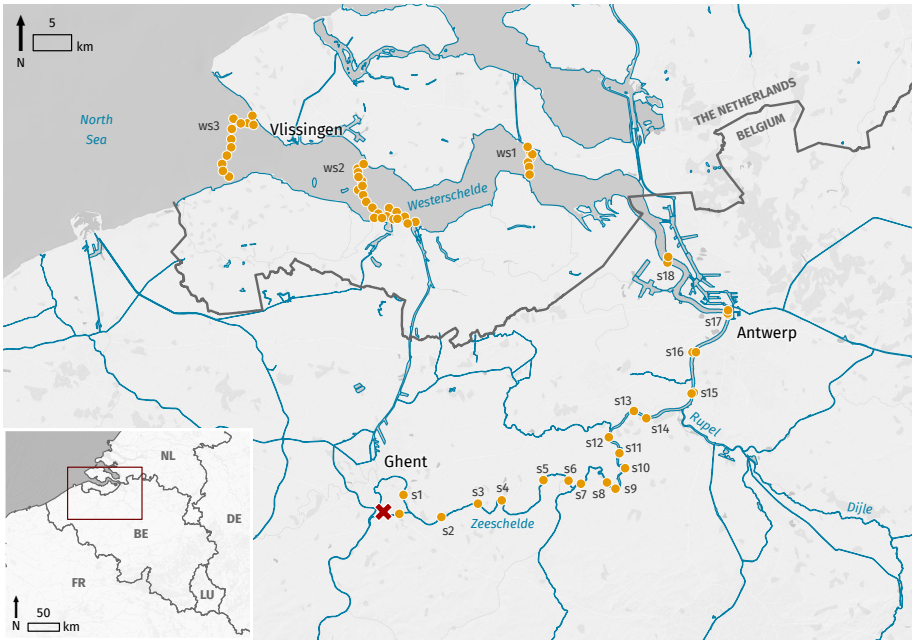

Figure 1: The Schelde Estuary comprises the Zeeschelde (Ghent-Antwerp) and Westerschelde (Antwerp-Vlissingen). The receivers are represented as orange circles. The gates are indicated as labels for different groups of receivers. The weir in Ghent where the eels were caught and released is depicted as a red cross. Detections at the three gates in the Westerschelde (ws1, ws2 and ws3) were not considered in this study because of their relatively low detection probabilities. Adapted from Bruneel et al. (2020).

2.4. Eel movement

Given the high average detection probability of 97.0 % in the Zeeschelde, the number of false zeros was likely limited. In addition, since the movement of eels is highly unidirectional once they have started migrating, eels not

being detected at one gate are likely to be detected at the next gate, causing some reduction in resolution, but still providing a reliable position estimate. Since false zeros, due to low detection probabilities, are unlikely, we decided to work with two-part models instead of zero-inflated models.

When a tagged eel was consecutively detected at two different gates, we considered the time lapse between these two detections as a movement interval and the distance between the two gates was determined. When a tagged eel was detected multiple times at a specific gate without being detected at any other gate, the time lapse between the earliest and last detection was considered as a residency interval and was assigned a distance value of zero. It should be noted that some short intervals might actually have been identified incorrectly as residency intervals. For example, a migrating eel might come within the detection range and have multiple detections while moving from one side of the gate to the other. Although considered as highly variable in literature (Breukelaar et al., 2009; Verbiest et al., 2012), we assumed an average migration speed of 0.25 m s^{-1} and a detection range of 250 meter, yielding an approximate threshold value of at least 30 minutes for residency intervals. To ensure that movement was not wrongly identified as residencies, this value was doubled and all residency intervals with a time span below 1 hour were omitted from the analysis (10.80 % of the residency intervals were retained for analysis).

It is possible that, before heading to the next gate, a tagged eel was resident between two gates without entering either gate's detection range.

Such unobserved or cryptic residencies are not directly apparent from the data as they are observed as being part of the movement interval. However, these cryptic residencies can be accounted for indirectly as they will cause a travel delay in the movement interval, negatively affecting the time necessary to reach the next gate.

Recognizing resident behavior in acoustic telemetry networks based on position estimates alone often remains a difficult objective (Cagua et al., 2015). In this specific study, indirect (i.e. through travel delays) and direct indications of apparent non-movement can be either the result of (i) fish choosing to be resident and to discontinue swimming or (ii) fish swimming against the currents without much net gain in distance covered. However, if there are clear signs of individual variation and enough individuals to account for it, a distinction between both can be made. When animals are resident, they cannot be distinguished from each other using position estimates alone. However, when they migrate, even against the current, the fastest individuals will reach higher swimming speeds and can as such be distinguished from slower individuals. It should be noted that throughout the manuscript, swimming speed represents the ground speed (i.e. geographical progress per unit of time) without correction for current speed. As each tag emits a signal at a unique frequency, individuals can be identified and individual variation in swimming speed can be determined and used to analyse the behavior associated with apparent non-movement. Since European eel and other migrating fish have been found to apply different strategies to

save energy, it is much more likely that eels would choose the most energy efficient option and choose to be resident when facing currents rather than to swim without much net gain in distance covered (Arnold and Cook, 1984; Glebe and Leggett, 1981; Metcalfe et al., 1990). Therefore, we consider measurements of apparent non-movement as residencies and evaluate afterwards whether this choice was justified based on the outcomes of the models.

To normalize the data, distances were divided by time, yielding swimming speed. Residency and movement intervals with a time lapse higher than one full tidal cycle were removed (13.42 % data removal) as they do not allow to contribute movement behaviour to either ebb tide, flood tide or a combination of both (see section 2.5). To account for telemetry detection errors that might cause unrealistic swimming speeds, movement intervals with a swimming speed of 1.5 interquartile ranges (IQRs) below the first quartile or above the third quartile were considered outliers and removed from the data set (Tukey, 1977). In practice, all movement intervals with a swimming speed higher than 2.7 or lower than -1.5 m/s were omitted from the analysis (additional 2.20 % data removal). In summary, first 89.20 % of residency intervals were removed, followed by a 13.42 % removal from the entire data set (movement intervals + residency intervals), followed by a 2.20 % removal from the entire data set. The final data set contained 19.24 and 80.76 % residency and movement intervals, respectively.

2.5. Environmental data

As the biological response in this study was analysed at a relatively fine spatiotemporal resolution (Bultel et al., 2014; Verhelst et al., 2018), a sound coupling of biological and environmental data would have been challenging and use of daily averages would have yielded inconclusive results on within-day movement patterns. Therefore only variables were included that were fixed in time (i.e. distance from source), fixed in space (i.e. day phase), known to be accurate at high spatial and temporal resolutions (i.e. period of flooding and period of ebbing), or known to be well represented by daily averages (i.e. moon and tidal phase). It should be noted that the main aim of this study was to assess the potential of alternative ecological models rather than to identify all environmental factors affecting eel migration. To obtain a more comprehensive understanding of these environmental factors, more fine-scale measurements and/or simulations of potentially important environmental variables, such as discharge, temperature, salinity and precipitation could be used to fine-tune the developed models.

The following description of the collection and processing of tidal data was adapted from Verhelst et al. (2018): To account for the distances between the locations of the gates and of the tidal measuring stations (Hydraulic Information Centre, Belgium), a weighted average method was applied to estimate the precise moments of low and high water at the gates. The closest upstream and downstream tidal measuring stations were assigned to each gate. Based on the distances between these tidal stations and the gate, lin-

ear weights were assigned to both tidal stations. When tidal data at the respective upstream or downstream tidal station was absent or of questionable quality (e.g. outliers and known periods of malfunctioning measuring devices) at the time interval of interest, the next upstream or downstream tidal station was chosen. This allowed us to estimate the duration of ebbing and flooding for each movement and residency interval.

The ratio of period flood tide (minutes) over total period of the interval (minutes) was determined and used as a predictor, i.e. flood ratio. Per gate, the ratio of the maximum difference in water level of the concerning day over the median of the maximum difference in water level per day of the entire study period was used as a proxy for tidal phase, with low values being associated with neap tide and large values with spring tide. Moon phase was a numerical value representing the degree of illumination of the moon, ranging from new moon (0) to full moon (1). Time of day was a categorical variable with the classes Day, Night, Dusk and Dawn. Distance from source gave the distance (km) from the most upstream gate to the detecting gate.

2.6. Model construction and evaluation

All analyses were performed using the R software (R Core Team, 2019). To construct the different models, the *stats* (R Core Team, 2019), *nnet* (Venables and Ripley, 2002) and *ranger*(Wright and Ziegler, 2014) packages were used.

255 *2.6.1. Model construction*

256 In the one-part, two-part, three-part and random forest models, swim-
257 ming speed was used as response variable, while flood ratio, tidal phase,
258 moon phase, day phase and distance from source were evaluated as poten-
259 tial predictors. Linear weights were introduced in model construction and
260 evaluation to account for the different numbers of observations between eels.
261 The weights of the observations were determined for each eel independently
262 as $1/n_k$, with n_k the number of observations of eel k. Hence, for each eel the
263 sum of the weights was one. As a consequence, each eel contributed equally
264 to the constructed models. In the one-part, two-part and three-part models,
265 these weights were used to determine weighted likelihoods. In the random
266 forests, these weights represented the probability with which observations
267 were selected in the bootstrap. First, a one-part model was constructed for
268 the entire data set which consisted of a multiple linear regression model with
269 Gaussian distribution.

270 Second, continuous two-part models were constructed which consisted of
271 two sub-models (adapted from Belotti et al. (2015)): (1) A binomial model
272 for the entire data set, with movement and residency as contrasts,

$$Pr(y \neq 0|\mathbf{x}) = F(\mathbf{x}^{\mathbf{T}}\boldsymbol{\alpha}) \tag{1}$$

273 where y is the response variable, $\mathbf{x}$ is a vector of predictors ($\mathbf{x} = (1, x_1, \ldots, x_k)$,
274 with k the number of predictors), $\boldsymbol{\alpha}$ is the corresponding vector of parameters

to be estimated ($\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)$, with k the number of parameters), and F is the cumulative distribution function of an independent and identically distributed error term from a probit model. (2) A multiple linear model with Gaussian distribution solely for the movement data,

$$\theta(y|y \neq 0, \mathbf{x}) = h(\mathbf{x^T}\boldsymbol{\beta}) \tag{2}$$

where θ is the probability density function, $\boldsymbol{\beta}$ is the corresponding vector of parameters to be estimated, and h is a Gaussian density function for y with expectation $x^T\beta$ and some constant variance σ^2. The likelihood contribution for an observation can be written as,

$$\theta(y) = \left\{1 - F(\mathbf{x^T}\boldsymbol{\alpha})\right\}^{i(y=0)} \times \left\{F(\mathbf{x^T}\boldsymbol{\alpha})h(\mathbf{x^T}\boldsymbol{\beta})\right\}^{i(y\neq 0)} \tag{3}$$

where *i(.)* denotes the indicator function. Then, the log-likelihood contribution is,

$$ln(\theta(y)) = i(y=0)ln\left\{1 - F(\mathbf{x^T}\boldsymbol{\alpha})\right\} + i(y \neq 0)[ln\left\{F(\mathbf{x^T}\boldsymbol{\alpha})\right\} + ln\left\{h(\mathbf{x^T}\boldsymbol{\beta})\right\}] \tag{4}$$

Because the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters are additively separable in the log-likelihood contribution for each observation, the models for the full data set and the non-zeros can be estimated separately. Predictions of y_i, $\widehat{y_i}|x_i$, were obtained by multiplying the predictions from each part of the model for the corresponding

observations,

$$\widehat{y_i}|x_i = (\widehat{p_i}|\mathbf{x}_i) \times (\widehat{y_i}|y_i \neq 0, \mathbf{x}_i) \tag{5}$$

where $\widehat{p_i}|\mathbf{x}_i$ is the predicted probability that $y_i \neq 0$. To obtain the most parsimonious model, each part of the model was constructed using a step-wise approach with AIC as selection criteria,

$$AIC = -2lnL + 2k \tag{6}$$

where L is the maximum value of the likelihood function and k the number of estimated parameters.

By definition, two-part models assume that both parts of the model are independent. However, this should not always necessarily be the case. Therefore, the added value of accounting for any dependence between both parts was also assessed. This type of model is referred to as a selection model in literature, and can be constructed using a two-stage estimation procedure: (1) The Inverse Mills Ratio (IMR) is determined from the binomial model for the full data set, and (2) the linear regression model for the movement data is constructed with IMR as additional covariate (Heckman, 1979). The IMR is,

$$IMR(\mathbf{x}) = \frac{\phi(\mathbf{x})}{\Phi(\mathbf{x})} \tag{7}$$

with ϕ the standard normal density, Φ the standard normal cumulative distribution function and $\mathbf{x}$ the vector of linear predictors of the binomial model.

306 To assess whether further distinction between upstream and downstream
307 movement would improve the predictions, a three-part model was constructed.
308 This model consisted of 1) a multinomial model (via neural networks) with
309 three contrasts: residency, upstream movement and downstream movement;
310 2) a linear model of the upstream movement; and 3) a linear model of the
311 downstream movement.

312 One could argue that the few upstream intervals (3.7 % of the total
313 amount of intervals per eel), actually represented residency intervals gone
314 wrong (i.e. eel trying to stay resident are in fact slightly pushed back up-
315 stream; see also section 3.1). Therefore additional one-part and two-part
316 models were constructed after transformation of the few upstream movement
317 intervals into residency intervals, i.e. they were given a value 0.

318 Additionally, a three-part model was constructed as an attempt to ac-
319 count for the bimodal pattern of the data (See section 3.1). The three parts
320 in this model were: 0 vs 0 to threshold vs threshold to 2.7 m s^{-1}. After assess-
321 ing the predictive performance of models with different thresholds (threshold
322 interval selection based on inspection of stacked density plots in section 3.1)
323 from 0.3 to 0.7 with a step-size of 0.01 and 10^4 Monte-Carlo cross-validations,
324 the threshold that yielded the model with the highest predictive performance
325 was retained (threshold = 0.45 m s^{-1}; see section 3.2).

326 Finally, conditional inference random forests were used to analyse both
327 data sets, i.e. with and without upstream movement intervals. Different
328 parameter settings were assessed, but since default parameters gave slightly

higher performances, only these results were reported.

2.6.2. Model performance

To assess the performance of the models, Monte Carlo cross-validations were performed with 10^6 repeats, during which some individuals were used for training and some for testing. Different ratios (2/3, 3/4, 4/5, 5/6, 6/7, 7/8, 8/9 and 9/10 for training) were assessed but since very similar results were obtained within each model, e.g. 0.1 % difference in Root Mean Square Error (RMSE), only results for a ratio of 9/10-1/10 for training-testing, were reported. Per repeat, a step-wise approach with AIC as selection criterion was used to arrive at the most parsimonious model. Per repeat the RMSE was calculated as given in Eq. 8, with m the number of eels in the test data set, n_k the number of observations of eel k, y_j the actual value and $\widehat{y_j}|x_j$ the predicted value of the swimming speed. Finally, the average RMSE over all repeats was determined.

$$RMSE = \frac{1}{m}\sum_{k=1}^{m}\sqrt{\frac{1}{n_k}\sum_{j=1}^{n_k}(y_j - \widehat{y_j}|x_j)^2} \tag{8}$$

2.6.3. Model validation

To quantify the uncertainty of the parameter estimates, bootstrap confidence intervals were determined. While standard parametric inferences rely on a-priori assumptions of the underlying distribution of the population, the non-parametric resampling approach of bootstrapping provides an estimate

of the statistic's sampling distribution using within-sample variation. More specifically, by considering the sample distribution as representative for the population distribution, bootstrapping can be used to estimate the quality of the predictive model. First, to develop the most parsimonious models, model selection was performed using the procedure described by Austin and Tu (2004), based on bootstrap samples, backwards elimination and AIC (n=10^4). Second, the coefficient estimates of the retained variables and their 95% bootstrap percentile confidence intervals were determined (n=10^4) (Davison and Hinkley, 1997). Linear bootstrap sampling weights were used to account for the different number of observations between eels.

2.6.4. Extension to one-part and two-part mixed models

One major advantage of telemetry is its ability to provide data on the level of individuals and therefore mixed models that account for individual correlation are commonly used. Therefore, we also compared the explanatory power of one-part mixed models and two-part mixed models. Both models had eel ID as random intercept. The RMSE values were used as proxies of explanatory power. Since in the two-part models independence between parts is assumed, we did not account for any correlation across both fixed effects and random effects from the different parts of the two-part model (i.e. the random effects of the binomial model and those of the linear model were determined independently).

3. Results

3.1. Exploratory analysis

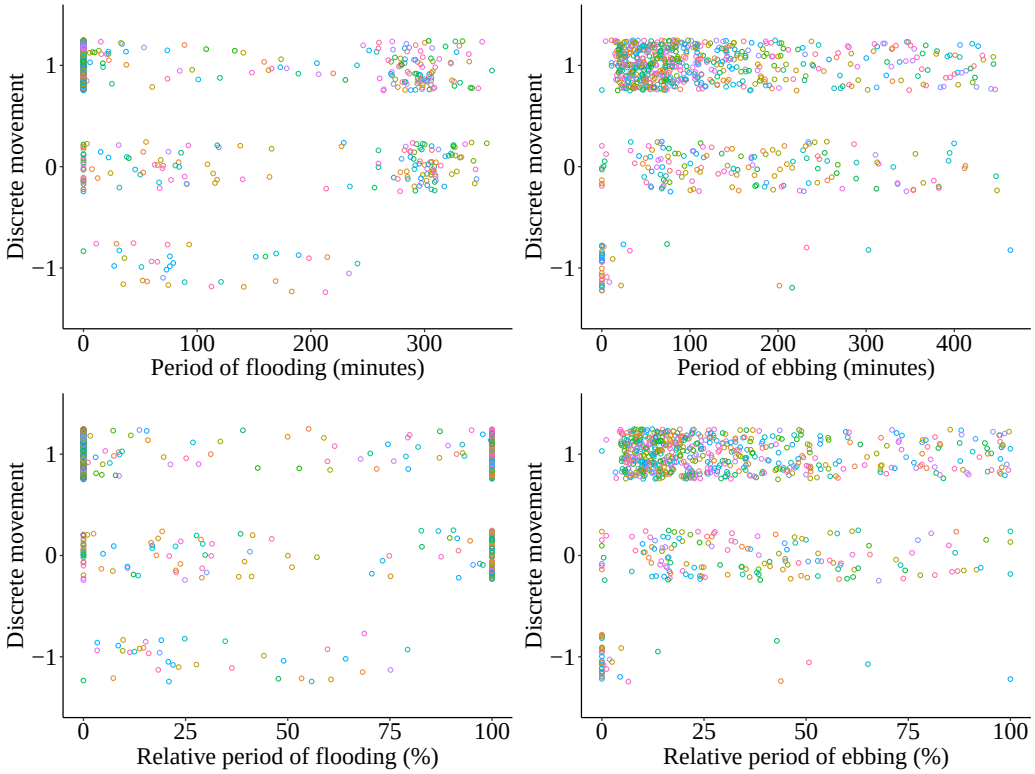

Figure 2: Graphs of discrete movement. Downstream movement (1); upstream movement (-1); residency (0) versus the relative (%) and actual (minutes) period of flooding and ebbing. All movement and residency intervals are depicted. Different colors represent different eels.

An exploratory analysis of the data suggests that downstream movement intervals generally took place during ebb tide (Figs 2 and C.1). The normalized duration of flood tide in the downstream movement intervals was either 0 or to a lesser extent 100 % (Fig. 2), suggesting that downstream movement intervals contained either no flooding at all or a full flood cycle.

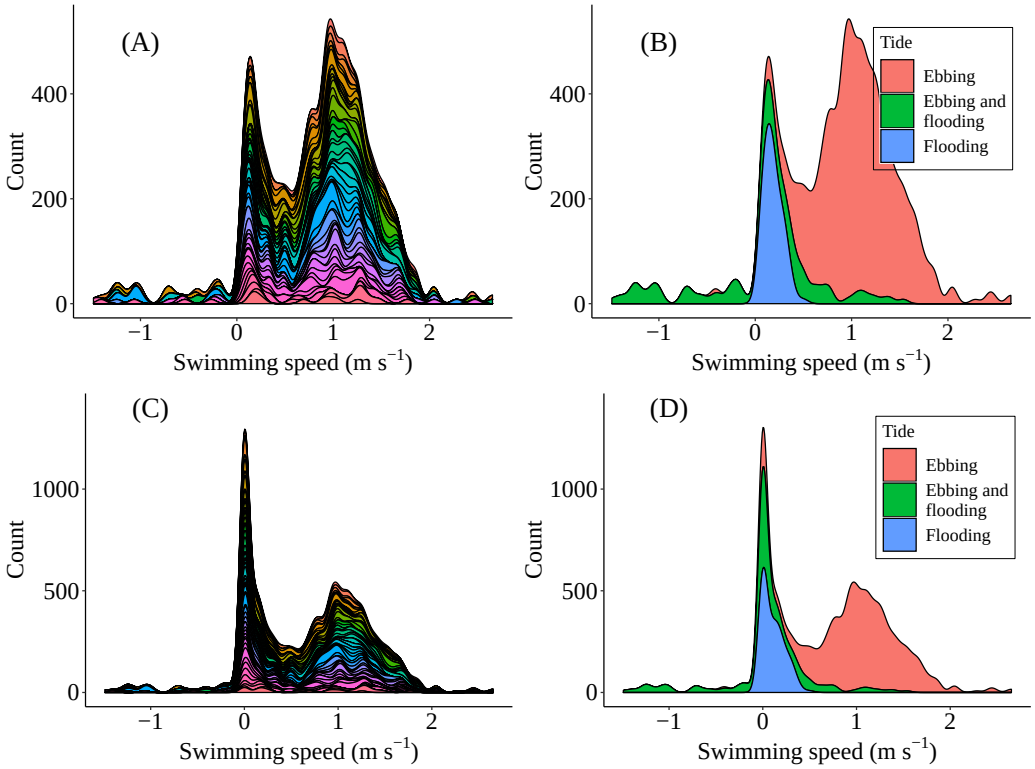

Figure 3: Transformed stacked density plots of eel swimming speed (m s^{-1}). To determine the count of the stacked density plots, there are several steps involved. First, the x-axis, which represents the swimming speed, is subdivided in swimming-speed intervals with a width of 0.05 m s^{-1}. Second, the amount of movement (and residency) intervals for each swimming-speed interval is divided by the swimming-speed interval width. For example, in the swimming-speed interval centering the value 1 m s^{-1}, 25 movement intervals were found. Hence, 25 movement intervals divided by a width of 0.05 m s^{-1} yield a count of 500. In A and B the density plots of all movement intervals are given. The different colors in A depict the different eels, while the different colors in B depict whether movement intervals occurred during flooding, ebbing or a combination of both. In C and D the density plots of all residency intervals and movement intervals are given. The different colors in C depict the different eels, while the different colors in D depict whether residency and movement intervals occurred during flooding, ebbing or a combination of both.

On the other hand, upstream movement intervals typically took place during flood tide (Figs 2 and C.1). Finally, residencies seemed to occur more often during flood tide than during ebb tide (Fig. 2).

Transformed stacked density plots of swimming speed gave additional insights into the distribution of the data (Fig. 3). It is clear from these figures that the bimodal pattern in the data is the result of different tidal conditions rather than of individual differences. Most eels have swimming speeds ranging from 0 to 2 m s^{-1}, but swimming speeds from 0 to approximately 0.45 m s^{-1} typically occurred during pure flooding or a combination of flooding and ebbing, while swimming speeds of approximately 0.45 to 2 m s^{-1} typically occurred during pure ebbing events. This suggests that movement intervals with a swimming speed below approximately 0.45 m s^{-1} are likely to contain cryptic residencies, causing a delay in travel time.

3.2. Model construction and evaluation

For the original data set, Monte-Carlo cross-validations indicated that the three-part model, which compartmentalized predictions into (1) residencies and (2) downstream and (3) upstream movement, had the highest predictive performance (RMSE = 0.4055), followed by the two-part model (RMSE = 0.4073), which compartmentalized predictions in (1) residencies and (2) movement, the selection model (RMSE = 0.4132) and the one-part model (RMSE = 0.4165) (Table 1). After transformation of the upstream movement intervals to residency intervals, Monte-Carlo cross-validations indicated that the three-part model, which compartmentalized predictions into classes of (1) 0, (2) 0 to 0.45 and (3) 0.45 to 2.7 m s^{-1}, had the highest predictive performance (RMSE = 0.3653) followed by the two-part model (RMSE =

Data	Model	RMSE
Original data set	One-part model	0.4165
	Two-part model: 0 vs not 0 m s^{-1}	0.4073
	Selection model: 0 vs not 0 m s^{-1}	0.4132
	Three-part model: 0 vs 0 vs 0 m s^{-1}	0.4055
	Conditional inference random forests	0.3941
No upstream movement	One-part model	0.4051
	Two-part model: 0 vs not 0 m s^{-1}	0.3804
	Selection model: 0 vs not 0 m s^{-1}	0.5410
	Three-part model: 0 vs 0-0.45 vs 0.45-2.7 m s^{-1}	0.3653
	Conditional inference random forests	0.3669

Table 1: RMSE values (Eq. 8) after Monte Carlo cross-validations (10^4 permutations) for different models and different data subsets.

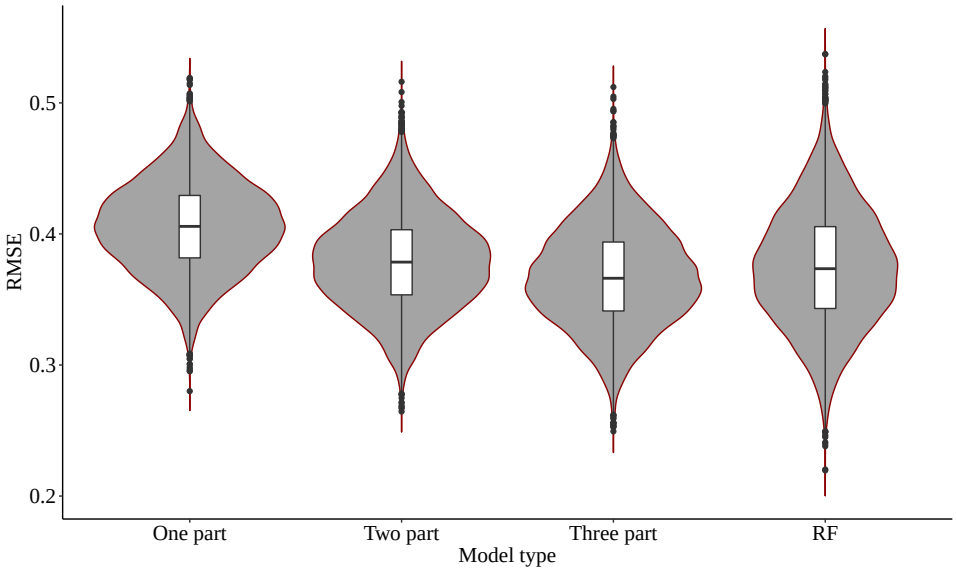

Figure 4: Violin plots representing the distribution of RMSE values obtained through cross-validation (n=10^4) for the different models. RMSE distributions are given for the one-part model, two-part model (0 vs not 0 m s^{-1}), three-part model (0 vs 0-0.45 vs 0.45-2.7 m s^{-1}) and random forests model (RF). The data set without upstream intervals was used to construct the models.

0.3804), which compartmentalized predictions into (1) residencies and (2) movement, one-part model (RMSE = 0.4051) and selection model (RMSE = 0.5410) (Table 1). Since the three-part model performed best, it was retained for further analysis (Table 2).

The results of the multinomial model of the three-part model indicated that the distinction between <0.45 and >0.45 m s^{-1} was significantly better than the distinction between 0 and 0 to 0.45 m s^{-1}. The relative risk ratio for a one-percentage increase in the flood ratio was 0.987 for being between 0 and 0.45 m s^{-1} versus 0 m s^{-1} and 0.907 for being between 0.45 and 2.7 m s^{-1} versus 0 m s^{-1}. The higher the flood ratio, the higher the probability

				Intercept	Flood ratio	Distance	Moon phase	Tidal phase
One-part model			Estimate	0.704	-0.0124	$4.94*10^{-3}$	0.0885	
			CI	[0.629 0.780]	[-0.0133 -0.0116]	[$3.57*10^{-3}$ $6.32*10^{-3}$]	[0.0105 0.166]	
			p-value	0	0	0	0	
Two-part model	Binomial model		Estimate	1.45	-0.0195			
			CI	[1.31 1.60]	[-0.0231 -0.0160]			
			p-value	0	0			
	Linear model		Estimate	0.795	-0.0137	$5.16*10^{-3}$	0.0915	
			CI	[0.725 0.866]	[-0.0151 -0.0123]	[$3.90*10^{-3}$ $6.43*10^{-3}$]	[0.0101 0.174]	
			p-value	0	0	0	0.00258	
Three-part model	Multinomial model	0-0.45 vs 0 m s^{-1}	Estimate	0.507	-0.0134			
			CI	[0.0830 0.973]	[-0.0212 -0.00595]			
			p-value	0.105	$8.00*10^{-4}$			
		0.45-2.7 vs 0 m s^{-1}	Estimate	2.96	-0.0987			
			CI	[2.57 3.43]	[-0.117 -0.0835]			
			p-value	0.126	0.00152			
	Gamma model	0-0.45 m s^{-1}	Estimate	-2.2	-0.00548			0.870
			CI	[-3.31 -1.12]	[-0.00893 -0.00204]			[-0.192 1.94]
			p-value	0.650	0.00167			0.648
	Linear model	0.45-2.7 m s^{-1}	Estimate	0.82	-0.00455	$7.22*10^{-3}$	0.0425	
			CI	[0.750 0.889]	[-0.00690 -0.00219]	[$5.93*10^{-3}$ $8.53*10^{-3}$]	[-0.0436 0.127]	
			p-value	$3.00*10^{-4}$	0	0	0.0011	

Table 2: Parameter estimates, 95% percentile confidence intervals (CI) and p-values of the one-part, two-part and three-part models obtained using a weighted bootstrap approach (n=10^4). The models had swimming speed as response and predictors were selected using a bootstrap selection procedure based on backwards elimination and AIC. The considered predictors were flood ratio (% percentage flood over total period), distance from source (km), moon phase (degree of moon illumination ranging from 0 to 1), tidal phase (ratio of the maximum difference in water level of the concerning day over the median of the maximum difference in water level per day of the entire study period) and day phase (categorical: day, night, dusk or dawn). The data set without upstream intervals was used to construct the models.

411 of an observed residency interval (0 m s^{-1}) and the lower the probability
412 of a movement interval with a swimming speed above 0.45 m s^{-1}. The
413 probability of a movement interval with a swimming speed below 0.45 m
414 s^{-1} shows an increasing trend with flood ratio similar to the probability of
415 residency intervals until a flood ratio of approximately 40 %, after which the
416 probability decreases (Fig. 5).

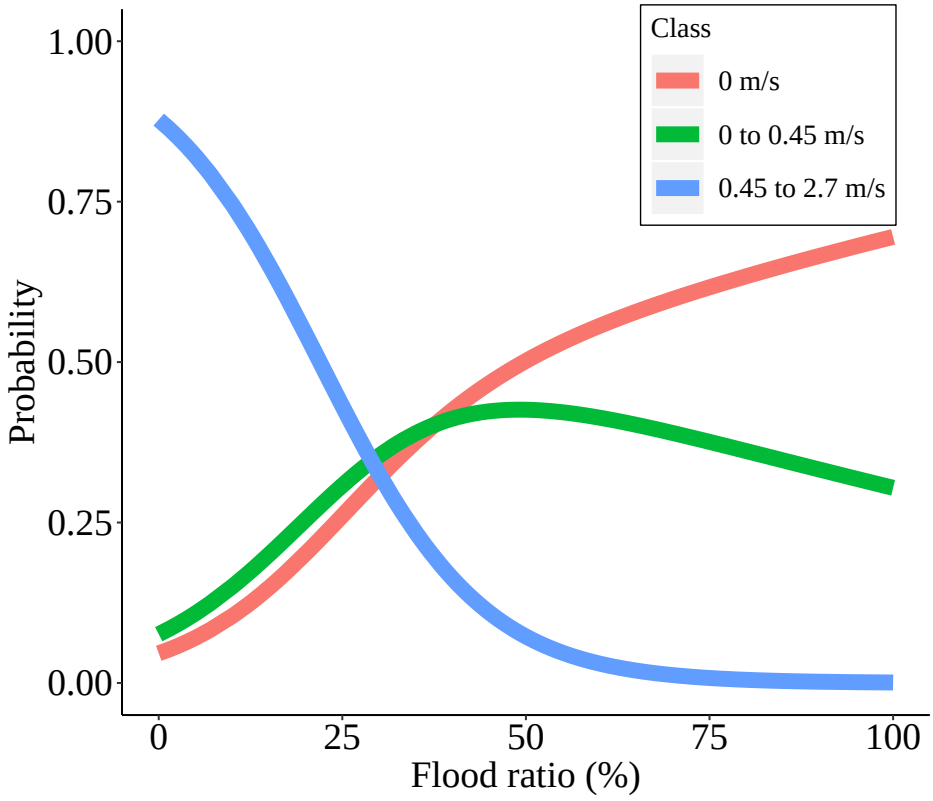

Figure 5: Output of the most parsimonious multinomial model with as response the three categories: 0, 0 to 0.45, 0.45 to 2.7 m s^{-1} and as predictor the flood ratio. The probability of each class is given as a function of the flood ratio.

417 However, distinction between swimming speeds of 0 and 0 to 0.45 m s^{-1}

was necessary in order to fit a generalized linear model with gamma distribution through the data. Using a binomial model with contrasts <0.45 and >0.45 m s^{-1} followed by two linear models yielded a lower predictive performance (RMSE = 0.3712) and would have violated model assumptions. The multinomial model on its own provided a relatively low predictive performance (RMSE=0.3940), but addition of a generalized linear model with gamma distribution from 0 to 0.45 m s^{-1} and a linear model from 0.45 to 2.7 m s^{-1} increased the predictive performance with 7.3 % (RMSE=0.3653). The gamma model from 0 to 0.45 m s^{-1} indicated a significant negative effect of flood ratio. However, it should be noted that the model fit was relatively poor as using a null model instead decreased the overall predictive performance with only 1.1 % (RMSE = 0.3693). More benefit was gained from the linear model for the part of 0.45 to 2.7 m s^{-1} as its omission reduced overall predictive performance with 6.7 % (RMSE = 0.3898). The flood ratio and moon phase had a significantly negative and positive effect on the swimming speed, respectively, but were found to be far less important than the significant positive effect of the distance to source. During ebbing tide, eels closer to the North Sea had relatively higher swimming speeds. Finally, all full model-parts of this three-part model were offered the variable eel ID as fixed factor in the model selection process, but it was only retained in the latter linear model from 0.45 to 2.7 m s^{-1}. This suggests that individual differences were important to predict swimming speeds from 0.45 to 2.7 m s^{-1}, but not to distinguish between classes (1) 0, (2) 0 to 0.45 and (3) 0.45

441 to 2.7 m s^{-1} or to predict the swimming speed from 0 to 0.45 m s^{-1}.

442 Similar predictors with reliable parameter estimates were retained in the
443 different models (Table 2). For the binomial part of the two-part models
444 only flood ratio was retained, while the one-part models and linear parts of
445 the two-part models retained, in order of decreasing importance, the factors
446 flood ratio, distance to source and moon phase. The variable importance
447 provided by the conditional inference random forests indicated that flood ra-
448 tio (0.3733) was most important, followed by distance from source (0.0609),
449 moon phase (0.0218), tidal phase (0.0204) and day phase (0.00796). The
450 conditional inference random forests performed better (2.8 %) than the best
451 statistical model when considering upstream movement intervals, but per-
452 formed slightly worse (0.4 %) than the best statistical model when upstream
453 movement intervals were not considered.

454 RMSE values of the one-part and two-part mixed models for the data
455 set without upstream movement intervals were 0.373 and 0.347 respectively.
456 Hence, the two-part mixed model explained patterns in the data 7.0 % better
457 than the one-part mixed model.

458 4. Discussion

459 *4.1. Evaluating one-part, two-part and three-part (mixed) models*

460 Movement decisions have been assessed in depth for a wide range of ani-
461 mals (Berdahl et al., 2017; Dechmann et al., 2017; O'Neal et al., 2018), but
462 the number of studies combining movement decisions with movement inten-

sity, e.g. swimming speed or distance covered, has been limited (Brodersen et al., 2008). Because zero values describe a unique behavioral aspect in movement behavior, i.e. residencies, defining observed zeros and identifying cryptic zeros in telemetry data sets allowed to improve predictive performance and to obtain more detailed ecological insights. The predictive performances of the original three-part and two-part models were higher (between 2.2 and 9.8 %) than those of the one-part models, suggesting that the conditions that affect the movement decision are not necessarily the same as the conditions that affect the movement intensity. Taking into consideration that both processes might be correlated did not improve predictions as the selection models had a lower predictive performance. This is in concordance with many econometric studies in which accounting for potential dependencies between both parts of the model did not seem to add to the quality of the predictions (Madden, 2008; Smith, 2003).

Although distinguishing between movement and residencies provided clearly better predictions, further distinction between upstream and downstream movement only provided marginally better predictive performances (0.4 %). This might be because of the limited amount of upstream movement intervals and the limited amount of individuals exhibiting upstream movement, causing only a limited increase in explanatory power in the test set. However, the poor gain in explanatory power of the model may also be the result of the similar conditions in which upstream movement and residencies occurred. Indeed, considering upstream movement as residencies gone wrong,

resulted in a 6.6 % and 2.7 % increase in performance for the two-part and one-part model respectively. This suggests that some eels are unsuccessful in remaining resident during flooding as they are pushed back, or that they mistake flooding for ebbing when moving along with the current. A final improvement of model performance was apparent from further compartmentalization. Distinction between swimming speeds of (1) 0, (2) 0 to 0.45 and (3) 0.45 to 2.7 m s^{-1} caused predictions of swimming speed to be 9.8 % better. This model improvement was mainly the result of the contrasting tidal conditions before and after 0.45 m s^{-1}, with eels facing or not facing a flooding event respectively. Hence, compartmentalization was successful because it adequately classified observed residencies (0 m s^{-1}), cryptic residencies (0 to 0.45 m s^{-1}) and movement intervals (0.45 to 2.7 m s^{-1}).

The results of the three-part model suggest that the movement decision depends only on the tides, while the swimming speed is dependent on the tides and the distance from source. The larger the contribution of flood, the more likely a specific time lapse will be a residency interval rather than a movement interval. In addition, eels which migrated during ebb tide and which were already close to the sea, typically had the highest swimming speed. The conditions during which the movement intervals of the first peak of the bimodal pattern ($<$0.45 m s^{-1}) occurred were actually more closely related to those of residency intervals than those of movement intervals of the second peak of the bimodal pattern ($>$0.45 m s^{-1}). Within the observed movement intervals characterized by a swimming speed below 0.45 m

509 s^{-1}, cryptic or undetected residencies were invoked by flooding events. Dur-
510 ing these flooding events, eels had to interrupt their journey, causing lower
511 observed swimming speeds. For swimming speeds above 0.45 m s^{-1}, the dis-
512 tance to the North Sea seemed to play a more important role than the tides.
513 In addition, individual variation was significantly more important for swim-
514 ming speeds above than below 0.45 m s^{-1} and also the movement decision
515 did not show any significant individual variation. This suggests that all eels
516 stay resident during flood, but also that some eels swim faster or slower than
517 others once the decision to continue their migration has been made. The
518 simple position estimates of a single individual would have made it difficult
519 to classify apparent non-movement as either (i) residencies or (ii) movement
520 without net gain in distance covered. However, the ability to quantify indi-
521 vidual variation from a large number of tagged individuals provided evidence
522 in favor of the first option. More specifically, as there were clearly faster
523 and slower swimming individuals, the second option would have resulted in
524 meaningful differences between individuals across all parts of the model (i.e.
525 some individuals would be pushed back while others would advance during
526 flood). This was, however, not the case.

527 One major advantage of telemetry is its ability to provide data on the
528 level of individuals, and therefore mixed models that account for individual
529 correlation are commonly used (Gillies et al., 2006; Hooten et al., 2017).
530 Two-part and three-part models can be easily extended to include mixed
531 effects in order to provide a higher explanatory power. In this study, the

explanatory power of mixed two-part models was 7.0 % higher than their one-part equivalents. However, it should be noted that potential dependencies between the elements of random and fixed factors across the different parts were not considered. If correlation between the random effects across the different parts is expected, a joint maximization of the likelihood functions would be required. More research is needed to evaluate the added value of such an approach as its importance is likely to be case-specific.

Eels have already been shown to exhibit selective tidal stream transport (STST), as they make use of the tides to reach their destination with as little energy expenditure as possible (Barry et al., 2016; Verhelst et al., 2018). However, by comparing one-part with two-part and three-part models, we illustrated that migrating fish exhibit complex behaviour and that models initially constructed to assess human customer behavior, might also be of use to study other animals (Farewell et al., 2017).

4.2. Statistical models versus machine learning

Statistical models are generally preferred over machine learning when the number of available predictors is limited and the main purpose is to infer ecological knowledge, while the contrary is true if predictive performance is deemed more important than inference. Since researchers often seek to optimize both knowledge and predictions, a mutually exclusive approach should be avoided. In this study we started off with a simple linear regression (i.e. one-part model), then moved further to a two-part model which combined a

554 binomial regression with linear regression, and finally ended up with a three-
555 part model which combined a multinomial model (via neural networks, i.e.
556 machine learning), generalized linear regression with gamma distribution and
557 linear regression. Because each step of the model improvement was supported
558 by ecological knowledge, i.e. being aware that the conditions that cause eels
559 to reside or to move might be different, and methodological considerations,
560 i.e. residencies taking place between gates are not directly observed but do
561 cause a travel delay, the final three-part model remained interpretable. The
562 conditional inference random forests provided similar results, though less in-
563 formative, and had only slightly higher or lower predictive performances than
564 the developed three-part models. Hence, appreciating the potential complex-
565 ity of animal behaviour and awareness towards the statistical framework that
566 machine learning algorithms are built upon, will provide researchers with the
567 best machine learning has to offer without compromising the lessons learnt
568 from statistical models.

569 *4.3. Recommendations for future studies*

570 In order for zero values to provide useful information, a good under-
571 standing of the meaning of zeros in the data is required. In this study we
572 considered all observed zeros to be true zeros, which is a plausible assumption
573 given the high detection probability of the network and mainly unidirectional
574 movement of migrating eel. In contrast, in case detection probabilities are
575 low, many zero values might actually be false zeros as the result of impor-

tant design and/or observer errors, and hence the probability of a false zero should be explicitly integrated in the model. Since the detection probability is affected by the network design, transmission intervals and detection range, which in turn is affected by environmental conditions (Reubens et al., 2019b), an elaborate addition to the two-part models may be required to deal with high levels of false zeros. In addition, a good understanding of the detection range variability is also necessary to estimate any difference between the observed and actual biological response. For instance, in this study, the observed swimming speed of eel likely differed from the actual swimming speed because of the unaccounted detection range variability. Furthermore, the factors known to affect the detection range, i.e. tides (Mathies et al., 2014), also seem to be affecting the movement behaviour of eel, introducing not only noise but even a potential bias in the data. Independent range tests at different locations along the estuary and at different moments within the tidal cycle are a necessary addition to quantify the noise and/or bias associated with detection range variability (Kessel et al., 2014).

It should also be noted that some limitations are inherent to the applied technique of passive telemetry and can only be resolved by additional data collection. For example, when eels are between gates and there seem to be travel delays during flood, apparent from reduced swimming speed, it is difficult to tell whether eels (i) remained stationary near the bottom to preserve energy or (ii) swam against the currents without much gain in distance covered. Although the constructed models indicated that the first

option is much more likely than the second, depth profiles and actual swimming speed measurements, obtained through archival tags with depth sensors and accelerometers, would provide more direct estimates of specific animal behavior and would allow to validate the results of this study.

5. Conclusion

In this study we illustrated how accounting for both well-defined and cryptic residencies provides a better insight into the movement behaviour of migrating eel. Two-part and three-part models turned out to be promising tools to deal with zero-inflated telemetry data, underlining the complex behaviour of migrating fish. Nevertheless, a sound assessment of the detection range variability in combination with more fine-scale measurements of environmental variables, is necessary in order to confirm the observed patterns in eel movement and its relationship with environmental variables. Although only data from one species, one telemetry network and one telemetry technique was used, the proposed model framework can be used for study cases with other species, networks and techniques (e.g. passive integrated transponder and radio telemetry).

Acknowledgements

This work was supported by the Flemish branch of the LifeWatch ESFRI observatory. P. Verhelst acknowledges the support of the Flemish Agency for Innovation and Entrepreneurship (VLAIO), now under the auspices of the

National Science Fund FWO, during a large part of this study. R. Baeyens, N. De Maerteleire, S. Franquet, E. Gelaude, T. Lanssens, S. Pieters, K. Robberechts, T. Saerens, R. van der Speld, S. Vermeersch and Y. Verzelen assisted with the data collection. B. Lonneville aided in the creation of the map. This work makes use of data and infrastructure provided by VLIZ and INBO and funded by Research Foundation - Flanders (FWO) as part of the Belgian contribution to LifeWatch. We would also like to thank the Royal Belgian Institute of Natural Sciences, Operational Directorate Natural Environment (RHIB Tuimelaar) for infrastructure provision and Rijkswaterstaat (The Netherlands) for their cooperation and the permission to use their marine buoys. This research has benefitted from a statistical consult with Ghent University FIRE (Fostering Innovative Research based on Evidence)

Authors' contributions

S.B. conceived the ideas and designed methodology, analyzed the data and led the writing of the manuscript; P.V., J.R. and S.B. collected the data; All authors contributed critically to the drafts and gave final approval for publication.

Data accessibility

The R code, subset of data and documentation are available on Mendeley Data: http://dx.doi.org/10.17632/vtrxw2m9wp.1

6. References

References

Arnold, G. P., Cook, P. H., 1984. Fish Migration by Selective Tidal Stream Transport: First Results with a Computer Simulation Model for the European Continental Shelf. In: McCleave, J. D., Arnold, G., Dodson, J. J., Neill, W. H. (Eds.), Mechanisms of Migration in Fishes. Springer, Boston, pp. 227–261.

Austin, P. C., Tu, J. V., 2004. Bootstrap Methods for Developing Predictive Models. American Statistician 58 (2), 131–137.

Barry, J., Newton, M., Dodd, J. A., Lucas, M. C., Boylan, P., Adams, C. E., 2016. Freshwater and coastal migration patterns in the silver-stage eel Anguilla anguilla. Journal of Fish Biology 88 (2), 676–689.

Belotti, F., Deb, P., Manning, W. G., Norton, E. C., 2015. twopm: Two-part models. Stata Journal 15 (1), 3–20.

Berdahl, A., Westley, P. A., Quinn, T. P., 2017. Social interactions shape the timing of spawning migrations in an anadromous fish. Animal Behaviour 126, 221–229.

Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., Castells, E., 2019. What does a zero mean? Understanding false, random and structural zeros in ecology. Methods in Ecology and Evolution 10, 949–959.

660 Breukelaar, A. W., Ingendahl, D., Vriese, F. T., De Laak, G., Staas, S.,
661 Klein Breteler, J. G. P., 2009. Route choices, migration speeds and daily
662 migration activity of European silver eels Anguilla anguilla in the River
663 Rhine, north-west Europe. Journal of Fish Biology 74 (9), 2139–2157.

664 Brodersen, J., Nilsson, P. A., Ammitzbøll, J., Hansson, L. A., Skov, C.,
665 Brönmark, C., 2008. Optimal swimming speed in head currents and effects
666 on distance movement of winter-migrating fish. PLoS ONE 3 (5), 1–7.

667 Brownscombe, J., Lédée, E., Graham, R., Struthers, D., Gutowsky, L. F. G.,
668 Nguyen, V. M., Young, N., Stokesbury, M. J. W., Holbrook, C. M., Bren-
669 den, T. O., Vandergoot, C. S., Murchie, K. J., Whoriskey, K., Mills, J.,
670 Steven, F., Krueger, C. C., Cooke, S. J., 2019. Conducting and interpret-
671 ing fish telemetry studies : considerations for researchers and resource
672 managers. Reviews in Fish Biology and Fisheries 29, 369–400.

673 Bruneel, S., Verhelst, P., Reubens, J., Baetens, J. M., Coeck, J., Moens, T.,
674 Goethals, P., 2020. Quantifying and reducing epistemic uncertainty of pas-
675 sive acoustic telemetry data from longitudinal aquatic systems. Ecological
676 Informatics 59.

677 Bultel, E., Lasne, E., Acou, A., Guillaudeau, J., Bertier, C., Feunteun, E.,
678 2014. Migration behaviour of silver eels (Anguilla anguilla) in a large estu-
679 ary of Western Europe inferred from acoustic telemetry. Estuarine, Coastal
680 and Shelf Science 137 (1), 23–31.

681 Bzdok, D., Altman, N., Krzywinski, M., 2018. Points of Significance: Statis-
682 tics versus machine learning. Nature Methods 15 (4), 233–234.

683 Cagua, E. F., Cochran, J. E. M., Rohner, C. A., Prebble, C. E. M., Sinclair-
684 taylor, T. H., Pierce, S. J., Berumen, M. L., 2015. Acoustic telemetry
685 reveals cryptic residency of whale sharks. Biology letters 11.

686 Cornet, E., Vereecken, H., Deschamps, M., Verwaest, T., Mostaert, F., 2016.
687 Hydrologisch jaarboek 2016. Tech. rep., Waterbouwkundig Laboratorium,
688 Antwerpen.

689 Davison, A. C., Hinkley, D. V., 1997. Bootstrap Methods and their Applica-
690 tion.

691 Deb, P., Holmes, A., 2002. Estimates of Use and Costs of Behavioural Health
692 Care: a comparison of Standard and Finite Mixture Models. In: Jones, A.,
693 O'Donnell, O. (Eds.), Econometric Analysis of Health Data. John Wiley
694 & Sons, Ch. 6.

695 Dechmann, D., Wikelski, M., Ellis-Soto, D., Safi, K., O'Mara, M., 2017. De-
696 terminants of spring migration departure decision in a bat. Biology letters
697 13 (9), 1–5.

698 Durif, C., Dufour, S., Elie, P., 4 2005. The silvering process of Anguilla an-
699 guilla: a new classification from the yellow resident to the silver migrating
700 stage. Journal of Fish Biology 66 (4), 1025–1043.

Farewell, V. T., Long, D. L., Tom, B. D. M., Yiu, S., Su, L., 2017. Two-Part and Related Regression Models for Longitudinal Data. Annual Review of Statistics and Its Application 4, 283–315.

Gillies, C. S., Hebblewhite, M., Nielsen, S. E., Krawchuk, M. E. G. A., Aldridge, C. L., Jacqueline, L., Saher, D. J., Stevens, C. E., Jerde, C. L., 2006. Application of random effects to the study of resource selection by animals. Journal of Animal Ecology 75, 887–898.

Glebe, B. D., Leggett, W. C., 1981. Temporal, Intra-population Differences in Energy Allocation and Use by American Shad (Alosa sapidissima) During the Spawning Migration . Canadian Journal of Fisheries and Aquatic Sciences.

Heckman, J., 1979. Sample Specification Bias as a Selection Error. Econometrica 47, 153–161.

Hooten, M. B., Johnson, D. S., McClintock, B. T., Morales, J. M., 2017. Animal Movement: Statistical Models for Telemetry Data, 1st Edition. CRC Press.

Joseph, J., Torney, C., Kings, M., Thornton, A., Madden, J., 2017. Applications of machine learning in animal behaviour studies. Animal Behaviour 124, 203–220.

Joseph, L. N., Conservancy, A. W., Elkin, C. M., Martin, T. G., 2009. Modeling abundance using N-mixture models : The importance of considering

722 Modeling abundance using N -mixture models : the importance of consid-
723 ering ecological mechanisms (May 2019).

724 Kessel, S. T., Cooke, S. J., Heupel, M. R., Hussey, N. E., Simpfendorfer,
725 C. A., Vagle, S., Fisk, A. T., 2014. A review of detection range testing in
726 aquatic passive acoustic telemetry studies. Reviews in Fish Biology and
727 Fisheries 24 (1), 199–218.

728 Kraus, R. T., Holbrook, C. M., Vandergoot, C. S., Stewart, T. R., Faust,
729 M. D., Watkinson, D. A., Charles, C., Pegg, M., Enders, E. C., Krueger,
730 C. C., 2018. Evaluation of Acoustic Telemetry Grids for Determining
731 Aquatic Animal Movement and Survival. Methods in Ecology and Evo-
732 lution, 1–14.

733 Levy, Y., Plancke, Y., Peeters, P., Taverniers, E., Mostaert, F., 2014. Het
734 getij in de Zeeschelde en haar bijrivieren: langjarig overzicht van de voor-
735 naamste getijkarakteristieken. Tech. rep., Waterbouwkundig Laboratorium
736 (Antwerpen, België).

737 Madden, D., 2008. Sample selection versus two-part models revisited: The
738 case of female smoking and drinking. Journal of Health Economics 27 (2),
739 300–307.

740 Mathies, N. H., Ogburn, M. B., McFall, G., Fangman, S., 2014. Environ-
741 mental interference factors affecting detection range in acoustic telemetry

742 studies using fixed receiver arrays. Marine Ecology Progress Series 495,
743 27–38.

744 Melnychuk, M., 2012. Detection efficiency in telemetry studies: definitions
745 and evaluation methods. In: Adams, N., Beeman, J., Eiler, J. (Eds.),
746 Telemetry techniques: A user guide for fisheries research. American Fish-
747 eries Society, pp. 339–357.

748 Metcalfe, J. D., Arnold, G. P., Webb, P. W., 1990. The energetics of migration
749 by selective tidal stream transport: An analysis for plaice tracked in the
750 southern north sea. Journal of the Marine Biological Association of the
751 United Kingdom.

752 Neelon, B., O'Malley, A. J., Smith, V. A., 2016. Modeling zero-modified
753 count and semicontinuous data in health services research Part 1: back-
754 ground and overview. Statistics in Medicine 35 (27), 5070–5093.

755 O'Neal, B. J., Stafford, J. D., Larkin, R. P., Michel, E. S., 2018. The effect
756 of weather on the decision to migrate from stopover sites by autumn-
757 migrating ducks. Movement Ecology 6 (1), 1–9.

758 Perry, R. W., Castro-Santos, T., Holbrook, C. M., Sandford, B. P., 2012.
759 Using Mark-Recapture Models to Estimate Survival from Telemetry Data.
760 In: Adams, N., Beeman, J., Eiler, J. (Eds.), Telemetry Techniques: A User
761 Guide for Fisheries Research. American Fisheries Society, pp. 453–475.

762 Pohlmeier, W., Ulrich, V., 1995. An econometric model of the two-part de-
763 cisionmaking process in the demand for healthcare. Journal of Human
764 Resources 30 (2), 339–361.

765 R Core Team, 2019. R: A Language and Environment for Statistical Com-
766 puting.

767 Reubens, J., Verhelst, P., Knaap, I. V. D., Wydooghe, B., Milotic, T.,
768 Deneudt, K., Hernandez, F., Pauwels, I., 2019a. The need for aquatic
769 tracking networks : the Permanent Belgian Acoustic Receiver Network.
770 Animal Biotelemetry 7 (2), 1–6.

771 Reubens, J., Verhelst, P., van der Knaap, I., Deneudt, K., Moens, T., Hernan-
772 dez, F., 2019b. Environmental factors influence the detection probability
773 in acoustic telemetry in a marine environment: results from a new setup.
774 Hydrobiologia 845, 81–94.

775 Smith, M. D., 2003. On dependency in double-hurdle models. Statistical
776 Papers 44 (4), 581–595.

777 Steckenreuter, A., Hoenner, X., Huveneers, C., Simpfendorfer, C., Bus-
778 cot, M. J., Tattersall, K., Babcock, R., Heupel, M., Meekan, M., Van
779 Den Broek, J., McDowall, P., Peddemors, V., Harcourt, R., 2017. Opti-
780 mising the design of large-scale acoustic telemetry curtains. Marine and
781 Freshwater Research 68 (8), 1403–1413.

Thorstad, E. B., Rikardsen, A. H., Alp, A., Okland, F., 2013. The Use of Electronic Tags in Fish Research - An Overview of Fish Telemetry Methods. Turkish Journal of Fisheries and Aquatic Sciences 13, 881–896.

Tukey, J. W., 1977. Exploratory Data Analysis. Pearson.

Venables, W. N., Ripley, B. D., 2002. Modern Applied Statistics with S, 4th Edition. Springer, New York.

Verbiest, H., Breukelaar, A., Ovidio, M., Philippart, J. C., Belpaire, C., 2012. Escapement success and patterns of downstream migration of female silver eel Anguilla anguilla in the River Meuse. Ecology of Freshwater Fish 21 (3), 395–403.

Verhelst, P., Bruneel, S., Reubens, J., Coeck, J., Goethals, P., Oldoni, D., Moens, T., Mouton, A., 2018. Selective tidal stream transport in silver European eel (Anguilla anguilla L.) – Migration behaviour in a dynamic estuary. Estuarine, Coastal and Shelf Science 213, 260–268.

Wang, G., 2019. Machine learning for inferring animal behavior from location and movement data. Ecological Informatics 49, 69–76.

Warton, D. I., 2005. Many zeros does not mean zero inflation: Comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics 16 (3), 275–289.

Whoriskey, K., Martins, E. G., Auger-Méthé, M., Gutowsky, L. F., Lennox, R. J., Cooke, S. J., Power, M., Mills Flemming, J., 2019. Current and

emerging statistical techniques for aquatic telemetry data: A guide to analysing spatially discrete animal detections. Methods in Ecology and Evolution 10 (7), 935–948.

Wright, M. N., Ziegler, A., 2014. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software 77 (1), 1–17.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., Smith, G. M., 2009. Mixed Effects Models and Extensions in Ecology with R, 1st Edition. Springer-Verlag New York, New York.

Appendix A. Tagging procedure

The following description is adopted from Verhelst et al. (2018). 100 Eels were caught and tagged at the tidal weir in Merelbeke in the Zeeschelde during late summer and autumn (September–November) of three consecutive years (2015 till 2017) using double fyke nets. After periods of heavy rain, water flows over the sluices allowing eels to swim over the sluices. Placing the fyke nets behind the sluices and during periods of heavy rain, allowed to coordinate capture events and improve the chance of capturing eel. Several morphometric features were measured in order to determine the eel maturation stage (Durif et al., 2005): Total length (TL, to the nearest mm), body weight (W, to the nearest g), the vertical and horizontal eye diameter (EDv and EDh respectively, to the nearest 0.01 mm) and the length of the pectoral fin (FL, to the nearest 0.01 mm) (Table A.1). Only females were tagged, since males are smaller than the minimum size handled in this study ($<$ 450 mm (Durif et al., 2005)). Eels of three different maturation stages were tagged: premigrant (F3, n = 51) and the two migrant stages F4 and F5 (n = 21 and n = 28, respectively). The eels were tagged with V13 coded acoustic transmitters (13 x 36 mm, weight in air 11 g, frequency 69 kHz, ping frequency: 60–100 s; estimated battery life: 1021-1219 days (battery life time depended on specific transmitter settings), (Table A.2)) from VEMCO Ltd (Canada). After anaesthetizing them with 0.3 ml/L clove oil, tags were implanted with permanent monofilament (Thorstad et al., 2013). Eels recovered in a quarantine reservoir for approximately one hour and were subsequently released

Stage	Number	TL (mm)	BW (g)	EDh (mm)	EDv (mm)	FL (mm)
F3	51	702±57 (568 - 835)	674±165 (324 - 1106)	8.08±0.57 (6.77 - 9.08)	7.55±0.60 (6.20 - 9.70)	32.92±3.29 (26.76 - 40.32)
F4	21	810±57 (707 - 932)	1162±217 (771 - 1830)	10.41±0.92 (9.13 - 12.49)	9.66±0.78 (8.60 - 11.86)	40.86±4.32 (30.84 - 48.18)
F5	28	662±56 (575 - 775)	585±144 (417 - 912)	9.33±0.80 (8.14 - 11.18)	8.80±0.79 (7.62 - 10.39)	34.41±3.68 (28.97 - 45.37)

Table A.1: Number of all tagged female eels per stage with the different morphometrics: total length (TL), body weight (BW), horizontal and vertical eye diameters (EDh and EDv, respectively) and pectoral fin length (FL). Mean, standard deviation and range (between brackets) are indicated (Adopted from Verhelst et al. (2018)).

Number of transmitters	Step 1			Step 2			Battery life (days)
	PO	Ping frequency (s)	Duration (days)	PO	Ping frequency (s)	Duration (days)	
20	L	60 - 100	1216	NA	NA	NA	1216
40	H	60 - 100	120	L	60 - 100	901	1021
40	H	60 - 100	120	L	60 - 100	902	1022

Table A.2: The number and settings of the transmitters of all tagged eels per step: power output (PO; L = low power output, H = high power output), ping frequency (s) and the time duration (days) per step as well as the total battery life time (days). (Adopted from Verhelst et al. (2018))

at the nearest receiver.

Appendix B. Telemetry network

Appendix C. Figures

gate name	Distance (km)	Deployment date	Number of receivers	Receiver inactivity	Det. prob. (%)
s1	0.0	31/03/2015	1		100.0
s2	6.6	20/03/2016	1		100.0
s3	12.1	20/03/2016	1		97.1
s4	16.8	20/04/2015	1		97.4
s5	26.7	31/03/2015	1		99.1
s6	30.6	2/04/2015	1		98.7
s7	33.0	24/03/2016	1	17/10/2017 - 24/11/2017	96.7
s8	39.3	24/03/2016	1		81.6
s9	40.8	20/04/2015	1		99.9
s10	44.1	20/04/2015	1		99.3
s11	46.5	27/04/2015	1		100.0
s12	49.0	2/04/2015	1		98.4
s13	53.8	2/04/2015	1		93.2
s14	55.6	2/04/2015	1		100.0
s15	63.3	2/04/2015	2		100.0
s16	68.6	2/04/2015	2		100.0
s17	75.8	30/09/2015	3		100.0
s18	88.2	3/09/2015	2		77.8
ws1	112.8	22/09/2015	6		91.3

Table B.1: List of gates, with distance from Ghent (km), deployment date, number of included receivers, period of receiver inactivity and detection probability. Receiver inactivity represents the period during which one receiver of the gate was inactive. Adapted from Bruneel et al. (2020).

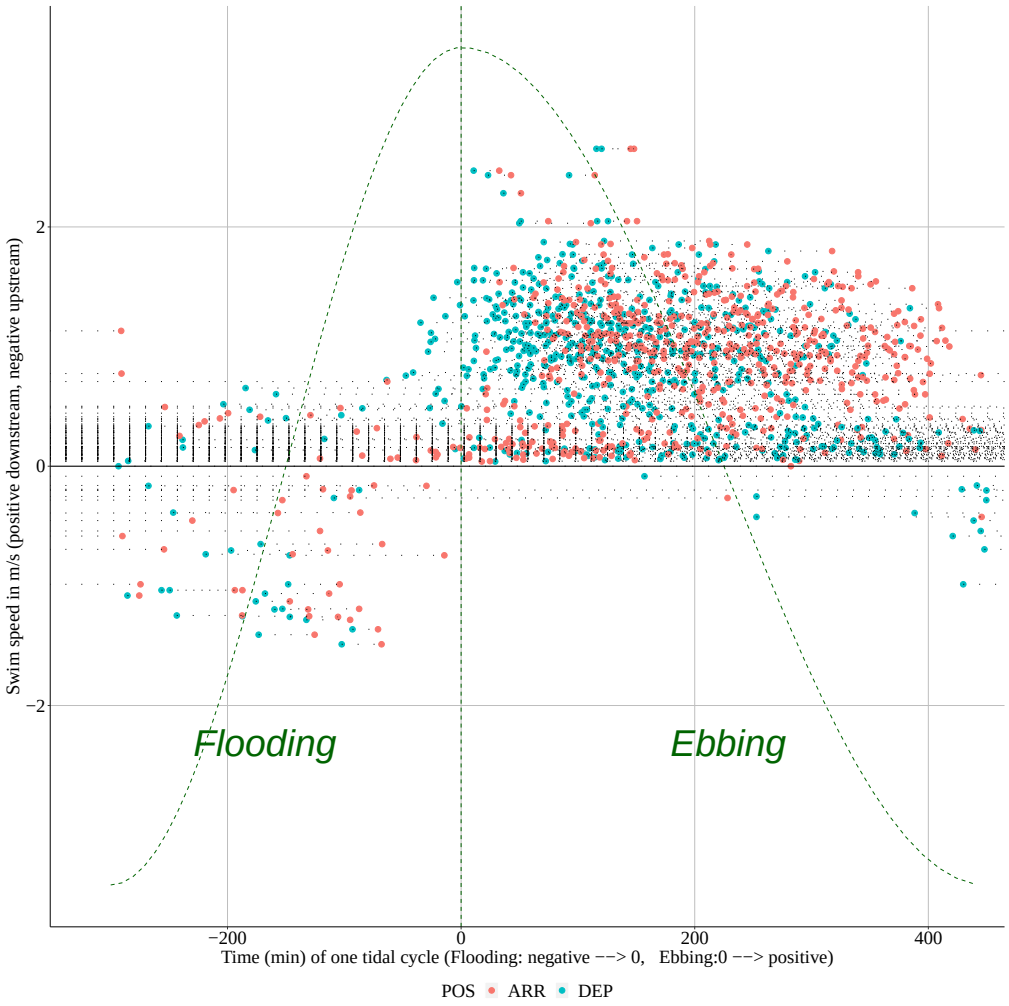

Figure C.1: Movement intervals of all tagged eels depicted by the departure (DEP) from a receiver and arrival (ARR) at another receiver. The swimming speed (m s^{-1}) during a movement interval is given in function of the moment within the tidal cycle. In the ZS, the period of ebbing is larger than the period of flooding, with differences being most pronounced upstream. However, for visualization purposes the average period of flooding (300 minutes) and period of ebbing (450 minutes) of the city of Dendermonde (in the center of the ZS) were used to rescale the TMIs (Levy et al., 2014)