

Analysis of local habitat selection and large-scale attraction/avoidance based on animal tracking data: is there a single best method?

Moritz Mercker (✉ info@bionum.de)

BIONUM <https://orcid.org/0000-0002-7685-6193>

Philipp Schwemmer

Christian-Albrechts-Universität zu Kiel

Verena Peschko

Christian-Albrechts-Universität zu Kiel

Leonie Enners

Christian-Albrechts-Universität zu Kiel

Stefan Garthe

Christian-Albrechts-Universität zu Kiel

Research

Keywords: animal movement, autocorrelation, bio-logging, habitat selection, point process, resource selection, species distribution, telemetry, avoidance

Posted Date: April 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-21556/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Movement Ecology on April 23rd, 2021. See the published version at <https://doi.org/10.1186/s40462-021-00260-y>.

RESEARCH

Analysis of local habitat selection and large-scale attraction/avoidance based on animal tracking data: is there a single best method?

Moritz Mercker^{1*}, Philipp Schwemmer², Verena Peschko², Leonie Enners² and Stefan Garthe²

*Correspondence: info@bionum.de

¹BIONUM statistical consulting office, Hamburg, Finkenwerder Norderdeich 15 A, Hamburg, Germany

Full list of author information is available at the end of the article

[†]Equal contributor

Abstract

Background: New wildlife telemetry and tracking technologies have become available in the last decade, leading to a large increase in the volume and resolution of animal tracking data. These technical developments have been accompanied by various statistical tools aimed at analysing the data obtained by these methods.

Methods: We used simulated habitat and tracking data to compare some of the different statistical methods frequently used to infer local resource selection and large-scale attraction/avoidance from tracking data. Notably, we compared the performances of spatial logistic regression models (SLRMs), point process models (PPMs), and integrated step selection models ((i)SSMs) and their interplays with habitat, tracking-device, and animal movement properties.

Results: We demonstrated that SLRMs were inappropriate for large-scale attraction studies and prone to bias when inferring habitat selection. In contrast, PPMs and (i)SSMs showed comparable (unbiased) performances for both habitat selection and large-scale effect studies. However, (i)SSMs had several advantages over PPMs with respect to robustness, user-friendly implementation, and computation time.

Conclusions: We recommend the use of (i)SSMs to infer habitat selection or large-scale attraction/avoidance from animal tracking data. This method has several practical advantages over PPMs and additionally extends SSMs, thus increasing its predictive capacity and allowing the derivation of mechanistic movement models.

Keywords: animal movement; autocorrelation; bio-logging; habitat selection; point process; resource selection; species distribution; telemetry; avoidance

1 Introduction

The identification of factors that influence species distribution and resource selection is an important ecological issue [1] that has traditionally been addressed using appropriate regression methods based on presence-absence or count data [2, 3, 4]. However, recent technical developments involving radio- and telemetry-based approaches (e.g., Argos-, global positioning system-, or very-high frequency-based methods [5, 6, 7, 8]) allow single-animal tracks to be recorded, thus providing a new and highly valuable alternative source of information to help answer such questions. These approaches allow the locations and movements of species to be analysed at a much finer spatio-temporal scale than previously possible.

12 Frequent ecological questions associated with animal tracks concern either the se-
13 lection/avoidance of a certain resource/habitat or structure, or alternatively changes
14 in behaviour related to such covariates. Both questions can again be applied at dif-
15 ferent spatial scales. On a small scale, the direct interaction of an animal with its
16 surroundings can be investigated; e.g., by evaluating if certain habitats are used
17 more intensively than others. On a large-scale however, certain areas or structures
18 may assert effects on animal movement, for example as a result of direct perception
19 (e.g., over the sea possible for $\geq 10 km$ [9]) or spatial memory [10, 11]. Animal
20 movement is often eventually shaped by a complex combination of all of the above-
21 mentioned types of perceptions and memories interacting at various spatio-temporal
22 scales [8, 12, 13, 14].

23
24 In contrast to count- or presence-absence data, tracking data only contain infor-
25 mation about single points in time and space where the animal is present, with
26 no information for any other locations [15]. The statistical analysis of such 'point
27 observation data' is often challenging, and various approaches have previously been
28 developed and discussed (e.g., summarized in [8]). In summary, there are two dis-
29 tinctly different approaches for the analysis of animal tracking data.

30
31 The first approach is given by animal movement models that rely solely on track-
32 ing data points, e.g., by evaluating step sizes and turning angles. Variations on these
33 models mainly differ from each other in the way in which time and space dependen-
34 cies, as well as latent behavioural states, are entered into the model [8, 16, 17, 18].
35 Indeed, one of the main strengths of these approaches is the analysis of animal
36 behaviour, possibly changing in short-range or large-scale interaction with the en-
37 vironment. 'Discrete-time hidden Markov models' are a prominent and increasingly
38 used example [17, 18, 19, 20, 21]. However, habitat selection cannot be directly as-
39 sessed using such movement models, given that they lack quantification of habitat
40 availability.

41
42 The second approach augments tracking points with an additional set of arti-
43 ficially created points ('pseudo-absences', 'dummy points', or 'available steps') to
44 quantify habitat availability. The tracking points thus represent the used habitats,
45 while dummy points are chosen optimally to measure/represent how much of each
46 habitat type is available. This approach is frequently used in the context of spatial
47 point observation data, and various sophisticated strategies have been used to select
48 appropriate pseudo-absences (e.g., [22, 23]), frequently followed by an analysis us-
49 ing spatial regression or machine learning techniques, such as MAXENT or spatial
50 logistic regression models (SLRMs) (e.g., [24, 25, 26, 27]).

51
52 In the context of animal tracking data, such 'dummy point approaches' must
53 however be used with care: first, the artificial generation of pseudo-absence points
54 is an *ad hoc* approach and thus associated with several statistical disadvantages
55 and criticisms: e.g., the choice of location and number of these points are often not
56 straightforward, even though the regression results may sensitively depend on these
57 choices [26, 27, 28]. Second, strong spatial, temporal, and angular autocorrelations

58 are frequent challenges presented by tracking data [29, 30]: a tracking point is most
59 likely to appear in spatial and temporal proximity to the previous point, and turn-
60 ing angles often correlate with their predecessors, inducing directional persistence.
61 The importance of taking account of autocorrelations increases with the sampling
62 frequency relative to the velocity of the animal [8], and neglecting these issues may
63 distinctly bias the results [31, 32]. The correct implementation of autocorrelation
64 in the context of ecological modelling is however complex: e.g., when considering
65 spatial autocorrelation based on mixed modelling, fixed and random spatially vary-
66 ing covariates may be collinear ('spatial confounding'), potentially causing biased
67 interference [33, 34, 35], especially if covariates and residual structures act on sim-
68 ilar spatial scales [36].

69
70 Two new methods have become available during the last decade, providing a
71 rational basis for the choice of dummy points on the one hand, and integrating ap-
72 propriate correlation structures on the other. First, point process models (PPMs)
73 naturally and automatically resolve many of the questions and pitfalls arising from
74 previous techniques [8, 26, 27, 37]. In particular, PPMs allow the role and number
75 of dummy points to be deduced purely mathematically by aiming an efficient esti-
76 mation of an integral as a part of the PPM likelihood [26, 37]. Additionally, PPMs
77 represent a generalisation of many other frequently used methods [26, 38, 39],
78 and the PPM likelihood can be approximated using standard generalised linear
79 modelling regression software [27, 40] (possibly including mixed and/or additive
80 modelling). The latter ensures a flexible and individual implementation, including
81 an appropriate integration of spatio-temporal autocorrelation [40]. The second is
82 step-selection models (SSMs) [41], which have been developed from the viewpoint
83 of an individual, in contrast to the population viewpoint adopted in PPMs and
84 related models [42]. SSMs have recently been combined with simultaneous animal
85 movement models estimating movement and resource selection parameters, leading
86 to integrated step selection analysis (iSSM) methods [43]. The above-mentioned
87 different viewpoints of PPM and (i)SSM approaches become noticeable e.g., when
88 looking at autocorrelations in the data. In (i)SSM methods, these dependencies
89 are used to create appropriate data stratification and to choose reasonable dummy
90 point locations, whereas in PPMs, these correlations are considered to be a statis-
91 tical nuisance [42, 43]. It is therefore not surprising that both methods can lead to
92 divergent conclusions [42]. Finally, all three methods (PPMs, SSMs, and (i)SSMs)
93 can be used not only to test hypotheses, but also to estimate utilization distribu-
94 tions respectively for predictions [8, 43, 44].

95
96 However, the variety of available statistical tools makes it difficult to select the
97 most appropriate one for analysing a particular set of tracking data and addressing
98 a specific research question. Although partial comparisons between the different ap-
99 proaches have been presented (e.g., [42, 43, 45]), a more comprehensive comparison
100 of the different frequently used methods in terms of their interplays with habitat,
101 movement and tracking-device properties has not, to the best of our knowledge, yet
102 been presented.

104 In the current study, we used simulated data to systematically analyse and compare
105 the performances of different statistical approaches (namely SLRMs, PPMs,
106 SSMS, and (i)SSMs) with respect to local resource/habitat selection, as well as
107 large-scale attraction/avoidance processes. We analysed and compared the statisti-
108 cal powers (i.e., rate of detecting an existent effect) and false-positive rates (type
109 I error rates) in the interplay with habitat, animal movement, and tracking-device
110 properties (Fig. 1). Based on these results, we provide practical guidance for the
111 conditions under which each of these methods works reliably.

112 2 Methods

113 2.1 Overview

114 To cover the potential variety of animal tracking data, we simulated a broad range of
115 reasonable animal tracking data, resulting from an interplay among various habitat
116 types, animal movement properties, and tracking-device outcomes. To investigate
117 the statistical power of different frequently used methods, different variations of
118 SLRMs, PPMs, and (i)SSMs were applied to these simulated tracking data, the
119 latter influenced by local habitat selection and/or large-scale attractions of varying
120 strength. In contrast, if the rate of false-positive findings (type I error rate) was
121 the focus, they were applied to simulated data with no underlying attraction effects.

122
123 The above framework produced four different data frames for final compara-
124 tive analyses: two data frames for the analysis of local habitat selection (with vs.
125 without underlying simulated attraction effect), and two data frames focussing on
126 large-scale attraction (also with vs. without underlying simulated attraction effect).
127 Each data frame consisted of approximately 7,000 rows nested in 400 different an-
128 imal tracks (i.e., each track has been analyzed with various methods), which we
129 considered to be an appropriate order of magnitude to detect practically relevant
130 differences between the compared methods. In particular, each row depicted one
131 unique method-track combination comprising information about (1) the statistical
132 method used, (2) several parameters related to habitat, movement, and tracking-
133 device properties (including the respective strengths of the above-mentioned two
134 attraction effects), and (3) the outcome variable (*effect*) defined by the binomial
135 response 'significant effect detected' vs. 'no significant effect detected' (where the
136 $\alpha = 0.05$ significance level has been used).

137
138 In a final step, these four final data frames were further analysed using appropri-
139 ate regression model selection techniques, e.g., to evaluate and compare the average
140 statistical powers and false-positive rates of the different methods in interplay with
141 the several predictors related to the simulated tracking and environmental data. For
142 the sake of clarity, in the following, variable names starting with *Meth...* concern
143 the applied statistical method of interest (i.e., a specific variation of a SLRM, PPM,
144 or (i)SSM approach), variables starting with *Hab...* represent habitat properties,
145 *Mov...*, variables define properties related to animal movement, and variables start-
146 ing with *Dev...* concern tracking-device properties. An overview of all considered
147 variables is given in Tab. 1. A graphical overview of the workflow applied in this
148 study is given in Fig. 1.

149 2.2 Computation and software

150 All statistical analyses were performed using the open source software R [46]. In
 151 particular, spatial methods and spatial visualizations were mainly based on the
 152 package *raster* [47], and all other statistical plots were performed using *ggplot2*
 153 [48]. Additional R-packages and functions are detailed below. All computations
 154 were performed using a non-parallelized code on an Intel(R) 4-Core(TM) i5-6600
 155 with 3.30 GHz and 16 GB RAM.

156 2.3 Simulated data

157 2.3.1 Habitat data

158 Habitat raster data were generated based on the *raster()*-function in the R-package
 159 *raster* [47], where habitat values (given by the variable *hab*) were randomly assigned
 160 (uniformly distributed between 0 and 1) to $500 \times 500 = 250,000$ spatial 2D pixels
 161 $\vec{x} = (x_1, x_2)$. Possible spatial autocorrelation (i.e., local spatial clustering of certain
 162 habitat types) was quantified by the continuous variable $Hab_auto \in \{1, \dots, 60\}$,
 163 where a value of 1 represents no autocorrelation (i.e., the *hab* value for each pixel
 164 does not depend on the *hab* values in the surroundings) and values of $Hab_auto > 1$
 165 represent an increasing isotropic autocorrelation. In contrast, anisotropic autocorre-
 166 lation was represented by the continuous variable $Hab_anis \in [0.1, 1]$, where a value
 167 of 0.1 represents an isotropic landscape, up to strongly anisotropic landscapes with
 168 maximal values of $Hab_anis = 1.0$, the latter resulting in habitats stretched along
 169 the x-axis. Both types of autocorrelation were generated using a 'moving window'
 170 approach based on the R-function *focal()* in the *raster* package. Furthermore, the
 171 smoothness of the transition between different habitat types was defined by the cat-
 172 egorical variable *Hab_smooth*, which can generate sharp, medium, or smooth/blurry
 173 transitions, also realized in the context of the above-mentioned *focal()*-function. Fi-
 174 nally, after generating autocorrelation and transition smoothness, in approximately
 175 50 % of cases, the continuous variable *hab* was transformed into a binomial one
 176 (based on the threshold of 0.5) where the final type of *hab* was represented by
 177 the variable *Hab_type* differentiating between continuous and categorical habitat
 178 data. Some examples for simulated categorical habitat data (permuted over the
 179 variables *Hab_auto*, *Hab_anis* and *Hab_smooth*) are shown in Fig. S1, and further
 180 examples including continuous habitats are presented in Fig. S2 and Fig. 1.

181 2.3.2 Animal movement

182 Simulated animal movement is strongly related to (and extends) the 'stepping-stone'
 183 algorithm, as presented by Avgar *et al.* [43]. In particular, we simulated movement
 184 on the finest available spatial scale, i.e., on a scale of pixels. For each simulation,
 185 the starting location and the location of the attraction centre were chosen randomly
 186 within the virtual study area, and a maximum of 5,000 steps per trip was simulated,
 187 unless otherwise stated. For each simulated time step t with location $\vec{x} = (x_1, x_2)$,
 188 the probability of choosing the neighbouring pixel $\vec{y} = \vec{x}_{t+1}$ out of the eight nearest
 189 neighbours \mathcal{N} as the next point was given by

$$p(x_{t+1} = \vec{y}) = \frac{F(\vec{x})}{\sum_{\vec{y} \in \mathcal{N}} F(\vec{y})},$$

with

$$F(\vec{y}) = N(0, Mov_SD) + \exp\left(Mov_w \cdot hab(\vec{x}) - \mu \|\vec{y} - \vec{x}\| - Mov_s \cdot \alpha_{att} - f(Mov_ran, Mov_ran2) \cdot \alpha_{pers}\right),$$

190 where a more detailed motivation of the general structure of this stepping stone
 191 algorithm is given by Avgar *et al.* [43]. Here, $N(0, M_{SD})$ represents a normally
 192 distributed random component in attraction strength, i.e., scaling the strength of
 193 random movement vs. directed/biased movement in the animal path, quantified by
 194 the movement standard deviation $Mov_SD \in [0, 2.5]$. $Mov_w \in [0, 1.0]$ quantifies the
 195 strength of resource selection, while the term $\mu \|\vec{y} - \vec{x}\|$ with $\mu = 1.8$ penalizes larger
 196 Euclidean distances to neighbouring pixels. Furthermore, $Mov_s \in [0, 0.1]$ penal-
 197 izes angular deviations α_{att} from the direct path between \vec{z} and the attraction centre
 198 (and thus introduces directional bias towards the centre, 'biased random walk'
 199 [8]), and $f(Mov_ran, Mov_ran2)$ finally penalizes angular deviations α_{pers} from the
 200 direction of the foregoing movement step, thus leading to directional persistence
 201 ('correlated random walk' [8]). Notably, $f(Mov_ran, Mov_ran2) = Mov_ran / (1 +$
 202 $Mov_ran2 \cdot \overline{hab}(\vec{y}))$ includes the variable $Mov_ran \in [0, 2.5]$ for the general strength
 203 of directional persistence, but also the variable $Mov_ran2 \in [0, 1.0]$ antagonizing
 204 this effect if local habitat values (averaged over all neighbours) are high. The latter
 205 effect thus induced a less-directed and more-random search behaviour in appropriate
 206 habitats. To avoid boundary effects, simulations were stopped as soon as a pixel be-
 207 longed to the boundary of the simulated area. However, tracks with a size of ≤ 2000
 208 points were discarded. Finally, after generating the animal track, a virtual time t
 209 was added to the data. In particular, for each time point t the next time point $t + 1$
 210 was generated by $t + 1 = 30 + pois\left(\lambda = 30 + 30 \cdot (1 + Mov_ran3 \cdot \overline{hab}(\vec{y}))\right)$ sec. Here,
 211 $Mov_ran3 \in [0, 1.0]$ reflects the possibility that animals slow down their movement
 212 in appropriate habitats, the latter again represented by high values of $\overline{hab}(\vec{y})$. I.e.,
 213 for $Mov_ran3 = 0$, time intervals were on average 1 *min* long and independent of
 214 local habitats (but intervals may display some temporal stochasticity due to natural
 215 variation in animal velocity). In contrast for $Mov_ran3 \geq 0$, the average interval
 216 length between steps increases linearly with local average *hab* values. Some example
 217 tracks with varying values for Mov_w , Mov_s and Mov_ran are given in Fig. S2.

218 2.3.3 Tracking device and final tracking data

219 The simulated animal movement data at the spatial pixel scale (c.f., previous subsec-
 220 tion) were subsequently reduced to a much coarser temporal resolution, mimicking
 221 the data collected by a tracking device. A total of 150 tracking points were selected
 222 from the approximately 5,000 raw data points. For this selection, the temporal
 223 regularity of tracking points was represented by the categorical variable *Dev_reg*,
 224 where a value of 1 represents temporally equidistant tracking points (with respect
 225 to the virtual time t), and levels 2 and 3 represent increasing random irregular-
 226 ity, the latter realized via increasing variance of a random normal deviation from
 227 selected equidistant time points. Due to the inherent temporally discrete nature
 228 of simulated animal movement data, the nearest existing time point rather than
 229 the exact selected one was chosen. After selection of the spatio-temporal subset

230 ('tracking data') from the raw animal movement data, as described above, spa-
231 tial bias was added, quantified by the variable $Dev_{bias} \in [0, 3]$ depicting the
232 standard deviation of a normally distributed random error separately added to
233 each point and coordinate. These final animal tracking data were used to gener-
234 ate the following additional variables: total spatial extent of the trip defined by
235 $Dev_{ext} = ((max(x) - min(y)) + (max(x) - min(y)))/2$, and total number of
236 tracking points defined by the variable Dev_N .

237 2.4 Statistical analysis of simulated animal tracks

238 As noted in the Introduction, we applied and compared three main classes of statis-
239 tical methods (SLRMs, PPMs, and (i)SSMs, and several variations) using simulated
240 tracking data. To obtain optimal comparability, several measures during tracking
241 data analysis were unified. First, the same level of dummy point numbers per track-
242 ing point (namely $N_{dummy} \in \{8, 80, 224\}$) were used in all models. Second, central
243 predictors were entered in all models in a similar manner: *hab* as a linear predic-
244 tor for habitat selection studies, and *attrac_dist* for the evaluation of large-scale
245 attraction measuring the Euclidean distance to the attraction centre. As an alter-
246 native to *attrac_dist*, the variable *attrac_ang* was used, calculating the cosine of
247 the angular deviation from a straight line between the previous tracking data point
248 and the attraction centre, thus representing the directional bias towards the centre.
249 *attrac_ang* was not used for SLRMs, because a pure spatial approach does not
250 allow the consideration of angular deviations depending on the temporal order of
251 tracking points. Finally, all compared models were nested in simulated tracks (i.e.,
252 each simulated animal track was analysed with all models) to minimize the level of
253 unexplained variance in the final comparative analyses.

254 *Spatial logistic regression models*

255 SLRMs were implemented using the `gam()` function in the R-package *mgcv* [49]
256 with a binomial error distribution and logit-link function, where true tracking points
257 served as presence points in the outcome variable, and dummy points were treated as
258 true absences. In particular, dummy points were selected in two different manners:
259 first, randomly chosen within the minimum convex polygon (MCP) around the
260 tracking data (calculated by the *mcp()*-function from the *sp*-package [50]) depicting
261 a simple and frequently used approach to estimate home ranges (e.g., Ref. [51, 52])
262 and leading to the model *Meth_SLRM_mcp*; and second, based on spatio-temporal
263 tracking data (i.e., step size and turning angle distributions) as provided by the *amt*-
264 package, leading to the model *Meth_SLRM_amt*. Notably, step lengths were based
265 on a gamma distribution, whereas turning angles relied on a von Mises distribution
266 [41, 43]. Although we used SLRMs to investigate the performance of pure spatial
267 methods with respect to the analysis of tracking data, we used this spatio-temporal
268 dummy point selection approach to obtain an 'optimal set' of spatially distributed
269 dummy points, assuming that sophisticated pure spatial selection methods would
270 be at least as good as this spatio-temporal selection method. Finally, to account
271 for possible spatial autocorrelation, we optionally added a spatial 2D thin plate
272 regression spline $s(x_1, x_2)$ to the predictor (where the optimal number of knots
273 was estimated based on generalized cross-validation [49]), leading to the models
274 *Meth_SLRM_amt_ac* and *Meth_SLRM_mcp_ac*.

275 *(Integrated) step selection models*

276 SSM and (i)SSM analyses relied on conditional logistic regression functions
 277 (*fit_ssf()* respectively *fit_issf()*), as provided by the R-package *amt* [44]. In par-
 278 ticular, data were stratified by the time points of tracking points ('used steps')
 279 and temporally associated dummy points ('available steps') [41, 43]. Similar to
 280 *Meth_SLRM_amt*, dummy points were generated based on step lengths and turn-
 281 ing angles, assuming a gamma von Mises distribution [43, 44]. However, several
 282 predictor formulations are possible considering temporal, spatial, and directional
 283 correlations [44]. We therefore performed a model selection step based on the Akaike
 284 information criterion (AIC) [53] comparing 10 different models reflecting different
 285 combinations of main effects and interactions in relation to the different autocorre-
 286 lation terms, namely temporal distance to the foregoing tracking point, logarithm
 287 of the spatial distance to the forgoing tracking point, and cosine of the turning an-
 288 gle [44]. The best models from this model selection step were termed *Meth_SSM_1*
 289 and *Meth_iSSM_1*, respectively. A second variant of these models was generated
 290 by augmenting the above-mentioned model selection by a second AIC-based model
 291 selection step, where the above models were compared with models where the log-
 292 arithm of the spatial distance to the forgoing tracking point ($\log(d_x)$) was replaced
 293 by d_x , $\sqrt{d_x}$ and $\log(\sqrt{d_x})$, respectively, leading to the models *Meth_SSM_2* and
 294 *Meth_iSSM_2*. Finally, in order to estimate the general importance of model selec-
 295 tion within the (i)SSM framework, we fitted the (i)SSM model without any model
 296 selection procedure using $\log(d_x)$, the cosine of the turning angle, as well as their
 297 interaction term as autocorrelation-related predictors (*Meth_iSSM_no_ms*).

298 *Point process models*

299 The applied PPMs mainly relied on the approach of Johnson et al. [40]. In particular,
 300 we approximated the spatio-temporal PPM likelihood by the expression

$$\sum_{j,k} w_{jk} \left(u_{jk} \ln(\lambda(t_j, \vec{x}_{jk})) - \lambda(t_j, \vec{x}_{jk}) \right)$$

301 with point process intensity $\lambda()$ depending on time points t_j and spatial 2D points
 302 \vec{x}_{jk} . Furthermore, $u_{jk} = 1/w_{jk}$ for tracking points and $u_{jk} = 0$ for dummy points
 303 with appropriate quadrature weights w_{jk} [27, 40], where the index j always refers to
 304 discrete time points and the index k to the discretisation in 2D space. This expres-
 305 sion is proportional to a weighted Poisson likelihood with weights w_{jk} and observa-
 306 tions u_{jk} , such that standard generalized linear/additive modelling (GLM/GAM)
 307 software can be used [27, 40]. Dummy point generation was based on the R-package
 308 *mvQuad* [54]. In particular, for each tracking point, we created a rectangular grid of
 309 dummy points (centred around the previous tracking point and randomly rotated)
 310 including quadrature weights based on the rectangle rule. The latter outperformed
 311 several other possible quadrature weights with respect to maximal PPM statistical
 312 power, while showing a reasonable type I error rate (results not shown). However,
 313 the optimal spatial extent of the dummy point grid in PPMs is *a priori* not clear,
 314 in contrast to SSM methods where dummy points represent 'available steps' and
 315 are thus defined by average spatial step lengths of the tracking point time series. In

316 contrast, for each time point t_j in PPMs, the grid has to be large enough to cover
 317 sufficient area where the conditional intensity surface substantially differs from zero
 318 on the one hand, but small enough to obtain a satisfactory spatial resolution on
 319 the other. We compared two different approaches to achieve this goal. In the first
 320 model (*Meth_PPM_1*), the side length l of the dummy point grid associated with
 321 tracking point x_{jk} was based on the distance between x_{jk} and the foregoing track-
 322 ing point $x_{(j-1)k}$ by $l = 2\|x_{jk} - x_{(j-1)k}\|$. The idea was that, similar to (i)SSMs,
 323 dummy point grid size was locally adapted to the spatial step length. In the second
 324 model (*Meth_PPM_2*), we determined the optimal l value empirically by fitting
 325 several models using the values $l = 5, 10, 15, \dots, 180$, and finally choosing the l -value
 326 that led to the highest z-score with respect to the predictor of interest (e.g., *hab* if
 327 habitat selection has been analysed) for further PPM analysis. I.e., we selected the
 328 model in which the most reliable attraction effect was detected.

329

330 All PPMs were fitted based on GAM-software [55] using the *bam()*-function in the
 331 *mgcv* package [49]. Additive models allow the investigation and formulation of non-
 332 linear relationships between different variables [49, 55], which we used to formulate
 333 possibly nonlinear autocorrelation-related predictors within the PPM, extending
 334 the linear approach of Johnson et al. [40]. In particular, autocorrelation has been
 335 integrated as a Markov process (similar to the spatio-temporal PPM approach of
 336 [40]), and considering a possibly nonlinear dependency of the tracking data on step
 337 duration, step length, and step heading, the same variables are suggested to in-
 338 tegrate possible autocorrelation within the (i)SSM framework [43]. In particular,
 339 optimal models were selected based on the AIC (with AIC calculations based on
 340 the PPM-likelihood approximation and not on the standard AIC provided by the
 341 *mgcv* package), comparing several linear and tensor-smooth combinations of these
 342 variables. Additionally, as proposed by Hooten et al [8], we optionally augmented
 343 the final selected models with a spatial 2D thin plate spline $s(x_1, x_2)$ representing
 344 the large-scale home range and leading to the model *Meth_PPM_2_hr*. To investigate
 345 the general importance of model selection within the PPM framework, we also fitted
 346 a model without any model selection procedure using a 3D-tensor-smooth for the
 347 above-mentioned autocorrelation-related variables (*Meth_PPM_2_no_ms*). Finally,
 348 in addition to the PPMs using a Poisson distribution, we alternatively fitted PPMs
 349 using a negative binomial (*Meth_PPM_2_nb*) and a quasi-Poisson (*Meth_PPM_2_qp*)
 350 distribution, possibly accounting for over- or underdispersion [8].

351 2.5 Final comparative analysis of model performance

352 As explained in the introduction to the Methods section, the final analysis compar-
 353 ing the performances of the different statistical methods was based on four different
 354 data frames: two analysing habitat and two analysing large-scale attraction. No-
 355 tably, in each case, the analysis was based either on simulated tracking data with
 356 an existing underlying attraction effect (for evaluating statistical power) or with
 357 no underlying effect (for evaluating false-positive rates). In all cases, the binomial
 358 variable *effect* ('detected significant effect' vs. 'no significant effect detected') was
 359 evaluated, likely depending on some of the 16 different variants of the applied statis-
 360 tical methods, but also possibly interplaying with the 15 different predictors related

361 to habitat, animal movement, and tracking-device properties (c.f., Tab. 1).
362

363 For the final comparative analyses, generalized estimating equation (GEE) mod-
364 els [56] with an 'exchangeable correlation structure' [2] and a binomial probability
365 distribution were used, realized with the *geeglm()*-function in the R-package *geep-*
366 *ack* [57]. We favoured GEEs over logistic generalized linear mixed models (GLMMs
367 [2, 58, 59]), because residuals related to the trip IDs (used as a random intercept)
368 strongly violated the normality assumption in GLMMs. Here, GEEs are known to
369 be much more robust against miss-specification in correlation structures [2].
370

371 For the comparison of average values of statistical power in terms of false-positive
372 rates between the different statistical methods, the variable *method* was used as the
373 only predictor in GEEs. If the model performance in interplay with several habi-
374 tat, movement, and tracking device-related properties was investigated instead, a
375 very high number of possible predictors resulted from the variables presented in
376 Tab. 1, because not only the main effects, but also the interaction terms between
377 method- and tracking-related variables, were considered. We therefore applied an
378 efficient model selection technique before GEE analysis, namely the 'least abso-
379 lute shrinkage and selection operator' (LASSO) technique [60, 61] (based on the
380 R-package *glmnet* [62]). This technique is known to perform reliable model selection
381 even if predictor numbers are high [4]. Notably, we used logistic regression models
382 within the LASSO procedure and combined them with cross-validation to select a
383 promising set of predictors [63]. The cross-validation method is a systematic ap-
384 proach to adjust a certain degree of freedom (here, predictor number and choice) to
385 maximize the generalization of the model to independent data (i.e., its predictive
386 performance) [64]. In a second step, the selected predictors were analysed by GEEs
387 as described above, finally allowing us to determine the circumstances under which
388 each method performed best.
389

390 We analysed the impact of the tracking data size on model performance and com-
391 puting time largely as described above (c.f. also Fig. 1). However, only the size
392 of the simulated animal tracks was varied, such that the final tracking data size
393 ranged between $N = 33$ and 300 tracking points; all other variables (related to
394 habitat, movement, and tracking-device properties) were set to constant average
395 values across all simulations. The latter aim was to minimize the unexplained vari-
396 ance. The final comparative analysis was again based on GEEs if statistical power
397 or false-positive rates were investigated, while appropriate GAMs were separately
398 applied for each method for the analysis of computation time [49, 55].

399 **3 Results**

400 **3.1 Habitat selection**

401 *3.1.1 Average performance*

402 We first compared the average performances of the 15 different variants of the three
403 main statistical methods using 80 dummy points per tracking point. Corresponding
404 GEE results are shown in Fig. 2.
405

variable name	explanation
<i>Dev_reg</i>	temporal regularity of tracking points
<i>Dev_bias</i>	strength of spatial location error
<i>Dev_ext</i>	spatial extent of the animal trip
<i>Dev_step</i>	average spatial step size of the animal trip
<i>Dev_var</i>	variance of the spatial step size of the animal trip
<i>Hab_auto</i>	strength of spatial habitat autocorrelation
<i>Hab_anis</i>	anisotropy of spatial habitat autocorrelation
<i>Hab_smooth</i>	smoothness of transition between habitats
<i>Hab_type</i>	continuous or categorical habitat
<i>Mov_SD</i>	strength of randomness in animal movement
<i>Mov_σ</i>	movement bias towards attraction centre
<i>Mov_ω</i>	habitat selection strength
<i>Mov_ran</i>	strength of directional persistence
<i>Mov_ran2</i>	reducing directional persistence in preferred habitats
<i>Mov_ran3</i>	introducing slower movement in preferred habitats
<i>Meth_SLRM_mcp</i>	spatial logistic regression model, dummy points are randomly generated inside the minimal convex polygon
<i>Meth_SLRM_amt</i>	spatial logistic regression model, sophisticated generation of dummy points based on spatio-temporal information
<i>Meth_SLRM_mcp_ac</i>	as for <i>SLRM_mcp</i> but with spatial 2D smooth aiming to reduce spatial autocorrelation
<i>Meth_SLRM_amt_ac</i>	as for <i>SLRM_amt</i> but with spatial 2D smooth aiming to reduce spatial autocorrelation
<i>Meth_SSM_1</i>	step selection model including model selection with respect to the autocorrelation terms
<i>Meth_SSM_2</i>	as <i>SSM_1</i> but with refined model selection with respect to spatial autocorrelation
<i>Meth_iSSM_1</i>	integrated step selection model including model selection with respect to the autocorrelation terms
<i>Meth_iSSM_2</i>	as <i>iSSM_1</i> but with refined model selection with respect to spatial autocorrelation
<i>Meth_iSSM_no_ms</i>	as <i>iSSM_1</i> but without model selection procedure
<i>Meth_PPM_1</i>	spatio-temporal point process model including model selection, the spatial scale of the dummy point grid is related to local tracking data step size
<i>Meth_PPM_2</i>	spatio-temporal point process model including model selection, the optimal spatial scale of dummy point grid is empirically determined
<i>Meth_PPM_2_hr</i>	as <i>PPM_2</i> but augmented with spatial 2D smooth representing home range
<i>Meth_PPM_2_nb</i>	as <i>PPM_2</i> but using a negative binomial instead of a Poisson probability distribution
<i>Meth_PPM_2_qp</i>	as <i>PPM_2</i> but using a quasi-Poisson instead of a Poisson probability distribution
<i>Meth_PPM_2_no_ms</i>	as <i>PPM_2</i> but without model selection procedure

Table 1 Possible predictors for statistical power and type I error rates in the context of habitat selection and large-scale attraction based on animal tracking data. Above the double line, predictors related to habitat (Hab...), animal movement (Mov...) and tracking-device (Dev...) properties are shown; below the double line predictors related to the statistical method applied to simulated tracking data (Meth...) are presented.

406 SLRMs without consideration of spatial autocorrelation appeared to show up-
407 ward biased type I error rates, especially if dummy points were selected based on
408 the simple MCP approach. In the model *Meth_SLRM_mcp*, the type I error rate was
409 30 %, which is far above the nominal level of 5 %. The application of more sophis-
410 ticated schemes for dummy point selection (namely the spatio-temporal method
411 used in model *Meth_SLRM_amt*) appeared to reduce this bias. However, our results
412 suggested that the type I error rate was still above the nominal level, and could
413 be even more pronounced if sophisticated but pure spatial methods for dummy
414 point selection were applied. We therefore do not recommend using SLRMs with-
415 out consideration of spatial autocorrelation to infer habitat selection from tracking
416 data. This approach would only be appropriate if the tracking data were spatially
417 uncorrelated; however, the existence of home ranges means that this will be rarely
418 the case. PPM approaches using a quasi-Poisson distribution instead of a Poisson
419 distribution (model *Meth_PPM_2_qp*) also showed distinctly upward biased type
420 I error rates, making this approach also inappropriate for the analysis of habitat
421 selection.

422
423 With respect to the statistical power (currently excluding the above models with
424 biased type I error rates), (i)SSM approaches showed the highest average power,
425 which appeared to be robust against the different variants of these models, such as
426 different model selection procedures or SSMS vs. iSSMs. PPMs generally showed
427 equally robust but slightly reduced power, except for the model *Meth_PPM_1* where
428 the dummy point grid extension was based on the local distance between subse-
429 quent tracking points. The latter approach led to distinctly reduced power and type
430 I error rate, indicating that the empirical choice of the dummy point grid extension
431 was critical for good PPM performance. All the remaining PPM variants (possibly
432 including a negative binomial instead of a Poisson distribution, an additional 2D
433 home range smooth in the predictors, or lacking the model selection procedure)
434 showed similar magnitudes of statistical power to each other and to the (i)SSMs.

435
436 Finally, SLRM approaches including a spatial smooth to take account of spatial
437 autocorrelation showed the lowest power of all the tested models with appropri-
438 ate false-positive rates, where the sophisticated methods for dummy point selection
439 (*Meth_SLRM_amt_ac*) performed better than the model with the simple MCP pro-
440 cedure (*Meth_SLRM_mcp_ac*), as expected. Notably however, the total differences
441 in average statistical power among the three main methods were relatively small:
442 PPM approaches differed by < 5 % from (i)SSM approaches, and the difference
443 between (i)SSM and SLRM approaches was < 10 %.

444 3.1.2 Dummy point numbers

445 To investigate the influence of dummy point numbers on model performance,
446 we compared the statistical powers and false-positive rates among four well-
447 performing representative models (*Meth_PPM_2*, *Meth_SSM_1*, *Meth_iSSM_1* and
448 *Meth_SLRM_amt_ac*) using three different levels of dummy points (N_{Dummy}
449 $\in \{8, 80, 224\}$) per tracking point (Fig. 3). For all the compared methods and
450 dummy point numbers, type I error rates were within or below the acceptable nom-
451 inal range of 5 %. With respect to the statistical power, only the PPM approach

452 showed a drastic decrease if only eight dummy points were used, however, there was
 453 no significant difference for 80 vs. 224. Although all the other models showed similar
 454 qualitative behaviours, SSMs, iSSMs, and SLRMs had much higher robustness with
 455 respect to *N_Dummy*.

456 3.1.3 Interplay between method and tracking-data properties

457 To analyse the interplay between the methods and variables related to environmen-
 458 tal, movement, and tracking-device properties (c.f., Tab. 1), we first restricted the
 459 final data frame for the power evaluation to the four above-mentioned representative
 460 methods (based on 80 dummy points per tracking point) (c.f., previous subsection).
 461 As described in more detail in subsection 2.5, we then applied a LASSO-based
 462 model selection in combination with cross-validation to the data frame to select a
 463 set of promising predictors, which were further analysed in a third step using appro-
 464 priate GEEs. We were interested in the factors influencing the statistical power of
 465 the different methods in a similar way. Thus, in the first GEE, we only investigated
 466 the main effects of the variables presented in Tab. 1. In contrast, the main effects
 467 were augmented in the second GEE with interaction terms between model-related
 468 (Tab. 1 below the double line) and non-model-related (Tab. 1 above the double line)
 469 variables. In particular, the interaction terms represented conditions in which the
 470 different methods performed differently. In Tab. 2, all significant terms (to the level
 471 of $\alpha = 0.05$) in the first main effect GEE (above) and the significant interaction
 472 terms in the second GEE (below) are presented.

473

Parameter name	Estimate	SE	p
Meth_PPM_2	-0.2130	0.0716	0.0029
Meth_SLRM_amt_ac	-0.4249	0.0688	0.0000
Mov_ ω	4.0999	0.4601	0.0000
Hab_auto	-0.0236	0.0068	0.0005
Dev_bias	-0.4458	0.1279	0.0005
Hab_smooth	0.2925	0.1417	0.0390
Hab_type	-1.0178	0.2336	0.0000
Mov_ran2	0.8928	0.4333	0.0394
Mov_ran3	0.9753	0.3891	0.0122
Dev_step	-0.1114	0.0417	0.0076
Meth_modelSLRM_amt_ac:Hab_auto	-0.0100	0.0044	0.0240
Meth_modelPPM_2:Hab_type	-0.6554	0.1592	0.0000
Meth_modelPPM_2:Hab_anis	0.7750	0.2582	0.0027
Meth_modelSLRM_amt_ac:Hab_anis	1.1949	0.2789	0.0000

Table 2 Significant main effects (above the double line) and interaction terms (below the double line) driving the statistical power during habitat selection studies. Results are based on GEE analyses in combination with LASSO-based model selection.

474 In accordance with subsection 3.1.1, the main effects in Tab. 2 revealed that the
 475 PPM and more notably the SLRM showed reduced power on average, compared
 476 with the (i)SSM (the latter served as the baseline category in the model-related
 477 predictor variable *model*). Furthermore, the strength of local habitat attraction
 478 (*Mov_ ω*) and decrease in directional persistence and movement velocity in appro-
 479 priate habitats (*Mov_ran2* respectively *Mov_ran3*) positively influenced the power.
 480 This was expected, given that all three processes resulted in an increased proportion
 481 of tracking points (relative to dummy points) in preferred habitats. Less expectedly,
 482 gradual continuous habitat changes appeared to have a positive influence on the

483 statistical power: the average power was lower in categorical (*Hab.type*) than in
484 continuous habitats, and blurry boundaries (*Hab.smooth*) between different habitat
485 types resulted in a higher power than sharp transitions. Generally, this observation
486 suggests that both categorical habitats as well as sharp boundaries provide less local
487 information on animal–habitat interactions than continuous and spatially graded
488 variables, affecting the power of the statistical method finally applied to these vari-
489 ables. Finally, increasingly strong habitat autocorrelation (*Hab.auto*) or location
490 error (*Dev.bias*) and large average step sizes (*Dev.step*) decreased the power, all
491 of which were intuitively expected.

492

493 The interaction terms in Tab. 2 revealed that each method had specific weaknesses
494 with respect to habitat properties. (i)SSM methods coped poorly with strong habi-
495 tat anisotropy, since the statistical powers of PPM and SLRM approaches decreased
496 significantly less strongly with increasing anisotropy compared to the baseline level
497 ‘iSSM’. In contrast, the power of PPMs decreased strongly for categorical habitats,
498 whereas the decrease for SLRMs was strongest for autocorrelated habitats.

499

500 In accordance with the above observations, if we restricted the power-related
501 data frame during GEE analysis to continuous habitats with high anisotropy
502 (*Hab.anis* > 0.5), i.e., optimal for PPM performance, PPMs had significantly
503 higher power on average (70 %) compared with SLRMs (63 %). If we restricted
504 the habitats to those with moderate autocorrelation (*Hab.auto* < 30) and high
505 anisotropy (*Hab.anis* > 0.5), i.e., optimal for SLRMs, the latter showed a non-
506 significant (by 1 %) higher statistical power than (i)SSMs. The issue of which
507 method shows the highest statistical power may thus depend on the properties
508 of the habitats of interest, and the optimal method could be chosen accordingly.

509 3.2 Long-range attraction

510 3.2.1 Average performance

511 The GEE results with respect to the average performance of the different methods
512 in the context of inferring large-scale attraction are shown in Fig. 4. In particular,
513 models ending in ‘...ang’ evaluated the angular bias towards the attraction centre,
514 whereas those ending in ‘...dist’ considered the Euclidean distance to the centre in-
515 stead. Overall, the statistical powers and false-positive rates of (i)SSMs and PPMs
516 were similar to the analyses of local habitat selection (c.f., subsection 3.1.1). (i)SSM
517 approaches showed rather robust behaviour with statistical power and type I er-
518 ror rates similar to most of the PPM approaches. Furthermore, only PPMs using a
519 quasi-Poisson instead of a Poisson distribution (models *Meth.PPM_2_qp...*) showed
520 upward biased type I error rates, and PPMs without empirical determination of the
521 optimal extent of the dummy point grid (models *Meth.PPM_1...*) showed strongly
522 reduced power.

523

524 Interestingly, in PPMs, the above-mentioned angular approach for detecting long-
525 range attraction invariably performed better than evaluating the Euclidean dis-
526 tance, whereas the opposite was true in (i)SSMs. Overall, (i)SSM methods using
527 the distance-based approach showed the highest statistical power. However, similar

528 to analyses of habitat selection, absolute differences in power between the different
 529 appropriate PPM and (i)SSM approaches were mostly moderate ($< 10\%$).

530

531 In contrast to the habitat selection studies, all SLRM-based approaches performed
 532 badly. SLRMs without consideration of spatial autocorrelation showed strongly up-
 533 ward biased false-positive rates, whereas SLRMs incorporating these spatial corre-
 534 lations had drastically reduced statistical power.

535 3.2.2 Dummy point numbers

536 Similar to the habitat selection analyses (c.f., subsection 3.1.2), we also assessed the
 537 required number of dummy points in the context of large-scale attraction effects
 538 based on a selection of representative models. The corresponding results (Fig. 5)
 539 were highly comparable to those for the habitat selection studies (c.f., Fig. 3). We
 540 observed a robust dependency of (i)SSM performance with respect to all considered
 541 dummy point numbers. Combining PPMs with the angular approach to detect
 542 large-scale bias, a low number ($N_Dummy = 8$) of dummy points led to a distinct
 543 drop in statistical power compared with $N_Dummy \in \{80, 224\}$, even stronger
 544 than for resource selection PPMs (Fig. 3). However, this effect was not observed
 545 when evaluating the Euclidean distance. Finally, SLRMs performed approximately
 546 equally badly for all considered dummy point numbers.

547 3.2.3 Interplay between method and tracking data properties

548 Similar to the resource selection analysis (c.f., subsection 3.1.3), the interplay be-
 549 tween statistical methods and tracking data properties was evaluated by combin-
 550 ing a LASSO-based approach with final GEE evaluation, separately evaluated in
 551 a model using only main effect predictors on one hand, and in a second model in-
 552 cluding various interaction terms on the other. However, variables related to habitat
 553 properties were omitted, given that the focus was on the detection of a large-scale at-
 554 traction effect possibly interplaying with movement and tracking device properties.
 555 Despite the model-related terms, which were in good accordance with subsection
 556 3.2.1 (the model *Meth_PPM2_ang* served as the baseline level), the variable *Mov_σ*
 557 was selected as the only additional significant predictor in the main effect GEE
 558 (c.f., Tab. 3), demonstrating that the statistical power with respect to large-scale
 559 attraction did not sensitively depend on animal movement or tracking-device prop-
 560 erties, but only on the strength of the attraction itself. Furthermore, in the GEE
 561 including interaction terms, no significant interaction occurred, suggesting that this
 562 robustness held for all considered methods equally.

	Estimate	SE	p
Meth_iSSM_1_dist	0.7074	0.1331	0.0000
Meth_PPM_2_ang	0.4982	0.1176	0.0000
Meth_SLRM_amt_ac	-3.4786	0.3029	0.0000
Meth_SSM_1_dist	0.5869	0.1250	0.0000
<i>Mov_σ</i>	63.1623	5.5495	0.0000

Table 3 Significant main effects driving the statistical power during large-scale attraction studies. Results are based on GEE analyses in combination with LASSO-based model selection.

563 3.3 Tracking data size and computation time

564 All the above analyses were performed based on relatively small tracking data
 565 sizes of $N \approx 150$ tracking points. However, both the performance of the differ-
 566 ent methods and their computation times may depend critically on N . We thus
 567 investigated the influences of tracking data size $N \in [33, 300]$ on model perfor-
 568 mance and computation time, restricting the analysis again to the four represen-
 569 tative and well-performing models *Meth_PPM_2*, *Meth_SSM_1*, *Meth_iSSM_1* and
 570 *Meth_SLRM_amt_ac*. GEE analysis using the *model* variable and N as predictors
 571 revealed that there were no significant differences in false-positive rates between the
 572 PPM (baseline level) and the other models, and no significant dependency on N . In
 573 contrast, and in accordance with the results presented above, GEE results revealed
 574 that iSSMs and SSMs deviated positively (but non-significantly) from the PPM
 575 in terms of statistical power, whereas the SLRM showed a significant reduction in
 576 power. Additionally, as expected, the positive dependency of statistical power on N
 577 was highly significant. Adding the interaction term "*model* : N " to the model did
 578 not result in any significant interactions, suggesting that the N -dependent increase
 579 in power was of the same order of magnitude in all models (c.f., Tab. 4).

580

	Estimate	SE	p
Meth_iSSM_1	0.2354	0.2475	0.3416
Meth_SSM_1	-0.0329	0.3651	0.9282
Meth_SLRM_amt_ac	-0.5161	0.3932	0.1894
N	0.0184	0.0043	0.0000
Meth_iSSM_1:N	-0.0007	0.0014	0.6319
Meth_SSM_1:N	0.0024	0.0034	0.4711
Meth_SLRM_amt_ac:N	-0.0003	0.0027	0.9098

Table 4 GEE results analyzing the statistical power depending on the different methods in interplay with the tracking data size N . Here, PPM.2 served as the baseline level in the variable specifying the statistical method.

581 With respect to the computation time (referring to non-parallelized computations
 582 on a desktop computer), GAMs separately applied for each of the four above-selected
 583 methods revealed that the time increased approximately linearly with the number
 584 of tracking points for all models (c.f., Fig 6). Here, iSSMs and SSMs were con-
 585 densed because of the strong similarity in results. PPMs showed distinctly longer
 586 computation times than the other methods.

587 4 Discussion

588 Our simulation study demonstrated that the spatio-temporal (i)SSM and PPM
 589 approaches performed better than the pure spatial SLRMs, with respect to bias,
 590 statistical power, and robustness. In particular, regarding the analyses of large-scale
 591 attraction effects, SLRMs appear to be strongly biased and thus inappropriate. This
 592 result is not unexpected, given that the large-scale attraction process takes place
 593 on the spatial scale of the animal track, whereas contrasting dummy points are
 594 selected on a small scale. Thus, if considered as a purely spatial problem, dummy
 595 points show similar distances to the attraction centre on average, compared with
 596 the tracking points, and contrasting both therefore does not lead to reliable results.
 597 Relative differences between tracking and dummy points would thus only lead to
 598 meaningful results if the data were evaluated in a time-dependent manner, i.e., by

599 locally grouping tracking points with dummy points (the latter centred around the
600 foregoing tracking point) by their associated time point. In the habitat selection
601 case, SLRMs only perform acceptably if spatial autocorrelation is appropriately
602 considered and sophisticated dummy point selection methods are used. But even
603 here, SLRMs coped particularly poorly with strong habitat autocorrelation, which
604 could be caused by spatial confounding, given that autocorrelation in SLRMs is
605 not introduced on the scale of single steps but as a 2D smooth at the scale of the
606 trip. This could match the scale of habitat autocorrelation, finally leading to spatial
607 confounding [36]. In summary, we therefore strongly recommend the use of (i)SSM
608 or PPM approaches (e.g., as presented in [40, 43, 44]) to avoid biased results, espe-
609 cially if the tracking data are spatio-temporally correlated.

610
611 With respect to the performances of PPM and (i)SSM approaches, there was only
612 a small difference in their statistical powers (if a few points with respect to PPM
613 implementation were taken into account; c.f., below), regardless of whether local
614 habitat selection or large-scale attraction/avoidance effects are investigated. No-
615 tably, (i)SSMs generally had slightly higher power than PPMs, but if the habitats
616 of interest are strongly anisotropic, the power of PPMs during habitat selection
617 studies can exceed that of (i)SSMs. However, to maximize the power of PPMs and
618 (i)SSMs, several factors need to be considered. Compared with the other methods,
619 PPMs appeared to require a higher minimum number of dummy points in order to
620 maximize their potential statistical power. Hence, sufficient dummy points (≥ 80)
621 per tracking point should be used. This is probably because dummy points in PPMs
622 are used to approximate a spatio-temporal integral [40, 27], which process may be
623 strongly biased for low dummy point numbers, especially if the integral is complex.
624 Additionally, the PPM power depends critically on an optimal spatial extension
625 of the dummy point grid, as determined empirically in this study. Furthermore, if
626 large-scale attraction/avoidance is evaluated, (i)SSMs should include distance to
627 the structure of interest as a predictor, whereas PPMs should use the angular bias
628 to this structure.

629
630 Despite these small differences in statistical power between PPMs and (i)SSMs,
631 we recommend in general using (i)SSMs over PPMs for the following reasons: (1) the
632 statistical power of (i)SSMs is more robust with respect to dummy point numbers
633 compared with PPMs; (2) (i)SSMs do not need the initial empirical determination
634 of critical measures (the latter becomes increasingly complex if more than one pre-
635 dictor is considered); (3) computation times for (i)SSMs are much shorter than for
636 PPMs, which is especially important in the context of large data sets; and finally
637 (4) (i)SSM model implementation is user-friendly and well-documented because of
638 the provided R-package *amt*, in contrast to spatio-temporal PPMs for which there
639 is currently (to the best of our knowledge) no available R-package.

640
641 The above mentioned difference in computation time between PPMs and (i)SSMs
642 was probably on the one hand because the initial empirical procedure to deter-
643 mine the optimal spatial extent of the dummy point grid required several time-
644 consuming model fits. On the other hand, our PPM implementation may have been

645 programmed in a less time-efficient manner compared with the SSM and (i)SSM
646 codes provided within the well-established R-package. Further advances in non-
647 empirical methods for determining optimal PPM dummy point grid extension, as
648 well as a more time-efficient code for PPMs could therefore strongly reduce the de-
649 scribed differences between PPMs and the other approaches. Indeed, time-efficient
650 methods for estimating PPM integrals have recently been discussed by Hooten et
651 al. [8]. Interestingly, overall, not pure spatial SLRMs but (i)SSMs showed the lowest
652 computational costs, possibly because of the computationally costly estimation of
653 the spatial smooth (including a cross-validation procedure [49]) in SLRMs.

654
655 Despite the general rather practical motivated recommendation using (i)SSMs
656 over PPMs, there could be situations where PPMs could be nevertheless preferred:
657 first, as mentioned above, if habitats are strongly anisotropic (e.g. associated with
658 thin borders, such as coast lines) PPMs may show a higher statistical power. Sec-
659 ond, if the focus of the study is on population behaviour rather than on the analysis
660 of single animals, PPMs may lead to more reliable population-level estimates com-
661 pared to (i)SSMs due to the Eulerian approach adopted in PPMs (in contrast to the
662 Lagrangian viewpoint in (i)SSMs) [42].

663
664 Finally, all comparisons performed within this study only demonstrated virtual
665 differences between SSM and (i)SSM approaches, and we were therefore unable
666 to make any clear recommendations. However, Avgar et al. noted additional ad-
667 vantages of iSSMs over SSMs [43], including their predictive capacity (e.g., for
668 landscapes different from the landscape used for the model fit), and the ability to
669 derive and parametrize a mechanistic movement model. We therefore finally recom-
670 mend the use of iSSMs.

671

672 **5 Conclusions**

673 To conclude, our work provides an extensive simulation study comparing the sta-
674 tistical powers and false-positive rates of different statistical methods frequently
675 used to infer local habitat selection or large-scale attraction/avoidance from animal
676 tracking data. We compared several variants of spatio-temporal PPMs, (i)SSMs, and
677 pure spatial SLRMs, and evaluated their performances in interplay with a broad
678 range of simulated habitat, movement, and tracking-device properties. Our results
679 suggest that SLRMs are inappropriate for detecting large-scale attraction/avoidance
680 effects, and only produce unbiased habitat selection inference if spatial autocorre-
681 lation and dummy point choice are considered with care. In contrast, PPMs and
682 (i)SSMs can produce robust and unbiased results for both habitat selection and
683 large-scale attraction studies, with only slight differences in their average statisti-
684 cal powers. For habitat selection studies, (i)SSMs can cope better with categorical
685 habitats than PPMs, whereas PPMs have higher statistical power than (i)SSMs in
686 strongly anisotropic habitats. When large-scale attraction/avoidance is investigated
687 using PPMs, the angular deviation from the straight connection to the structure of
688 interest should be evaluated, whereas the Euclidean distance to this structure per-
689 forms better in (i)SSMs. In general, we recommend the use of (i)SSMs over PPMs,

690 mainly because of their more user-friendly implementation, greater robustness (e.g.,
 691 with respect to dummy point numbers), and faster computation times. In particu-
 692 lar, (i)SSMs (the recent extension of SSMs) provide the additional advantage of an
 693 increased predictive capacity in combination with the derivation of a parametrized
 694 mechanistic movement model.

695 List of abbreviations

696 A list of abbreviations used in this study is given in Tabl 5.

abbreviation	explanation
SLRM	'spatial logistic regression model'; pure spatial approach for analysing tracking data
PPM	'point process model'; spatio-temporal approach for analysing tracking data, developed from an Eulerian point of view
SMM	'step selection model' (also called 'step selection function /SSF' or 'step selection analysis / SSA'); spatio-temporal approach for analysing tracking data, developed from a Lagrangian point of view
(i)SMM	'integrated step selection model'; integrating the SSM approach with a mechanistic movement model
MCP	'minimal convex polygon'; minimal convex polygon containing all tracking points, for estimation of the minimum home rage
GLM	'generalized linear model'; extension of linear regression models for non-normally distributed data
GAM	'generalized additive model'; extension of GLM to describe non-linear ('additive') dependencies
GEE	'generalized estimation equation'; estimation method for parameters of a GLM with possible unknown correlation between outcomes
GLMM	'generalized linear mixed model'; extension of GLM to describe outcome correlation with random effects
LASSO	'least absolute shrinkage and selection operator'; technique for model selection that can especially cope with a high number of possible predictors

Table 5 List of abbreviations used in this study

697 Declarations

698 Ethics approval and consent to participate

699 Not applicable.

700 Consent for publication

701 Not applicable.

702 Availability of data and materials

703 All data are available within this publication and the supporting material.

704 Competing interests

705 The authors declare that they have no competing interests.

706 Funding

707 Support for the statistical framework was obtained from the Federal Ministry for
 708 Economic Affairs and Energy according to a decision of the German Bundestag
 709 (project HELBIRD, 0325751).

710 **Authors' contributions**

711 MM developed the statistical methods, programmed the R-code, and performed
712 most of the statistical tests and graphical analyses. PS, VP, and LE discussed
713 and developed the required features of the analysis, provided ecological input and
714 research questions, and participated in study design and result validation. PS and
715 SG conceived the study and participated in its coordination. All authors helped to
716 draft the manuscript and read and approved the final manuscript.

717 **Acknowledgements**

718 Not applicable.

719 **Figure titles and legends**

720 **Figure 1:** Sketch of the approach applied to compare the performances of dif-
721 ferent statistical methods in interplay with simulated habitat, animal movement,
722 and tracking-device properties. In particular, local habitat selection and large-scale
723 attraction were evaluated separately, and each was further split into the evalua-
724 tion of statistical power (using movement data with underlying attraction effect)
725 and false-positive rate (using movement data without underlying attraction effects).

726

727 **Figure 2:** GEE results showing the comparative performances of different models
728 inferring habitat selection. The average statistical power is shown in green and the
729 false-positive rate in red. Eighty dummy points per tracking point were used in
730 all cases. Bars represent 95 % confidence intervals, points correspond to regression
731 coefficient estimates.

732

733 **Figure 3:** GEE results showing four representative models inferring habitat selec-
734 tion with a focus on the dependency of model performance on dummy point number
735 (the latter represented as circles vs. triangles vs. squares). The average statistical
736 power is shown in green and the false-positive rate in red. Bars represent 95 %
737 confidence intervals, symbols correspond to regression coefficient estimates.

738

739 **Figure 4:** GEE results comparing the performances of different models inferring
740 long-range attraction. The average statistical power is shown in green and the
741 false-positive rate in red. Eighty dummy points per tracking point were used in all
742 cases. Bars represent 95 % confidence intervals, symbols correspond to regression
743 coefficient estimates. Circles indicate that long-range attraction was measured by
744 the angular bias towards the attraction centre, and triangles indicate use of the
745 Euclidean distance.

746

747 **Figure 5:** GEE results showing four representative models inferring long-range
748 attraction with a focus on the dependency of model performance on dummy point
749 number (the latter represented as circles vs. triangles vs. squares). The average
750 statistical power is shown in green and the false-positive rate in red. Bars represent
751 95 % confidence intervals, symbols correspond to regression coefficient estimates.

752

753 **Figure 6** Non-parallel computation times for dummy point generation and model
754 fit for three representative models, in relation to the number of tracking points.

755 Shaded areas correspond to 95 % confidence bands.

756

757 **Figure S1:** Examples of simulated categorical habitat data permuted over the
758 variables *Hab_auto* (strength of spatial autocorrelation), *Hab_anis* (strength of au-
759 tocorrelation anisotropy), and *Hab_smooth* (blurry vs. sharp transition between
760 habitat boundaries).

761

762 **Figure S2:** Examples of simulated continuous habitat data permuted over the
763 variables *Mov_ω* (habitat selection strength), *Mov_σ* (strength of bias towards the
764 attraction centre), and *Mov_ran* (strength of directional persistence). Blue point
765 represents location of the attraction centre.

766

767 **Author details**

768 ¹BIONUM statistical consulting office, Hamburg, Finkenwerder Norderdeich 15 A, Hamburg, Germany. ²Research
769 and Technology Centre (FTZ), Hafentörn 1., 25761 Büsum, Germany.

770 **References**

- 771 1. Elith, J., Leathwick, J.R.: Species distribution models: Ecological explanation and prediction across space and
772 time. *Annual Review of Ecology, Evolution, and Systematics* **40**, 677–697 (2012)
- 773 2. Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M.: *Mixed Effect Models and Extensions in*
774 *Ecology* with R. Springer, New York, USA (2009)
- 775 3. Zuur, A.F.: *A Beginner's Guide to Generalized Additive Models* with R. Highland Statistics Ltd., New York,
776 USA (2012)
- 777 4. Korner-Nievergelt, F., Roth, T., von Felten, S., Guelat, J., Almasi, B., Korner-Nievergelt, P.: *Bayesian Data*
778 *Analysis in Ecology Using Linear Models with R, BUGS, and Stan*. Elsevier, London, UK (2015)
- 779 5. Austin, D., McMillan, J., Bowen, W.: A three-stage algorithm for filtering erroneous argos satellite locations.
780 *Marine Mammal Science* **19**(2), 371–383 (2003)
- 781 6. Hebblewhite, M., Haydon, T.D.: Distinguishing technology from biology: a critical review of the use of gps
782 telemetry data in ecology. *Philosophical transactions R. Soc. B* **365**, 2303–2312 (2010)
- 783 7. White, G.C., Garrott, R.A.: *Analysis of Wildlife Radio-tracking Data*. Elsevier (London), Boca Raton, Florida,
784 USA (2012)
- 785 8. Hooten, M.B., Johnson, D.S., McClintock, B.T., Morales, J.M.: *Animal Movement - Statistical Models for*
786 *Telemetry Data*. CRC Press, Boca Raton, Florida, USA (2017)
- 787 9. Mendel, B., Schwemmer, P., Peschko, V., Müller, S., Schwemmer, H., Mercker, M., Garthe, S.: Operational
788 offshore wind farms and associated ship traffic cause profound changes in distribution patterns of loons (*gavia*
789 *spp.*). *Journal of environmental management* **231**, 429–438 (2019)
- 790 10. Collett, M., Chittka, L., Collett, T.S.: Spatial memory in insect navigation. *Current biology : CB* **23**, 789–800
791 (2013). doi:10.1016/j.cub.2013.07.020
- 792 11. Poulter, S., Hartley, T., Lever, C.: The neurobiology of mammalian navigation. *Current biology : CB* **28**,
793 1023–1042 (2018). doi:10.1016/j.cub.2018.05.050
- 794 12. Phillips, J.B., Schmidt-Koenig, K., Muheim, R.: True navigation: sensory bases of gradient maps. Brown MF,
795 Cook R, G., editors. *Animal Spatial Cognition: Comparative, Neural & Computational Approaches*:
796 *Comparative Cognition Society* (2006)
- 797 13. Cheung, A., Zhang, S., Stricker, C., Srinivasan, M.V.: Animal navigation: the difficulty of moving in a straight
798 line. *Biological cybernetics* **97**(1), 47–61 (2007)
- 799 14. et al, S.A.: *Animal Movement Across Scales*. Oxford university press, Oxford, UK (2014)
- 800 15. Aarts, G., Fieberg, J., Matthiopolous, J.: Comparative interpretation of count, presence-absence and point
801 methods for species distribution models. *Methods in Ecology and Evolution* **3**, 177–187 (2012)
- 802 16. Gurarie, E., Ovaskainen, O.: Characteristic spatial and temporal scales unify models of animal movement. *The*
803 *American naturalist* **178**, 113–123 (2011). doi:10.1086/660285
- 804 17. McClintock, B.T., King, R., Thomas, L., Matthiopolous, J., McConnell, B.J., Morales, J.M.: A general
805 discrete-time modeling framework for animal movement using multistate random walks **82**, 335–349 (2012).
806 doi:10.1890/11-0326.1
- 807 18. McClintock, B.T., Johnson, D.S., Hooten, M.B., Ver Hoef, J.M., Morales, J.M.: When to be discrete: the
808 importance of time formulation in understanding animal movement. *Movement ecology* **2**, 21 (2014).
809 doi:10.1186/s40462-014-0021-6
- 810 19. Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E., Fryxell, J.M.: Extracting more out of relocation data:
811 building movement models as mixtures of random walks. *Ecology* (2004)
- 812 20. Jonsen, I.D., Flemming, J.M., Myers, R.A.: Robust state–space modeling of animal movement data. *Ecology*
813 **86**(11), 2874–2880 (2005)
- 814 21. Langrock, R., King, R., Matthiopolous, J., Thomas, L., Fortin, D., Morales, J.M.: Flexible and practical
815 modeling of animal telemetry data: hidden markov models and extensions. *Ecology* **93**, 2336–2342 (2012).
816 doi:10.1890/11-2241.1
- 817 22. Pearce, J.L., Boyce, M.S.: Modelling distribution and abundance with presence-only data. *Journal of Applied*
818 *Ecology* **43**, 405–412 (2006)
- 819 23. Barbet-Massin, B., Jiguet, F., Albert, C.H., Thuiller, W.: Selecting pseudo-absences for species distribution
820 models: how, where and how many? *Methods in Ecology and Evolution* **3**, 327–338 (2012)
- 821 24. S.J., P., Dudik, M.: Modeling of species distributions with maxent: new extensions and a comprehensive
822 evaluation. *Ecography* **31**, 161–175 (2008)
- 823 25. Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.J.: A statistical explanation of maxent for
824 ecologists. *Diversity and Distributions* **17**, 43–57 (2011)
- 825 26. Warton, D.I., Shepherd, L.C.: Poisson point process models solve the "psuedo-absence problem" for
826 presence-only data. *The Annals of Applied Statistics* **4**(3), 1383–1402 (2010)
- 827 27. Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G., Warton, D.I.: Poin
828 process models for presence-only analysis. *Methods on Ecology and Evolution* **6**, 366–379 (2015)
- 829 28. VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E.: Selecting pseudo-absence data for presence-only
830 distribution modeling: How far should you stray from what you know. *Ecological Modelling* (2009)
- 831 29. Fieberg, J., Matthiopolous, J., Hebblewhite, M., Boyce, M.S., Frair, J.L.: Correlation and studies of habitat
832 selection: problem, red herring or opportunity? *Philosophical transactions of the Royal Society B* **365**,
833 2233–2244 (2010)
- 834 30. Fleming, C.H., Fagan, W.F., Mueller, T., Olson, K.A., Leimgruber, P., Calabrese, J.M.: Estimating where and
835 how animals travel: An optimal framework for path reconstruction from autocorrelated tracking data. *Ecology*
836 **97**(3), 576–82 (2016)

- 837 31. Hoeting, J.A.: The importance of accounting for spatial and temporal correlation in analyses of ecological data.
838 *Ecological Applications* **19**(3), 574–577 (2009)
- 839 32. F. Dormann, C., M. McPherson, J., B. Araújo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A.,
840 Jetz, W., Daniel Kissling, W., *et al.*: Methods to account for spatial autocorrelation in the analysis of species
841 distributional data: a review. *Ecography* **30**(5), 609–628 (2007)
- 842 33. Hodges, J.S., Reich, B.J.: Adding spatially correlated errors can mess up the fixed effect you love. *The*
843 *American Statistician* **64**, 325–334 (2011)
- 844 34. Hanks, E.M., Schliep, E.M., Hooten, M.B., Hoeting, J.A.: Restricted spatial regression in practice:
845 geostatistical models, confounding, and robustness under model misspecification: Spatial confounding in
846 geostatistical models. *Environmetrics* **26**(4), 10–10022331 (2015)
- 847 35. Hefley, H.M.B.H.E.M.R.R.E. T.J., Walsh, D.P.: Dynamic spatio-temporal models for spatial data. *Spatial*
848 *statistics* **20**, 206–220 (2017)
- 849 36. Paciorek, C.J.: The importance of scale for spatial-confounding bias and precision of spatial regression
850 estimators. *Statist. Sci.* **25**(1), 107–125 (2010)
- 851 37. Warton, D., Aarts, G.: Advancing our thinking in presence-only and used-available analysis. *Journal of Animal*
852 *Ecology* **82**, 1125–1134 (2013)
- 853 38. Aarts, G., Fieberg, J., Matthiopolous, J.: Comparative interpretation of count, presence-absence and point
854 methods for species distribution models. *Methods in Ecology and Evolution* **3**, 177–187 (2012)
- 855 39. Johnson, C., Thomas, D.L., Ver Hoef, J.M., Christ, A.: A general framework for the analysis of animal resource
856 selection from telemetry data. *Biometrics* **64**, 968–976 (2008)
- 857 40. Johnson, D.S., Hooten, M.B., Kuhn, C.E.: Estimating animal resource selection from telemetry data using
858 point process models. *Journal of Animal Ecology* **82**, 1155–1164 (2013)
- 859 41. Thurffjell, H., Ciuti, S., Boyce, M.S.: Applications of step-selection functions in ecology and conservation **2**, 4
860 (2014). doi:10.1186/2051-3933-2-4
- 861 42. Michelot, T., Blackwell, P.G., Matthiopolous, J.: Linking resource selection and step selection models for
862 habitat preferences in animals. *Ecology* **100**, 02452 (2019). doi:10.1002/ecy.2452
- 863 43. Avgar, T., Potts, J.R., Lewis, M.A., Boyce, M.S.: Integrated step selection analysis: bridging the gap between
864 resource selection and animal movement *Methods in Ecology and Evolution* **7**, 619–630 (2016).
865 doi:10.1111/2041-210x.12528
- 866 44. Signer, J., Fieberg, J., Avgar, T.: Animal movement tools (amt): R package for managing tracking data and
867 conducting habitat selection analyses. *Ecology and evolution* **9**, 880–890 (2019). doi:10.1002/ece3.4823
- 868 45. Duchesne, T., Fortin, D., Rivest, L.-P.: Equivalence between step selection functions and biased correlated
869 random walks for statistical inference on animal movement. *PloS one* **10**, 0122947 (2015).
870 doi:10.1371/journal.pone.0122947
- 871 46. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical
872 Computing, Vienna, Austria. ISBN 3-900051-07-0 URL <http://www.R-project.org/> (2016)
- 873 47. Hijmans, R.J., van Etten, J.: Raster: Geographic analysis and modeling with raster data. r package version
874 2.0-12. <http://CRAN.R-project.org/package=raster> (2012)
- 875 48. Wickham, H.: *Ggplot2: Elegant Graphics for Data Analysis*. Springer, New York, USA (2009)
- 876 49. Wood, S.: *Generalized Additive Models: An Introduction with R*. CRC press, Boca Raton, Florida, USA (2017)
- 877 50. Pebesma, R.S.B. E.J.: Classes and methods for spatial data in r. *R News* **5** (2) (2005)
- 878 51. Worton, B.J.: A convex hull-based estimator of home-range size. *Biometrics*, 1206–1215 (1995)
- 879 52. Burgman, M.A., Fox, J.C.: Bias in species range estimates from minimum convex polygons: implications for
880 conservation and options for improved planning. In: *Animal Conservation Forum*, vol. 6, pp. 19–28 (2003).
881 Cambridge University Press
- 882 53. Akaike, H.: Information theory and an extension of the maximum likelihood principle. *International Symposium*
883 *on Information Theory Second Edition*, 267–281 (1973)
- 884 54. Weiser, C.: Methods for multivariate quadrature (r package version 1.0-6). URL
885 <http://CRAN.R-project.org/package=mvQuad> (2016)
- 886 55. Hastie, T., Tibshirani, R.J.: *Generalized Additive Models*. Chapman and Hall, London, UK (1990)
- 887 56. Hardin, J., Hilbe, J.: *Generalized Estimating Equations*. Chapman and Hall/CRC, London, UK (2003)
- 888 57. Halekoh U, Y.J. Hojsgaard S: The r package geepack for generalized estimating equations. *Journal of*
889 *Statistical Software* **15**(2), 1–11 (2006)
- 890 58. Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S.:
891 *Generalized linear mixed models: a practical guide for ecology and evolution*. *Trends Ecol Evol* **24**(3), 127–135
892 (2009). doi:10.1016/j.tree.2008.10.008
- 893 59. Zuur, A.F., Saveliev, A.A., Ieno, E.N.: *Zero Inflated Models and Generalized Linear Mixed Models With R*.
894 Highland Statistics Ltd., New York, USA (2012)
- 895 60. Tibshirani, R.: Regression shrinkage and selection via the lasso. *J Royal Statist Soci B* **58**, 267–288 (1996)
- 896 61. Tibshirani, R.: The lasso method for variable selection in the cox model. *Stat Med* **16**(4), 385–395 (1997)
- 897 62. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate
898 descent. *Journal of statistical software* **33**, 1–22 (2010)
- 899 63. Usai, M.G., Goddard, M.E., Hayes, B.J.: Lasso with cross-validation for genomic selection. *Genet Res (Camb)*.
900 **91**(6), 427–36 (2009)
- 901 64. Browne, M.W.: Cross-validation methods. *Journal of mathematical psychology* **44**(1), 108–132 (2000)

Figures

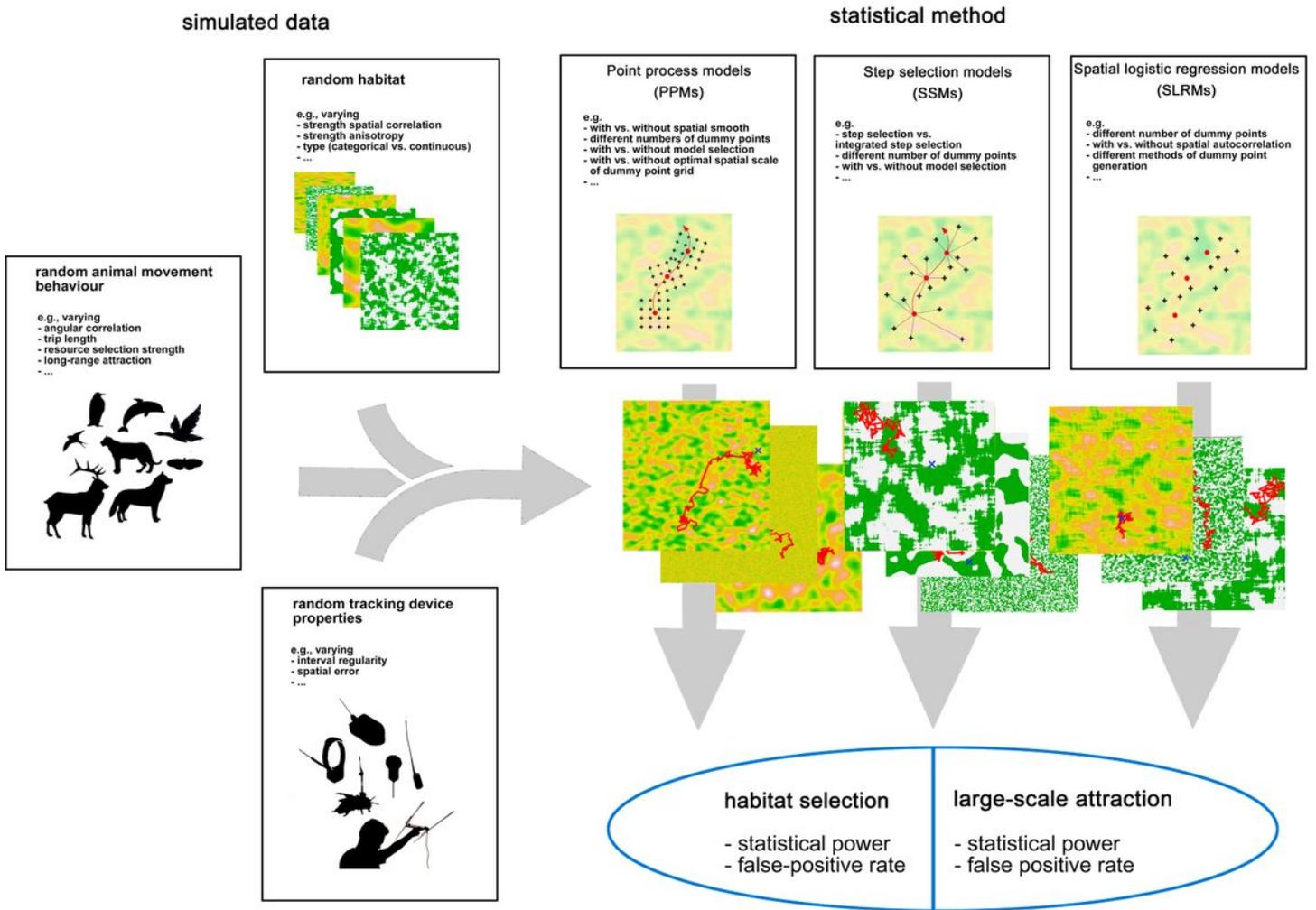


Figure 1

Sketch of the approach applied to compare the performances of different statistical methods in interplay with simulated habitat, animal movement, and tracking-device properties. In particular, local habitat selection and large-scale attraction were evaluated separately, and each was further split into the evaluation of statistical power (using movement data with underlying attraction effect) and false-positive rate (using movement data without underlying attraction effects).

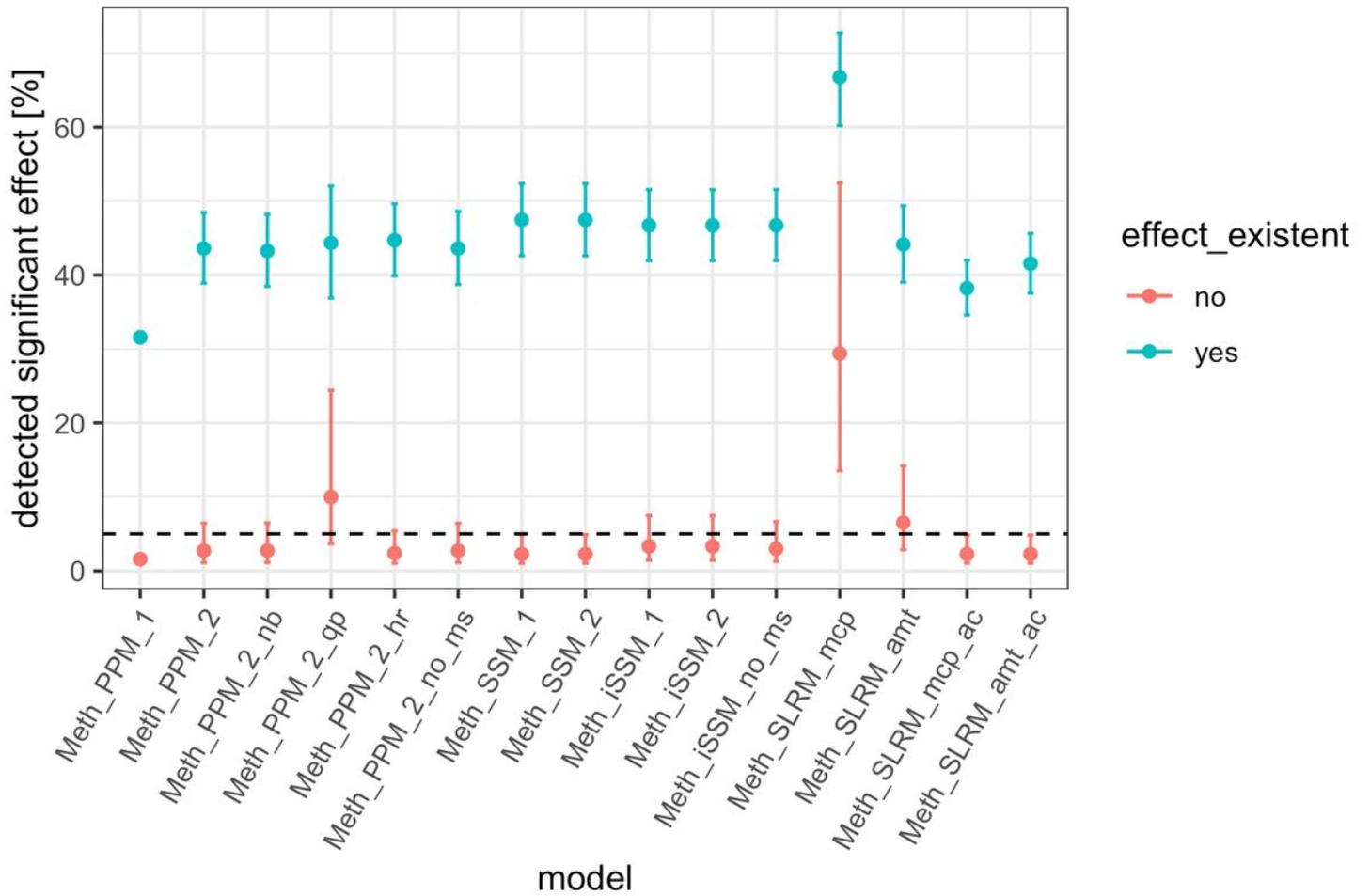


Figure 2

GEE results showing the comparative performances of different models inferring habitat selection. The average statistical power is shown in green and the false-positive rate in red. Eighty dummy points per tracking point were used in all cases. Bars represent 95% confidence intervals, points correspond to regression coefficient estimates.

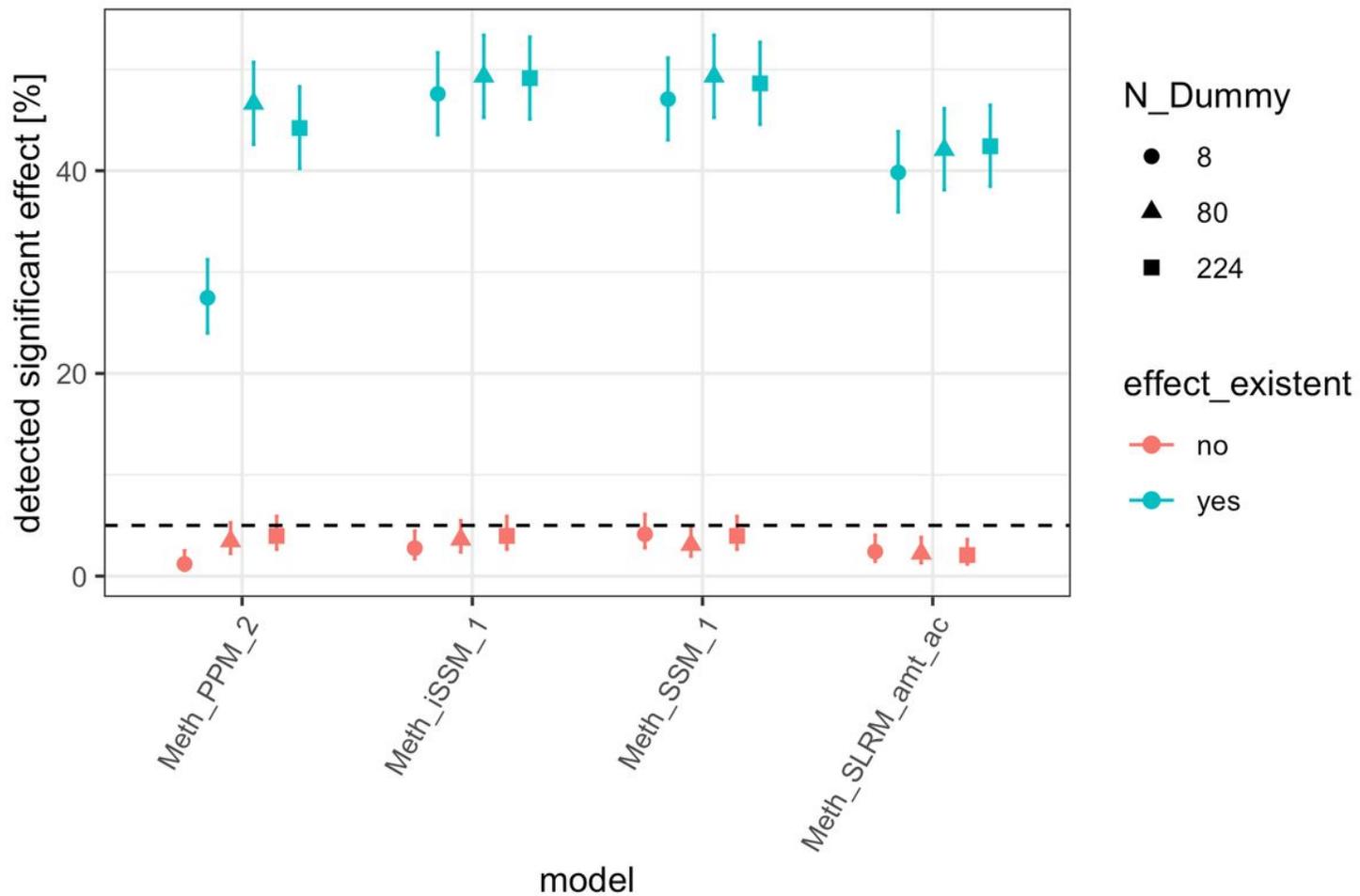


Figure 3

GEE results showing four representative models inferring habitat selection with a focus on the dependency of model performance on dummy point number (the latter represented as circles vs. triangles vs. squares). The average statistical power is shown in green and the false-positive rate in red. Bars represent 95 % confidence intervals, symbols correspond to regression coefficient estimates.

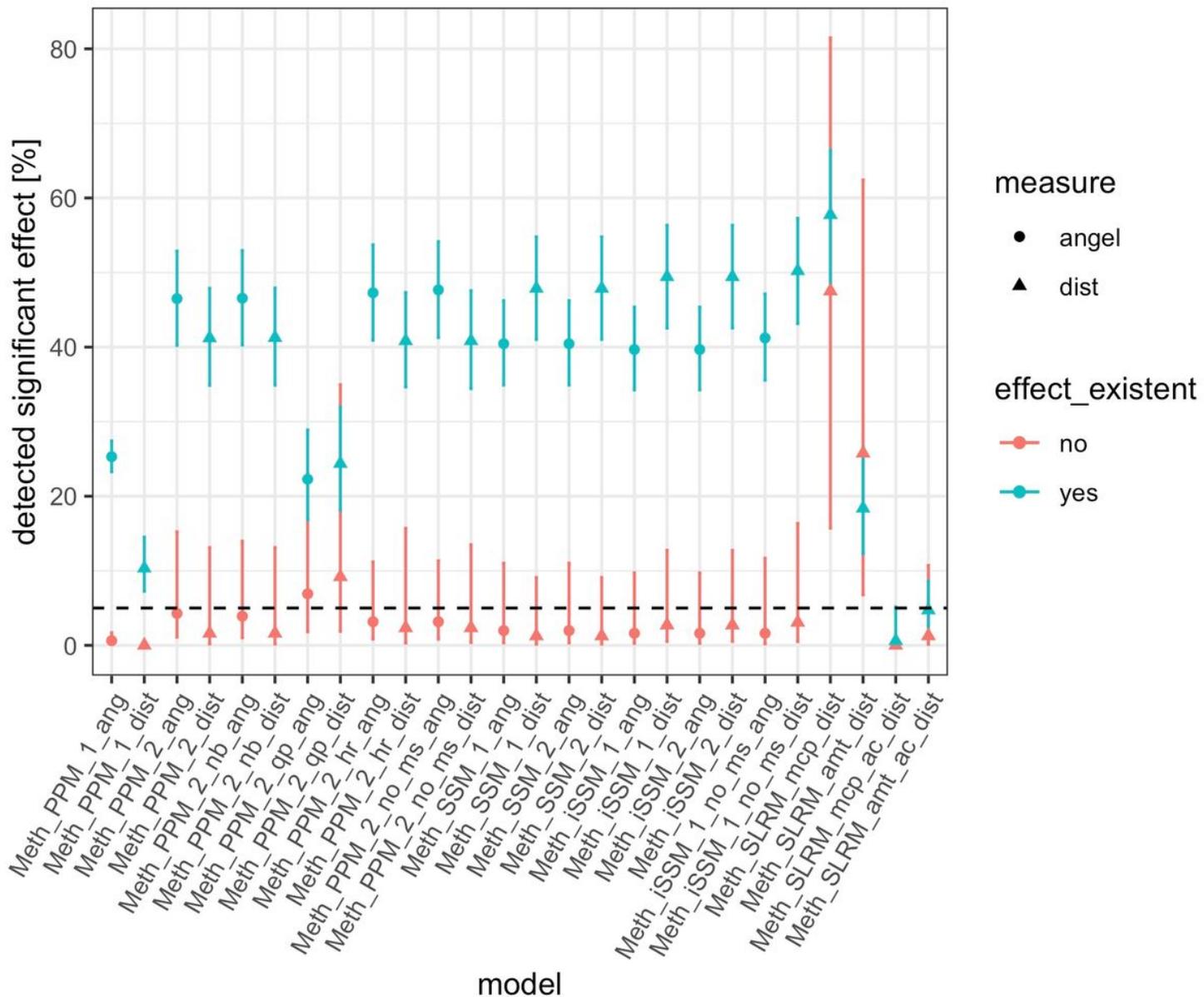


Figure 4

GEE results comparing the performances of different models inferring long-range attraction. The average statistical power is shown in green and the false-positive rate in red. Eighty dummy points per tracking point were used in all cases. Bars represent 95 % confidence intervals, symbols correspond to regression coefficient estimates. Circles indicate that long-range attraction was measured by the angular bias towards the attraction centre, and triangles indicate use of the Euclidean distance.

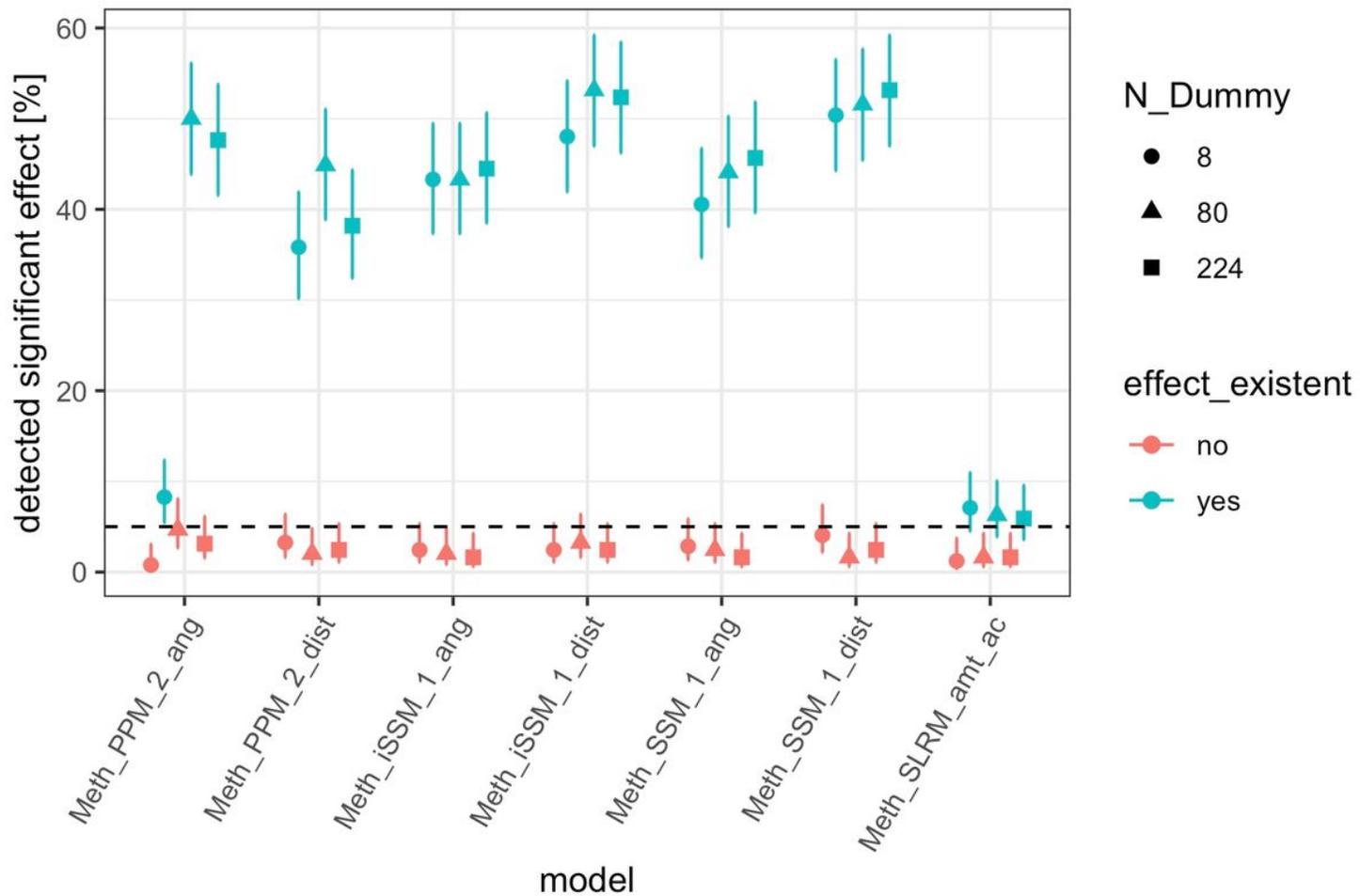


Figure 5

GEE results showing four representative models inferring long-range attraction with a focus on the dependency of model performance on dummy point number (the latter represented as circles vs. triangles vs. squares). The average statistical power is shown in green and the false-positive rate in red. Bars represent 95 % confidence intervals, symbols correspond to regression coefficient estimates.

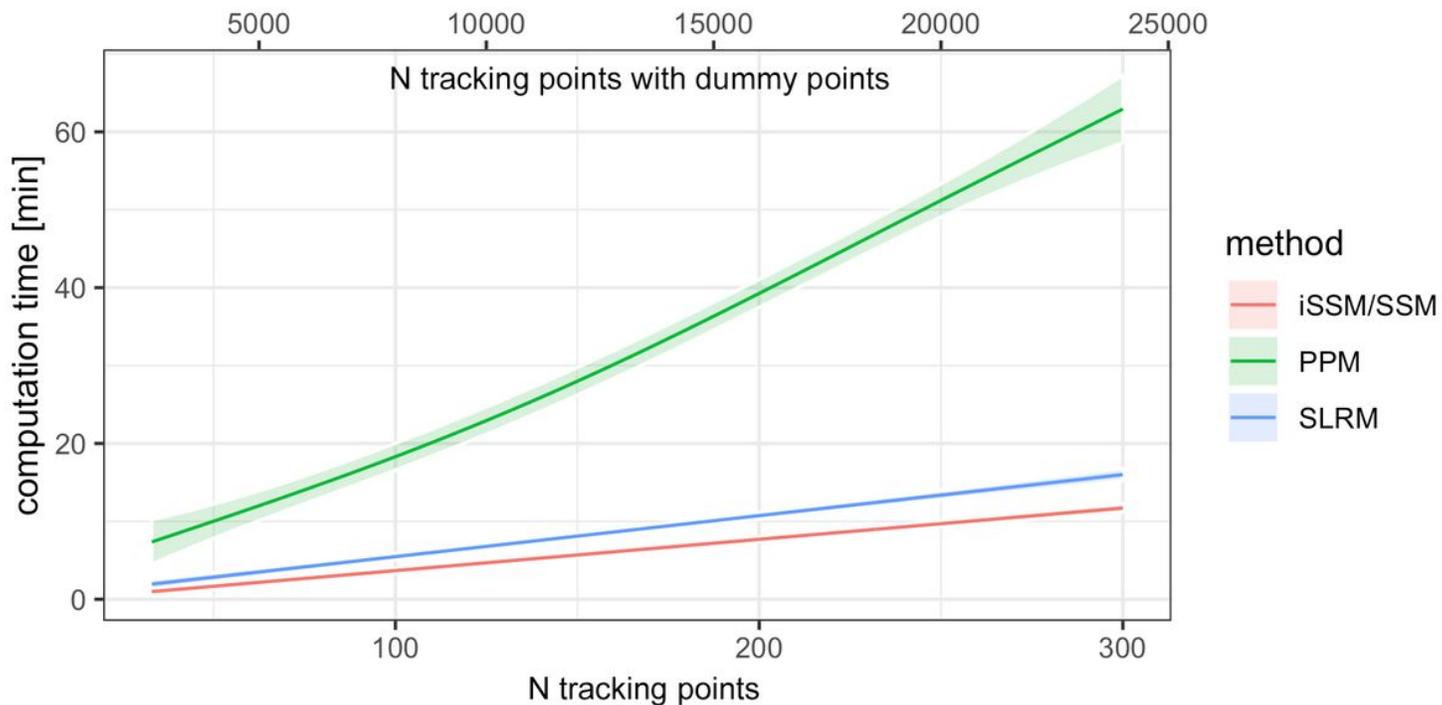


Figure 6

Non-parallel computation times for dummy point generation and model fit for three representative models, in relation to the number of tracking points. Shaded areas correspond to 95 % confidence bands.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SIFig01.jpg](#)
- [SIFig02.jpg](#)