The interplay between animal location accuracy and the decorrelation length scale of environmental variables when investigating environmental selection in marine organisms

Jérôme Pinti[1,*], Matthew Shatley[1], Helga S. Huntley[2], Aaron Carlisle[1], Matthew J. Oliver[1]

[1] College of Earth, Ocean, and Environment, University of Delaware, Lewes, DE 19958, USA

[2] Rowan University, Glassboro, NJ 08028, USA

* Corresponding author: jpinti@udel.edu

Running header: Limits of environmental selection detectability

Abstract

Many large pelagic organisms appear to select specific oceanic conditions, probably due to physiological, energetic, reproductive, or other life history needs. However, prior to characterizing these dynamics and determining their underlying drivers, the selection itself must be reliably identified. Our ability to do so depends on the quality of animal locations and on the heterogeneity of the environmental conditions driving selection. To draw meaningful conclusions about environmental selection of large organisms and therefore about their basic ecology, distribution, and ultimately their potential exploitation and conservation, limits of selection detectability must be established. Here, we investigate how animal location accuracy and environmental variable decorrelation length scales impact the ability to detect environmental selection by marine organisms. We create synthetic tracks cuing on environmental variables potentially relating to life history demands like sea surface temperature, chlorophyll-a concentration, and Lagrangian coherent structures. By artificially imposing different animal selection strengths and location accuracies, we assess how well environmental selection can be detected statistically in different cases. We found medium or strong selection to be reliably detected, even with relatively small samples and large position uncertainty, while weak selection presents similarly to no selection, especially for large sample sizes and position uncertainties. We therefore recommend using a selection strength threshold, which can significantly reduce the number of false positives while only increasing the risk of false negatives in cases of very weak selection, which are also less meaningful ecologically. We provide criteria to use when assessing confidence in environmental selection results for real marine organisms.

Keywords— movement ecology, habitat selection, environmental selection, biotelemetry, decorrelation length scale

1 Introduction

Marine ecosystems provide humanity with a number of important ecosystem services, including tourism and furnishing food for a growing portion of the world's population (Peterson & Lubchenco 1997). However, marine ecosystems (and consequently the services they provide) are increasingly threatened by pollution, overexploitation, habitat destruction, and climate change (Dulvy et al. 2003, Lotze et al. 2006, Halpern et al. 2008). Preserving marine ecosystems in the face of continued and increasing human exploitation requires careful spatial planning and management. Effective management is only possible if researchers and management agencies are able to observe and characterize the spatial ecology of marine species and ecosystems. Importantly, assessing the spatial ecology of oceanic species has historically been very difficult due to the sheer scale of their habitats (i.e. the open ocean), their highly mobile nature, the dynamic nature of oceanic ecosystems, and, perhaps most importantly, the logistical and technological difficulties associated with actually studying these animals in the wild (Block et al. 2002).

Understanding how mobile marine organisms react to various oceanographic features is critical for informing management of these species and ecosystems. Marine organisms may favour and select more or less strongly for specific environmental conditions, such as specific ranges of sea surface temperature or chlorophyll-a concentration that can be tracked remotely from satellites (Abrahms et al. 2019, Lee et al. 2021, Pinti et al. 2022). By associating the distribution of marine species with particular environmental conditions, we can characterize environmental selection in these animals and predict environmental mismatches that may arise because of climate change (Pörtner & Knust 2007, Møller et al. 2008). The first step to predicting the ecosystem's adaptive capacity is to identify the environmental selection (Kai et al. 2009, Oliver et al. 2019, Hazen et al. 2021, Fahlbusch et al. 2022) and to separate the selection signal from observation errors. These errors may arise from position uncertainty for the animal locations (Braun et al. 2023), or from a mismatch of the scales at which animals react to environmental variables and those at which they are observed (Scales et al. 2017).

To estimate animal locations, a wide array of electronic tag technologies are available, each with their own specificities and with different spatio-temporal resolutions and accuracies (Hussey et al. 2015). For example, acoustic telemetry records precisely when an animal carrying an acoustic tag is within the range of an acoustic receiver, but does not provide any information when animals are outside those detection ranges (Matley et al. 2021). Pop-up satellite archival tags (PSATs) rely on light levels to compute time of dawn, dusk, and day duration (sometimes coupled with sea surface temperature or other measurements) to estimate the position of individuals, but that means that position uncertainties routinely reach hundreds of kilometers (Gunn 1994, Sibert et al. 2003). Positions obtained from ARGOS tags are generally much more precise ($\sim$100s of meters to kms), although their uncertainty can sometimes reach 10s to 100s of kilometers, too. Additionally, ARGOS tags require the animal to stay at the surface for some period of time so that multiple ARGOS satellites can triangulate the tag position (Douglas et al. 2012). The development of Fastloc GPS tags has facilitated the collection of much more precise location data with uncertainties only up to a few meters, even for animals that only surface briefly (Dujon et al. 2014, Thomson et al. 2017). While these technologies all provide similar types of data, the large range of uncertainties associated with their respective location estimates means that they cannot be used interchangeably and that the ecological question that is being investigated should dictate the kind of biologging technology used.

As pop-up satellite archival tags are often the only possibility for studies of pelagic organisms that do not surface, large uncertainties are a particular challenge for assessing habitat selection in marine organisms. Using tracks with very large location uncertainty may yield incorrect results or blur selection signals.

Environmental conditions also vary in space and time, and their measurements come with their own uncertainties. Decorrelation length scales (DCLS) estimate the spatial distance over which a variable (e.g., temperature) remains correlated. Formally, the DCLS is the distance at which the correlation between the time series of a variable (or its anomaly) drops below a certain value, typically $1/e$ (Kuragano & Kamachi 2000, Hosoda & Kawamura 2004, De Benedetti & Moore 2017). It can be thought of as the typical scale of a coherent feature (such as a front or eddy), as captured by the dataset. These length scales vary greatly between different environmental variables; they also vary based on the resolution of the observations. We hypothesize that decorrelation length scales impact our ability to investigate habitat selection in marine organisms. To detect selection, animals would have to traverse length scales longer than the decorrelation length scale in a given region. We also postulate that the environmental selection of dynamic fields with very small decorrelation length scales needs to be studied at a fine scale, and thus requires more precise estimates of animal location and higher resolution environmental fields than the investigation of animals selecting for less dynamic environmental variables with larger decorrelation length scales.

Here, we investigate how the interplay between animal location accuracy, selection strength, and decorrelation length scales of environmental variables impacts our ability to statistically detect habitat selection by marine organisms. To do so, we create synthetic animal tracks with predetermined selection strengths for different environmental conditions. The specific environmental variables used are sea surface temperature (SST), chlorophyll-a concentration, and a Lagrangian metric, Finite-Time Lyapunov Exponents (FTLE), that captures the tendency of ocean currents to aggregate or disperse passive tracers (Haller & Yuan 2000, Shadden et al. 2009). By comparing these synthetic tracks with generated pseudo-absences that mimic the movements of similar but environmentally naive individuals, we are able to determine what level of location accuracy is required to definitively identify selection for the studied environmental variables. In addition, we introduce the notion of "effective selection" (combining both the test statistic and the significance level of Kolmogorov-Smirnov tests), significantly decreasing false-positive results in the absence of selection while only marginally increasing the rate of false-negative results at high sample sizes.

2 Methods

2.1 Environmental variables and decorrelation length scale

Three different environmental fields are used in this study: sea surface temperature (SST), chlorophyll-a concentration, and Finite-Time Lyapunov Exponent (FTLE) (figure 1 A-C).

SST and chlorophyll-a concentration are taken from MODIS-Aqua observations (JPL/OBPG/RSMAS 2020, NASA Goddard Space Flight Center et al. 2022). These fields have a native 9 km spatial resolution. The native time resolution is 1 day, but the data are processed so that each day consists of a backward rolling average of the previous 8 days to increase data coverage, by decreasing missing data due to clouds and incomplete daily satellite

coverage.

FTLE are a tool for differentiating regions of the ocean subject to high dispersal from those prone to accumulation of passive tracers (e.g. Waugh et al. 2012, Sulman et al. 2013, Peacock & Haller 2013, Allshouse & Peacock 2015, Haller 2015, Hadjighasem et al. 2017, Callies & von Storch 2022). They measure relative dispersion, i.e. how far nearby water parcels separate or come together over a specified time interval. FTLE are a function of ocean currents, which are taken here from a global run of the Hybrid Coordinate Ocean Model (HYCOM), experiment 19.1, carried out by the Naval Research Laboratory at Stennis, MS, and archived by the HYCOM consortium at `hycom.org`. The FTLE are computed using an along-trajectory velocity gradient integration on the model archive grid, which has a $1/12.5^\circ$ ($\sim$ 9 km) native resolution (Huntley et al. 2015).

We hypothesize that some marine organisms are attracted to regions of the ocean whose currents were conducive to accumulation of particles (e.g., plankton) over the recent past (Della Penna et al. 2015, Oliver et al. 2019, Lieber et al. 2023). Therefore, FTLE are derived from an integration backward in time over 3 days. The sign convention adopted here has large positive FTLE values correspond to highly attracting regions – water parcels that started far apart end up close together at the point in question and at the index time.

In addition, we also investigate the influence of data coverage by creating an environmental product corresponding to the FTLE product with the same data coverage as chlorophyll-a. In practice, we removed FTLE data where chlorophyll-a data was missing, effectively testing whether data gaps would affect our results.

For these three environmental variables (SST, chlorophyll-a, and FTLE), we compute the zonal (in the East-West direction) and meridional (in the North-South direction) decorrelation length scale (DCLS), defined as the e-folding scale of the variable anomalies (Hosoda & Kawamura 2004, De Benedetti & Moore 2017), i.e., the distance at which the correlation between two time series of the same variable drops below $1/e$. In practice, we start by subtracting the climatological signal (computed as the average for each calendar day over the period 2000-2010) from the daily signal. Then, for each reference point, we compute the correlation between the time series at that reference point and nearby time series. The distance at which the correlation drops below 1/e (in either the positive or negative direction) is taken as the DCLS at the reference point. The time series used here are 80 days long, spanning from May 1^{st} 2006 to July 19^{th} 2006, which is the same time period as the one over which the animal tracks are generated (see below).

2.2 Generation of synthetic animal tracks

For each of the three environmental variables, we created an array of 100 80-day long tracks, hereafter referred to as synthetic tracks. We vary the strength κ of environmental selection ($\kappa = 0$ for no selection, $\kappa = 0.25, 0.5, 0.75, 1, 2, 5, 10, 20$ for increasing selection strengths for high values of the target environmental variable)and the level of uncertainty σ assigned to the locations within these tracks. The standard deviations associated with the track locations are $\sigma = 0$ (exact location), 1 km, 10 km, 25 km, 50 km, and 111 km ($\sim 1^\circ$). Our total dataset of synthetic tracks then consists of 9 x 6 x 3 (9 different κ values and 6 different σ values for the 3 environmental variables used) sets of 100 tracks, i.e. 16,200 tracks.

The synthetic tracks of daily positions are generated as biased random walks, following Pinti et al. Pinti et al. (2022), although here step lengths are not fixed, to better mimic the daily movements of marine organisms.

For each track, a random starting location is picked in the North East Pacific, with latitude between 10 and 40°N and longitude between 140 and 130°W. Then, for each time step, the location of the highest value of the target variable within a 50 km radius is found, and the distance and bearing to that location are computed. The bearing of the actual step is then computed by modulating this bearing by a random angle drawn from a Von Mises distribution with mean 0 and concentration κ (figure 2A). The step length is the distance to the location of highest value modulated by a random distance drawn from a normal distribution with mean 0 and standard deviation $\frac{10}{\kappa+0.5}$ (figure 2B), except for $\kappa = 0$. The case $\kappa = 0$ mimics environmentally naive organisms, and to be sure that no information about the environment can be used for pseudo-absences generation, the step length is only drawn from a normal distribution with mean 50 km and standard deviation $\frac{10}{\kappa+0.5} = 20$ km.

2.3 Generation of pseudo-absences

Following Pinti et al. (Pinti et al. 2022), three different kinds of pseudo-absences are generated: Brownian motion, Correlated Random Walks, and Joint Correlated Random Walks. Pinti et al. also explored using Lévy walks as a null model, but concluded that they could lead to high rates of false positive results – hence our use of only three different kinds of null models here. For each synthetic track, 100 tracks of each null models are generated, for a total of 300 pseudo-absences for each presence record from the synthetic tracks. Uncertainties associated with pseudo-absence locations are the same as the uncertainties associated with the generated synthetic locations.

Each pseudo-absence track starts at the starting location of its corresponding synthetic presence track, and is reset to the actual synthetic animal location at the beginning of every month, following Pinti et al.'s analysis (Pinti et al. 2022). Therefore, pseudo-absence tracks are maximum 31 days long. This ensures that pseudo-absences are constrained to the area around the real track and could still be reasonably considered to be in the same environment (Pinti et al. 2022).

Each null model generates locations based on step lengths and turning angles. The three null models differ only in the way they generate these step lengths and turning angles. For Brownian motion, the turning angle is drawn from a uniform distribution bound between -180° and 180°, and the step length is drawn from a normal distribution with mean and standard deviation equal to the mean and standard deviation of the step length distribution of the synthetic track. Correlated Random Walks and Joint Correlated Random Walks are simulated by drawing turning angles and step lengths directly from the empirical distribution of the synthetic track. The difference between these two is that for correlated random walks step length and turning angles are drawn independently, whereas for joint correlated random walks the ordered pair (step length, turning angle) is drawn from a single step. Resulting pseudo-absences tracks for $\kappa = 1$ are pictured in figure S3.

2.4 Statistical analysis

Environmental data from the three variables are matched to each animal presence and pseudo-absence. When location uncertainty is greater than 0, however, it is not clear that the value at the recorded location is representative of the animal's environment. To account for this uncertainty in animal presence data, we average the environmental variables around each location, assuming a 2D Gaussian error distribution (i.e., the closer the observation is to the estimated location, the stronger the weight of this observation). The standard deviation for

the distribution is set equal to the position uncertainty, hence accounting for the higher probability of having the animal in locations close to the most likely position estimate while still capturing the environmental variability within the range.

To test if there is evidence of selection, we perform one-sided Kolmogorov-Smirnov (KS) tests, as implemented in ks.tests of the R stats package (R: A language and environment for statistical computing, R Development Core Team). KS tests compare two cumulative distributions, with the test statistic D being the maximum distance between the two cumulative distributions. Here, the cumulative distributions correspond to the distribution of environmental variables matched to presence and pseudo-absence tracks, respectively. As such, D can be used as a proxy to estimate selection strength. The null hypothesis is that the cumulative distribution of the target environmental variable for presences is "not less than" (or "not greater than") the cumulative distribution generated from pseudo-absences. For the "not less than" test, if the null hypothesis is rejected, it means that the presence cumulative distribution function is below that of the pseudo-absence. The environmental variable distributions are shifted toward higher values, and animals select for high values of the environmental variable (e.g. higher temperatures or areas with higher chlorophyll-a concentration) compared to environmentally naive organisms. Conversely, for the "greater than" test, it means that animals select for lower values than environmentally naive organisms. Throughout this manuscript and unless specified otherwise, the significance level is set at $\alpha = 0.05$. Practically, very weak selection can in many cases not be distinguished statistically from no selection based on a finite sample (nor is it ecologically relevant). Therefore, we also impose a threshold on the magnitude of the difference itself: we consider that animals are effectively selecting for higher (or lower) values of the environmental variable if the test result is significant ($p \leq 0.05$) and the test statistic D is greater than 0.05. This threshold was chosen as to decrease the rate of false-positive results in the absence of selection. This notion of effective selection allows us to discard false-positive results at high sample sizes in the absence of selection or at very weak selection strengths (see e.g. left columns of figures 5 and S8).

3 Results

3.1 Tracks and environmental variables

The three environmental products considered (SST, chlorophyll-a concentration, and FTLE) have different spatial characteristics (figure 1). First, the SST field is smoother than the chlorophyll-a field, which is smoother than the FTLE field. In the North East Pacific, the decorrelation length scales (DCLS) of the SST field vary between $\sim$50 and 800 km, while the DCLS of chlorophyll-a and FTLE vary between 0-400 km and 0-100 km, respectively. The median DCLS of the three products in the study region captures the differences in their spatial variability. SST has a median zonal DCLS of 130 km and a meridional DCLS of 111 km, while chlorophyll-a has median zonal and meridional DCLS of 46 and 47 km, and both FTLE DCLS are 27 km. Generally, zonal DCLS have higher maximum values than meridional DCLS, even though the means are very similar for chlorophyll-a and FTLE.

Because of the different spatial structures of these variables, the tracks generated based on the environmental selection of these fields are qualitatively different, especially at higher selection strengths (figure 3; for more samples see S2). Tracks selecting for higher SST tend to be more spread and south-ward, while tracks strongly

selecting for chlorophyll-a and FTLE organize around ridges. Pseudo-absence tracks do not have a well defined structure as they are not cuing on any environmental variable (figure S3).

The resulting distribution of SST, chlorophyll-a and FTLE for tracks strongly selecting for higher values of the environmental variables are shown in figure 4 (and histograms of environmental variable distributions for tracks with different selection strengths are plotted in figures S4, S5, and S6). For high selection strengths, as expected, the distribution of presence data is shifted towards higher values of the environmental variable compared to pseudo-absences. As κ decreases, the distribution of presence data shifts towards the distribution of pseudo-absences, all the way to $\kappa = 0$ where synthetic presence and pseudo-absence points have very similar distributions, consistent with the fact that this corresponds to animals not selecting for environmental variables. As σ increases (i.e. as location accuracy decreases), the overlap between distributions generated by animal presences and animal pseudo-absences increases, especially for FTLE and chlorophyll-a concentration (figure 4). This is because the larger scale averaging means that all values are more similar to each other, especially for environmental products with lower DCLS. The statistical analysis unravels the differences between these distributions as κ and σ vary.

3.2 Statistical analysis

To understand the interplay between position uncertainty σ, selection strength κ, and environmental variable DCLS, we have run a total of 162 experiments (6 values of σ, 9 values of κ, 3 environmental fields with different DCLS). In the following, we will focus on a selection of the results while describing the general patterns.

Figure 5 shows the value of D for selection for higher values as a function of sample size (number of tracks). The corresponding data for selection for lower values is given in figure S8). As suggested by the probability density functions of environmental variables, the stronger the selection and the higher the location accuracy, the higher the test statistic. In addition, the FTLE field with gaps did not reveal results qualitatively different from the complete FTLE field (figures S7 A-D and S8 M-P).

As the sample size increases, D decreases and converges to a point that depends on the environmental variable considered, the strength of the selection, and the location accuracy. In the absence of selection ($\kappa = 0$), some tests yield differences in distributions that are statistically significant results, yet with very small D (< 0.05).

A possibility to decrease these false-positive results would be to decrease the significance level below $\alpha = 0.05$. However, p-values usually decrease as sample sizes increase (figures S9-S16) and at large sample sizes they can reach values as low as 10^{-5} in the absence of selection (e.g. figures S10 O, P and S15 M, N, O, P). Decreasing the level of significance to 10^{-5} would strongly decrease the rate of false-positive results in the absence of selection, but it would also increase consequently the rate of false-negative results in the presence of selection with limited sample size or with low accuracy. For example, setting the significance limit to 10^{-5} would require more than 30 tracks to reliably detect selection for higher SST in the case $\kappa = 0.25$, even when the tracks are reported with perfect accuracy (figure S9 I). If the tracks have an uncertainty of 50km, which is not unusual for animals tracked with ARGOS tags (Douglas et al. 2012), detecting selection for high chlorophyll-a concentration requires more than 20 tracks in the case of strong selection ($\kappa = 2$ or more), and more than 80 tracks in the case of weak selection ($\kappa = 0.25$) (figure S10 C and K). These very large sample sizes are very rare in ecological studies, and setting up such a stringent threshold would hinder our ability to detect selection in most practical cases. The case

$\kappa = 0$ with perfect position accuracy reveals that different animals following different movement models are likely to sample slightly different environments even without a selection bias. These very small differences show up even with very large sample sizes, but they are not ecologically relevant. Thus, we decided to evaluate *effective selection* here, i.e. selection with a p-value below 0.05, but with $D \geq 0.05$. This ensures that statistically significant results are not wrongly categorized as "selecting" for higher (or lower) values of environmental variables, while still being able to detect selection when it is strong enough.

The ability to detect effective selection depends on sample size, selection strength, and location accuracy (figures 6, S26 and S17-S20 for selection for high values of environmental variables, and figure S25 for selection for low values of environmental variables). For FTLE with and without data gaps (figures 6 I-L and S26 A-D), the results are similar, suggesting that the method is not too sensitive to incomplete data coverage.

In the absence of selection, a large enough sample size allows us to confidently rule out effective selection (figure 6 A, E, and I). In the case of SST, more than $\sim$ 20 tracks (or 5 years of data) allows us to rule out effective selection irrespective of the location accuracy. For both chlorophyll-a and FTLE, the sample size needed to rule out selection varies between 13-18 tracks (3-4 years of data) and more than 45 tracks ($\sim$ 10 years of data) depending on location accuracy (the more uncertain the track, the more data are needed). At the other extreme, when the selection is strong ($\kappa \geq 2$), it is possible to detect effective environmental selection even with high location uncertainty and limited sample size (figure 6 D, H, and L). As selection strength is weakened($\kappa = 0.75$), large position uncertainty begins to have an impact on selection identification skill, at least for the variables with shorter DCLS (figure 6 C, G, and K). It is at weak selection strength ($\kappa = 0.25$) that the relationship between location accuracy, sample size, and identification skill becomes more complex (figure 6 B, F, and J). For chlorophyll-a and FTLE, the general pattern of higher detection skills for larger sample sizes holds, except for cases with high position uncertainty. For SST, regardless of position uncertainty, figure 6 B shows the counter-intuitive pattern that *smaller* sample sizes allow for a more reliable identification of effective selection bias. This is a direct result of the D thresholding (figure 5 B, F, and J): at small selection strength, the SST sample distributions for presence and pseudo-absence tracks differ only slightly for large sample sizes (and similarly for chlorophyll-a and FTLE in cases with large position uncertainty). The differences are insufficient to be picked up as indicators of bias. The larger D values for smaller sample sizes mean that our algorithm detects selection bias. However, we caution that these larger values are not an indication of the strength of the bias but an artefact of undersampling.

4 Discussion

4.1 Detecting and quantifying environmental selection

In this paper, we explored our ability to detect marine animal environmental selection using KS tests, depending on a number of environmental and technical limitations. The ability to detect environmental selection depends on the DCLS of the environmental variable considered, on the selection strength of the animal, and on the sample size and accuracy of the tracking dataset available. While we focus on SST, chlorophyll-a, and FTLE in this study, our results are transferable to other environmental variables. Once the DCLS of other variables are computed, one can refer to the variable with a similar DCLS in

figure 6 to assess confidence in their selection results.

Setting a threshold on the test statistic D enables us to detect effective selection, thus considerably decreasing the risk of false-positive results while moderately decreasing our ability to detect very weak selection. We found that setting the D threshold at 0.05 and the significance threshold at 0.05 yields to reliable selection detection (termed here effective selection), given enough data. The amount of data needed to have full confidence in a positive result depends on the DCLS of the investigated environmental variable, and can be defined as the amount of data necessary to fully rule out environmental selection in naive marine organisms (figure 6 A, E, and I). As an upper limit, more than 10 years of daily tracking data ($\sim 3,650$ daily location estimates) was enough to have no false-positive results, no matter the DCLS of the environmental variables investigated here.

Another important result of this study is the possibility to quantify selection using the test statistic D as a proxy for selection strength. Indeed, for a given track accuracy, the stronger the selection strength, the higher the asymptotic D value as sample size increases (figure 5). Provided with a large enough dataset of animal tracks, performing KS tests on increasing fractions of the dataset can reveal if the test statistic obtained for the entire data set is close to a converged value or not – and thus if the selection strength can be confidently quantified.

4.2 The different types of animal tracking data and their impact on environmental selection detection

A major discrepancy between real animal tracking datasets and the presence tracks generated here is that observed data can have varying accuracy and are not necessarily evenly spaced in time. Processing the data with a state-space model allows us to resample the track evenly (Jonsen et al. 2005). However, state-space models usually assume that animals follow correlated random walks without accounting for environmental selection, thereby biasing processed data against selection detection. In addition, the spatial uncertainties associated with processed points temporally far away from observed points is usually greater – further increasing the range of location uncertainties within tracks. As we assumed here that location uncertainty was homogeneous within the entire dataset, adapting this method to real animal location may require discarding location estimates that are too uncertain.

This consideration is particularly important for more uncertain tracking methods. Fastloc GPS tags provide accuracy within 700 m and most of the time within 50 m (Dujon et al. 2014), well below the resolution of the environmental products considered here. However, ARGOS and PSATs have larger error ranges. ARGOS tags provide locations with a quality location class (Douglas et al. 2012), allowing the filtering of data to keep only the most precise – at the expense of sample size (Thomson et al. 2017). PSATs, because they mostly geolocate animals thanks to light levels and SST (even though other variables such as bathymetry, magnetic field, or temperature profiles can be used (Nielsen et al. 2019, 2020, Braun

et al. 2018)), have an even wider range, with errors in actual positions routinely around 1° in longitude and latitude (Wilson et al. 2007). Yet, for highly migratory marine species, it appears that PSATs can perform better than ARGOS tags when building species distribution models (using other methods than ours, such as generalized linerar or additive mixed models, or boosted regression trees) using variables with high DCLS (Braun et al. 2023). This highlights the importance of considering the scale of the question being investigated. Following the results of this study, we recommend using either ARGOS or GPS tag records when investigating environmental selection in marine organisms for variables with low DCLS. In general, generating the most accurate location data possible is often the safest way to get reliable results. In particular, when limited to PSATs data because of the ecology of the animal, relying on manufacturer-provided geolocation models often yield inaccurate results relative to more advanced geolocation models (Braun et al. 2018, Nielsen et al. 2023). More precise locations will naturally lead to more accurate ecological results, and will also enable investigating ecological questions requiring more precise data, such as the selection of environmental variables with comparatively lower DCLS than would be possible without these data.

4.3 Scale, resolution, and accuracy of environmental variables

In addition to the resolution at which animal locations are observed, the scale at which we acquire oceanographic data is important when investigating environmental selection. The datasets we use here have a fixed 9 km resolution (8 km in the case of FTLE), smoothing variability at smaller scales and making sub-mesoscale features undetectable. Spatial resolution can be finer if environmental data are acquired differently. For example, Lagrangian coherent structures have been computed from high frequency radars from 6 km down to 500 m depending on the frequency used (Kim et al. 2011, Berta et al. 2014, Oliver et al. 2019), but usually at the expense of spatial coverage. Spatially averaging the environmental data increases spatial coverage, creating a trade-off between data coverage and spatial resolution. In addition to technical considerations regarding this trade-off, this raises ecological questions as to what scale is appropriate when investigating selection of marine organisms. A larger scale may be necessary to have a good spatial coverage, but may blur small variations that animals may use and select for. As it is not possible to detect selection happening at a finer scale than the resolution of the data, it is important to match the scale at which animals can move and the scale at which we can acquire environmental data as well as possible.

The way oceanographic data are acquired dictates their resolution, but also their accuracy. While data observed in situ (and, to a lesser extent, remotely sensed data) have a relatively strong spatio-temporal accuracy, data generated by models may have lower accuracy. For example, the FTLE used here are computed from HYCOM model outputs and not directly observed. Several studies have found that the uncertainties in predictions of trajectories and Lagrangian Coherent Structures computed from

HYCOM model outputs are, on average, 50 km, with outliers as large as 100 km (Muscarella et al. 2015, Huntley et al. 2011, Thoppil et al. 2021). Here, this discrepancy is irrelevant as the tracks are synthetic and generated using the model FTLE outputs directly, but it is of importance when dealing with real tracks of animals that are cuing on actual Lagrangian structures. Incorporating the environmental data uncertainty in this method will likely have the same effect as considering the track uncertainty – it will average out specific values that animals may cue on with background conditions, thus decreasing our ability to detect selection.

Data coverage is another factor that might influence our ability to detect selection. The amount of missing data differs for the three environmental fields (figure S1). As the FTLE field is a model output, there are no data gaps, while there are some in the SST and even more in the chlorophyll-a fields. Thus, not all points could be matched to all environmental variables (table 1). While the same number of tracks were used with all three variables, this translates into a different number of data points with matched environmental variables, with FTLE having the most data points and chlorophyll-a the least. The number of data points also increases as the assumed accuracy of location estimates decreases, as less precise estimates mean that the average is performed across a larger area and thus more likely to encompass existing environmental data.

4.4 Increasing sample size: beware of heterogeneity

Larger animal tracking datasets lead to improved confidence in selection detection. Therefore, it is tempting to aggregate data from different sources, geographic locations, time of the year, or even from different years to increase sample size. However, disregarding the variability that can exist in datasets may yield to inaccurate results. Populations from different geographic areas may experience a different range of environmental conditions. Thus, different populations might be adapted to and select for different environmental conditions – such as shark species living in pelagic environments and very clear waters in the Pacific Ocean but near the coasts in turbid environments in the Atlantic Ocean (Merson & Pratt, Jr. 2001, Papastamatiou et al. 2006). Animals may change behavior and selection strategy as they grow (ontogenetic niche partitioning, (Grubbs 2010)) or throughout the year, for example when they migrate to colder, lower latitudes (Horton et al. 2011), or when they move between offshore and inshore grounds with different productivity levels (Weng et al. 2007). Finally, conditions may change over time, for example as a result of climate change – resulting in changing experienced environmental conditions for organisms, such as a decrease in available prey (Meyer-Gutbrod et al. 2023).

Depending on the flexibility of animal movements, these longer-term, multi-year changes in environmental conditions may be chosen or imposed on marine organisms. Migrating organisms can follow specific cues and select for particular environmental conditions, or they can also use memory to reproduce movement patterns of previous years – potentially resulting in a mismatch with their optimal environmen-

tal conditions (Abrahms et al. 2019, Fagan 2019). How the changing baseline will impact environmental selection in marine organisms is yet to be determined and will depend on how much animals rely on memory vs. environmental cues. To this end, it would be useful to compare animals tracks not only to instantaneous environmental products but also to past variables and variables averaged at different temporal scales to understand the importance of conditions experienced in the past in shaping marine animals' movements and environmental niches.

5 Conclusion

The accuracy of animal location, as well as the spatial variability in environmental variables (measured through decorrelation length scales) impacts our ability to detect selection in marine organisms. As previously shown (Pinti et al. 2022), increasing sample size allows for more robust detection of environmental selection, except when selection is very weak. Setting a significance level at 0.05 and a cumulative distribution difference threshold level at 0.05 to detect effective selection in marine organisms significantly reduces the risk of false-positive results, while only increasing the risk of false-negative results in the case of very weak selection strength.

The amount and accuracy of tracking data needed to detect selection depends on the DCLS of the environmental variable tested. In practice, data acquired with Fastloc GPS (and ARGOS) tags are the most precise and thus the most efficient to use when it comes to detecting and quantifying selection strength, even though a large fraction of marine organisms do not surface and therefore cannot be studied with such technology.

Finally, it is important to mention that we are detecting correlation but not necessarily causation. While we can confidently say whether organisms preferentially associate with certain environmental variables thanks to this method, we cannot ascertain if they actively target these variables or variables correlated to them, or whether they target these conditions thanks to immediate cues or thanks to memory and movement patterns acquired through social learning. Testing for environmental selection with past environmental conditions and at different temporal scales might help answer this question.

6 Acknowledgements

This work was supported by NASA Earth Science division, ROSES 2020 Program (award number 80NSSC21K1145). We thank three anonymous reviewers for their comments that helped us improve this manuscript.

7 Competing interests

The authors declare that they have no competing interests.

8 Availability of data and materials

All code used to perform the analysis is available at `https://github.com/JeromeAqua/Selection_accuracy/`.

References

Abrahms B, Hazen EL, Aikens EO, Savoca MS, Goldbogen JA, Bograd SJ, Jacox MG, Irvine LM, Palacios DM, Mate BR (2019) Memory and resource tracking drive blue whale migrations. Proc Natl Acad Sci USA 116(12):5582–5587

Allshouse MR, Peacock T (2015) Refining finite-time Lyapunov exponent ridges and the challenges of classifying them. Chaos 25(8)

Berta M, Bellomo L, Magaldi MG, Griffa A, Molcard A, Marmain J, Borghini M, Taillandier V (2014) Estimating Lagrangian transport blending drifters with HF radar data and models: Results from the TOSCA experiment in the Ligurian Current (North Western Mediterranean Sea). Prog Oceanogr 128:15–29

Block BA, Costa DP, Boehlert GW, Kochevar RE (2002) Revealing pelagic habitat use: The tagging of Pacific pelagics program. Oceanol Acta 25(5):255–266

Braun CD, Arostegui MC, Farchadi N, Alexander M, Afonso P, Allyn A, Bograd SJ, Brodie S, Crear DP, Culhane EF, Curtis TH, Hazen EL, Kerney A, Lezama-Ochoa N, Mills KE, Pugh D, Queiroz N, Scott JD, Skomal GB, Sims DW, Thorrold SR, Welch H, Young-Morse R, Lewison R (2023) Building use-inspired species distribution models: using multiple data types to examine and improve model performance. Ecol Appl (February):1–20

Braun CD, Galuardi B, Thorrold SR (2018) HMMoce: An R package for improved geolocation of archival-tagged fishes using a hidden Markov method. Methods Ecol Evol 9(5):1212–1220

Callies U, von Storch H (2022) Extreme separations of bottle posts in the southern Baltic Sea –tentative interpretation of an experiment-of-opportunity. Oceanologia 1886

De Benedetti M, Moore GW (2017) Impact of Resolution on the Representation of Precipitation Variability Associated With the ITCZ. Geophys Res Lett 44(24):12,519–12,526

445 Della Penna A, De Monte S, Kestenare E, Guinet C, D'Ovidio F (2015) Quasi-planktonic behavior of
446 foraging top marine predators. Sci Rep 5(May):1–10

447 Douglas DC, Weinzierl R, C Davidson S, Kays R, Wikelski M, Bohrer G (2012) Moderating Argos location
448 errors in animal tracking data. Methods Ecol Evol 3(6):999–1007

449 Dujon AM, Lindstrom RT, Hays GC (2014) The accuracy of Fastloc-GPS locations and implications for
450 animal tracking. Methods Ecol Evol 5(11):1162–1169

451 Dulvy NK, Sadovy Y, Reynolds JD (2003) Extinction vulnerability in marine populations. Fish Fish
452 4(1):25–64

453 Fagan WF (2019) Migrating whales depend on memory to exploit reliable resources. Proc Natl Acad Sci
454 USA 116(12):5217–5219

455 Fahlbusch JA, Czapanskiy MF, Calambokidis J, Cade DE, Abrahms B, Hazen EL, Goldbogen JA (2022)
456 Blue whales increase feeding rates at fine-scale ocean features. Proc Royal Soc B 289(1981)

457 Grubbs RD (2010) Ontogenetic shifts in movements and habitat use. In: *Sharks and their relatives II*,
458 page 32. CRC Press

459 Gunn J (1994) The development and use of archival tags for studying the migration, behavior and
460 physiology of southern bluefin tuna, with an assessment of the potential for transfer of technology to
461 groundfish research. Proc ICES Symp Fish Migration 21:1–23

462 Hadjighasem A, Farazmand M, Blazevski D, Froyland G, Haller G (2017) A critical comparison of La-
463 grangian methods for coherent structure detection. Chaos 27(5)

464 Haller G (2015) Lagrangian coherent structures. Annu Rev Fluid Mech 47:137–162

465 Haller G, Yuan G (2000) Lagrangian coherent structures and mixing in two-dimensional turbulence.
466 Physica D 147(3-4):352–370

467 Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C, Bruno JF, Casey KS, Ebert C,
468 Fox HE, Fujita R, Heinemann D, Lenihan HS, Madin EMP, Perry MT, Selig ER, Spalding M, Steneck
469 R, Watson R (2008) A Global Map of Human Impact on Marine Ecosystems. Science 319(5865):948–952

470 Hazen EL, Abrahms B, Brodie S, Carroll G, Welch H, Bograd SJ (2021) Where did they not go? Con-
471 siderations for generating pseudo-absences for telemetry-based habitat models. Mov Ecol 9(1):1–13

472 Horton TW, Holdaway RN, Zerbini AN, Hauser N, Garrigue C, Andriolo A, Clapham PJ (2011) Straight
473 as an arrow: Humpback whales swim constant course tracks during long-distance migration. Biol Lett
474 7(5):674–679

475 Hosoda K, Kawamura H (2004) Global space-time statistics of sea surface temperature estimated from
476 AMSR-E data. Geophys Res Lett 31(17):1–5

477 Huntley HS, Lipphardt BL, Jacobs G, Kirwan AD (2015) Clusters, deformation, and dilation: Diagnostics
478 for material accumulation regions. J Geophys Res Oceans 120(10):6622–6636

479 Huntley HS, Lipphardt BL, Kirwan AD (2011) Lagrangian predictability assessed in the East China Sea.
480 Ocean Model 36(1-2):163–178

481 Hussey NE, Kessel ST, Aarestrup K, Cooke SJ, Cowley PD, Fisk AT, Harcourt RG, Holland KN, Iverson
482 SJ, Kocik JF, Flemming JE, Whoriskey FG (2015) Aquatic animal telemetry: A panoramic window
483 into the underwater world. Science 348(6240):1255642

484 Jonsen I, Flemming JM, Myers R (2005) Robust State-Space Modeling of Animal Movement Data.
485 Ecology 86(11):2874–2880

486 JPL/OBPG/RSMAS (2020) GHRSST Level 2P Global Sea Surface Skin Temperature from the Moderate
487 Resolution Imaging Spectroradiometer (MODIS) on the NASA Aqua satellite (GDS2). MODIS Aqua
488 L2P swath SST data set ver 20190

489 Kai ET, Rossi V, Sudre J, Weimerskirch H, Lopez C, Hernandez-Garcia E, Marsac F, Garçon V (2009)
490 Top marine predators track Lagrangian coherent structures. Proc Natl Acad Sci USA 106(20):8245–
491 8250

492 Kim SY, Terrill EJ, Cornuelle BD, Jones B, Washburn L, Moline MA, Paduan JD, Garfield N, Largier JL,
493 Crawford G, Kosro PM (2011) Mapping the U.S. West Coast surface circulation: A multiyear analysis
494 of high-frequency radar observations. J Geophys Res Oceans 116(3):1–15

495 Kuragano T, Kamachi M (2000) Global statistical space-time scales of oceanic variability estimated from
496 the TOPEX/POSEIDON altimeter data. J Geophys Res Oceans 105(C1):955–974

497 Lee KA, Butcher PA, Harcourt RG, Patterson TA, Peddemors VM, Roughan M, Harasti D, Smoothey AF,
498 Bradford RW (2021) Oceanographic conditions associated with white shark (*Carcharodon carcharias*)
499 habitat use along eastern Australia. Mar Ecol Prog Ser 659:143–159

500 Lieber L, Füchtencordsjürgen C, Hilder RL, Revering PJ, Siekmann I, Langrock R, Nimmo-Smith WAM
501 (2023) Selective foraging behavior of seabirds in small-scale slicks. Limnol Oceanogr Lett 8(2):286–294

502 Lotze HK, Lenihan HS, Bourque BJ, Bradbury RH, Cooke RG, Kay MC, Kidwell SM, Kirby MX,
503 Peterson CH, Jackson JBC (2006) Depletion, Degradation, and Recovery Potential of Estuaries and
504 Coastal Seas. Science 312(5781):1806–1809

505 Matley JK, Klinard NV, Barbosa Martins AP, Aarestrup K, Aspillaga E, Cooke SJ, Cowley PD, Heupel
506 MR, Lowe CG, Lowerre-Barbieri SK, Mitamura H, Moore JS, Simpfendorfer CA, Stokesbury MJ,
507 Taylor MD, Thorstad EB, Vandergoot CS, Fisk AT (2021) Global trends in aquatic animal tracking
508 with acoustic telemetry. Trends Ecol Evol 37(1):79–94

509 Merson RR, Pratt, Jr HL (2001) Distribution, movements and growth of young sandbar sharks, *Car-*
510 *charhinus plumbeus*, in the nursery grounds of Delaware Bay. Environ Biol Fishes 61:13–24

511 Meyer-Gutbrod EL, Davies KT, Johnson CL, Plourde S, Sorochan KA, Kenney RD, Ramp C, Gosselin
512 JF, Lawson JW, Greene CH (2023) Redefining North Atlantic right whale habitat-use patterns under
513 climate change. Limnol Oceanogr (68):S71–S86

514 Møller AP, Rubolini D, Lehikoinen E (2008) Populations of migratory bird species that did not show a
515 phenological response to climate change are declining. Proc Natl Acad Sci USA 105(42):16195–16200

516 Muscarella P, Carrier MJ, Ngodock H, Smith S, Lipphardt BL, Kirwan AD, Huntley HS (2015) Do
517 assimilated drifter velocities improve lagrangian predictability in an operational ocean model? Mon
518 Weather Rev 143(5):1822–1832

519 NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group (2022)
520 Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Ocean Color Data

521 Nielsen JK, Bryan DR, Rand KM, Arostegui MC, Braun CD, Galuardi B, McDermott SF (2023) Geolo-
522 cation of a demersal fish (Pacific cod) in a high-latitude island chain (Aleutian Islands, Alaska). Anim
523 Biotelemetry 11(1):1–22

524 Nielsen JK, Mueter FJ, Adkison MD, Loher T, McDermott SF, Seitz AC (2019) Effect of study area
525 bathymetric heterogeneity on parameterization and performance of a depth-based geolocation model
526 for demersal fishes. Ecol Modell 402(April):18–34

527 Nielsen JK, Mueter FJ, Adkison MD, Loher T, McDermott SF, Seitz AC (2020) Potential utility of
528 geomagnetic data for geolocation of demersal fishes in the North Pacific Ocean. Anim Biotelemetry
529 8(1):1–20

530 Oliver MJ, Kohut JT, Bernard K, Fraser W, Winsor P, Statscewich H, Fredj E, Cimino M, Patterson-
531 Fraser D, Carvalho F (2019) Central place foragers select ocean surface convergent features despite
532 differing foraging strategies. Sci Rep 9(1):157

533 Papastamatiou YP, Wetherbee BM, Lowe CG, Crow GL (2006) Distribution and diet of four species
534 of carcharhinid shark in the Hawaiian Islands: Evidence for resource partitioning and competitive
535 exclusion. Mar Ecol Prog Ser 320:239–251

536 Peacock T, Haller G (2013) Lagrangian coherent structures: The hidden skeleton of fluid flows. Phys
537 Today 66(2):41–47

538 Peterson C, Lubchenco J (1997) Marine ecosystem services. In: *Nature's Services: Societal Dependence*
539 *on Natural Ecosystems*, pages 177–194. Island Press, Washington DC. ISBN 978-1559634762

540 Pinti J, Shatley M, Carlisle A, Block BA, Oliver MJ (2022) Using pseudo - absence models to test for
541 environmental selection in marine movement ecology : the importance of sample size and selection
542 strength. Mov Ecol 10(60):1–17

543 Pörtner HO, Knust R (2007) Climate Change Affects Marine Fishes Through the Oxygen Limitation of
544 Thermal Tolerance. Science 315(5808):95–97

545 Scales KL, Hazen EL, Jacox MG, Edwards CA, Boustany AM, Oliver MJ, Bograd SJ (2017) Scale of
546 inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of
547 environmental data. Ecography 40(1):210–220

548 Shadden SC, Lekien F, Paduan JD, Chavez FP, Marsden JE (2009) The correlation between surface
549 drifters and coherent structures based on high-frequency radar data in Monterey Bay. Deep Sea Res 2
550 Top Stud Oceanogr 56(3-5):161–172

551 Sibert JR, Musyl MK, Brill RW (2003) Horizontal movements of bigeye tuna (*Thunnus obesus*) near
552 Hawaii determined by Kalman filter analysis of archival tagging data. Fish Oceanogr 12(3):141–151

553 Sulman MH, Huntley HS, Lipphardt BL, Jacobs G, Hogan P, Kirwan AD (2013) Hyperbolicity in tem-
554 perature and flow fields during the formation of a Loop Current ring. Nonlinear Process Geophys
555 20(5):883–892

556 Thomson JA, Börger L, Christianen MJ, Esteban N, Laloë JO, Hays GC (2017) Implications of location
557 accuracy and data volume for home range estimation and fine-scale movement analysis: comparing
558 Argos and Fastloc-GPS tracking data. Mar Biol 164(10):1–9

559 Thoppil PG, Frolov S, Rowley CD, Reynolds CA, Jacobs GA, Joseph Metzger E, Hogan PJ, Barton N,
560 Wallcraft AJ, Smedstad OM, Shriver JF (2021) Ensemble forecasting greatly expands the prediction
561 horizon for ocean mesoscale variability. Commun Earth Environ 2(1):1–9

562 Waugh DW, Keating SR, Chen ML (2012) Diagnosing ocean stirring: Comparison of relative dispersion
563 and finite-time Lyapunov exponents. J Phys Oceanogr 42(7):1173–1185

564 Weng KC, Boustany AM, Pyle P, Anderson SD, Brown A, Block BA (2007) Migration and habitat of
565 white sharks (*Carcharodon carcharias*) in the eastern Pacific Ocean. Mar Biol 152(4):877–894

566 Wilson SG, Stewart BS, Polovina JJ, Meekan MG, Stevens JD, Galuardi B (2007) Accuracy and precision
567 of archival tag data: A multiple-tagging study conducted on a whale shark (*Rhincodon typus*) in the
568 Indian Ocean. Fish Oceanogr 16(6):547–554

9 Tables

Table 1: Fraction of animal presence location estimates with environmental data for different location standard error σ. Each cell of this table corresponds to 72,000 data points (100 80-day tracks for 9 different values of κ).

σ	**0** km	**1** km	**10** km	**25** km	**50** km	**111** km
chlorophyll-a	40.7 %	67.3 %	87.7 %	96.2 %	99.6 %	100 %
SST	86.5 %	95.5 %	99.6 %	99.9 %	100 %	100 %
FTLE	100 %	100 %	100 %	100 %	100 %	100 %

10 Figures

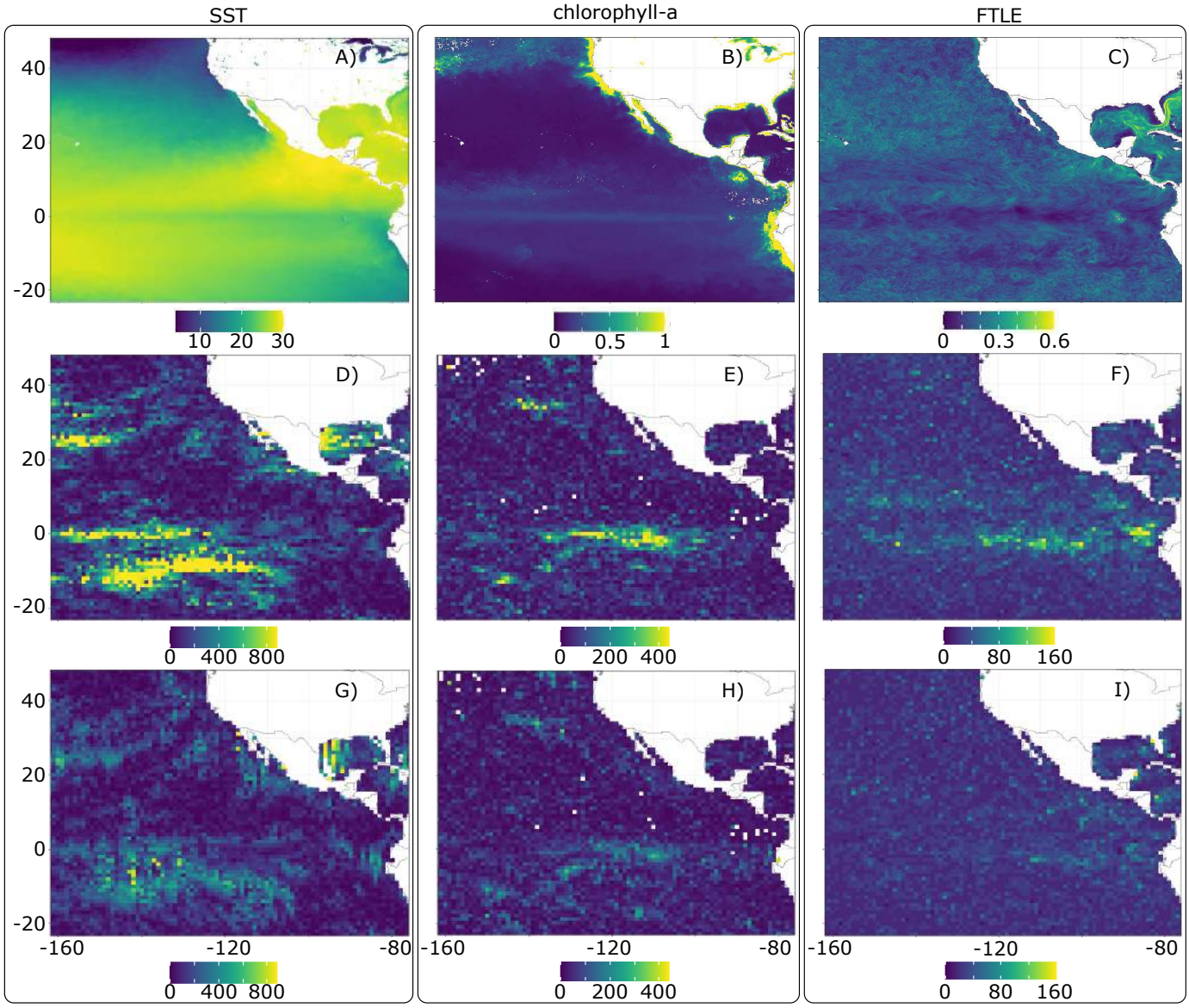

Figure 1: (A)-(C): Environmental conditions averaged over the period May 1^{st} 2006 - July 19^{th} 2006, corresponding to the time period of the simulated tracks. (A) Sea Surface Temperature (SST, in °C), (B) chlorophyll-a concentration ($mg\ m^{-3}$), (C) Finite-Time Lyapunov Exponent (FTLE, in day^{-1}). (D)-(F): Zonal (East-West) decorrelation length scale (in km) for (D) SST, (E) chlorophyll-a concentration, and (F) FTLE. (G)-(I): Meridional (North-South) decorrelation length scale (in km) for (G) SST, (H) chlorophyll-a concentration, and (I) FTLE. x and y axes of all panels are longitude and latitude, respectively.

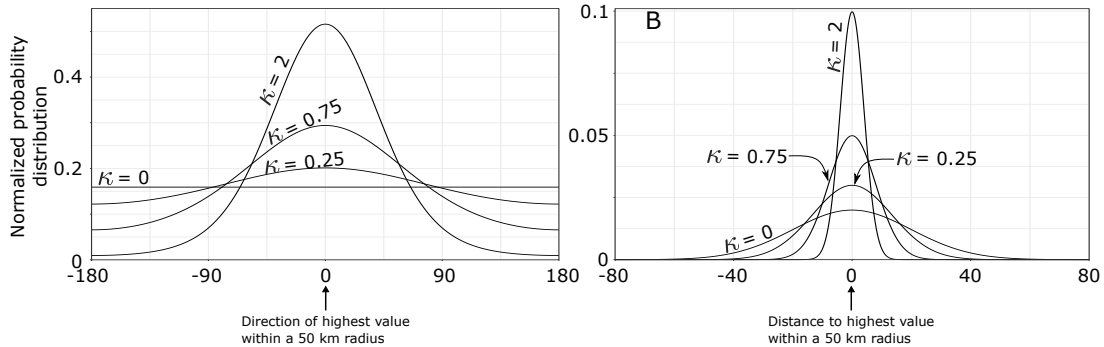

Figure 2: Von Mises distribution of relative bearing (A) and normal distribution of step length (B) for movement models biased toward the highest value of the environment variable within a 50km radius for different values of κ.

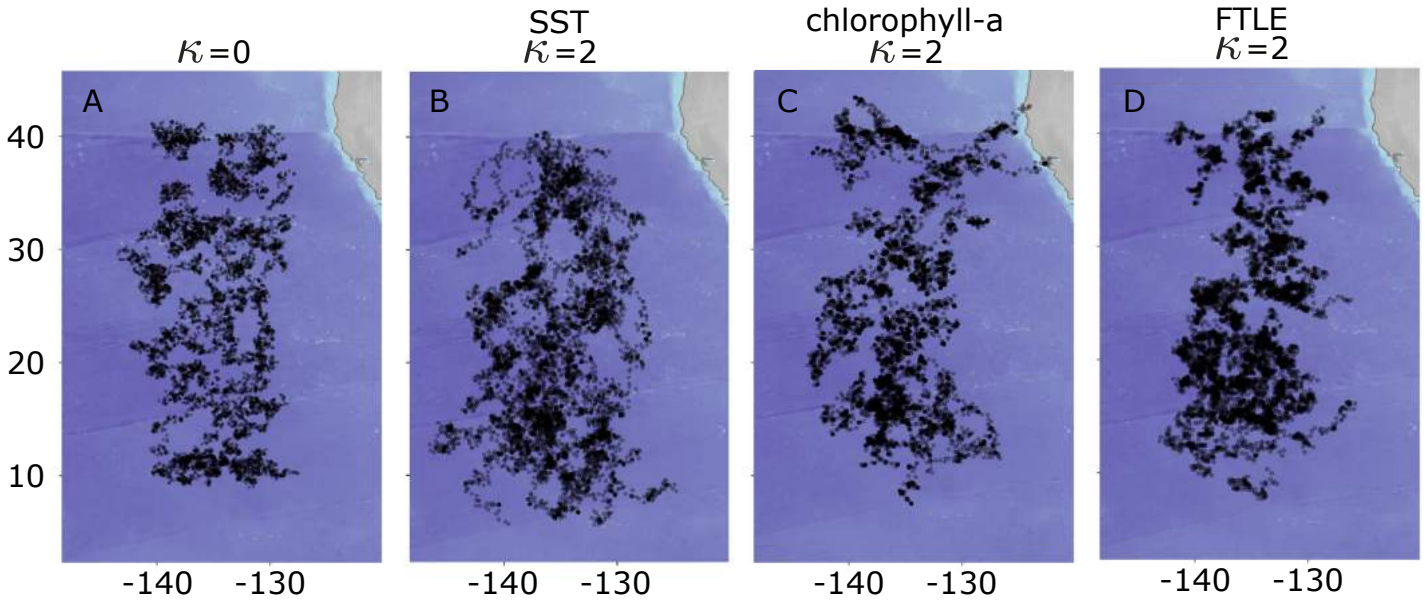

Figure 3: Synthetic tracks of animals selecting for no environmental variable (A), and strongly selecting ($\kappa = 2$) for high SST (B), high chlorophyll-a concentration (C), and high FTLE values (D). There are 100 different 80-day tracks per panel. As the case $\kappa = 0$ corresponds to no environmental selection, the same set of tracks was used for all three environmental variables. x and y axes of all panels are longitude and latitude, respectively. Tracks for additional κ values are shown in figure S2.

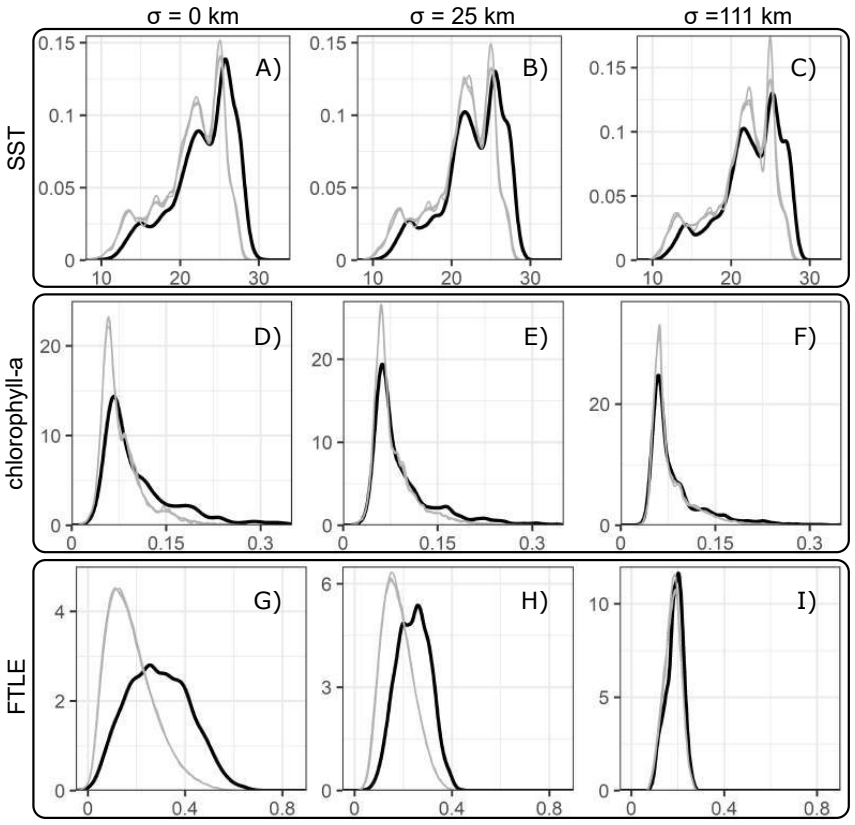

Figure 4: Probability density functions of SST (A, B, C, in $°C$), of chlorophyll-a concentration (D, E, F, in $mg\ m^{-3}$) and of FTLE (G, H, I, in day^{-1}) at synthetic presence (black) and pseudo-absence (grey) points, for presence tracks biased towards high values of the environmental variables (κ =2). x axes represent the environmental variable (SST, chlorophyll-a concentration, or FTLE), y axes represent the probability density function. The first column (A, D, G) corresponds to locations known with perfect accuracy, the second column (B, E, H) to locations known with moderate accuracy (standard deviation of 25 km), and the third column (C, F, I) to locations known with weak accuracy (standard deviation of 111 km).

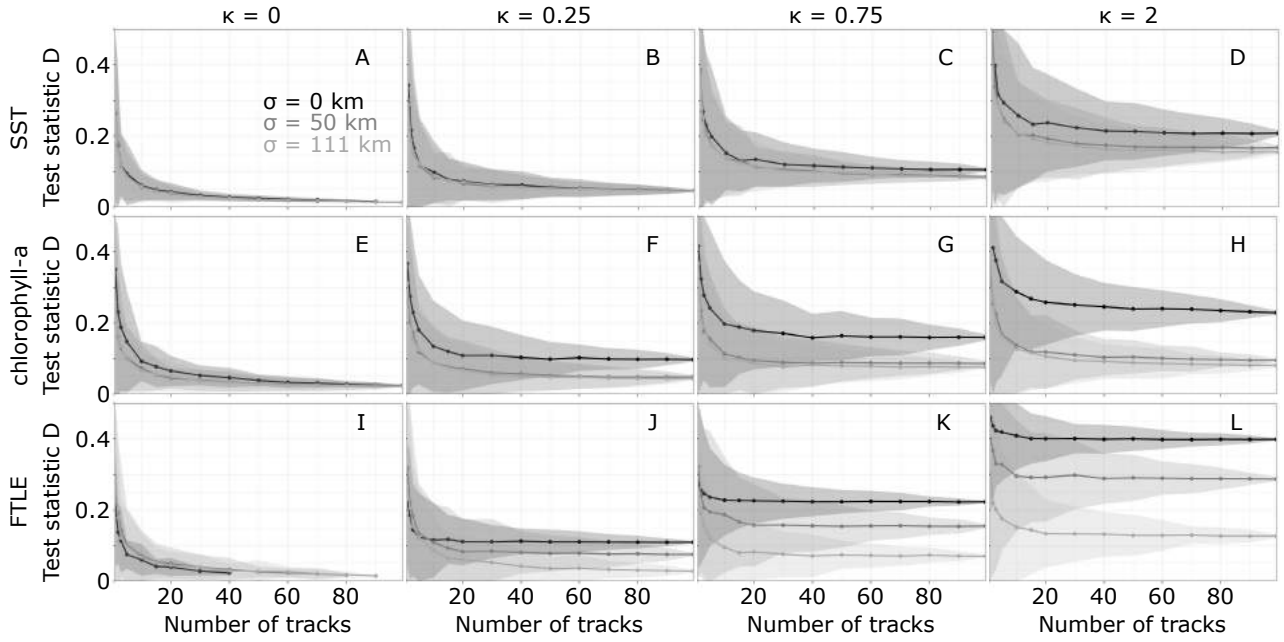

Figure 5: Statistic D of tests of selection for higher values of SST (A, B, C, D), chlorophyll-a (E, F, G, H), and FTLE (I, J, K, L) at different selection strengths (in columns) and for three different accuracies of geolocation estimates (shades of grey). Shaded areas show the ensemble-mean estimate ± 3 standard deviations. Only test statistics of tests with $p < 0.05$ are included.

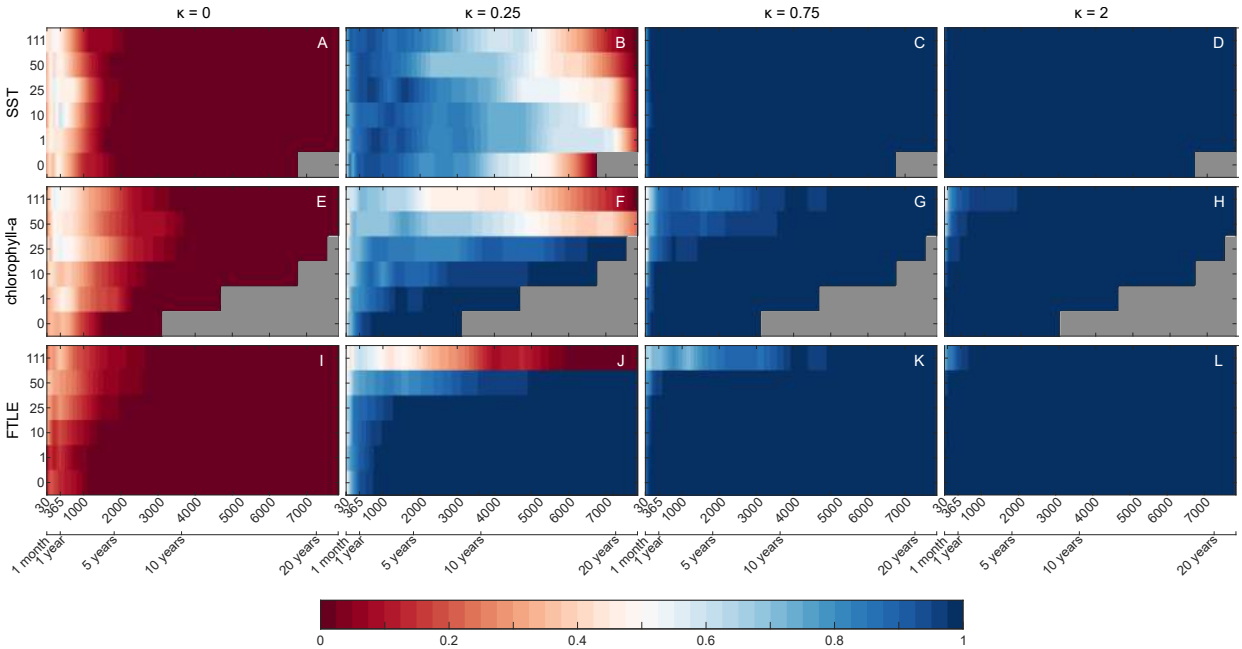

Figure 6: Fraction of tests showing effective selection for higher values of SST (A, B, C, D), chlorophyll-a (E, F, G, H), and FTLE (I, J, K, L) at different selection strengths (in columns), location accuracy (rows of each panel) and sample sizes (columns of the each panel). Sample sizes are plotted as number of data points for which environmental data is available (in both number of data points and corresponding time span). At high location accuracy, we average the environmental variables over small areas, resulting in many track locations with missing data points for environmental products acquired with satellites. This lack of data is apparent in panels A-H, where grey patches correspond to missing data. We would need to generate more than 8,000 days of tracking data to have 8,000 of tracking data with matching environmental conditions.

Supporting Information 1 for "The interplay between animal location accuracy and the decorrelation length scale of environmental variables when investigating environmental selection in marine organisms"

A Supplementary figures

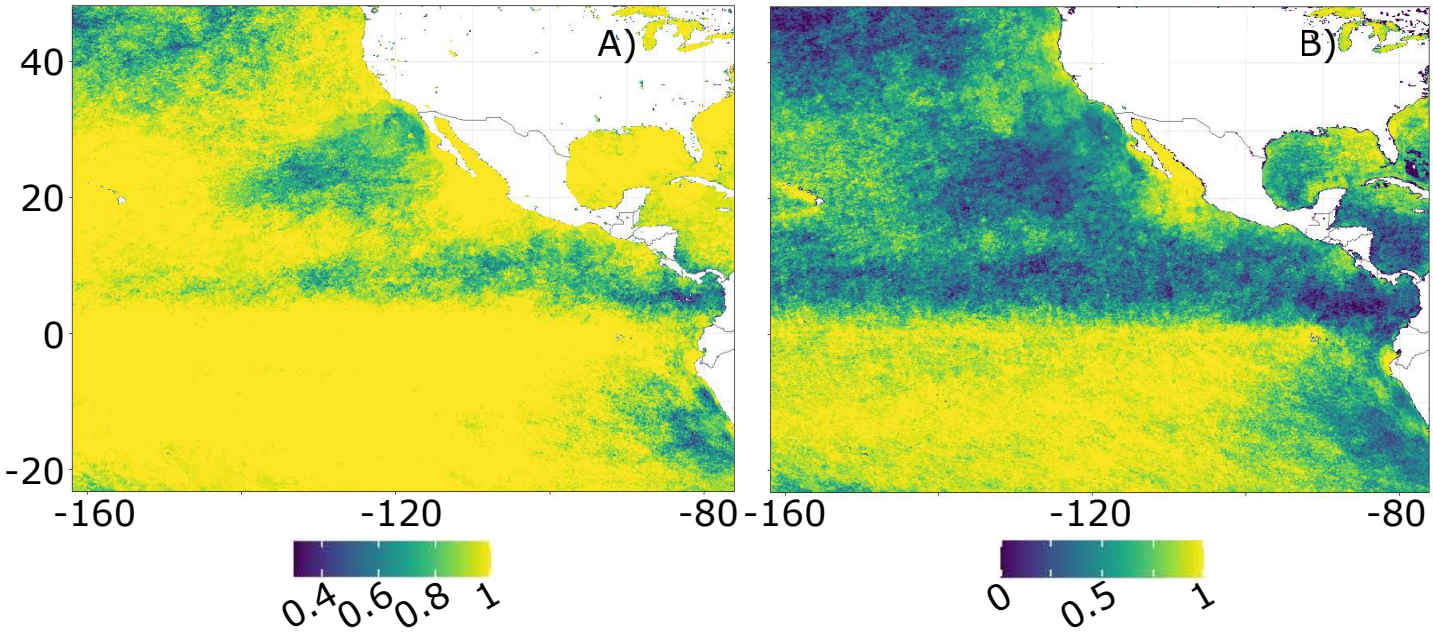

Figure S1: Fraction of the 80-day period with data for SST (A) and chlorophyll-a concentration (B) after 8-day averaging. As FTLE is a model output, the data coverage is perfect and thus not plotted here.

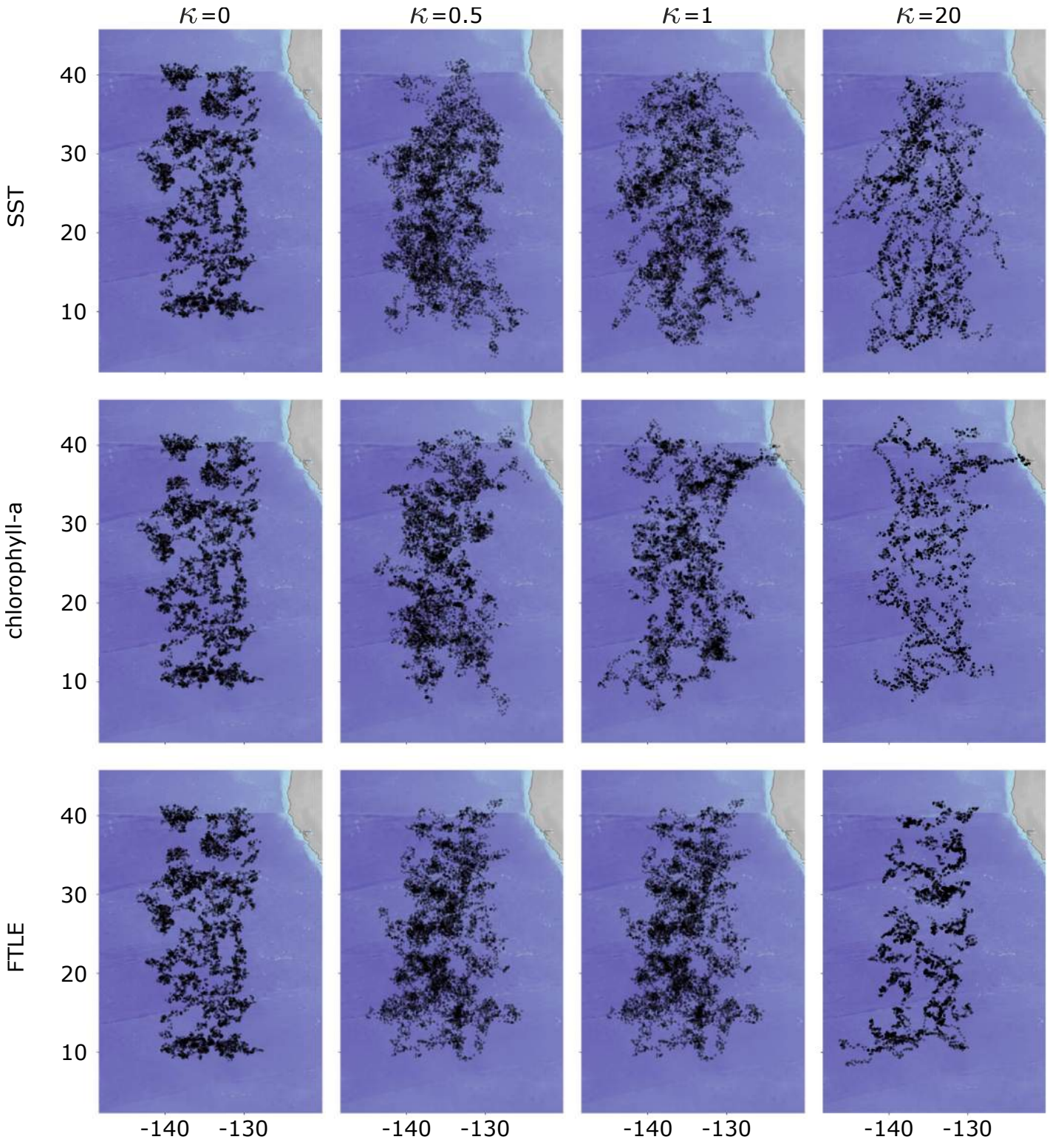

Figure S2: Synthetic tracks of animals selecting for different environmental variables (rows) and with different selection strengths (columns). There are 100 different 80-day tracks per panel. As the case $\kappa = 0$ corresponds to no environmental selection, the same set of tracks was used for all three environmental variables.

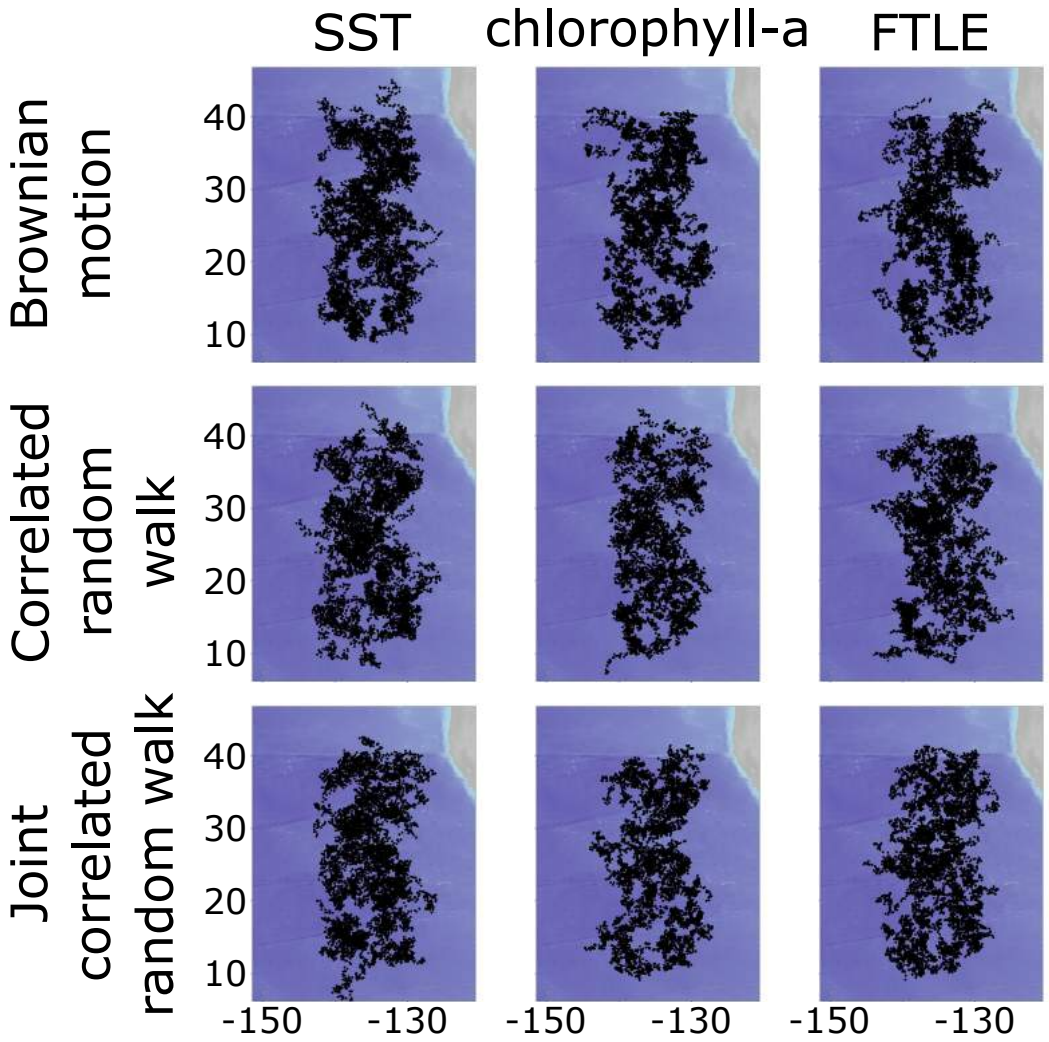

Figure S3: Pseudo-absence tracks for all three null models, based on the synthetic tracks for each of the three different environmental variables selected, with selection strength $\kappa = 1$. Pseudo-absence tracks based on tracks with different selection strengths are qualitatively similar and thus not pictured here.

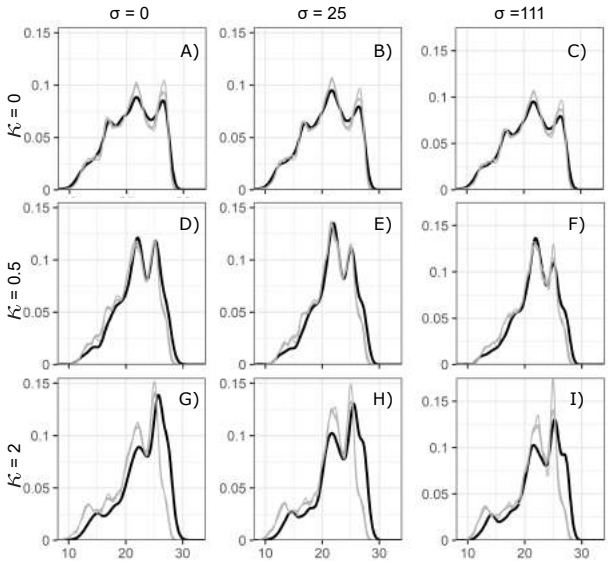

Figure S4: Density distribution of SST (in $°C$) at synthetic presence (black) and pseudo-absence (grey) points, for presence tracks not biased towards high SST (A, B, C), weakly biased towards high SST (D, E, F), and for presence tracks moderately biased towards high SST (G, H, I). The first column (A, D, G) corresponds to locations known with perfect accuracy, the second column (B, E, H) to locations known with moderate accuracy (standard error of 25 km), and the third column (C, F, I) to locations known with weak accuracy (standard error of 111 km). Note the different vertical axes of the third column.

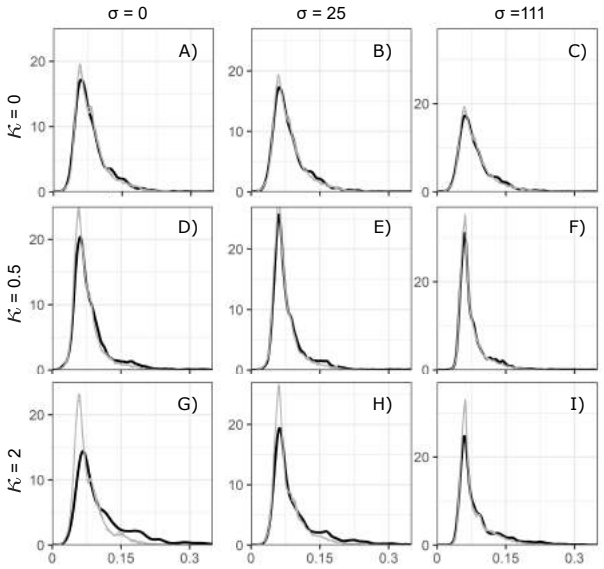

Figure S5: Density distribution of chlorophyll-a concentration (in $mg\ m^{-3}$) at synthetic presence (black) and pseudo-absence (grey) points, for presence tracks not biased towards high chlorophyll-a (A, B, C), weakly biased towards high chlorophyll-a (D, E, F), and for presence tracks moderately biased towards high chlorophyll-a (G, H, I). The first column (A, D, G) corresponds to locations known with perfect accuracy, the second column (B, E, H) to locations known with moderate accuracy (standard error of 25 km), and the third column (C, F, I) to locations known with weak accuracy (standard error of 111 km). Note the different vertical axes of the three columns.

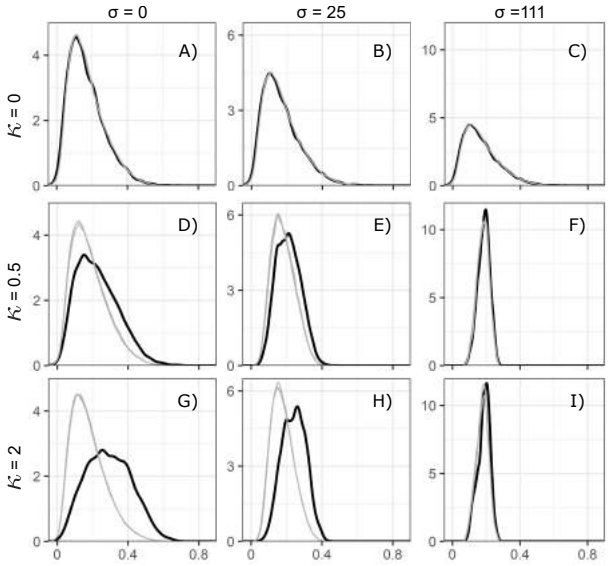

Figure S6: Density distribution of FTLE (in day^{-1}) at synthetic presence (black) and pseudo-absence (grey) points, for presence tracks not biased towards high FTLE (A, B, C), weakly biased towards high FTLE (D, E, F), and for presence tracks moderately biased towards high FTLE (G, H, I). The first column (A, D, G) corresponds to locations known with perfect accuracy, the second column (B, E, H) to locations known with moderate accuracy (standard error of 25 km), and the third column (C, F, I) to locations known with weak accuracy (standard error of 111 km). Note the different vertical axes of the three columns.

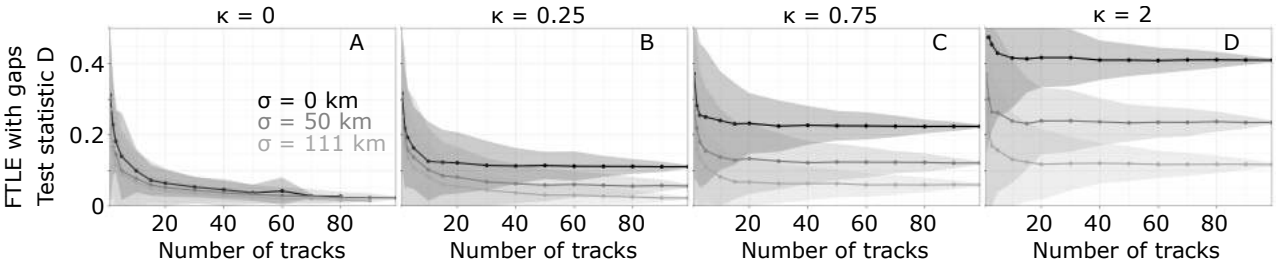

Figure S7: Test statistic D for selection for higher values of FTLE with the same data gaps as chlorophyll-a (A, B, C, D) at different selection strengths (in columns) and for three different accuracies of geolocation estimates (shades of grey). Only test statistics from tests with $p < 0.05$ are included. Shaded areas around each line indicate the bootstrapped 95% confidence intervals.

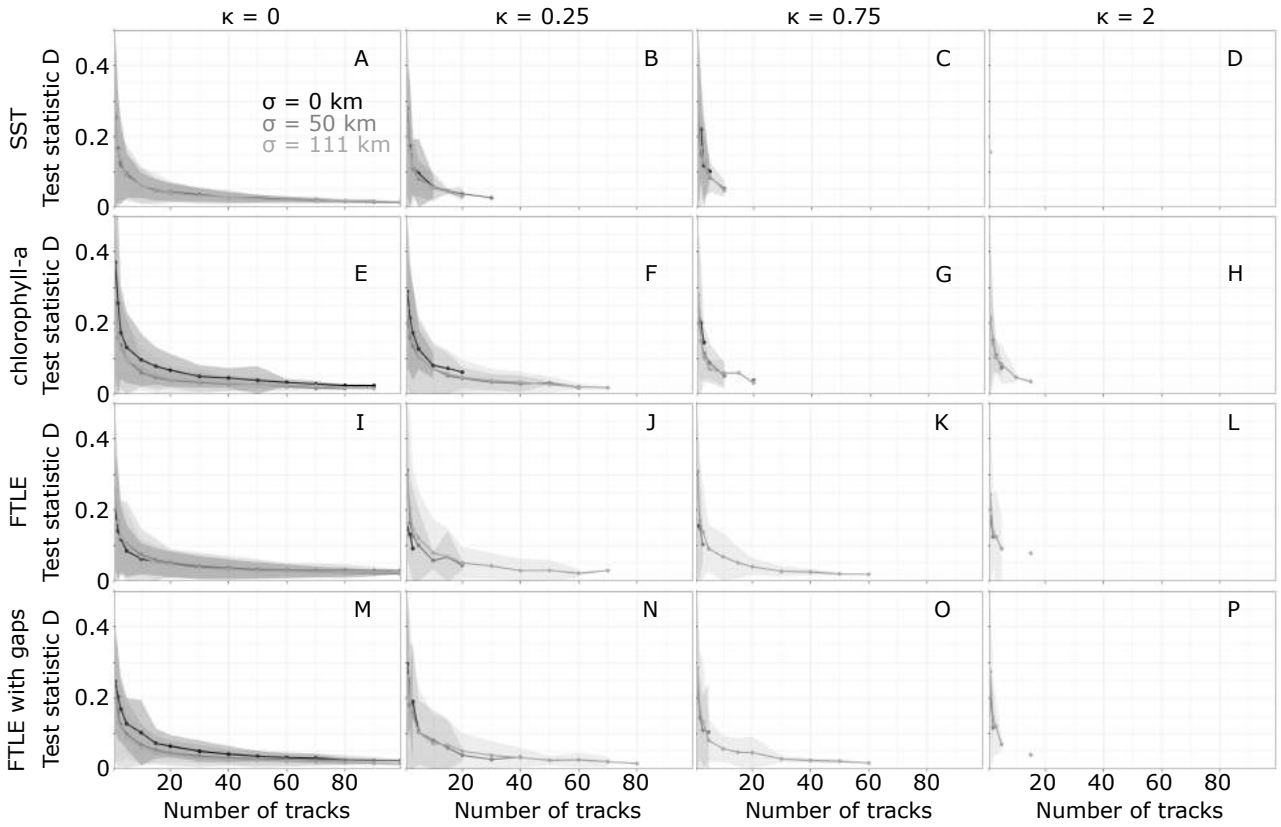

Figure S8: Test statistic D for selection for lower values of SST (A, B, C, D), chlorophyll-a (E, F, G, H), FTLE (I, J, K, L), and FTLE with the same data gaps as chlorophyll-a (M, N, O, P) at different selection strengths (in columns) and for three different accuracies of geolocation estimates (shades of grey). Only test statistics from tests with $p < 0.05$ are included. Shaded areas around each line indicate the bootstrapped 95% confidence intervals.

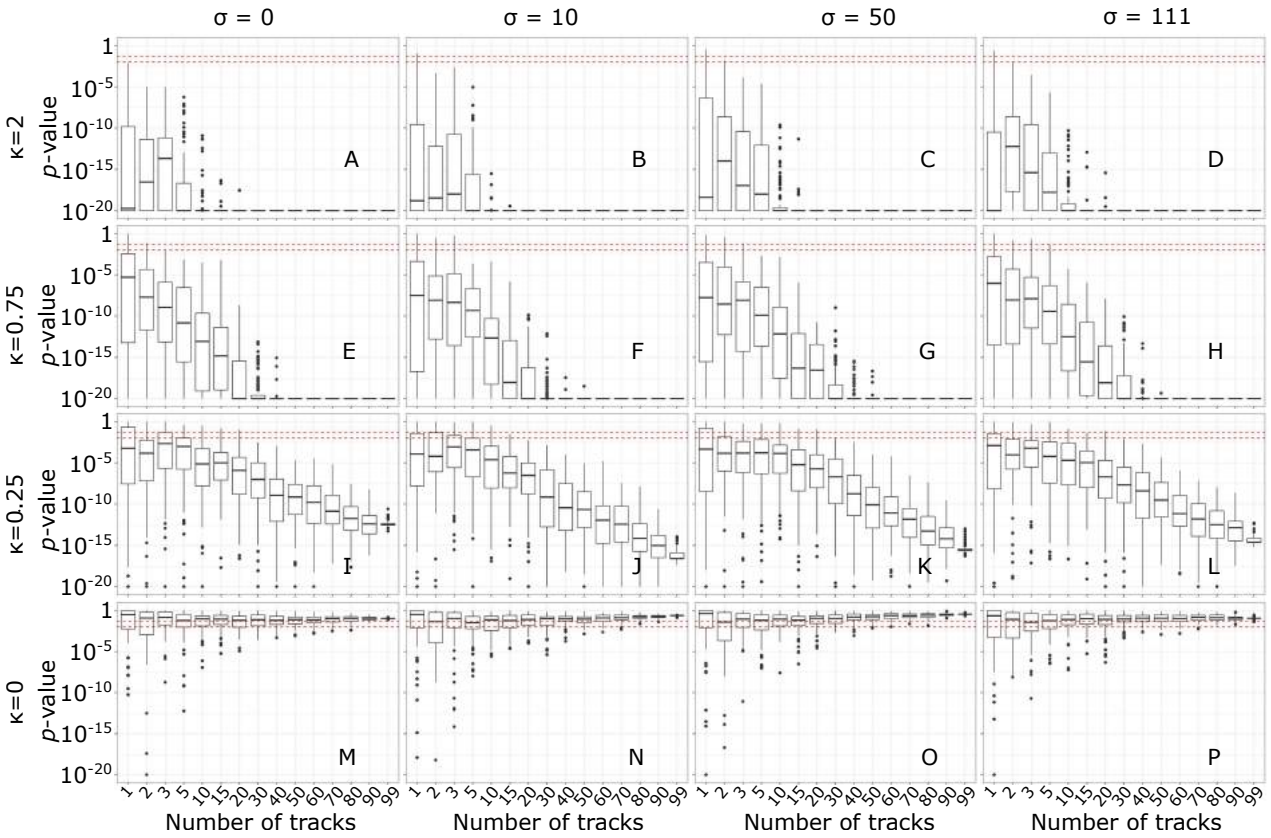

Figure S9: Box plots of the distribution of p-values of tests testing selection for higher values of SST as a function of the number of tracks considered for different values of κ (rows) and σ (columns). The two dotted red lines indicate $p = 0.05$ and $p = 0.01$, and p-values lower than 10^{-20} are plotted as 10^{-20}. Results shown for all null models aggregated only.

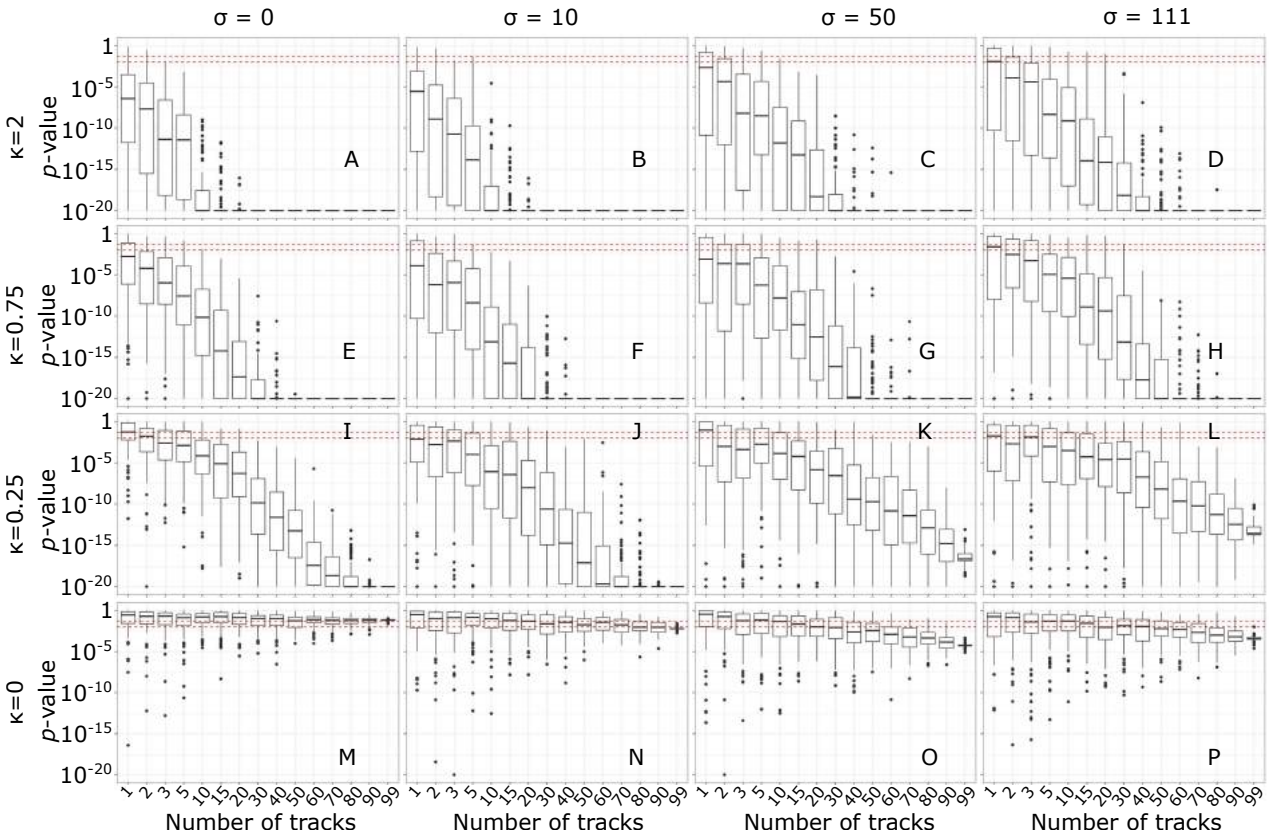

Figure S10: Box plots of the distribution of p-values of tests testing selection for higher values of chlorophyll-a concentration as a function of the number of tracks considered for different values of κ (rows) and σ (columns). The two dotted red lines indicate $p = 0.05$ and $p = 0.01$, and p-values lower than 10^{-20} are plotted as 10^{-20}. Results shown for all null models aggregated only.

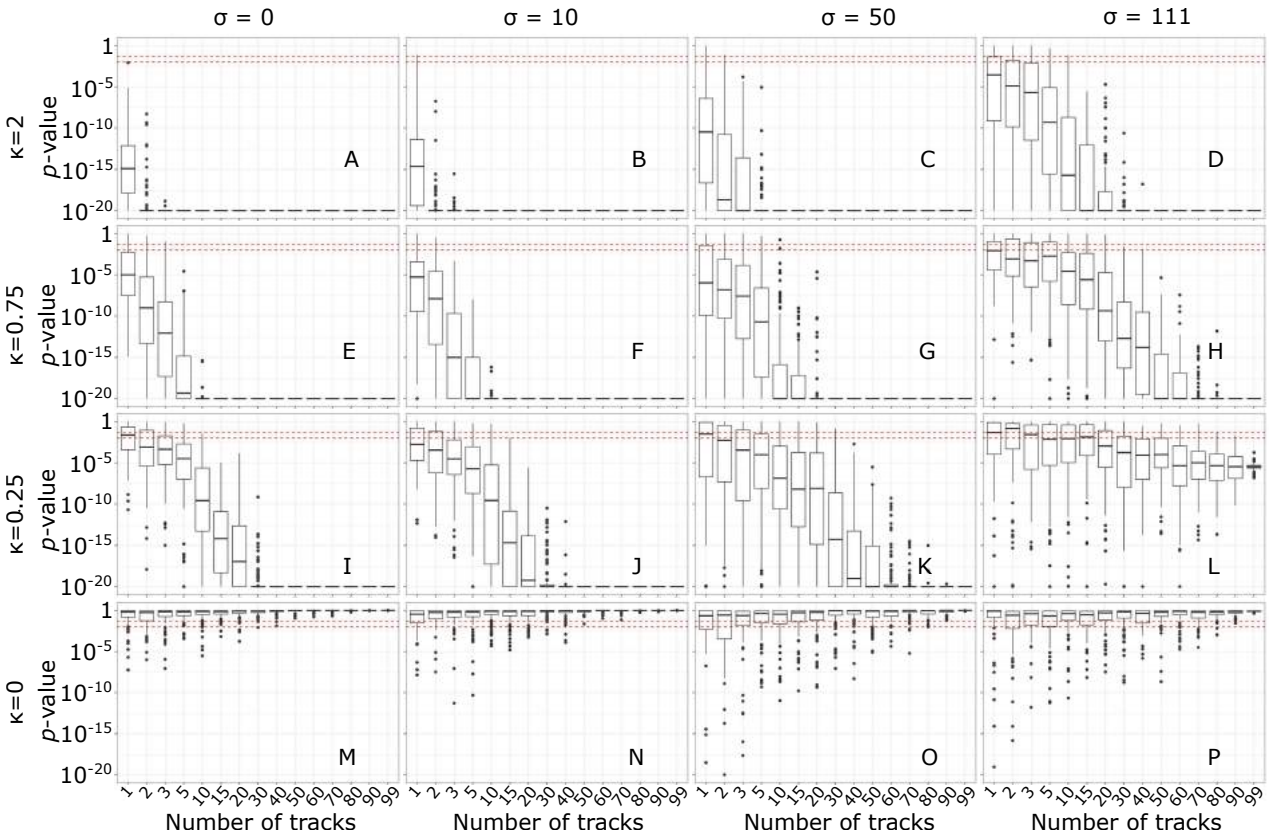

Figure S11: Box plots of the distribution of p-values of tests testing selection for higher values of FTLE as a function of the number of tracks considered for different values of κ (rows) and σ (columns). The two dotted red lines indicate $p = 0.05$ and $p = 0.01$, and p-values lower than 10^{-20} are plotted as 10^{-20}. Results shown for all null models aggregated only.

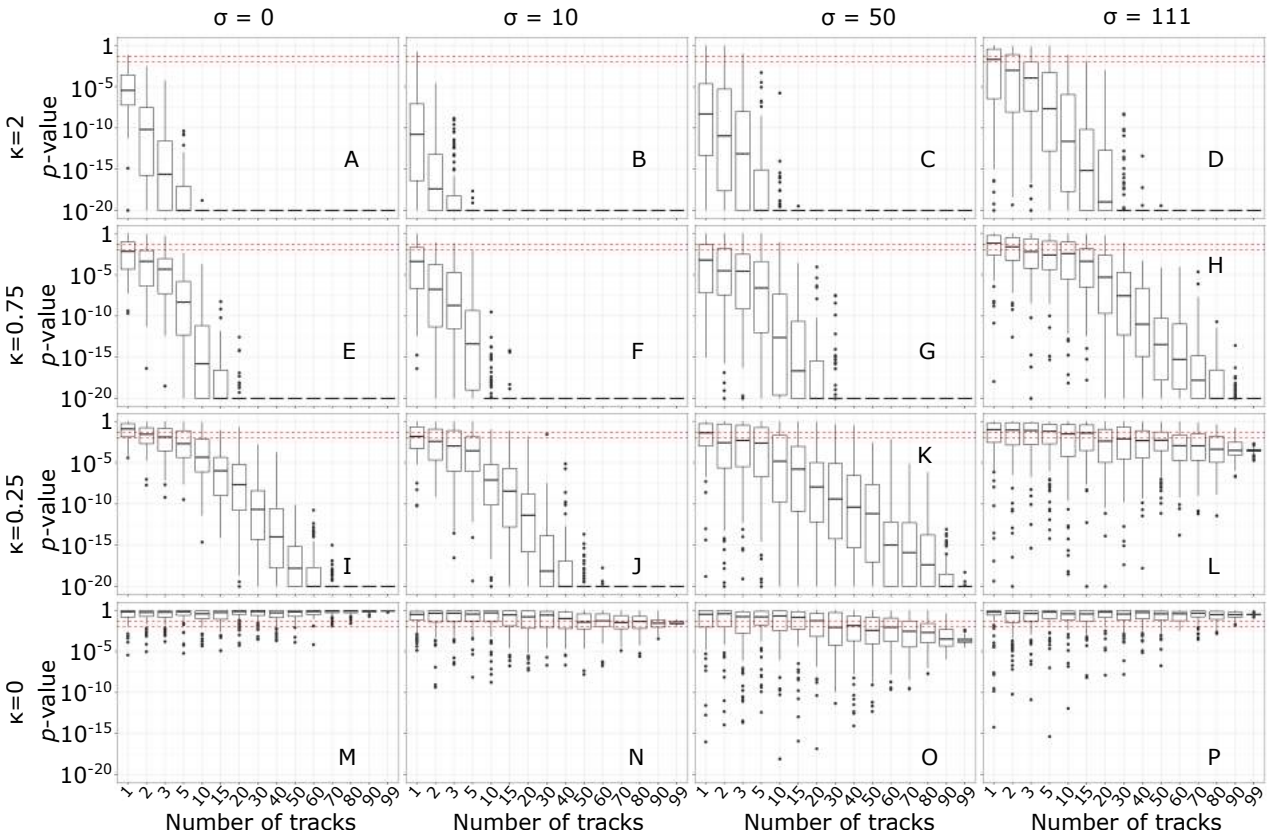

Figure S12: Box plots of the distribution of p-values of tests testing selection for higher values of FTLE (with the same data gaps as chlorophyll-a) as a function of the number of tracks considered for different values of κ (rows) and σ (columns). The two dotted red lines indicate $p = 0.05$ and $p = 0.01$, and p-values lower than 10^{-20} are plotted as 10^{-20}. Results shown for all null models aggregated only.

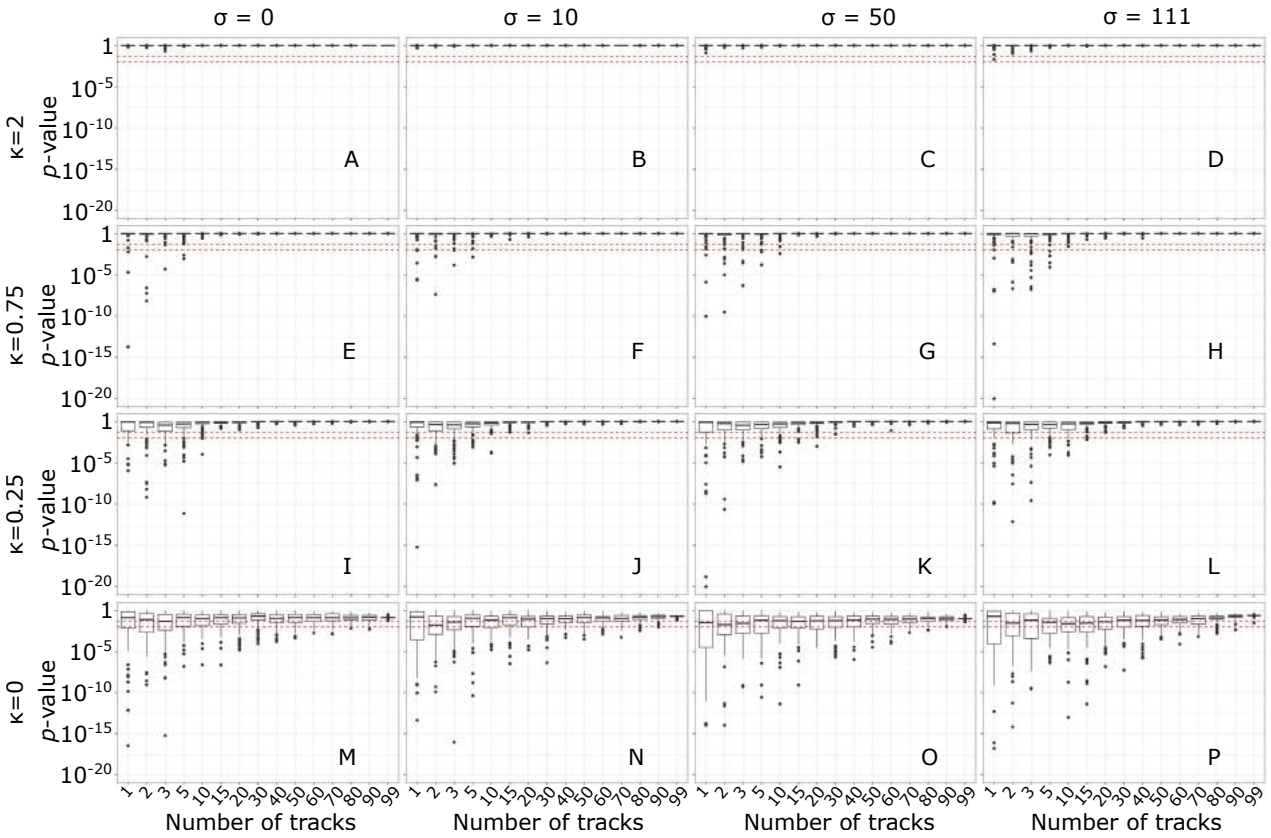

Figure S13: Box plots of the distribution of p-values of tests testing selection for lower values of SST as a function of the number of tracks considered for different values of κ (rows) and σ (columns). The two dotted red lines indicate $p = 0.05$ and $p = 0.01$, and p-values lower than 10^{-20} are plotted as 10^{-20}. Results shown for all null models aggregated only.

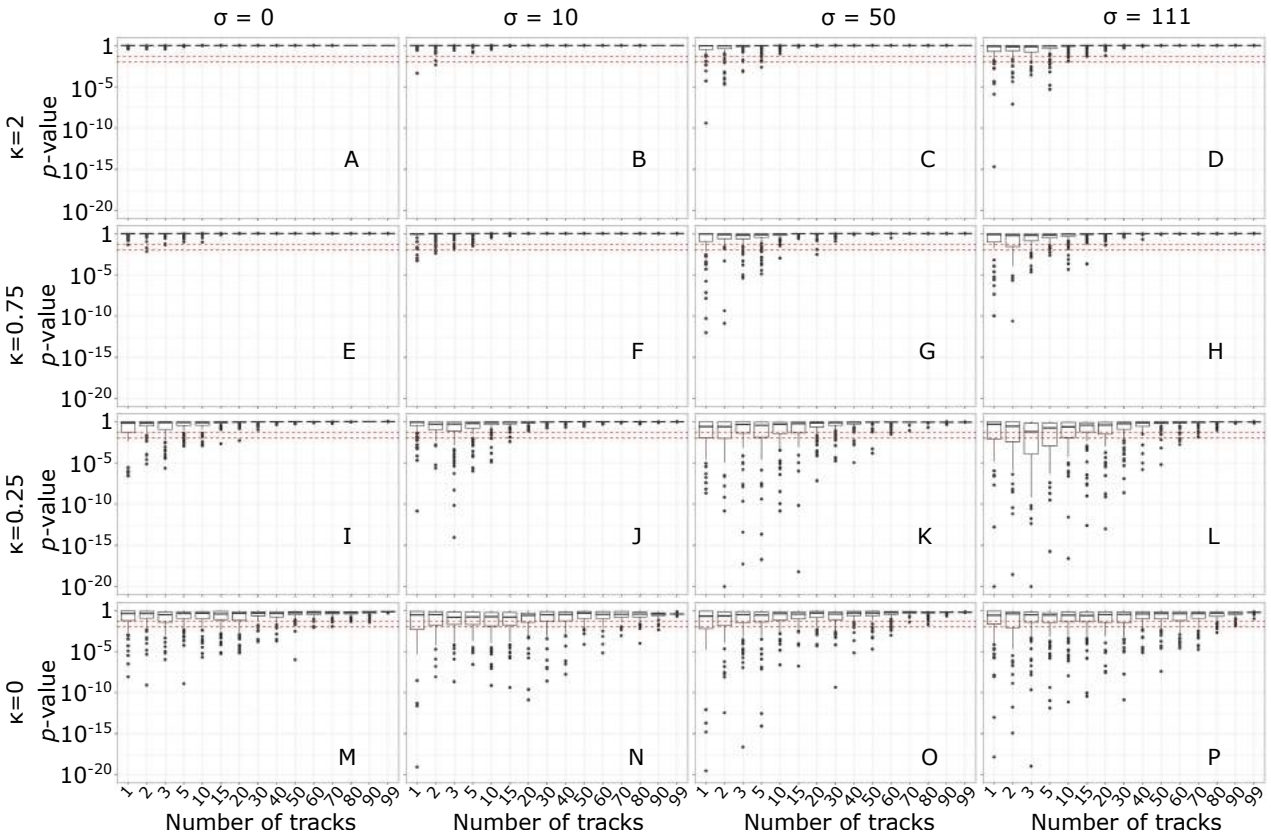

Figure S14: Box plots of the distribution of p-values of tests testing selection for lower values of chlorophyll-a concentration as a function of the number of tracks considered for different values of κ (rows) and σ (columns). The two dotted red lines indicate $p = 0.05$ and $p = 0.01$, and p-values lower than 10^{-20} are plotted as 10^{-20}. Results shown for all null models aggregated only.

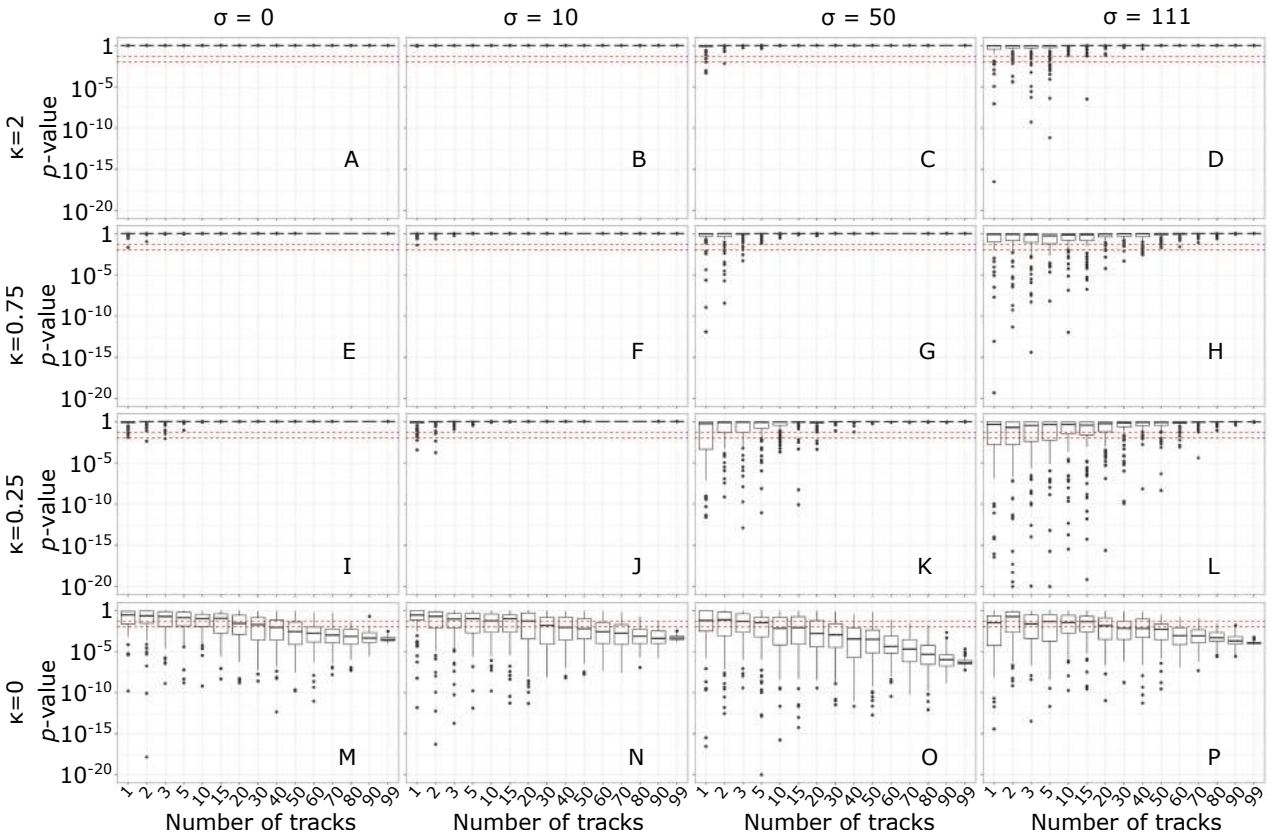

Figure S15: Box plots of the distribution of p-values of tests testing selection for lower values of FTLE as a function of the number of tracks considered for different values of κ (rows) and σ (columns). The two dotted red lines indicate $p = 0.05$ and $p = 0.01$, and p-values lower than 10^{-20} are plotted as 10^{-20}. Results shown for all null models aggregated only.

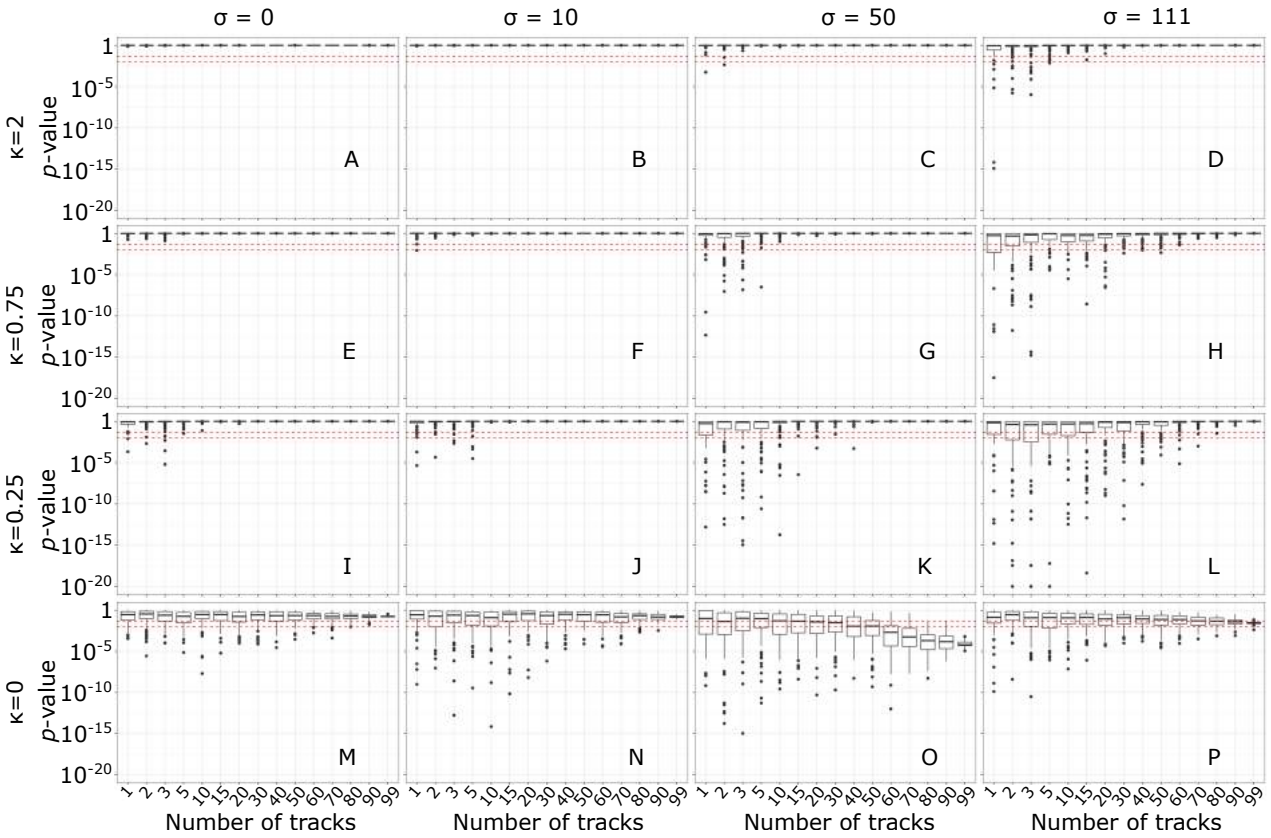

Figure S16: Box plots of the distribution of p-values of tests testing selection for lower values of FTLE (with the same data gaps as chlorophyll-a) as a function of the number of tracks considered for different values of κ (rows) and σ (columns). The two dotted red lines indicate $p = 0.05$ and $p = 0.01$, and p-values lower than 10^{-20} are plotted as 10^{-20}. Results shown for all null models aggregated only.

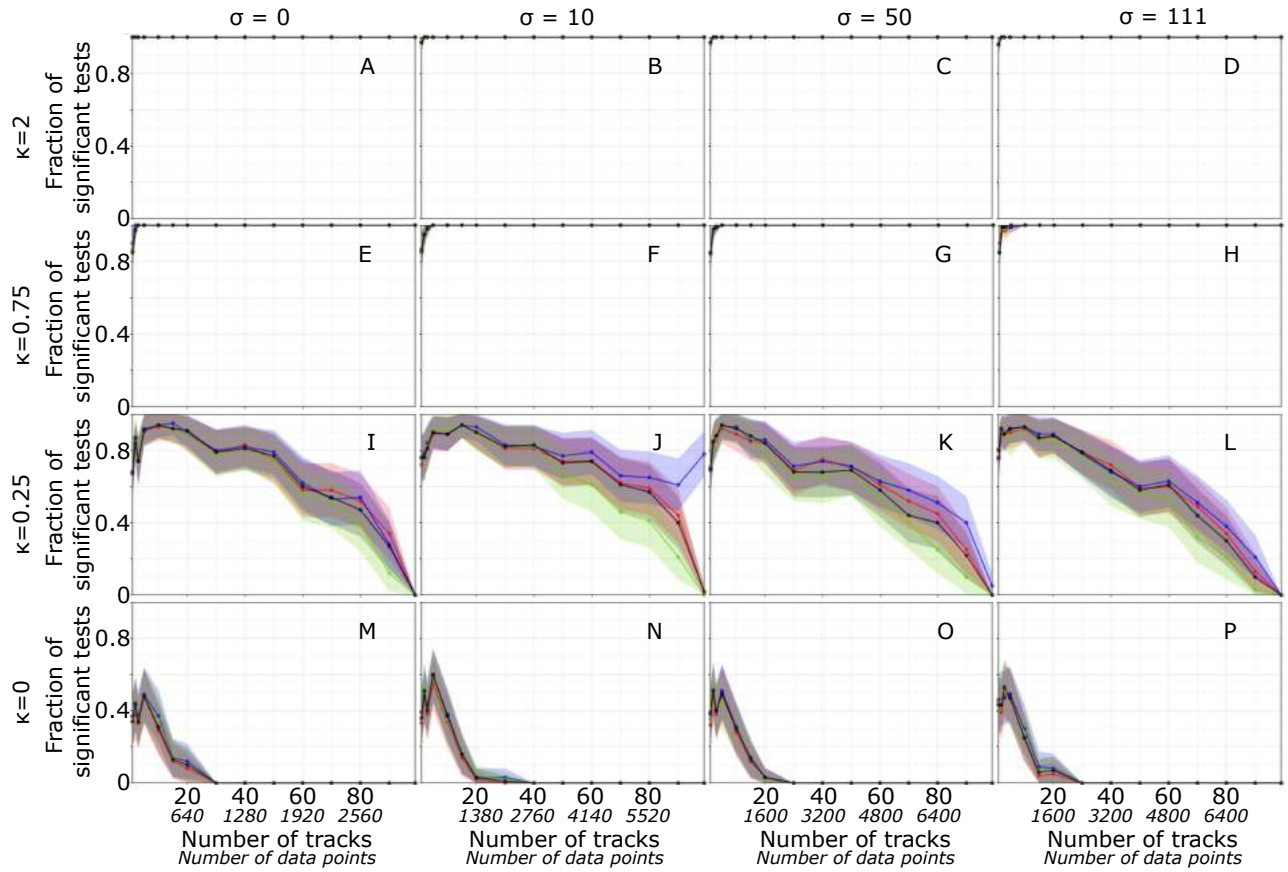

Figure S17: Fraction of test results showing effective selection for higher values of SST as a function of the number of tracks considered for different values of κ (rows) and σ (columns). Shaded areas around each line indicate the bootstrapped 95% confidence intervals. Black lines show results for all null models aggregated, red lines results from Brownian walks only, Blue lines from Correlated random walks only, and green lines from Joint correlated random walks only.

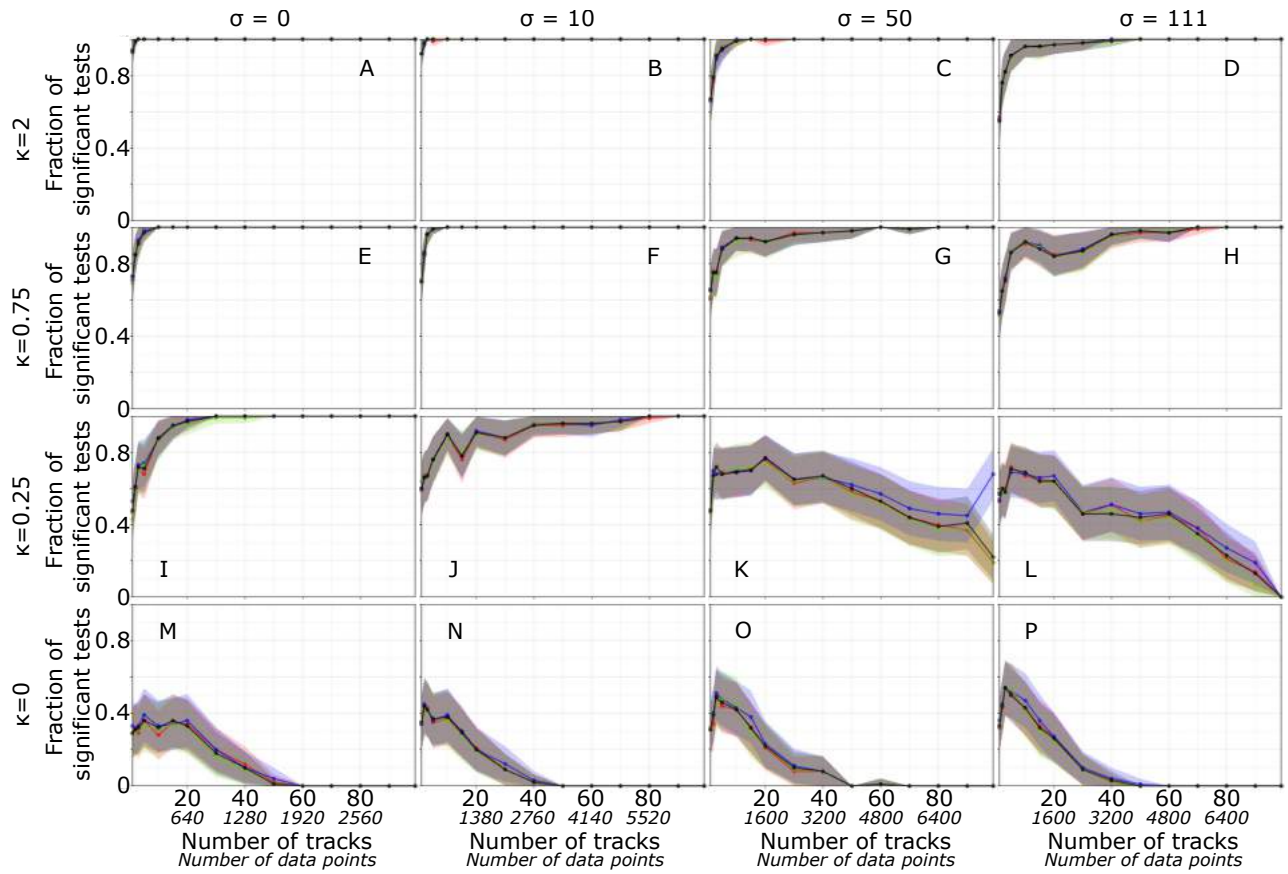

Figure S18: Fraction of test results showing effective selection for higher values of chlorophyll-a concentration as a function of the number of tracks considered for different values of κ (rows) and σ (columns). Shaded areas around each line indicate the bootstrapped 95% confidence intervals. Black lines show results for all null models aggregated, red lines results from Brownian walks only, Blue lines from Correlated random walks only, and green lines from Joint correlated random walks only.

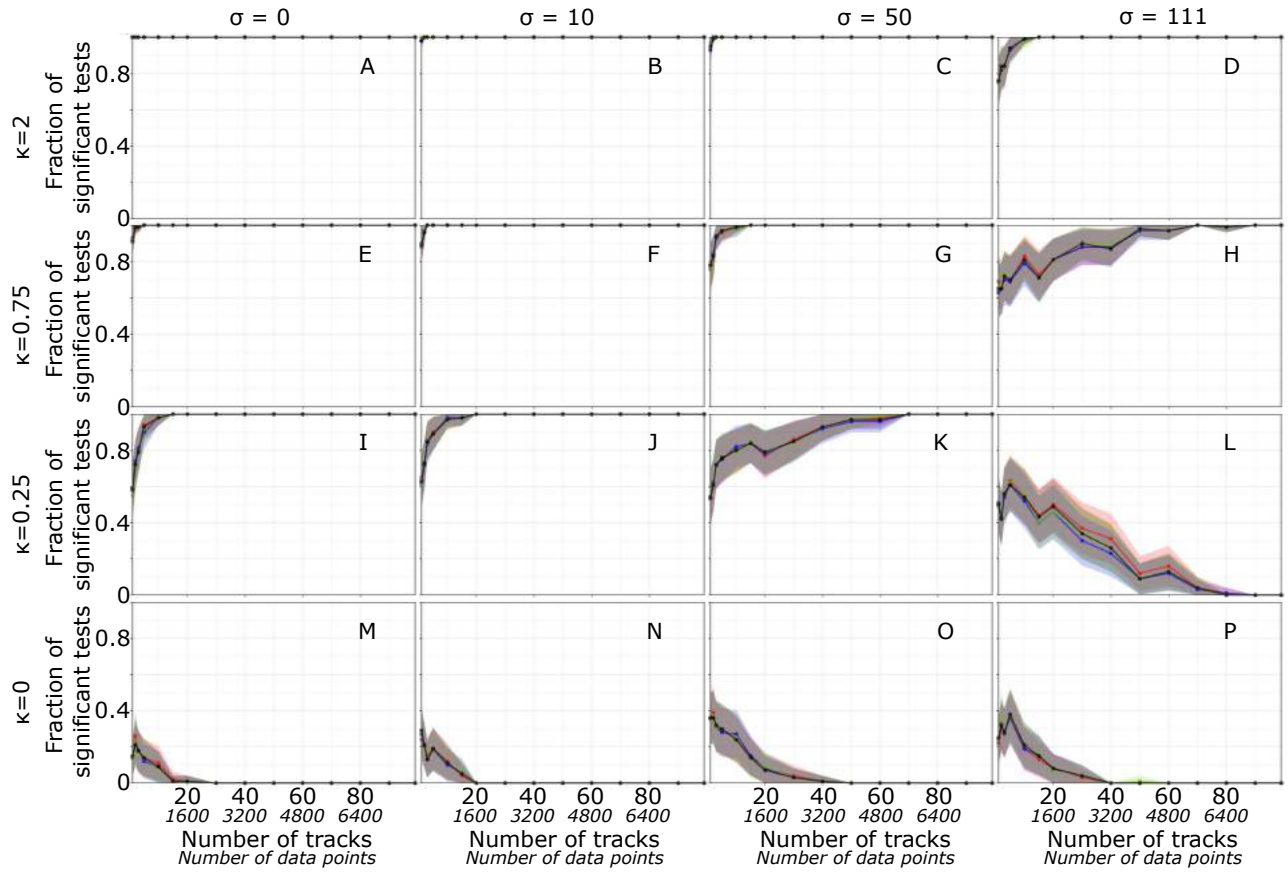

Figure S19: Fraction of test results showing effective selection for higher values of FTLE as a function of the number of tracks considered for different values of κ (rows) and σ (columns). Shaded areas around each line indicate the bootstrapped 95% confidence intervals. Black lines show results for all null models aggregated, red lines results from Brownian walks only, Blue lines from Correlated random walks only, and green lines from Joint correlated random walks only.

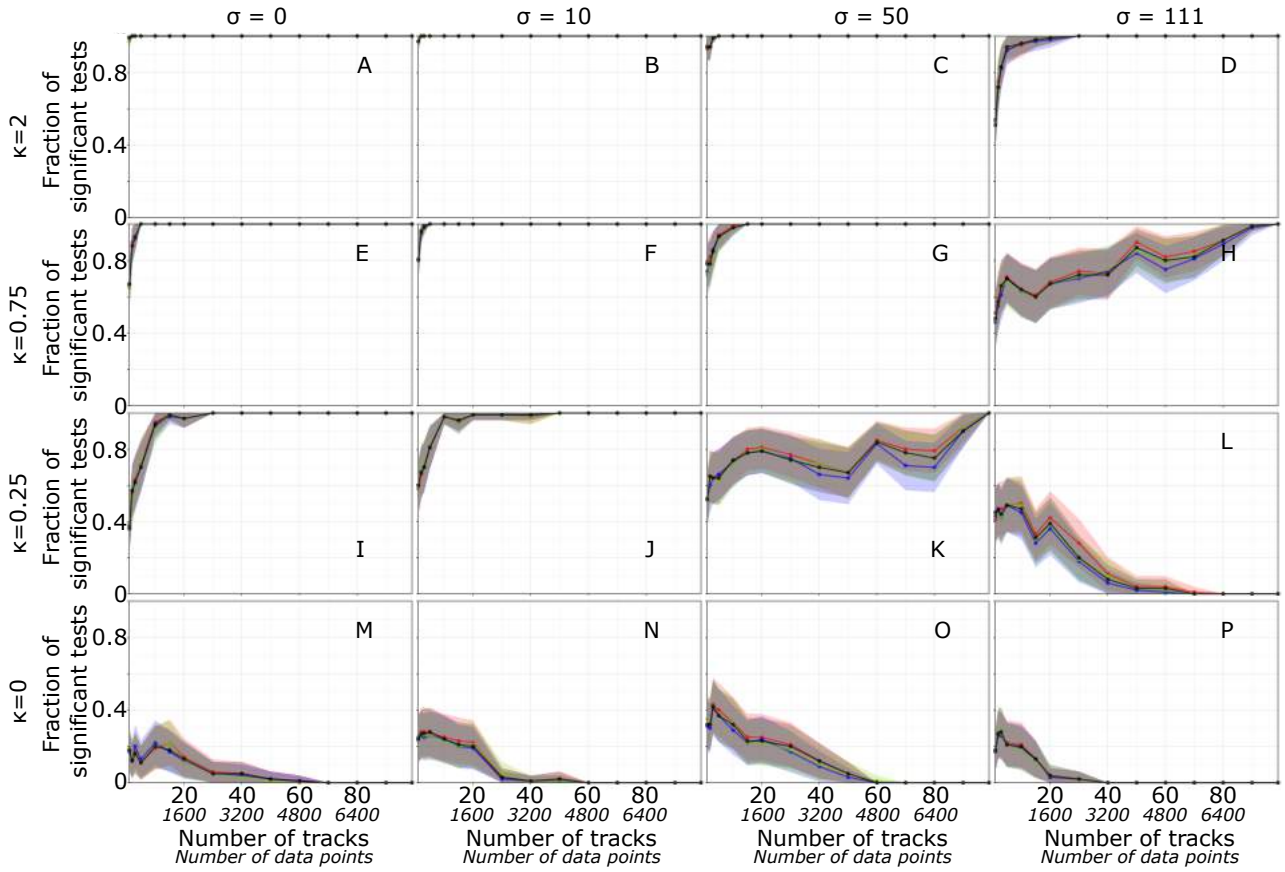

Figure S20: Fraction of test results showing effective selection for higher values of FTLE (with the same data gaps as chlorophyll-a) as a function of the number of tracks considered for different values of κ (rows) and σ (columns). Shaded areas around each line indicate the bootstrapped 95% confidence intervals. Black lines show results for all null models aggregated, red lines results from Brownian walks only, Blue lines from Correlated random walks only, and green lines from Joint correlated random walks only.

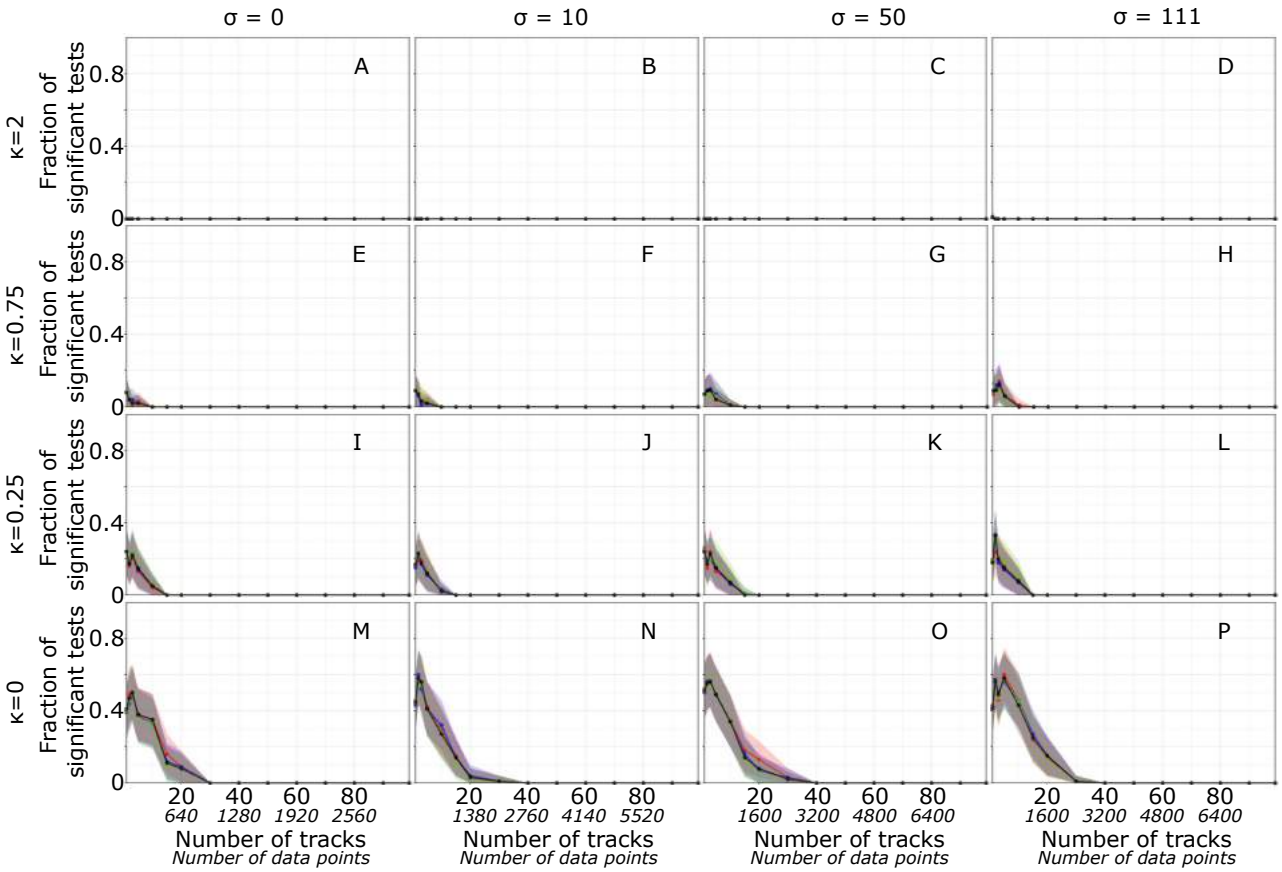

Figure S21: Fraction of test results showing effective selection for lower values of SST as a function of the number of tracks considered for different values of κ (rows) and σ (columns). Shaded areas around each line indicate the bootstrapped 95% confidence intervals. Black lines show results for all null models aggregated, red lines results from Brownian walks only, Blue lines from Correlated random walks only, and green lines from Joint correlated random walks only.

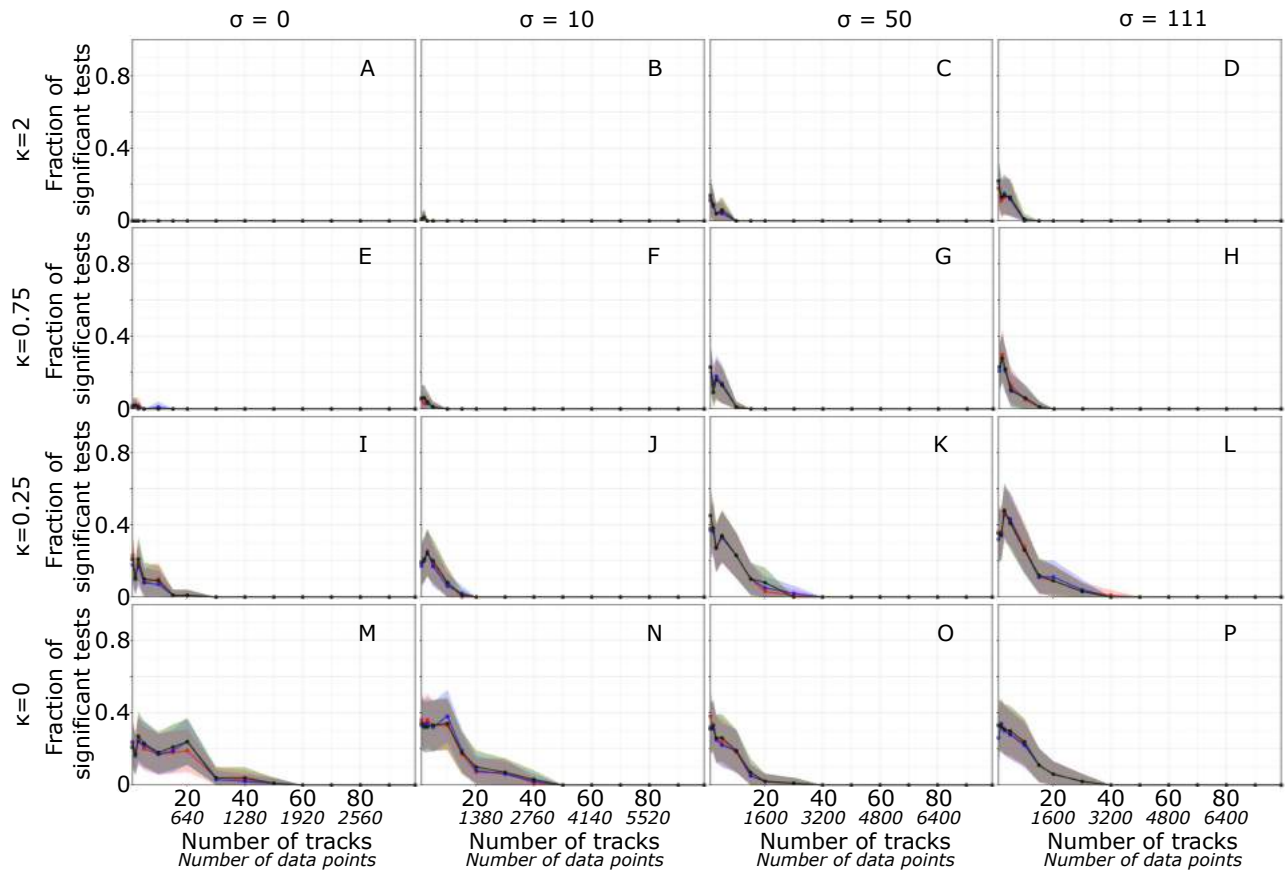

Figure S22: Fraction of test results showing effective selection for lower values of chlorophyll-a concentration as a function of the number of tracks considered for different values of κ (rows) and σ (columns). Shaded areas around each line indicate the bootstrapped 95% confidence intervals. Black lines show results for all null models aggregated, red lines results from Brownian walks only, Blue lines from Correlated random walks only, and green lines from Joint correlated random walks only.

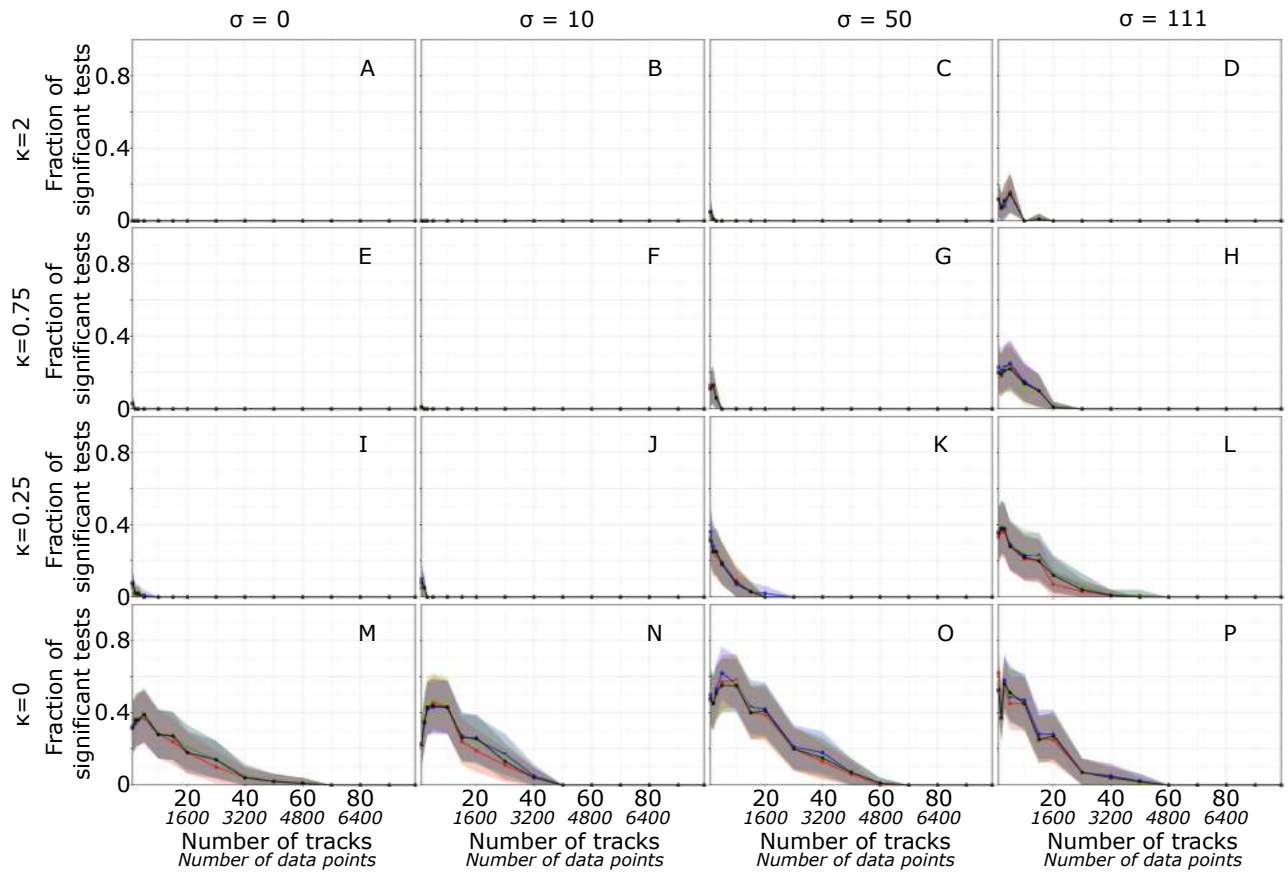

Figure S23: Fraction of test results showing effective selection for lower values of FTLE as a function of the number of tracks considered for different values of κ (rows) and σ (columns). Shaded areas around each line indicate the bootstrapped 95% confidence intervals. Black lines show results for all null models aggregated, red lines results from Brownian walks only, Blue lines from Correlated random walks only, and green lines from Joint correlated random walks only.

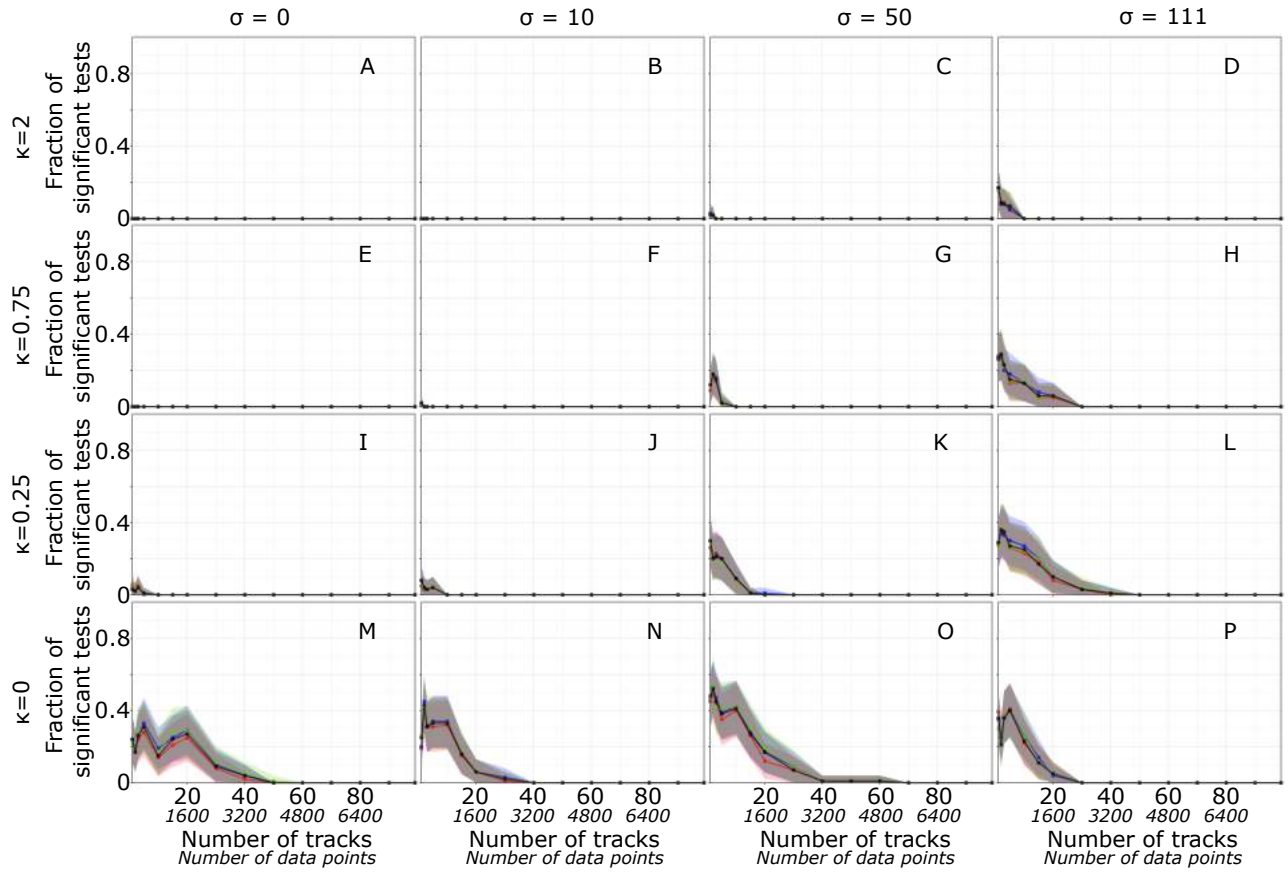

Figure S24: Fraction of test results showing effective selection for lower values of FTLE (with the same data gaps as chlorophyll-a) as a function of the number of tracks considered for different values of κ (rows) and σ (columns). Shaded areas around each line indicate the bootstrapped 95% confidence intervals. Black lines show results for all null models aggregated, red lines results from Brownian walks only, Blue lines from Correlated random walks only, and green lines from Joint correlated random walks only.

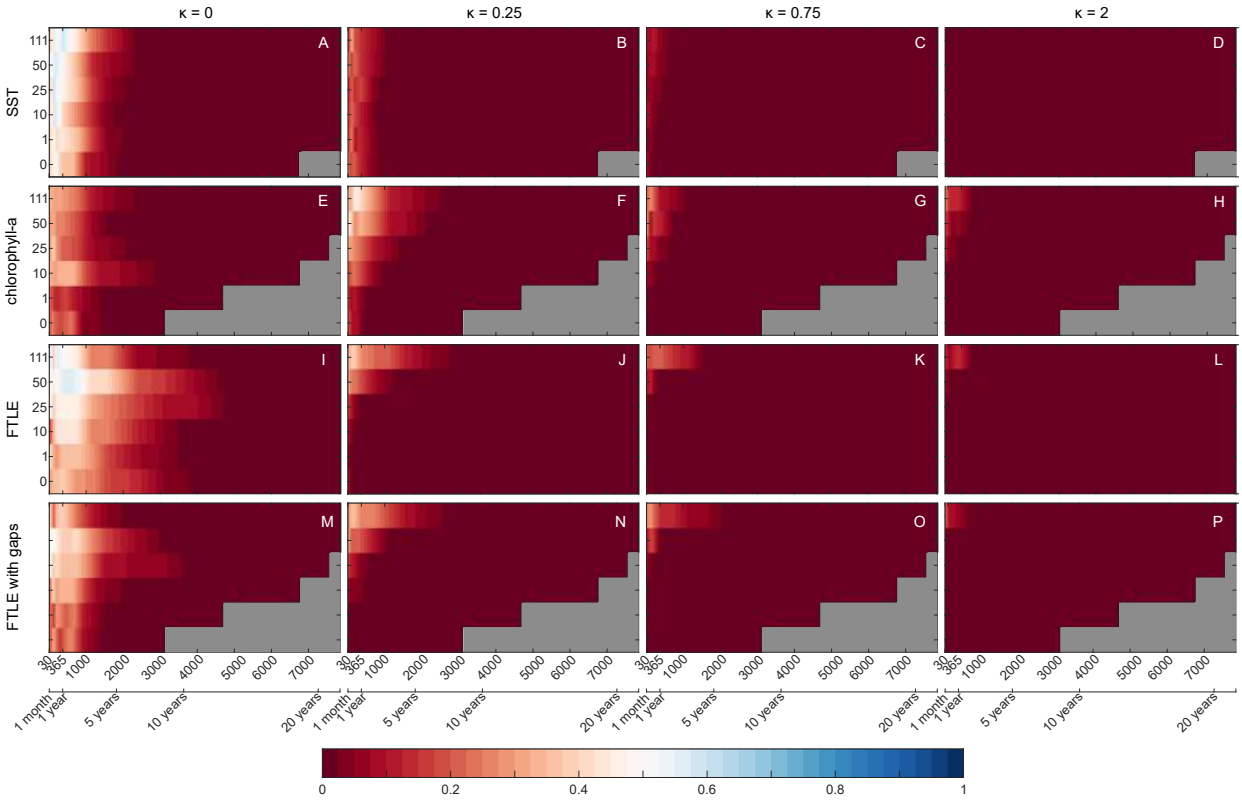

Figure S25: Fraction of test showing effective selection for lower values of SST (A, B, C, D), chlorophyll-a (E, F, G, H), FTLE (I, J, K, L), and FTLE with the same data gaps as chlorophyll-a (M, N, O, P) at different selection strengths (in columns), geolocation accuracy (rows of the panels) and sample sizes (columns of the panels). Sample sizes are plotted as sample size with environmental data (in both number of data points and its corresponding time span), hence the grey patches in panels A-H and M-P.

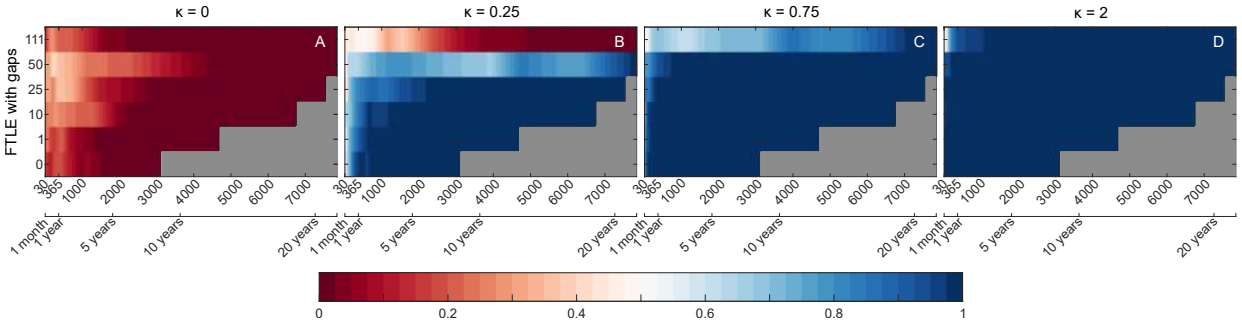

Figure S26: Fraction of test showing effective selection for higher values of FTLE with the same data gaps as chlorophyll-a (A, B, C, D) at different selection strengths (in columns), geolocation accuracy (rows of the panels) and sample sizes (columns of the panels). Sample sizes are plotted as sample size with environmental data (in both number of data points and its corresponding time span), hence the grey patches.