



Applying a Multiverse to Habitat Analyses

Benjamin Michael Marshall*¹ and Alexander Bradley Duthie**¹

¹ Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, UK

*benjaminmichaelmarshall@gmail.com

**alexander.duthie@stir.ac.uk

Abstract

Researchers are intrinsically part of the research process. While we may strive for objectivity, there are always judgement calls required during research. When you ask 10 researchers to answer the same question with the same dataset, you will receive 10 different answers. This variation in answers has been linked to several disciplines' replication crises. Here, we explore whether answers from movement ecology, specifically habitat preference, studies vary as a result of differing analytical choices. We conducted a multiverse analysis to explore over 15,000 ways of determining habitat preference from animal tracking data. By using simulated virtual animals with a known preference, we were able to show which decisions during analysis could lead to more variable estimates of habitat preference. The multiverse revealed that data quantity (i.e., tracking frequency and duration) was more important to obtaining reliable answers than any analysis choice. Overall, the pattern of estimates shows the majority of analysis pathways provide similar final results, particularly for modern analysis methods. The pattern reflects findings from other disciplines, indicating that while movement ecology is not immune to issues of non-replicability stemming from researcher choice, it is also not at any greater risk than other disciplines.

Keywords

Movement ecology, simulation

1 Introduction

Researchers are intrinsically part of the research process (Levins & Lewontin, 1985; Tang-Martínez, 2020), and our expectations can shape the answers we find (Holman et al., 2015). While we can strive to conduct research objectively, there are frequent moments during research that require judgement calls (Steege et al., 2016). Such calls, choices, or decisions occur throughout the research process, encompassing everything from study design (e.g., sample size, sampling intensity, sample stratification) to analysis (e.g., Bayesian or frequentist, ex/inclusion of outliers). We draw on our own experience and the input of peers to try and ensure the best choices are made to produce robust and reliable results, but we also contend with data, expertise, and interpretability constraints (Liu, Althoff & Heer, 2020).

Researchers are also influenced by the incentive system around them (Anderson et al., 2007). We cannot undertake research in a vacuum; we often require institutions, funders, and scientific journals to produce and share research. These bodies can influence what research is conducted; they are in a position to incentivise or disincentivise research of certain topics or methodologies (Fanelli, 2010a; Ware & Munafò, 2015; Smaldino & McElreath, 2016). The use of impact factor (and other similar metrics) is an example of citations being used as a measure of quality or research worth. But when examined closer, the impact factor appears detached from the robustness or reliability of the research (Brembs, 2018). This decrease in robustness can be seen in the increases in effect sizes inflation, p-value misreporting, among a number of other measures of quality (Brembs, 2018). Similarly, novelty has been fetishised by journals to the detriment of replications studies (Vinkers, Tjldink & Otte, 2015; Forstmeier, Wagenmakers & Parker, 2017; Brembs, 2019), despite widespread agreement on the importance of replications (Fraser et al., 2020). There is a bias towards positive, statistically significant results (Jennions & Møller, 2002; Cassey et al., 2004). Due to the nature of frequentist statistical significance, a prioritisation of significant results can elevate underpowered studies and boost false positive rates (Forstmeier & Schielzeth, 2011; Albers, 2019).

Unfortunately, there is evidence that the system of incentives trickle down to impact the judgement calls and decisions of researchers while undertaking research and when publishing those results. The more detrimental of these decisions have been termed questionable research practices (Fraser et al., 2018; Bishop, 2019). They chiefly come in three forms: HARKing, cherry-picking, and p-hacking. Hypothesising After Results Known (HARKing) is where the research can present the results as a confirmatory result, despite originally there being no or contrary hypothesis. HARKing can sometimes be further enabled and rationalised by hindsight bias, where unexpected results are perceived as more likely once they have been observed (Gelman & Loken, 2013; Forstmeier, Wagenmakers & Parker, 2017). Cherry-picking is the removal or non-reporting of data points or (co-)variables, that did not yield significant results. P-hacking is the repeated use of statistical tests, with different settings, to achieve statistical significance. Arguably, the existence of p-hacking is enabled by an over-reliance on p-value thresholds, rather than flexible p-value thresholds that are predefined based on effect size of interest, sample size, and desired accuracy of estimation (Lakens et al., 2018). Questionable research practices can be viewed as methods to achieve a neat, statistically significant, and publishable narrative (O'Boyle, Banks & Gonzalez-Mulé, 2014); in the worse cases, narratives could be prioritised over transparently reported results.

There is a fear that questionable research practices, and the broader incentives they are connected to, are responsible for the replicability crisis (often referred to as the “reproducibility crisis”). Across many disciplines, there are examples of replication studies being unable to replicate prior research (Freedman, Cockburn & Simcoe, 2015; Open Science Collaboration, 2015; Kelly, 2019). Often these replication efforts are conducted with larger sample sizes, or rely on the consolidation of many independent studies (often in the form of meta-analyses). The implication is not that the original studies were necessarily flawed; but – in the absence of questionable research practices – sufficient variation exists in the study subjects to obscure a consistent effect [i.e., variation beyond variation stemming from sampling; Simonsohn (2015)].

However, variation can also stem from analysis flexibility (i.e., the presence of many ways to analyse the same data to answer the same question). This flexibility helps enable questionable research practices (Fraser et al., 2018), and is potentially steered by publication bias (Jennions & Møller, 2002; Cassey et al., 2004) if results that produce more publishable results are prioritised/rewarded over less exciting but robust results. Given the prominence of questionable research practices and publication bias, the inconsistencies between initial and replication studies warrant investigation (especially when analysis flexibility is also implicated in potentially flawed replications Bryan, Yeager & O'Brien, 2019). It is key to note that analysis flexibility can still lead to variable results in the absence of any undesirable incentives simply as the result of researchers considering different approaches of differing validities for a given dataset (Gelman & Loken, 2013).

Scientific progress requires building upon past results, and therefore requires confidence in past results. Issues arise when subsequent research is based upon weak foundations –i.e., studies with a limited capacity to be replicated because of questionable practices or over-generalisation. Early significant results can dictate the direction of research and grow resistant to later contradictory results (Barto & Rillig, 2012); therefore, early diagnoses of overly confident results or previously unknown sources of variation becomes a priority.

In medical fields, a lack of replicability comes with direct monetary and well-being costs (Freedman, Cockburn & Simcoe, 2015). Like the medical field, ecological studies can come with well-being costs to the study subjects [e.g., direct surgery/markings of the animal (Reinert & Cundall, 1982; Winne et al., 2006)], as well as impacts on stakeholders stemming from management decisions. There are fears that the lack of replicability will feed distrust of science more generally (Anvari & Lakens, 2018). Therefore, maximising replicability in ecology is key to minimising research waste (Grainger et al., 2019) and the negative impacts on systems and subjects studied.

The impacts on the study subjects, paired with the often high monetary costs of ecological studies (particularly bio-logging where animals may undergo surgery, Weaver, Westphal & Taylor, 2021) means that replications can be more difficult to justify. When paired with fact that ecological systems are complex and in constant flux –often frustrating perfect replications due to changes in space and time (Nakagawa & Parker, 2015; Schnitzer & Carson, 2016) –we are left with a distinct lack of direct replications in ecology (Kelly, 2019).

The low prevalence of replications in ecology make it difficult to assess the overall irreproducibility situation in ecology (Kelly, 2019); but there are several examples that suggest irreproducibility is something ecologists should be wary of (Wang et al., 2018; Sánchez-Tójar et al., 2018; Roche et al., 2020; Clark et al., 2020). The potential for irreproducibility is further supported by evidence of positive publication bias (Fanelli, 2010b, 2012), and links between smaller sample sizes and inflated effect sizes (Lemoine et al., 2016).

In the absence of direct replications, ecology is often left to assess replicability via conceptual replications (Fraser et al., 2020) or efforts broadly referred to as quasi-replications (Palmer, 2000). Replications range in intensity. Direct (or exact) replications are attempts to replicate a tightly defined concept/hypothesis while duplicating of all characteristics of the original study. Partial replications are a step looser, where the concept/hypothesis tested is less clearly defined (e.g., applicable to a broader scale) but efforts are made to repeat the same methodology. The most general category are conceptual replications, where the subject and method of study varies from the original study, but the replication targets a the same concept/hypothesis (Nakagawa & Parker, 2015; Kelly, 2019). Both partial and conceptual can be classed as quasi-replications if the concept and scale is broadly defined (Nakagawa & Parker, 2015).

Conceptual replications are extremely valuable, but rely on our ability to compare replication efforts to previous findings. An important aspect of those comparisons is accounting for factors differing between the studies that are not salient to the effect of interest (Forstmeier, Wagenmakers & Parker, 2017); e.g., those linked to sampling differences (Simonsohn, 2015). An example of sampling differences leading to differences in final results can be seen in the case of reptile space use. Silva et al. (2020) showed how frequently a reptile was located by a researcher interacted with the space-use estimation method, leading to large differences in area estimates even when using the same estimation method. What is revealing is not only how the choices during analysis (e.g., choice of area estimation method) impacted results, but how the error introduced by those choices changed depending on the sampling. It presents a scenario where the *correct* choice was dependent on preceding decisions when designing the study; therefore, highlighting the need to explore the impacts of multiple decisions simultaneously.

As seen in the reptile space use example, the choices made by the researcher [researcher degrees of freedom; Simmons, Nelson & Simonsohn (2011)] is a key source of variation among studies. It would be advantageous to understand which choices have a significant impact and whether we can account for differences in choice during comparisons. An understanding of choice could better guide decisions during a study and potentially be used to gauge the robustness of a given dataset in answering a given question.

Research degrees of freedom [or flexibility in analysis; Forstmeier, Wagenmakers & Parker (2017)] have been elegantly demonstrated by a number of “many analysts” studies (e.g., Silberzahn et al., 2018; Huntington-Klein et al., 2021). In these studies, a number of researchers, or research groups, are tasked with answering the same question. Naturally each participant takes a slightly different approach, both in how the question is interpreted (Auspurg & Brüderl, 2021), and the analysis approach chosen (Gelman & Loken, 2013; Bastiaansen et al., 2020), resulting in different final results. The variation in final results can be considered originating from six sources of uncertainty/variation (Hoffmann et al., 2021): measurement (randomness from the act of measuring), data preprocessing (decisions on data inclusion/exclusion and transforming), parameter (decisions on which parameters used as covariates/predictors), model (decisions on model structure and specification), method (decisions on method choice and parameterisation), and sampling (randomness as a result of sampling a wider population). Several sources of variation (data preprocessing, parameter, model, and method) are likely to be particularly key to defining researcher degrees of freedom post data collection. In some cases, the cause behind the variability in results is hard to diagnose (Brezna et al., 2021), or will be less likely to be questioned because of the agreement with existing theory (Gelman & Loken, 2013). There are examples where the variation in results is sufficient to change the final conclusions (Salis, Lena & Lengagne, 2021), and others where it alters the strength of an estimated effect (Desbureaux, 2021). The importance of the effect size variation is context specific, i.e., how variation relates to the overall effect size, and can impact results pertaining to real-world scenarios (Desbureaux, 2021).

1.1 Multiverse analysis

A rising approach to address the unknown impacts of undisclosed researcher degrees of freedom is to fully explore all plausible or reasonable analysis choices open to researchers – to explore a multiverse of design choices (Steege *et al.*, 2016). This multiverse analysis – closely linked to vibration of effects (Patel, Burford & Ioannidis, 2015), multi-model analysis (Young & Holsteen, 2017), and specification curve analysis (Simonsohn, Simmons & Nelson, 2020)– has the potential to demonstrate and quantify the variation stemming from researcher’s analyses choices (Rijnhart *et al.*, 2021). Choices can include everything from sample sizes and splits (e.g., Webb & Demeyere, 2021) to measurement and summary statistics (e.g., Parsons, 2020), but crucially should only include options that are reasonable (Simonsohn, Simmons & Nelson, 2020; Del Giudice & Gangestad, 2021). What counts as reasonable is not necessarily simple, and inclusion of irrelevant choices can easily mask important choices because of the multiplicative nature of a branching path network (Del Giudice & Gangestad, 2021) (Fig. 1). Construction of a multiverse requires justification of which decisions are treated as variable, and why there is not an *a priori* and defensible single solution (Del Giudice & Gangestad, 2021). A multiverse populated with well-justified decisions allows the exploration of which choices inflate variation between analysis universes, while also offering insights into how to deflate variation [e.g., refinement of initial study design, the removal of ambiguities like tightening categories definitions; Steege *et al.* (2016)].

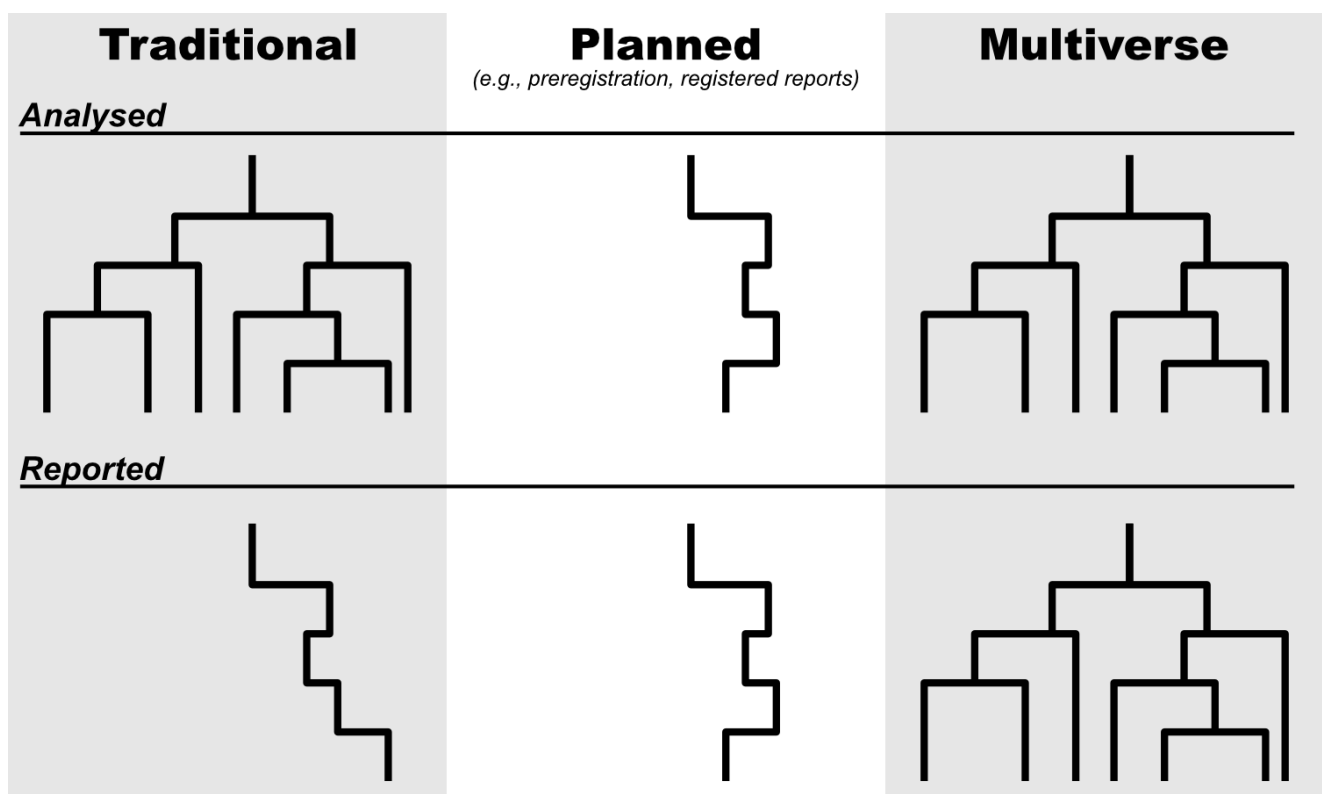


Figure 1. Diagram showing how multiverse analysis differs from other approaches. Each branch node represents a choice made during analysis

Ecological systems are complex to study and frustrate replication efforts (Nakagawa & Parker, 2015; Schnitzer & Carson, 2016), and in the case of movement ecology, the data analysed (i.e., derived data, such as step length, speed, and turn angle derived from timestamped coordinate data) require multiple stages of preprocessing. Therefore, multiverse analysis is an avenue to explore causes of variation between studies without the additional costs of practical studies, while also being capable of exploring data processing decisions that may not have immediately apparent impacts on final results.

If the data entropy (the process in which as data ages the chances of irreversible loss increase, Vines *et al.*, 2014) and resistance to data sharing (Miyakawa, 2020; Tedersoo *et al.*, 2021) can be overcome, we will be able to retroactively explore the impact of researcher degrees of freedom on ecological studies (Rijnhart *et al.*, 2021). Such retroactive assessment is an attractive option when other methods to explore false positive rates (Hoffmann *et al.*, 2021), such as preregistration and registered reports (Kaplan & Irvin, 2015; Scheel, Schijen & Lakens, 2021), will require more time to yield results. Ideally we can use multiverse analysis with preregistrations to boost transparency surrounding the inclusion of decisions and the rationale behind others exclusions (Dragicevic *et al.*, 2019; Simonsohn, Simmons & Nelson, 2020). Given the success of meta-analyses to overcome short-comings in the publication record [e.g., p-hacking;

Head et al. (2015)], multiverse analysis may aid the direction of future research efforts by providing a means of meeting calls to replicate results before collecting more (Nuijten et al., 2018).

However, as not all choices are equally valid, so multiverse analysis cannot simply provide a correct answer (Steege et al., 2016) –the “average” result is not necessarily the closest to the truth. If we were to undertake a multiverse analysis in a scenario with a “known truth”, i.e., using a simulated dataset (Bastiaansen et al., 2020), we may be able to detect identify the amount of variation from different sources [e.g, biological variation vs study design variation; Breznau et al. (2021)], and potentially the systematic biases stemming from specific choices (potentially via Bayesian Causal Forests Bryan, Yeager & O’Brien, 2019).

Use of simulated data is an established way to explore the robustness of methodologies (Minchin, 1987; Silva et al., 2020); and this project will harness the benefits of simulated data to assess the impacts of researcher degrees of freedom on the results garnered from animal movement datasets via a multiverse approach.

Here we describe the initial plan to undertake a multiverse approach to explore how decisions concerning the design and analysis of an animal tracking study can impact the findings on habitat selection. We describe the simulated scenarios we plan to base the study upon and a preliminary multiverse branching path consisting of three analysis approaches, as well as several broad hypotheses.

2 Methods

2.1 Simulating the Scenarios

We simulated three scenarios/species, comprising of different animals and landscapes. We simulated the landscapes using the NLMR v.1.1.1 package (Sciaini et al., 2018), and the animal movement using abmAnimalMovement v.0.1.3.0 (Marshall & Duthie, 2022). The abmAnimalMovement simulations are a discrete time, agent-based modelling approach for simulating animal movement. Further details of how the scenarios were parametrised can be found in (Marshall & Duthie, 2022), and also in the abmAnimalMovement github repository (abmAnimalMovement Github).

In brief: - Species 1, Badger: site fidelity to two shelter sites, a low movement speed constrained by terrestrial environment and territoriality, an 8-12 hour activity cycle with seasonal shifts. - Species 2, Vulture: medium site fidelity via the use of multiple roosting/resting sites a high and variable movement speed with minimal landscape resistance, an 8-12 hour activity cycle with seasonal shifts. - Species 3, King cobra: lower site fidelity making use of many shelter sites, a medium movement speed through a landscape with high resistance barriers, an 8-12 hour activity cycle combined with a approximately weekly forage-digest cycle and a weak seasonal cycle.

Each simulation generated a years worth of movement data, with a data point recorded every minute.

Our landscapes comprise of three elements, which are considered by the animal differently depending on the behavioural state the animal is in (resting, exploring, foraging). The three elements were matrices where each cell describes either foraging quality, shelter site quality, and movement ease. The three are non-independent, and based on a single initial random generation using a Gaussian field.

From the initial Gaussian field we altered the values to exaggerate the difference between high resource areas and low value areas. Broadly, we created: core resource areas (e.g., forests) with higher foraging quality but lower movement ease, edge areas that overlap with the forest with better movement ease, and more barren areas with high movement ease but with minimal foraging value. All shelter sites occurred within the higher quality core resource areas.

While our landscapes comprise of continuous values that determine the probability of a simulated animal using a cell or not, many habitat selection studies use categorical habitat metrics or land use types. Therefore, to simplify the interpretation of the multiverse we simplify out landscape into three distinct categories for analysis (Fig. 2): a low quality area where resources are low but movement is easy (class 0 in the figure), an edge habitat with middling resources and high movement ease (class 1), and a high resource habitat with low movement ease (class 2). While the categorisation of landscapes is a major decision during the analysis of habitat preference, we elected to keep the landscape classification constant to maintain an feasible multiverse analysis.

2.2 Sampling and Analysis Options

Ultimately number of variations and choices we were able to explore is dictated by computational costs and time. No multiverse can ever be entirely comprehensive, but we endeavoured to include the major decisions that are required during analysis that may not have immediately apparent “correct” answers.

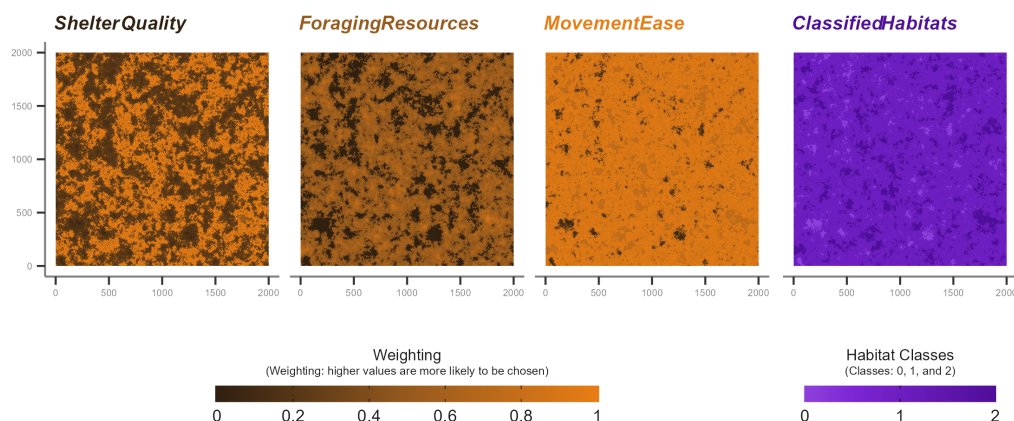


Figure 2. An example of the underlying environmental matrices used during simulation, and the subsequent habitat classes to be used in the habitat selection analysis.

2.2.1 Sampling

The first decisions concern data quantity. We varied tracking duration and tracking frequency primarily, while keep consistency fixed. While tracking consistency, or random/systematic data loss during tracking, is an element affecting data quantity the numerous ways of defining consistency means we did not explore tracking consistency at this time. Tracking timing is also an important consideration; for example, recording the locations of a diurnal animal only at night is highly unlikely to be a viable way of determining foraging choices. For this exploration we assumed that the researcher has sufficient knowledge about the animal's ecology to prioritise tracking during active hours. Timing only becomes a consideration when tracking frequency lowers to the point where the tracks performed during a day could occur entirely outside the animal's active period. Therefore, for the lower tracking frequencies we ensured that daily tracking was centred around midday (as all our simulated species had diurnal activity patterns).

Data or location quality is another point of sampling variation. Different bio-logging equipment, terrain, animal behaviour, and weather can all impact the location error when tracking an animal. As the causes and measurement of error, as well as the solutions to, location error are numerous we will not be exploring the impact of error during this exploration in favour of more variation in other decision nodes.

In summary, the sampling decisions will cover variation in species, tracking frequency, and tracking duration (Fig. ??A).

2.2.2 Analysis

For the analysis decisions we focused entirely on decisions that result from an R analysis workflow. R presents the number one tool for analysing animal movement data (Joo et al., 2022). Joo et al. (2020) provides a review of the R packages available to analysis movement data in R, and we used this review as a resource for determining the options for habitat analysis. Specifically, we explored a subset of decisions made during the workflows using adehabitatHS v.0.3.16 (Calenge, 2006), and amt v.0.1.7 (Signer, Fieberg & Avgar, 2019) R packages.

Combined the packages offer many options for habitat analysis, we focus on three that tackle individual habitat selection. One from the adehabitatHS package: resource selection ratios (Wides). And two from the amt package: resource selection function (RSF), step selection function (SSF) Each of the methods require downstream decisions resulting in a "garden of forking paths" to a final estimate of habitat selection (Fig. ??B).

The Wides and RSF methods share many analysis decisions, principally concerned with how available habitat is defined. The first decisions is whether to approach the habitat selection analysis as a type I, II or III design, as we are concerned about an individuals' habitat selection we will ignore type I. While type II habitat selection can identify individual preference, it requires habitat availability to be defined on a population or landscape scale (i.e., a universal availability). As

movement data presents an opportunity to estimate individual preference, and we are not simulating a true interacting population, we focus on type III. Type III and the need to individually define available habitat opens up many different analytically decisions concerning how that availability is defined.

We used a range of methods to build areas from the recorded locations of the animal: Minimum convex polygons (MCPs), Kernel Density Estimations (KDEs), Autocorrelated Kernel Density Estimations (AKDEs), and dynamic Brownian Bridge Movement Models (dBBMMs). Minimum convex polygons (MCPs) simply creates a polygon based on the out most locations, where the interior angles sum to less than 180 degrees.

Kernel Density Estimations (KDEs) use kernel smoothing to build a heat map of use or utilisation. Critical to KDEs operation is a bandwidth or smoothing factor (h) that alters the resulting area treated as used by the animal (Silva et al., 2020). For this analysis we use the reference bandwidth (h_{ref}) as it is commonly used (Crane et al., 2020), while also quickly and consistently calculable on a individual by individual basis. An alternative common method to determine smoothing factor is Least Squares Cross Validation (LSCV). However, we avoided the LSCV method for two reasons: 1) the LSCV method can fail to converge on a smoothing factor thereby introducing an unpredictable fail state in any potential multiverse, 2) the more limited area estimate it produces make it less suitable for defining habitat availability.

The limitations of MCPs and KDEs have prompted the development of newer methods of area use estimation (Silva et al., 2020) that better account for the non-independence and autocorrelative structures within animal movement data (Noonan et al., 2019). We used two of these methods to define availability.

First is the Autocorrelated Kernel Density Estimations (AKDEs) (Fleming et al., 2015; Fleming & Calabrese, 2017) from the *ctmm* package (Calabrese, Fleming & Gurarie, 2016; Fleming & Calabrese, 2021). The *ctmm* package provides a workflow for creating an area estimate based on a number of movement models. The movement models include different levels of autocorrelative structure: Ornstein-Uhlenbeck (OU) models account for central tendency, the Ornstein-Uhlenbeck Foraging (OUF) build on this by also accounting for autocorrelation in movement speed, while the Independent Identically Distributed (IID) models assume location independence meaning they are similar to traditional kernel density approaches. Each model type is also fitted in isotropic (circular) or anisotropic (elliptical) forms. The movement model used would constitute a decision during analysis; however, the *ctmm* package allows for model comparisons (using AICc) to chose a single best model. Therefore, we use the guidance from (Silva et al., 2022) to generate weighted AKDEs using perturbative hybrid residual maximum likelihood method (pHREML), and select the best performing by AICc for inclusion in further analysis. Using weighted AKDEs means the area estimates better address gaps between locations (Silva et al., 2022), with no apparent drawbacks nor large computational cost.

The second movement-specific method we used was dynamic Brownian Bridge Movement Models (dBBMMs) (Kranstauber et al., 2012) from the *move* package (Kranstauber, Smolla & Scharf, 2022), which estimates movement capacity of the animal to calibrate repeated random walks between known locations. DBBMMs require a window and margin size that defines the number of data points over which movement capacity (motion variance) is calculated (Kranstauber et al., 2012). Window and margin are defined by a number of data points; therefore, to keep the time they represent the same between different tracking frequencies, we changed their values for each tracking frequency. As our most infrequent tracking is 168 hours (1 week), we set the window to the number of data points collected over 168 hours, and a margin of 48 hours. The broader window and margin sizes helped reduce computational costs, a major concern when so many dBBMMs are being calculated.

Unlike the AKDEs, dBBMMs do not produce an utilisation distribution. As dBBMMs are estimating the uncertainty surrounding potential movement pathways, the resulting distribution is better described as an occurrence distribution. This occurrence distribution is describing the within-sample uncertainty rather than the potential areas available to the animal beyond the sampling period (i.e., it possess little to no predictive capabilities). Therefore, using dBBMMs to define habitat availability means rather than comparing the habitat use (derived from the movement data) to availability (broader area predicted by the movement data), we are more closely comparing comparing habitat use to an alternative measure of habitat use. We included dBBMMs to examine whether this arguably easy-to-make conceptual mistake impacts the final results. [ALSTON newer paper on unifying terms needs to be cited here]

What is considered available to an animal is not cleanly translated from biological concept to a statistical one. While the area estimation methods (MCPs, KDEs, AKDEs, dBBMMs) described above make –to varying degrees– use of the movement information contained within the dataset they remain abstractions of the underlying drivers of animal movement (i.e., selection and perception). Therefore, the researcher must make a some-what arbitrary choice regarding what is included/excluded and counted as available; Each of these area estimation methods require a choice to be made regarding the outermost boundary. We explored the impact of using a 90, 95, and 99% contour for all the methods. MCPs areas are generated based a percentage of the location points. Whereas the KDE, AKDE, and dBBMM methods require a contour to be extracted from the utilisation distribution (occurrence distribution for dBBMMs). AKDEs also provide a



95% confidence interval surrounding any chosen contour; for the purposes of simplicity we only use the point estimate for each % contour.

Once we had created a defined availability area, we generated points at which the habitat type is recorded to estimate the relative availability. How many and in what alignment these availability points are generated poses two additional points of analytical choice. For this study we varied the number of availability points as a multiple of the number of animal locations in the dataset: [ADD IN R CODE TO PULL THE CHOICE FROM THE TARGETS SCRIPT]. For the alignment of the points, points were either random or stratified within the available previously defined.

As mentioned resource selection function (RSF) share the above decisions on availability with the Wides methods. In addition, for the RSF we also explored the impact of varying the weighting of the available points when the RSF (i.e., generalized linear model / logistic regression) model is run, which impacts the fitting process. There are suggestions that altering the weighing can improve model performance and decrease the uncertainty surrounding habitat preference estimates [NEED THE FIEBERG CITE HERE]; we included this decision to examine whether the pursuit of a more confident answer is biasing the point estimate.

Other than the area based methods of Wides and RSF, we also explored of Step Selection Functions (SSF). Instead of estimating habitat availability using area, SSFs use observed step lengths and turn angles to generate available locations at each time step (i.e., for each data point in the dataset; strata). As a result SSFs have a number of decisions not shared with the area methods.

We explored three decisions associated with the generation of random (available but unused) locations for each step. The first is similar to the area methods, we varied the number of random locations generated per step from ## to ##. The second and third are concerning the distributions used to generate the random step and turn angles. Step lengths were tested with a Gamma and Exponential distributions, while turn angle was tested with Von-Mises and Uniform distributions.

The other decisions we explored in SSFs was whether to run the model as a standard step selection or an integrated step selection function. The standard step selection model is a conditional logistic regression, where we aim to estimate the used/non-used by the habitat values ($\text{case_} \sim \text{values} + \text{strata}(\text{step_id_})$). The integrated step selection model is very similar, apart from the predictors also included step length and turn angle. The addition of these two components is meant to reduce bias by better accounting for the selection that may simply be an artifact of the movements of the animal rather than active habitat preference [CITE HERE].

We have neglected to explore an expansive aspect of SSFs regarding high resolution data and the use of “bursts”. Bursts are used when the researcher wants to look at selection at a given timeframe; therefore, locations are grouped into bursts and the bursts become the strata in the model rather than timestep. Such grouping of data deserves investigation, but the myriad of ways bursts can be defined and their interaction with tracking duration, frequency, and consistency would warrant a separate multiverse analysis. As such we have restricted this exploration to step selection functions where the datapoints are equal to the strata in the model.

2.3 Assessing the multiverse

To generate the multiverse we used a meta-programming approach via the targets and tarchetypes R packages. The targets package was created to facilitate reproducible workflows, but is also paired with functionality enabling multiple analysis routes and parallel processing. Targets’ ability to track analysis task progress and intermediate R objects reduces repeat computation a complex workflow may not be able to be run in a single sitting. We created a branching analysis workflow based on the decisions we wished to explore, then used targets to iterate over all combinations of all decisions and compile all analysis end points ready for further examination.

Further examination was complicated by the three methods described above producing habitat preference values on different scales with different decisions associated with them. We elected to analyse the impact of the decisions separately. For each method we ran two Bayesian regression models to explore: 1) which decisions best explain the absolute deviation from the median estimated preference, 2) which decisions best explain the transformed deviation from the median estimated preference. The first model provides us with an idea of the decisions that can lead to random, but potentially unbiased noise surrounding the median estimate; whereas the second model highlights decisions that lead to systematic over or under estimation of a habitat preference.

We elected to have the model try and predict the median estimate of habitat preference because the simulation cannot provide a direct analogous value to the outputs of all three methods. Although the simulation is a discrete time model, where the animal makes decisions based upon a number of options, is similar to a step selection model, the decision making of the animal is made on two time frames breaking our ability to directly compare simulation values with step

selection outputs. Further, the simulated preference of the animal is balanced against other demands such as site fidelity, movement resistance, and avoidance of certain locations. The closest analogous values we could generate came from running RSF and iSSF models directly on the simulated decisions made by the animal. We ran these models for both time scales: movement decisions every time step, and destination decisions every behavioural state switch. These model results, while not strictly comparable to the other outputs, help confirm the animal was correctly preferring the chosen habitat (bare in mind the animal had a no preference exploratory state that would weaken the effect), and support the use of the median estimate to explore the cause behind the most deviant estimates.

We assessed Bayesian model convergence using rhat values, acf, and trace plots. Based on these assessments we modified the running parameters [FINAL VALUES HERE]

3 Results

The completion of the multiverse resulted in 614 datasets representing an individual animals movements sampled via a distinct tracking regime (i.e., a combination of tracking frequency and duration). The datasets covered 3 virtual species, 4 individuals per species, frequencies from 0.006 to 2 locations per hour, and durations from 7 to 240 days. Each of these datasets underwent every combination of analysis choices resulting in 281,952 estimates of habitat preference (Wides: 38,016 , RSF: 228,096 , SSF: 15,840).

3.1 General patterns

We can get an initial impression of the pattern of the final estimates of preference using a specification curve [CITE HERE]. In a scenario where the decisions make little difference, all estimates should group together around a common answer, with the majority of deviation from that common answer the result of sampling or individual variation.

Starting with the Wides results, we see a quite gentle *S* shape [FIGURE LINK HERE], with most estimates falling in the range of 0.5 to 2. While there is a clear agreement of positive selection for habitat 2 as expected, the exact estimate appears relatively sensitive to the choices made. Looking at the specific choices, we can see several weak patterns. The first is in tracking duration where we see slight reduction in the spread of estimates with longer tracking durations. The same is not clear in tracking frequency, because the lower tracking frequency could not be combined with lower tracking durations, resulting in fewer estimates for low frequency scenarios. Despite this, we do see a slight reduction in estimate variation from frequencies of 0.17 locations per hour and higher. Similar to the tracking intensity choices, the contour used during the area generation appears to affect the spread of estimates, with a larger contour leading to a greater agreement between estimates. So to does the multiplier of available points, with more points slightly reducing the spread of estimates, although we see no movement in the median estimate. Higher numbers of available points, and a higher tracking intensity did appear to lead to a greater chance of a successful estimate being produced (see numbers to the left of the 0 line in [LINK TO FIGURE]). The choice of area method (MCP = Minimum Convex Polygon, AKDE = Autocorrelated Kernel Density Estimate based on the best performing movement model, KDEhref = Kernel Density Estimate using reference bandwidth selection, dBBMM = dynamic Brownian Bridge Movement Model) appears to suggest more consistently similar estimates with AKDE and KDEhref methods compared to MCPs and dBBMMs.

The Resource Selection Function (RSF) specification curve shows a more ideal *S* shape, with the majority of estimates clearly converging on a common answer [FIGURE LINK HERE]. Similar to the Wides analysis the majority of the answers, and medians detected preference for habitat 2; and while there are answers suggesting avoidance of habitat 2, all but two medians connected to a specific choice (the lowest two tracking frequencies) suggest preference. To best see the changes in variation connected to a given choice, we can look at the rates and position of the outliers (highlighted in purple and orange). Although weak patterns, tracking duration, frequency, and multiplier of available points see slight reductions in the number of estimates very high or very low. Other patterns are difficult to discern, except for the weighting of used points that has a clear impact on the outlying low estimates. The distinct step-like pattern is a result of the logarithmic scale the weighting varies over (required to capture such a wide range of possible values), which can be seen reflected in other choices' groupings of estimates. Overall the weighting does not appear to impact the overall median, with weighting only appearing to dramatically impact the outlying results.

For the RSF analyses we also recovered the range of the 95% confidence intervals surrounding each estimate. When plotted against the estimate we see a clear groupings where those most extreme estimates are also, by a significant margin, the most uncertain.

The Step Selection Function (SSF) specification curves appears even more successful than the RSF. There is a decently clear *S* shape indicating many analysis end points resulting in similar estimates of preference. There are also considerably fewer estimates that indicate avoidance of habitat 2. The stand out patterns are more easily seen in the outlying

high estimates. Once again there are indications of fewer outlying high estimates when data quantity increases, both with duration and frequency. The number of points per step (or strata) appears to intuitively follow that more available steps leads to more consistent estimations, and also the outliers that exist appear to become less extreme. The decisions concerning the distributions used to model step lengths and turn angles, as well as whether the model is integrated or not appear to make little difference to the final estimates.

Very much like the RSF explorations of the raw data, the 95% confidence intervals are dramatically larger (between 100 and 1000 fold) for the outlying estimates than those nearer the median estimate.

3.2 Model exploration

While the specification curves are a good way of getting an intuitive sense of variation, the scale and range of estimates can skew ones' interpretation of which choices have the largest impact. Our Bayesian models cut through the noise displayed in [FIGURES XXXXXXXX] to reveal more generalised patterns and absorb variation stemming from the stochastic nature of the animal simulations (i.e., variation between individuals and species).

For the decisions connected to Wides results, the absolute deviation was curtailed most effectively by increases in data quantity, with increases in tracking duration (β -0.014, 95% median Highest-Density Continuous Interval CrI -0.016 – -0.011) being marginally more effective than tracking frequency (β -0.008, 95% CrI -0.011 – -0.006). Of the other decisions only the multiplier of available points appeared to decrease absolute deviation unambiguously (β -0.003, 95% CrI -0.004 – -0.002), and its effect was much smaller than those concerning data quantity. Choice of area method as well as the pattern used to generate random points seem to make little clear difference, with all credible intervals connected to those choices overlapping zero. Contour choice increases the absolute deviation from the median estimate *beta* 0.007, 95% CrI 0.006 – 0.008, that appears counter to the initial specification curve plot and highlights the difficulty of identifying patterns from the raw data alone.

Every decision examined by the RSF absolute effect model appears to be make a difference when estimating preference. Like the Wides analysis, data quantity is most important. The two largest effects are from tracking duration (β -0.819, 95% CrI -0.863 – -0.779) and frequency (β -1.499, 95% CrI -1.535 – -1.457), showing that more data leads to more stable estimates of habitat preference (i.e., more likely to get an answer closer to the overall median estimate).

However, unlike with Wides the other analysis decisions also matter, although to varying extents. It appears that larger contours (β -0.253, 95% CrI -0.272 – -0.233) and a greater multiplier of available points (β -0.239, 95% CrI -0.259 – -0.219) decreases the deviation from the median estimate. To a small extent using a stratified available points pattern lead to lower deviation from the median than a purely random available points pattern (β -0.059, 95% CrI -0.1 – -0.018). Choice of area method can alter results in two ways. Compared to using MCPs both AKDE (β -0.236, 95% CrI -0.292 – -0.18) and KDEhref (β -0.249, 95% CrI -0.301 – -0.192) produce estimates more likely to be closer to the median estimate, and to similar degrees. Dynamic Brownian Bridge Movement Models on the other hand are more likely to lead to greater variation in estimates than those analyses that use MCPs (β 0.307, 95% CrI 0.249 – 0.363).

Examining the interactions between tracking duration and frequency with the area methods further reveals the connection between dBBMMs are more variable estimates. While the other area methods interact with tracking duration in a similar fashion, dBBMMs are more sensitive meaning that at shorter tracking durations the variation created by choosing dBBMMs over another method is greater. When interacting with tracking frequency the sensitivity is less pronounced, instead we see MCPs being the most sensitive to a lowering of tracking frequency.

The SSF absolute deviation model matches very closely the previous two models, with tracking quantity both in terms of duration (β -1.471, 95% CrI -1.552 – -1.398) and frequency (β -1.676, 95% CrI -1.758 – -1.6) dominate in regards to the explaining why estimates tend to be closer to the median estimate. Unlike the other analysis all other decisions made negligible impact to an estimate's absolute proximity to the median estimate.

4 Discussion

While the overall pattern shows data quality supremacy, this cannot be used as an excuse to ignore the decisions during analysis. Two aspect of the results highlight that complacency is unwise. First RSF and SSF approach appear to be more consistent than the more basic and older Wides ratio based approach. Both RSF and SSF have improved capacity for accounting for the peculiarities movement data structure (e.g., via random effects) than the Wides approach. Second is the evidence for conceptual mistakes to have impacts on final results, implying that in some cases there are more correct answers to certain analysis decisions. The use of dBBMMs to define availability is not strictly correct as the (confidence) area generated by dBBMMs are reflected of uncertainty surrounding use. This renders the down stream analysis of habitat preference a comparison of use (measured by points) to use (measured by area of uncertainty surrounding

movement trajectories); as opposed to the desired use versus availability. The dBBMM results exemplify the need to maintain a connection between abstractions (e.g., an area) of the data and the intended use. A similar argument can be made for data quantity, where mismatches between frequency and desired behaviours or and inadequate duration will obscure answers regardless of the analysis applied. Our simulations are diurnal and the tracking spacing was skewed towards those diurnal hours for lower tracking frequency regimes. Failure to do this, instead allowing locations to be recorded equally at night and day, would have skewed estimates towards the habitat used for shelter rather than foraging. Matching question to analysis approach remains important, but the cost of mistakes may be less dramatic than feared.

Further good news is that some decisions that are connected to more stable estimates have very little reason not to be maximised. For example, the points multiplier or points per step decision can be maximised to whatever extent the computational resources available to the researcher allows. While we did not see the improvements in points per step in the SSF analysis, previous work has shown that the benefits of high available steps is more pronounced when working with rare habitat types. More available points reduces the chances of missing small areas of rare habitat; our analysis would not have revealed this because of the abundance of the preferred habitat type.

4.1 Limitations of the multiverse approach and future directions

Practical limitations simply prevent from all possible variations of all decisions to be explored. Not just because of time and computational limits, but the granularity of decision making examined can be ever increased.

Beyond the aforementioned limitations to the multiverse approach there is scope to expand the exploration. Our next step is to reapply the same meta-programming pipeline approach to population targeting habitat analyses. A focus on these approaches will allow us to explore the third key component of data quantity –sample size– as well as how well analyses tackle individual variability either in the analysis itself or during summarising of individual habitat preference estimates.

4.2 Conclusions

5 Acknowledgements

6 Software availability

We used R v.4.2.2 (R Core Team, 2022) via RStudio v.2023.3.0.386 (RStudio Team, 2022), and used rmarkdown v.2.19 (Xie, Allaire & Golemund, 2018; Xie, Dervieux & Riederer, 2020; Allaire et al., 2022), bookdown v.0.33 (Xie, 2016, 2022a), tinytex v.0.44 (Xie, 2019, 2022b), and knitr v.1.41 (Xie, 2014, 2015, 2022c) packages to generate type-set outputs.

Wolten et al. (2020)

We generated R package citations with the aid of grateful v.0.1.13 (Rodríguez-Sánchez, Jackson & Hutchins, 2022).

7 Data availability

References

- Albers C. 2019. The problem with unadjusted multiple and sequential statistical testing. *Nature Communications* 10:1921. DOI: [10.1038/s41467-019-09941-0](https://doi.org/10.1038/s41467-019-09941-0).
- Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R. 2022. *Rmark-down: Dynamic documents for r*.
- Anderson MS, Ronning EA, De Vries R, Martinson BC. 2007. The Perverse Effects of Competition on Scientists' Work and Relationships. *Science and Engineering Ethics* 13:437–461. DOI: [10.1007/s11948-007-9042-5](https://doi.org/10.1007/s11948-007-9042-5).
- Anvari F, Lakens D. 2018. The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology* 3:266–286. DOI: [10.1080/23743603.2019.1684822](https://doi.org/10.1080/23743603.2019.1684822).
- Auspurg K, Brüderl J. 2021. Has the Credibility of the Social Sciences Been Credibly Destroyed? Reanalyzing the “Many Analysts, One Data Set” Project. *Socius: Sociological Research for a Dynamic World* 7:237802312110244. DOI: [10.1177/23780231211024421](https://doi.org/10.1177/23780231211024421).
- Barto EK, Rilling MC. 2012. Dissemination biases in ecology: Effect sizes matter more than quality. *Oikos* 121:228–235. DOI: [10.1111/j.1600-0706.2011.19401.x](https://doi.org/10.1111/j.1600-0706.2011.19401.x).
- Bastiaansen JA, Kunkels YK, Blaauw FJ, Boker SM, Ceulemans E, Chen M, Chow S-M, Jonge P de, Emerencia AC, Epskamp S, Fisher AJ, Hamaker EL, Kuppens P, Lutz W, Meyer MJ, Moulder R, Oravecz Z, Riese H, Rubel J, Ryan O,

- Servaas MN, Sjobeck G, Snippe E, Trull TJ, Tschacher W, Veen DC van der, Wichers M, Wood PK, Woods WC, Wright AGC, Albers CJ, Bringmann LF. 2020. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal of Psychosomatic Research* 137:110211. DOI: [10.1016/j.jpsychores.2020.110211](https://doi.org/10.1016/j.jpsychores.2020.110211).
- Bishop D. 2019. Rein in the four horsemen of irreproducibility. *Nature* 568:435–435. DOI: [10.1038/d41586-019-01307-2](https://doi.org/10.1038/d41586-019-01307-2).
- Brembs B. 2018. Prestigious Science Journals Struggle to Reach Even Average Reliability. *Frontiers in Human Neuroscience* 12:1–7. DOI: [10.3389/fnhum.2018.00037](https://doi.org/10.3389/fnhum.2018.00037).
- Brembs B. 2019. Reliable novelty: New should not trump true. *PLOS Biology* 17:e3000117. DOI: [10.1371/journal.pbio.3000117](https://doi.org/10.1371/journal.pbio.3000117).
- Breznau N, Rinke EM, Wuttke A, Adem M, Adriaans J, Alvarez-Benjumea A, Andersen HK, Auer D, Azevedo F, Bahnsen O, Balzer D, Bauer G, Bauer P, Baumann M, Baute S, Benoit V, Bernauer J, Berning C, Berthold A, Bethke FS, Biegert T, Blinzler K, Blumenberg J, Bobzien L, Bohman A, Bol T, Bostic A, Brzozowska Z, Burgdorf K, Burger K, Busch K, Castillo JC, Chan N, Christmann P, Connelly R, Czymara CS, Damian E, Ecker A, Edelmann A, Eger MA, Ellerbrock S, Forke A, Forster AG, Gaasendam C, Gavras K, Gayle V, Gessler T, Gnambs T, Godefroidt A, Grömping M, Groß M, Gruber S, Gummer T, Hadjar A, Heisig JP, Hellmeier S, Heyne S, Hirsch M, Hjerm M, Hochman O, Hövermann A, Hunger S, Hunkler C, Huth N, Ignacz Z, Jacobs L, Jacobsen J, Jaeger B, Jungkunz S, Jungmann N, Kauff M, Kleinert M, Klinger J, Kolb J-P, Kołczyńska M, Kuk JS, Kunißen K, Sinatra DK, Greinert A, Lersch PM, Löbel L-M, Lutscher P, Mader M, Madia JE, Malancu N, Maldonado L, Marahrens H, Martin N, Martinez P, Mayerl J, Mayorga OJ, McManus P, Wagner K, Meeusen C, Meierrieks D, Mellon J, Merhout F, Merk S, Meyer D, Micheli L, Mijs JJB, Moya C, Neunhoeffer M, Nüst D, Nygård O, Ochsenfeld F, Otte G, Pechenkina A, Prosser C, Raes L, Ralston K, Ramos M, Roets A, Rogers J, Ropers G, Samuel R, Sand G, Schachter A, Schaeffer M, Schieferdecker D, Schlueter E, Schmidt K, Schmidt R, Schmidt-Catran A, Schmiedeberg C, Schneider J, Schoonvelde M, Schulte-Cloos J, Schumann S, Schunck R, Schupp J, Seuring J, Silber H, Slegers WWA, Sonntag N, Staudt A, Steiber N, Steiner N, Sternberg S, Stiers D, Stojmenovska D, Storz N, Striessnig E, Stroppe A-K, Teltemann J, Tibajev A, Tung BB, Vagni G, Van Assche J, Linden M van der, Noll J van der, Van Hootegeem A, Vogtenhuber S, Voicu B, Wagemans F, Wehl N, Werner H, Wiernik BM, Winter F, Wolf C, Yamada Y, Zhang N, Ziller C, Zins S, Zóltak T, Nguyen HHV. 2021. Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty. *MetaArXiv*. DOI: [10.31222/osf.io/cd5j9](https://doi.org/10.31222/osf.io/cd5j9).
- Bryan CJ, Yeager DS, O'Brien JM. 2019. Replicator degrees of freedom allow publication of misleading failures to replicate. *Proceedings of the National Academy of Sciences* 116:25535–25545. DOI: [10.1073/pnas.1910951116](https://doi.org/10.1073/pnas.1910951116).
- Calabrese JM, Fleming CH, Gurarie E. 2016. Ctm: An R Package for Analyzing Animal Relocation Data As a Continuous-Time Stochastic Process. *Methods in Ecology and Evolution* 7:1124–1132. DOI: [10.1111/2041-210X.12559](https://doi.org/10.1111/2041-210X.12559).
- Calenge C. 2006. The package adehabitat for the r software: Tool for the analysis of space and habitat use by animals. *Ecological Modelling* 197:1035.
- Cassey P, Ewen JG, Blackburn TM, Møller AP. 2004. A survey of publication bias within evolutionary ecology. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271. DOI: [10.1098/rsbl.2004.0218](https://doi.org/10.1098/rsbl.2004.0218).
- Clark TD, Raby GD, Roche DG, Binning SA, Speers-Roesch B, Jutfelt F, Sundin J. 2020. Ocean acidification does not impair the behaviour of coral reef fishes. *Nature* 577:370–375. DOI: [10.1038/s41586-019-1903-y](https://doi.org/10.1038/s41586-019-1903-y).
- Crane M, Silva I, Marshall BM, Strine CT. 2020. Lots of movement, little progress: A review of reptile home range literature. *bioRxiv* 2020.12.03.409292. DOI: [10.1101/2020.12.03.409292](https://doi.org/10.1101/2020.12.03.409292).
- Del Giudice M, Gangestad SW. 2021. A Traveler's Guide to the Multiverse: Promises, Pitfalls, and a Framework for the Evaluation of Analytic Decisions. *Advances in Methods and Practices in Psychological Science* 4:251524592095492. DOI: [10.1177/2515245920954925](https://doi.org/10.1177/2515245920954925).
- Desbureaux S. 2021. Subjective modeling choices and the robustness of impact evaluations in conservation science. *Conservation Biology* 35:1615–1626. DOI: [10.1111/cobi.13728](https://doi.org/10.1111/cobi.13728).
- Dragicevic P, Jansen Y, Sarma A, Kay M, Chevalier F. 2019. Increasing the Transparency of Research Papers with Explorable Multiverse Analyses. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland UK: ACM, 1–15. DOI: [10.1145/3290605.3300295](https://doi.org/10.1145/3290605.3300295).
- Fanelli D. 2010b. “Positive” Results Increase Down the Hierarchy of the Sciences. *PLoS ONE* 5:e10068. DOI: [10.1371/journal.pone.0010068](https://doi.org/10.1371/journal.pone.0010068).
- Fanelli D. 2010a. Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data. *PLoS ONE* 5:e10271. DOI: [10.1371/journal.pone.0010271](https://doi.org/10.1371/journal.pone.0010271).
- Fanelli D. 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90:891–904. DOI: [10.1007/s11192-011-0494-7](https://doi.org/10.1007/s11192-011-0494-7).
- Fleming CH, Calabrese JM. 2017. A new kernel density estimator for accurate home-range and species-range area estimation. *Methods in Ecology and Evolution* 8:571–579. DOI: [10.1111/2041-210X.12673](https://doi.org/10.1111/2041-210X.12673).
- Fleming CH, Calabrese JM. 2021. *Ctm: Continuous-time movement modeling*.
- Fleming CH, Fagan WF, Mueller T, Olson KA, Leimgruber P, Calabrese JM. 2015. Rigorous home range estimation with movement data: A new autocorrelated kernel density estimator. *Ecology* 96:1182–1188.
- Forstmeier W, Schielzeth H. 2011. Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and

- the winner's curse. *Behavioral Ecology and Sociobiology* 65:47–55. DOI: [10.1007/s00265-010-1038-5](https://doi.org/10.1007/s00265-010-1038-5).
- Forstmeier W, Wagenmakers E-J, Parker TH. 2017. Detecting and avoiding likely false-positive findings – a practical guide: Avoiding false-positive findings. *Biological Reviews* 92:1941–1968. DOI: [10.1111/brv.12315](https://doi.org/10.1111/brv.12315).
- Fraser H, Barnett A, Parker TH, Fidler F. 2020. The role of replication studies in ecology. *Ecology and Evolution* 10:5197–5207. DOI: [10.1002/ece3.6330](https://doi.org/10.1002/ece3.6330).
- Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F. 2018. Questionable research practices in ecology and evolution. *PLOS ONE* 13:e0200303. DOI: [10.1371/journal.pone.0200303](https://doi.org/10.1371/journal.pone.0200303).
- Freedman LP, Cockburn IM, Simcoe TS. 2015. The Economics of Reproducibility in Preclinical Research. *PLOS Biology* 13:e1002165. DOI: [10.1371/journal.pbio.1002165](https://doi.org/10.1371/journal.pbio.1002165).
- Gelman A, Loken E. 2013. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. :17.
- Grainger M, Bolam FC, Stewart G, Nilsen EB. 2019. Evidence synthesis for tackling research waste. *EcoEvoRxiv*. DOI: [10.32942/osf.io/42fkh](https://doi.org/10.32942/osf.io/42fkh).
- Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. 2015. The Extent and Consequences of P-Hacking in Science. *PLOS Biology* 13:e1002106. DOI: [10.1371/journal.pbio.1002106](https://doi.org/10.1371/journal.pbio.1002106).
- Hoffmann S, Schönbrodt F, Elsas R, Wilson R, Strasser U, Boulesteix A-L. 2021. The multiplicity of analysis strategies jeopardizes replicability: Lessons learned across disciplines. *Royal Society Open Science* 8:rsos.201925, 201925. DOI: [10.1098/rsos.201925](https://doi.org/10.1098/rsos.201925).
- Holman L, Head ML, Lanfear R, Jennions MD. 2015. Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording. *PLOS Biology* 13:e1002190. DOI: [10.1371/journal.pbio.1002190](https://doi.org/10.1371/journal.pbio.1002190).
- Huntington-Klein N, Arenas A, Beam E, Bertoni M, Bloem JR, Burli R, Chen N, Grieco R, Ekpe G, Pugatch T, Saavedra M, Stopnitzky Y. 2021. The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry* 59:944–960. DOI: [10.1111/ecin.12992](https://doi.org/10.1111/ecin.12992).
- Jennions MD, Møller AP. 2002. Publication bias in ecology and evolution: An empirical assessment using the 'trim and fill' method. *Biological Reviews of the Cambridge Philosophical Society* 77:211–222. DOI: [10.1017/S1464793101005875](https://doi.org/10.1017/S1464793101005875).
- Joo R, Boone ME, Clay TA, Patrick SC, Clusella-Trullas S, Basille M. 2020. Navigating through the r packages for movement. *Journal of Animal Ecology* 89:248–267. DOI: [10.1111/1365-2656.13116](https://doi.org/10.1111/1365-2656.13116).
- Joo R, Picardi S, Boone ME, Clay TA, Patrick SC, Romero-Romero VS, Basille M. 2022. Recent trends in movement ecology of animals and human mobility. *Movement Ecology* 10:26. DOI: [10.1186/s40462-022-00322-9](https://doi.org/10.1186/s40462-022-00322-9).
- Kaplan RM, Irvin VL. 2015. Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. *PLOS ONE* 10:e0132382. DOI: [10.1371/journal.pone.0132382](https://doi.org/10.1371/journal.pone.0132382).
- Kelly CD. 2019. Rate and success of study replication in ecology and evolution. *PeerJ* 7:e7654. DOI: [10.7717/peerj.7654](https://doi.org/10.7717/peerj.7654).
- Kranstauber B, Kays R, Lapoint SD, Wikelski M, Safi K. 2012. A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology* 81:738–746. DOI: [10.1111/j.1365-2656.2012.01955.x](https://doi.org/10.1111/j.1365-2656.2012.01955.x).
- Kranstauber B, Smolla M, Scharf AK. 2022. *Move: Visualizing and analyzing animal track data*.
- Lakens D, Adolphi FG, Albers CJ, Anvari F, Apps MAJ, Argamon SE, Baguley T, Becker RB, Benning SD, Bradford DE, Buchanan EM, Caldwell AR, Van Calster B, Carlsson R, Chen S-C, Chung B, Colling LJ, Collins GS, Crook Z, Cross ES, Daniels S, Danielsson H, DeBruine L, Dunleavy DJ, Earp BD, Feist MI, Ferrell JD, Field JG, Fox NW, Friesen A, Gomes C, Gonzalez-Marquez M, Grange JA, Grieve AP, Guggenberger R, Grist J, Harmelen A-L van, Hasselman F, Hochard KD, Hoffarth MR, Holmes NP, Ingre M, Isager PM, Isotalus HK, Johansson C, Juszczak K, Kenny DA, Khalil AA, Konat B, Lao J, Larsen EG, Lodder GMA, Lukavský J, Madan CR, Manheim D, Martin SR, Martin AE, Mayo DG, McCarthy RJ, McConway K, McFarland C, Nio AQX, Nilsson G, Oliveira CL de, Xivry J-JO de, Parsons S, Pfuhl G, Quinn KA, Sakon JJ, Saribay SA, Schneider IK, Selvaraju M, Sjoerds Z, Smith SG, Smits T, Spies JR, Sreekumar V, Steltenpohl CN, Stenhouse N, Świątkowski W, Vadillo MA, Van Assen MALM, Williams MN, Williams SE, Williams DR, Yarkoni T, Ziano I, Zwaan RA. 2018. Justify your alpha. *Nature Human Behaviour* 2:168–171. DOI: [10.1038/s41562-018-0311-x](https://doi.org/10.1038/s41562-018-0311-x).
- Lemoine NP, Hoffman A, Felton AJ, Baur L, Chaves F, Gray J, Yu Q, Smith MD. 2016. Underappreciated problems of low replication in ecological field studies. *Ecology* 97:2554–2561. DOI: [10.1002/ecy.1506](https://doi.org/10.1002/ecy.1506).
- Levins R, Lewontin R. 1985. *The dialectical biologist*. Harvard University Press.
- Liu Y, Althoff T, Heer J. 2020. Paths Explored, Paths Omitted, Paths Obscured: Decision Points & Selective Reporting in End-to-End Data Analysis. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, 1–14. DOI: [10.1145/3313831.3376533](https://doi.org/10.1145/3313831.3376533).
- Marshall BM, Duthie AB. 2022. *abmAnimalMovement: An r package for simulating animal movement using an agent-based model*. *F1000* 0:0.
- Minchin PR. 1987. Simulation of multidimensional community patterns: Towards a comprehensive model. *Vegetatio* 71:145–156.
- Miyakawa T. 2020. No raw data, no science: Another possible source of the reproducibility crisis. *Molecular Brain* 13:24, s13041-020-0552-2. DOI: [10.1186/s13041-020-0552-2](https://doi.org/10.1186/s13041-020-0552-2).

- Nakagawa S, Parker TH. 2015. Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology* 13:88. DOI: [10.1186/s12915-015-0196-3](https://doi.org/10.1186/s12915-015-0196-3).
- Noonan MJ, Tucker MA, Fleming CH, Akre T, Alberts SC, Ali AH, Altmann J, Antunes PC, Belant JL, Beyer D, Blaum N, Böhning-Gaese K, Cullen L, Paula RC de, Dekker J, Drescher-Lehman J, Farwig N, Fichtel C, Fischer C, Ford A, Goheen JR, Janssen R, Jeltsch F, Kauffman M, Kappeler PM, Koch F, LaPoint S, Catherine Markham A, Medici EP, Morato RG, Nathan R, Oliveira-Santos LGR, Olson KA, Patterson BD, Paviolo A, Ramalho EE, Roesner S, Schabo D, Selva N, Sergiel A, Xavier da Silva M, Spiegel O, Thompson P, Ullmann W, Zieba F, Zwijacz-Kozica T, Fagan WF, Mueller T, Calabrese JM. 2019. A comprehensive analysis of autocorrelation and bias in home range estimation. *Ecological Monographs*. DOI: [10.1002/ecm.1344](https://doi.org/10.1002/ecm.1344).
- Nuijten MB, Bakker M, Maassen E, Wicherts JM. 2018. Verify original results through reanalysis before replicating. *Behavioral and Brain Sciences* 41:e143. DOI: [10.1017/S0140525X18000791](https://doi.org/10.1017/S0140525X18000791).
- O'Boyle EH, Banks GC, Gonzalez-Mulé E. 2014. The Chrysalis Effect: How Ugly Initial Results Metamorphosize Into Beautiful Articles. *Journal of Management* 43:376–399. DOI: [10.1177/0149206314527133](https://doi.org/10.1177/0149206314527133).
- Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349:aac4716–aac4716. DOI: [10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716).
- Palmer AR. 2000. Quasi-Replication and the Contract of Error: Lessons from Sex Ratios, Heritabilities and Fluctuating Asymmetry. *Annual Review of Ecology and Systematics* 31:441–480. DOI: [10.1146/annurev.ecolsys.31.1.441](https://doi.org/10.1146/annurev.ecolsys.31.1.441).
- Parsons S. 2020. Exploring reliability heterogeneity with multiverse analyses: Data processing decisions unpredictably influence measurement reliability. *PsyArXiv*. DOI: [10.31234/osf.io/y6tcz](https://doi.org/10.31234/osf.io/y6tcz).
- Patel CJ, Burford B, Ioannidis JPA. 2015. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology* 68:1046–1058. DOI: [10.1016/j.jclinepi.2015.05.029](https://doi.org/10.1016/j.jclinepi.2015.05.029).
- R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reinert HK, Cundall D. 1982. An Improved Surgical Implantation Method for Radio-Tracking Snakes. *Copeia* 1982:702. DOI: [10.2307/1444674](https://doi.org/10.2307/1444674).
- Rijnhart JJM, Twisk JWR, Deeg DJH, Heymans MW. 2021. Assessing the Robustness of Mediation Analysis Results Using Multiverse Analysis. *Prevention Science*. DOI: [10.1007/s11121-021-01280-1](https://doi.org/10.1007/s11121-021-01280-1).
- Roche DG, Amcoff M, Morgan R, Sundin J, Finnøen MH, Lawrence MJ, Henderson E, Speers-Roesch B, Brown C, Clark TD, Bshary R, Jutfelt F, Binning SA. 2020. Behavioural lateralisation in a detour test is not repeatable in fishes. *EcoEvoRxiv*:62. DOI: [10.32942/osf.io/6kcwa](https://doi.org/10.32942/osf.io/6kcwa).
- Rodríguez-Sánchez F, Jackson CP, Hutchins SD. 2022. *Grateful: Facilitate citation of r packages*.
- RStudio Team. 2022. *RStudio: Integrated development environment for r*. Boston, MA: RStudio, PBC.
- Salis A, Lena J-P, Lengagne T. 2021. How Subtle Protocol Choices Can Affect Biological Conclusions: Great Tits' Response to Allopatric Mobbing Calls. *Animal Behavior and Cognition* 8:152–165. DOI: [10.26451/abc.08.02.05.2021](https://doi.org/10.26451/abc.08.02.05.2021).
- Sánchez-Tójar A, Nakagawa S, Sánchez-Fortún M, Martín DA, Ramani S, Girndt A, Bókony V, Kempnaers B, Liker A, Westneat DF, Burke T, Schroeder J. 2018. Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias. *eLife* 7:e37385. DOI: [10.7554/eLife.37385](https://doi.org/10.7554/eLife.37385).
- Scheel AM, Schijen MRMJ, Lakens D. 2021. An Excess of Positive Results: Comparing the Standard Psychology Literature With Registered Reports. *Advances in Methods and Practices in Psychological Science* 4:251524592110074. DOI: [10.1177/25152459211007467](https://doi.org/10.1177/25152459211007467).
- Schnitzer SA, Carson WP. 2016. Would Ecology Fail the Repeatability Test? *BioScience* 66:98–99. DOI: [10.1093/biosci/biv176](https://doi.org/10.1093/biosci/biv176).
- Sciaini M, Fritsch M, Scherer C, Simpkins CE. 2018. *NLMR and landscapetools: An integrated environment for simulating and modifying neutral landscape models in r*. *Methods in Ecology and Evolution* 00:1–9.
- Signer J, Fieberg J, Avgar T. 2019. *Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses*. *Ecology and Evolution* 9:880–890.
- Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, Bahník Š, Bai F, Bannard C, Bonnier E, Carlsson R, Cheung F, Christensen G, Clay R, Craig MA, Dalla Rosa A, Dam L, Evans MH, Flores Cervantes I, Fong N, Gamez-Djokic M, Glenz A, Gordon-McKeon S, Heaton TJ, Hederos K, Heene M, Hofelich Mohr AJ, Högden F, Hui K, Johannesson M, Kalodimos J, Kaszubowski E, Kennedy DM, Lei R, Lindsay TA, Liverani S, Madan CR, Molden D, Molleman E, Morey RD, Mulder LB, Nijstad BR, Pope NG, Pope B, Prenoveau JM, Rink F, Robusto E, Roderique H, Sandberg A, Schlüter E, Schönbrodt FD, Sherman MF, Sommer SA, Sotak K, Spain S, Spörlein C, Stafford T, Stefanutti L, Tauber S, Ullrich J, Vianello M, Wagenmakers E-J, Witkowski M, Yoon S, Nosek BA. 2018. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science* 1:337–356. DOI: [10.1177/2515245917747646](https://doi.org/10.1177/2515245917747646).
- Silva I, Crane M, Marshall BM, Strine CT. 2020. Reptiles on the wrong track? Moving beyond traditional estimators with dynamic Brownian Bridge Movement Models. *Movement Ecology* 8:43. DOI: [10.1186/s40462-020-00229-3](https://doi.org/10.1186/s40462-020-00229-3).
- Silva I, Fleming CH, Noonan MJ, Alston J, Foltá C, Fagan WF, Calabrese JM. 2022. Autocorrelation-informed home range estimation: A review and practical guide. *Methods in Ecology and Evolution* 13:534–544. DOI: [10.1111/2041-210X.13786](https://doi.org/10.1111/2041-210X.13786).
- Simmons JP, Nelson LD, Simonsohn U. 2011. False-Positive Psychology: Undisclosed Flexibility in Data Collection and

- Analysis Allows Presenting Anything as Significant. *Psychological Science* 22:1359–1366. DOI: [10.1177/0956797611417632](https://doi.org/10.1177/0956797611417632).
- Simonsohn U. 2015. Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychological Science* 26:559–569. DOI: [10.1177/0956797614567341](https://doi.org/10.1177/0956797614567341).
- Simonsohn U, Simmons JP, Nelson LD. 2020. Specification curve analysis. *Nature Human Behaviour* 4:1208–1214. DOI: [10.1038/s41562-020-0912-z](https://doi.org/10.1038/s41562-020-0912-z).
- Smaldino PE, McElreath R. 2016. The natural selection of bad science. *Royal Society Open Science* 3:160384. DOI: [10.1098/rsos.160384](https://doi.org/10.1098/rsos.160384).
- Steen S, Tuerlinckx F, Gelman A, Vanpaemel W. 2016. Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science* 11:702–712. DOI: [10.1177/1745691616658637](https://doi.org/10.1177/1745691616658637).
- Tang-Martínez Z. 2020. The history and impact of women in animal behaviour and the ABS: A North American perspective. *Animal Behaviour* 164:251–260. DOI: [10.1016/j.anbehav.2019.12.011](https://doi.org/10.1016/j.anbehav.2019.12.011).
- Tedersoo L, Küngas R, Oras E, Köster K, Eenmaa H, Leijen Ä, Pedaste M, Raju M, Astapova A, Lukner H, Kogermann K, Sepp T. 2021. Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data* 8:192. DOI: [10.1038/s41597-021-00981-0](https://doi.org/10.1038/s41597-021-00981-0).
- Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, Franklin MT, Gilbert KJ, Moore J-S, Renaut S, Rennison DJ. 2014. The Availability of Research Data Declines Rapidly with Article Age. *Current Biology* 24:94–97. DOI: [10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014).
- Vinkers CH, Tjldink JK, Otte WM. 2015. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis. *BMJ*:h6467. DOI: [10.1136/bmj.h6467](https://doi.org/10.1136/bmj.h6467).
- Wang D, Forstmeier W, Ihle M, Khadraoui M, Jerónimo S, Martin K, Kempnaers B. 2018. Irreproducible text-book “knowledge”: The effects of color bands on zebra finch fitness: COLOR BANDS HAVE NO EFFECT ON FITNESS IN ZEBRA FINCHES. *Evolution* 72:961–976. DOI: [10.1111/evo.13459](https://doi.org/10.1111/evo.13459).
- Ware JJ, Munafò MR. 2015. Significance chasing in research practice: Causes, consequences and possible solutions: Significance chasing. *Addiction* 110:4–8. DOI: [10.1111/add.12673](https://doi.org/10.1111/add.12673).
- Weaver SJ, Westphal MF, Taylor EN. 2021. Technology wish lists and the significance of temperature-sensing wildlife telemetry. *Animal Biotelemetry* 9:29. DOI: [10.1186/s40317-021-00252-0](https://doi.org/10.1186/s40317-021-00252-0).
- Webb SS, Demeyere N. 2021. Multiverse Analysis: A Method to Determine Researcher Degrees of Freedom in Test Validation. *PsyArXiv*. DOI: [10.31234/osf.io/nhrwq](https://doi.org/10.31234/osf.io/nhrwq).
- Winne CT, Willson JD, Andrews KM, Reed RN. 2006. Efficacy of marking snakes with disposable medical cautery units. *Herpetological Review* 37:52–54.
- Wolen AR, Hartgerink CHJ, Hafen R, Richards BG, Soderberg CK, York TP. 2020. osfr: An R interface to the open science framework. *Journal of Open Source Software* 5:2071. DOI: [10.21105/joss.02071](https://doi.org/10.21105/joss.02071).
- Xie Y. 2014. *Knitr: A comprehensive tool for reproducible research in R*. In: Stodden V, Leisch F, Peng RD eds. *Implementing reproducible computational research*. Chapman; Hall/CRC.
- Xie Y. 2015. *Dynamic documents with R and knitr*. Boca Raton, Florida: Chapman; Hall/CRC.
- Xie Y. 2016. *Bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman; Hall/CRC.
- Xie Y. 2019. *TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX distribution based on TeX live*. *TUGboat*:30–32.
- Xie Y. 2022a. *Bookdown: Authoring books and technical documents with r markdown*.
- Xie Y. 2022c. *Knitr: A general-purpose package for dynamic report generation in r*.
- Xie Y. 2022b. *Tinytex: Helper functions to install and maintain TeX live, and compile LaTeX documents*.
- Xie Y, Allaire JJ, Golemund G. 2018. *R markdown: The definitive guide*. Boca Raton, Florida: Chapman; Hall/CRC.
- Xie Y, Dervieux C, Riederer E. 2020. *R markdown cookbook*. Boca Raton, Florida: Chapman; Hall/CRC.
- Young C, Holsteen K. 2017. Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis. *Sociological Methods & Research* 46:3–40. DOI: [10.1177/0049124115610347](https://doi.org/10.1177/0049124115610347).