Different definitions of species presence to refine estimates of local species richness

Running title: daily site use richness vs seasonal maximum richness

List of authors: Alejandro Ruete[1,2, a], Tomas Pärt[2], Åke Berg[3], Jonas Knape[2], Debora Arlt[2]

Affiliations

1. Greensway, Ulls väg 29A. SE-756 51 Uppsala, Sweden.
2. Department of Ecology, Swedish University of Agricultural Sciences, Box 7044, SE-750 07 Uppsala, Sweden.
3. Swedish Biodiversity Centre, Swedish University of Agricultural Sciences, Box 7016, SE- 750 07 Uppsala, Sweden.

[a] Corresponding author. Telephone: +46 18 672453, E-mail: aleruete@gmail.com

Abstract

Aim

To improve predictions of spatial and temporal patterns of species richness it is important to consider how species presence at a site is defined. This is because this definition affects our estimate of species richness, which should be aligned with the aims of the study, e.g. estimating richness of the breeding community. Here we explore the sensitivity of species richness estimates to criteria for defining presence of species (e.g. in relation to number of days present during the breeding season) at 107 wetlands.

Innovation

We use opportunistic citizen science data of high density (a total of 151,817 observations of 77 wetland bird species; i.e. about 16 observations per day) to build site-occupancy models calculating occupancy probabilities at a high temporal resolution (e.g. daily occupancies) to derive probabilistic estimates of seasonal site use of each species. We introduce a new way for defining species presence by using different criteria related to the number of days the species are required to be present at local sites. We compared patterns of species richness when using these different criteria of species inclusions.

Main conclusion

While estimates of local species richness derived from high temporal resolution occupancy models are robust to observational bias, these estimates are sensitive to restrictions concerning the number of days of presence required during the breeding season. Unlike complete local species lists, summaries of seasonal site use and different presence criteria allow identifying differences between sites and amplifying the variability in species richness among sites. Thus, this approach allows filtering out species according to their phenology and

migration behaviour (e.g. passer-by species) and could improve the explanatory power of environmental variables on predictive models.

Keywords: biodiversity informatics, citizen science data, GBIF, migratory birds, occupancy model, opportunistic observations, presence-only data, site use, Swedish Species Gateway

Introduction

Measures of biodiversity are of central interest to many disciplines in ecology, from community and macroecology to functional ecology and conservation (Chapin *et al.*, 2000; Sala *et al.*, 2000; Hubbell, 2001; Leibold *et al.*, 2004; Isbell *et al.*, 2017). Biodiversity measures (a.k.a. biodiversity variables) are elaborations upon primary data such as species observations (Schmeller *et al.*, 2017). These measures are the species presence, species richness and community composition (or species lists) from which richness of functional traits could also be derived. However, these summary metrics are clearly dependent on the definition of the presence status of a species at a site that determines a species' inclusion in the summary metric. In other words, when is a species part of a local community? Is it enough for it to be present at a site only once during the season?

Here, we propose that in order to improve our understanding of species assemblages, we should first focus on how we define the presence of a species at a site, i.e. the data that is being fed into predictive models, before we can focus on how species distributions and local species richness are explained and predicted. So far, much focus has been set on the ability of predictive models to estimate species richness at unvisited sites (Dubuis *et al.*, 2011; Guisan & Rahbek, 2011; Calabrese *et al.*, 2014; Pollock *et al.*, 2014; Distler *et al.*, 2015; Zurell *et al.*, 2016). However, although mean richness levels are accurately predicted, these models typically over-predict low local species richness and under-predict high local species richness (Zurell *et al.*, 2016). Typical biases in richness predictions seem to be affected by species' prevalence and site use (e.g. habitat and resource requirements), which can only partially be amended by adequate predictor choice and resolution (Zurell *et al.*, 2016).

Based on recent modelling advancements (Ruete *et al.*, 2017), we introduce a new way of looking at species counts that challenges previous definitions of species presence in light of how each species use a site. So far, most modelling approaches use summaries of presence/absence or presence-only data collected during a rather long time period (e.g. typically, the population closure is assumed over a season, but could be a decade too) to infer about the presence of a species at a site. As we previously showed (Ruete *et al.*, 2017), seasonal closure periods may over-simplify the species phenology and day-to-day site use patterns, making it impossible to e.g. discriminate passer-by species from species attempting to breed in the area. Furthermore, mobile species, such as birds, might have large home ranges and use several sites for foraging and locally breeding species will therefore not be present at a site at all visits, i.e. hence it could be temporarily absent from a site during the season. The question that arises then is how to define the presence of a species, and in extension the species richness at a site that better reflects the species' more permanent site use in terms of diet, foraging and breeding habitat. Different definitions of species presence at a site are also of interest in communities with rapid succession within each season where some ephemeral species are observable during very short periods of time but may determine the succession of species during the rest of the season (e.g. algae, insects or vascular plants; Southwood *et al.*, 1979; Bond *et al.*, 1984; Berntsson & Jonsson, 2003; Whalen *et al.*, 2016).

One way of summarizing richness over a time window is by setting requirements on the amount of time units (i.e. days) a species needs to be present for it to be counted in estimates of local species richness. Such criteria are very much the same as used for several standardised inventory methods (e.g. territory mapping; Bibby, 2000). In this way, different criteria can be added for different species to filter species to be included as permanent members of the breeding community.

Techniques for stacking species-specific occupancies or distributions are popular among ecologists because they are intuitive and easy to apply (Calabrese *et al.*, 2014), by summing either binary states (presence/absence) or occupancy probabilities. However, these techniques have not been confronted with multiple potential definitions of species presence. With the advent of big biodiversity data (e.g. from citizen science), site-occupancy models can be used to estimate occupancy probabilities at a high temporal resolution, e.g. daily during a season (Ruete *et al.*, 2017), and to derive probabilistic estimates of seasonal site use. Here we explore how time series of site use can be used to constrain definitions of local species presence using biologically informed criteria to discriminate e.g. breeding species from passer-by's. We also propose that this approach allows refining species richness estimates and local species lists before using them in predictive models.

We used high density opportunistic citizen science data from popular birding wetlands in Sweden to explore the sensitivity of probabilistic stacking techniques to alternative definitions of species presence based on sequences of daily occupancy probabilities of wetland birds (mostly migrant species) in Sweden during a breeding season. We specifically analysed how different criteria for species presence lead to differences in species richness, and if they can be used to objectively filter out e.g. passer-by and vagrant species from local species list.

Methods

Data

We obtained data from Artportalen (Swedish Species Gateway, www.artportalen.se) by using the Swedish LifeWatch Analysis portal (Leidenberger *et al.*, 2016). This data is also available at the Global Biodiversity Information Facility (www.gbif.org). We extracted a total of 151,817 opportunistic presence-only species observations collected during 28,207 independent visits over 90 days (April to June) during 2014 for 77 wetland bird species at 107 wetland sites in Sweden (Fig. S1). A visit was defined as all observations made by each observer at a site during a day. The sampled area was restricted to a convex hull of all the locations from which observations were reported within a predefined polygon over the wetland extension used to query the observations from the Analysis portal. The area of the convex hull was later used to analyse if there were any effects of sampled area on species-richness estimates.

Species richness – using different definitions of presence

We used a site-use model, that assumes each day as closure periods (Ruete *et al.*, 2017), to estimate daily site- and species-specific occupancy probabilities. To explore the temporal variation in species richness we calculated daily, monthly and seasonal estimates of species richness based on the daily estimates of occupancy probability (p_{dij}, for day d, species i and site j) obtained from the site-use model.

Daily local richness (S_{jd}^{day}) was calculated following Calabrese *et al*. (2014) summing daily estimates of occupancy probability, p_{dij}, over species (Fig. 1). To summarise local species richness over time periods longer than the model resolution (i.e. >1 day) we used two approaches: 1) based on mean site use, and 2) based on different definitions for presence using different criteria for the number of days a species is required to be present.

Using the mean site use as a criterion for presence (Ruete *et al.*, 2017) we estimated mean richness over a time period (here months and the whole season): we first computed the mean of each species' estimated daily occupancy probability over the specified time window at each site (p_{ij}^{mean} = mean site use; Ruete *et al.*, 2017). We then computed estimates of species richness per site (S_j^{mean}, henceforth) by summing the mean probabilities (p_{ij}^{mean}) over all species j (Fig. 1). Therefore, S_j^{mean} estimates the number of species expected to co-occur at site j on any given day during the time period. This approach does not allow to identifying species with different phenologies (e.g. species arriving and breeding at different times) and behaviours (e.g. species passing by), and therefore is not possible to construct species lists as species assemblage may change from day to day.

To be able to distinguish between different phenologies and behaviours we also summarized richness over a time window by setting requirements on the amount of time units (i.e. days) a species needs to be present for it to be included in estimates of local species richness. We used criteria for the number of days a species needs to be present, either in any sequence (spread) or strictly on consecutive (continuous) days within the season. We used thresholds of 1, 20, 30 and 45 days during our 90 day (3 month) season to test for sensitivity to these criteria, see Table 1 and Fig. 1. Since species-specific models were fitted within the Bayesian framework, the model results comprise more than one estimate of daily occupancies (i.e. posterior distributions of daily occupancy status). Hence, the criteria for seasonal presence were assessed for each of the 2,000 posterior samples (N) of local sequences of daily occupancies summarising model uncertainty. The occupancy probability of species i at site j during a longer time window is then estimated as $p_{ij} = \sum_{n=1}^{N} O_{ijn}/N$, where the inclusion parameter O_{ijn} is 1 if a given criterion as defined above (Fig. 1) is fulfilled, and 0 otherwise, evaluated for each posterior sample n. Local richness at site j under a given criterion, e.g. 1 day, is then defined as $S_j^{1d} = \sum_{i=1}^{R} p_{ij}$ where R = 77 species. $S_j^{'criterion'}$ estimates richness as the total number of species in the community that fulfils certain presence criteria. Because this approach is binary regarding the fulfilment of a criteria it can produce local species lists.

As a baseline for comparison we compiled daily, monthly and seasonal lists of detected species at each site (observed species richness), solely based on the opportunistic primary observational data (i.e. number of species per month and season that were observed at least once under that period). For comparison with the criterion-based estimates, we also compiled species lists, using the same criteria for presence of a species, i.e. given each species was observed at least 1, 20, 30 or 45 days.

To test for the sensitivity of estimates of local richness to the variability in site- and time-specific sampling effort, we fitted logarithmic regression models testing for the effect of the number of visits and sampled area on estimates of site-specific richness. We assumed the number of species at site i to follow a Poisson distribution with mean λ_i, which was further modelled as $\ln(\lambda_i) = \alpha + \beta X_i$; where α is an intercept parameter, X_i is a matrix of explanatory variables, β is a vector of associated effect size parameters. For this we used the *glm(..., family = "poisson")* command in R (R Development Core Team, 2014).

Results

Species richness varied across the season. The monthly richness peaked in May (observed richness and estimated richness S^{1d}, Fig. 2), i.e. a time period when both migrants and breeders may be present at the wetlands. The difference between observed and estimated S^{1d} richness decreased as the time periods over which the data is summarised increased (from 1 day to month and to season). That is because many species will be detected at least once during a longer time period, although this also depends on the temporal patterns in number of visits per site (e.g. the difference was greater in June when the number of visits decayed). Although sampling effort in opportunistic observations of birds typically decreases at the end of the season (thus the lower number of detected species, Fig. 2), the occupancy model can correct for this tendency (see estimated richness in Fig. 2 and Ruete *et al.*, 2017). $S^{mean.Season}$ is the number of species expected to co-occur daily at a site, which is mathematically the same as the mean over the season of S^{day}. Due to the turnover in species diversity over the season, the number of species co-occurring daily at a site ($S^{mean.Season}$) was, as expected, much lower than $S^{1d.Season}$, which summarizes the total number of species present at least once at each site (Fig. 2).

Sensitivity of estimated richness to different definitions of species presence

Estimates of seasonal richness were very sensitive to the number of days required for considering a species to be present at a site (Fig. 3). Compared to the criterion *1d* criterion (i.e. observed at least once during April to June), naturally, the number of species considered as present decreased markedly (30% and 50% depending on the continuity criterion) when the species needed to occupy the site at least 20 days during the same time period (Fig. 3). Further increasing the restrictions on number of days present also reduced the number of species but less dramatically so (Fig. 3). Similarly, estimates under the continuous criterion (i.e. presences during consecutive days) were at least 20% lower than estimates under the spread criterion (i.e. presence on a given number of days in any sequence during a time period).

Sensitivity of estimated richness to sampling effort (number of visits and sampled area)

As expected, more visits resulted in greater observed and estimated species richness. In comparison to observed richness this effect was, however, widely reduced on richness estimates from the occupancy models (Fig. 4), with the estimated rate of increase in richness per added visit ranging between only 1.8% ($p > 0.05$) for S^{1d} and 6.3% ($p < 0.01$) for S^{45d} (Table S1). We also found a positive effect of sampled area on estimated richness, but only when richness was based on primary observations or the stricter 30 and 45 day criteria (S^{30d} and S^{45d}). However, adding sampled area to a model already accounting for the effect of number of visits did not improve the model fit (Table S1).

Discussion

Occupancy models, under the right circumstances, allow correcting for detection errors and predicting occupancy using presence-only data more precisely and presumably in a less biased way than other modelling techniques (Kéry & Royle, 2008; Dubuis *et al.*, 2011). However, occupancy models to estimate occupancy of breeding populations typically assume closure periods over the whole breeding season, neglecting effects of phenology and migration behaviour on the presence of species and hence occupancy probabilities across the season (Ruete *et al.*, 2017). For mobile species, assuming closure periods over the whole breeding season will also include species that are only passing by either during migration or during distant foraging trips. By making use of a large number of opportunistic observations spread out during the breeding season and specifically modelling the daily variability in occupancy probabilities for 77 wetland bird species at 107 sites, we were able to use different criteria for including species in each local species lists which allows to distinguish between species according to their phenology and migration behaviour. Here, we suggest using corrected occupancy data with high-temporal resolution together with biologically reasonable criteria to appropriately filter local species lists as the initial step before further analysing spatial and temporal patterns of species diversity.

An important question then is: What richness estimate to use in a given study. Here we compared i) daily species richness (S^{day}), ii) the average number of species daily co-occurring at a site over a chosen period (here month and season, S^{mean}), and iii) species richness where species are filtered based on criteria for presence (S^{1d} - S^{45d}), i.e. the number of days a species is required to be present at a site to be considered present in the season (here we chose 1, 20, 30 and 45 days). Richness estimated by summarising means of daily probabilities of occupancy (S^{mean}) describes the average site use of the whole bird community over time, that is, how many species are expected to share a site at a given time (day). However, since in this approach species richness is the results of added occupancy probabilities, the species actually being present cannot be identified and hence does not allow constructing local species lists or separating between species with different phenologies and behaviours (e.g. species passing by). Consider the case of three species with an occupancy probability across the period p_{ij}^{mean} = 0.27. These species may contribute equally little to the estimated richness S^{mean}. However, species 1 has a constant daily occupancy probability p_{dij} = 0.27 across the season, species 2 has a low p_{dij} = 0.1 during the first 70 days but p_{dij} = 0.9 during the last 20 days, while species 3 is likely present on 20 random days with p_{dij} = 0.9 and most likely absent the rest of the days with p_{dij} = 0.1. Biological criteria are needed to decide about the inclusion of a species, e.g. species most likely attempting to breed at a site. For example, one may want to use restrictive criteria as e.g. required presence during at least 45 days, to estimate the richness of species only using the site regularly during the breeding season (thus filtering out species passing by).

Richness estimates using a permissive criterion e.g. S^{1d} (i.e. species observed at least once during the time period) result in complete lists of species that could be observed at a site, including species using the site only temporary (e.g. passer-by and vagrant species). Cases where the inclusion of such species is of interest are surveys of e.g. stop-over sites or communities where species occur with rapid succession within a season and some ephemeral species are observable only during very short periods of time (e.g. algae, insects or vascular plants; Southwood *et al.*, 1979; Bond *et al.*, 1984; Berntsson & Jonsson, 2003; Whalen *et al.*, 2016). Then, using a permissive criterion will produce estimates of richness that are more suitable to understand the composition of those communities. However, in cases where the interest is the breeding community of migrant species more restrictive criteria are to define the presence of each species.

Different criteria need, however, to be chosen with care. Using continuous-presence criteria resulted in lower estimated richness compared to the spread-presence criteria (Fig. 3) indicating that there is a substantial group of vagrant species or species having large home ranges and thus being likely temporarily missing at a site. On the other hand, when continuous sequences of species presence may not be correctly identified, spread-presence criteria may safeguard from uncertainties in the occupancy model (e.g. due to very low detection probabilities; see Table S2: many of the 77 wetland species had low detection probabilities).

As expected, the variability in richness among sites increased when increasing the restrictions of inclusion of species in local lists (Fig. 3), highlight the differences between sites (Ruete *et al.*, 2017). Thus, this approach could improve the explanatory power of environmental variables describing the site use and breeding habitat selection reducing the risk of over- and underprediction error of predictive models (Calabrese *et al.*, 2014; Zurell *et al.*, 2016).

Although the number of visits at a site had strong positive effect on the number of species observed, when correcting for variability in the detection probability by occupancy models, the number of visits were only marginally positively associated with estimated species richness (Fig. 4). Given that richer sites may attract more voluntary observers the remaining effect of sampling effort (i.e. number visits) may be correlational and not strictly causal. These results suggest that occupancy models are robust for estimating richness compared to approaches using primary observational data, supporting the findings of Isaac *et al.* (2014).

Conclusion

We show how species presence can be defined in several ways when exploring species observation data at high temporal resolution, and how derived biodiversity variables, like estimates of species richness, are sensitive to it. In light of the results presented here, we suggest that the following questions should be considered when using species observations to define species presence: 1) how likely is it for different species to violate the closure assumption? That is, e.g. what possible dynamics in the species presence may be relevant to include in the occupancy model that is not captured by a assumed closure period of three or four months? 2) How relevant is the inclusion of e.g. vagrant species? That is, what richness (e.g. daily site use or seasonal maximum) is of interest? When the aim is to estimate the complete lists of species that could be visiting sites, even those occurring briefly, and if multiple observations per day are common we suggest to use S^{1d}. When the aim is to estimate the number of breeding species we suggest more restrictive approaches (such as S^{45d} or S^{20dC}). When the main interest is in summarising expected number of species simultaneously using the site then averages, such as S^{mean}, are recommended. Following those decisions, the filtered estimates of local species richness produced can be analysed using macroecological models (Dubuis *et al.*, 2011; Guisan & Rahbek, 2011). Alternatively, the presence criteria filters could be used per species if the intention is to stack species-specific distribution models either independently (Calabrese *et al.*, 2014) or integrated in multispecies occupancy models (MSOMs; Dorazio *et al.*, 2006; Iknayan *et al.*, 2014). In case local species lists are required for further analysis, e.g. community similarity indices or diversity of functional traits, criteria-based approaches (e.g. S^{1d} - S^{45d}) could be used. Then, bootstrapping techniques are recommended in order to account for and propagate the model uncertainty.

Given that opportunistic species observations are accumulating at a high rate in biodiversity databases (especially for birds; Graham *et al.*, 2004; Amano *et al.*, 2016) and the potential of these data continues to be uncover (Tulloch *et al.*, 2013; Theobald *et al.*, 2015; Ruete *et al.*, 2017), we envision that methods like the ones presented here will become more and more common.

Acknowledgements

We are thankful to Michał Żmihorski for fruitful discussions at early stages of this work. The research was financially supported by the Swedish Research Council VR (JK) and FORMAS (DA, TP).

Data accessibility

All the primary biodiversity data used in this study is available through Analysportalen.se and GBIF.org. It is also deposited as a .csv file in figshare.com (DOI:10.6084/m9.figshare.5178253) together with .shp files with the polygones describing the area of each wetland site.

References

Amano, T., Lamming, J.D.L. & Sutherland, W.J. (2016) Spatial gaps in global biodiversity information and the role of citizen science. *BioScience*, biw022.

Berntsson, K.M. & Jonsson, P.R. (2003) Temporal and spatial patterns in recruitment and succession of a temperate marine fouling assemblage: A comparison of static panels and boat hulls during the boating season. *Biofouling*, **19**, 187–195.

Bibby, C.J. (2000) *Bird Census Techniques*, Elsevier.

Bond, W.J., Volk, J. & Viviers, M. (1984) Variation in seedling recruitment of cape Proteaceae after fire. *Journal of Ecology*, **72**, 209–221.

Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014) Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.

Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E., Mack, M.C. & Diaz, S. (2000) Consequences of changing biodiversity. *Nature (London)*, **405**, 234–242.

Distler, T., Schuetz, J.G., Velásquez-Tibatá, J. & Langham, G.M. (2015) Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *Journal of Biogeography*, **42**, 976–988.

Dorazio, R.M., Royle, J.A., Söderström, B. & Glimskär, A. (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, **87**, 842–854.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122–1131.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, **19**, 497–503.

Guisan, A. & Rahbek, C. (2011) SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*, Princeton University Press, Princeton, New Jersey.

Iknayan, K.J., Tingley, M.W., Furnas, B.J. & Beissinger, S.R. (2014) Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, **29**, 97–106.

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P. & Roy, D.B. (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, **5**, 1052–1060.

Isbell, F., Gonzalez, A., Loreau, M., Cowles, J., Díaz, S., Hector, A., Mace, G.M., Wardle, D.A., O'Connor, M.I., Duffy, J.E., Turnbull, L.A., Thompson, P.L. & Larigauderie, A. (2017) Linking the influence and dependence of people on biodiversity across scales. *Nature*, **546**, 65–72.

Kéry, M. & Royle, J.A. (2008) Hierarchical Bayes estimation of species richness and occupancy in spatially replicated surveys. *Journal of Applied Ecology*, **45**, 589–598.

Leibold, M.A., Holyoak, M., Mouquet, N. & Amarasekare, P. (2004) The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.

Leidenberger, S., Käck, M., Karlsson, B. & Kindvall, O. (2016) The Analysis Portal and the Swedish LifeWatch e-infrastructure for biodiversity research. *Biodiversity Data Journal*, **4**, e7644.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, **5**, 397–406.

R Development Core Team (2014) *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ruete, A., Pärt, T., Berg, Å. & Knape, J. (2017) Exploiting opportunistic observations to estimate changes in seasonal site use: An example with wetland birds. *Ecology and Evolution*, n/a-n/a.
Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M. & Wall, D.H. (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.
Schmeller, D.S., Mihoub, J.-B., Bowser, A., Arvanitidis, C., Costello, M.J., Fernandez, M., Geller, G.N., Hobern, D., Kissling, W.D., Regan, E., Saarenmaa, H., Turak, E. & Isaac, N.J.B. (2017) An operational definition of essential biodiversity variables. *Biodiversity and Conservation*, 1–6.
Southwood, T.R.E., Brown, V.K. & Reader, P.M. (1979) The relationships of plant and insect diversities in succession. *Biological Journal of the Linnean Society*, **12**, 327–348.
Theobald, E.J., Ettinger, A.K., Burgess, H.K., DeBey, L.B., Schmidt, N.R., Froehlich, H.E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M.A. & Parrish, J.K. (2015) Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, **181**, 236–244.
Tulloch, A.I.T., Possingham, H.P., Joseph, L.N., Szabo, J. & Martin, T.G. (2013) Realising the full potential of citizen science monitoring programs. *Biological Conservation*, **165**, 128–138.
Whalen, M.A., Aquilino, K.M. & Stachowicz, J.J. (2016) Grazer diversity interacts with biogenic habitat heterogeneity to accelerate intertidal algal succession. *Ecology*, **97**, 2136–2146.
Zurell, D., Zimmermann, N.E., Sattler, T., Nobis, M.P. & Schröder, B. (2016) Effects of functional traits on the prediction accuracy of species richness models. *Diversity and Distributions*, **22**, 905–917.

Table 1: Number of days present and the temporal continuity of presences (over consecutive or spread days) required for the seven different estimates of species richness over the season.

Continuity	Number of days required to be present			
	1	20	30	45
Spread	S^{1d}	S^{20d}	S^{30d}	S^{45d}
Consecutive (C)		S^{20dC}	S^{30dC}	S^{45dC}

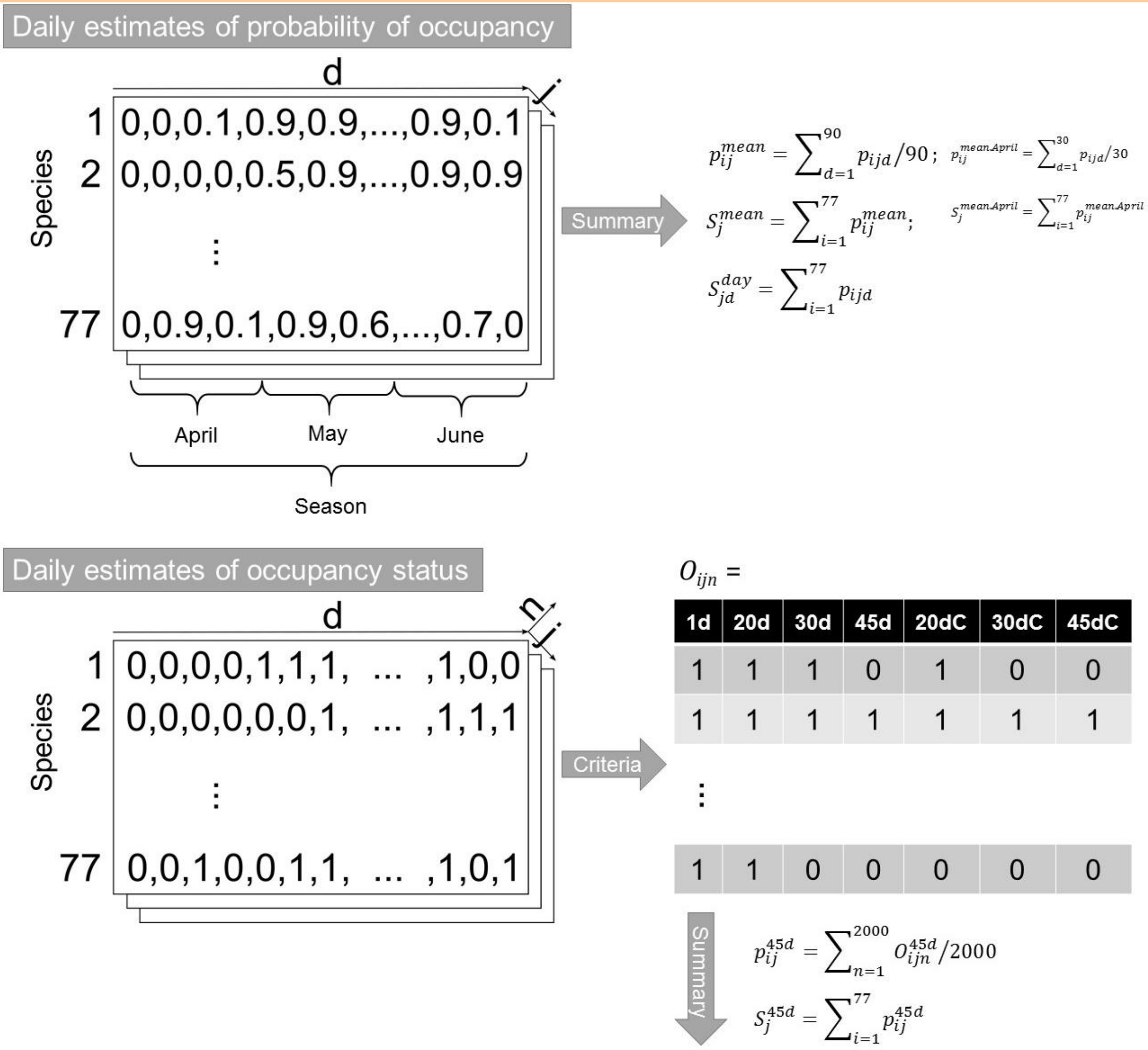

Figure 1. Diagram of the summarizing procedures for daily estimates of occupancy probabilities and status obtained from Ruete *et. al.* (2017). The dimensions described in these summaries are species (*i; I = 77*), days (*d*; *D* = 90 for season or D = 30 for months), sites (*j*; *J* = 107) and monitored MCMC replicates (*n*; *N* = 2000). p_{ij}^{mean} = mean period-specific occupancy probabilities (e.g. 90 days season or April); S_j^{mean} = mean period-specific species richness, S_{jd}^{day} = daily species richness. 20d = 20 spread days criterion, 20dC = 20 consecutive days criterion. O_{ijn} = inclusion parameter, i.e. fulfilment of a given criterion. p_{ij}^{45d} = occupancy probability given the fulfilment of a given criterion (e.g. 45d).

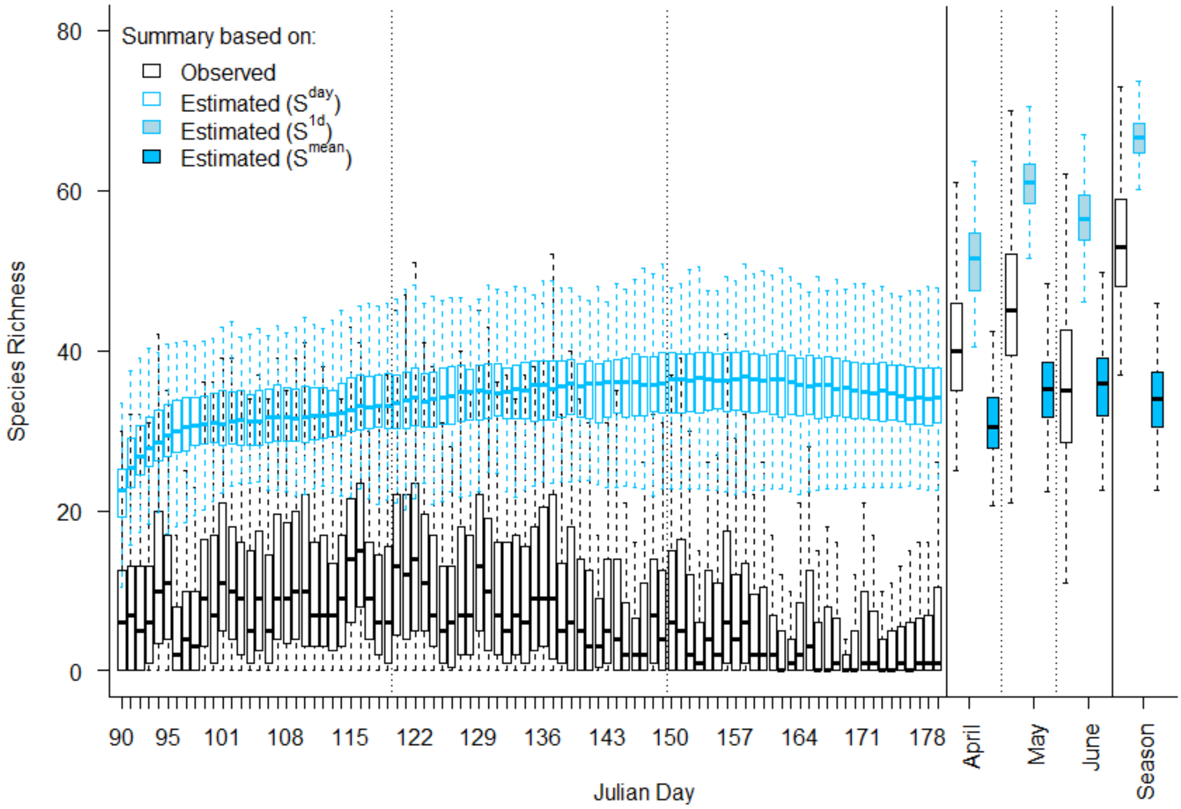

Figure 2. Observed and estimated (daily S^{day}, mean period S^{mean} *and criterion-based* S^{1d}) species richness. Boxplots summarize species richness across all (wetland) sites. Dashed vertical lines divide months.

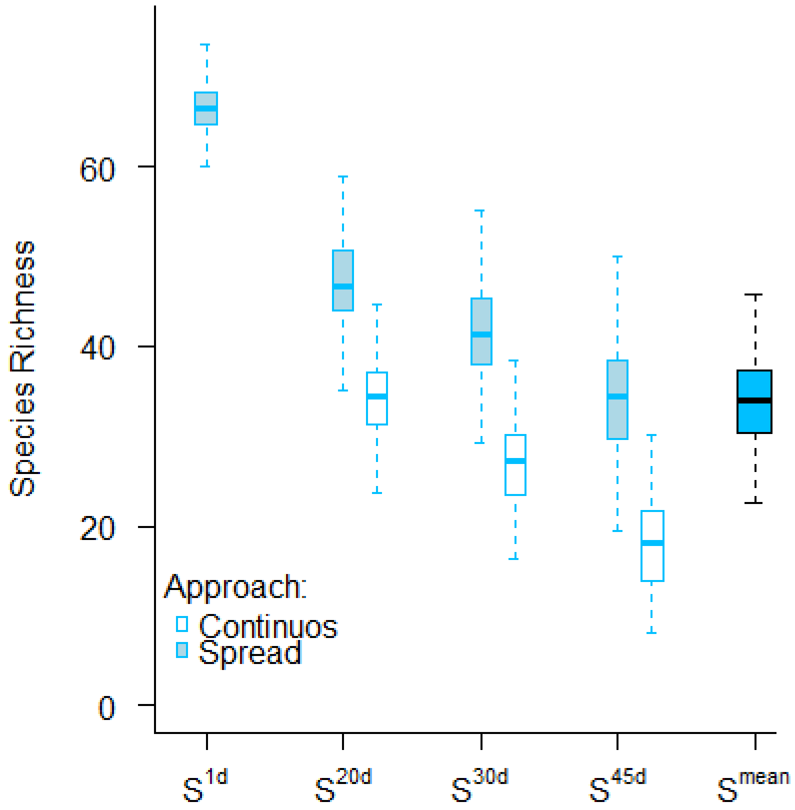

Figure 3. Species richness estimates for the 2014 breeding season (April - June), as a function of number of days and spread of these days (continuous vs spread days) required for inclusion of a species in the richness estimates. The seasonal mean richness estimate (S^{mean}) is included for comparison.

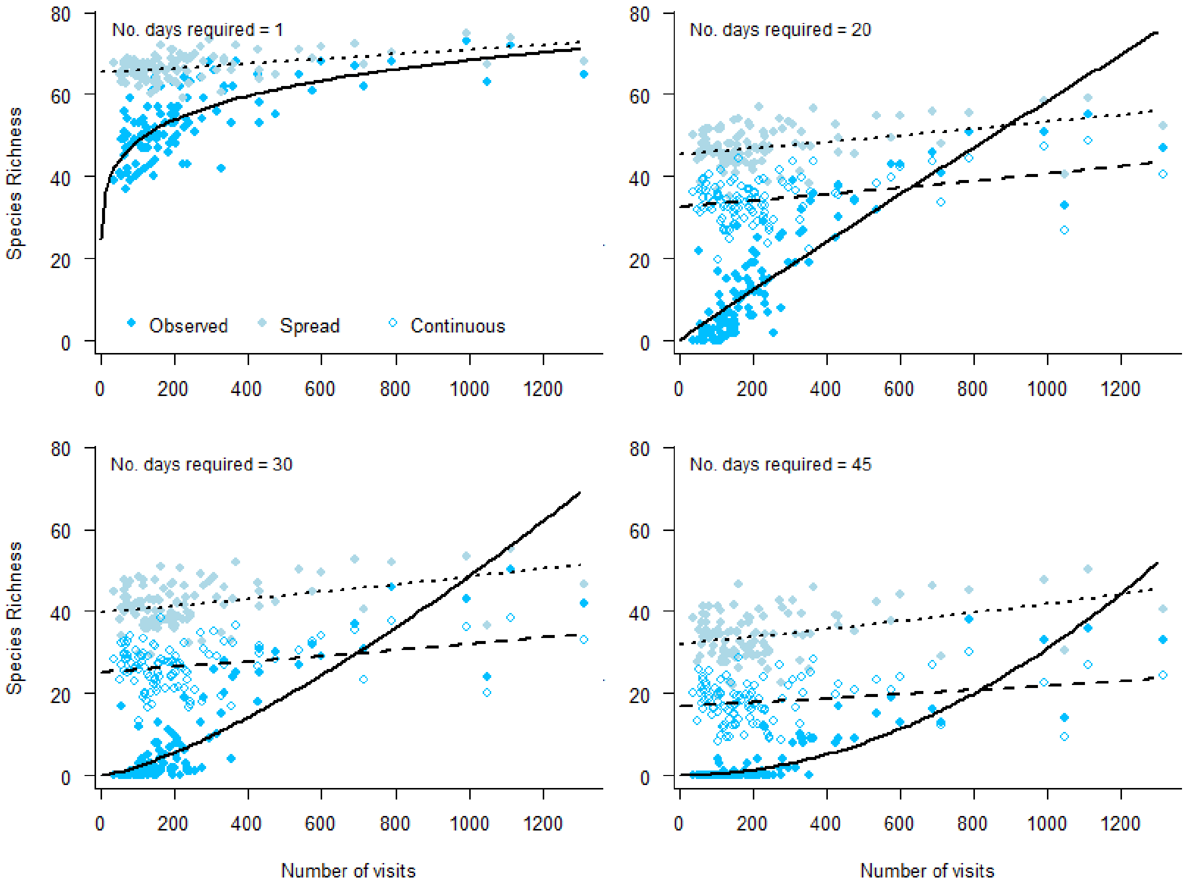

Figure 4. Effect of number of visits (sampling effort) on observed and estimated richness for different number of required days during April to June (1, 20, 30 or 45 days, either spread $S^{1d - 45d}$, or continuous $S^{20dC - 45dC}$) for species classified to be present at a site. Solid, dotted and dashed lines show the effect of number of visits on observed species richness, and estimated species richness under the spread and continuous estimates of local species richness, respectively.