Supporting Material

Extracting robust trends in species' distributions from unstructured opportunistic data: a comparison of methods

Nick J.B. Isaac, Arco J. van Strien, Tom A. August, Marnix P. de Zeeuw & David B. Roy

Appendix S1

1. An empirical basis for the simulation structure
2. Summary statistics on simulated datasets
3. Statistical description of the methods compared by simulation
4. Detailed results of the simulation study
5. References

Note: Instructions for reproducing the simulation results are in a separate document (Appendix S2)

1. An empirical basis for the simulation structure

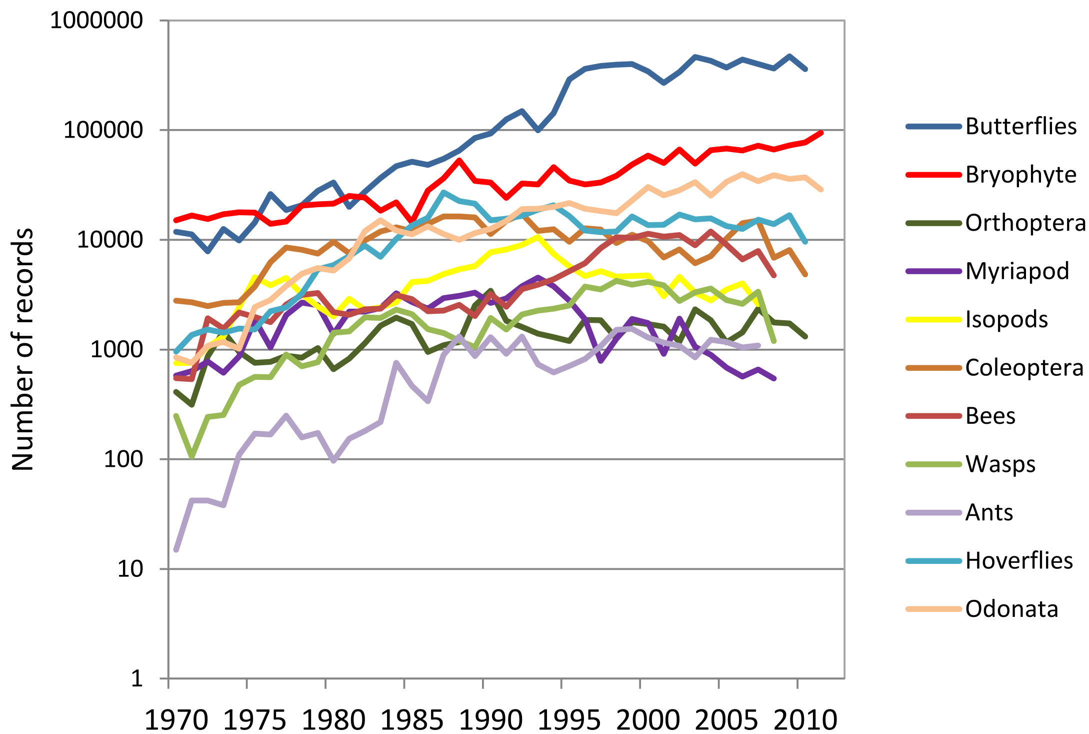
The sampling approach used in our simulations was designed to reflect the structure of real datasets collected by volunteer naturalists. Here we present the major patterns evident within opportunistic datasets in the UK and the Netherlands.

1.1 Biases in opportunistic data

In the manuscript we describe four forms of bias in opportunistic data: 1) uneven recording intensity over time, measured as the number of visits per year, 2) uneven spatial coverage, 3) uneven sampling effort per visit, and 4) uneven detectability. Here we present data from the UK recording schemes from 1970-2010 to illustrate the first three of these forms of bias.

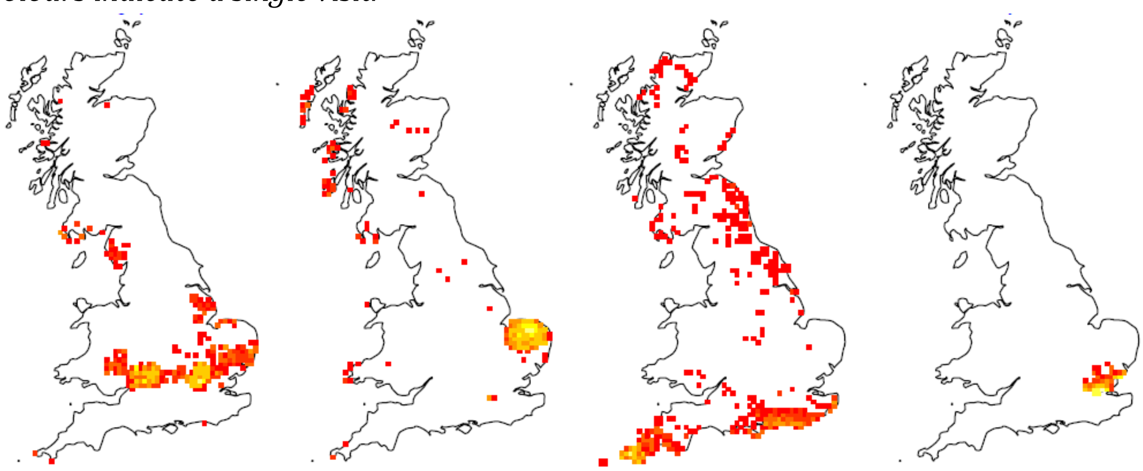
Uneven sampling over time is the best-known form of bias. The number of records being generated has increased markedly in recent years, and for many groups is growing exponentially (figure S1). Thus, for our *MoreVisits* scenario we simulated a doubling in the number of visits over a ten year period.

Figure S1: Number of records per year for 11 taxonomic groups in the UK, 1970-2010



Uneven spatial coverage occurs because most recorders tend to submit records within a well-defined geographic area (figure S2), and because a small number of very prolific recorders generate a disproportionate amount of the total data. The four most prolific Orthoptera recorders (of which there are over 2000) generated 14% of the total visits; half the visits were from just 38. The patterns evident in Orthoptera are typical of the datasets in figure S1. Moreover, since recorders are active at different times, the spatial intensity of recording is uneven over time. Our *MoreVisits+Bias* scenario attempted to capture this phenomenon.

Figure S2: Recording footprints of the four most prolific recorders of Orthoptera in Great Britain, 1970-2010. Yellow colours indicate grid cells with the highest number of visits; red colours indicate a single visit.



Uneven sampling effort reflects two phenomena in opportunistic data: one is the fact that surveys are highly variable in duration; the other is that an unknown proportion of visits are not surveys at all, but rather a collection of 'incidental records'. No direct information on this phenomenon, but we can infer sampling effort indirectly from the list length, L (the number of species recorded on a visit). For the British datasets we examined, at least 40% of all species lists contain just a single species (figure S3), which are unlikely to be the subject of thorough searching. Similar values are apparent in the dataset of Dutch Odonata (van Strien et al, 2010). For this reason, visits in our simulations do not automatically produce records for all species that are actually present.

For some groups, the proportion of single species lists varies markedly across time: the proportion of visits with L=1 has increased for isopods and Coleoptera, but has decreased for hoverflies and remained relatively constant for bryophytes (figure S4). Our *LessEffortPerVisit* scenario was therefore designed to simulate this pattern of systematic trends in survey effort per visit.
Figure S3: The distribution of list lengths among 11 datasets in the UK, 1970-2010. A visit is defined as a unique combination of date and grid reference. Taxa are ordered from highest to lowest proportions of single species visits.

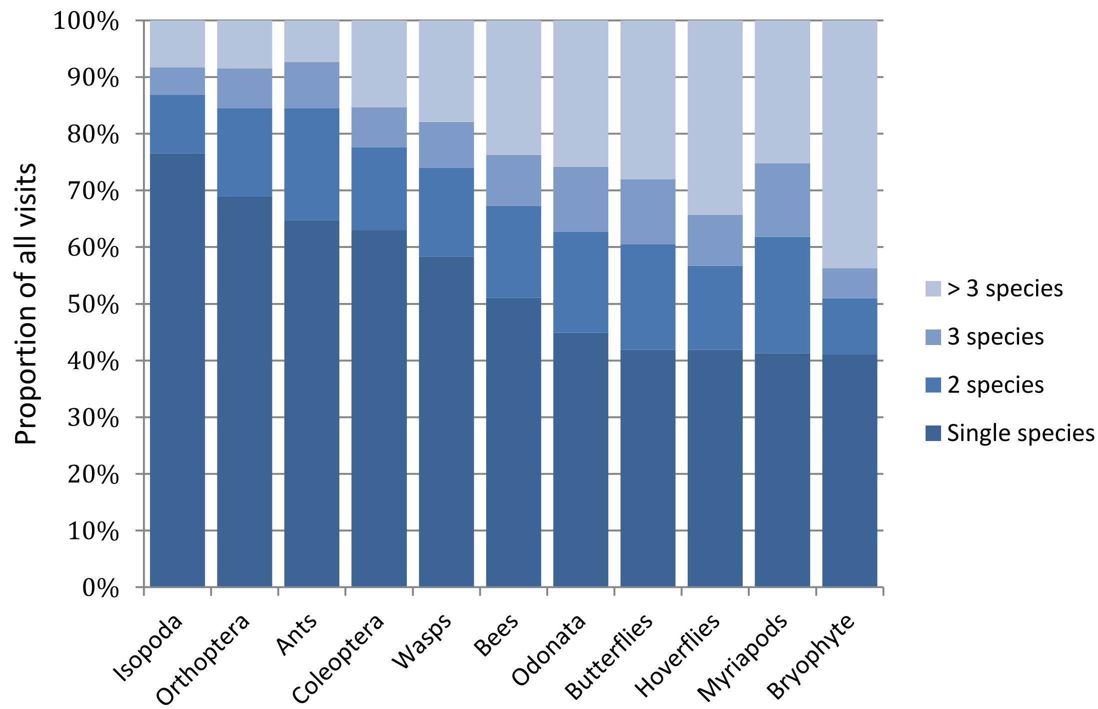
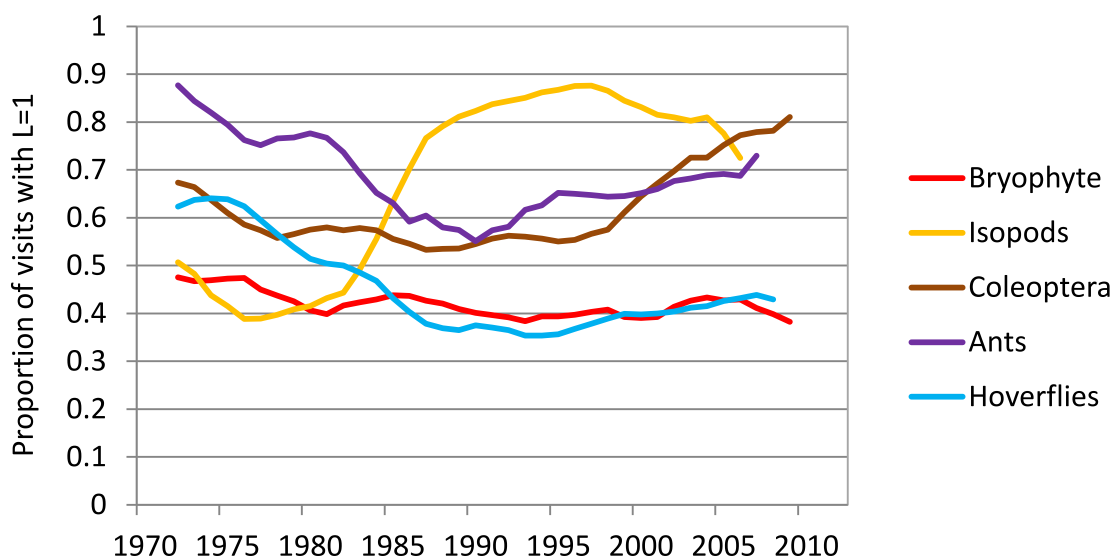


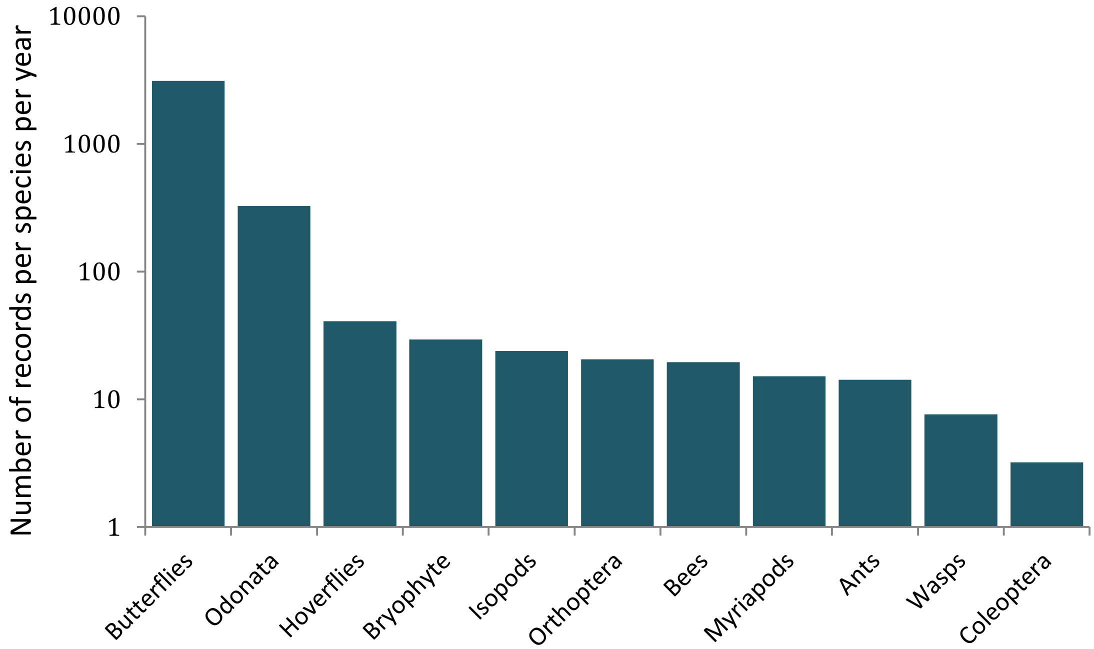
Figure S4: the proportion of visits returning a single species for five taxonomic groups in the UK, 1970-2010. Lines are plotted as moving average over three years.



1.2 Recording intensity

Real datasets differ markedly in overall recording intensity among taxa (figure S5).

Figure S5: Recording rates among 11 opportunistic datasets in the UK, 1970-2010. Numbers are expressed as records per species per year, and are directly comparable with numbers in tables S2-S4.



In addition, recording intensity varies markedly in space. The distribution of visits among sites is highly uneven: some sites receive many visits (figure S6). Capturing this pattern is important for all the methods that use visit-based information to estimate trends. It's especially important for Occupancy-Detection models, in which only sites with repeat visits contribute to the estimated detection probability.

Across 13 British and Dutch opportunistic datasets, the distribution of visits among sites each year is characterised by a power law (figure S6). We expressed the number of sites (defined as $1km^2$) as a proportion of the total that received any visits between 1990-2010, in order that the patterns are directly comparable with our simulated datasets of just 1000 sites. We then modelled the probability of a site receiving *n* visits as $a.n^b$, where *a* defines the proportion of sites receiving a single visit (in our simulation code, this parameter is known as *pSVS*, which stands for the proportion of single visit sites).

Figure S6: The frequency distribution of visits among sites among 11 opportunistic datasets in the UK and the Netherlands for the year 2000. A site is defined here as a $1km^2$ grid cell. Note log-log axes.

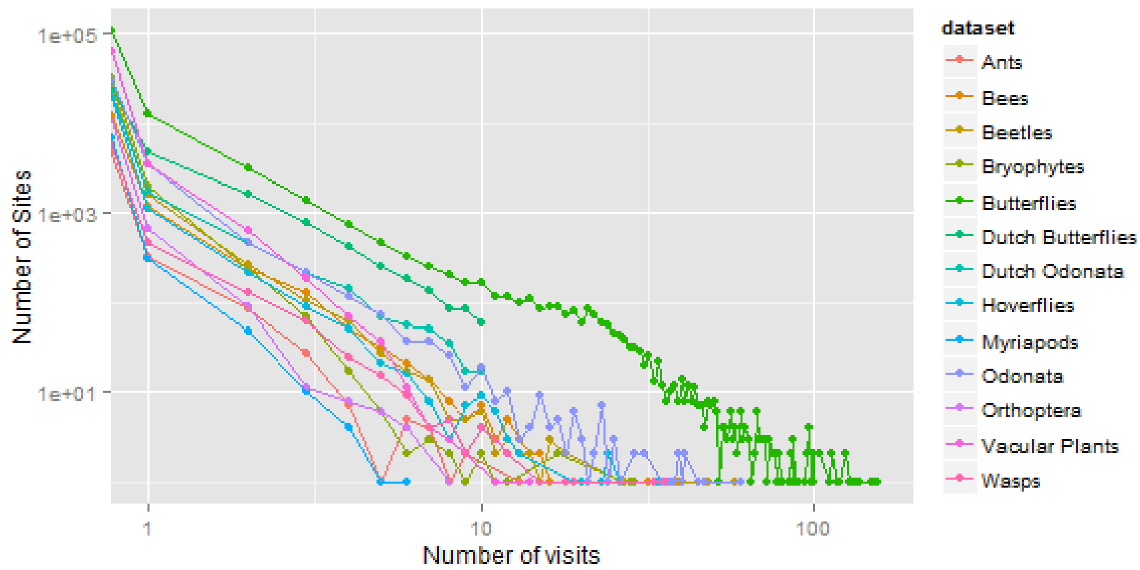


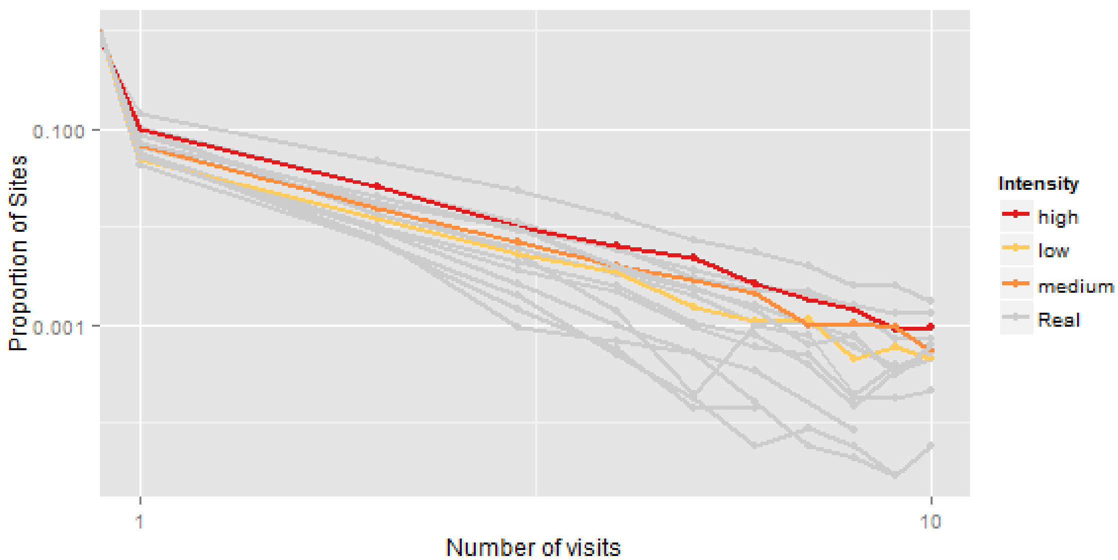
Table S1: Summary statistics for the power law distribution of visits among sites for 13 British and Dutch datasets, using records gathered between 1990-2010. We excluded the small number of sites that received >10 visits in any one year.

	a	b	Total Sites
Ants	0.042	-2.491	5049
Bees	0.051	-2.370	13679
Beetles	0.053	-2.575	28500
Bryophytes	0.049	-3.445	34263
Butterflies	0.085	-2.007	126969
Dutch Butterflies	0.136	-1.771	33822
Dutch Odonata	0.087	-2.022	23753
Hoverflies	0.048	-2.365	20444
Myriapods	0.025	-2.689	7067
Odonata	0.070	-2.259	35099
Orthoptera	0.034	-2.929	11723
Vascular Plants	0.052	-3.011	68316
Wasps	0.058	-2.304	6357

Parameters for the power law for real datasets are shown in table S1. The exponent, b, is remarkably consistent across datasets (median=-2.37, mean=-2.48, sd=0.457). We selected a value of -2 for all our simulations, since this corresponds to a situation in which the number of sites receiving n visits is 4 times greater than the number receiving $2n$ visits. With a single value of b, parameter a is a straightforward measure of overall sampling intensity. Our realised datasets have values of a between 0.025 and 0.136, but most are in the range 0.05-0.07 (median=0.0521, mean=0.0607, sd=0.0287). We therefore selected three values of a for our simulations: 0.05 (low intensity), 0.07 (medium) and 0.1 (high).

Our simulated datasets do a reasonable job of capturing the distribution of visits among sites, especially for those receiving 1, 2 and 3 visits (figure S7). Thus our high intensity simulated datasets resemble those for Butterflies, medium intensity datasets resemble Odonata, and low intensity datasets resemble those of ants, bees, beetles, bryophytes, hoverflies, Orthoptera and wasps.

Figure S7: The frequency distribution of visits among sites among real and simulated datasets. Real datasets are as shown in figure S6, but replotted on a proportional scale. Simulated data come from a sample of 20 years sampling under the Control scenario.



2. Summary statistics on simulated datasets

Table S2: Summary statistics under low recording intensity. Numbers are mean values across 1000 simulated datasets. Focal commonness refers to the rank frequency of the focal species (out of 26).

	Control	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Sites per year	77.3	60.0	63.6	77.3	77.3	77.4
List Length/visit	6.87	6.87	6.91	2.88	6.85	6.32
Records/species/year	38.6	29.9	30.1	16.2	38.5	35.6
Total records	10042.5	7784.2	7819.7	4215.9	10007.2	9250.4
Focal records	381.7	293.7	323.9	160.5	346.4	382.2
Focal commonness	10.5	10.7	9.5	10.5	11.8	9.4

Table S3: Summary statistics under medium recording intensity. Details as in table S2.

	Control	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Sites per year	108.3	83.9	89	108.3	108.3	108.4
List Length/visit	6.89	6.89	6.92	2.89	6.86	6.34
Records/species/year	54.2	42.0	42.2	22.7	54.0	50.0
Total records	14099.9	10924.2	10976.2	5914.7	14050.3	12998.8
Focal records	535.0	411.0	454.0	224.7	485.5	536.6
Focal commonness	10.7	10.8	9.6	10.6	11.8	9.5

Table S4: Summary statistics under high recording intensity. Details as in table S2.

	Control	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Sites per year	154.8	119.9	127.3	154.8	154.8	154.8
List Length/visit	6.91	6.91	6.94	2.89	6.88	6.36
Records/species/year	77.8	60.3	60.6	32.6	77.5	71.6
Total records	20225.4	15668.1	15743.3	8478.6	20154.7	18613.8
Focal records	764.1	587	648.7	320.9	693.3	765.0
Focal commonness	10.7	10.8	9.6	10.6	11.9	9.6

3. Statistical description of the methods compared by simulation

Here we define the details and mathematical notation for each of the methods used. This includes information the *Grain size* (the basic unit, defining how many rows of data are in the analysis), the *Response variable* in the model, the conceptual quantity that is being modelled (‘*what is being modelled*’), the class of *Statistics* used (e.g. GLM) and any other relevant details.

In each simulation, we tested the null hypothesis of no change in the distribution of the focal species. For *Telfer* and *Frescalo_P*, there are specific tests for null hypothesis that are described below. For all other methods the null hypothesis was tested using the estimate of parameter *b* and its standard error (or Bayesian credible intervals).

y: the response variable (defined separately for each model)
T: vector of years (or time periods)
L: the list length (the number of species per visit), or a vector of list lengths
a: intercept in a linear model
b: slope term from a linear model. Generally refers to the annual trend in the response variable
i: species identity
j: site identity
t: identifies the year (or time period)
u: random effect of mean zero, whose variance is estimated from the data.
v: visit identity
z: occupancy status of site *j* (present or absent)

Figure S8 (on page 13) is a conceptual diagram showing how seven of the eleven methods are related to one another.

3.1 Naïve model

Grain size: year
Response variable: number of sites on which the focal species was recorded in year t.
What is being modelled: a trend in the expected number of sites.
Statistics: Poisson generalised linear model (GLM).
Equation: $y_t \sim \text{Poisson}(\lambda_t)$; $\log(\lambda_t) = a + b.\text{T}$
Details: λ is the intensity of the Poisson distribution

3.2 Telfer’s Index

Grain size: species

Response variable: The proportion of sites on which the species i was recorded in the second time period, p_{2i}.

What is being modelled: A change in the relative distribution of species *i*.

Statistics: Logistic GLM

Equation: $\text{logit}(p_{2i}) – a - m.\text{logit}(p_{1i})$

Details: The equation above defines the unstandardised residual from a double logistic regression of the proportion (p_{1i}, p_{2i}) of sites that are occupied by the focal species in two time periods (only sites with records in both periods are considered). In fact, the index is based on the standardised residual (Telfer, Preston, & Rothery, 2002), so the two-tailed p-value can be calculated by assuming these residuals are drawn from a standard normal distribution. Parameter *m* defines the slope of the interspecific relationship in relative distribution between time periods: if total recording intensity is higher in the second time period then $m > 1$ and/or $a > 0$.

3.3 Frescalo_P

Grain size: time-period

Response variable: Not applicable

What is being modelled: A change in the reporting rate of species *i*, r_{it}, relative to the change in reporting rate of benchmark species.

Statistics: Z-test on the difference in reporting rates.

Equation: $Z = (r_{i2} – r_{i1}) / (\sigma_{i1}^2 + \sigma_{i2}^2)$

Details: The Frescalo method was implemented as described in Hill (2012), with Φ= 0.74, R^*=0.27 and a neighbourhood size of 100 cells. We used the same set of neighbourhood weights for all our simulated datasets: the file containing these weights is included in the data on our Github repository. The weights were calculated using the `dist` function in R, from a species occurrence matrix generated by the same rules used in our simulation. We used two equal time periods, in order to produce estimates of r_{it} and associated standard deviations (σ_{it}). The null hypothesis is rejected if $|Z| > 1.96$. Note that r_{it} is referred to in Hill (2012) as a 'time factor', but we prefer the more informative term 'relative reporting rate' (following Fox *et al.*, 2014).

3.4 Frescalo_Y

Grain size: year

Response variable: The reporting rate of the focal species, relative to that of benchmark species, r_{it}.

What is being modelled: A trend in the relative reporting rate of species *i*.

Statistics: Quasi-binomial GLM

Equation: $y_t \sim \text{Bin}(1, p_t)$; $\text{logit}(p_t) = a + b.T$

Details: As Frescalo_P, except that the number of time periods was set equal to the number of years (i.e. 10). This produced estimates of r_{it} for each year, which we ran through a GLM.

3.5 Reporting Rate

Grain size: year

Response variable: A binomial response ('successes' and 'failures') for each combination. Successes are the number of visits on which the focal species was recorded and failures are the number of visits on which it was not recorded.

What is being modelled: a trend in the probability, p, of being recorded on the average visit.

Statistics: Binomial GLM

Equation: $y_t \sim \text{Bin}(N_t, p_t)$; $\text{logit}(p_t) = a + b.\text{T}$

Details: p defines the probability of being recorded per visit; N is the number of visits in year *t*.

3.6 RR+SF

Identical to *ReportingRate*, except that the data were first filtered to exclude sites that were visited in just one year.

3.7 RR+Site

Grain size: site:year combination

Response variable: A binomial response ('successes' and 'failures') for each combination. Successes are the number of visits on which the focal species was recorded and failures are the number of visits on which it was not recorded.

What is being modelled: a trend in the probability of being recorded on the average visit.

Statistics: Binomial GLMM

Equation: $y_{jt} \sim \text{Bin}(N_{jt}, p_{jt})$; $\text{logit}(p_{jt}) = a + b.\text{T} + u_j$

Details: p defines the probability of being recorded per visit; N is the number of visits to site *j* in year *t*.

3.8 RR+LL

Grain size: visit

Response variable: Logical (was the focal species recorded or not).

What is being modelled: a trend in the probability, p, of being recorded on the average visit.

Statistics: Logistic GLM

Equation: $y_{jtv} \sim \text{Bernoulli}(p_{jt})$; $\text{logit}(p_{jt}) = a + b.\text{T} + c.\log(L_{jtv})$

Details: L is the list length and parameter *c* estimates how the probability of being observed scales with sampling effort (a measure of species' detectability).

3.9 RR+SF+LL+ Site

Grain size: visit.

Response variable: Logical (was the focal species recorded or not).

What is being modelled: a trend in the probability, p, of being recorded on the average visit.

Statistics: Logistic GLMM

Equation: $y_{jtv} \sim \text{Bernoulli}(p_{jt})$; $\text{logit}(p_{jt}) = a + b.\text{T} + c.\log(L_{jtv}) + u_j$

Details: L is the list length and parameter *c* estimates how the probability of being observed scales with sampling effort (a measure of species' detectability). As with *RR+SF*, the data were first filtered to exclude sites that were visited in just one year.

3.10 OccDetSimple

Grain size: visit.

Response variable: Logical (was the focal species recorded or not), conditional on being present.

What is being modelled: A trend in the probability ψ_{jt} that an average site is occupied

Statistics: Hierarchical Bayesian

Equations: State process: $z_{jt} \sim \text{Bernoulli}(\psi_{jt})$; $\text{logit}(\psi_{jt}) = b_t$

Observation process: $y_{jtv}|z_{jt} \sim \text{Bernoulli}(z_{jt} * p_{jtv})$; $\text{logit}(p_{jtv}) = a_t$

Details: The model consists of two hierarchically coupled submodels, one governing the true state of sites (presence-absence) and the other governing the observations (detection-nondetection), in which p_{ijv} is the conditional probability of detection when present. Both state and observation processes are modelled as binary variables. Subscript *t* on parameters *a* and *b* indicates that year was modelled as a categorical variable, rather than continuous. The linear trend in occupancy was estimated using a two-step process: first extracting the posterior distributions of b_t for each year, then estimating a trend through those estimates (with full error propagation), as implemented in van Strien *et al.* (2013). This procedure has the advantage that any missing values in the data were imputed in a subtle way, namely from year to year changes rather than from a linear trend across all years as when linear trends were estimated in one step. Especially when changes from one year to another are not gradual, our trend slope procedure may describe the data better. In our case, however, simulated trends were either stable or gradual, so the results are similar to what we would have produced if linear trends were estimated in one step. BUGS code used to fit this model, including the prior distribution, is available at https://github.com/BiologicalRecordsCentre/RangeChangeSims in a file named 'Occ_Simple.bugs'.

3.11 OD+SF+LL+ Site

Grain size: visit.

Response variable: Logical (was the focal species recorded or not), conditional on being present.

What is being modelled: A trend in the probability ψ_{jt} that an average site is occupied.

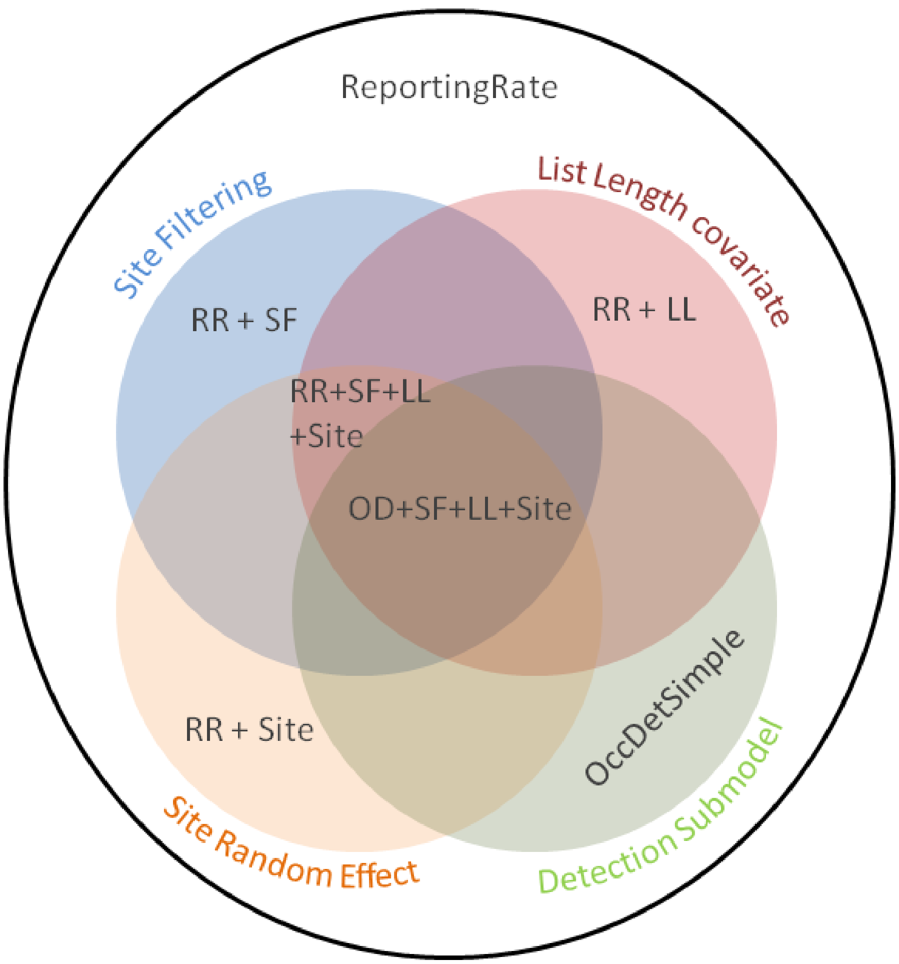
Statistics: Hierarchical Bayesian

Equations: State process: $z_{jt} \sim \text{Bernoulli}(\psi_{jt})$; $\text{logit}(\psi_{jt}) = b_t + u_j$

Observation process: $y_{jtv}|z_{jt} \sim \text{Bernoulli}(z_{jt} * p_{jtv})$; $\text{logit}(p_{jtv}) = a_t + c.\log(L_{jtv})$

Details: Differs from *OccDetSimple* in three ways: 1) the addition of a site-level random effect (u_j) on the state model, 2) the addition a list length coefficient on the observation model (as in *RR+LL* and *RR+SF+LL+Site*), and 3) the data were first filtered to exclude sites that were visited in just one year. The file containing BUGS code used to fit this model is called 'Occ_LL_Site.bugs'.

Figure S8: Conceptual framework of statistical approaches for dealing with variation in recorder activity in opportunistic data, showing how six of the methods in this study are generalisations of the simple ReportingRate model. Coloured words refer to four 'components' described in the main text for circumventing bias in the data. Words in black text are the names of methods used in this study. Note that the Naïve model, Telfer's method, Frescalo_P and Frescalo_Y fall outside this framework.



4. **Detailed results of the simulation study**

Table S5: Type I error rates under low intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.048	0.747	0.262	0.292	0.089	0.040
Telfer	0.02	0.019	0.043	0.020	0.058	0.019
Frescalo_P	0.027	0.018	0.067	0.048	0.062	0.063
Frescalo_Y	0.052	0.046	0.124	0.083	0.099	0.092
ReportingRate	0.114	0.128	0.245	0.596	0.347	0.119
RR+SF	0.103	0.125	0.159	0.395	0.208	0.112
RR+LL	0.115	0.135	0.241	0.075	0.339	0.241
RR+Site	0.057	0.058	0.171	0.472	0.252	0.059
RR+SF+LL+Site	0.051	0.070	0.092	0.050	0.166	0.159
OccDetSimple	0.059	0.056	0.340	0.033	0.066	0.062
OD+SF+LL+Site	0.012	0.020	0.036	0.052	0.026	0.021

Table S6: Type II error rates under low intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.686	1.000	0.984	0.345	0.880	0.700
Telfer	0.759	0.813	0.573	0.902	0.881	0.980
Frescalo_P	0.726	0.807	0.543	0.758	0.894	0.840
Frescalo_Y	0.661	0.715	0.489	0.692	0.851	0.791
ReportingRate	0.429	0.520	0.201	0.129	0.809	0.418
RR+SF	0.616	0.695	0.466	0.372	0.857	0.621
RR+LL	0.448	0.536	0.213	0.740	0.810	0.747
RR+Site	0.484	0.625	0.255	0.244	0.868	0.502
RR+SF+LL+Site	0.641	0.743	0.592	0.882	0.884	0.840
OccDetSimple	0.567	0.601	0.186	0.904	0.691	0.562
OD+SF+LL+Site	0.661	0.775	0.550	0.83	0.797	0.670

Table S7: Results from the test of power under low intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.266	0	0	0.363	0.031	0.260
Telfer	0.221	0.168	0.384	0.078	0.061	0.001
Frescalo_P	0.247	0.175	0.39	0.194	0.044	0.097
Frescalo_Y	0.287	0.239	0.387	0.225	0.050	0.117
ReportingRate	0.457	0.352	0.554	0.275	0	0.463
RR+SF	0.281	0.180	0.375	0.233	0	0.267
RR+LL	0.437	0.329	0.546	0.185	0	0.012
RR+Site	0.459	0.317	0.574	0.284	0	0.439
RR+SF+LL+Site	0.308	0.187	0.316	0.068	0	0.001
OccDetSimple	0.374	0.343	0.474	0.063	0.243	0.376
OD+SF+LL+Site	0.327	0.205	0.414	0.118	0.177	0.309

Table S8: Type I error rates under medium intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.035	0.889	0.364	0.367	0.103	0.047
Telfer	0.019	0.019	0.051	0.011	0.067	0.015
Frescalo_P	0.017	0.017	0.070	0.050	0.065	0.068
Frescalo_Y	0.046	0.051	0.152	0.068	0.133	0.117
ReportingRate	0.114	0.125	0.333	0.727	0.421	0.126
RR+SF	0.100	0.114	0.202	0.516	0.313	0.111
RR+LL	0.117	0.135	0.315	0.057	0.425	0.293
RR+Site	0.054	0.06	0.213	0.616	0.367	0.067
RR+SF+LL+Site	0.051	0.058	0.08	0.033	0.292	0.229
OccDetSimple	0.055	0.066	0.429	0.034	0.069	0.060
OD+SF+LL+Site	0.009	0.014	0.036	0.079	0.031	0.015

Table S9: Type II error rates under medium intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.605	1.000	0.971	0.196	0.841	0.591
Telfer	0.662	0.750	0.457	0.864	0.835	0.984
Frescalo_P	0.642	0.760	0.409	0.682	0.873	0.820
Frescalo_Y	0.563	0.667	0.318	0.591	0.815	0.724
ReportingRate	0.354	0.440	0.102	0.035	0.792	0.334
RR+SF	0.475	0.591	0.275	0.147	0.819	0.488
RR+LL	0.355	0.449	0.129	0.666	0.803	0.704
RR+Site	0.331	0.491	0.150	0.071	0.819	0.333
RR+SF+LL+Site	0.447	0.595	0.385	0.812	0.859	0.780
OccDetSimple	0.495	0.539	0.076	0.854	0.658	0.440
OD+SF+LL+Site	0.426	0.604	0.290	0.672	0.660	0.428

Table S10: Results from the test of power under medium intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.360	0	0	0.437	0.056	0.362
Telfer	0.319	0.231	0.492	0.125	0.098	0.001
Frescalo_P	0.341	0.223	0.521	0.268	0.062	0.112
Frescalo_Y	0.391	0.282	0.530	0.341	0.052	0.159
ReportingRate	0.532	0.435	0.565	0.238	0	0.540
RR+SF	0.425	0.295	0.523	0.337	0	0.401
RR+LL	0.528	0.416	0.556	0.277	0	0.003
RR+Site	0.615	0.449	0.637	0.313	0	0.600
RR+SF+LL+Site	0.502	0.347	0.535	0.155	0	0
OccDetSimple	0.450	0.395	0.495	0.112	0.273	0.500
OD+SF+LL+Site	0.565	0.382	0.674	0.249	0.309	0.557

Table S11: Type I error rates under high intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.036	0.964	0.476	0.527	0.124	0.033
Telfer	0.020	0.02	0.078	0.014	0.086	0.013
Frescalo_P	0.022	0.023	0.111	0.045	0.067	0.076
Frescalo_Y	0.060	0.055	0.22	0.104	0.167	0.134
ReportingRate	0.115	0.133	0.408	0.872	0.523	0.087
RR+SF	0.103	0.116	0.283	0.776	0.435	0.088
RR+LL	0.113	0.121	0.386	0.072	0.489	0.396
RR+Site	0.060	0.059	0.200	0.816	0.509	0.046
RR+SF+LL+Site	0.060	0.061	0.108	0.042	0.429	0.342
OccDetSimple	0.040	0.053	0.552	0.068	0.058	0.044
OD+SF+LL+Site	0.014	0.015	0.041	0.114	0.039	0.009

Table S12: Type II error rates under high intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.447	1.000	0.973	0.074	0.778	0.460
Telfer	0.469	0.622	0.254	0.802	0.760	0.987
Frescalo_P	0.490	0.638	0.241	0.598	0.805	0.785
Frescalo_Y	0.394	0.513	0.165	0.476	0.741	0.670
ReportingRate	0.202	0.287	0.026	0.010	0.764	0.211
RR+SF	0.296	0.426	0.095	0.027	0.792	0.297
RR+LL	0.217	0.310	0.038	0.583	0.762	0.647
RR+Site	0.159	0.296	0.073	0.017	0.779	0.180
RR+SF+LL+Site	0.252	0.383	0.191	0.697	0.809	0.699
OccDetSimple	0.278	0.380	0.025	0.814	0.511	0.307
OD+SF+LL+Site	0.139	0.335	0.072	0.445	0.387	0.163

Table S13: Results from the test of power under high intensity recording

Method	*Control*	*MoreVisits*	*MoreVisits +Bias*	*LessEffort PerVisit*	*More Detectable*	*NonFocal Declines*
Naive	0.517	0	0	0.399	0.098	0.507
Telfer	0.511	0.358	0.668	0.184	0.154	0
Frescalo_P	0.488	0.339	0.648	0.357	0.128	0.139
Frescalo_Y	0.546	0.432	0.615	0.42	0.092	0.196
ReportingRate	0.683	0.58	0.566	0.118	0	0.702
RR+SF	0.601	0.458	0.622	0.197	0	0.615
RR+LL	0.670	0.569	0.576	0.345	0	0
RR+Site	0.781	0.645	0.727	0.167	0	0.774
RR+SF+LL+Site	0.688	0.556	0.701	0.261	0	0
OccDetSimple	0.682	0.567	0.423	0.118	0.431	0.649
OD+SF+LL+Site	0.847	0.650	0.887	0.441	0.574	0.828

Figure S9. Results from the test of Validity under all scenarios (note square root scale on y-axis). The x-axis shows low (L), medium (M) and high (H) levels of recording intensity, as defined in the text. The solid and dashed lines indicate α=0.05 and α=0.1 respectively

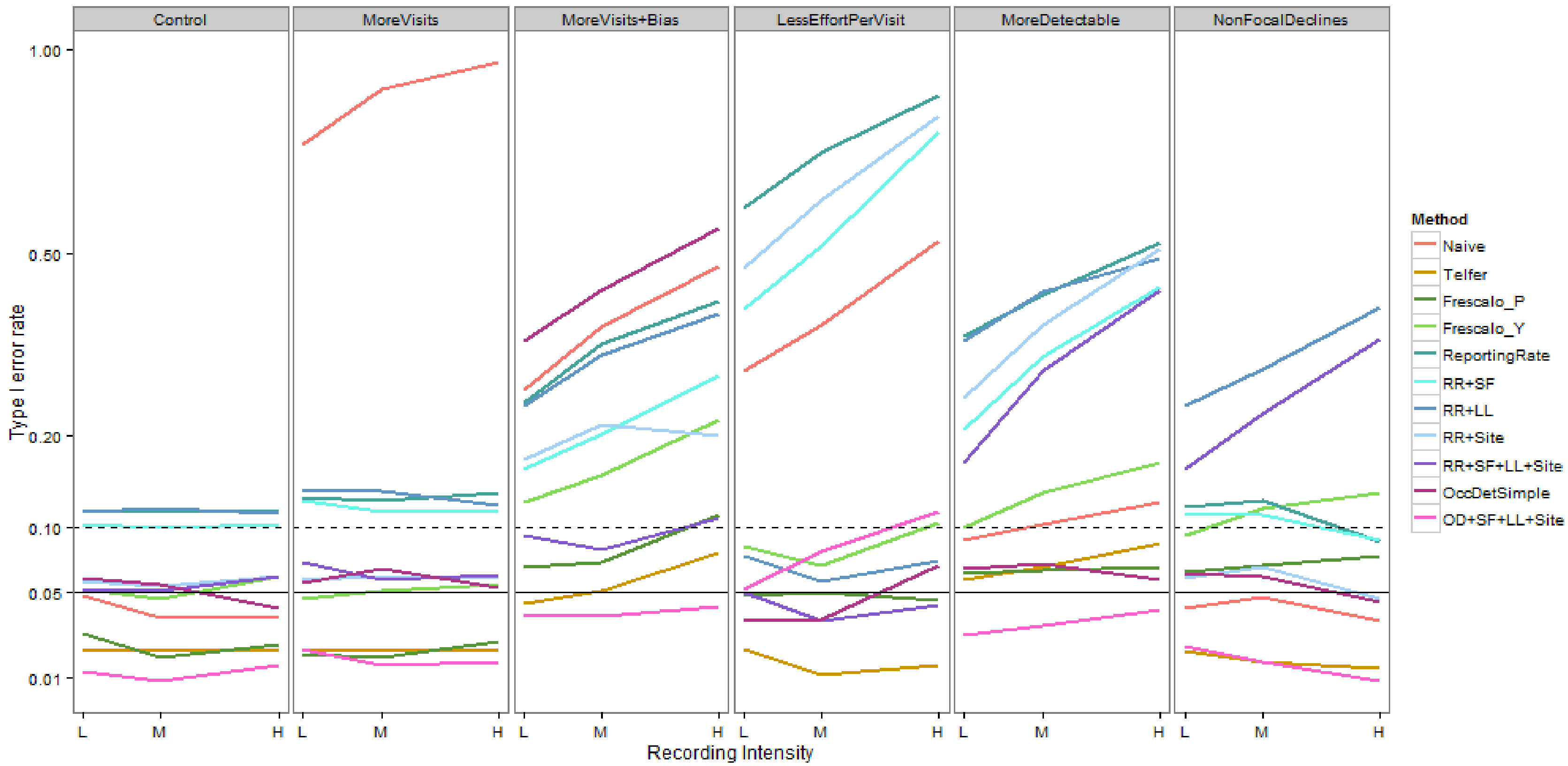
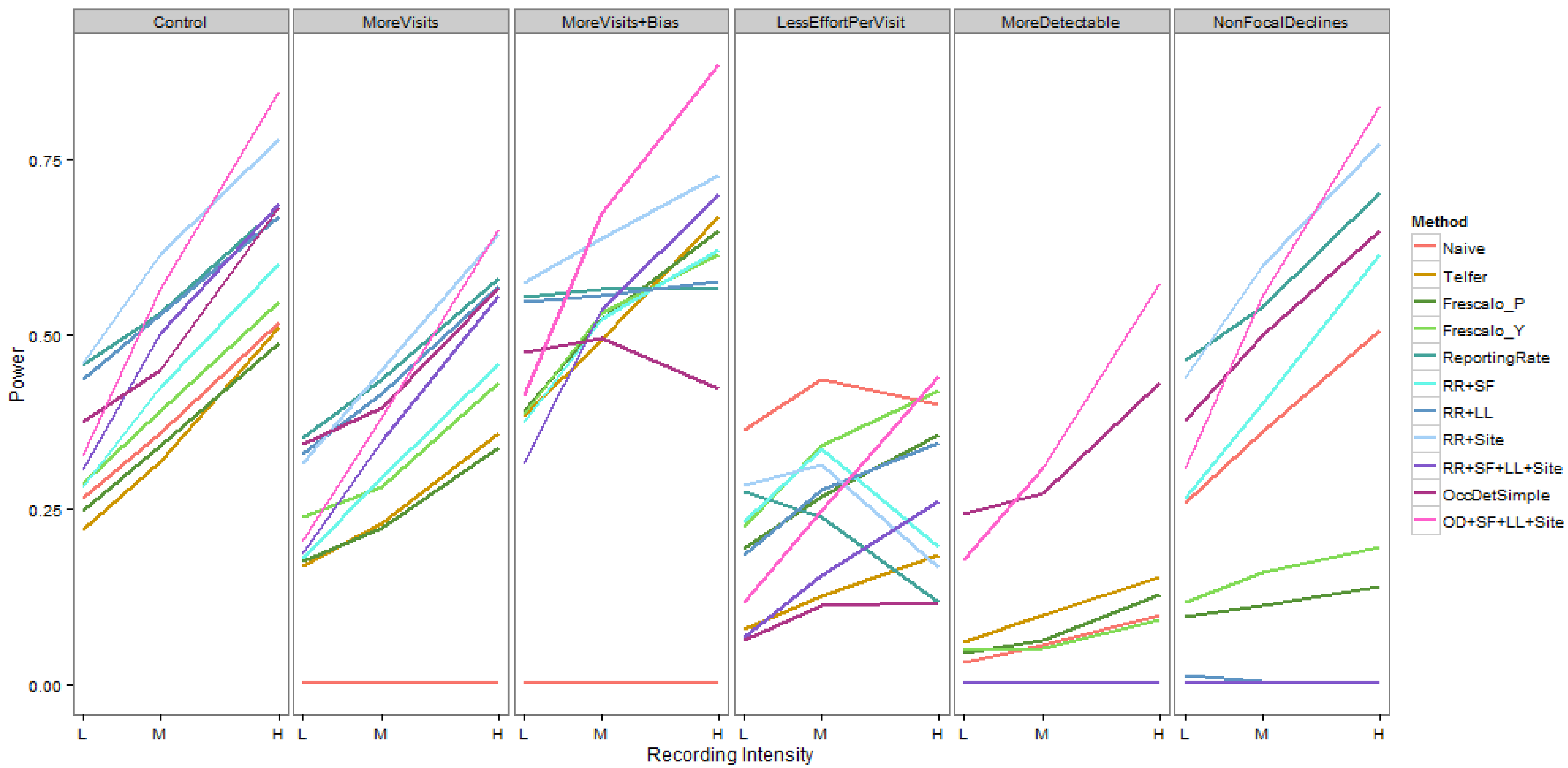


Figure S10. Power of all methods under all scenarios. The x-axis shows low (L), medium (M) and high (H) levels of recording intensity, as defined in the text. Power generally increases with recording intensity, except in cases where the Type I error rate is more sensitive to recording intensity than the Type II error rate.



5. References

Fox, R., Oliver, T. H., Harrower, C., Parsons, M. S., Thomas, C. D., & Roy, D. B. (2014). Long-term changes to the frequency of occurrence of British moths are consistent with opposing and synergistic effects of climate and land-use changes. Journal of Applied Ecology, n/a–n/a. doi:10.1111/1365-2664.12256

Hill, M. O. (2012). Local frequency as a key to interpreting species occurrence data when recording effort is not known. Methods in Ecology and Evolution, 3(1), 195–205. doi:10.1111/j.2041-210X.2011.00146.x

Telfer, M. G., Preston, C. D., & Rothery, P. (2002). A general method for measuring relative change in range size from biological atlas data. Biological Conservation, 107(1), 99–109. doi:10.1016/S0006-3207(02)00050-2

Van Strien, A. J., Termaat, T., Groenendijk, D., Mensing, V., & Kéry, M. (2010). Site-occupancy models may offer new opportunities for dragonfly monitoring based on daily species lists. *Basic and Applied Ecology*, *11*(6), 495–503. doi:10.1016/j.baae.2010.05.003

Van Strien, A. J., van Swaay, C. A. M., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. Journal of Applied Ecology, 50(6), 1450–1458. doi:10.1111/1365-2664.12158