

The impact of data quality filtering of opportunistic citizen science data on species distribution model performance

Camille Van Eupen^{a,b,*}, Dirk Maes^c, Marc Herremans^d, Kristijn R.R. Swinnen^d, Ben Somers^{b,#}, Stijn Luca^{a,#}

^a Ghent University, Department of Data Analysis and Mathematical Modelling, Coupure Links 653, B-9000 Ghent, Belgium

^b KU Leuven, Department of Earth and Environmental Sciences, Division Forest Nature and Landscape, Celestijnenlaan 200E, B-3001 Heverlee, Belgium

^c Research Institute for Nature and Forest (INBO), Species Diversity Group, Herman Teirlinckgebouw, Havenlaan 88, bus 73, B-1000 Brussels, Belgium

^d Natuurpunt Studie, Coxiestraat 11, 2800 Mechelen, Belgium

ARTICLE INFO

Keywords:

Data quality filtering
Maxent
Opportunistic data
Presence-only
Sample size
Species distribution models

ABSTRACT

Opportunistically collected species occurrence data are often used for species distribution models (SDMs) when high-quality data collected through standardized recording protocols are unavailable. While opportunistic data are abundant, uncertainty is usually high, e.g. due to observer effects or a lack of metadata. To increase data quality and improve model performance, we filtered species records based on record attributes that provide information on the observation process or post-entry data validation. Data filtering does not only increase the quality of species records, it simultaneously reduces sample size, a trade-off that remains relatively unexplored. By controlling for sample size in a dataset of 255 species, we were able to explore the combined impact of data quality and sample size on model performance. We applied three data quality filters based on observers' activity, the validation status of a record in the database and the detail of a submitted record, and analyzed changes in AUC, Sensitivity and Specificity using Maxent with and without filtering. The impact of stringent filtering on model performance depended on (1) the quality of the filtered data: records validated as correct and more detailed records lead to higher model performance, (2) the proportional reduction in sample size caused by filtering and the remaining absolute sample size: filters causing small reductions that lead to sample sizes of more than 100 presences generally benefitted model performance and (3) the taxonomic group: plant and dragonfly models benefitted more from data quality filtering compared to bird and butterfly models. Our results also indicate that recommendations for quality filtering depend on the goal of the study, e.g. increasing Sensitivity and/or Specificity. Further research must identify what drives species' sensitivity to data quality. Nonetheless, our study confirms that large quantities of volunteer generated and opportunistically collected data can make a valuable contribution to ecological research and species conservation.

1. Introduction

Appropriate conservation measures must mitigate the alarming declines of biodiversity caused by global pressures such as climate change (Urban et al., 2016), invasive species (Early et al., 2016) and intensifying land use (Newbold et al., 2015). Choosing proper conservation measures requires evidence on the state of biodiversity and species' distributions. Ideally, such evidence is gathered through standardised protocols, performed by trained observers and with a clear description of both data collection and project objectives (Kosmala et al., 2016).

Such highly structured data, however, is rarely available for a wide range of species, nor extensive periods or geographical areas (Urban et al., 2016).

In response, less structured but bulky occurrence data with varying information content, often collected by volunteers participating in citizen science initiatives (Theobald et al., 2015), are being explored for biodiversity conservation purposes (Guisan et al., 2013). The value of data with information on detectability or information on absences is indisputable and their applications are abundant, e.g. for species distribution models (SDMs) (Guisan and Zimmermann, 2000; van Strien

* Corresponding author.

E-mail address: camille.vaneupen@kuleuven.be (C. Van Eupen).

Joint last author.

<https://doi.org/10.1016/j.ecolmodel.2021.109453>

Received 15 September 2020; Received in revised form 9 December 2020; Accepted 16 January 2021

Available online 2 February 2021

0304-3800/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

et al., 2013; Wood et al., 2018) or Red List compilations (e.g. Maes et al., 2015). In contrast, the value of data with little information on the observation process is uncertain and conservation applications are limited (Dobson et al., 2020; Guillera-Aroita et al., 2015). When such unstructured occurrence data consist of occasional observations of species presences, they are termed opportunistic presence-only (Giraud et al., 2016) or presence-background data (Wang and Stone, 2019). They are generally used in SDMs (e.g. Maxent (Phillips et al., 2006) or point process models (Renner et al., 2015)) that contrast available environmental conditions in the study area (the background), with the conditions at locations where the species was observed (Elith et al., 2011).

Using opportunistic presence-only data for SDMs has both advantages and disadvantages. The main advantage is the abundance of the available data, because an easy data collection leads to the coverage of a large number of species over large geographical areas, at a fine scale and over potentially long periods (Kosmala et al., 2016). Online platforms and smartphone applications facilitate an easy recording of species for a volunteer observer, and the number of active observers on data platforms such as iRecord in the United Kingdom (<https://www.brc.ac.uk/irecord/>), *waarnemingen.be* in Flanders (northern Belgium; <https://www.waarnemingen.be>) or iNaturalist worldwide (<https://www.inaturalist.org/>) is indeed growing by the hundreds (e.g. *waarnemingen.be*) or even thousands (e.g. iNaturalist) every year. Since the quantity and extent of this data can never be reached by standardised monitoring schemes, opportunistic data can make a valuable contribution to science if processed correctly (Giraud et al., 2016; Soroye et al., 2018). Two major disadvantages of opportunistic presence-only data limiting their application potential (Dobson et al., 2020; Guillera-Aroita et al., 2015), however, are the incapability of delivering probabilistic model outputs (Guillera-Aroita et al., 2015) and a high risk of bias and error (Bird et al., 2014; Isaac and Pocock, 2015). The awareness of these uncertainties reflects in the scepticism towards data quality of opportunistic observations or citizen science data in general (Burgess et al., 2017), because when disregarded in the modeling or decision-making process, these disadvantages can lead to misguided conservation measures (Isaac et al., 2014).

Different strategies are applied to increase the quality of opportunistic datasets. A first strategy is rather bottom-up, where the underlying protocol of a citizen science project is changed (Kosmala et al., 2016). This requires a regime shift and takes time, but can be fruitful (e.g. eBird - Sullivan et al., 2014). A second and promising strategy is data-integration (Miller et al., 2019), where multiple sources of opportunistic presence-only data are combined (Lin et al., 2017) or presence-only data is treated as complementary to structured presence-absence data (Robinson et al., 2019). A third strategy, integrated into many national citizen science databases, is data validation, where the species' identification is verified, often together with the spatial and temporal plausibility of a record. It is common practice in, for example, eBird (Sullivan et al., 2009), *waarnemingen.be* (Swinnen et al., 2018) and iRecord (<https://www.brc.ac.uk/irecord/records-verified>). However, even with the best experts and state-of-the-art methods (e.g. image recognition), it is challenging to verify thousands of records entering data repositories every day, particularly those without corroborating picture evidence. As a result, many researchers apply a fourth strategy, where data reliability is maximised by data filtering or data cleaning. This can be done by error detection (e.g. Serra-Diaz et al., 2017), outlier removal (e.g. Kallimanis et al., 2017), by filtering in geographical or environmental space (e.g. Varela et al., 2014), or by deleting species records based on data attributes (e.g. Rutten et al., 2019), so-called "stringent filtering" (Steen et al., 2019).

The desired effect of stringent filtering is an increase in quality, by reducing bias and error (Steen et al., 2019). Yet, sample size is inevitably reduced by filtering, and as sample size is known to have a major influence on model performance (Gábor et al., 2019; Wisz et al., 2008), stringent filtering leads to a trade-off between data quality and sample size. To our knowledge, the combined impact of data quality and sample

size in stringent filtering on the performance of SDMs remains under-explored. Studies that explored the impact of stringent data filters found a negligible effect on bird occurrence predictions when retaining only structured survey data (Kamp et al., 2016) or data from observers with higher expertise (Steen et al., 2019). On the other hand, predictions were more accurate when using only records validated as correct for a butterfly genus prone to misidentification (Vantieghem et al., 2017), or by using only eBird checklists of observers who travelled larger distances to do their observations (Steen et al., 2019).

In our study, we will expand on previous findings by applying different quality filters on a regional species occurrence database '*waarnemingen.be*'. The database consists of both structured and unstructured recordings in Flanders since 2008 and currently holds more than 44 million species records and one of the densest collections of species records in Europe (Herremans et al., 2018). Our aim is to identify which quality filters increase the discrimination accuracy of Maxent and to formulate recommendations based on taxonomic group and data characteristics. Every citizen science database is unique and while the considered taxonomic groups in *waarnemingen.be* are blessed with a relatively high proportion of quality data, this might not be the case in all data repositories. The properties of *waarnemingen.be* allowed us to evaluate the impact on model performance of different changes in data quality, for a wide range of changes in sample size. This does not only provide more insight into the trade-off between data quality and sample size in stringent filtering but also ensures the transferability of our results to datasets of lower quality and/or record density.

2. Material and methods

2.1. Dataset and quality filters

We assessed the impact of data quality filtering on opportunistic citizen science data gathered in the Flemish species occurrence database '*waarnemingen.be*'. The dataset contained both "structured data" or observations supported by guidelines or a protocol (varying from standardized monitoring schemes to small project observations), and "unstructured data" or incidental observations. For a detailed description of our data selection and model testing procedure see Section 2.2 and Appendix A. 161,782 structured records were separated for model testing, to measure the performance of the SDMs (see Section 2.3). Another 5,547,750 unstructured records were used for model training. We adopted the ODMAP protocol (v1.0, Zurell et al., 2020) and describe the different steps (Overview, Data, Model, Assessment and Prediction) in Appendix B.

We selected three dichotomous filters as a measure for data quality, based on available metadata (Table A.1). The first filter "ACTIVITY" refers to the annual average number of active recording days of an observer, in the study period 2014-2019. We calculated the individual activity rate of observers, including the observers with the highest number of records first and stopped when we reached the observers that cumulatively collected 80% of the data. The threshold for a high activity rate was set to the first quartile of the activity rate of this group, i.e. 92 recording days in one year. We considered this a proxy for observer experience, known to lead to lower rates of both false-negative and false-positive errors (Farmer et al., 2012; Kallimanis et al., 2017; Kelling et al., 2015). The second filter "DETAIL" reflects whether observers provide information beyond the default date, location and species name, such as species behaviour, photographs or additional comments. Records submitted with more effort are of higher quality, if effort is defined by the 'distance travelled for a checklist' (Steen et al., 2019). Because we applied filters to unstructured data only, we used record detail as a measure for effort instead. The third filter "VALSTAT" is based on the status of a record in the internal validation system of the database, indicating if it was evaluated as correct or as uncertain. Records marked as correct are meant to contain no misidentification errors (e.g. Vantieghem et al., 2017), even though an occasional human or software

error might occur. Records marked as uncertain have either not been validated or were hard to judge correctly, due to lack of additional information (Swinnen et al., 2018).

2.2. Data selection

Records from four well-studied taxonomic groups in Flanders, i.e. birds, butterflies, dragonflies and plants were subjected to some initial data restrictions: (1) records were limited to our study area, the Flemish region of Belgium, (2) observations dated from January 2014 to September 2019, (3) we included only records with sufficient geographical precision (≤ 500 metres), (4) for birds, only birds that breed in Flanders were used (Vermeersch et al., 2020), and (5) we removed absences (zero-counts) and entries validated as incorrect.

After the initial selection, we divided the data into records for model training and records for model testing (also see Appendix A and Fig. A.1). Structured data were used solely for model testing and never for model training, and were further reduced to high-quality testing records. This was done by selecting only structured records that were validated as correct and from observers with a high activity rate. The model training records consisted of unstructured presence-only data, a data type found in many large-scale datasets of opportunistic species records (e.g. GBIF, <https://www.gbif.org>). Model training records were subjected to the three quality filters and their combinations, resulting in seven filtered datasets (Fig. A.2).

Per species, training and testing records were aggregated in a 1×1 km grid, a frequently used resolution in Flemish biodiversity research (e.g. Demolder et al., 2014; Rutten et al., 2019; Vantieghem et al., 2017), resulting in one presence per grid cell per species. This aggregation of records is also known as ‘spatial thinning’ or ‘spatial filtering’, a common technique to reduce spatial bias (Kramer-Schadt et al., 2013) and improve model performance (Boria et al., 2014). The high-quality presences of the model testing set were complemented with absences derived from grid cells with high search effort for the associated taxonomic group, but where the target species was not observed. We kept only species with at least 50 presences in the testing set, and at least one filtered training set with at least 100 presences. This resulted in a dataset of 255 species in four taxonomic groups (full list in Table C.1).

2.3. Species distribution model

We evaluated the impact of stringent filtering on the performance of Maxent (software version 3.4.1, implemented in the R package ‘dismo’ v1.1.4 (Hijmans et al., 2017)). Maxent is a commonly used presence-only algorithm (Elith et al., 2011; Phillips et al., 2006), which models a relative probability of occurrence based on a species’ presence records and background points. Background points are used to define the contrast between what is available in the environment and what is used by the species (Elith et al., 2011). We included all of the 13,552 cells in our study area as background and did not adjust the background selection to correct for sampling bias (e.g. Phillips et al., 2009; Vollerling et al., 2019), to ensure comparability of our models (Merow et al., 2013). Comparability was further supported by allowing only linear, quadratic and product features for every model, by setting a minimum sample size of 100 to ensure that the regularization coefficient was kept to 0.05, and by using identical predictors in all Maxent models.

The predictor set represented the range of environmental conditions in our study area and comprised twelve continuous predictors and two factor variables (see Table C.2 for a summary). We aggregated the land use in Flanders in eleven classes: agriculture, forest, semi-natural grassland, scrub, heathland, saltmarshes, wetlands, dunes, urban areas, water and other green areas (i.e. green areas outside the urban area that are not mapped as agricultural or natural land use) (Poelmans and Van Daele, 2014). The area of these classes in each 1×1 km cell was calculated and cells were removed if the cumulative area of land use was less than 50% of the total area (i.e. cells close to regional borders). We

removed one class ‘agriculture’ from the set because of the relatively high collinearity with other classes and because of the problem with perfect multicollinearity in compositional data (Aichison, 2003). The ten other land use classes were used to describe the variation in the extremely fragmented landscape in Flanders (Antrop, 2004). Two additional continuous predictors were the mean annual temperature and mean annual precipitation, BIO1 and BIO12 from WorldClim2 respectively (Fick and Hijmans, 2017). The first factor variable was a grid cell’s dominant soil texture class (Maréchal and Tavernier, 1974), a direct or indirect influencer of a species’ microclimate (Titeux et al., 2009). The second was ‘Ecoregion’ (Couvreur et al., 2004), which is a region with similar biotic and abiotic conditions. Since Flanders has limited geographical and environmental gradients (e.g. 240 km across, 0 to 288 m elevation and relatively uniform climatic conditions) and species use similar biotopes throughout the region, we assumed that the environmental response of a species was similar across the entire study area (Chen et al., 2020).

2.4. Model evaluation

For model evaluation, we chose three metrics: the area under the receiver operating characteristic (ROC) curve (AUC), Sensitivity (i.e. true positive rate) and Specificity (i.e. true negative rate) (Fielding and Bell, 1997), based on three rationales. First, using AUC alone as a summary metric of the ROC curve would lead to a loss of information about model performance (Jiménez-Valverde, 2012). Second, these metrics are measures of model discrimination and independent of species prevalence, which is unknown in presence-background situations (Lawson et al., 2014). Third, we evaluated our models on an external testing test that contained both presences and absences, enabling a reasonable calculation of the two threshold-dependent metrics (Sensitivity and Specificity) and justifying the use of these metrics for model evaluation (Jiménez-Valverde, 2012; Jiménez and Soberón, 2020). Sensitivity and Specificity were calculated by transforming the continuous model predictions, resulting from the different training sets, into a binary response. The threshold was set to the value that maximized the sum of Sensitivity and Specificity calculated on the species’ testing set, thereby minimizing misclassification errors (Kaivanto, 2008). The difference in model performance (Δ AUC, Δ Sensitivity and Δ Specificity) was used to evaluate the impact of data quality filtering. Four choices facilitated the comparison of evaluation metrics within one species (Elith et al., 2011; Lobo et al., 2008; Merow et al., 2013): (1) an identical testing set, (2) identical Maxent settings (features and regularization coefficient), (3) identical background selection and (4) identical predictors.

2.5. The impact of data quality on model performance

We repeatedly (20 times) selected a random sample from the unfiltered and filtered training sets, at six predefined levels of 100, 250, 500, 1000, 2000 and 4000 presences (also see Fig. A.3). Model evaluation metrics were compared between training sets of constant fixed sample size but with different quality, resulting from the application of the different filters.

For the evaluation of data quality, species were divided into one of the six sample size levels, based on two conditions. Firstly, the sample size level was bounded from above by the number of available presences for all filtered training sets, including the 3-filter combination ACTIVITY-DETAIL-VALSTAT (ADV). This facilitated a comparison of all filters without the influence of inter-species differences. Secondly, species were classified at the highest level possible, based on the number of available presences in the original ADV training set. In other words, sample size was kept as close as possible to the number of recorded presences in the database. This way we prevented that large differences between the sample size of the unchanged training set (i.e. the actual occurrence in the data) and the fixed sample size would impact model

performance (Hanberry et al., 2012).

2.6. The impact of absolute sample size on model performance

For the evaluation of absolute sample size, we included models from different fixed sample sizes per species. We kept data quality constant by comparing results per filter and not between filters. Per filter, species were grouped in one out of six intervals of sample size that indicate the sample size of the original training sets: [100, 250[or [250, 500[or [500, 1000[or [2000, 4000[or ≥ 4000 . Species were thus constant across absolute sample sizes but not across filters nor intervals.

2.7. The combined impact of data quality and sample size on model performance

The impact on model performance of a change in data quality and a change in sample size will occur simultaneously. To evaluate this combined impact, we analysed 30,724 combinations of unfiltered and filtered training sets, with different changes in quality and sample size. We used all training sets of fixed sample size (at the six predefined levels) that we could obtain for each species, together with the original training sets, with sample size equal to the number of aggregated presences from the dataset.

Model evaluation metrics were averaged across the 20 repetitions for the fixed sample sizes (i.e. per species, filter type and sample size level) and we looked at the mean differences in model performance (Δ AUC, Δ Sensitivity, Δ Specificity) between models of an unfiltered training set and the filtered training sets. To fully capture the impact of the change in sample size, we assessed two 'sample size variables': the remaining sample size after filtering and the proportional reduction in sample size. The latter is defined as the proportion of presences removed from an unfiltered training set by applying a single filter or a combination of filters. See Fig. A.3 for an example of how many different datasets we could extract for one species and filter.

The combined impact of data quality and sample size on the difference in model performance was assessed using Generalized Additive Mixed Models (GAMMs) with species as a random effect, implemented in the 'mgcv' R package v1.8–31 (Wood, 2017). To account for the doubly-bounded character of our response variable, we rescaled Δ AUC, Δ Sensitivity and Δ Specificity to fall between 0 and 1 and used the 'betareg' family with logit-link. Smoothing functions were used to fit both sample size variables, with cubic spline method and $k = 5$ to reduce overfitting. We included interactions by allowing different smoothers per filter and by including the product of the remaining sample size and the proportional reduction in the equation. Per taxonomic group, the model which best explained the difference in model performance while keeping model complexity low was selected, by comparing the Akaike's Information Criterion (AIC) (Burnham et al., 2011) of multiple a priori GAMMs (full list in Appendix F) in the R package 'MuMIn' v1.43.17 (Barton, 2019). The relative importance of data quality (filter type) and sample size (sample size after filtering and proportional reduction) was assessed by comparing the proportion of explained deviance of those variables in the best model identified by our model selection.

We performed all analyses for the three evaluation metrics (AUC, Sensitivity and Specificity) across all species and within species groups and show the main results for AUC in the main text. All other results can be found in Appendices D through H in Supplementary Information 1. Models and statistical analyses were run in R v4.0.1 (R Core Team, 2020).

3. Results

Throughout the results section, the filters will be referred to as ACTIVITY (A): retaining records collected by observers with a high activity rate, DETAIL (D): retaining records that were submitted with information beyond the default date, location and species name, and VALSTAT

(V): retaining records marked as 'correct' in the data platform's validation system.

3.1. The impact of data quality on model performance

Fig. 1 shows that for all species, filtered data could deliver higher AUCs than unfiltered data, but with differences among sample size levels. Smaller sample sizes of filtered data were more likely to result in higher AUCs compared to large sample sizes of filtered data. At 100 presences, all filters could result in a higher AUC, while at 250 and 500 presences VALSTAT and DETAIL could deliver positive results. For larger sample sizes, VALSTAT and its combinations (at 1000 presences) or no filters at all (at 2000 and 4000 presences) benefitted model performance.

Plants were most sensitive to data quality, where DETAIL and VALSTAT, and also ACTIVITY at 100 presences, resulted in higher AUCs throughout. Birds were sensitive to data quality at the low and intermediate sample sizes, where the best option was VALSTAT. At 500 and 1000 presences, VALSTAT alone already increased AUC. At 100 and 250 presences, VALSTAT had to be combined with at least one other filter. For butterflies, AUCs increased when using ACTIVITY: alone or in combination with one or two other filters at 4000 presences, or in combination with VALSTAT at 1000 presences. For dragonflies, single filters were not powerful enough to increase AUC. Combining DETAIL with VALSTAT at 500 presences or with ACTIVITY at 1000 presences did deliver higher AUCs.

Similar results to AUC were found for Specificity, but mostly for plants at small sample sizes of 100 presences (all filters increased Specificity) and 250 presences (DETAIL, VALSTAT, A + D and A + V increased Specificity). At 500 presences, we noted increases in Specificity for dragonflies (A + D and A + D + V) and decreases in Specificity for plants (DETAIL, A + D and D + V). At larger sample sizes of 1000 presences or more, a higher Specificity was found only for birds (filter combinations). Data quality did not impact Specificity for butterflies (Fig. D.2).

Results for Sensitivity showed more negative impacts of using filtered data compared to AUC and Specificity, yet also increases in Sensitivity were noted for plants at 250 presences (DETAIL and its combinations) and 500 presences (all filters except ACTIVITY and A + V), and for butterflies at 4000 presences (ACTIVITY and its combinations). A lower Sensitivity was found for plants at 100 presences (VALSTAT and its combinations), for dragonflies at 500 presences (A + D and A + D + V) and for birds at 100 presences (A + D), 2000 presences (ACTIVITY and combinations with VALSTAT) and 4000 presences (DETAIL and its combinations) (Fig. D.1).

3.2. The impact of absolute sample size on model performance

Fig. 2 shows that reducing absolute sample size beyond a certain level always impacted AUCs negatively. This level depended more on the original sample size than on the applied filter. At lower original sample sizes (< 2000 presences), reducing sample size by 50% did not cause significant decreases in AUC for most filters, with exceptions for DETAIL, VALSTAT, A + D and A + V at 500 to 1000 presences. At larger original sample sizes (> 2000 presences), sample size could be reduced by 75% for most filters, with exceptions for VALSTAT and D + V at 2000 to 4000 presences. Reducing sample size to 100 presences, no matter what the original sample size was, always resulted in lower model performance. For birds and butterflies, the impact of sample size on AUC was similar to that of all species (Fig. E.3 and E.4). Dragonfly and plant models appeared less sensitive to sample size (Fig. E.5 and E.6).

Similar to AUC, the impact of smaller sample sizes on Specificity was generally negative across all species with a higher tolerance for larger reductions when original sample sizes were high, yet with more variation among filters (Fig. E.2). Specificity of butterfly and plant models (Fig. E.12 and E.14) appeared more sensitive to smaller sample sizes

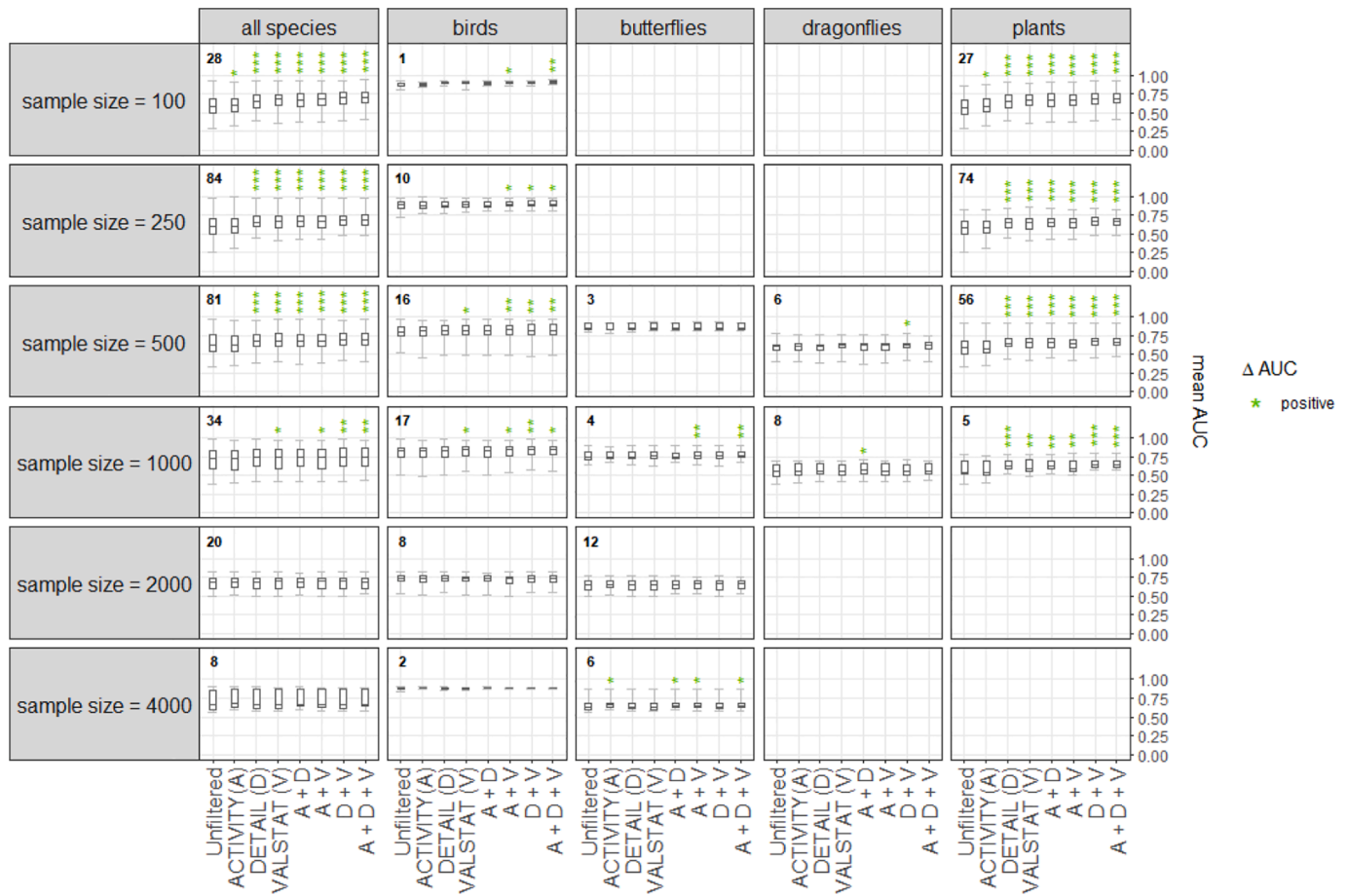


Fig. 1. The impact of data quality on AUC for all species and per taxonomic group, when absolute sample size is constant at six levels: 100, 250, 500, 1000, 2000 and 4000 presences. Per level, species were limited to those that could be modelled with all filters at the considered level, including the 3-filter combination ACTIVITY-DETAIL-VALSTAT. Species were subsequently classified at the highest level possible, meaning that AUC results cannot be compared between sample size levels, because species are different. The number of species in each comparison is presented in the top left corner of the graphic areas. Not all levels could be assessed for all taxonomic groups, because for example for butterflies there were no species with less than 500 presences in our dataset, so all species were classified at level 500 or higher. Boxplots represent medians, upper and lower quartiles with whiskers extending to the minimum and maximum values. Asterisks show significant differences in AUC compared to the unfiltered data, tested by a multiple comparison test with Benjamini & Hochberg (1995) correction (** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). Colours indicate only positive changes (green) for AUC. Results for the impact of data quality on Sensitivity and Specificity are found in Fig. D.1 and D.2 respectively.

compared to bird and dragonfly models (Fig. E.11 and E.13).

In contrast with results for AUC and Specificity, the impact of smaller sample sizes on Sensitivity is generally positive. Significant increases in Sensitivity were more likely to occur for higher quality data (filter combinations) at lower original sample sizes and for lower quality data (unfiltered data and single filters) at higher original sample sizes (Fig. E.1). For butterflies, dragonflies and plants, Sensitivity generally increased (Fig. E.8, E.9 and E.10) when Specificity decreased (Fig. E.12, E.13 and E.14). For birds, this contrast was less pronounced and we even noted more decreases in Sensitivity than increases when sample size was reduced (Fig. E.7).

3.3. The combined impact of data quality and sample size on model performance

Up to this point, the absolute sample size of unfiltered and filtered data remained identical. In reality, however, sample size usually decreases when applying quality filters. Therefore, the impact of sample size was quantified with two variables in this section: the ‘proportional reduction in sample size’ and the ‘sample size after filtering’ (also called ‘remaining sample size’). A detailed summary per species of all the filters and their impact on model performance showed that model performance mostly increased after filtering (depending on the applied filter, for 55 to

80% of the species for AUC, 49 to 55% for Sensitivity and 51 to 58% for Specificity), but that various filter-species combinations also show a negative impact on model performance (Table in Supplementary Information 2).

Per taxonomic group, we selected the ‘best’ GAMM (Appendix F), i.e. the model with the least parameters and a small difference in AIC ($\Delta AIC < 1$) compared to the top model, to evaluate the combined impact of data quality and sample size on the change in model performance caused by filtering. Fig. 3 shows the relative importance of the variables in the model for ΔAUC . Considering the averages across species (boxplots), the change in quality (the filter type) explained most of the variation in ΔAUC for plants and dragonflies, yet with high variability in percentage deviance explained (%DE) among species. The interaction between proportional reduction and sample size after filtering explained the most variation in ΔAUC for bird and butterfly models and is also important for dragonfly models. For plants, however, more variation in ΔAUC was explained by the interaction between quality and sample size after filtering. This interaction was also more important when considering the variation in Δ Sensitivity and Δ Specificity, and the differences between the proportional%DE for the variables ‘filter’, ‘interaction RxS’ and ‘interaction SxF’ became smaller. The filter type remained the most important variable for plants for predicting both Δ Sensitivity and Δ Specificity, yet with less variability among species compared to AUC

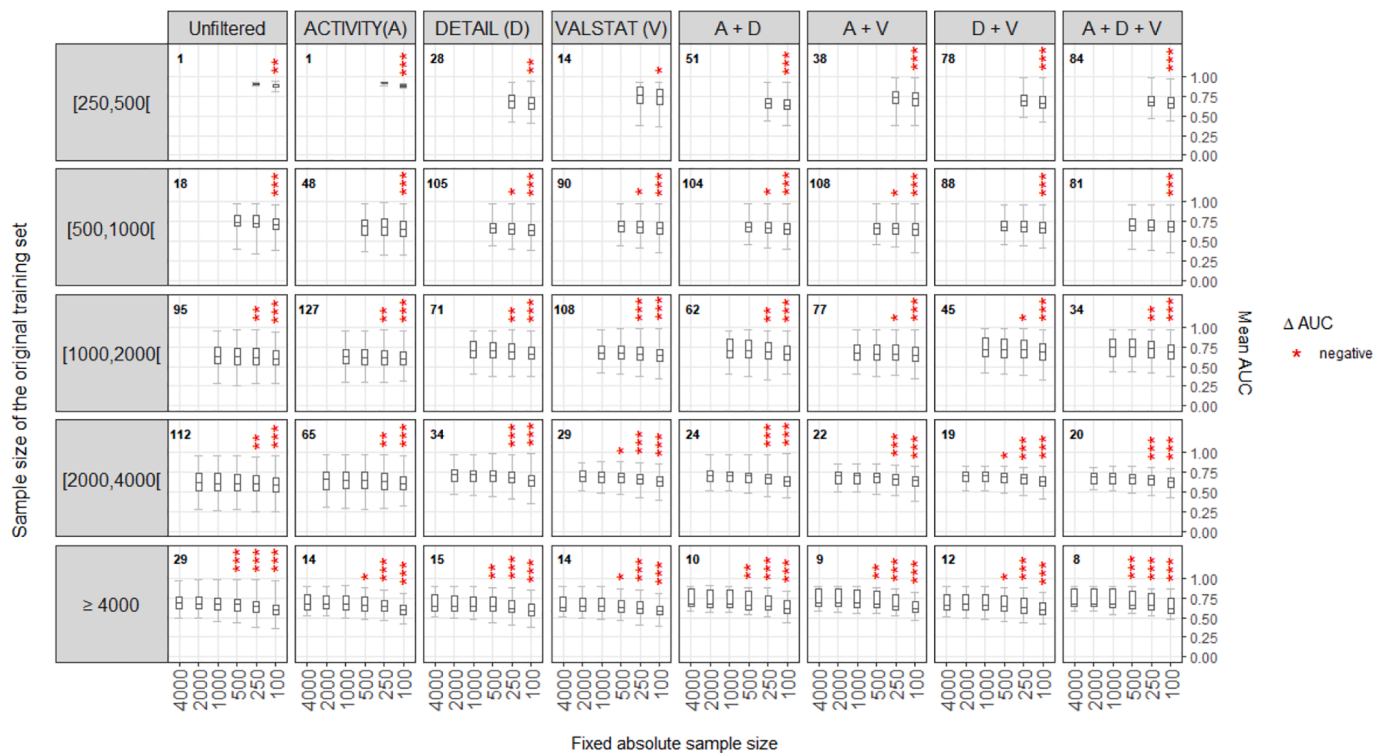


Fig. 2. The impact of absolute sample size on AUC for all species when data quality is constant. Per filter, species were grouped in one of the six specified intervals of sample size (left) that indicate the available sample sizes of the original training sets. AUCs were compared between models resulting from a repeated and random selection of different fixed sample sizes. Because species differ, results can only be compared within the graphic areas, i.e. between fixed sample sizes, but not between filters (horizontal) nor intervals (vertical). The number of species in each comparison is presented in the top left corner of the graphic areas. Boxplots represent medians, upper and lower quartiles with whiskers extending to the minimum and maximum values. Asterisks show significant differences in AUC compared to the highest sample size, tested by a multiple comparison test with Benjamini & Hochberg (1995) correction (** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). Colours indicate only negative changes (red) for AUC ($\Delta \text{AUC} < 0$). Results for the impact of absolute sample size on Sensitivity and Specificity are found in Fig. E.1 and E.2 respectively.

results (Fig. G.1 and G.2).

The predictions for ΔAUC of the best GAMM are presented in Fig. 4, along a continuous scale of proportional reduction and for three sample sizes after filtering, that we chose based on data availability: 100, 500 and 1000 presences. Predictions for $\Delta \text{Sensitivity}$ and $\Delta \text{Specificity}$ are found in Appendix H. The combined impact of filtering varies among taxonomic groups and we find the highest impacts for plant models (AUC and Sensitivity) and dragonfly models (Sensitivity), with the largest differences in model performance among filters. The predictions for birds and plants in Fig. 4 show that the best filters (i.e. the filters leading to increases in AUC) can differ between remaining sample sizes, confirmed by the relatively higher importance of the interaction between filter and sample size after filtering (Fig. 3). For plants, for example, the best filter was $A + D + V$ at small, but $D + V$ at large remaining sample sizes. Similar patterns were detected for Sensitivity (birds, dragonflies and plants in Fig. G.1 and H.1) and for Specificity (all groups in Fig. G.2 and H.2). In general, filters that resulted in high-quality data usually increased model performance (Figs. 1, D.1 and D.2). The proportional reduction in sample size could also be higher for those filters, before a negative impact on model performance was detected.

Overall, filtering increased AUCs and Sensitivity for plants (i.e. $\Delta > 0$) and decreased Sensitivity for birds (i.e. $\Delta < 0$), while in other cases, both increases and decreases in model performance were noted. Different trends described the impact of proportional reduction on model performance. The shape of the trend depended on the remaining sample size, with different trend slopes for all taxonomic groups and even different trend directions for birds (Sensitivity), butterflies (AUC), dragonflies (Sensitivity and Specificity) and plants (Sensitivity).

For AUC and Specificity, trends at small remaining sample sizes of 100 presences were negative, and filtering decreased model performance (i.e. $\Delta < 0$) beyond a certain maximal threshold of proportional reduction. Depending on the filter, maximum reductions in sample size could range from 0–35% (AUC) for birds, 20–60% (AUC) or 10–30% (Specificity) for butterflies, 55–85% (AUC) or 35–65% (Specificity) for dragonflies and 5–85% (Specificity) for plants. For Sensitivity, trends at a remaining sample size of 100 presences were positive, except for birds. Depending on the filter, reductions had to be at least 0–10% for butterflies and 35–70% for dragonflies before an increase in model performance was noted.

For larger remaining sample sizes of 500 and 1000 presences, trends in the impact of proportional reduction on ΔAUC and $\Delta \text{Specificity}$ remained negative for birds. For butterflies, trends for ΔAUC flattened with increasing sample size after filtering and ΔAUC s became largely positive, except for DETAIL, VALSTAT and $D + V$ at reductions above 45%. We even saw a positive trend when reductions above 70% resulted in larger sample sizes of 1000 presences. For dragonflies, trends were flattened for AUC and Specificity at larger remaining sample sizes and, except in the case of Specificity and VALSTAT, model performance generally increased after filtering. Trends even became positive for Specificity at larger remaining sample sizes of 1000 presences and reductions above 20%. For Sensitivity, however, trends became more negative for dragonflies at higher remaining sample sizes and only VALSTAT, at 500 presences and reductions below 70%, lead to increases in model performance.

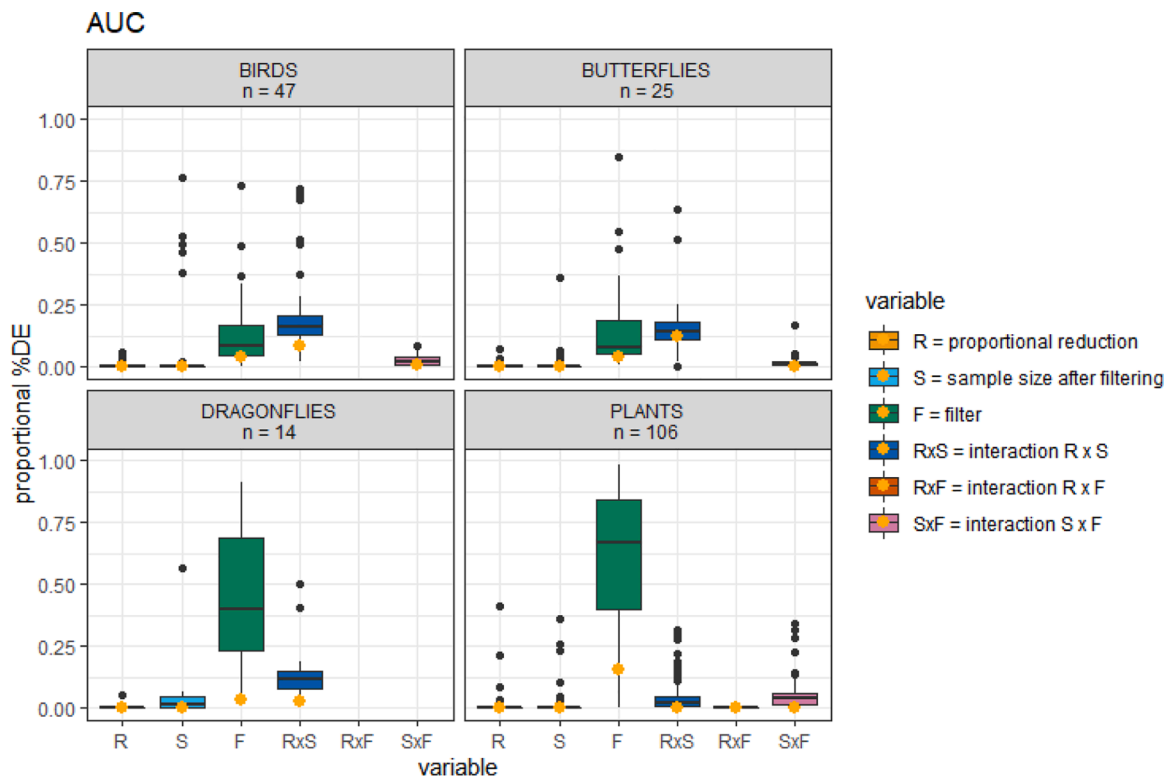


Fig. 3. The relative variable importance for the impact of data quality and sample size on Δ AUC, based on the proportion of the percentage of deviance explained (%DE) by the different explanatory variables in the best GAMM (Generalized Additive Mixed Model) per taxonomic group (orange dots), and the relative variable importance across species, in the GAMs (Generalized Additive Models) where the random species effect was excluded (but with identical smoothing parameters), relative to the %DE of the full model to summarize effects across n species. Species where the full model could not be estimated due to convergence issues were excluded from the summary. The relative variable importance for the impact on Δ Sensitivity and Δ Specificity are found in Fig. G.1 and G.2 respectively.

4. Discussion

We applied three dichotomous filters to opportunistic species records of citizen scientists as single filters and in combinations to test their impact on species distribution model (SDM) performance. We retained records from more active observers (ACTIVITY), detailed records, i.e. submitted with information beyond the default date, location and species name (DETAIL) and validated records, i.e. marked as 'correct' in the data platform's validation system (VALSTAT). Results indicated that the impact of stringent filtering on model performance (measured by changes in AUC, Sensitivity and Specificity) depended on the quality of the filtered data, both the proportional reduction in sample size caused by filtering and the remaining absolute sample size, and the taxonomic group.

A recurring pattern was that Specificity results (true negative rates) generally agreed more with AUC results than Sensitivity results (true positive rates). Moreover, Specificity usually increased when Sensitivity decreased and vice versa, which happens when evaluating model predictions on an external data set (Jiménez-Valverde, 2012). In the discussion that follows, we will focus on AUC results and we refer to the different results for Specificity and Sensitivity in the results section and Supplementary Information. The reader must keep in mind that the choice of an optimal threshold for threshold-dependent metrics depends on the characteristics of the SDM study (e.g. the goal of the study or the availability of information on species prevalence) (Jiménez-Valverde and Lobo, 2007) and that this choice might influence the recommendations for the most suited approach for quality filtering.

The quality of validated and detailed records was generally higher than the quality of records from more active observers. Luckily, validation of occurrence data entering large repositories, by synergies between human experts and computer intelligence, has been common

practice (e.g. in eBird, Kelling et al., 2013). The main benefits for data quality of such an internal validation system are (i) the quick and relatively easy identification and correction of false-positive errors, as they can impact model performance negatively (Costa et al., 2015), and (ii) an increased observer skill by the interaction between data managers and users (Sullivan et al., 2009).

Metadata cannot only hold important information to improve SDMs by overcoming problems with imperfect detection (e.g. Kéry et al., 2009) or other types of systematic bias (e.g. Johnston et al., 2017), but our results also indicate that the very act of supplying additional information can benefit data quality. We therefore agree that observer dedication and effort (linked to DETAIL) are more fit measures of data quality than observer experience and recording rates (linked to ACTIVITY) (Henckel et al., 2020; Steen et al., 2019). Like in several other studies on data quality, it remains tough to detect changes in model performance due to observer related measures of quality (e.g. observer skill and reporting consistency in Henckel et al. (2020) or observer expertise in Steen et al. (2019)). Combining multiple observer characteristics in observer profiles (Boakes et al., 2016; Isaac and Pocock, 2015) might be of added value here. Nonetheless, selecting data from active observers did significantly increase data quality for eight butterfly species that were among the most observed species in our dataset. We hypothesize that these common species are susceptible to misidentification by the inexperienced observer (Farmer et al., 2012), because of their highly familiar names in Dutch (*Aglais io* L., *Gonepteryx rhamni* L. and *Vanessa atalanta* L.) or because they are hard to distinguish from congeners (*Pieris rapae* L., *Maniola jurtina* L. and *Pararge aegeria* L.) (Vantieghe et al., 2017).

When deciding whether or not to filter, it is not only important to consider the obtained data quality, but also both the proportional reduction in sample size and the remaining absolute sample size after

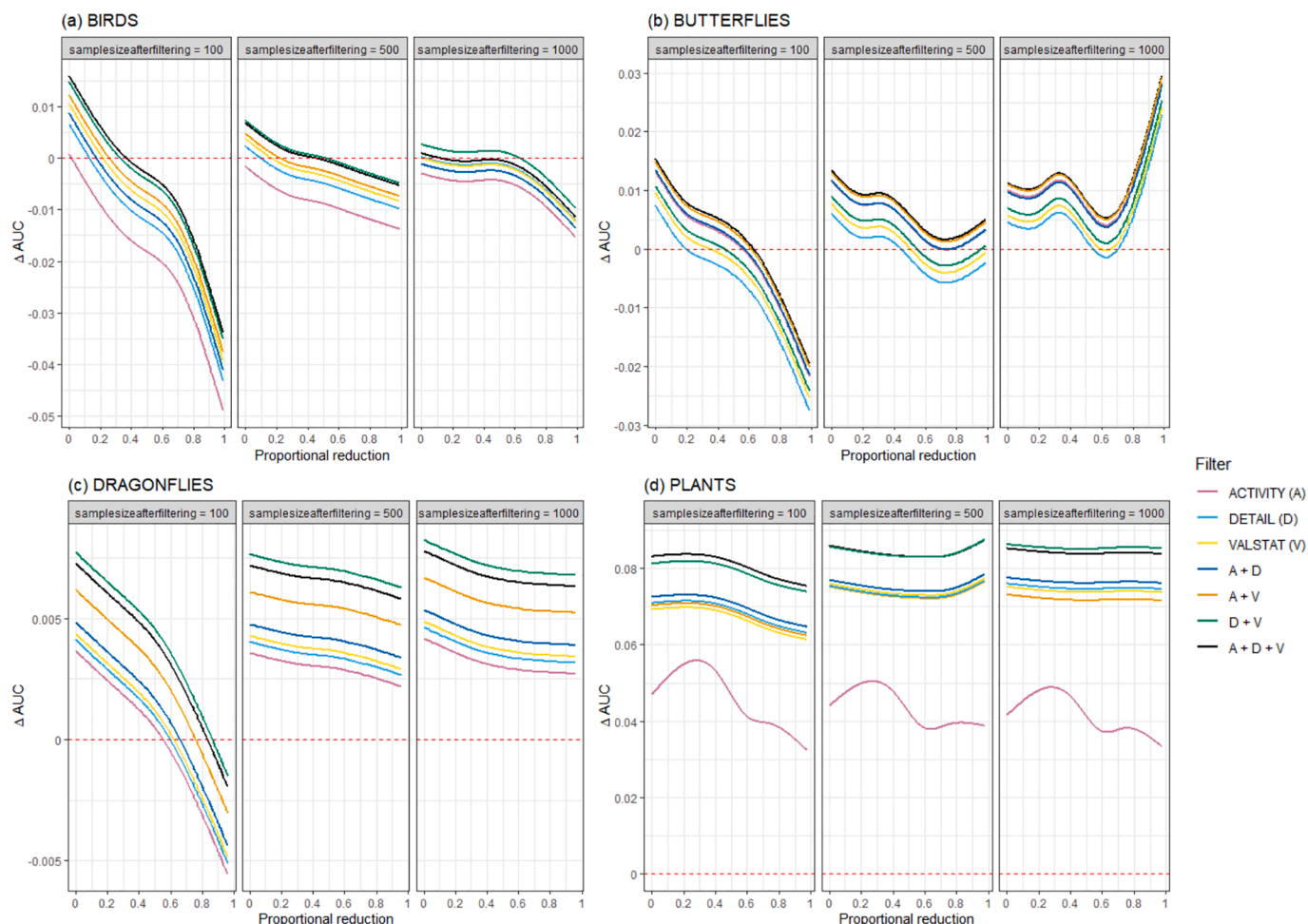


Fig. 4. The combined impact of data quality and sample size on Δ AUC per taxonomic group. The full lines are the predictions for Δ AUC ($AUC_{\text{filtered data}} - AUC_{\text{unfiltered data}}$) from the ‘best’ GAMM (Generalized Additive Mixed Model) along a continuous scale of proportional reduction in sample size and for three sample sizes after filtering that we chose based on data availability: 100, 500 and 1000 presences. Colours represent the different filters (data quality). The red dotted line equals a Δ AUC of 0, i.e. filtering did not impact model performance. We used the REML-method (restricted maximum likelihood) in the ‘gam’ function of the ‘mgcv’ R package v 1.8–31 (Wood, 2017) to model our data. Filter type was modelled as factor variable and species as random effect. Smoothing functions were used to fit both sample size variables (proportional reduction and sample size after filtering), with cubic spline method and $k = 5$. Δ AUC was rescaled to fall between 0 and 1, so that we could use the ‘betareg’ family with logit-link, because of the double-bounded character of the response variable (Δ AUC). The combined impact of data quality and sample size on Δ Sensitivity and Δ Specificity are shown in Fig. H.1 and H.2 respectively.

filtering. Large reductions or small remaining sample sizes do not always cause lower model performance, and while we agree that small sample sizes generally lead to worse models (Jiménez-Valverde et al., 2009; Liu et al., 2019), the relative change in sample size must not be ignored (Hanberry et al., 2012). Both measures of sample size co-define which filters are suited for model performance improvement. They have a limited impact on the selection of the best or worst filters based on AUC results, as the relative impact on AUC of the different filters remained largely constant across different changes in sample size. However, here we must mention that when the goal is to increase Sensitivity or Specificity, the remaining sample size after filtering does need to be considered (Appendix G).

The different drivers of model performance make the interpretation complex, but also highlight the importance of analysing multiple aspects of data manipulation together (Gábor et al., 2019). We add data quality to the list of drivers that can notably impact model performance, such as species characteristics, modeling technique and sample size (Gábor et al., 2019; Tessarolo et al., 2014). Compared to these factors, previous studies found marginal importance of the impact of sampling bias (Gábor et al., 2019; Tessarolo et al., 2014) and we have no reason to contest this finding based on our results (but note that we partially controlled for sampling bias by spatial thinning (Kramer-Schadt et al.,

2013)). Disentangling the different drivers of model performance in stringent filtering could be more feasible in a virtual species setting (Hirzel et al., 2001; Meynard et al., 2019), however, we argue that the simulation of filtered data of different quality is not trivial. This would require a more profound understanding of how data quality is impacted by data and species characteristics.

We can recommend stringent filtering for taxonomic groups where model performance is more impacted by data quality and less by sample size, such as the plants and dragonflies in this study. For plant models, we even observed that an increase in quality can mitigate the negative impact on AUC of reducing sample size to 100 presences (Fig. E.6 and Fig. 4). For the other taxonomic groups, this is only true below certain proportional reductions. Models from species with specific habitat conditions, such as dragonflies, are less sensitive to sample size and also profit from data quality increase. Such species have a more distinct link with their habitat and are easier to model compared to species with a broader niche (Hernandez et al., 2006). Nevertheless, caution is needed, because the impact of data quality on model performance shows large variation among plant and dragonfly species (Fig. 3) and is different when considering other evaluation metrics (Appendix E).

For taxonomic groups where model performance is more impacted by sample size and less by data quality, such as the birds and butterflies

in this study, we advise being more careful. We observed that filtering is less beneficial for these groups, probably because their abundant data already leads to relatively high model performance. Especially for birds, unfiltered data appeared very suited for modeling and filtering did not improve AUCs, certainly when less than 50% of the sample size remained. For these groups, even filters that do not cause large reductions nor lead to a small sample size could cause model performance to decrease. Nonetheless, choosing the right filter can mitigate the negative impact of sample size if the obtained quality is high enough (e.g. extracting data from active observers for butterflies or combining validated and detailed records for birds).

In this study, we focussed on the combined impact of data quality and sample size in stringent filtering, but we acknowledge that other factors, such as environmental filtering (Gabor et al., 2019), scale (Connor et al., 2018; Gottschalk et al., 2011), species traits (Hernandez et al., 2006; McPherson and Jetz, 2007) and SDM technique (Liu et al., 2019) will probably impact the sensitivity of a dataset to stringent filtering as well. For example, the proportion of high-quality data in a model training set is scale-dependent, because a coarse resolution gives a higher chance that at least one high-quality observation falls in a grid cell. Spatial thinning is therefore not only a way to remove spatial bias (Boria et al., 2014), but also to reduce other sources of uncertainty (Kramer-Schadt et al., 2013), such as the presence of data with uncertain quality. We also detected variation among species, and as taxonomic groups still show plenty variation in species traits (Maes et al., 2019), it might be more efficient to formulate recommendations for stringent filtering based on species traits rather than on taxonomy. Species prone to misidentification, for example, can benefit from retaining only records validated as correct based on photos supplied by the observer (Vantieghem et al., 2017) and we have indications that, for example, habitat-specificity, mobility and popularity impact the sensitivity of a species to data quality filtering as well.

Our recommendations are limited to the discrimination accuracy of Maxent. As Maxent usually comes out as a relatively more robust SDM technique (Thibaud et al., 2014), our conclusions are likely to be conservative. We therefore expect at least a similar, if not a larger, impact of data quality filtering for other SDM techniques.

5. Conclusions

We conclude that data quality filtering has the potential to improve predictions of species distributions, especially for species where SDMs are less sensitive to decreases in sample size. However, data quality should not be pursued at any cost, because filtering can also impact model performance negatively, e.g. for species with abundant data or when filtering leads to low sample sizes or causes high sample size reductions. We encourage the further development and adoption of techniques that can increase the availability of high-quality data, to be able to fully profit from the benefits of opportunistic citizen science data. The value of a database-integrated validation system demonstrates the potential of bulky datasets from platforms and applications where the focus is on the identification and validation of species observations, such as iNaturalist (<https://www.inaturalist.org/>), Pl@ntnet (<https://www.plantnet.org>) or ObsIdentify (Hogeweg et al., 2019). We advise to always “Think before you shrink” because volunteer generated data can make valuable contributions to science if processed correctly.

Data accessibility statement

The full dataset (unfiltered and filtered species presences for model training, model testing sets and model predictors) are available in Dryad Digital Repository at a 1 × 1 km resolution (<https://doi.org/10.5061/dryad.jwstqj83>).

Declaration of Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We foremost thank the thousands of volunteers for collecting the millions of records that supported this study; *Natuurpunt Studie* for making the data available for this research, in particular Joeri Belis and Karin Gielen for composing the dataset; and Wouter Vanreusel, Pieter Vanormelingen and Wouter van Landuyt for their help with the interpretation of the data. This work was supported by the Flemish Research Foundation FWO-SB [grant number 1S92118N].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ecolmodel.2021.109453](https://doi.org/10.1016/j.ecolmodel.2021.109453).

References

- Aichison, J., 2003. *The Statistical Analysis of Compositional Data*. Blackburn Press, Caldwell.
- Antrop, M., 2004. Landscape change and the urbanization process in Europe. *Landsc. Urban Plan.* 67, 9–26. [https://doi.org/10.1016/S0169-2046\(03\)00026-4](https://doi.org/10.1016/S0169-2046(03)00026-4).
- Barton, K., 2019. *MuMIn: Multi-Model Inference*. R Package Version 1.
- Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J.F., Pecl, G.T., Barrett, N., Frusher, S., 2014. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* 173, 144–154. <https://doi.org/10.1016/j.biocon.2013.07.037>.
- Boakes, E.H., Gliozzo, G., Seymour, V., Harvey, M., Smith, C., Roy, D.B., Haklay, M., 2016. Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording. *Behaviour*. Sci. Rep. 6 <https://doi.org/10.1038/srep33051>.
- Boria, R.A., Olson, L.E., Goodman, S.M., Anderson, R.P., 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol. Modell.* 275, 73–77. <https://doi.org/10.1016/j.ecolmodel.2013.12.012>.
- Burgess, H.K., Debey, L.B., Froehlich, H.E., Schmidt, N., Theobald, E.J., Ettinger, A.K., Hillerislambers, J., Tewksbury, J., Parrish, J.K., 2017. The science of citizen science: exploring barriers to use as a primary research tool. *Biol. Conserv.* 208, 113–120. <https://doi.org/10.1016/j.biocon.2016.05.014>.
- Burnham, K.P., Anderson, D.R., Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* 65, 23–35. <https://doi.org/10.1007/s00265-010-1029-6>.
- Chen, Q., Yin, Y., Zhao, R., Yang, Y., Teixeira da Silva, J.A., Yu, X., 2020. Incorporating local adaptation into species distribution modeling of *Paeonia mairai*, an endemic plant to China. *Front. Plant Sci.* 10, 1717. <https://doi.org/10.3389/fpls.2019.01717>.
- Connor, T., Hull, V., Viña, A., Shortridge, A., Tang, Y., Zhang, J., Wang, F., Liu, J., 2018. Effects of grain size and niche breadth on species distribution modeling. *Ecography (Cop.)* 41, 1270–1282. <https://doi.org/10.1111/ecog.03416>.
- Costa, H., Foody, G., Jiménez, S., Silva, L., 2015. Impacts of species misidentification on species distribution modeling with presence-only data. *ISPRS Int. J. Geo-Information* 4, 2496–2518. <https://doi.org/10.3390/ijgi4042496>.
- Couvreur, M., Menschaert, J., Sevenant, M., Ronse, A., Van Landuyt, W., De Blust, G., Antrop, M., Hermy, M., 2004. Ecodistricten en ecoregio's als instrument voor natuurstudie en milieubeleid. *Natuur.Focus* 3, 51–58.
- by Demolder, H., Schneiders, A., Spanhove, T., Maes, D., Landuyt, W., Van, Adriaens, T., 2014. In: Stevens, M. (Ed.). *Instituut voor Natuur- en Bosonderzoek, Brussels*. by.
- Dobson, A.D.M., Milner-Gulland, E.J., Aebischer, N.J., Beale, C.M., Brozovic, R., Coals, P., Critchlow, R., Dancer, A., Greve, M., Hinsley, A., Ibbett, H., Johnston, A., Kuiper, T., Le Comber, S., Mahood, S.P., Moore, J.F., Nilsen, E.B., Pocock, M.J.O., Quinn, A., Travers, H., Wilfred, P., Wright, J., Keane, A., 2020. Making messy data work for conservation. *One Earth* 2, 455–465. <https://doi.org/10.1016/j.oneear.2020.04.012>.
- Early, R., Bradley, B.A., Dukes, J.S., Lawler, J.J., Olden, J.D., Blumenthal, D.M., Gonzalez, P., Grosholz, E.D., Ibanez, I., Miller, L.P., Sorte, C.J.B., Tatem, A.J., 2016. Global threats from invasive alien species in the twenty-first century and national response capacities. *Nat. Commun.* 7, 1–9. <https://doi.org/10.1038/ncomms12485>.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>.
- Farmer, R.G., Leonard, M.L., Horn, A.G., 2012. Observer effects and avian-call-count survey quality: rare-species biases and overconfidence. *Auk* 129, 76–86. <https://doi.org/10.1525/auk.2012.11129>.

- Serra-Diaz, J.M., Enquist, B.J., Maitner, B., Merow, C., Svenning, J.C., 2017. Big data of tree species distributions: how big and how good? For. Ecosyst. 4 <https://doi.org/10.1186/s40663-017-0120-0>.
- Soroye, P., Ahmed, N., Kerr, J.T., 2018. Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research. Glob. Chang. Biol. 24, 5281–5291. <https://doi.org/10.1111/gcb.14358>.
- Steen, V.A., Elphick, C.S., Tingley, M.W., 2019. An evaluation of stringent filtering to improve species distribution models from citizen science data. Divers. Distrib. 25, 1857–1869. <https://doi.org/10.1111/ddi.12985>.
- Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Damoulas, T., Dhondt, A.A., Dieterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J.W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W.M., Iliiff, M.J., Lagoze, C., La Sorte, F.A., Merrifield, M., Morris, W., Phillips, T.B., Reynolds, M., Rodewald, A.D., Rosenberg, K.V., Trautmann, N.M., Wiggins, A., Winkler, D.W., Wong, W., Wood, C.L., Yu, J., Kelling, S., 2014. The eBird enterprise: an integrated approach to development and application of citizen science. Biol. Conserv. 169, 31–40. <https://doi.org/10.1016/j.biocon.2013.11.003>.
- Sullivan, B.L., Wood, C.L., Iliiff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: a citizen-based bird observation network in the biological sciences. Biol. Conserv. 142, 2282–2292. <https://doi.org/10.1016/j.biocon.2009.05.006>.
- Swinnen, K.R.R., Vercayie, D., Vanreusel, W., Barendse, R., Boers, K., Bogaert, J., Dekeukeleire, D., Driessens, G., Dupriez, P., Jooris, R., Steeman, R., van Asten, K., van den Neucker, T., van Dorsselaer, P., van Vooren, P., Wysmantel, N., Gielen, K., Desmet, P., Herremans, M., 2018. Waarnemingen.be-Non-native plant and animal occurrences in Flanders and the Brussels Capital Region. Belgium. BioInvasions Rec. 7, 335–342. <https://doi.org/10.3391/bir.2018.7.3.17>.
- Tessarolo, G., Rangel, T.F., Araújo, M.B., Hortal, J., 2014. Uncertainty associated with survey design in species distribution models. Divers. Distrib. 20, 1258–1269. <https://doi.org/10.1111/ddi.12236>.
- 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. Biol. Conserv. 181, 236–244. <https://doi.org/10.1016/j.biocon.2014.10.021>.
- Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C., Guisan, A., 2014. Measuring the relative effect of factors affecting species distribution model predictions. Methods Ecol. Evol. 5, 947–955. <https://doi.org/10.1111/2041-210x.12203>.
- Titeux, N., Maes, D., Marmion, M., Luoto, M., Heikkinen, R.K., 2009. Inclusion of soil data improves the performance of bioclimatic envelope models for insect species distributions in temperate Europe. J. Biogeogr. 36, 1459–1473. <https://doi.org/10.1111/j.1365-2699.2009.02088.x>.
- Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.B., Pe'er, G., Singer, A., Bridle, J.R., Crozier, L.G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J.J., Holt, R.D., Huth, A., Johst, K., Krug, C.B., Leadley, P.W., Palmer, S.C.F., Pantel, J.H., Schmitz, A., Zollner, P.A., Travis, J.M.J., 2016. Improving the forecast for biodiversity under climate change. Science (80-) 353. <https://doi.org/10.1126/science.aad8466> aad8466.
- 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. J. Appl. Ecol. 50, 1450–1458. <https://doi.org/10.1111/1365-2664.12158>.
- Vantieghe, P., Maes, D., Kaiser, A., Merckx, T., 2017. Quality of citizen science data and its consequences for the conservation of skipper butterflies (Hesperiidae) in Flanders (northern Belgium). J. Insect Conserv. 21, 451–463. <https://doi.org/10.1007/s10841-016-9924-4>.
- Varela, S., Anderson, R.P., García-Valdés, R., Fernández-González, F., 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography (Cop.) 37, 1084–1091. <https://doi.org/10.1111/j.1600-0587.2013.00441.x>.
- Vermeersch, G., Devos, K., Everaert, J., Feys, S., Herremans, M., Onkelinx, T., Stienen, E., T'Jollyn, F., 2020. Broedvogels in Vlaanderen 2013-2018. Recente status en trends van in Vlaanderen broedende vogelsoorten. Instituut voor Natuur- en Bosonderzoek, Brussels. <https://doi.org/10.21436/inbor.18794135>.
- Vollering, J., Halvorsen, R., Auestad, I., Rydgren, K., 2019. Bunching up the background betters bias in species distribution models. Ecography (Cop.) 42, 1717–1727. <https://doi.org/10.1111/ecog.04503>.
- Wang, Y., Stone, L., 2019. Understanding the connections between species distribution models for presence-background data. Theor. Ecol. 12, 73–88. <https://doi.org/10.1007/s12080-018-0389-9>.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., 2008. Effects of sample size on the performance of species distribution models. Divers. Distrib. 14, 763–773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>.
- Wood, K.A., Stillman, R.A., Hilton, G.M., 2018. Conservation in a changing world needs predictive models. Anim. Conserv. 21, 87–88. <https://doi.org/10.1111/acv.12371>.
- Wood, S., 2017. Generalized Additive Models: An Introduction With R, 2nd ed. Chapman and Hall, London.
- Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Aroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitão, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., Merow, C., 2020. A standard protocol for reporting species distribution models. Ecography (Cop.) 43, 1–17. <https://doi.org/10.1111/ecog.04960>.