ORIGINAL RESEARCH

Ecology and Evolution

WILEY

How citizen science could improve species distribution models and their independent assessment

Florence Matutini[1] [ID] | Jacques Baudry[2] | Guillaume Pain[1] | Morgane Sineau[3] |
Joséphine Pithon[1]

[1]BAGAP, INRAE, Institut Agro, ESA, Angers, France

[2]BAGAP, INRAE, Institut Agro, ESA, Rennes, France

[3]URCPIE Pays de la Loire, Nantes, France

Correspondence
Florence Matutini, BAGAP, INRAE Institut Agro, ESA, 55 Rue Rabelais, Angers 49000, France.
Email: florence.matutini@gmail.com

Abstract

1. Species distribution models (SDM) have been increasingly developed in recent years, but their validity is questioned. Their assessment can be improved by the use of independent data, but this can be difficult to obtain and prohibitive to collect. Standardized data from citizen science may be used to establish external evaluation datasets and to improve SDM validation and applicability.

2. We used opportunistic presence-only data along with presence–absence data from a standardized citizen science program to establish and assess habitat suitability maps for 9 species of amphibian in western France. We assessed Generalized Additive and Random Forest Models' performance by (1) cross-validation using 30% of the opportunistic dataset used to calibrate the model or (2) external validation using different independent datasets derived from citizen science monitoring. We tested the effects of applying different combinations of filters to the citizen data and of complementing it with additional standardized fieldwork.

3. Cross-validation with an internal evaluation dataset resulted in higher AUC (Area Under the receiver operating Curve) than external evaluation causing overestimation of model accuracy and did not select the same models; models integrating sampling effort performed better with external validation. AUC, specificity, and sensitivity of models calculated with different filtered external datasets differed for some species. However, for most species, complementary fieldwork was not necessary to obtain coherent results, as long as the citizen science data were strongly filtered.

4. Since external validation methods using independent data are considered more robust, filtering data from citizen sciences may make a valuable contribution to the assessment of SDM. Limited complementary fieldwork with volunteer's participation to complete ecological gradients may also possibly enhance citizen involvement and lead to better use of SDM in decision processes for nature conservation.

KEYWORDS
amphibians, biodiversity conservation, data culling, data filtering, external evaluation, habitat suitability modeling, sampling effort

1 | INTRODUCTION

In the current context of biodiversity loss, a stronger relationship between conservation science and citizen participation could help to make conservation actions more effective (Forrester et al., 2017; Lewandowski & Oberhauser, 2017). Availability of data from citizen sciences has considerably increased over the past few decades (Dickinson et al., 2010; McKinley et al., 2017). This data have great potential because (a) large quantities of data can be collected over large areas, which would be difficult and expensive for researchers to collect; (b) data may be collected over long time periods, which is especially useful for studying the effects of climate and landscape changes on population dynamics at large scales; (c) citizens are involved in the research process, thereby gaining knowledge, and their involvement might lead to improved implementation of biodiversity conservation action (Dickinson et al., 2010; McKinley et al., 2017). However, quality of data from participatory sciences is heterogeneous and different methods have been developed to boost data accuracy and account for bias, including interactive project development, volunteer training, expert data validation, and statistical modeling improvement (Kosmala et al., 2016). Although researchers have been skeptical about the value of datasets from citizen science, recent publications show that some could be as valid as data collected by professional scientists (Kosmala et al., 2016). This is conditional on such data being judged in context (i.e. according to the sampling methods used, program objectives, and applications) on the use of rigorous data sorting and analyses (Isaac et al., 2014; Robinson et al., 2020; Steen et al., 2019).

Opportunistic presence-only data collected by citizens at large scales have contributed to the expansion of species distribution models (SDM) over the past twenty years, particularly for biological conservation applications (Guisan & Thuiller, 2005). The validity of presence-only SDM is, however, increasingly questioned as well as their applicability (Barve et al., 2011). Presence-only data come from different source databases reduced to simple species presence records and mostly collected in an unstandardized way by volunteers. In contrast to presence–absence data, they are abundant but have poor quality, few metadata and come from different sources (Robinson et al., 2020). This introduces numerous sources of bias that need to be assessed and accounted for in modeling processes (Guillera-Arroita et al., 2015; Phillips et al., 2009). Common problems are heterogeneous sampling effort, conditions and methods, imprecise spatial and temporal resolutions, and different levels of expertise among observers (Dickinson et al., 2010; McKinley et al., 2017; Phillips et al., 2009; Schulman et al., 2007). Different methods have been developed to correct these biases, including sorting or weighting presence-data to reduce identification errors and pseudo-replication linked to sampling effort (Guisan & Theurillat, 2000; Phillips et al., 2009) and/or using sampling effort assumptions to establish pseudo-absence sampling strategies (Barbet-Massin et al., 2012). Understanding the structure and intensity of sampling effort in space is essential to determine whether an undetected species is truly absent. For example, it may be conditioned by site

accessibility (Kadmon et al., 2004; Phillips et al., 2009), site attractiveness, or observer distribution (Phillips et al., 2009; Robinson et al., 2018). Not accounting for heterogeneous sampling effort or using erroneous assumptions to define it can lead to overassessment of model accuracy and/or false interpretation (Guillera-Arroita et al., 2015; Phillips et al., 2009; Schulman et al., 2007).

SDM validation is challenging (Vaughan & Ormerod, 2005) but is a crucial step for applying results to conservation objectives. There is still debate about SDM validity, especially when presence-only data are used to calibrate models (Barve et al., 2011). Using data with the same spatial bias to calibrate and assess a model tends to overestimate prediction accuracy, by modeling observation processes more than ecological processes therefore producing erroneous results. Currently, testing model accuracy with a fully independent dataset is considered the most robust method for assessing SDM (Araujo et al., 2005; Guisan et al., 2017). However, obtaining an external dataset for large-scale studies is often cost prohibitive and exploiting data from standardized citizen science programs may in some cases provide the solution. For example, Robinson et al. (2020) have shown that using filtered large-scale citizen science data for SDM calibration can improve model accuracy. Alternatively, detection–nondetection data from more standardized citizen sciences programs which are rarer than opportunistic data but have higher quality could provide presence–absence sets for external validation of presence-only SDM. In addition, using presence-only and presence–absence data at different stages of the modeling process could be a method for combining different datasets with heterogeneous quality which is a current challenge to improve SDM validity (Robinson et al., 2020; Zipkin & Saunders, 2017).

Amphibians are among the most threatened taxa in the world with rapid and widespread population declines (Stuart et al., 2004). They are particularly sensitive to fragmentation and habitat loss (Cushman, 2006) because they need different resources during their life cycle involving movements (seasonal migration and dispersion) between aquatic sites (usually ponds) and terrestrial areas (Cushman, 2006; Sinsch, 1990). Many citizen science programs have been initiated for monitoring amphibian species (De Solla et al., 2005; Schmeller et al., 2009), and data collected have been used in some conservation studies to describe population trends (Petrovan & Schmidt, 2016), road effects (Cosentino et al., 2014), climate change (Préau et al., 2019), and large-scale species distributions (Brown et al., 2016). Despite abundant literature on amphibian ecology and the availability of several citizen science databases, few studies have attempted predictive amphibian distribution models at large scales (Brown et al., 2016; Graham & Hijmans, 2006; Préau et al., 2019). Therefore, amphibian data could be suitable for testing the capacity of different types of citizen data (presence–absence or opportunist) for calibrating and assessing SDM.

Here, we compare the predictive performance of presence-only SDM for nine amphibian species using different types of data (*internal* presence-only or *external* presence–absence) from citizen science programs for their assessment. We also test the opportunity to use filtered standardized citizen science data to constitute

the independent dataset for external evaluation. We hypothesized that (1) the type of data used for validation (internal or external) would influence the assessment of model accuracy; (2) standardized citizen science datasets might be used as independent data for external evaluation of SDM using data filters and/or complementary fieldwork.

2 | MATERIALS AND METHODS

2.1 | Study area

Our study was performed in Pays de la Loire (western France), a region covering 32,082 km^2 with low relief and bordering on the Atlantic Ocean to the west. The region has an extensive hydrographic network organized around the River Loire and its tributaries, influencing local climate and landscape configuration. Agricultural landscapes dominate the region and traditional hedgerow network landscapes associated with extensive livestock farming are recognized for their conservation value. Such mosaics of small pastures delimited by hedgerows and small woods and generally associated with dense pond systems (Baudry et al., 2000) are favorable for many organisms including endangered species such as some amphibians species (Boissinot et al., 2019). With 21 known species (for 43 species recorded in France), the region has a high responsibility for the preservation of amphibians and their habitats, including traditional hedgerow landscape and wetlands.

2.2 | Biological data

We studied habitat suitability of 9 amphibian species: *Bufo spinosus*, *Hyla arborea*, *Rana dalmatina*, *Rana temporaria*, *Triturus cristatus*, *Triturus marmoratus*, *Lissotriton helveticus*, *Salamandra Salamandra*, and *Pelodytes punctatus*. Two types of amphibian data were used: (1) opportunistic data from a citizen database with presence-only records for model calibration and internal evaluation and (2) standardized detection–nondetection data from a citizen science program and complementary field work for external evaluations. A more detailed description of the datasets and complementation strategies is available in Appendix S1.

2.2.1 | Opportunistic presence data (calibration and cross-validation dataset)

We accessed presence-only occurrences from a regional database for the period 2013–2019. 86% of the dataset was collected by citizens and recorded online (website or associated mobile application) and 14% by various professional organizations involved in nature protection. All data were compiled for the regional Atlas of amphibians by a French nongovernmental organization (French BirdLife partner—LPO). See Appendix S1: Table S1 for data sources.

We retained only species with enough data according to number of predictors used (i.e. at least 477 presence cells; see Appendix S1: Table S2). We selected only precise GPS records (precision of the observation under 50 m), and we checked all data for anomalies in geographical location or species identification.

For each species, we sorted data to reduce spatial autocorrelation by projecting presence data on a 500 m resolution grid and retaining only cells containing at least one occurrence as presence cells for the analyses (Guisan & Theurillat, 2000). We chose a 500 m resolution as it is the mean size of species' home ranges (Semlitsch & Bodie, 2003). Finally, we excluded all opportunist data intersecting cells used for external validation described below to increase independency between calibration and validation sets.

2.2.2 | Standardized detection–nondetection data (external validation datasets)

For external validation, we first extracted *detection–nondetection* amphibian data for the period 2013–2019 collected as part of a citizen science program called "Un Dragon dans mon Jardin" (Appendix S1: Section 1.2). We retained 576 sites which were monitored at least 3 times between February and June during at least one year and following a standard method commonly used for amphibian community surveys (Boissinot et al., 2019). We called this dataset *CS.0* (see Figure 1). Some large areas of the region were not sampled due to lack of observers so that data were clustered near cities, with spatial autocorrelation. Therefore, with help from volunteers, we completed this dataset with some additional fieldwork and applied filters.

To complete and filter *CS.0*, different strategies were used. First, we organized complementary fieldwork in 2018 and 2019 to complete two landscape gradients (woody element density and pond density) which are two variables known to strongly affect amphibian distribution and which are relevant in our regional context. All sites were selected randomly but so as to maximize and decorrelate the two landscape gradients in different areas (see Appendix S1: Section 1.3). In total, 263 sites were monitored: 132 sites by experts in 2018–2019 (called *PRO*, see Figure 1) and 131 by 75 volunteers in 2019 (called *VOL*, see Figure 1). All data (*CS.0*, *PRO* and *VOL*) were projected on the same 500 m resolution grid. One further problem, common in citizen science programs (Geldmann et al., 2016), is that only aquatic sites are surveyed while areas known to be very unsuitable for amphibians such as urban areas and intensive agriculture are generally excluded. To reduce this source of bias, we randomly selected 5% more 500 m grid cells in totally urbanized areas without aquatic sites and 5% more grid cells in homogeneous croplands without trees or ponds and we attributed "absence" values to each after field checks (called *ABS*, see Figure 1). These landscapes represent 9% of the total area of the region.

Second, we applied different filter combinations to establish subsets from *CS.0*, *PRO*, and *VOL* (see Figure 1):
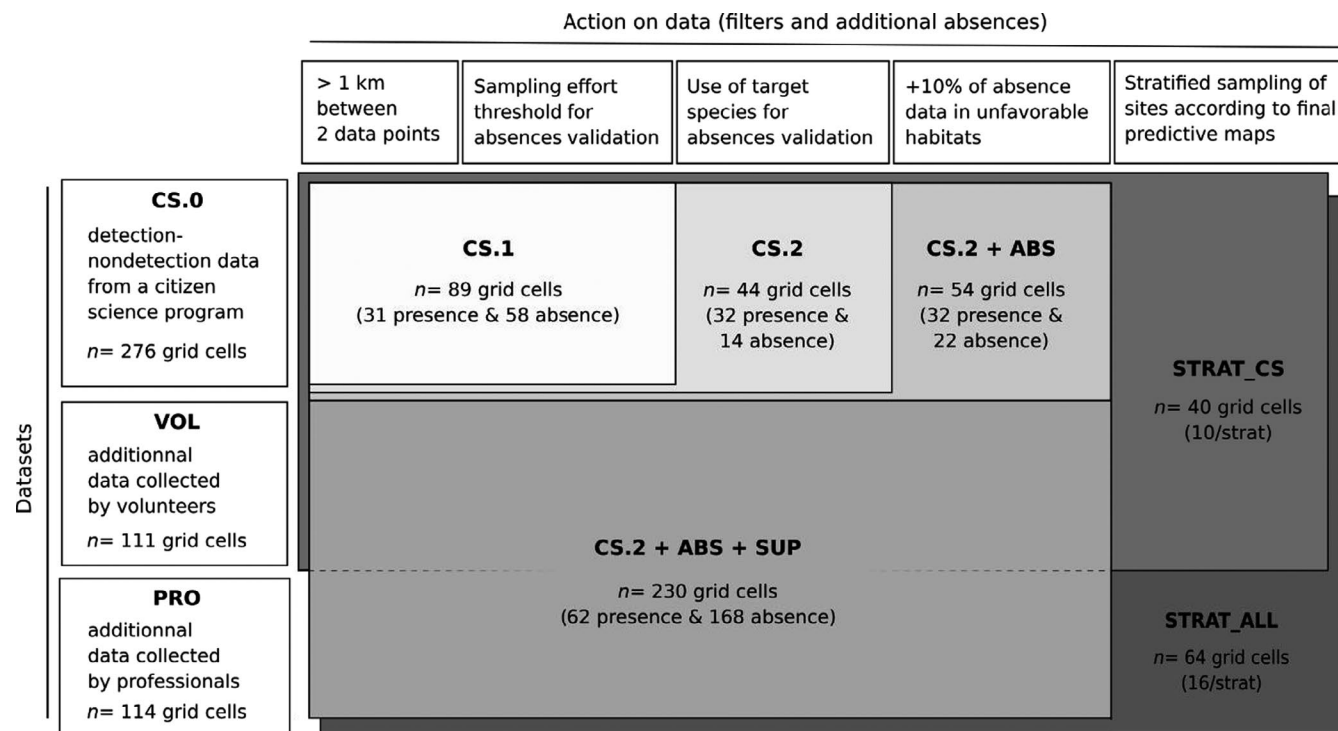
FIGURE 1 Description of datasets, filters and complementation used for external evaluation derived. Number of data (*n* grid cells) given for *T. marmoratus*, for 1 iteration only and s3. CS: data derived from the citizen science data set; SUP: additional data collected by volunteers (VOL) and professionals (PRO)

1. A minimum distance of 1 km between two grids containing data;

2. Threshold values to validate nondetection as absence data and exclude under-sampled sites, defined as a minimum sampling effort required to detect a species based on the species' detectability group and observer level of expertise. Four species detectability groups were defined from occupancy studies in France (Boissinot, 2009; Petitot et al., 2014) and the UK (Sewell et al., 2010). Observers were classed as either *expert*, *intermediate*, or *novice* using 3 criteria: number of years of participation, number of species observed and permit holder for amphibian capture. A "*novice*" observer was considered more likely to miss or misidentify a species which was actually present than an "*expert*" observer for the same considered survey effort. In addition, novice observers did not use sampling nets, influencing detectability, in particular of Urodeles. Based on our observer classes and sampling methods used (e.g., acoustic, visual and/or fishing), we set threshold values for the minimal sampling effort needed to validate absence data, depending on species detectability (see Appendix S1: Section 1.4 for details) and according to the results from Boissinot (2009) for minimum sampling effort needed to detect a focal species (with 95% probability) in a similar biological and landscape context.

3. Target species to valid nondetection as absence, as recommended by Phillips et al. (2009). So, if *species A* with the same detectability as *species B* is detected at a site, then *species B* is likely to be truly absent (see Appendix S1: Section 1.5 for target species list).

4. Stratified sampling on final prediction maps (see STRAT_CS and STRAT_ALL in Figure 1). We established independent datasets, stratified by model predictions for each species (Newbold et al., 2010; Guisan et al., 2017) with equal random sampling of validation cells with presence or absence data in 4 habitat suitability classes predicted (i.e. [0;0.3[, [0.3;0.5[, [0.5;0.7[and [0.7;1]). We obtained an equal number of validation cells by predicted suitability class (see 1.4.3 for predictive map used).

2.3 | Environmental dataset and variables

We assembled environmental data relevant to amphibian ecology and of importance in the study region (Guisan et al., 2017). A more detailed description of variables with associate references is available in Appendix S2: Table S1.

Bioclimatic variables were accessed from a compilation of climate data for the period 1950–2000 at a spatial resolution of 5 km^2 (Hijmans et al., 2005). An altitude variable was derived from the U.S. Geological Survey's Hydro-K dataset, at the same spatial resolution. We performed a principal-components analysis (PCA) on 11 bioclimatic variables relevant for amphibians and the altitudinal layer to produce 2 uncorrelated axes (see Appendix S2: Table S4 and Figure S2). Land-cover data were downloaded from the highly detailed vector database OCS GE 1.1 (IGN 2019), the Theia OSO Land Cover Map 2017 (available at www.theia-land.fr), and from BDTopo (IGN 2019). This was coupled with a more detailed

regional inventory of hedgerows (from 2005 to 2008) and ponds (2012) and a national farming database from the EU LPIS (Land Parcel Identification System 2016) used to classify agricultural areas (see Table 1).

We calculated land-cover variables in windows composed of a 500 m grid cell with a buffer of 300 m (see Table 1). This took into account landscape context based on species' dispersal capacities as well as the resolution of the species dataset (Guisan & Thuiller, 2005). Distance and home range differ among amphibian species but a 1 km circle may be accepted as an average maximum range (Collins & Fahrig, 2017). Collins and Fahrig (2017) and Boissinot et al. (2019) show that landscape variables affect Anuran occupancy and diversity at this scale in agriculture-dominated regions. We use the same environmental variables for all species (see Table 1) except *B. spinosus* for which "pond density" (water point $<5,000 \text{ m}^2$) was substituted by "water point density" because of this species' ability to reproduce in larger water bodies with fish (Boissinot et al., 2019).

All predictive variables were centered and scaled. The spatial correlation between environmental predictors was investigated using the variance inflation factor (VIF) as a measure of multicollinearity and Pearson's correlation tests with VIF < 6 and r < 0.6 as advised by O'Brien, 2007 (see Appendix S2: Tables S2 and S3).

2.4 | Habitat suitability modeling

2.4.1 | Statistical models

Different modeling algorithms can lead to varying results according to heterogeneous sensitivities and calculation processes (Thuiller et al., 2009). Therefore, consensus models based on multi-modeling approaches (ensemble modeling) can improve final results by reducing "noise" associated with individual model errors (Araujo et al., 2005; Meller et al., 2014; Thuiller et al., 2009). For each species,

we used one regression-based approach (Generalized Additive Models, GAM) and one machine-learning algorithm (Random Forest, RF) to predict and assess habitat suitability within the studied region with 50 bootstrap replicates. Presence points were randomly split 50 times into a training set (70% of the whole dataset) and the remaining 30% were used as testing set for internal evaluation (see 1.4). To construct these models, we used biomod2 package (Thuiller et al., 2009) in R environment v. 3.5.3 (R Development Core Team, 2019).

2.4.2 | Background data and pseudo-absence selection

Modeling habitat suitability for a species with GAM or Random Forest requires both presence and absence data. In order to overcome the problem of missing absence data needed for most SDM, pseudo-absence selection strategies have been developed to select absence data where real absence is most likely (Barbet-Massin et al., 2012; Phillips et al., 2009). We tested three strategies for generating artificial absence points: (s1) simple random selection of background points within the studied region (Guisan et al., 2017); (s2) random pseudo-absence selection excluding known presence points (Engler et al., 2004); (s3) random pseudo-absence selection constrained to take sampling effort into account (see Appendix S3 for method). The latter strategy aimed to select pseudo-absences where true absences were more likely. For this strategy, we considered three main sources of bias in pseudo-absence selection: accessibility, linked to distance from roads or urban areas (Barbet-Massin et al., 2012; Kadmon et al., 2004), attractiveness, relating to oversampling in protected sites or tourist areas (Phillips et al., 2009; Robinson et al., 2018) and observer effort, because certain administrative areas are covered by particularly active nature protection organizations (see Appendix S3). For each strategy, the number of artificial absences was fixed equal

TABLE 1 Environmental variables used for species distribution modeling of each amphibian species in Pays de la Loire region. Associated references are available in the Appendix S2: Table S1

Variable category	Code	Variable description	Original resolution
Climatic	CLIM_1	First axis from a PCA on 12 worldclim variables and altitude	2.5 arc-min/5 km
	CLIM_2	Second axis from a PCA on 12 worldclim variables and altitude	2.5 arc-min/5 km
Land cover	%WOOD_DM	Proportion of deciduous and mixed forest	5 m
	%WOOD_C	Proportion of coniferous forest	5 m
	%CROP	Proportion of crop	20 m
	%PASTURE	Proportion of permanent pasture	20 m
	NB_PONDS	Pond density (or water point density)	5 m
	L_HEDGE	Hedgerow density	5 m
	L_ROAD_1ST	Primary road density outside urban areas	5 m
	L_ROAD_2ND	Secondary road density outside urban areas	5 m
	L_RIVER	Canal and river density	5 m
	%URBAN	Proportion of urban area	20 m
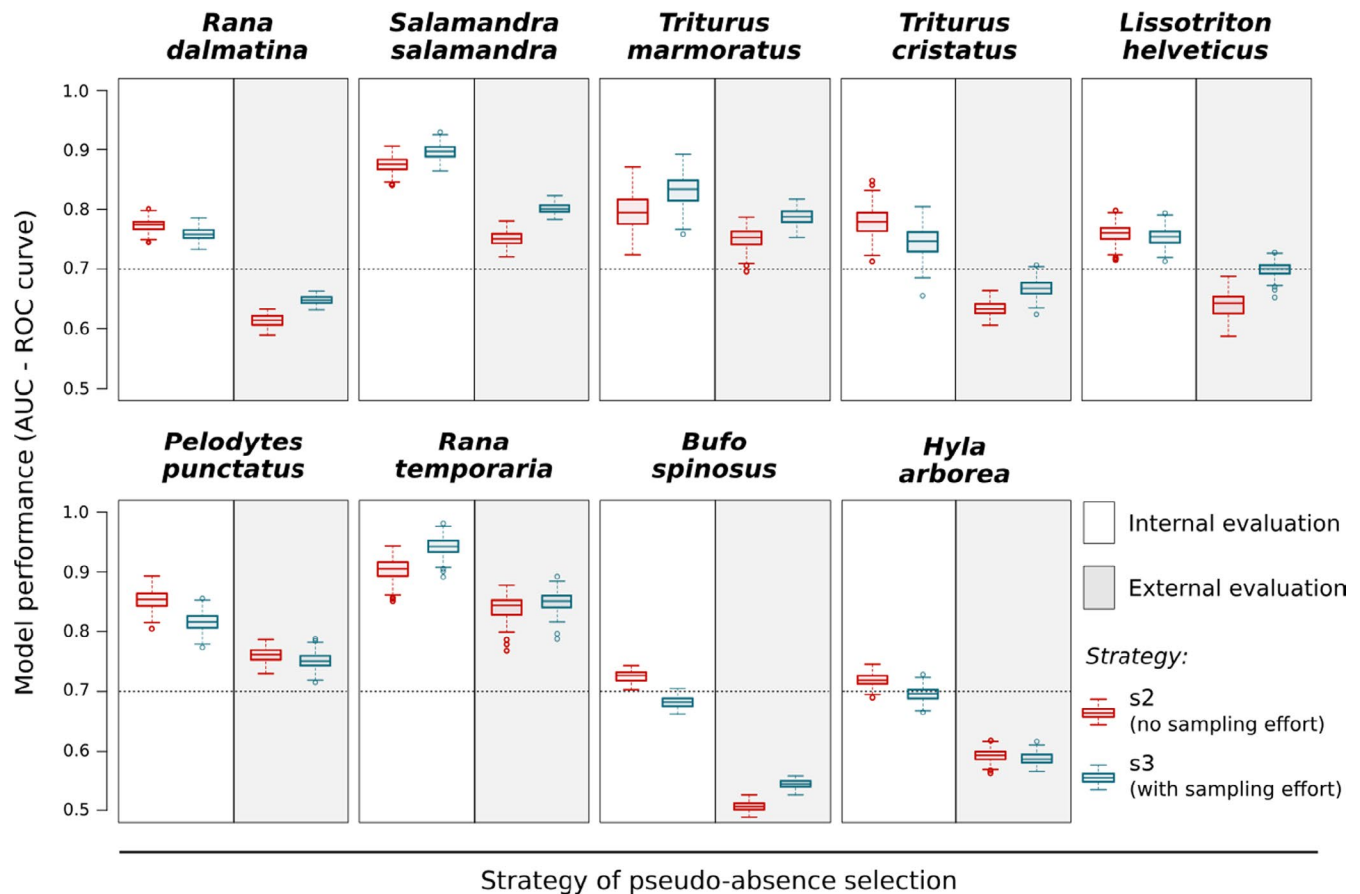
FIGURE 2 Model performance for the 9 studied species assessed by external or internal data using different pseudo-absence selection strategies. Assessment by AUC under the ROC for GAM only are shown (see Appendix S1–S4). Artificial absence sampling strategies shown are s2 (random pseudo-absence selection excluding known presence points) and s3 (random pseudo-absence selection excluding known presence points and adjusted to consider site accessibility and sampling effort). Per strategy, 10 replicates of the artificial absence points generation processes with 50 bootstraps for the random selection of the straining set (70%) and the internal testing set (30%). Black dotted line indicates the 0.70 threshold above which models have an acceptable level of accuracy (Swets 1988)

to the number of presence data (Barbet-Massin et al., 2012; Liu et al., 2019) and we performed 10 replicates of the artificial absence generation processes.

2.4.3 | Ensemble modeling

Finally, we conducted ensemble modeling by calculating the median value of (1) all individual maps generated by GAM and Random forest (i.e. 500 maps/algorithm) (Thuiller et al., 2009) to compare internal versus external evaluation for each species. Second, we also calculated median values from ensemble maps calibrated with 100% of presence-only data to compare different external evaluations sets (i.e. 10 maps/algorithm).

2.5 | Internal and external model validation

We first use a cross-validation method using a 30% random split of the whole set to asses each model (for pseudo-absence

selection strategies s1, s2 and s3) with 50 bootstraps repeated 10 times. We calculated the area under the curve (AUC) of a receiver operating characteristic (ROC) plot of the predicted model habitat suitability scores with (1) the 30% test dataset for internal validation and (2) with the larger filtered external independent dataset (e.g., *CS.2 + ABS + SUP*, see Figure 1) using Biomod2 package (Thuiller et al., 2009). AUC is the most common metric used in SDM studies, as it has the advantage of being threshold and prevalence independent and has been accepted as the standard measure for assessing SDM accuracy (Guisan et al., 2017). AUC > 0.50 signifies that the model has better prediction than a random model.

Second, we calculated AUC values, specificity (true negative rate), and sensitivity (true positive rate) of ensemble models calibrated with 100% of the presence-only data using different evaluation sets derived from the global external dataset used in the previous stage (see Figure 1). These calculations (with 100 bootstraps) were performed using PresenceAbsence package (Freeman 2012) with a standard threshold value for presence–absence discrimination fixed at 0.5.
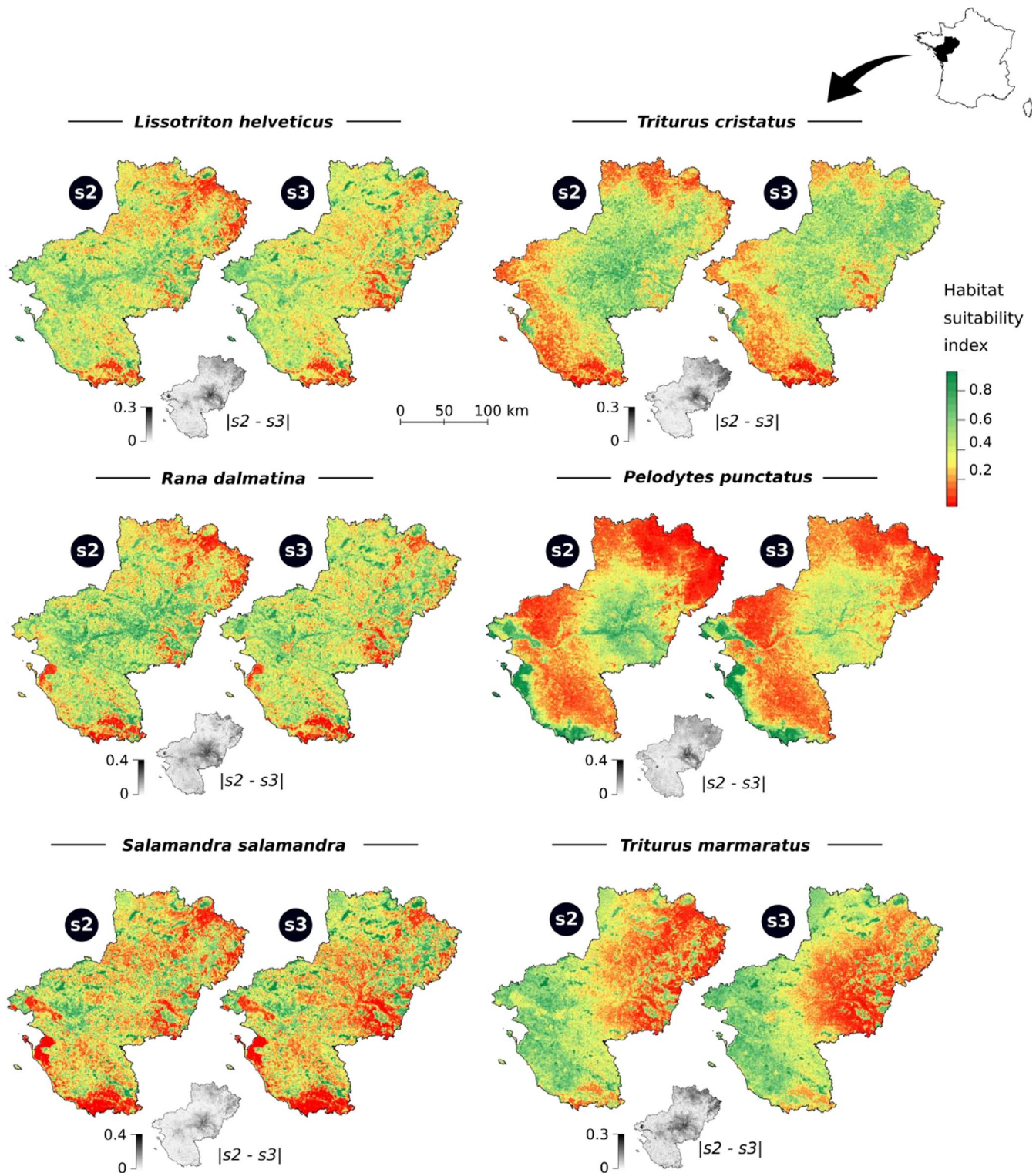
FIGURE 3 Habitat suitability maps for six studied species produced using two form of pseudo-absence selection: s2 (random pseudo-absence selection excluding known presence points) and s3 (random pseudo-absence selection excluding known presence points and constrained to account sampling effort). The black and white map under each pair shows net difference between s2 and s3. Map resolution is 500 m

3 | RESULTS

3.1 | Model performance and selection

For each species, the median AUC was higher with internal validation than external validation for all three pseudo-absence selection

strategies (s1, s2 and s3), both for GAM and Random Forest (see Figure 2 and Appendix S4) with a delta-AUC ranging from 0.05 (*T. marmoratus*) to 0.21 (*B. spinosus*). With internal evaluation (cross-validation), all models had excellent (AUC > 0.90) very good (0.80–0.90) or good accuracy (0.70–0.80) except for the model of *B. spinosus* and *H. arborea* including sampling effort parameters (s3).

However, with external evaluation, only four species had a high level of accuracy (AUC > 0.70): *S. salamandra*, *T. marmoratus*, *P. punctatus*, and *R. temporaria*. For *R. dalmatina*, *T. cristatus*, and *L. helveticus*, model accuracies were poor (0.60 < AUC < 0.70) and for *B. spinosus* and *H. arborea* even poorer (AUC < 0.60). The strategy s1 (background data) was not selected neither with internal nor external evaluation.

The method used for pseudo-absence selection influenced the predictive performance of models, but differences between AUC values were minimal (Figure 2). However, s2 (uncorrected sampling bias) was the best strategy for six species when internal validation was used, while s3 was best for seven species when external validation was used. Results for RF can be found in the supplementary material but do not differ greatly (Appendix S4: Figure S1).

3.2 | Impact of model selection on final habitat suitability map

Internal or external validation resulted in different models being selected, based on AUC comparison. Therefore, the final habitat suitability maps selected by each of these two assessment methods would lead to different interpretation and conservation decisions (Figure 3). Maps for *H. arborea* and *B. spinosus* are not shown because of poor accuracy (see supplementary results Appendix S4). All response curves and associated variable contributions can be found in the supplementary material (Appendix S4: Figures S2 and S3).

3.3 | Comparison of external evaluation sets

Values of AUC, sensitivity, and specificity to four species are shown in Table 2 (two Anurans and two Urodeles; one forest species and one generalist specie each). Results for other species and CS.1 (similar to CS.2) are presented in Appendix S4: Table S3. Considering AUC values, evaluation with the external dataset from participative science without filter data (*CS.0*) show more similar model selection results than internal cross-validation except for *R. temporaria*. Sorting presence data led to decreased sensitivity and increased specificity for all species except for *S. Salamandra*. The models selected (s2 or s3) were similar for most species whether using stratified data from volunteers' only (STRAT_CS) or stratified data with added professional observations (STRAT_ALL), or professional data only (PRO). See Table 2 and Appendix S4. We excluded the s1 model from the comparison because this model is never selected, either with internal or external evaluation.

4 | DISCUSSION

External evaluation with independent data generated lower AUC values than cross-validation, which calls into question the validity of models validated by commonly used selection threshold values such

as AUC > 0.70. According to Araujo et al. (2005), internal evaluation with nonindependent data always leads to overoptimistic assessment of model performance. Even if cross-validation is better than substitution procedures (Araujo et al., 2005; Edwards et al., 2006; Vaughan & Ormerod, 2005), split data for internal validation are nonindependent and do not avoid the main limits of correlative models in SDMs because of spatial or temporal autocorrelation, especially when sampling effort is highly heterogeneous (Edwards et al., 2006; Roberts et al., 2017). Our result supports criticisms of certain types of SDM and further highlights the need to be careful in their general interpretation and assessment (Lobo et al., 2008).

The difference between internal-AUC and external-AUC is particularly pronounced for the most common and generalist species in spite of the large number of available data, especially for *B. spinosus* and *H. arborea* whose models failed to attain an acceptable level of accuracy with external evaluation. Brotons et al. (2004) highlighted the difficulty of predicting distributions of the most generalist species. However, for such species, the use of filters increases specificity values considerably and the results are coherent with these species' wide ranging and ubiquitous distributions. Using external presence–absence data also makes it possible to exploit the whole presence-only dataset for calibration and to use stricter filters to reduce sampling bias or data culling to retain higher quality data (Isaak et al., 2014; Steen et al., 2019). Our study shows that it is possible to apply strong filters (e.g., STRAT_CS) but finally retain reasonable sample sizes for most species.

It should be noted that, for four species, *R. temporaria*, *T. marmoratus*, *S. salamandra*, and *P. punctatus*, our results were ambiguous. For the first three, all forest-dwelling species or very closely related to woodlands (Boissinot et al., 2019), both internal and external validation methods selected models with sampling effort integrated (s3). Two main reasons could explain these results: first, presence data may have been insufficient for *R. temporaria*, and second these species' affinity for forest habitats. *R. temporaria* is a rare species and is more dependent on wet forest, flood meadows, and small streams as breeding sites than other species (Boissinot et al., 2015). This species has a patchy distribution (i.e. locally abundant but regionally rare) and is difficult to detect. Hence, presence data are few in the studied area both in the opportunist dataset (*N* = 477 presence-cells) and in the independent dataset (*N* = 13 presence-cells). As highlighted by c, such factors can lead to model over fit, even with a relatively small number of variables, resulting in high AUC values. In addition, according to Brotons et al. (2004), low-density habitat (i.e. forest habitat in our region) may be overweighted and it can be difficult to assess between good or bad suitability without adapted presence–absence data. Monitored forest sites are rare in our validation dataset, and the assumptions we used to define sampling effort may not be well adapted for forest specialists. Finally, *P. punctatus* is a rare species but abundant on the Atlantic coast. Unlike the other species, it is a pioneer, adapted to open areas, especially primary unvegetated habitats such as sand dunes and mudflats with frequent physical disturbance (Joly et al., 2005). These habitats are mainly located near the coast and along the main regional floodplains (Loire Valley),

TABLE 2 Model performance according to different filters and complementary fieldwork applied to the external evaluation dataset

		Triturus marmoratus			*Lissotriton helveticus*			*Hyla arborea*			*Rana temporaria*		
		SEN	SPE	AUC	SEN	SPE	AUC	SEN	SPE	AUC	SEN	SPE	AUC
CS.0	s2	0.63	0.41	0.58	**0.69**	0.29	0.53	**0.82**	0.43	0.68	0.71	**0.87**	0.87
	s3	**0.65**	**0.45**	**0.61**	0.64	**0.34**	0.54	0.78	**0.49**	0.68	**0.86**	0.84	0.88
CS.2 + ABS	s2	0.59	0.78	0.81	**0.67**	0.76	0.80	**0.71**	0.61	0.73	0.64	**0.88**	0.86
	s3	0.58	**0.87**	0.81	0.62	**0.88**	**0.86**	0.68	**0.64**	0.74	**0.82**	0.84	0.86
CS.2 + ABS + SUP	s2	0.58	**0.82**	0.82	**0.51**	0.77	0.68	**0.63**	**0.56**	0.63	0.59	**0.91**	0.85
	s3	**0.62**	0.78	0.81	0.49	0.78	**0.74**	0.57	0.54	0.63	**0.76**	0.86	0.85
PRO	s2	0.58	**0.83**	**0.85**	**0.41**	0.76	0.61	**0.59**	**0.39**	0.51	*0.50*	*0.95*	*0.71*
	s3	**0.79**	0.77	0.83	0.37	**0.79**	**0.64**	0.51	0.33	**0.54**	*0.50*	*0.92*	*0.71*
STRAT_CS	s2	0.71	**0.72**	**0.81**	0.67	0.81	0.82	0.72	0.64	0.73	0.79	0.57	0.69
	s3	0.70	0.69	0.78	0.68	0.82	**0.88**	0.71	0.63	**0.75**	**0.95**	**0.60**	**0.79**
STRAT_ALL	s2	**0.80**	**0.74**	**0.87**	0.66	0.76	0.78	**0.68**	**0.62**	0.70	0.84	0.57	0.73
	s3	0.78	0.70	0.83	**0.71**	**0.83**	**0.89**	0.66	0.60	**0.72**	**0.96**	0.58	**0.81**

Note: External datasets used were (see Figure 1): *CS.0* (all data from the standardized citizen science dataset); *CS.2 + ABS* (*CS.2* with 10% supplementary absence cells in very unfavorable habitats); *PRO* (data collected by professionals only in 2018–2019); *CS.2 + ABS + SUP* (citizen science data cited before adding all complementary fieldwork by professionals and volunteers); *STRAT_CS* (stratified data selection from *CS.2 + ABS* with complementary fieldwork by volunteers); *STRAT_ALL* (stratified data selection from *CS.2 + ABS + SUP*). Models assessed: s2 (random pseudo-absence selection excluding known presence points) and s3 (random pseudo-absence selection constrained to account sampling effort and correct sampling bias). SEN, sensitivity; SPE, specificity. Bold values show best values between s2 et s3 with delta >0.02 and italic values show species with less than 2 presence data. All analyses with a random sampling in presence selection with a distance condition or a stratified random selection were performed using 100 bootstraps (mean calculation).

with a high local density of presence data. So this species is also patchily distributed and models may be affected by the same bias as *R. temporaria*. Alternatively, the similarity between AUC values may also relate to sampling effort bias along the Atlantic coast (e.g., Fithian et al., 2015). There results highlight the need to adapt methods and filters used for each species.

4.1 | Using heterogeneous data from citizen science in SDM

Our results show that it is possible to obtain useful external and independent datasets for model validation from filtered standardized citizen science data. Indeed, the use of filters has successfully reduce bias and noise in citizen science datasets for SDM in others studies (Isaac et al., 2019; Robinson et al., 2018; Steen et al., 2019). In addition, filtered evaluation dataset showed coherent results according to Phillips et al. (2009). Indeed, choosing pseudo-absence data with the same bias as occurrence data improved model performance.

Since external independent data are necessary for more robust assessment of SDM (Araujo et al., 2005), but prohibitive to collect, filtering low quality but large datasets from monitoring to obtain more standard and independent data may be worthwhile. In addition, AUC appears to be more informative when presence–absence data are used to assess and compare models than when presence-background data alone is used (Jiménez-Valverde, 2012). However, using detection–nondetection citizen data without filters may also lead to erroneous results because of overlapping sources of bias in

both datasets (e.g., *CS.0* selects the same model as cross-validation). The large amount of available data allows us to strongly select data according to our research objective. Our results using *PRO* datasets are inconclusive for rare species perhaps because their detection was insufficiently frequent (e.g., only two observations of *R. temporaria* for 132 sampled sites). Finally, we found that general rules to guide data sorting were difficult to define. Our results were sensitive to the type of data used, and the species studied, reinforcing the need to define filters according to available data and species' ecology (Steen et al., 2019).

Independence between training and testing sets is an essential criterion, but data should also be unbiased or corrected. Selection methods have been developed to try to divide the opportunistic dataset strategically to increase the independence of the testing set for cross-validation (see Block-cross-validation in Robert et al., 2017). However, this method does not make it possible to escape from the general biases linked to sampling effort and/or can create extrapolation problems (see Roberts et al., 2017). Using a more standardized dataset from a participatory science program (e.g., CS0) for the evaluation makes it easier to understand the sources of bias (presence of metadata and nondetection data), to better control them and to obtain more robust information on the absence data. However, these data may also share biases with the opportunistic dataset used for calibration. In our case, the sampling of the monitored sites (*CS0*) was partly biased towards volunteers' place of residence and areas with a higher density of observers. These biases were reduced through additional sampling involving volunteers. Our results show that certain filters, as well as targeted complementary fieldwork,

make it possible to reduce the biases identified and produce conclusive results. In addition, the use of a stratified sampling of the testing set along the suitability gradient from the SDM results (e.g., our STRAT_CS and STRAT_ALL datasets) appears to be a particularly interesting method showing stable and consistent results according to Phillips et al. (2009).

However, our method may be not applicable in all cases. In our study, external data came from a program with general population monitoring objectives, using standard methods designed to be accessible to a wide audience (e.g., novice and professionals). This program concerns all amphibians and their habitats, whereas many citizen science programs are limited to a particular taxonomic group or habitat type (cities, gardens or farms…) and would therefore be difficult to extrapolate to wider contexts.

4.2 | Involved stakeholders and citizens in conservation research

Our study was part of a wider project for amphibian conservation in the Pays-de-la-Loire region of western France. Involving citizens in the SDM evaluation process may make conservation action easier to implement, through both better shared knowledge and stronger personal involvement. Forrester et al. (2017) and Lewandowski and Oberhauser (2017) highlighted an increase in conservation advocacy among participants of citizen science projects that might improve access to evidence for conservationists and decision makers (Sutherland & Wordley, 2017). Maps are a specially a good tool for improving communication between researchers and volunteers in the context of citizen science (Zapponi et al., 2017). Indeed, many nature protection organizations are already involved in distribution atlas projects and naturalists are aware of data collection methods and local species distributions. They seek out ways to prioritize field observations; making a useful contribution to developing SDMs to guide conservation action can be a source of motivation, making scientist-volunteer interactions easier.

AUTHOR CONTRIBUTIONS
Florence Matutini: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Funding acquisition (supporting); Investigation (lead); Methodology (lead); Visualization (lead); Writing-original draft (lead); Writing-review & editing (lead). **Jacques Baudry:** Conceptualization (equal); Methodology (supporting); Supervision (equal); Writing-original draft (supporting); Writing-review & editing (supporting). **Guillaume Pain:** Conceptualization (equal); Funding acquisition (lead); Methodology (supporting); Supervision (equal); Writing-original draft (supporting); Writing-review & editing (supporting). **Morgane Sineau:** Data curation (equal); Funding acquisition (supporting); Writing-review & editing (supporting). **Josephine Pithon:** Conceptualization (equal); Funding acquisition (supporting); Methodology (supporting); Supervision (equal); Writing-original draft (supporting); Writing-review & editing (supporting).

ORCID
Florence Matutini https://orcid.org/0000-0002-7013-929X

REFERENCES
Araujo, M., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology*, *11*(9), 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.001000.x

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, *3*(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., & Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, *222*(11), 1810–1819. https://doi.org/10.1016/j.ecolmodel.2011.02.011

Baudry, J., Bunce, R. G. H., & Burel, F. (2000). Hedgerows: An international perspective on their origin, function and management. *Journal of Environmental Management*, *60*(1), 7–22. https://doi.org/10.1006/jema.2000.0358

Boissinot, A. (2009). *Influence de la structure du biotope de reproduction et de l'agencement du paysage, sur le peuplement d'amphibiens d'une région bocagère de l'ouest de la France, Diplôme de l'Ecole Pratique des Hautes Etudes*. Laboratoire de Biogéographie et d'ecologie Des Vertébrés (EPHE/CEFE/CNRS).

Boissinot, A., Besnard, A., & Lourdais, O. (2019) Agriculture, Ecosystems and Environment Amphibian diversity in farmlands: Combined in

fluences of breeding-site and landscape attributes in western France. *Agriculture, Ecosystems and Environment*, *269*, 51–61. https://doi.org/10.1016/j.agee.2018.09.016

Boissinot, A., Grillet, P., Besnard, A., & Lourdais, O. (2015). Small woods positively influence the occurrence and abundance of the common frog (*Rana temporaria*) in a traditional farming landscape. *Amphibia Reptilia*, *36*(4), 417–424. https://doi.org/10.1163/15685381-00003013

Brotons, L., Thuiller, W., Araújo, M. B., & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, *27*(4), 437–448. https://doi.org/10.1111/j.0906-7590.2004.03764.x

Brown, J. L., Sillero, N., Glaw, F., Bora, P., Vieites, D. R., & Vences, M. (2016). Spatial biodiversity patterns of Madagascar's amphibians and reptiles. *PLoS One*, *11*(1), 1–26. https://doi.org/10.1371/journal.pone.0144076

Collins, S. J., & Fahrig, L. (2017). Responses of anurans to composition and configuration of agricultural landscapes. *Agriculture, Ecosystems and Environment*, *239*, 399–409. https://doi.org/10.1016/j.agee.2016.12.038

Cosentino, B. J., Marsh, D. M., Jones, K. S., Apodaca, J. J., Bates, C., Beach, J., Beard, K. H., Becklin, K., Bell, J. M., Crockett, C., Fawson, G., Fjelsted, J., Forys, E. A., Genet, K. S., Grover, M., Holmes, J., Indeck, K., Karraker, N. E., Kilpatrick, E. S., … Willey, A. (2014). Citizen science reveals widespread negative effects of roads on amphibian distributions. *Biological Conservation*, *180*, 31–38. https://doi.org/10.1016/j.biocon.2014.09.027

Cushman, S. A. (2006). Effects of habitat loss and fragmentation on amphibians: A review and prospectus. *Biological Conservation*, *128*(2), 231–240. https://doi.org/10.1016/j.biocon.2005.09.031

de Solla, S. R., Shirose, L. J., Fernie, K. J., Barrett, G. C., Brousseau, C. S., & Bishop, C. A. (2005). Effect of sampling effort and species detectability on volunteer based anuran monitoring programs. *Biological Conservation*, *121*, 585–594. https://doi.org/10.1016/j.biocon.2004.06.018

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Edwards, T. C., Cutler, D. R., Zimmermann, N. E., Geiser, L., & Moisen, G. G. (2006). Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling*, *199*, 132–141. https://doi.org/10.1016/j.ecolmodel.2006.05.016

Engler, R., Guisan, A., & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *The Journal of Applied Ecology*, *41*, 263–274. https://doi.org/10.1111/j.0021-8901.2004.00881.x

Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, *6*(4), 424–438. https://doi.org/10.1111/2041-210X.12242

Forrester, T. D., Baker, M., Costello, R., Kays, R., Parsons, A. W., & McShea, W. J. (2017). Creating advocates for mammal conservation through citizen science. *Biological Conservation*, *208*(6), 98–105. https://doi.org/10.1016/j.biocon.2016.06.025

Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B. O., Olsen, K., Rahbek, C., & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, *22*(11), 1139–1149. https://doi.org/10.1111/ddi.12477

Graham, C. H., & Hijmans, R. J. (2006). A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, *15*(6), 578–587. https://doi.org/10.1111/j.1466-8238.2006.00257.x

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, *24*(3), 276–292. https://doi.org/10.1111/geb.12268

Guisan, A., & Theurillat, J.-P. (2000). Assessing alpine plant vulnerability to climate change: A modeling perspective. *Integrated Assessment*, *1*(4), 307–320. https://doi.org/10.1023/A:1018912114948

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R, ecology, biodiversity and conservation*. The University of Chicago Press.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O. L., Schmucki, R., Simmonds, E. G., & O'Hara, R. B. (2019). Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology and Evolution*, *35*(1), 56–67. https://doi.org/10.1016/j.tree.2019.08.006

Isaac, N. J. B., Strien, A. J., August, T. A., Zeeuw, M. P., & Roy, D. B. (2014). Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, *5*(10), 1052–1060. https://doi.org/10.1111/2041-210X.12254

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, *21*(4), 498–507. https://doi.org/10.1111/j.1466-8238.2011.00683.x

Joly, P., Morand, A., Plenet, S., & Grolet, O. (2005). Canalization of size at metamorphosis despite temperature and density variations in *Pelodytes punctatus*. *Herpetological Journal*, *15*(1), 45–50.

Kadmon, R., Farber, O., & Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, *14*(2), 401–413. https://doi.org/10.1890/02-5364

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *14*(10), 551–560. https://doi.org/10.1002/fee.1436

Lewandowski, E. J., & Oberhauser, K. S. (2017). Butterfly citizen scientists in the United States increase their engagement in conservation. *Biological Conservation*, *208*, 106–112. https://doi.org/10.1016/j.biocon.2015.07.029

Liu, C., Newell, G., & White, M. (2019). The effect of sample size on the accuracy of species distribution models: Considering both presences and pseudo-absences or background sites. *Ecography*, *42*(3), 535–548. https://doi.org/10.1111/ecog.03188

Lobo, J. M., Jiménez-valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*, 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., Evans, D. M., French, R. A., Parrish, J. K., Phillips, T. B., Ryan, S. F., Shanley, L. A., Shirk, J. L., Stepenuck, K. F., Weltzin, J. F., Wiggins, A., Boyle, O. D., Briggs, R. D., Chapin, S. F., … Soukup, M. A. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, *208*, 15–28. https://doi.org/10.1016/j.biocon.2016.05.015

Meller, L., Cabeza, M., Pironon, S., Barbet-Massin, M., Maiorano, L., Georges, D., & Thuiller, W. (2014). Ensemble distribution models in conservation prioritization: From consensus predictions to consensus reserve networks. *Diversity and Distributions*, *20*(3), 309–321. https://doi.org/10.1111/ddi.12162

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Petitot, M., Manceau, N., Geniez, P., & Besnard, A. (2014). Optimizing occupancy surveys by maximizing detection probability: Application to amphibian monitoring in the Mediterranean region. *Ecology and Evolution*, 4(18), 3538–3549. https://doi.org/10.1002/ece3.1207

Petrovan, S. O., & Schmidt, B. R. (2016). Volunteer conservation action data reveals large-scale and long-term negative population trends of a widespread amphibian, the common toad (*Bufo bufo*). *PLoS One*, 11(10), 1–12. https://doi.org/10.1371/journal.pone.0161943

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197. https://doi.org/10.1890/07-2153.1

Préau, C., Isselin-Nondedeu, F., Sellier, Y., Bertrand, R., & Grandjean, F. (2019). Predicting suitable habitats of four range margin amphibians under climate and land-use changes in southwestern France. *Regional Environmental Change*, 19(1), 27–38. https://doi.org/10.1007/s10113-018-1381-z

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. https://doi.org/10.1111/ecog.02881

Robinson, O. J., Rui-Gutierrez, V., Reynolds, M. D., Golet, G. H., Strimas-Mackey, M., & Fink, M. (2020) Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. *Diversity and Distributions*, 2019, 806547. https://doi.org/10.1101/806547

Robinson, O. J., Ruiz-Gutierrez, V., & Fink, D. (2018). Correcting for bias in distribution modelling for rare species using citizen science data. *Diversity and Distributions*, 24(4), 460–472. https://doi.org/10.1111/ddi.12698

Schmeller, D. S., Henry, P.-Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., Lengyel, S., Nowicki, P., Déri, E., Budrys, E., Kull, T., Tali, K., Bauch, B., Settele, J., Van swaay, C., Kobler, A., Babij, V., Papastergiadou, E., & Henle, K. (2009). Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology*, 23(2), 307–316. https://doi.org/10.1111/j.1523-1739.2008.01125.x

Schulman, L., Toivonen, T., & Ruokolainen, K. (2007). Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. *Journal of Biogeography*, 34(8), 1388–1399. https://doi.org/10.1111/j.1365-2699.2007.01716.x

Semlitsch, R. D., & Bodie, J. R. (2003). Biological criteria for buffer zones around wetlands and riparian habitats for amphibians and reptiles. *Conservation Biology*, 17(5), 1219–1228. https://doi.org/10.1046/j.1523-1739.2003.02177.x

Sewell, D., Beebee, T. J. C., & Griffiths, R. A. (2010). Optimising biodiversity assessments by volunteers: The application of occupancy modelling to large-scale amphibian surveys. *Biological Conservation*, 143(9), 2102–2110. https://doi.org/10.1016/j.biocon.2010.05.019

Sinsch, U. (1990). Migration and orientation in anuran amphibians. *Ethology Ecology and Evolution*, 2(1), 65–79. https://doi.org/10.1080/08927014.1990.9525494

Steen, V. A., Elphick, C. S., & Tingley, M. W. (2019). An evaluation of stringent filtering to improve species distribution models from citizen science data. *Diversity and Distributions*, 25(12), 1857–1869. https://doi.org/10.1111/ddi.12985

Stuart, S. N., Chanson, J. S., Cox, N. A., Young, B. E., Rodrigues, A. S. L., Fischman, D. L., & Waller, R. W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, 306(5702), 1783–1786. https://doi.org/10.1126/science.1103538

Sutherland, W. J., & Wordley, C. F. R. (2017). Evidence complacency hampers conservation. *Nature Ecology & Evolution*, 1(9), 1215–1216. https://doi.org/10.1038/s41559-017-0244-1

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD - A platform for ensemble forecasting of species distributions. *Ecography*, 32(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Vaughan, I. P., & Ormerod, S. J. (2005). The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, 42(4), 720–730. https://doi.org/10.1111/j.1365-2664.2005.01052.x

Zapponi, L., Cini, A., Bardiani, M., Hardersen, S., Maura, M., Maurizi, E., Redolfi De Zan, L., Audisio, P., Bologna, M. A., Carpaneto, G. M., Roversi, P. F., Sabbatini Peverieri, G., Mason, F., & Campanaro, A. (2017). Citizen science data as an efficient tool for mapping protected saproxylic beetles. *Biological Conservation*, 208, 139–145. https://doi.org/10.1016/j.biocon.2016.04.035

Zipkin, E. F., & Saunders, S. P. (2017). Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation*, 217, 240–250. https://doi.org/10.1016/j.biocon.2017.10.017

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Matutini F, Baudry J, Pain G, Sineau M, Pithon J. How citizen science could improve species distribution models and their independent assessment. *Ecol Evol*. 2021;11:3028–3039. https://doi.org/10.1002/ece3.7210