Article (refereed) - postprint

Balance between site fidelity and habitat preferences in colony site selection by herons and egrets

Luis Carrasco[1], Yukihiko Toquenaga[2] and Miyuki Mashiko[3]

[1]Centre for Ecology and Hydrology, Library Avenue, Lancaster Environmental Centre, Lancaster, Bailrigg LA1 4AP, UK.

[2]Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki 305-8572, Japan.

[3]Wildlife Management Office, Inc., 1-10-13 Oyamagaoka, Machida, Tokyo 194-0215, Japan.

Corresponding author: Luis Carrasco, Centre for Ecology and Hydrology, Library Avenue, Lancaster Environmental Centre, Lancaster, Bailrigg LA1 4AP, UK. E-mail: luirra@ceh.ac.uk

Habitat selection in avian species is a hierarchical process driven by different factors acting at multiple scales. Habitat preferences and site fidelity are two main factors affecting how colonial birds choose their breeding locations. Although these two factors affect how colonial species choose their habitats, previous studies have only focused on one factor at a time to explain the distribution of species at regional scales. Here we used 28 years of colony location data of herons and egrets around Ibaraki prefecture in Japan in order to analyze the relative importance of habitat preferences and colony site fidelity for selecting breeding locations. We used Landsat satellite images together with a ground survey-based map to create land-use maps for past years and determine the habitats surrounding the herons and egrets colonies. Combining the estimated colony site fidelity with the habitat data, we used a random forest algorithm to create habitat selection models, which allowed us to analyze the changes in the importance of those factors over the years. We found high levels of colony site fidelity for each year of study, with its relative importance as a predictor for explaining colony distribution increasing drastically in the most recent five years. The increase in collective site fidelity could have been caused by recent changes in the population size of grey herons *Ardea cinerea*, a key species for colony establishment. We observed a balance between habitat preferences and colony site fidelity: habitat preferences were a more powerful predictor of colony distribution until 2008, when colony site fidelity levels were lower. Considering changes in the relative importance of these factors can lead to a better understanding of the habitat selection process and help to analyze bird species' responses to environmental changes.

Habitat selection in avian species is a hierarchical process involving various decisions at different scales (Cody 1985, Wiens 1989, Jones 2001). Birds can select their habitats by assessing the quality or quantity of the different available habitat patches (Fretwell and Lucas 1969, Esely and Bollinger 2001, Tarvin and Garvin 2002). Other factors that can affect habitat selection are conspecific or heterospecific attraction (Danchin et al. 1998, Thomson et al. 2003), avoiding competition (Immelmann 1972, Petit and Petit 1996, Haynes et al. 2014), or anti-predation measures (Cody 1985, Martin 1993). Site fidelity, i.e., returning to natal sites or previously used habitats (Greenwood 1980, Greenwood and Harvey 1982), can also be adopted by a wide range of avian species, due to the benefits of previous knowledge of the site and for avoiding assessment efforts (Forbes and Kaiser 1994).

Habitat preferences and site fidelity are habitat selection factors that can be studied with distribution data only; they may also combine to affect how species choose their living habitats (Jones 2001). As it is often difficult to grasp both factors at once (Bollinger 1995, Ganter and Cooke 1998), generally due to the costly efforts of gathering site fidelity information, many ecologists have tried to explain the distribution of the species at regional scales by focusing on only one of them (Warnock and Takekawa 1995, Ganter and Cooke 1998, Fox et al. 2005). However, studying the relative importance of habitat preferences and site fidelity simultaneously can be helpful in understanding more broadly the habitat selection process. Differences in environmental factors, such as temporal changes in habitats and population densities of birds, could affect the relative importance of each of the two factors within a species.

Combining satellite images with up-to-date land-use data enables us to investigate changes in habitat selection over time with the help of long-term data on species distributions (Magurran et al. 2002). Studies on diverse avian families showed changes in habitat preferences (Clark and Shutler 1999, Dale and Christiansen 2010) or changes in site fidelity (Calder et al. 1983, Ganter and Cooke 1998) over the years. However, none of them examined how the balance between habitat preferences and site fidelity, the relative importance of each of these factors for habitat selection, changes over time. If habitat preferences and site fidelity change over the years, their effect on habitat selection could compensate each other. Therefore, the study of the balance between these two factors is important, especially in changing landscapes, in which adverse or preferable habitats for the species are rapidly altered, such that drastic changes in avian population dynamics can occur.

Colonial bird species exhibit habitat preferences and site fidelity when choosing their breeding sites (Lack 1968, Cézilly et al. 2000, Bried et al. 2003). Colonial birds employ collective decision-making, i.e., consensus decisions in which the group makes a single collective choice, in establishing new colony locations and returning to previously used colonies (Deneubourg and Goss 1989, Couzin 2009). As direct observation of collective decisions is difficult (Kerth 2010), many authors have studied how new breeding locations were selected by modeling preferred or avoided habitats for different spatial scales (Heinänen et al. 2008, Kelly et al. 2008) or for different spatial configurations (Carrasco et al. 2015). Despite the fact that many colonial birds, such as many seabirds and ardeids

species, show breeding site fidelity, none of the aforementioned studies include site fidelity within their statistical models.

Many seabirds reuse the same colony sites at high rates due to the limitation of available breeding locations (Lack 1968). Throughout this paper we use the term "colony site fidelity" to refer to the repeated use of the same breeding location year after year. Despite high colony site fidelity, individual site fidelity for some seabirds is low. For example, herring gulls *Larus argentatus* have shown a 30% return tendency among young survivors to natal colony locations (Chabrzyk and Coulson 1976), while black-legged kittiwake *Rissa tridactyla* showed a 23% rate in a study spanning 36 years (Coulson and Coulson 2008). Individual site fidelity rates for ardeids (terrestrial species such as herons and egrets) are even lower: from 23.8% in great egrets *Ardea alba* to 7.9% in little blue herons *Egretta caerulea* (Melvin et al. 1999). Available breeding sites for ardeids are abundant, as they can build their nests on the top of vegetation or in the ground (Kushlan and Hancock 2005). Despite the abundance of available breeding sites and the low individual return rates, the colonies of most of ardeid species can last for years in the same locations (Kushlan and Hancock 2005). In a study spanning 51 years of colony locations of herons and egrets, Mashiko and Toquenaga (submitted) showed that two thirds of the colonies exhibited persistent use of the same or nearby breeding sites. Although different intensity levels in colony site fidelity could result on differences in colony site selection, how these levels change over time have not been well documented. Studying changes in colony site fidelity and its relationship to the habitat preferences is important in clarifying how this factor can affect the colony site selection process.

To study the relationship between habitat preference and colony site fidelity in colonial birds, we analyzed the relative importance of environmental variables and the reuse of colony locations in heron and egret communities over a span of 28 years. We used long-term data for colony locations of herons and egrets around Ibaraki prefecture, in Japan, together with satellite images from previous years to reconstruct the land uses surrounding the colonies. First, we expected colony site fidelity to be the main factor for explaining colony site selection, as herons and egrets show persistent use of breeding sites. Second, we hypothesized that the relative importance between environmental variables and colony site fidelity would stay constant over the years.

Methods

Study area and species

The study area is located in the eastern part of the Kanto Plain in central Japan (Fig. 1). The area covers most of Ibaraki prefecture and part of some bordering parts of Tochigi, Saitama and Chiba prefectures, extending about 10,000 km². The landscape in the study area is a complex mosaic of paddy fields, crops, forests, water bodies, urban areas and other strongly human-influenced habitat patches.

Six species of herons and egrets settle their annual mixed-species colonies across the study area from late February to mid May. The grey heron, the species not present in the study area until the mid 1990's (Mashiko and Toquenaga 2014), is the first species to arrive at the breeding locations, followed by the great egret, little egret *E. garzetta* and black-crowned night heron *Nycticorax nycticorax*. Finally, the intermediate egret *E. intermedia* and cattle egret *Bubulcus ibis*, the

migrant species, arrive by early May. They build their nest on the top of trees and in bamboo thickets and forage mainly in paddy fields and other flooded fields (Environmental Agency of Japan 1994). Colonies break up from August to September and no colony sites were used for winter roosting for most of the years of study.

Colony locations

Colony-location data from 1986 to 2013 in the study area (Fig. 1) were used. Colony locations from 1986 to 1998 were obtained from Koshida (2007). The study area was region-wide aerial and ground surveyed annually during the breeding seasons from 1999 to 2013 (Fujioka et al. 2001, Mashiko 2014) and the total number of annual colonies and each colony location was recorded. From March to May, most of the colonies were found by checking previous years locations. If a colony had been abandoned, the surroundings were surveyed to determine whether other colonies had been established nearby. For some years, additional colony locations were identified by referring to local literature or to personal communications of local amateur avian researchers. Every year, the majority of colonies were reestablished at the same locations as previous years, and other colonies were relocated near to those of previous years or disappeared. Relocation or extinction of colonies can be due to human disturbance (Mashiko and Toquenaga 2013), but some colonies disappear in subsequent years even though there was no human disturbance or obvious changes in nesting vegetation at the previous colony locations. The number of colonies recorded each year fluctuated from 14 to 29 (20.4 ± 4.0).

Land-use maps

Five late spring or summer Landsat TM/ETM+ images of the years 1989, 1993, 1999, 2005 and 2011 (years with cloud-free images for the desired period) were selected for creating land-use maps of the study area. The Landsat images were calibrated and geometrically rectified and then an atmospheric correction was applied using a dark subtraction algorithm (Chavez 1996). Clouds were identified with the thermal band of the Landsat images, as clouds are much less reflective than the rest of the land uses at those wavelengths, and then masked. The *landsat* package (Goslee 2011) in R 3.0.2 (R Development Core Team 2013) was used for image processing.

A supervised classification was used to create land-use maps of the study area from the processed Landsat images. Four land-use variables were selected due to their high identifiability from satellite images and to their relevance for habitat preferences modelling of herons and egrets from previous studies (Kelly et al. 2008, Fasola et al. 2010, Carrasco et al. 2014): water bodies, paddy fields, forests and urban areas. Based on an available land-use map of 2011 (provided by the Japanese Aerospace Exploration Agency), we identified samples of pixels (regions of interests, *i.e.* ROI) representing each land-use class on the 2011 Landsat satellite image. The spectral signatures of the four land uses in the Landsat image of 2011 were used to extrapolate and identify ROI for the rest of the Landsat images. Some of the images presented regions with snow; in these cases snowy regions were also identified and classified, although this land cover was not used in posterior analysis. Once the ROI were selected, a land-use classification of the whole image

for each year was applied using a regression tree algorithm (Breiman et al. 1984). The flow of the land-use maps classification is represented in Fig. A1 (Supplementary material Appendix 1). Selection of ROI and extraction of the spectral information was done using Bio7 software (Austenfeld and Beyschlag 2012) and the regression tree classification was performed using the package *rpart* (Therneau and Atkinson 2007) in R.

Colony site selection models

Predictor variables

We created predictive models of colony site selection by comparing two different types of variables between colony sites and pseudo-absences (randomly selected unoccupied sites). The first variable was a measure of the site fidelity of each colony in respect to the previous year's colony distribution. An estimate based on the nearest neighbor distances (*NND*, see Clark and Evans 1954) of a target colony from other colonies in previous years, was constructed. We used the *NND* of a colony in year t against all colonies in year $t-1$. *NND* is zero if the colony persisted at the same site in both year $t-1$ and *t*. If the colony did not persist at the same site in year $t-1$, *NND* is the nonzero distance from the colony in year t and the nearest colony in year $t-1$. *NND* is a decreasing function of the number of colonies, so we standardized *NND* with the following equation based on the Poisson expectation (Clark and Evans 1954) with a slightly adjusting term:

$$NND' = NND - \frac{2}{\sqrt{\rho}} - \frac{4}{\rho},$$

where ρ is the density of colonies or number of colonies per area unit.

The second type of variables were proportional areas of land use around colonies. We created buffer zones of different scales (0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, and 40 km) around each colony and calculated the percentage of each of the four land uses obtained in the Landsat maps: water bodies, paddy fields, forests and urban areas. As we did not have a land-use map for every year of the study, we used map for the nearest year to extract land-use information. The time difference between the studied year and the land-use map was never greater than three years, so we avoided significant differences between the actual land use distribution of the studied year and the map information.

Model creation

We built yearly predictive site selection models to compare colony sites and unoccupied sites. To do so, we created random sets of pseudo-absences within the study area (delimited by an arbitrary rectangular area) in locations where colonies could, potentially, be established. Forest patches below an altitude of 100 m were considered as available locations, as colonies are seldom found at higher elevations in this area of Japan (Fig. 1). In order to avoid the problems associated with unbalanced data, the number of pseudo-absences included in a model was the same as the number of colonies detected for the studied year. Therefore, the sample size of each model was equivalent to the double of the total number of colonies for the studied year (Table 1). The response variables were the presences or absences of colonies.

We used site fidelity and land uses surrounding the real and hypothetical random colony sites as explanatory variables for a random forest (RF)

classification model (Breiman 2001). The RF algorithm combines the ideas of classification trees and bagging methods (Breiman 1984) and it is used for predicting the membership of cases in the classes of categorical dependent variable (colony or absence of a colony in our case), by measuring one or more predictor variables. A RF model is built by creating training sets with random sampling (bootstrap sample). For each bootstrap sample, a classification tree is built. The final prediction of the RF is the class with maximum votes among all trees. The RF algorithm is a great tool for modelling complex interactions among predictor variables (Prasad et al. 2006), including autocorrelation (Oliveira et al. 2012), and is becoming very popular in ecology due to its predictive power and its capacity to measure variable importance (Cutler et al. 2007). This methodology is especially suitable for studying the distribution of avian species in complex landscapes (Carrasco et al. 2014). In order to reduce the problems of spatial autocorrelation in our study, yearly models were used, and therefore the sample sizes became too small for traditional linear methodologies. The RF algorithm, however, can handle small number of observations and small input/predictor ratios (Cutler et al. 2007). An additional advantage of using a RF algorithm is that it provides a value for the predictive accuracy of each of the models without performing any cross validation. This is because the RF algorithm preserves some of the input data (about one third) each time one of the classification trees is built (500 trees were built for each RF model in our study), so we can use the left-out data, also called out of bag data (OOB), for estimating accuracy. The proportion of times among all the trees that the classification result is equal to the true class,

averaged among all cases, is the predictive accuracy. The predictive accuracy give us an estimate of the predictor variables' ability to explain colony distribution.

Randomized models were also created by using pseudo-presences, and comparing them with different sets of pseudo-absences. Surrounding land uses were analyzed and a RF algorithm was trained for each scale and its accuracy recorded. This process was done with ten sets of pseudo-absences (considered colonies) and each of these sets was compared with ten other sets of pseudo-absences (considered non-colony sites). Thus, a total of 100 estimations of random accuracy were calculated for each scale. To analyze the influence of both colony site fidelity and environmental variables on colony site selection, models including the site fidelity variable and the four land-use variables, as well as models that included only the land-use variables were created. Predictive accuracy of the real data higher than the randomized models' accuracy represents non-random distribution of the colonies. Consult Carrasco et al. (2014) for details on the creation of RF models.

Predictors' effects on the response variable

Partial dependence plots (Friedman 2001) were used to analyze the relationships between the predictor variables (land uses and site fidelity) and response variable (absence or presence of a colony). Partial dependence plots are based on predictions from the RF model, and they are very useful to graphically represent the responses to the predictors' variations. We can interpret the vertical axis of a partial dependence plot as the predicted colony presence probability for a certain value of the studied variable (proportion of the land-use area at certain scale or

NND), averaged across the observed values of the rest of predictors. In addition, similarity between partial dependence plots for different years can indicate the robustness of our models. In order to visually analyze the robustness of our model's responses behavior, we represented the partial dependence for every year model in the same plot.

Changes in variable importance

The RF algorithm also offers some measurements of variable importance, such as the mean decrease accuracy, or permutation accuracy (PA) importance (Breiman 2001). To calculate the PA index, the predictive accuracy for the OOB is recorded for each tree of the RF model. Then, the variable is permuted and the accuracy is recorded again. The differences between the two accuracies are then averaged over all trees and normalized by calculating the standard deviation. We used the PA index for analyzing changes in the importance of the different variables of the colony site selection models over the years of study. Again, we studied these changes for the models including and excluding site fidelity.

Changes in predictive accuracy

We analyzed the yearly changes of the predictive accuracy of the models created with and without colony site fidelity as an input value. The difference in accuracy between models with and without colony site fidelity was also calculated. This difference is a relative measure of how important the colony site fidelity variable is for explaining colony distribution. We examined yearly changes of this difference and fitted the results into three different models to try to explain

changes in the importance of colony site fidelity over the years: a null model (no variation over the years, linear model with slope equal to zero), a linear model (colony site fidelity importance varied among years) and a step function (representing any abrupt variation found in the importance of colony site fidelity). We studied which of these models better explained those yearly changes by comparing their AIC values and analyzing their generalized coefficient of determination (Nagelkerke 1991).

We also used the population data from Mashiko and Toquenaga (2013) to analyze the correlation between colony site fidelity levels (difference in accuracy between RF models with and without site fidelity) and the population sizes of the six species present in the study area colonies during the last 11 years of our study (the period with available population data).

Results

Predictors' effect on colony presence

Partial dependence plots of each studied predictor variable followed similar patterns for every study year (Fig. 2). Very low proportions of water bodies affected negatively the colony presence probability, while higher proportions had null or very little effect for most of the years (Fig. 2a). Small amounts of forests affected positively the colony presence probability, but proportions higher than 0.2 strongly decreased it (Fig. 2b). Responses to paddy fields and urban areas were very similar, showing negative effects for small proportions, and slightly positive effects for proportions higher than 0.3 (Fig. 2c, d). Finally, high site fidelity (NND values lower than 0.1) had a strong positive effect on the colony presence (Fig. 2e).

The response patterns were quite consistent among the different scales, although for larger scales the effects (positive or negative) of the land-use variables were weaker. Although the responses were similar for different years, we observe higher variability for high values of forest proportions and NND values (Fig. 2b, e) and for low values of paddy and urban proportions (Fig. 2c, d). Nevertheless, for all these cases the effect on the colony presence was negative, never showing drastic variations from negative to positive values.

Colony site fidelity vs land use preferences

Colony site fidelity was the most important variable for explaining colony sites almost every year for the models that included this variable (Fig. 3). Figure 3 also suggests a drastic increase in the importance of site fidelity from the year 2009. At the 1-km scale, the forests variable was second in importance, followed by urban areas (Fig. 3a). At the 10-km scale, in general, all land uses were less important for explaining colony distribution (Fig. 3b), in parallel with the example of the year 2000, where the predictive accuracy of the models without colony site fidelity decreased with the increase of the scale (Fig. 4). Also, at the 10-km scale, forests became the most important land-use variable for most of the years (Fig. 3b). Models without colony site fidelity showed similar patterns, with forests being the main land-use that increased the models' accuracies at small scale (Fig. 3c) and urban areas becoming more important at large scale (Fig. 3d). From 2008 to 2013, almost the same period during which we observe the large increase in the influence of colony site fidelity on the predictive models, the relative importance

of forests for the models without colony site fidelity increased drastically at the 1-km scale (Fig. 3c).

Changes in predictive accuracy

Models at all scales and for every year of the study showed a predictive accuracy higher than that for randomized models (ranging from a 10% to a 45% higher). This was true for both models with and without colony site fidelity as an explanatory variable. Figure 4 shows an example of the predictive accuracy graph of the RF model for each scale of study for the year 2000 (a central year in our study). The rest of the years showed similar patterns. Models including colony site fidelity had very high accuracy (over 85% in many of the scales). This is because a high percentage of the colonies reused their locations year after year, so models including colony site fidelity were able to easily predict the presence or absence of a colony based on its presence or absence in the previous year. Models without colony site fidelity also had accuracy values higher than random, but their accuracies were strongly scale dependent. Filled triangles in Figure 4 show that small scales were the best for explaining colony distribution, but there existed other scales at which we observed plateaus of high accuracy (typically at 4-5 km and 10 km).

Values of accuracy for models that included colony site fidelity fluctuated from 88% to almost 100% over the years for all scales (Fig. 5). Also, variance with scale was relatively small, as colony site fidelity measurements do not depend on the scale of the study. On the other hand, variance for the accuracy of the models without colony site fidelity was larger due to the dependence on scale. Averaged

values for the accuracies for these models fluctuated between 73% and 85%, except for the year 2002, in which the model obtained very high predictive accuracy (97%). We also observe a decrease of the accuracy of models without colony site fidelity throughout the years, although this decrease was not successfully explained with a linear regression model (decrease of 2.8% accuracy per year; p=0.016; r²=0.2).

In order to measure the influence of the colony site fidelity variable in our models, we calculated the difference between the values of accuracy of the models with and without colony site fidelity for each year (Fig. 6). None of the regression models successfully explained the variance of the data (step function: r²=0.006; linear function: r²=0.002) but the step function explained the change better than the simple linear model (Table 2). The step function model revealed an increase of almost 10% in the influence of colony site fidelity on predictive accuracy in the most recent five years of study (Fig. 6). This result supports the observed increase of the relative importance of site fidelity obtained from the PA analysis (Fig. 3)

The analysis of the relationship between colony site fidelity levels and species population sizes showed that grey heron population size had a significant positive correlation with site fidelity levels (Fig. 7a). Little egret population size showed no correlation (Fig. 7f), and the rest of the species presented marginal levels of significance (Fig. 7b-e).

Discussion

Our novel methodology, estimating colony site fidelity to include it as a predictor in yearly colony site selection models, allowed us to study how this factor varied over the years in relation to other environmental variables. RF models showed consistent responses of colony presence probability to different levels of site fidelity and to different proportions of land-uses around the colonies. This stability across the study period, together with the generalized high accuracies indicates a good reliability of the modelling approach. Models that only considered land use variables had accuracies higher than random (from 73% to 85%, except for one year), but these values were lower than those for models that included colony site fidelity (from 88% to 100%) and varied, depending on the scale and year of study. This supports our first hypothesis that colony site fidelity plays a main role in colony site selection. The difference in accuracy between models with and without colony site fidelity increased since 2009, indicating a rise in the importance of colony site fidelity for predicting colony distribution. Variable importance analysis of the models confirmed the increase in the importance of colony site fidelity relative to the four land-use variables in the most recent years of study. These results contradict our second hypothesis that the relative importance between habitat preference and colony site selection stays constant over the years. Below, we will argue how changes in the population of one of the species within the heron and egret communities may have caused the observed increase in colony site fidelity levels. Then, we will discuss the observed balance between environmental variables and colony site fidelity, as well as the importance in studying the relationship between the different factors that affect habitat selection in birds.

Our results showed that the level of colony site fidelity increased in the most recent five years of study. Different heron and egret species have different individual site fidelity levels (Melvin et al. 1999), while different species can have a different effect on the persistence of a colony over the years. Mashiko and Toquenaga (2014) showed that the presence of grey herons increased the persistence of the colonies in Ibaraki prefecture. Furthermore, the grey heron population in the region has increased significantly in the last decade, with its distribution limited to only one colony up until 1999 (Mashiko and Toquenaga 2014). We found that colony site fidelity levels and the grey heron population were highly correlated (Fig. 7a). The increase in the population size of the grey heron, together with its influence on the persistence of colonies with grey herons, might be causing the observed increase in colony site fidelity levels. The grey heron is the first among the six species to arrive at the breeding sites. Although the proportion of individuals exhibiting site fidelity is less than 20% of this species (Fernández-Cruz and Campos 1993), these few first-comers could be regarded as spearheading the establishment of the colonies. The rest of the species might then join the colonies that were initially established by those site-faithful grey herons. This process could have led to the observed increase in colony site fidelity over the past few years in our study area.

Together with the increase in the importance of colony site fidelity in recent years, we showed that the importance of environmental variables, in terms of the analyzed land use, decreased during the study period relative to colony site fidelity. Population changes can affect habitat preferences of birds (Svardson 1949, Helle and Järvinen 1986), including birds that exhibit site fidelity to nesting sites

(McLandress 1983, Donald and Evans 1995). Similarly, other environmental alterations, such as climate or landscape changes, can modify habitat preferences of a species (Norriss and Wilson 1993, Warnock and Takekawa 1995, Fox et al. 2005). But the relation of the changes in habitat preferences with the intensity of site fidelity was not broadly discussed in the latter cases. Our study shows that changes in site fidelity levels could be also an important factor for explaining the observed changes in habitat preferences among birds.

We observed a balance between habitat preferences and colony site fidelity: habitat preferences were a more powerful predictor of colony distribution until 2008, when colony site fidelity levels were lower. The change in the balance of these two factors could be affected, as we discussed, by the changes in species composition of the colonies. The effect of both factors behind the selection of breeding sites could compensate each other. For those species that exhibit site fidelity, changes in site fidelity levels could be influencing our observations regarding habitat preferences, as some species could be relying more on reusing locations than on habitat assessments (Jones 2001, Chalfoun and Schmidt 2012). Changes in the balance between factors affecting habitat selection could have a significant impact upon avian species. For example, an increase in site fidelity could keep a species on an unsuitable site if the habitat quality decreases over the years (Knick and Rotenberry 2000, Osborne et al. 2007). Furthermore, considering changes to site fidelity intensity in terms of habitat selection is important in conservation planning, as predictive models are often based on habitat preferences only (Guisan and Zimmermann 2000), which means that the effects of habitat changes could be overestimating future distribution changes. More studies that

include other factors, such as conspecific or heterospecific attraction, competition and predation, are required for a better understanding of habitat selection and clarification of the influence of environmental alterations on the balance between those factors.

Our results suggest that examining relationships between different factors, which influence how birds choose their living habitats, can lead to a better understanding of the habitat selection process. Furthermore, these relationships are difficult to observe in short-term studies because of the potential existence of inter-annual variability, as shown in this study. The recent availability of satellite images from previous years allows us to combine land-use data from the past with distribution and population trends concerning long-term data compiled by ecologists (Kerr and Ostrovsky 2003, Rose et al. 2015). Our findings add evidence of the benefits of this combined approach to analyzing species' responses to environmental changes, as well as highlight the importance of long-term data in order to understand high inter-annual variable processes and the potential for studying slow processes or rare ecological events.

References

Austenfeld, M. and Beyschlag, W. 2012. A graphical user interface for R in a rich client platform for ecological modeling. – J. Stat. Softw. 49: 1–19.

Bollinger, E. K. 1995. Successional changes and habitat selection in hayfield bird communities. – Auk 112: 720–730.

Breiman, L. 2001. Random forests. – Mach. Learn. 45: 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone C. G. 1984. Classification and regression trees. – Wadsworth International Group, Belmong, California.

Bried, J., Pontier, D. and Jouventin, P. 2003. Mate fidelity in monogamous birds: a re-examination of the procellariiformes. – Anim. Behav. 65: 235–246.

Calder, W., Waser, N. M., Hiebert, S. M., Inouye, D. W. and Miller, S. 1983. Site-fidelity, longevity, and population dynamics of broad-tailed hummingbirds: a ten year study. – Oecologia 56: 359–364.

Carrasco, L., Mashiko, M. and Toquenaga, Y. 2014. Application of random forest algorithm for studying habitat selection of colonial herons and egrets in human-influenced landscapes. – Ecol. Res. 29: 483–491.

Carrasco, L., Toquenaga, Y. and Mashiko, M. 2015. Extrapolation of random forest models shows scale adaptation in egret colony site selection against landscape complexity. – Ecol. Complex. 24: 29–36.

Cézilly, F., Dubois, F. and Pagel, M. 2000. Is mate fidelity related to site fidelity? A comparative analysis in ciconiiforms. – Anim. Behav. 59: 1143–1152.

Chabrzyk, G. and Coulson, J. C. 1976. Survival and recruitment in the herring gull *Larus argentatus*. – J. Anim. Ecol. 45: 187–203.

Chalfoun, A. D. and Schmidt, K. A. 2012. Adaptive breeding-habitat selection: is it for the birds? – Auk 129: 589–599.

Chavez, P. S. 1996. Image-based atmospheric corrections revisited and improved. – Photogramm. Eng. Remote Sensing 62: 1025–1036.

Clark, P. J. and Evans, F. C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. – Ecology 35: 445–453.

Clark, R. G. and Shutler, D. 1999. Avian habitat selection: pattern from process in nest-site use by ducks? – Ecology 80: 272–287.

Cody, M. L. 1985. Habitat selection in birds. – Academic Press, San Diego.

Coulson, J. C. and Coulson, B. A. 2008. Measuring immigration and philopatry in seabirds; recruitment to black-legged kittiwake colonies. – Ibis 150: 288–299.

Couzin, I. D. 2009. Collective cognition in animal groups. – Trends Cogn. Sci. 13: 36–43.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. and Lawler, J. J. 2007. Random forests for classification in ecology. – Ecology 88: 2783–2792.

Dale, S. and Christiansen, P. 2010. Individual flexibility in habitat selection in the ortolan bunting *Emberiza hortulana*. – J. Avian Biol. 41: 266–272.

Danchin, E., Boulinier, T. and Massot, M. 1998. Conspecific reproductive success and breeding habitat selection: implications for the study of coloniality. – Ecology 79: 2415–2428.

Deneubourg, J. L. and Goss, S. 1989. Collective patterns and decision-making. – Ethol. Ecol. Evol. 1: 295–311.

Donald, P. F. and Evans, A. D. 1995. Habitat selection and population size of corn buntings *Miliaria calandra* breeding in Britain in 1993. – Bird Study 42: 190–204.

Environmental Agency of Japan 1994. Distribution and population status of colonies and communal roosts of 22 bird species from 1990 to 1992. Wild Bird Society of Japan and the Environmental Agency of Japan, Tokyo, Japan (In Japanese)

Esely, J. D. and Bollinger, E. K. 2001. Habitat selection and reproductive success of loggerhead shrikes in northwest Missouri: a hierarchical approach. – Wilson Bull. 113: 290–296.

Fasola, M., Rubolini, D., Merli, E., Boncompagni, E. and Bressan, U. 2010. Long-term trends of heron and egret populations in Italy, and the effects of climate, human-induced mortality, and habitat on population dynamics. – Popul. Ecol. 52: 59–72.

Fernández-Cruz, M. and Campos, F. 1993. The breeding of grey herons (*Ardea cinerea*) in western Spain: the influence of age. – Colon. Waterbird. 16: 53–58.

Forbes, L. S. and Kaiser, G. W. 1994. Habitat choice in breeding seabirds: when to cross the information barrier. – Oikos 70: 377–384.

Fox, A. D., Madsen, J., Boyd, H., Kujiken, E., Norriss, D. W., Tombre, I. M. and Stroud, D. A. 2005. Effects of agricultural change on abundance, fitness components and distribution of two artic-nesting goose populations. – Global. Change. Biol. 11: 881–893.

Franciesaz, C., Guilbault, E., Lebreton, J. D, Trouvilliez, J. and Besnard, A. 2016. Colony persistence in waterbirds is constrained by pond quality and land use. – Freshwater Biol. (In Press)

Fretwell, S. D. and Lucas, H. L. 1969. On territorial behaviour and other factors influencing habitat distribution in birds. – Acta. Biotheor. 14: 16–36.

Friedman, J. 2001. Greedy function approximation: a gradient boosting machine. – Ann. Stat. 29: 1189–1232.

Fujioka, M., Yoshida, H. and Toquenaga, Y. 2001. Research on the dynamic phase of biodiversity in geographic scale and its preservation, (3) Analysis of the dynamic phase wildlife population in geographic scale, (2) Research on the spatiotemporal dynamic phase of bird gathering places. – In: Tsubaki, Y. (ed.), Global environment research fund: analysis and conservation of biodiversity on a geographical scale. Ministry of the Environment, Tokyo, pp. 75–88. (In Japanese)

Ganter, B. and Cooke, F. 1998. Colonial nesters in a deteriorating habitat: site fidelity and colony dynamics of lesser snow geese. – Auk 115: 642–652.

Goslee, S. C. 2011. Analyzing remote sensing data in R: the LANDSAT package. – J. Stat. Softw. 43: 1–25.

Greenwood, P. J. 1980. Mating systems, philopatry and dispersal in birds and mammals. – Anim. Behav. 28: 1140–1162.

Greenwood, P. J. and Harvey, P. 1982. The natal and breeding dispersal of birds. – Annu. Rev. Ecol. Syst. 13: 1–21.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Modell. 135: 147–186.

Haynes, T. B., Shmutz, J. A., Lindberg, M. S., Wright, K. G., Uher-Kock, B. D. and Rosenberger, A. E. 2014. Occupancy of yellow-billed and Pacific loons: evidence for interspecific competition and habitat mediated co-occurrence. – J. Avian Biol. 45: 296–304.

Heinänen, S., Rönkä, M. and Numers, M. V. 2008. Modelling the occurrence and abundance of a colonial species, the arctic tern *Sterna paradisaea* in the archipelago of SW Finland. – Ecography 31: 601–611.

Helle, P. and Järvinen, O. 1986. Population trends of North Finnish land birds in relation to their habitat selection and changes in forest structure. – Oikos 46: 107–115.

Immelmann, K. 1972. Sexual and other long-term aspects of imprinting in birds and other species. – Adv. Stud. Behav. 4: 147–174.

Jones, J. 2001. Habitat selection studies in avian ecology: a critical review. – Auk 118: 557–562.

Kelly, J., Stralberg, D., Etienne, K. and McCaustland, M. 2008. Landscape influence on the quality of heron and egret colony sites. – Wetlands 28: 257–275.

Kerr, J. T. and Ostrovsky, M. 2003. From space to species: ecological applications for remote sensing. – Trends Ecol. Evol. 18: 299–305.

Kerth, G. 2010. Group decision-making in animal societies. – In: Kappeler, P. M. (ed.), Animal behaviour: Evolution and mechanisms. Springer, Heidelberg, pp. 241–265.

Knick, S. T. and Rotenberry, J. T. 2000. Ghosts of habitat past: contribution of landscape change to current habitats used by shrubland birds. – Ecology 81: 220–227.

Koshida, C. 2007. Colony vicissitudes of herons and egrets around southern Ibaraki Prefecture. Master's thesis, University of Tsukuba, Japan. (In Japanese)

Kushlan, J. A. and Hancock, J. A. 2005. The Herons. – Oxford University Press, New York.

Lack, D. 1968. Ecological adaptations for breeding in birds. – Methuen, London, UK.

Magurran, A. E. et al. 2002. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. – Trends Ecol. Evol. 25: 574–582.

Martin, T. E. 1993. Nest predation and nests sites: new perspectives on old patterns. – Bioscience 43: 523–532.

Mashiko, M. 2014. Colony dynamics of heron communities: ecological and historical perspectives. – PhD thesis, Univ. Tsukuba.

Mashiko, M. and Toquenaga, Y. 2013. Increasing variation in population size and species composition ratio in mixed-species heron colonies in Japan. – Forktail 29: 71–77.

Mashiko, M. and Toquenaga, Y. 2014. Grey heron (*Ardea cinerea*) expansion promotes the persistence of mixed-species heron colonies. – Waterbirds 37: 362–370.

McLandress, M. R. 1983. Temporal changes in habitat selection and nest spacing in a colony of ross' and lesser snow geese. – Auk 100: 335–343.

Melvin, S. L., Gawlik, D. E. and Scharff, T. 1999. Long-term movement patterns for seven species of wading birds. – Waterbirds 22: 411–416.

Nagelkerke, N. J. D. 1991. A note on a general definition of the coefficient of determination. – Biometrika 78: 691–692.

Norriss, D. W. and Wilson, H. J. 1993. Seasonal and long-term changes in habitat selection by greenland white-fronted geese *Anser albifrons flavirostris* in Ireland. – Wildfowl 44: 7–18.

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A. and Pereira, J. M. C. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. – Forest Ecol. Manag. 275: 117–129.

Osborne, P. E., Suárez-Seoane, S. and Alonso, J. C. 2007. Behavioural mechanisms that undermine species envelope models: the causes of patchiness in the distribution of great bustards *Otis tarda* in Spain. – Ecography 30: 819–828.

Petit, L. J. and Petit, D. R. 1996. Factors governing habitat selection by prothonotary warblers: field tests of the Fretwell-Lucas models. – Ecol. Monogr. 66: 367–387.

Prasad, A. M., Iverson, L. R. and Liaw, A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. – Ecosystems 9: 181–199.

Rose, R. A. et al. 2015. Ten ways remote sensing can contribute to conservation. – Conserv. Biol. 29: 350–359.

Svardson, G. 1949. Competition and habitat selection in birds. – Oikos 1: 157–174. Tarvin, K. A. and Garvin, M. C. 2002. Habitat and nesting success of blue jays (*Cyanocitta cristata*): importance of scale. – Auk 119: 971–983.

Therneau, T. M. and Atkinson, E. J. 2007. The rpart package. – R Package Version 2007 .

Thomson, R. L., Forsman, J. T. and Mönkkönen, M. 2003. Positive interactions between migrant and resident birds: testing the heterospecific attraction hypothesis. – Oecologia 134: 431–438.

Tsai, J. S., Reichert, B. E., Frederick, P. C. and Meyer, K. D. 2016. Breeding site longevity
and site characteristics have intrinsic value for predicting persistence of colonies of an
endangered bird. – Wetlands 36: 639–647.

Warnock, S. E. and Takekawa, J. Y. 1995. Habitat preferences of wintering shorebirds in a
temporally changing environment: western sandpipers in the San Francisco bay
estuary. – Auk 112: 920–930.

Wiens, J. A. 1989. Spatial scaling in ecology. – Funct. Ecol. 3: 385–397.

Supplementary material (Appendix JXXXXX at
<www.oikosoffice.lu.se/appendix>). Appendix 1.

Table Legends

Table 1. Sample sizes for each year model. The sample size is double to the number of colonies detected for that year.

Year	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Sample size	28	28	30	38	40	36	46	36	34	40	46	40	38	46

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Sample size	50	40	32	32	34	36	40	42	48	48	58	54	50	50

Table 2. AIC values of the fitted models for the data of Figure 6. A step function model, a linear model and a linear model with slope equal to zero (null model) were used. ΔAIC values represent differences to the minimum AIC model and wAIC values represent the rounded Akaike weights.

Model	AIC_i	$\Delta_i AIC$	$w_i AIC$
Step	169.5	0	.998
Linear	178.7	9.2	.009
Null	181.9	12.4	.002

Figure Legends

Figure 1. Map showing the location of 149 colonies in the study area around Ibaraki Prefecture, Japan, from 1986 to 2013. Broken lines represent prefectural boundaries and continuous lines represent the main rivers in the study area. White areas are lowlands with an altitude of 0 m to 100 m where heron colonies could be formed on trees and bamboo thickets that are distributed throughout the regions. Gray areas show an altitude greater than 100 m where the distribution of egrets is lower.
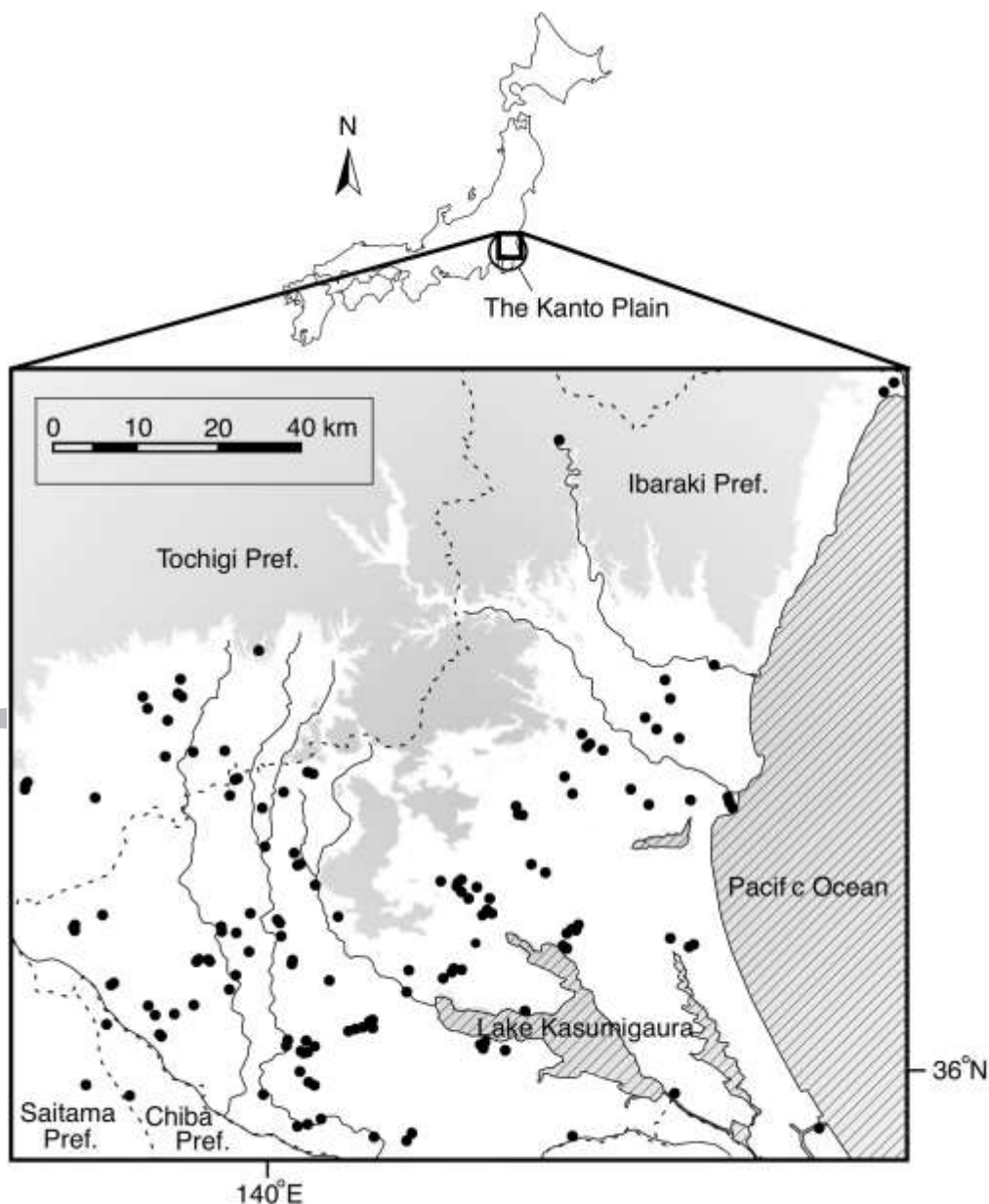
Fig 2. Partial dependence plots for random forest models that included site fidelity at 1-km scale. Each curve represents the partial dependence of a single year model, so that each graph contains the partial dependence plot every single studied year (*n* = 28). Partial dependence is the marginal effect of a predictor variable on the probability of presence of a colony.
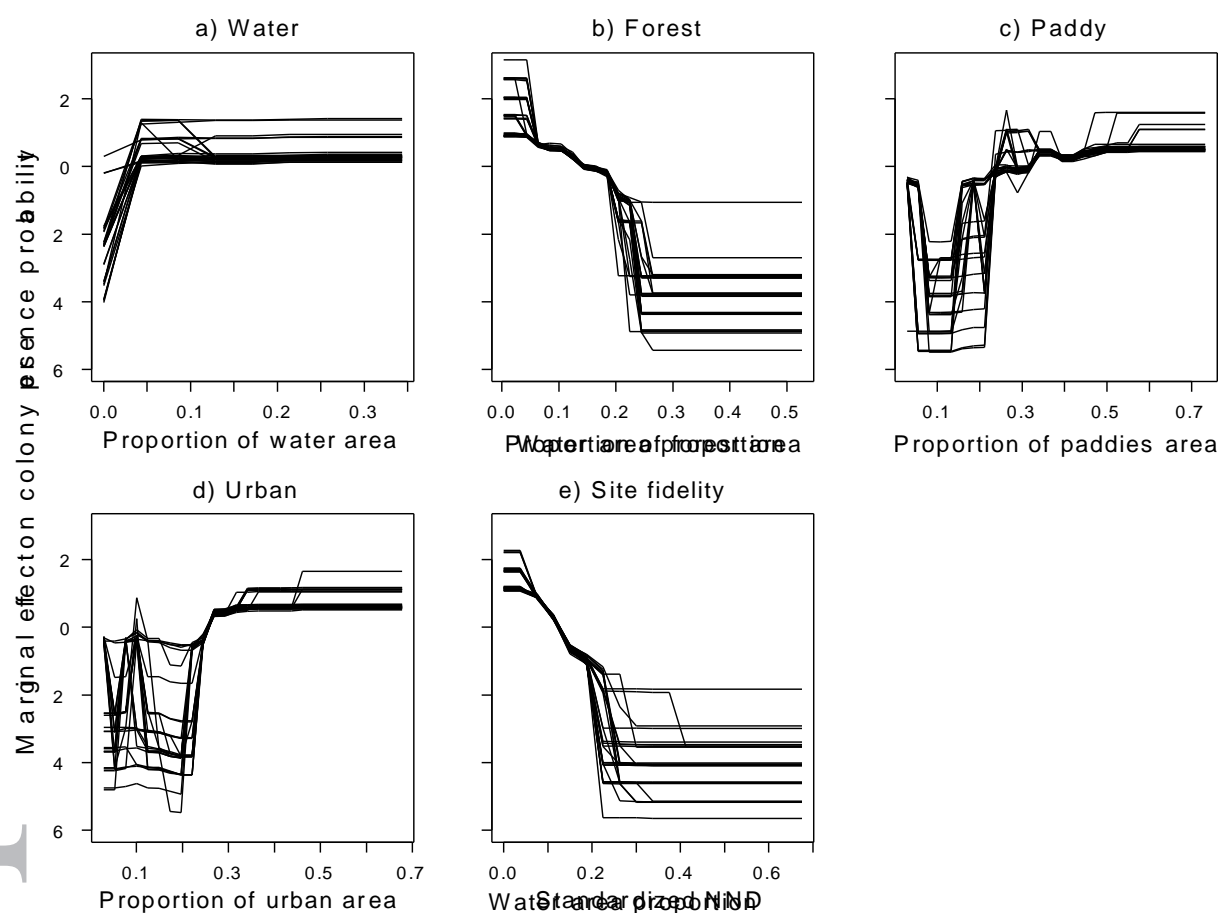
Figure 3. Importance of the variables measured with the permutation accuracy (PA) index for the random forest models with and without site fidelity, on two different scales (1 km and 10 km).
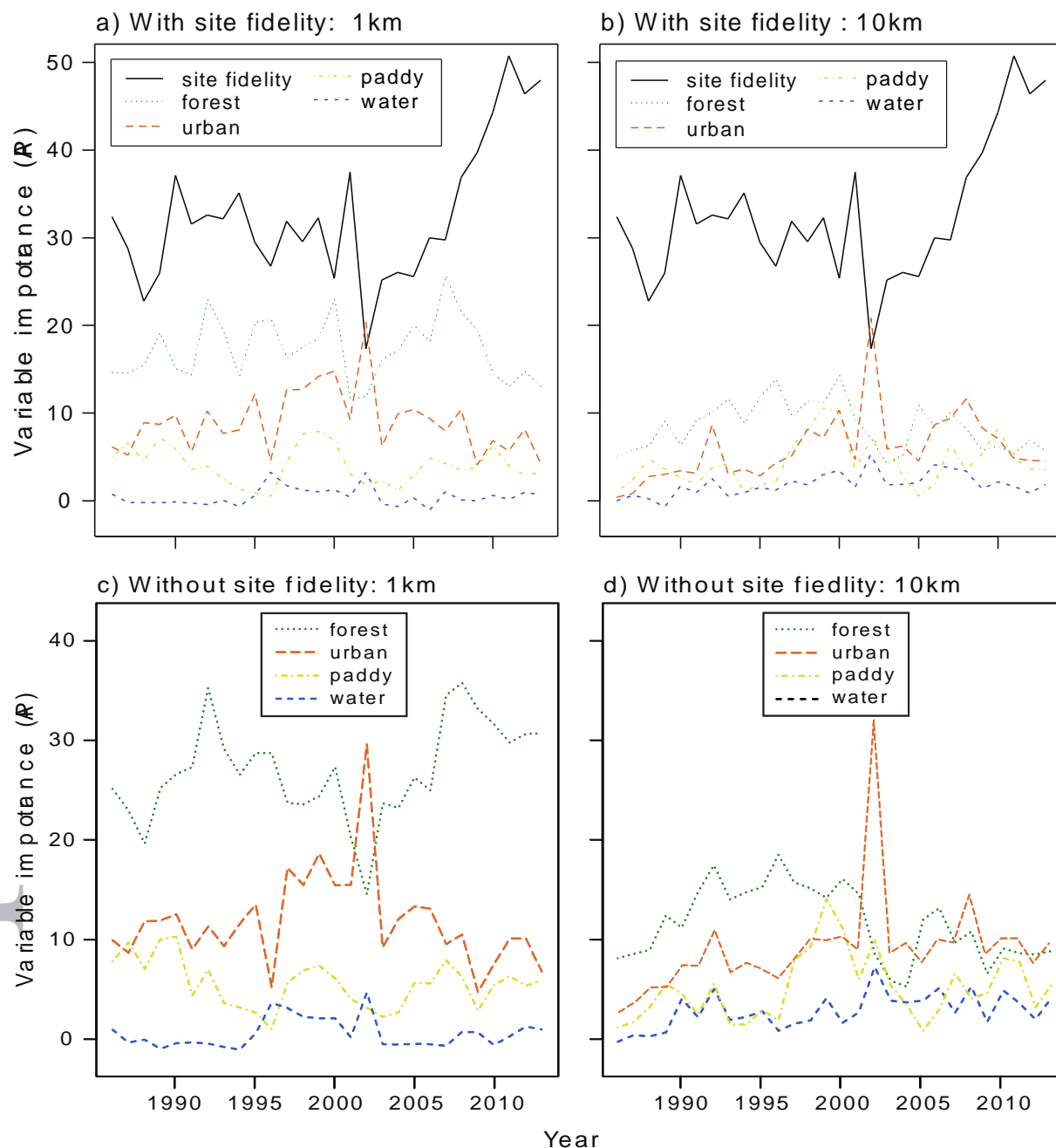
Figure 4. Predictive accuracies for models from the year 2000 for each scale of the study. Accuracy values for the models that included land uses and site fidelity are shown with filled triangles. Accuracy values for the models that included only land uses are shown with filled circles. Open triangles represent accuracies for the random model with site fidelity. Open circles represent accuracies for the random model without site fidelity. Mean ± 95% confidence intervals among 100 different random permutations of the response values are shown.
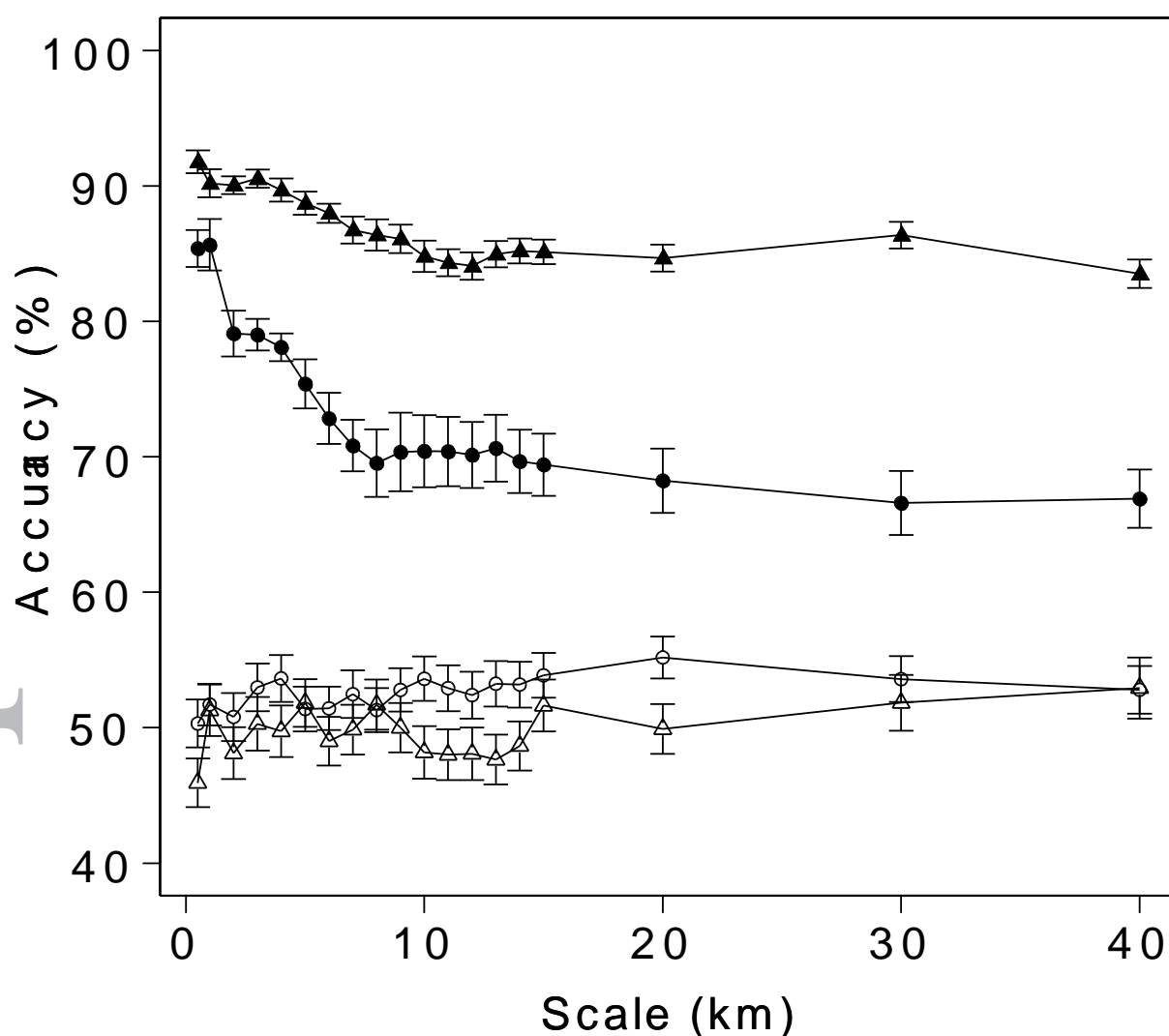
Figure 5. Mean predictive accuracy among all scales for models with site fidelity (triangles) and models without site fidelity (circles). Error bars represent a 95% confidence interval of the means.
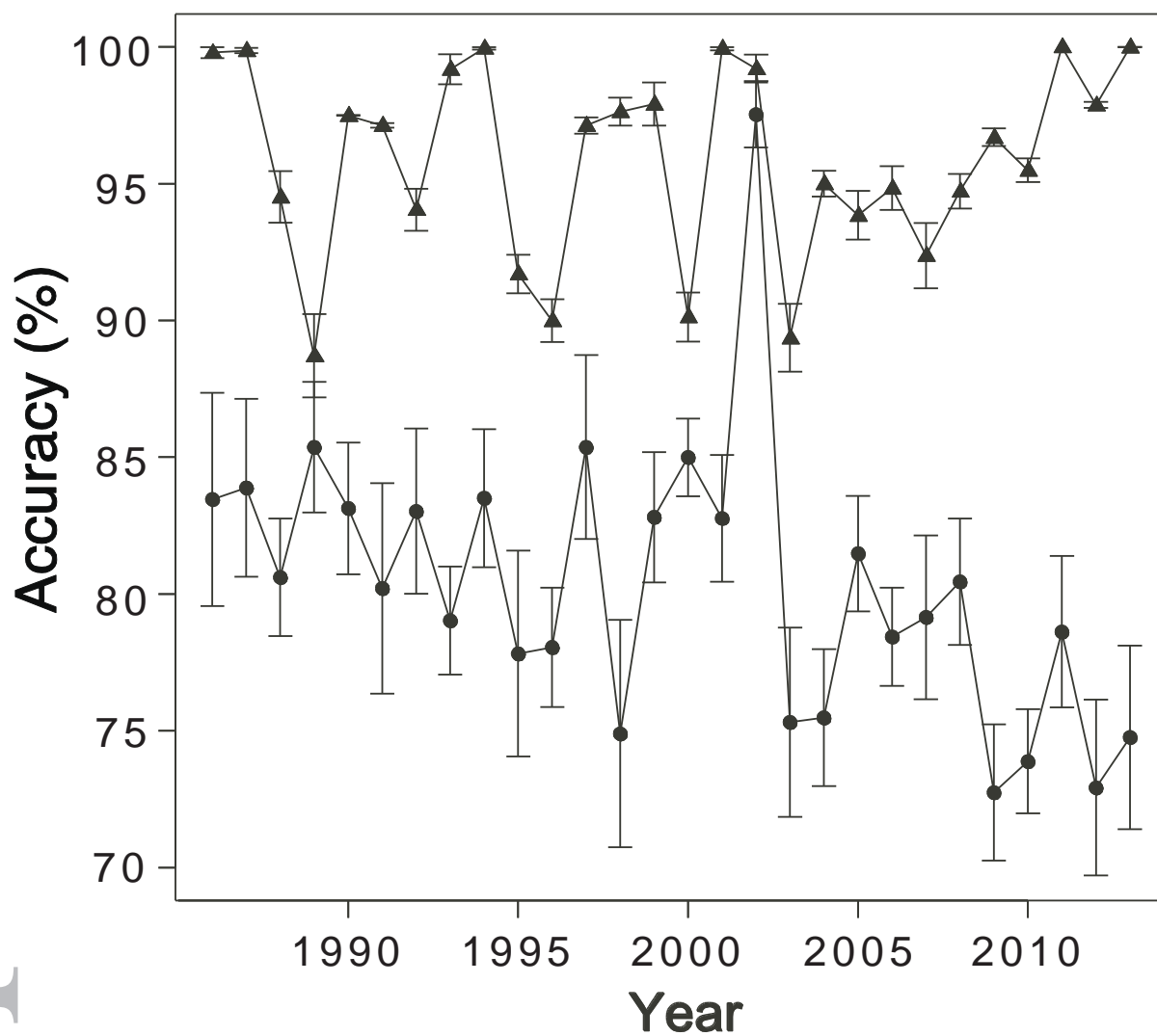
Figure 6. Accuracy differences between models with and without site fidelity. Error bars represent a 95% confidence interval of the differences. The line represents the fitted step function.
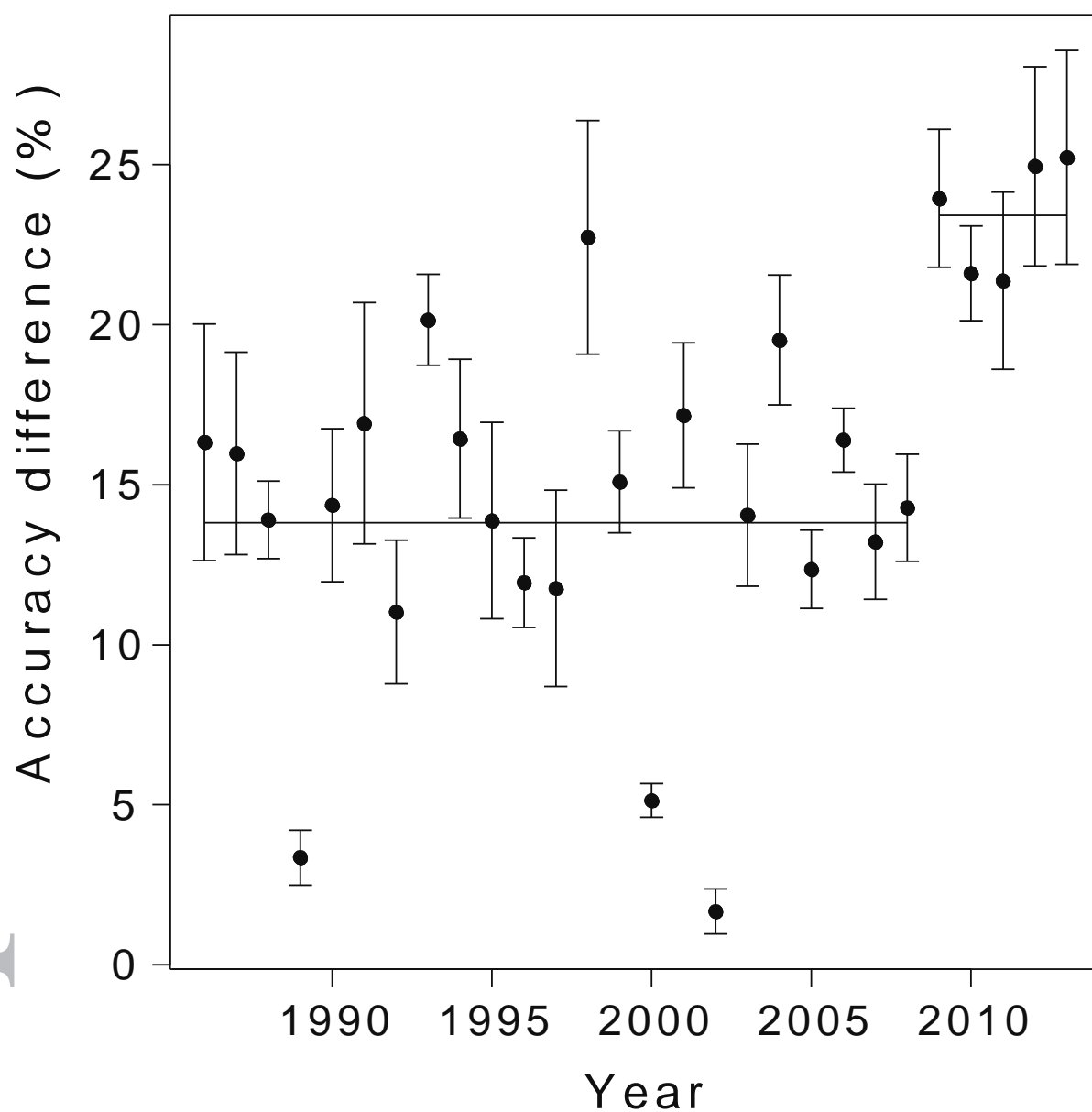
Figure 7. Relationships between colony site fidelity levels, estimated as the difference in predictive accuracy between site selection models with and without site fidelity, and population size of the six species of herons and egrets present in the study area. Each value corresponds to one year, from 2003 to 2013. The Pearson's correlation coefficient and its p values are shown.