

Colony Site Selection of Herons and Egrets: From Landscapes to Behavior

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Science
(Doctoral Program in Biological Sciences)

LUIS CARRASCO TORNERO

Contents

List of Tables	iii
List of Figures	iv
Abstract	1
General Introduction	3
1 Application of Random Forest Algorithm for Studying Habitat Selection of Colonial Herons and Egrets in Human-influenced Landscapes	7
1.1 Introduction	7
1.2 Methods	9
1.3 Results	14
1.4 Discussion	15
2 Extrapolation of Random Forest Models Shows Scale Adaptation in Heron and Egret Colony Site Selection against Landscape Complexity	20
3 Balance Between Site Fidelity and Habitat Preferences in Colony Site Selection by Herons and Egrets	22
General Discussion	24
Acknowledgements	27
References	28

List of Tables

1	Proportion of each land use of the study area	42
2	Pearson's correlation coefficient between land-use variables among all buffers of 4-km scale	43
3	Pearson's correlation coefficient between land-use variables among all buffers of 10-km scale	44

List of Figures

1	Locations of colonies in 2011 in the study area	36
2	Land-use map of the study area	37
3	Model accuracies for each scale	38
4	Land uses importance	39
5	Partial dependence plots	40

Abstract

Colonial breeding occurs in many avian species, being a very common form of group living among seabirds and waterbirds. Herons and egrets (family Ardeidae) gather during breeding seasons and form single- or mixed-species colonies. Available sites for establishing colonies are abundant for herons and egrets. However, most available sites remain unoccupied; only a portion of them is used for colony sites, and some are used repeatedly for decades. A better understanding of factors that influence colony site selection is fundamental to explain changes in distribution or population trends of these species.

Six species of herons and egrets, Grey Heron (*Ardea cinerea*), Great Egret (*A. alba*), Little Egret (*Egretta garzetta*), Intermediate Egret (*E. intermedia*), Cattle Egret (*Bubulcus ibis*) and Black-crowned Night Heron (*Nycticorax nycticorax*) breed mainly in mixed-species colonies on the Kanto Plain, in the eastern region of Japan. By combining colony distribution data with land-use maps, I studied colony site selection of herons and egrets in Ibaraki Prefecture and its changes across the landscapes and over the years.

In Chapter 1, I described the habitat preferences of herons and egrets for selecting their colony sites in 2011. I used a random forest algorithm to analyze the land uses surrounding the colonies at different scales. I found that there were two main scales at which herons and egrets selected their colony sites: medium (4 km) and large (10-15 km). Colonies were attracted to areas with large amounts of evergreen forests at the medium scale while at large scales they avoided high-density urban areas. Paddy fields, the main foraging habitat for herons and egrets, were not important for explaining colony sites. Previous studies used attractive factors, mainly foraging areas, to explain colony distributions, but my study is the first to show the major importance of repellent factors at large scales.

In Chapter 2, I compared colony site selection between Ibaraki and Chiba prefectures. I applied a cross-validation by creating a colony site selection model at a certain scale in a region and applied it in the other region at different scales, comparing habitat

preferences across scales. I found that herons and egrets shared the same habitat preferences for selecting the colony sites in both regions. However, higher complexity of the paddy fields in Chiba prefecture resulted in a reduction of the scale of selection in this region. Previous studies have shown how landscape complexity can affect the scale of foraging behavior, movement or habitat range, but this is the first study to show evidence of adaptation to scale for colony site selection.

In Chapter 3, I described the colony site selection of herons and egrets in Ibaraki Prefecture over 28 years. I analyzed the balance between habitat preferences and collective site fidelity for colony site selection and its changes over the years. I observed high levels of collective site fidelity for every year of the study, making this an important factor in explaining colony distribution in Ibaraki. I also found that the importance of collective site fidelity increased in comparison to habitat preferences in the last five years of the study. The observed increase of collective site fidelity was caused by the growth of the Grey Heron population in the last years.

The observed differences of colony site selection across scales, landscapes, and over time could explain the disparate results in previous studies on colony site selection of herons and egrets. Restricting the studies to one single scale or to one single year can enormously limit the ability of a habitat selection model to explain colony distribution. Additionally, researchers should include, whenever possible, collective site fidelity as a variable for explaining colony distribution, as the high tendency of herons and egrets to use the same colony sites year after year, strongly affects the colony site selection of these species, even when individual site fidelity can be very low. The use of habitat selection models at the landscape level was very useful to estimate some behavioral aspects of the colony site selection of herons and egrets.

Key-words

colonial birds, habitat selection, landscape complexity, landscape ecology, land-use maps, long-term data, random forest, scale, site fidelity

General Introduction

Colonial breeding, a form of group living in which individuals gather together to raise their offspring (Danchin and Wagner 1997), can be observed in many animal taxa like reptiles (Trillmich and Trillmich 1984), mammals (Terhune and Brillant 1996) and birds (Lack 1968). Coloniality may be the result of a limited number of breeding locations (Wittenberger and Hunt 1985; Rolland et al. 1998). But its evolution has also been explained in terms of its advantages, such as predation reduction (Lack 1968; Anderson and Hodum 1993), efficient mating and reproduction (Møller 1987; Wagner et al. 1996), and increased foraging efficiency through the provision of an information exchange center for food resources (Ward and Zahavi 1973; Richner and Heeb 1995).

Colonial breeding occurs in about 13% of avian species (Gill 2007), being very common among seabirds and waterbirds (Danchin and Wagner 1997). When potential breeding locations are abundant, colonial birds should select appropriate locations in order to optimize their fitness, but this colony site selection process is not well understood (Brown et al. 2000; Brown and Brown 2001). Birds could select their breeding sites by assessing the local resource availability (Fretwell 1969; Brown and Rannala 1995), or the quality of surrounding habitats (Tarvin and Garvin 2002). Some species can use the presence of other individuals of the same species (conspecific attraction) or even of different species (heterospecific attraction) as cues for choosing their breeding sites (Keister 1979; Danchin et al. 1998; Thomson et al. 2003). Predation avoidance has also been studied to explain breeding habitat choices in avian species (Cody 1985; Martin 1993). Alternatively, they could return to their natal breeding areas or previously used habitats (site fidelity) (Greenwood 1980; Doligez et al. 2003), reducing assessment efforts (Forbes and Kaiser 1994).

Hérons and egrets (family Ardeidae) are long-legged wading birds, widely distributed all around the world (Kushlan and Hancock 2005). They gather during the breeding season forming single- or mixed-species colonies that can consist of from a few individuals to thousands in some regions. In general, available sites for establishing

colonies are abundant for herons and egrets, as they can build their nests in trees, bushes or reeds (Kushlan and Hancock 2005). However, most available sites remain unoccupied; only a portion of them is used for colony sites, and some are used repeatedly for decades. Not a few studies have tried to explain the factors that affect the colony site selection of herons and egrets (Gibbs et al. 1987; Fasola and Alieri 1992; Tourenq et al. 2004; Boisteau and Marion 2007; Kelly et al. 2008; Fasola et al. 2010; Parkes et al. 2012). Those studies reported very disparate results in terms of the habitat preferences, as well as on the scale at which herons and egrets select their colony locations. A better understanding of the factors that influence colony site selection is fundamental, as it can help us to explain changes in distribution or population trends of heron and egret species.

Six species of herons and egrets, Grey Heron (*Ardea cinerea*), Great Egret (*A. alba*), Little Egret (*Egretta garzetta*), Intermediate Egret (*E. intermedia*), Cattle Egret (*Bubulcus ibis*) and Black-crowned Night Heron (*Nycticorax nycticorax*) breed mainly in mixed-species breeding colonies on the Kanto Plain, in the eastern region of Japan. They build their nests, from late February to May, on conifers, broad-leaf trees and in bamboo thickets, leaving the colonies at the end of summer. Colony distribution around Ibaraki Prefecture, in Japan, has been monitored by members of the Population Ecology Laboratory (PEL), University of Tsukuba, since 1983 to date. I joined the PEL in 2011 and helped with the colony surveys until 2014. The surveys in the area have been intense almost every year, so the probability of missing an existing colony is minimal, and the results of colony presence and absence data has been of great interest.

Distribution data combined with land-use maps of high resolution can be a useful tool for explaining how species choose their habitats (Ozesmi and Bauer 2002; Rose et al. 2014). Furthermore, good quality long-term data, such as the colony distribution data compiled by the PEL in the last decades, is crucial for studying species distribution and its changes over time (Magurran et al. 2010). In this study, I used a (ground-survey based) land-use map provided by the Japanese Aerospace Exploration Agency (JAXA)

in 2011 to study the relationship between the colony sites of herons and egrets and their surrounding habitats. This land-use data allowed me not only to study the habitat preferences of colonies in Ibaraki in 2011, but also to analyze differences of habitat preferences among landscapes, and was also the base for creating land-use maps of previous years.

By combining colony distribution data with land-use maps, I studied colony site selection of herons and egrets in Ibaraki and its changes across the landscapes and over the years. The Ibaraki landscape is highly fragmented due to human impact, and this increases the difficulties with studying the relationship between colony sites and their surrounding habitats. Linear methods, such as logistic regressions, are not ideal for dealing with this complex spatial data (Seppelt and Voinov 2002). In addition, these methods are very specific to data sets, making them difficult to extrapolate. Thus, for explaining colony distribution from land-use data, I used a non-linear algorithm, Random Forest (RF) (Breiman 2001), which is better suited for extrapolating results across landscapes, and is able to deal with the complex Ibaraki landscape.

In Chapter 1, I describe the RF algorithm to analyze the surroundings of the 2011 colonies in order to study the habitat preferences at several scales. I found that there were two main spatial scales that herons and egrets use to select their colony sites: medium scale (4 km) and large scale (10-15km). At the medium scale, colonies were attracted mainly to evergreen forests while at large scale they avoided high-density urban areas. In Chapter 2, I give a comparison of colony site selection between Ibaraki and Chiba prefectures. I found that the herons and egrets had the same habitat preferences, but that differences in habitat configuration resulted in changes in the scale of selection. In Chapter 3, I provide an examination of colony site selection over 28 years. I found that the balance between habitat preferences and collective site fidelity varied over time. The importance of collective site fidelity for selecting colony sites increased during the last five years of study, coinciding with the drastic increase in the Grey Heron population.

My study shows that herons and egrets can be attracted or repelled by different habitats when placing their colonies, depending on the scale. However, the scale can depend on landscape complexity. Collective site fidelity was very important in explaining colony site selection, but its importance varied over time. A long-term study was fundamental to understanding the balance between habitat preferences and collective site fidelity in choosing colony locations. This balance varied due to changes in the population dynamics of colonies as well as environmental variation. Analyzing colony distribution from a landscape-oriented perspective was very useful in clarifying behavioral aspects of the colony site selection of herons and egrets. This approach will be advantageous for studying the collective behavior of gregarious animals whose behaviors are very hard to analyze by direct observation.

Chapter 1

Application of Random Forest Algorithm for Studying Habitat Selection of Colonial Herons and Egrets in Human-influenced Landscapes

1.1 Introduction

Understanding the mechanisms of habitat selection is fundamental to the construction of proper conservation and management plans for many avian species. Choosing breeding sites is a crucial task for avian species, but it is still not clear how they undertake it. Approximately 13% of birds breed in spatially packed colonies (Gill 2007). Colony site selection is a more difficult problem than choosing an individual nest site because the site selection affects the fate of all members of the colonies.

Identifying those scales at which certain distribution pattern occur can help to clarify what mechanisms are involved in habitat selection. However, many studies on colonial birds have been conducted at a single spatial scale, so results about colony site selection and explanations for the mechanisms involved have been widely diverse. Most of them used linear models or simple correlations between landscape variables and the presence of the species for explaining the colony distribution (Fasola and Canova 1991; Tourenq et al. 2004). The interaction between the explanatory variables are very intricate, and a high correlation among scales makes this analysis even more

complicated, specially when studying mixed species colonies, when differential habitat selection among species could add more complexity.

Some authors created habitat suitability models for colonial birds (Kelly et al. 2008; Parkes et al. 2012), but their methods assumed linear responses between the dependent variable and the explanatory variables. All of these methods are generally appropriate when studying relatively simple variable interactions and when responses to the explanatory variables are linear. However, widely used methods such as logistic regression are often misapplied. In many cases, applying a logistic regression does not guarantee maximum-likelihood estimates and the odd ratios are not always proportional to the probability of presence of the species (Keating and Cherry 2004). For this reason, new methodologies that can successfully incorporate non-linear and complex-variables' relationships are needed to analyze differences in site selection for each scale.

In the last decades, the high human impact on natural landscapes has challenged scientists to improve their predictive models in order to create effective conservation plans for bird species that share habitat with human beings. The complexity of the optimization problem in ecological models increases with spatial complexity (Seppelt and Voinov 2002), so including a higher fragmentation of agricultural landscapes may add difficulty to the analysis of the relationships between the habitat variables and the colony locations. Moreover, landscape complexity can affect the ability of the species to assess the habitat and for the detection of resources (Wiens and Milne 1989). Furthermore, for some agricultural landscapes affected by urban development, the explanatory data are too complex, and it is necessary to use other techniques without assuming linearity, such as classification trees or machine learning methods. These techniques are better tools for extrapolating the response variables across landscapes and for analyzing the importance of the predictors than are other methods such as linear regressions (Prasad et al. 2006). The random forest (RF) technique (Breiman 2001) does not need to assume linearity. It allows for the modelling of complex interactions among predictor variables and is becoming widely used due to its predictive

power (in comparison with normal decision trees) and its capacity to measure variable importance (Cutler et al. 2007).

My objective was to detect the factors that affect, at different scales, breeding site selection of colonial birds in a human-influenced landscape. Japan is a good example of a highly human-influenced complex landscape where we can still find birds breeding in mixed-species colonies, and where we can obtain precise data of land uses and breeding locations distribution. I used location data for heron and egret colonies distributed in the fairly complex agricultural landscape of Ibaraki and surrounding prefectures in Japan in 2011, and compared the land types surrounding the colonies with those around unoccupied sites using geographic information systems (GIS) techniques. Then I applied a RF algorithm to analyze the importance of the different land-use variables at different scales for establishing a colony.

1.2 Methods

Study area and species

The study area was the central and southern regions of Ibaraki Prefecture and some bordering regions of Tochigi, Gunma, Saitama and Chiba prefectures in central Japan (Fig. 1). The region is limited by mountains to the north-west, by the Pacific Ocean to the east and by the Tone River to the south, with a total area of approximately 10,022 km². It is mainly a low altitude plain and its main geological feature is the presence of Lake Kasumigaura. The predominant human-influenced land use is agricultural, rice fields being the dominant cultivation (8.5% of the study area). There are residential areas of various sizes and forest patches spread all around the region. Six species of herons and egrets, Grey Heron (*Ardea cinerea*), Great Egret (*A. alba*), Little Egret (*Egretta garzetta*), Intermediate Egret (*E. intermedia*), Cattle Egret (*Bubulcus ibis*) and Black-crowned Night Heron (*Nycticorax nycticorax*), breed mainly in mixed-species colonies every year in the study area. They build their nests on conifers, broad-leaf trees

and in bamboo thickets (Environmental Agency of Japan 1994).

Colony locations

Twenty colony locations were recorded by ground surveys (Mashiko and Toquenaga 2013) in the study area during the breeding season, from March to August, of 2011. In my study, I aimed to analyze the land use surrounding the colonies, so I referred to the colony data for 2011 due to the limited availability of land-use maps with sufficient resolution up to this year. The site selection model for this study was based on differences in the areas surrounding colonies and the those surrounding unoccupied sites. For the statistical model to be consistent, I needed to compare the same number of colonies and unoccupied sites, so I randomly chose 20 locations, which corresponds with the number of observed colonies in this study, where a colony could, potentially, be formed. Locations available for colonization were defined as follows. First, a rectangular area of the study area was arbitrarily delimited ($35^{\circ} 52' 32''\text{N}$ - $36^{\circ} 35' 43''\text{N}$, $139^{\circ} 35' 36''\text{E}$ - $141^{\circ} 00' 00''\text{E}$). Second, forest areas below an altitude of 100 m were selected, as colonies are seldom found at higher elevations in this area of Japan (Fig. 1).

Unoccupied sites were then determined by randomly selecting numbers corresponding to the IDs of raster cells among the potentially available regions. Because the random selection of sites could lead to slightly different results, I created 30 different data sets of 20 points each, and analyzed the data for each set.

Landscape variables

To analyze the information on land use in the areas surrounding the colonies and unoccupied sites, I used a land-use map of Japan provided by the JAXA. This map was created with multi-satellite imagery from 2011. The final map had an approximately 45-m pixel size. The processing and classification details are explained at http://www.eorc.jaxa.jp/ALOS/lulc/lulc_jindex.htm.

Eight relevant land-use variables for herons and egrets were identified as follows:

bare land, evergreen forest, deciduous forest, grassland, crop land, paddy field, urban area and body of water (Fig. 2). The selection of the variables was based on previous knowledge of the ecology of the heron species (Tojo 1996; Lane and Fujioka 1998) and examination of the satellite images. Layers of circular buffer zones were created around the colonies and the 20 randomly selected points, and areas of the eight selected land-use variables were identified within each. The radii of the buffer zones were 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, and 40 km, distances that cover the inter-distance of colonies with a high resolution as well as areas at coarser scales (Fig. 1). The *raster* package (Hijmans and van Etten 2012) in R 2.15.2 (R Development Core Team 2011) was used to extract the land-use information.

Statistical analyses

Random forest model

A colony site selection model was created using a random forest (RF) algorithm (Breiman 2001), which relies on the ideas of classification and regression trees (CART) (Breiman et al. 1984), and on bagging methods (Breiman 1996). Classification trees are used for predicting the membership of cases in the classes of a categorical dependent variable by measuring one or more predictor variables. A basic classification tree algorithm uses splits of subsets (nodes) of the feature space (set of samples) into two descendant subsets. Recursively applies this splitting on each subset until a stop criteria is reached. RF algorithm applies the ideas of bagging methods to CART algorithms, by creating new training sets by random sampling (bootstrap sample). For each bootstrap sample, a different tree is performed. Besides, RF applies a second randomization, selecting a random subset of predictors for each split. The final prediction is the class with maximum votes among all trees.

For input data against models, I used the area of each landscape variable surrounding the 20 colonies and the 20 randomly chosen unoccupied sites. One model for each scale around the colonies (each buffer radius) was built. For example, for

creating the 3-km scale model the eight land-use variables measured at a buffer radius of 3 km were used. Randomized models were created to compare the predictive power of my colony distribution models with the predictive power of the models when the outputs of the training sets (presence and absences) were randomly permuted. One third of the data was left out for each bootstrap sample (the out-of-bag data, OOB) for each tree and 500 trees were created for each classification model. The process was repeated for each of the 30 random points data sets.

Predictive accuracies

I calculated the accuracy of the model as $1 - \text{OOB error estimate}$ (Breiman 2001). As some of the data is left out at each tree when training the RF algorithm, we do not need a cross-validation to get an unbiased estimate of the test set error. RF uses the one-third of the data left out of the bootstrap sample of a certain tree to get a classification. Each sample data will be tested then in one-third of the trees approximately. The proportion of times that the classification result is not equal to the true class, averaged among all cases, is the OOB error estimate. The results of the RF model can be slightly different each time it is performed, even when using the same parameters, so ten models were built for each scale. The mean value of the OOB error for the ten repetitions was used to measure the accuracy for one model. Disparate accuracy results are also obtained for the randomized models for different permutations of the presence-absence values, so the mean value of the accuracy out of 100 repetitions was used. The prediction accuracy of models tended to vary among random sets of hypothetical unoccupied colony locations. Each random set of hypothetical colonies plays a role similar to that of the data for supervised learning in a neural network model. Bad and good data could, respectively, cause low and high accuracies. So I calculated not only means but also maximum accuracies of the RF models for each scale. Mean accuracies represent the overall tendency of model performance across the scales, but the best performance of a RF model at each scale should be evaluated by the maximum accuracy values. This is

because I wanted to use the set that best explained the colony distribution, not being interested in one “average model” for analyzing the important explanatory variables.

Variable importance

For the scales with the best accuracy models, the importance of each land use variable was analyzed using the mean decrease accuracy index (Breiman 2001). To do this, I selected the data-set with the maximum accuracy for a specific scale among all sets. To calculate the mean decrease accuracy index, the prediction error of the OOB is recorded for each tree. Then, the variable is permuted and the error is recorded again. The difference between the two are then averaged over all trees and normalized by the standard deviation of the differences. As I found some variance each time the RF was constructed, the average of the mean decrease accuracy index over ten models was used.

Variable effect on colony presence

Partial dependence plots (Friedman 2001; Hastie et al. 2005) were used to graphically characterize relationships between individual predictor variables and predicted probabilities of colony presence (Cutler et al. 2007). The vertical axis of this plot is a measure of the marginal effect of a certain explanatory variable on the class probability. In my study, the vertical axis represents the effect of the area of each land use on the probability of colony presence. The horizontal axis represents the value of the variable for which partial dependence is sought. I interpreted that a colony “preferred” certain land use when the effect on the probability of colony presence was higher for higher areas of that land type. A colony, therefore, “avoided” a land type when the effect on the probability of colony presence decreased as that land-use area increased. The *randomForest* package (Liaw and Wiener 2002) in R was used for model creation and analysis.

1.3 Results

The accuracy values of the randomized models strongly depended on the random permutation of the output values (colony presence or absence) but, on average, all the scales showed an accuracy of between 47% and 48% (Fig. 3). The 4-km scale model had the best accuracy on average, followed by the 15-, 30-, 10- and 1-km scales. The 4-km model had a relatively high accuracy for all the sets but, in most cases, it was not the scale with the best accuracy within the set. In contrast, 10-km or 15-km scales performed the best in many of the sets. There is moderate collinearity among land-use variables and the proportions of land-use patterns around colonies at selected scales were not strongly biased from the average of the study area (Table 4, 5 and 6 in the Appendix 1).

The shape of the maximum accuracies graph was quite similar to that of the mean accuracies graph, but the 4-km models did not have the maximum values for accuracy (Fig. 3). The scale with the highest maximum accuracy was the 10-km scale, with almost 78% accuracy, 30% more accurate than its equivalent randomized model.

Some scales were ineffectual for explaining colony distribution, even when nearby scales were important. This was the case for the 0.5-, 2-, 5-, 6- and 7-km scales, showing low mean accuracy levels (Fig. 3). Variable importance was analyzed for the sets that showed maximum accuracies for the 4-, 10- and 15-km scale models (Fig. 4). Evergreen forest was the most important variable for explaining colony distribution at the 4-km scale, followed by urban and crop areas (Fig. 4a). The importance of the evergreen forests decreased as scale size increased, being very low for 15 km. For the 10- and 15-km scales, urban areas and bare land were the most important variables for the model (Fig. 4b and c). Paddy fields were revealed to have low importance for the model, although their importance increased with scale.

Partial dependence plots of each landscape variable were very similar for the different scales, so I chose the most important ones of the best explanatory scales to analyze the relationship between the area of each landscape predictor and the

probability of colony presence. Evergreen forest was an attractive land type for establishing a colony (Fig. 5). For the 4-km scale, regions made up of less than 7% evergreen forest were strongly avoided. On the other hand, areas made up of more than 35% evergreen forest were neither attractive nor repellant. I used the 10-km model to analyze the importance of urban areas, as this variable was the most important at this scale. Regions made up of at least 10% urban areas were strongly avoided (Fig. 5). Colonies tended to be established where urban areas made up between 5% and 10% of the area within the 10-km radius. Bare soil produced similar results to those of urban land types. Crops and paddy fields were attractive land types when they were relatively important variables in some models.

1.4 Discussion

I obtained highly accurate colony site selection models for 4-, 10- and 15-km scales. At the 4-km scale, evergreen forest was the most important variable, being an attractive factor. At the 10- and 15-km scales, urban areas and bare land were the main variables for explaining the models, both of which were avoided when they were present in high ratios. Paddy fields, the main foraging habitat for all the species, was not a high-importance variable for any of the scales, although its importance increased with scale. The highest accuracy obtained was 78% for the 10-km model.

It was revealed that there are two very distinct general scales by which site selection is most affected: the 4-km range (medium scale) and the 10-15 km range (large scale). The 4-km scale had the highest average accuracy for all sets, the 10-km scale had the highest maximum accuracy values and the 15-km scale had high average and maximum values. The 30-km scale also exhibited good performance on average and when analyzing the maximum accuracy. The variables that best explained the distribution of the colonies at 30 km were very similar to those at 15 km, so I consider this scale to be highly correlated with the 15-km scale. Herons and egrets could be mainly using these

two scales to decide where to establish their colonies. Scales in between these two and also small scales were of very low importance for colony site selection.

Scale dependence could explain the diversity of results obtained in previous studies on the accuracy of various models used for explaining the distribution of colonies of herons and egrets, and could also provide a different explanation for which factors affect colony distribution. For example, Gibbs and Kinkel (1997) used a 15-km scale to explain colony distribution. On the other hand, Fasola and Alieri (1992) and Tourenq et al. (2004) used a 5-km scale, while Boisteau and Marion (2007) used 25 km. In these cases, foraging habitats were used as predictors, so the scales were justified by the observed foraging ranges. My study reveals that studying one single scale could lead to models with low predictive power, and with potentially fatal consequences when erroneously considering explanatory variables as key factors, while other variables could be much more important at different scales.

Most of the past studies on herons and egrets analyzed areas not highly inhabited by humans. Scale is a crucial factor when studying colony site selection, and it could depend strongly on the landscape configuration. Among the studies that have considered different scales, Kelly et al. (2008) showed that the 1-km scale was best for explaining the colony distribution of herons and egrets in tidal marshes. Parkes et al. (2012) also found that small scale (below 1.5 km) was very important for the sites of Cattle Egret colonies in upland residential areas (although their model did not consider interaction between the explanatory variables). Bigger scales (from 1 to 10 km), however, were more important in rice field related colonies in France (Tourenq et al. 2004), depending on the study species. The scale at which the colony site selection is performed could strongly depend on the land types surrounding the colonies. Landscape configuration and different levels of fragmentation of important habitats could be crucial factors for determining which scales allow the species to assess the surroundings and choose optimal colony sites.

Evergreen forest, an attractive land type for herons and egrets, was the most

important variable at the medium scale, while repelling factors such as urban areas and bare land were most important at large scales. Evergreen forest includes bamboo thickets, the most important nest substrate for herons and egrets, followed by trees, which explains the higher importance of this land use over deciduous forest. Identifying regions with greater amounts of bamboo thickets, places available for the establishment of a colony, could be one of the most important steps of habitat assessment after the arrival of the individuals. The importance of the medium scale for colony site selection might reflect an effect of the study area landscape patterns, where high densities of evergreen forest could be easily detected at the 4-km scale. At different scales, the distribution patterns of this land use could make the habitat assessment more difficult for herons and egrets, being unable to discriminate high density forest regions at small or large scales. Colonies demonstrated, however, a preference for lower urban- and bare-area densities (the latter being highly related to human-influenced land types) at large scales. When food and forest availability does not determine colony locations, mechanisms such as avoiding disturbances become more important. There is evidence of heron and egret colonies avoiding urban areas at small scales (Fasola and Alieri 1992) but my current study is the first study to show this effect on the distribution of colonies at large scales. Avoiding large urban areas could be advantageous in terms of lower levels of disturbances for the colonies. In Saitama Prefecture, located in the south-west of my study area and in the northern suburbs of Tokyo, there has been a great deal of urban development since 1960; population sizes of herons and egrets have decreased and some colonies have disappeared, even where paddy fields and forest patches that are available for the establishment of colonies remain (Narusue 1992). The effect of the high density of urban areas at the large scale could have been the main reason for the extinction of heron and egret colonies in central Japan.

Paddy fields, the main foraging areas for the study species, was revealed to be unimportant for predicting colony site, in contrast to results of past studies on herons and egrets. However, its importance seemed to increase with scale. Previous studies

showed weak relationships between food habitats and the colony distribution of these species (Fasola and Alieri 1992; Boisteau and Marion 2007) in agricultural landscapes, but they included only attractive factors, and no repelling factors, in their models. Fasola and Canova (1991) had the same results for mixed-species colonies of gulls and terns, where foraging sites were not an important factor in colony location, allowing us to infer that this could also be true for other wading bird families. Landscape complexity in developed urban regions, as exist in the present study area, could lead to difficulties in the assessment of the quality of food habitats for herons and egrets. Difficulties for many species on the evaluation of food availability for the breeding period have been widely discussed (Orians and Wittenberger 1991; Fuller 2012). Also, the capacity for evaluating the amount and quality of foraging habitats could be damaged by surrounding urban landscapes (Battin and Lawler 2006). My study area is highly affected by urban development and some regions are experiencing very rapid landscape changes, so even resident species could have problems assessing habitat quality, despite that their evaluation process continues even during non-breeding seasons.

The study of positively and negatively associated factors on different scales revealed two main characteristics of colony distribution of herons and egrets: colony sites were established near large amounts of evergreen forests at medium-scale distances and where there are less urban areas at higher-scale distances. Including avoidance factors could improve the performance of predictive bird colony distribution models, especially when analyzing at large scales. Conservation and management of colonial wading birds living in human-influenced landscapes should not only focus on the maintenance of the available foraging and breeding habitats, but also on controlling urban development around the colonies.

In the present study, the predictive accuracy for the best model (78%) was higher than that of a predictive colony distribution model for mixed-species herons and egrets by Kelly et al. (2008) (68%) and similar to the predictive model for Cattle Egret by Parkes et al. (2012) (79%). This shows that the RF method can be a good tool for

predicting the colony sites of herons and egrets, and that it can handle the complexity of human-influenced landscapes. However, the predictive power of the RF model was lower than that of those using other colonial but non-wading and single-species data (Bustamante (1997): 84%; Lauver et al. (2002): 82%; Heinänen et al. (2008): area under the curve, AUC =0.91). One reason may be the relative simplicity of those data, where linear statistical tools could be sufficient for explaining the distribution data. Alternative explanation would be the lack of reliability in comparing different accuracy indicators. The choice of predictive ability indicator can have a large impact on the results, especially when comparing regression models with machine learning methodologies such as CART or RF (Myrtveit et al. 2005).

The use of GIS in combination with the newest classification techniques, such as RF, seems to be an appropriate analysis method for complex ecological data that includes the complexities of human-influenced agricultural landscapes. Logistic regressions are often powerless, especially with complicated land-use patterns. Resultant logistic models are often very specific to data sets, and the same model cannot be applied to other similar data sets. Neural networks and decision trees are alternatives for such complicated land use patterns, but my approach of using RF has two prominent advantages against neural networks and other decision trees. The first is that we can evaluate the relative importance of competing variables, or land-use types. The second is that RF can avoid over specialization and remain generalized for similar problems (Breiman 2001).

In summary, I applied a RF algorithm for analyzing the distribution of herons and egrets colonies in a strongly human-influenced landscape in Japan, and I was able to clarify some important characteristics of the colony site selection strategies of these species. I strongly believe that non-linear methods as RF are more appropriate when dealing with predictive suitability models for birds living in highly human-influenced landscapes than classic linear methods. These methodologies could be a big help to rethink and improve the conservation plans of those species threatened by the advance of the agricultural and urban landscapes.

Chapter 2

Extrapolation of Random Forest Models Shows Scale Adaptation in Heron and Egret Colony Site Selection against Landscape Complexity

Availability of certain habitats or landscape configurations can cause differential habitat selection in animal species. Landscape complexity can affect foraging scales, home ranges and movement, but its effect on habitat selection is little documented. I aimed to examine differences in colony site selection of herons and egrets in different regions. I studied if landscape complexities could affect their scale of selection and their habitat preferences.

I used colony distribution data of two nearby regions, Ibaraki and Chiba in Japan, combined with land-use maps and created random forest models for analyzing habitat preferences and important scales of selection. I made a cross-validation of models each of which adjusted for Ibaraki or Chiba land use map with changing scales.

The scales that best explained colony distribution were 1-, 4-, 10- and 15-km in Ibaraki, and 1- and 10-km in Chiba. Evergreen forest was the most important variable for Ibaraki at 4-km and for Chiba at 1-km. However, the importance of the other variables differed for other models. Cross-validation showed that herons and egrets had the same habitat preferences at 4-km scale in Ibaraki and at 1-km scale in Chiba. The

scale of selection was reduced at Chiba, in which the main foraging resource for herons and egrets presented higher complexity.

Differences in landscape complexities did not affect habitat preferences but resulted on differences on the scale of selection. Analyzing the effects of landscape configuration on habitat selection is useful for studying animal behavior difficult to observe directly, as collective decision-making.

Chapter 3

Balance Between Site Fidelity and Habitat Preferences in Colony Site Selection by Herons and Egrets

Habitat preferences and site fidelity can be affecting together how species choose their living habitats. It is often difficult to grasp both phenomena at once, so many ecologist tried to explain the distribution of the species at regional scales focusing on one of them. However, studying the relative importance of habitat preferences and site fidelity working simultaneously can be helpful for understanding the habitat selection process as a whole.

To study the relationship between habitat preference and colony site fidelity in colonial birds, I analyzed the relative importance of these two factors in heron and egret communities over a span of 28 years. To do this, I used long-term data for colony locations of herons and egrets around Ibaraki prefecture, Japan. I hypothesized that if the level of colony site fidelity increased, habitat preferences would become less important for explaining colony distribution, and vice versa. To quantify site fidelity exhibited by herons and egrets, I created an index measuring the site fidelity level of every colony for each year of the study. I used Landsat satellite images together with a ground-survey-based map of Ibaraki from 2011 to create land-use maps of the area of study for past years and to determine the land uses surrounding the colonies. Combining the estimated colony site fidelity with the land-use data, I created habitat selection predictive models using a random forest algorithm (Breiman 2001). Using the

predictive accuracies and the variable importance estimations of the predictive models for each year, I analyzed the changes in the importance of land-use preferences and colony site fidelity over the years.

I observed high levels of colony site fidelity every year of the study, making it a very important predictor for explaining colony distribution. Colony site fidelity increased drastically in its importance relative to habitat preferences in the last five years. The observed increase of collective site fidelity was caused by the growth of the Grey Heron population in the last years.

Land-use maps for past years along with long-term data for colony distribution were helpful in studying habitat selection not as a static mechanism, but as a changing balance between habitat preferences and site fidelity. Further long-term studies that include other factors, such as conspecific attraction or heterospecific attraction, competition and predation, are required for a better understanding of habitat selection and to clarify the influence of environmental alterations on the balance between those factors.

General Discussion

In this study, I combined colony distribution data with land-use maps to study colony site selection of herons and egrets in Ibaraki, Japan, and its changes across landscapes and over time. In Chapter 1, I described my study of the habitat preferences of herons and egrets for selecting their colony sites in 2011. I used a random forest (RF) algorithm to analyze the land uses surrounding the colonies at different scales. I found that there were two main scales at which herons and egrets selected their colony sites: medium (4 km) and large (10-15 km). Colonies were attracted to areas with large amounts of evergreen forests at the medium scale while at large scales they avoided high-density urban areas. Paddy fields, the main foraging habitat for herons and egrets, were not important for explaining colony sites. Previous studies used attractive factors, mainly foraging areas, to explain colony distributions, but my study is the first to show the major importance of repellent factors at large scales. The newest non-linear methodologies, such as RF, are needed when modeling complex variable interactions when organisms are distributed in complex landscapes.

In Chapter 2, I compared colony site selections between Ibaraki and Chiba prefectures. The novelty of my methodology was applying a cross-validation by creating a colony site selection model at certain scale and applying it to the other region at different scales, comparing habitat preferences across scales. I found that herons and egrets shared the same habitat preferences for selecting colony sites in both regions. However, an increase of complexity in the paddy fields of Chiba resulted in a reduction of the scale of selection in this region. Previous studies have shown how landscape complexity can affect the scale of foraging behavior, movement or habitat range (Stapp and van Horne 1997; Ritchie 1998; Haskell et al. 2002; DeJagger et al. 2011), but this is the first study to show evidence of adaptation to scale for colony site selection.

In Chapter 3, I described colony site selection of herons and egrets in Ibaraki over a span of 28 years. I analyzed the balance between habitat preferences and colony site fidelity for colony site selection and its changes over the time. High levels of colony site

fidelity was observed every year of the study, making this an important factor for explaining colony distribution in Ibaraki. The importance of colony site fidelity increased in relation to habitat preferences in the last five years of the study. The observed increase of colony site fidelity levels might have been caused by the growth of the Grey Heron population in the last years. The Grey Heron is a key species for colony establishment, so its population increase together with its influence on the persistence of a colony over the years, might have caused the increase in colony site fidelity.

Analyzing the relationship between land-use data and species distribution with a non-linear algorithm, such as RF, seems to be an appropriate methodology when dealing with complex ecological data like human-influenced agricultural landscapes. It allowed us to deal with the complex interactions among land-use data and their spatial correlation, and it provided an effective method for detecting the important scales of selection and for identifying the most important variables influencing colony site selection of herons and egrets. Also, this technique provides a good tool for extrapolating the models across landscapes (Prasad et al. 2006), so it allowed us to compare colony site selection models between different regions, as well as to analyze the changes of these models over the time. Non-linear methodologies, such as RF, can help to improve habitat selection models and their predictive abilities across space and time, especially for those species distributed in complex human-influenced landscapes.

The combination of colony distribution data with high-resolution land-use maps was fundamental for analyzing colony site selection of herons and egrets. Land-use data with sufficient spatial resolution is needed if we aim to analyze a species' interaction with its environment at a local scale (Rose et al. 2014). Furthermore, land-use data of good quality, based on ground surveys, such as the JAXA map of 2011 that I used in this study, is important for a precise analysis of the relationships of the species and their surrounding habitats, especially in fragmented and complex landscapes like Ibaraki. I also used this land-use map of 2011, combined with Landsat historical satellite images, to create land-use maps for previous years. Combining these maps with the long-term

data of colony locations compiled by the Population Ecology Laboratory during the last 28 years allowed me to observe changes in colony site selection of herons and egrets over the years. Long-term analysis of colony site selection is important as there exists high inter-annual variability, so restricting a study to several years could lead to a poor understanding of the site selection process. The recent possibility of joining long-term species distribution data with historical environmental maps is gaining importance in analyzing ecological processes that can vary year to year, as well as in analyzing the recent species responses to landscape and climate changes (Collins 2001; Rose et al. 2014).

Herons and egrets in Ibaraki selected their colony sites based on habitat preferences that depended on scale. I showed that this scale can be affected by landscape complexity. Site fidelity can also affect colony site selection, and its influence can vary over time. These differences across scales, landscapes, and over time, could explain the disparate results in previous studies on colony site selection of herons and egrets (Gibbs et al. 1987; Fasola and Alieri 1992; Tourenq et al. 2004; Boisteau and Marion 2007; Kelly et al. 2008; Fasola et al. 2010; Parkes et al. 2012). Studying one single scale can enormously limit the ability of a habitat selection model to explain colony distribution. Additionally, researchers should include, whenever possible, collective site fidelity as a variable for explaining colony distribution, as the high tendency of herons and egrets to use the same colony sites year after year, strongly affects the colony site selection of these species, even when individual site fidelity can be very low.

Studying site selection of colonial breeding birds is not an easy task, as it is difficult to directly observe animal behaviors that affect the whole group. Using habitat selection models at the landscape level was very useful for clarifying some behavioral aspects of the colony site selection of herons and egrets. Further long-term studies including other factors, such as conspecific or heterospecific attraction, competition, or predation avoidance, are needed for a better understanding of the habitat selection of colonial breeding avian species.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Associate Professor Yukihiro Toquenaga, for his scientific guidance, enthusiasm and patience. Not only has he taught me many tools that will be of great help for the rest of my career, but he has also always encouraged me to find the newest and most original approaches to face any scientific challenge. I would also like to thank my thesis defense committee members, Associate Professor Masahiro Fujioka, Associate Professor Mitsuru Hirota and Associate Professor Yuji Inagaki, for their interest and valuable comments on my work.

I am also very grateful to Assistant Professor Kazuharu Ohashi for his advice and kind support, and to the rest of the members of the Population Ecology Laboratory for their kindness and help in my everyday life. Also to Dr. Miyuki Mashiko, with whom I shared many hours of field research, without her this work would not have been possible to complete.

I would like to make special mention of all the people that provided data on the locations of herons and egrets colonies. Associate Professor Yukihiro Toquenaga started with the census of colonies in the early 80's and members of the Population Ecology Laboratory continued his work making yearly censuses: Mr. Kazuhide Hikosaka, Mr. Tomomi Anzai, Mr. Saburo Ohshio, Mr. Naoki Tanigaki, Ms. Ryoko Yamagishi, Mr. Takayuki Tohyama, Mr. Yakuo Yamaguchi, Mr. Haruki Takeda, Ms. Chieko Koshida, Ms. Akiko Abe, Mr. Kohey Nabeya and Dr. Miyuki Mashiko. Dr. Miyuki Mashiko compiled and organized a great amount of historical data of colonies in Ibaraki. Many other people provided information data on colony locations that made this work possible: Associate Professor Masahiro Fujioka, Mr. Susumu Ikeno, Mr. Tadao Kawasaki, Mr. Seiichi Kasuga, Mr. Takashi Uetake, Mr. Masahiro Seido, Mr. Kohei Takeda, Mr. Koju Ueda, Mr. Makoto Yamaguchi, Mr. Takeru Mikami and the Heron Research Group of the Approved Specified Nonprofit Corporation Friends of Gytoku Bird Observatory.

I would like to thank the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan for the Japanese Government Scholarship Program,

which supported my stay at the University of Tsukuba.

I would not have completed this journey without my parents, whose unconditional love and support travelled everyday thousands of kilometers to give me the peace and serenity to keep on with my work and do the best of myself. I am also thankful for the support of my brother, my uncles, cousins and my friends in Spain, I always felt like they were just around the corner. Finally to my friends in Tsukuba, who “suffered” me and helped me so much all these years and to whom I will always be grateful.

References

- Anderson, D. J. and P. J. Hodum (1993). Predator behavior favors clumped nesting in an oceanic seabird. *Ecology* *74*, 2462–2464.
- Battin, J. and J. J. Lawler (2006). Cross-scale correlations and the design and analysis of avian habitat selection studies. *The Condor* *108*, 59–70.
- Boisteau, B. and L. Marion (2007). Habitat use by the Grey Heron (*Ardea cinerea*) in eastern france. *Comptes rendus biologies* *330*(8), 629–34.
- Breiman, L. (1996). Bagging predictors. *Machine Learning* *24*, 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning* *45*, 5–32.
- Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984). *Classification and regression trees*. Belmont, California: Wadsworth International Group.
- Brown, C. R. and M. B. Brown (2001). Avian coloniality. *Current Ornithology* *16*, 1–82.
- Brown, C. R., M. B. Brown, and E. Danchin (2000). Breeding habitat selection in Cliff Swallows: the effect of conspecific reproductive success on colony choice. *Journal of Animal Ecology* *69*, 133–142.
- Brown, C. R. and B. Rannala (1995). Colony choice in birds: models based on temporally invariant site quality. *Behavioral Ecology and Sociobiology* *36*, 221–228.
- Bustamante, J. (1997). Predictive models for Lesser Kestrel *Falco naumanni* distribution, abundance and extinction in southern Spain. *Biological Conservation* *80*, 153–160.

- Cody, M. L. (1985). *Habitat selection in birds*. San Diego: Academic Press.
- Collins, S. L. (2001). Long-term research and the dynamics of bird populations and communities. *The Auk* 118, 583–588.
- Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler (2007). Random forests for classification in ecology. *Ecology* 88, 2783–2792.
- Danchin, E., T. Boulinier, and M. Massot (1998). Conspecific reproductive success and breeding habitat selection: implications for the study of coloniality. *Ecology* 79, 2415–2428.
- Danchin, E. and R. H. Wagner (1997). The evolution of coloniality: the emergence of new perspectives. *Trends in Ecology and Evolution* 12, 342–347.
- DeJagger, M., F. Weissing, P. Herman, B. Nolet, and J. van de Koppel (2011). Lévy walks evolve through interaction between movement and environmental complexity. *Science* 332, 1551.
- Doligez, B., C. Cadet, E. Danchin, and T. Boulinier (2003). When to use public information for breeding habitat selection? The role of environmental predictability and density dependence. *Animal Behaviour* 66, 973–988.
- 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.
- Fasola, M. and R. Alieri (1992). Conservation of heronry Ardeidae sites in North Italian agricultural landscapes. *Biological Conservation* 62, 219–228.
- Fasola, M. and L. Canova (1991). Colony site selection by eight species of gulls and terns breeding in the «Valli di Comacchio» (Italy). *Italian Journal of Zoology* 658, 261–266.
- Fasola, M., D. Rubolini, E. Merli, E. Boncompagni, and U. Bressan (2010). Long-term trends of heron and egret populations in Italy, and the effects of

- climate, human-induced mortality, and habitat on population dynamics. *Population Ecology* 52, 59–72.
- Forbes, L. S. and G. W. Kaiser (1994). Habitat choice in breeding seabirds: when to cross the information barrier. *Oikos* 70, 377–384.
- Fretwell, S. (1969). On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheoretica* 19, 45–52.
- Friedman, J. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29, 1189–1232.
- Fuller, R. J. (2012). *Birds and habitat: relationships in changing landscapes*. Cambridge, UK: Cambridge University Press.
- Gibbs, J. and L. Kinkel (1997). Determinants of the size and location of Great Blue Heron colonies. *Colonial Waterbirds* 20, 1–7.
- Gibbs, J., S. Woodward, M. Hunter, and A. Hutchinson (1987). Determinants of Great Blue Heron colony distribution in coastal Maine. *The Auk* 104, 38–47.
- Gill, F. (2007). *Ornithology*. New York: Freeman and Company.
- Greenwood, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* 28, 1140–1162.
- Haskell, J., M. Ritchie, and H. Olf (2002). Fractal geometry predicts varying body size scaling relationships for mammal and bird home ranges. *Nature* 418, 527–530.
- Hastie, T., R. Tibshirani, J. Friedman, and J. Franklin (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer* 27, 88–85.
- Heinänen, S., M. Rönkä, and M. V. Numers (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.

- Hijmans, R. J. and J. van Etten (2012). Raster: geographic analysis and modeling with raster data. *R package version 2.1-25*.
- Keating, K. and S. Cherry (2004). Use and interpretation of logistic regression in habitat selection studies. *Journal of Wildlife Management* 68, 774–789.
- Keister, A. R. (1979). Conspecifics as cues: a mechanism for habitat selection in the Panamanian Grass Anole (*Anolis aeneus*). *Behavioral Ecology and Sociobiology* 5, 323–330.
- Kelly, J., D. Stralberg, K. Etienne, and M. McCaustland (2008). Landscape influence on the quality of heron and egret colony sites. *Wetlands* 28, 257–275.
- Kushlan, J. A. and J. A. Hancock (2005). *The Herons*. New York: Oxford University Press.
- Lack, D. (1968). *Ecological adaptations for breeding in birds*. London, UK: Methuen.
- Lane, S. and M. Fujioka (1998). The impact of changes in irrigation practices on the distribution of foraging egrets and herons (Ardeidae) in the rice fields of central Japan. *Biological Conservation* 83, 221–230.
- Lauver, C. L., W. H. Busby, and J. L. Whistler (2002). Testing a GIS model of habitat suitability for a declining grassland bird. *Environmental management* 30, 88–97.
- Liaw, A. and M. Wiener (2002). Classification and regression by randomForest. *R news* 2, 18–22.
- Magurran, A. E., S. R. Baillie, S. T. Buckland, J. M. Dick, D. A. Elston, E. M. Scott, R. I. Smith, P. J. Somerfield, and A. D. Watt (2010). Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology and Evolution* 25, 574–582.
- Martin, T. E. (1993). Nest predation and nests sites: new perspectives on old patterns. *Bioscience* 43, 523–532.
- Mashiko, M. and Y. Toquenaga (2013). Increasing variation in population size and

- species composition ratio in mixed-species heron colonies in Japan. *Forktail* 29, 71–79.
- Møller, A. P. (1987). Advantages and disadvantages of coloniality in the swallow, *Hirundo rustica*. *Animal Behaviour* 35, 819–832.
- Myrtveit, I., E. Stensrud, and M. Shepperd (2005). Reliability and validity in comparative studies of software prediction models. *IEEE Transactions on Software Engineering* 31, 380–391.
- Narusue, M. (1992). Changes in the distribution and extent of breeding colonies of egrets in Saitama Prefecture. *Strix* 11, 189–209.
- Orians, G. and J. Wittenberger (1991). Spatial and temporal scales in habitat selection. *The American Naturalist* 137, S29–S49.
- Ozesmi, S. L. and M. E. Bauer (2002). Satellite remote sensing of wetlands. *Wetlands Ecology and Management* 10, 381–402.
- Parkes, M. L., M. A. Mora, and R. A. Feagin (2012). Using scale, cover type and GIS to evaluate nuisance Cattle Egret colony site selection. *Waterbirds* 35, 56–63.
- Prasad, A. M., L. R. Iverson, and A. Liaw (2006, March). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199.
- R Development Core Team (2011). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Richner, H. and P. Heeb (1995). Is the information center hypothesis a flop? *Advances in the Study of Behavior* 24, 1–45.
- Ritchie, M. (1998). Scale-dependent foraging and patch choice in fractal environments. *Evolutionary Ecology* 12, 309–330.
- Rolland, C., E. Danchin, and M. Fraipont (1998). The evolution of coloniality in birds in relation to food, habitat, predation and life-history traits: a comparative

- analysis. *The American Naturalist* 151, 514–529.
- Rose, R. A., D. Byler, J. R. Eastman, E. Fleishman, G. Geller, S. Goetz, L. Guild, H. Hamilton, M. Hansen, R. Headley, J. Hewson, N. Horning, B. A. Kaplin, N. Laporte, A. Leidner, P. Leimgruber, J. Morissette, J. Musinsky, L. Pintea, A. Prados, V. C. Radeloff, M. Rowen, S. Saatchi, S. Schill, K. Tabor, W. Turner, A. Vodacek, J. Vogelmann, M. Wegmann, D. Wilkie, and C. Wilson (2014). Ten ways remote sensing can contribute to conservation. *Conservation Biology* (*in press*).
- Seppelt, R. and A. Voinov (2002). Optimization methodology for land use patterns using spatially explicit landscape models. *Ecological Modelling* 151, 125–142.
- Stapp, P. and B. van Horne (1997). Response of Deer Mice (*Peromyscus maniculatus*) to shrubs in shortgrass prairie: linking small-scale movements and the spatial distribution of individuals. *Functional Ecology* 11, 644–651.
- Tarvin, K. A. and M. C. Garvin (2002). Habitat and nesting success of Blue Jays (*Cyanocitta cristata*): importance of scale. *The Auk* 119, 971–983.
- Terhune, J. M. and S. W. Brillant (1996). Harbour Seal vigilance decreases over time since haul out. *Animal Behaviour* 51, 757–763.
- Thomson, R. L., J. T. Forsman, and M. Mönkkönen (2003). Positive interactions between migrant and resident birds: testing the heterospecific attraction hypothesis. *Oecologia* 134, 431–438.
- Tojo, H. (1996). Habitat selection, foraging behavior and prey of five heron species in Japan. *Japanese Journal of Ornithology* 45, 141–158.
- Tourenq, C., S. Benhamou, N. Sadoul, A. Sandoz, F. Mesleard, J. Martin, and H. Hafner (2004). Spatial relationships between tree-nesting heron colonies and rice fields in the Camargue, France. *The Auk* 121, 193–202.
- Trillmich, F. and K. Trillmich (1984). The mating systems of pinnipeds and marine iguanas: convergent evolution of polygyny. *Biological Journal of the Linnean*

Society 21, 209–216.

Wagner, R. H., M. D. Schug, and E. S. Morton (1996). Condition-dependent control of paternity by female Purple Martins: implications for coloniality. *Behavioral Ecology and Sociobiology* 38, 379–389.

Ward, P. and A. Zahavi (1973). The importance of certain assemblages of birds as information-centers for food-finding. *Ibis* 115, 517–534.

Wiens, J. A. and B. T. Milne (1989). Scaling of ‘landscapes’ in landscape ecology, or, landscape ecology from a beetle’s perspective. *Landscape Ecology* 3, 87–96.

Wittenberger, J. F. and G. L. Hunt (1985). The adaptive significance of coloniality in birds. In D. S. Farner, J. R. King, and K. D. Parkes (Eds.), *Avian biology (Vol. 8)*, pp. 1–78. Academic Press.

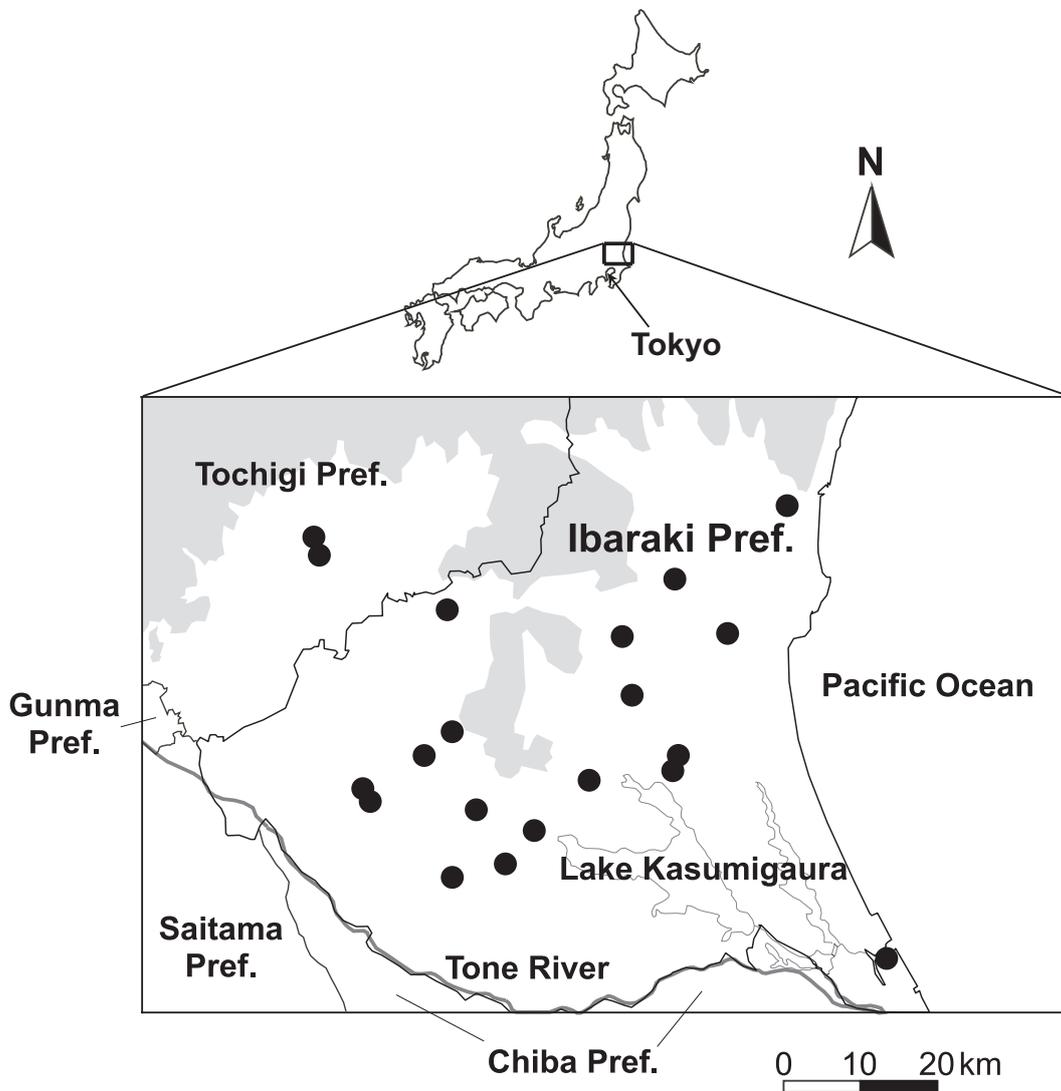


Figure 1 Locations of heron and egret colonies in 2011 in my study area. Each dot represents a colony location (20 colonies in total). Mean nearest neighbor distance between colonies was 9.97 km. Grey regions show an altitude greater than 100 m where the distribution of herons and egrets is much lower.

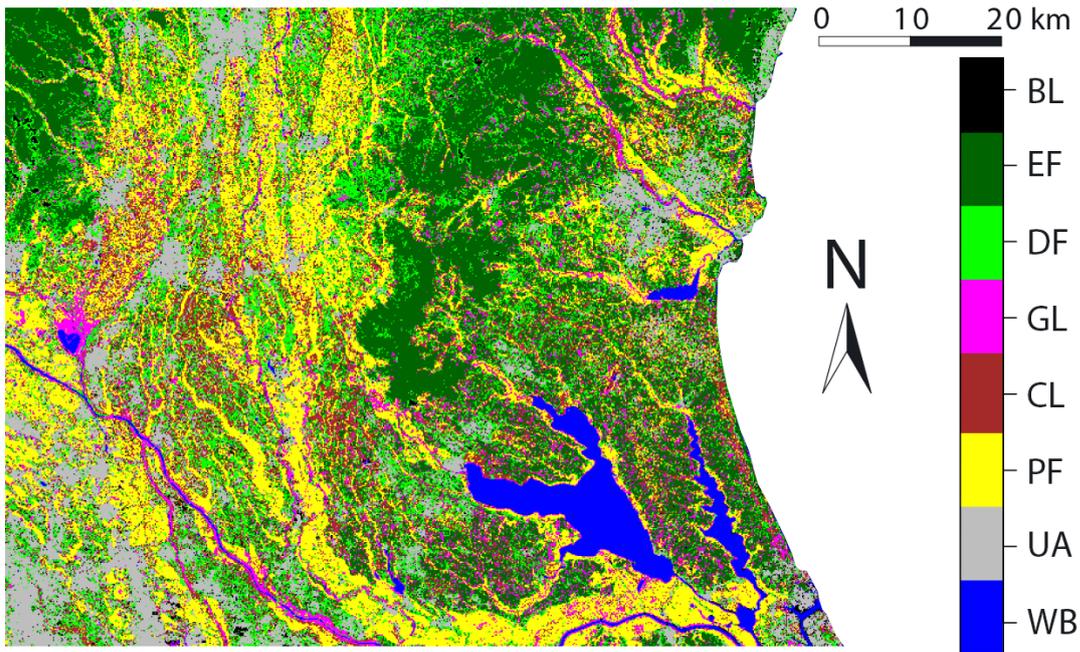


Figure 2 Land-use map of the study area. Map provided by JAXA combining ALOS satellite imagery and ground surveys from 2011. BL: bare land, EF: evergreen forest, DF: deciduous forest, GL: grassland, CL: crop land, PF: paddy field, UA: urban area, and WB body of water.

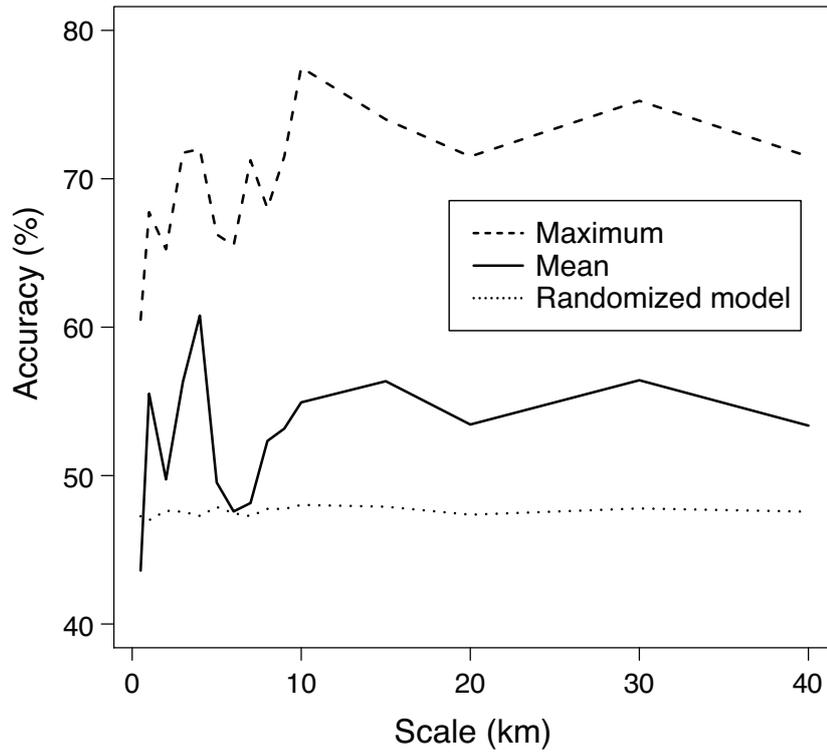


Figure 3 Model accuracies for each scale. The mean graph represents the average accuracy for all of the 30 random-points datasets. The maximum graph considers only the random set that provide the maximum accuracy for each scale. The randomized accuracy graph represents the average accuracy of the randomized models for all 30 random-points data sets.

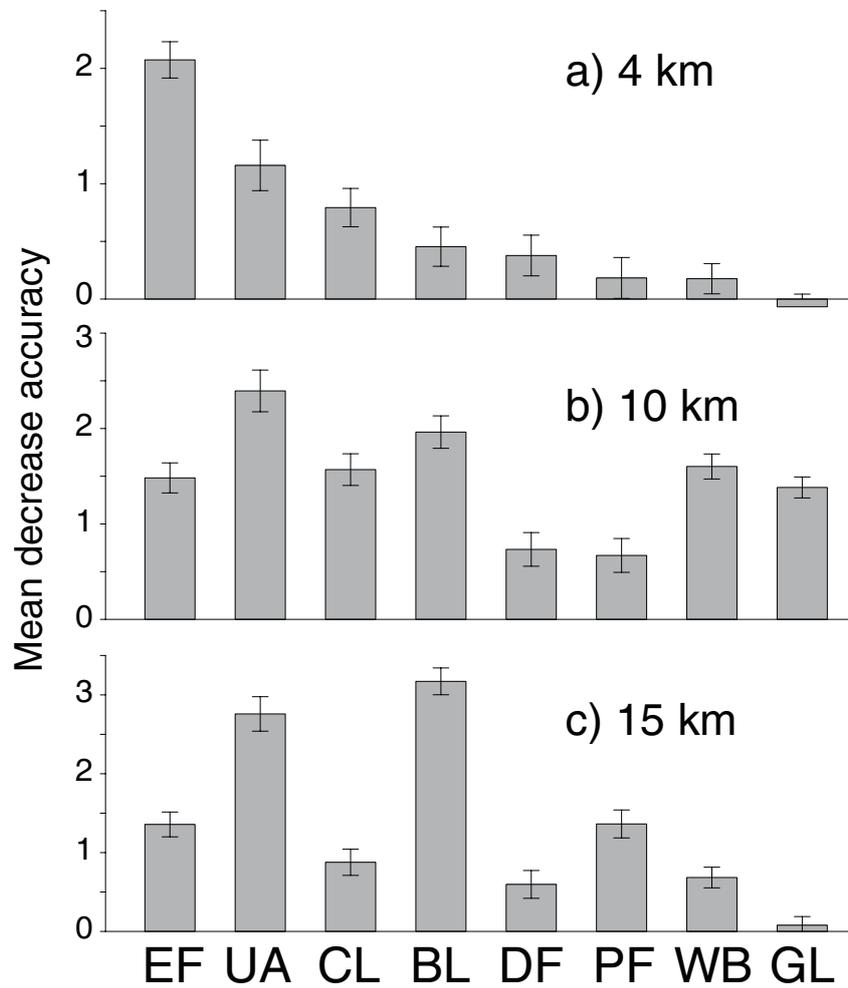


Figure 4 Variable importance of each land use for three scale models (4, 10 and 15 km) represented by the mean decrease accuracy index. Error bars represent a 95% confidence interval. EF: evergreen forest, UA: urban areas, CL: crop land, BL: bare land, DF: deciduous forest, PF: paddy field, WB: body of water, and GL grassland.

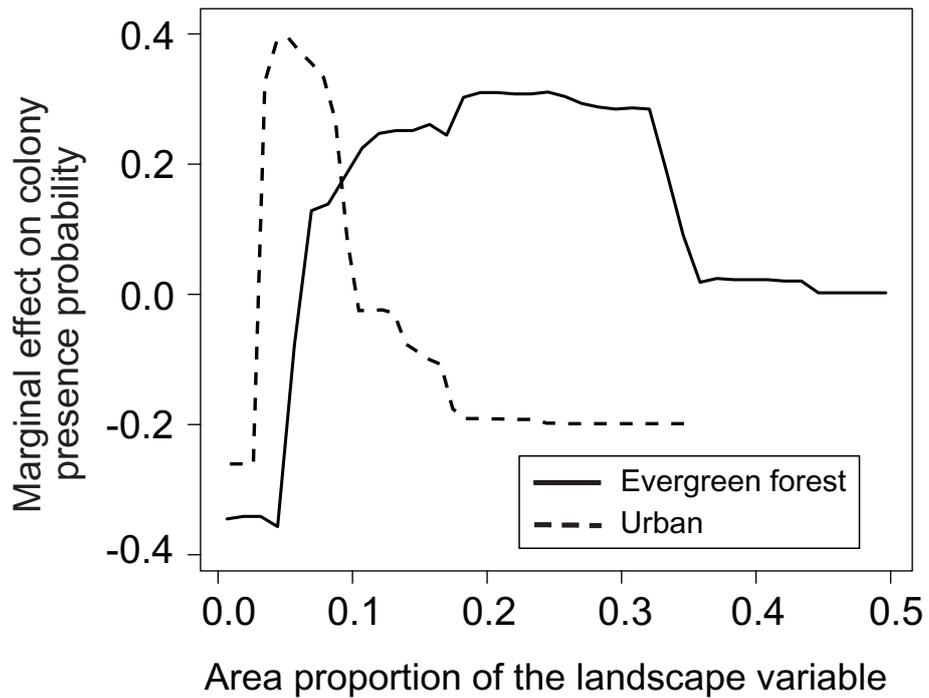


Figure 5 Partial dependence plot for evergreen-forest land use for the 4-km scale model and urban land use for the 10-km scale model. The horizontal axis represents the proportional area of the land use, of which the partial dependence was calculated within a circular buffer with a 10-km radius.

Appendices

Appendix 1

In order to study the range of each explanatory variable used in the samples used in this study models, we compared the percentages of the land use coverage of the whole study area with the average of the percentages of the land-use variables for the buffer areas surrounding the colonies for the most important scales (Table 1).

For a better understanding of the differences on the relative importance of each land-use variable, we studied the correlations between them among the most important scales (Table 2 and 3). We used the Pearson's correlation coefficient for creating a correlation matrix and comparing each explanatory variable relationship.

Table 1 Proportion of each land use (%) of the whole study area and of the average among all buffers surrounding the colonies for the two most important scales

Land use	Study area	4-km scale	10-km scale
Body of water	3.8	2.7	4.9
Urban areas	11.7	11.2	8.8
Paddy field	22.8	21.8	22.7
Crop land	13.2	17.2	16.0
Grassland	6.7	6.3	6.7
Decidious forest	17.1	20.5	17.9
Evergreen forest	22.7	18.0	20.5
Bare land	2.1	2.2	1.9

Table 2 Pearson's correlation coefficient between land-use variables among all buffers of 4-km scale

Land use	Urban areas	Paddy field	Crop land	Grassland	Decidious forest	Evergreen forest	Bare land
Body of water	-0.01	-0.46	-0.30	-0.18	-0.52	-0.06	0.42
Urban areas		0.07	-0.30	-0.43	0.01	-0.66*	0.71*
Paddy field			0.28	-0.11	0.06	-0.54	-0.15
Crop land				0.12	0.17	-0.19	-0.26
Grassland					-0.25	0.43	-0.38
Decidious forest						-0.09	-0.13
Evergreen forest							-0.58

*Significantly different from zero ($P \leq 0.05$)

Table 3 Pearson's correlation coefficient between land-use variables among all buffers of 10-km scale

Land use	Urban areas	Paddy field	Crop land	Grassland	Decidious forest	Evergreen forest	Bare land
Body of water	-0.18	-0.44	-0.38	-0.60	-0.77*	-0.06	0.13
Urban areas		0.02	-0.21	-0.47	-0.06	-0.61	0.81*
Paddy field			0.42	-0.10	0.43	-0.53	0.07
Crop land				0.26	0.54	-0.26	-0.09
Grassland					-0.02	0.32	-0.46
Decidious forest						-0.10	-0.03
Evergreen forest							-0.75

*Significantly different from zero ($P \leq 0.05$)