[redacted]

Kingfisher Birds and Water Quality in Danish Lakes

by

[redacted] and [redacted]

[redacted]

Contents

1 Introduction

The aim of this paper is to investigate the patterns in the geographical occurrence of the kingfisher in Denmark. Firstly, we investigate how the bird observations are clustered. Additionally, we investigate the relationship between the observations of the kingfisher and the water quality of the lakes near which the birds reside.

Our hypothesis is that the bird observations are clustered and that this clustering is related to water quality. Specifically, we hypothesize that *there is a negative correlation between bad water quality of lakes and the presence of the kingfisher in the nearby area*.

Finally, as a visual product of this project we create an interactive map in R (R Core Team 2020), displaying the relationship between the water quality of Danish lakes and the presence of the kingfisher between the years 2016 and 2020.

The product of this investigation, that is the R code as well as the `.html` rendering of the map reside in the following GitHub repository:

[redacted]

1.1 The Kingfisher in Denmark

The kingfisher (*Alcedo atthis*) is a primarily tropical bird which is a relatively rare sight in Denmark, both due to its few numbers and its general shy nature. The kingfisher's habitat is predominantly associated to clean, clear bodies of water with large populations of various fish that constitute the kingfisher's source of food (Skriver 2019). In ornithology the kingfisher is therefore often used as an indirect marker of the ecological state of lakes and streams (Skriver 2019).

In an effort to protect the kingfisher, the Danish Ornithological Association (DOA) has since 1986 worked politically to implement plans to improve the ecological quality of the water (Marczak 2019). Ornithologists believe that the improved maintenance of e.g. brooks and streams has lead to a more natural growth of waterweeds, which in turn leads to more biodiversity (Skriver 2019).

This effort has resulted in generally cleaner waters around Denmark, and thus also in a rise of the number of kingfisher breeding birds. Our project aims at contributing to the investigation of the relationship between the number of kingfisher birds and water quality.

2 Environmental and Cultural Relevance ([redacted])

The loss of species diversity in modern time is the so-called *Sixth Extinction*, which succeeds the widely recognized five big mass extinctions (Rampino and Shen 2019). Our project is primarily motivated by an interest in the impact of humans on the environment and thus on the loss of species. It is well known that human activities are contributing to the degradation of ecosystems world wide, and many have argued for the central role of human behaviour in reversing the decline in biodiversity (Nielsen et al. 2021: 550).

By examining the kingfisher and water quality our goal is to provide a tangible example of how human behaviour affects nature in terms of pollution. As stated, we hypothesize that bad water quality leads to fewer kingfishers.

Even if we do not find any evidence of a human impact on lake biodiversity and the population of kingfishers, we believe our project still has relevance for the public. The construction of an intuitive, easily accessible map can be helpful in terms of increasing awareness of both the development of the kingfisher population, as well as the development of Danish lake's water quality. Additionally, the map can be helpful in communicating where in Denmark chances are best of spotting the rare kingfisher.

3 Methods ([redacted])

In this section we will describe our data, explain the main steps we took during the pre-processing of the data, and present the methods we used in our analysis. Finally, we will introduce the processing steps taken to create the interactive map.

3.1 Data Acquisition ([redacted])

This investigation utilizes data from two main sources: data on the occurrence of the kingfisher in Denmark over time, and data on the ecological state of all lakes across Denmark.

3.1.1 Kingfisher Data

The data set on the kingfisher was extracted from the DOA data base. A `.csv` file containing all kingfisher observations was extracted for years 2016 and 2020. Each line in the `.csv` contains a reported kingfisher observation made by an ornithologist including the following information: the name of the observation spot; the geographical coordinates for the observation spot in decimal degrees; the number of birds observed; the behaviour of the birds, as well as, a number of additional features irrelevant to this analysis. In addition we extracted the bird observations from 2017, 2018 and 2019 for the interactive map.

3.1.2 Water Quality Data

The data set on the water quality of Danish lakes was provided by the Danish state's environmental protection agency's (EPA) own public database: https://mst.dk/service/miljoegis/hent-data/. This agency monitors the state of lakes by regularly analyzing water samples to measure the amount of heavy metals and pesticides from e.g. farming and factory waste (Miljøstyrelsen a 2021). The data were collected as part of their 2015-2021 water protection program. This program is made to ensure cleaner water in Danish seas, lakes, streams and ground water in accordance with the EU Water Framework Directive evidently to protect biodiversity (Miljøstyrelsen b 2021).

The provided data set is a shapefile (`.shp`) with multi-line polygon geometry. The polygons outline the boundaries of the lakes defined by geographic coordinates in decimal degrees. The data set also includes a combined measure of the ecological state of the lake and a number of other variables such as lake name, municipality number of the lake location and a series of ecological measures of the water quality. For this analysis, we have chosen to look at the combined ecological measure of the water quality, which summarises the information of the other measures.

Note that we only have data on water quality for 2016 and not for 2020, as these data have not been made public yet. Thus, the analysis of the relationship between the presence of birds and water quality is based on the water quality in 2016. In spite of this, we chose to include kingfisher observations ranging from 2016-2020, because there might be a delay in the effect of the improved water quality on the birds.

3.2 Pre-processing the Data ([redacted])

The following section outlines the principal pre-processing steps taken for both data sets.

3.2.1 Kingfisher Data

For this analysis we chose to only include the bird observations where their behaviour was either marked as 'nesting' or 'foraging'. These behaviours indicate that the observed birds were residing near the lake for a longer time period, thus making them good markers of the water quality (Skriver 2019). Bird observations where the behavior was of the type 'migrating' were excluded, as they might just be passing by the observation spot.

We also included the location name of the observation; the geographical coordinates for the place of observation; the observation year; observation date, and the number of observed birds. Finally, we made sure there was only one bird observation per row in the data set. Thus, we had 2118 observed kingfishers in 2016, and 4287 in 2020.

3.2.2 Water Quality Data

For the analysis the coordinates outlining the lakes and the condition of the water were subset from the original data set. Furthermore, the water quality rating was simplified from seven to three levels.

In the original data set the water quality of each lake was rated on a scale with seven categories (Table 1, 'Original meaning'). We chose to reduce these to three, by collapsing the two highest and the two lowest respectively. The main reasons for this were simplicity of analysis and ease of visualization. Since we color coded each water quality level, it would be visually challenging to distinguish between seven differently colored lake contours, as opposed to only three (Gimond 2021).

Abbreviation	Original meaning	Count	New Category Name
Ht	High ecological state	50	1
Gt	Good ecological state	114	1
Mt	Moderate ecological state	186	2
Rt	Poor ecological state	129	3
Dt	Bad ecological state	205	3
Mp	Moderate ecological potential	3	N/A
0	State unknown	170	N/A

Table 1: Original vs new water quality categories

Table 1 illustrates the original categories, and how they were collapsed. The last two categories marked as N/A were excluded either because they did not contain precise enough data ('Mp') or had no data ('0') on water quality. Thus, 173 lakes were excluded from the analysis. As displayed in Figure 1 we ended up with 164 lakes of category 1, 315 lakes of category 2 and 334 lakes of category 3.

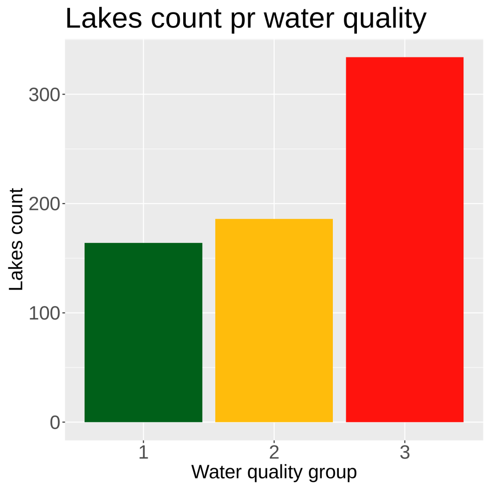

Figure 1: Total amount of lakes in each water quality category

3.3 Analysis ([redacted])

Firstly, we examined the clustering in the point pattern made up by the kingfisher observations using the `spatstat` package (Baddeley and Turner 2005). Using a freely available shapefile outlining the outer bounds of Denmark, an `owin` object was constructed to outline the area of the point pattern (Upadhyay 2021). The clustering was analyzed using the `Kest` function in R. We use border correction method 'border' as this is the most efficient when using a complex window shape in terms of computing speed (R Documentation). The Kest-function estimates Ripley's reduced second moment function `K(r)` from a point pattern in a window of arbitrary shape (R Documentation). Ripley's K equals the expected number of points within a distance r of a typical point in the point pattern.

For both years, we did a Monte Carlo test by computing 99 simulations of a theoretical point pattern with the window size and number of points in our `ppp` object and computed the function `K(r)` for these patterns. For both years these were then compared to the K function for the point pattern of the actual data.

Apart from the clustering analysis we investigated whether the water quality and the number of kingfisher observations were correlated. In order to conduct the correlation test we did a count of the bird observations near all lakes by computing buffers. A buffer defines an area around a given feature using a set distance from the center. Thus, we computed buffers around the centroids of all the lakes in the data set using the `sf` package (Pebesma 2018). In order to make the buffers around the lakes, we used a local projection for Denmark, namely `UTM32N`.

The computed buffers had a radius of 1000 m. This radius is based on the fact that kingfishers do not necessarily reside right next to lakes, but can also nest several hundred meters away (Dansk Ornitologisk Forening 2004-2021). Utilizing the buffers, the number of bird observations for all lakes for both 2016 and 2020 was calculated. The bird counts for both years is visualized in Figure 2. Finally, we did a Pearson's correlation test in R (R Core Team 2020) on the water quality and the bird count.

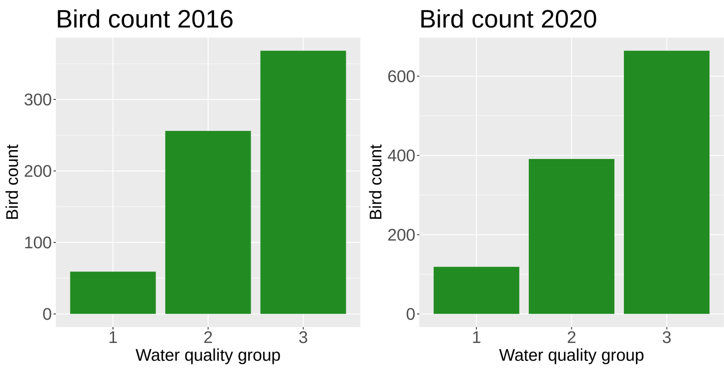

Figure 2: Bird count by water quality category

3.4 Making the Interactive Map ([redacted])

Visualizing spatial data in maps allows us to better understand spatial relationships (Bivand et al. 2013: 59). Following this principle, we chose to make the map using Leaflet, an open-source tool for interactive map-making in R (Agafonkin 2021). With its many styling features and layering options Leaflet allows the map-maker to make intuitive and illustrative interactive maps.

We chose to work with Esri-layering which provides the base layers for the map. The base layers are *map image layers* which allow the user to switch from e.g. satellite view to a more simplistic view of a geographical area which brings forward the bird observation layers. These layers are mutually exclusive i.e. it is not possible to show both satellite and simplistic view at the same time (Agafonkin 2021). As both the water and bird data sets used coordinates from Google maps, it was necessary to project the data files to EPSG 4226.

4 Results ([redacted])

4.1 Clustering analysis

Figure 3 shows a plot of the the K function for the 99 simulations together with the K function for the 2016 data.

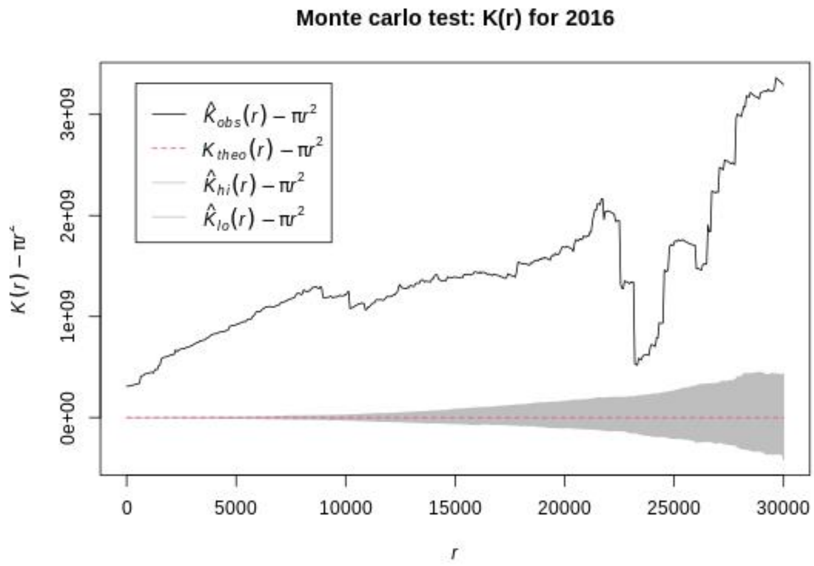

Figure 3: K(r) Monte Carlo test on the 2016 kingfisher observations.

Figure 4 shows the plot of the K function for the 99 simulations together with the K function for the 2020 data.

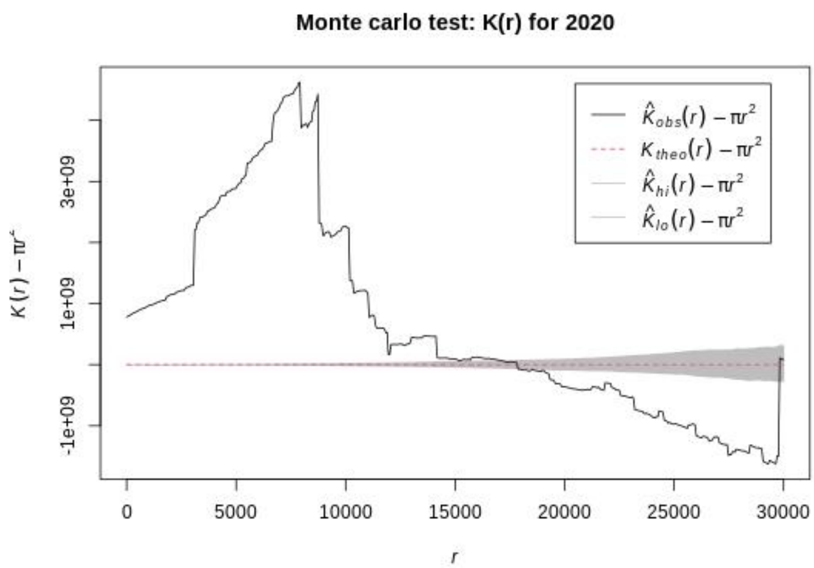

Figure 4: K(r) Monte Carlo test on the 2020 kingfisher observations.

4.2 Correlation test

Table 2 summarizes the results for both correlation tests.

Year	Correlation	p-value	Confidence Interval
2016	0.06	0.11	-0.01, 0.14
2020	0.06	0.09	-0.01, 0.14

Table 2: Correlation test results

4.3 The Map

The interactive map is a visual product of our analysis that visualizes the changes in kingfisher occurrences over time around the lakes.

The map consists of a total of nine layers: three base-layers and six overlay-layers. Our three base-layer options are: standard, simplistic and satellite. The overlay-layers allow the user to view several layers at the same time. One of these layers shows the contours of the lakes, color coded according to water quality (green for good, yellow for moderate, and red for bad). The remaining layers allow the user to pick a year between 2016-2020 to view where and when kingfisher birds were observed.

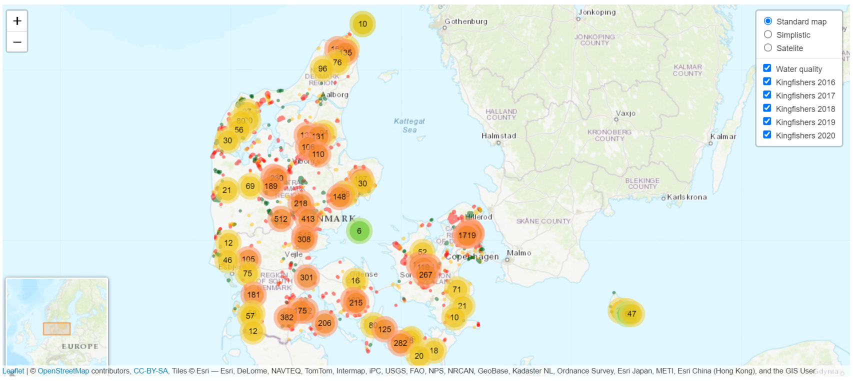

Figure 5: Interactive map

5 Discussion ([redacted])

As can be seen from Figures 3 and 4 the K function of the observed data from both 2016 and 2020 both lie above the K functions of the simulated point patterns for all distances up to 15.000 m. This indicates that the kingfisher observations are clustered within all radii up to this radius. The reason we did not consider a radius larger than 15 km is the small size of Denmark combined with the frequency of lakes. This makes it difficult not to hit an area around a different lake when looking at a specific radius. It would make the clustering irrelevant if we considered a radius of 30 km that encompassed e.g. two lakes 20 km apart.

The results displayed in Figures 3 and 4 could thus indicate that the kingfishers forage and breed in the same areas as the observations cluster together even when using relatively large radii. However, it is important to note that the bird observation data was reported by private individuals going into the field, so the location of the observations are influenced by what areas the ornithologists typically visit. This could to some extent explain why the point pattern in the kingfisher observations appears to be clustered.

The data is also influenced by how many trips the observers go on. As described earlier, there were 2118 observations in 2016, and 4287 in 2020, which could be due to people having generally more time to do the observations in 2020 during the lockdown. In an improved analysis one would ensure to include an equal amount of 'visits' of the observers to all the lakes; both visits where kingfishers were observed and where none were observed.

In addition to the clustering analysis we tested the correlation between the occurrence of kingfishers in a 1 km radius around the lakes, and the water quality of the lakes. We did this both for 2016 and 2020 to examine whether the clustering of the kingfishers could be related to water quality. For both 2016 and 2020 we see a positive correlation (Table 2), which indicates that there are fewer kingfishers around lakes with good water quality (e.g. water quality = 1), and more around those with bad water quality (e.g. water quality = 3). The correlation is however not significant as the confidence intervals of the correlation test are crossing 0 for both years. This shows that there might not be a relationship between the two variables at all. Thus, our results do not support our hypothesis that water quality has an impact on the number of observed kingfishers neither in 2016 nor in 2020. I conclusion we find evidence of clustering but no evidence that the clustering is related to the water quality of the lakes.

It is likely that other factors aside from lake water quality also impact the kingfisher. Firstly, the presence of noise from e.g. highways, cities, camping grounds etc., could be a hidden factor influencing the presence of the kingfisher. We speculate that e.g. a highway positioned right next to a lake could be disturbing for both nesting and foraging kingfishers. Given the kingfisher's shy nature it may also find the frequent presence of people disturbing, wherefore they might prefer lakes further away from cities.

Secondly, local weather conditions may also influence the population size of kingfishers in some areas of Denmark. For example, it is known that harsh winters can be deadly for kingfishers and therefore lead to a reduction in their population (Syndergaard 2021:12; Skriver 2019).

Lastly, it can be argued that the division of water quality into three categories is not the best way of modelling this variable. A simpler analysis design would be to divide water quality into the categories 'good' and 'bad'.

6 Conclusions

The results of the analysis indicate that the kingfisher observations are clustered within radius of 15 km. Nevertheless, we found no correlation between water quality and the number of kingfisher observations, indicating that this clustering cannot be explained only by the water quality of lakes. Instead, we have argued that other factors such as noise from roads and camping grounds as well as local weather conditions influence where the kingfisher forages and breeds in Denmark. The map is a visual confirmation of this result, meaning that we can see the birds clustering around lakes of all water quality levels in the period 2016-2020.

References

[1] Agafonkin, Vladimir. 2010-2021. Leaflet: an open-source JavaScript library for mobile-friendly interactive maps. https://leafletjs.com/index.html. Accessed on 1/6/2021.

[2] Baddeley Adrian, Rolf Turner. 2005. *spatstat: An R Package for Analyzing Spatial Point Patterns. Journal of Statistical Software*, 12(6). 1–42.

[3] Bivand, Roger S., Edzer Pebesma & Virgilio Gómez-Rubio. 2013. *Applied Spatial Data Analysis with R* 2nd Edition. Springer, New York, NY. https://doi-org.ez.statsbiblioteket.dk:12048/10.1007/978-1-4614-7618-4.

[4] Dansk Ornitologisk Forening. 2004-2021. Isfugl (alcedo atthis). *Fakta om fugle*. Accessed on 27/4/2021. https://dofbasen.dk/ART/art.php?art=08310&sorter=&arter=V%E6lg+art

[5] Marczak, Peter. 25/08/2019. *Isfugl - Danmarks flotteste fisker* . YouTube video published by BirdLife Danmark. https://www.youtube.com/watch?v=51tM8a3dV1U&ab_channel=BirdLifeDanmark. Accessed on 23/4/2021.

[6] Miljøstyrelsen a. 2021. *Søer*. https://mst.dk/natur-vand/overvaagning-af-vand-og-natur/soeer/. Accessed on 02/06/2021.

[7] Miljøstyrelsen b. 2021. *Vandområdeplaner*. https://mst.dk/natur-vand/vandmiljoe/vandomraadeplaner/vandomraadeplaner-2015-2021/vandomraadeplaner-2015-2021/. Accessed on 02/07/2021.

[8] Nielsen, Kristian Steen, Theresa Marteau, Jan M. Bauer, Richard B. Bradbury, Steven Broad, Gayle Burgess, Mark Burgman, Hilary Byerly, Susan Clayton, Dulce Espelosin, Paul J. Ferraro, Brendan Fischer, Emma Garnett, Julia PG. Jones, Mark Otieno, Stephen Polasky, Taylor H Ricketts, Rosie Trevelyan, Sander van der Linden, Diogo Veríssimo & Andrew Balmford. 2021. Biodiversity conservation as a promising frontier for behavioural science. *Nature human behaviour*. https://doi.org/10.1038/s41562-021-01109-5

[9] Pebesma, Edzer. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1). 439–446. doi: 10.32614/RJ-2018-009,https://doi.org/10.32614/RJ-2018-009.

[10] R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

[11] Rampino, Michael R. & Shu-Zhong Shen. 2019. The end-Guadalupian (259.8 Ma) biodiversity crisis: the sixth major mass extinction?. *Historical Biology*, 33(5). 716-722. https://doi.org/10.1080/08912963.2019.1658096

[12] Skriver, Jan. 18/12/2019. Gode tider for Danmarks smukkeste fugl: Den flyvende juvel, isfuglen, er i fremgang. *Dansk Ornitologisk Forening: Nyheder*. Accessed on 23/4/2021. https://www.dof.dk/om-dof/nyheder?nyhed_id=1784

[13] Syndergaard, Mads. 2021. Isfugle - der er flere end vi tror, in *Fugle og Natur*. Dansk Ornitologisk Forening. BirdLife.

[14] Upadhyay, Akshay. 2021. Download Shapefile of Denmark – Country boundary, Polygon. *GisMap*. https://www.igismap.com/download-shapefile-of-denmark-country-boundary-polygon/. Accessed on 31/05/2021.

Acknowledgements

We would like to say special thanks to [redacted], the administrator of DOA as well as an avid bird enthusiast, for introducing us to the DOA data base and providing the idea of investigating the occurrence of the kingfisher.

Metadata

Table 3 contains the variable names and their corresponding explanations for the used water quality data.

Variable Name	Variable Explanation
VRD_TIL	Overall ecological condition of the lake water. For example: amount of chlorophyll, fish, phytoplankton etc.
NAVN	Name of the lake, e.g.: Råbjerg Mile Sø
wtr_qlt	Collapsed water quality measures from 1-3. 1 is 'good', 2 is 'moderate' and 3 is 'bad'.
geometry	Geographic coordinates for lake location.

Table 3: Water quality variable explanations

Table 4 contains variable names and explanations for the used kingfisher data.

Variable Name	Variable Explanation
Year	Year when the bird was observed.
Antal	Amount of birds observed.
Adfbeskrivelse	Behavioral description. E.g.: nesting, fouraging etc.
Loknavn	Name of observation location.
lok_laengdegrad	Longitude in decimal degrees of observation location.
lok_breddegrad	Latitude in decimal degrees of observation location.

Table 4: Kingfisher observations variable descriptions