Urban nesting Herring Gull *Larus argentatus* and Lesser Black-backed Gull *Larus fuscus* population estimates: devising species-specific correction models for ground-based survey data.

Daisy Burnell

February 2021

This report should be cited as: Burnell, D (2021). Urban nesting Herring Gull *Larus argentatus* and Lesser Black-backed Gull *Larus fuscus* population estimates: devising species-specific correction models for ground-based survey data. Natural England publication ref: JNCC21_01

Author affiliation: Daisy Burnell works for the Joint Nature Conservation Committee (JNCC), as a seabird ecologist and project coordinator of the fourth breeding seabird census for Britain and Ireland, Seabirds Count.

Acknowledgments:
Thanks to BTO and Defra for the work carried out in 2018 and 2019 that this report builds upon. Thank also to the 'urban gull sub-group', with representatives from Natural Resources Wales (NRW), Natural England (NE), NatureScot, Department of Agriculture, Environment and Rural Affairs (DAERA), Royal Society for the Protection of Birds (RSPB), British Trust for Ornithology (BTO) and BirdWatch Ireland, for reviewing this work and contributing to the design and implementation of the new survey methods and sampling strategy.

A big thank you to all of the surveyors who volunteered their time and effort out in the field to help produce the data that has been instrumental in the production of population estimates presented in this report. And thanks also to the work conducted by BL-Ecology Ltd and their contracted surveyors, and that by Hi-Def Aerial Surveying Ltd to progress the accelerated urban nesting gull census in 2020, made possible through generous funding and support from Defra and Natural England.

A special thank you to Ivan Lakin (NE) for edits and comments prior to publication.

EQA:
This report is compliant with JNCC's Evidence Quality Assurance Policy
https://jncc.gov.uk/about-jncc/corporate-information/evidence-quality-assurance/

This has also been through Natural England's internal review process.

Executive Summary

This report was produced to inform *Seabirds Count* (2015-present) the fourth Britain and Ireland breeding seabird census. The previous census, *Seabird 2000* (1998-2002) was the first to produce separate inland, coastal, natural-nesting and roof-nesting population estimates of Herring Gull (HG) and Lesser Black-backed Gull (LB); *Seabirds Count* will be the second census to achieve this.

The Joint Nature Conservation Committee (JNCC) was commissioned by Natural England (NE) to produce two reports – one that presents gull breeding population estimates and the extrapolation analysis used to derive these estimates for urban nesting gulls in association with correction factor models[1]; and this report, which describes methods of obtaining the correction models based on comparative analysis of digital aerial survey (DAS) and ground based surveys (GBS) of HG and LB to account for detectability issues posed by the urban environment.

The urban environment is broadly recognised as increasingly important for supporting breeding populations of HG and LB in Britain and Ireland. Whilst simple and standardised monitoring methods (Walsh et al., 1995) are used to monitor these gull species nesting in more natural settings, there is yet to be a consistently accepted method to surveying these species in the urban environment, nor has there been found a consistent way to analyse survey data from which to accurately estimate urban, and therefore also, overall populations. Challenges to this are varying levels of detectability from GBS and then ably correcting for those detectability limitations.

Published national population estimates of HG and LB from the *Seabird 2000* census (Mitchell et al., 2004) are thought to be unreliable due to under-estimates of roof-nesting gulls that resulted from gaps in coverage and from detectability issues using the vantage point methods (identified by Coulson & Coulson, 2015).

An ‘urban gull sub-group’ was set up refine survey methodology and to outline comparative studies to explore the detectability issue. Defra funded and commissioned the BTO to undertake pilot surveys in 2018 and 2019 and associated analyses. Relationships between GBS, vantage-point surveys and DAS were investigated and the best fitting models were presented (Woodward et al., 2020).

In 2020, NE commissioned JNCC to build on previous work with the benefit of additional comparable GBS and DAS data from northeast England in 2020. These data were pooled with those from 2018 and 2019, increasing the analysis sample size from 162 to 235 1 km survey squares. Analyses focused on testing for consistent relationships between DAS and GBS data to identify correction factors for use in models to estimate population sizes.

The assumption is that the DAS apparently occupied nest (AON) total for each species, in each 1 km square, was true to reality. Testing of various models found conversion models with the best fit for GBS data were based on individual adults (IND) and two types of environmental predictors within each survey square – the predominant type of urban fabric (strata), and the proportion of each 1 km square covered by urban fabric (PROP). These environmental predictors were assumed to influence nesting numbers and also their GBS detectability.

---

[1] Burnell, D. 2021. Population estimates for urban and natural nesting Herring Gull *Larus argentatus* and Lesser Black-backed Gull *Larus fuscus* in England. Natural England publication ref: JNCC21_02

Although model fit was good for both species, these still produced large confidence limits (CL) around the estimated total AONs for each of the study areas (Birmingham, North Wales, Cleveland and Humberside) and combined. Recommendations and conclusions based on extrapolated estimates of HG and LB urban-nesting population should, therefore, be caveated and direct comparisons with previously published estimates should be avoided, due to differences in methodology and reliability.

To improve precision of future population estimates, this report recommends –

- recording and analysing more precise information about urban fabric, such as building height and roof structure;
- increasing sample sizes of poorly represented urban fabric types; and
- increasing sample sizes of combined GBS and DAS squares, particularly those that support very large numbers of nesting gulls.
- more comprehensive coverage of the urban environment within squares should also be sought, as the adjustment of DAS and GBS counts to 100% coverage could be an added source of error.

Contents

1 Introduction

Annual monitoring of breeding seabirds at a sample of colonies (by SMP) and periodic national censuses enable population trends and estimates, at both the temporal and spatial scale, to be produced. For HG and LB, robust population trends have been limited to the 'natural-nesting' cohort due to insufficient monitoring of urban-nesting colonies, and some previous population estimates are unreliable due to methodological limitation.

The lack of reliable, comprehensive, and up-to-date information about HG and LB breeding populations has presented an issue for policy makers and determining conservation status.

In the UK for *Seabird 2000* (1998-2002), approximately 16% of the estimated 127,349 HG AON and approximately 9% of the estimated 116,340 LB AON, were roof-nesting. These HG and LB population estimates differ from those in Mitchell *et al*. (2004) due to subsequent discovery and the correction of errors found in the original dataset. This was a separate issue to that identified by Coulson and Coulson (2015) whom demonstrated that traditional methods of using vantage point and GBS counts failed to account for variable detection rates within urban strata, which can lead to imprecise population estimates.

Given the conservation status of these two species (Eaton et al., 2015); the importance of the UK to their international populations (e.g. Mitchell et al., 2004); apparent divergence in the fortunes of roof-nesting and 'natural-nesting' populations (e.g. JNCC, 2020 online, Balmer et al., 2013 and Raven and Coulson, 1997); together with recently dynamic regulatory policy relating to these species (NE, *pers comm*., 2020), up to date and accurate breeding population estimates are required.

DAS offer a practical and accurate alterative to vantage point surveys to resolve detectability issues relating to elevation (Ross et al., 2016; Thaxter et al., 2017). However, these are expensive, and prohibitively so if estimates are to be produced regularly to understand temporal change (Woodward et al., 2020), or if considered for use to survey urban nesting gulls at national scales.

Unmanned aerial vehicle (UAV) surveys have shown promising results in Victoria, Canada, for urban nesting large gulls (Blight *et al*., 2019) and are less expensive than their DAS counterparts (Christie *et al*., 2016). However, restrictions on the use of UAVs in built-up areas in the UK make these surveys, at least for the near future, renders this option unattainable (Ross et al., 2016).

Vantage point surveys from Mobile Elevating Work Platforms (MEWPS, e.g. 'cherry pickers', can offer more accurate counts than GBS, but still underestimate actual AONs (Coulson and Coulson, 2015). Logistically, the use of such equipment for surveys to produce robust population estimates at county or national scales would be complex and expensive (Ross et al., 2016).

In 2015, at the start of *Seabirds Count*, the fourth breeding seabird census in Britain and Ireland, an 'urban gull sub-group' was convened; see 'acknowledgments' for representation. It was tasked with finding suitable ways to accurately estimate of urban-nesting gull populations. This group suggested that 'ground-truthed' DAS, from which conversion methods could be produced, was a potential solution. The analysis and modelling to generate accurate conversion methods would need to be representative, consistent and repeatable to ensure that subsequent census results could be comparable. However, the analysis and modelling required are particularly complex, and a specialist is required to do this. In 2018, Defra commissioned research into this potential solution.

DAS and GBS of sampled 1 km squares were conducted in Birmingham in 2018 and in north Wales in 2019. Analysis of these data found relationships between DAS and GBS, but models produced for HG were over-dispersed and additionally for both species, the confidence limits around the predicted AON were too wide to be recommended for use in producing population estimates (Woodward et al., 2020). A larger sample of comparable DAS and GBS squares was recommended to improve model fitness and to increase the confidence in population estimate predictions.

The aim of this report is to build on the work of Woodward *et al.* (2020) by combining the 2018 and 2019 data it used with an additional 2020 dataset to increase sample size and representation of urban strata. The objectives were to:

- refine relationships between GBS and DAS counts;
- explore if additional variables improve the fit of the models; and,
- determine if models can devise accurate conversion factors to produce robust population estimates of urban nesting HG and LB.

2 Methods

Urban environments are heterogenous in their structure and thus pose varying levels of nesting gull detectability from GBS. To understand the nature of these variations, and to ensure that survey design adequately accounted for these, a number of steps were taken in the analysis methodology. Each 1 km square in the UK was assigned one of five urban fabric strata, based on the CORINE dataset (CLC, 2012), where the survey square was predominantly either:

- **Rural**: less than 2% coverage of urban fabric;
- **Sub**: suburban, discontinuous, essentially residential urban fabric;
- **Ind:** industrial, commercial or transport urban fabric types;
- **Ind/Sub mix**: industrial/suburban mix, a relatively even mix of 'Sub' and 'Ind'; or
- **Most Urb**: most urban, high density and continuous urban fabric comprising a mix of urban fabric types.

In addition to these strata, each square had an estimated percentage cover of urban fabric (PROP) within the square, this was also derived from the CORINE (CLC, 2012). All squares assigned '**Rural'** were removed from survey planning and analysis. This habitat is instead surveyed using the standard methods for natural nesting HG and LB (Walsh et al., 1995). A detailed explanation of the process to obtain these strata from CORINE is presented in Woodward *et al*. (2020).

DAS and GBS were conducted at a sample of urban 1 km squares in England and north Wales. Although the objective of this study is to produce population estimates in England only, the use of sample squares located in Wales was only for the purposes of producing correction methods. It is assumed that relationships between urban fabric strata and nesting HG and LB are the same in England and Wales.

2.1 Digital aerial survey (DAS) data collection and validation

DAS data were collected in the breeding seasons of 2018, 2019 and 2020 by HiDef Aerial Surveying Limited. In 2018, 100 1 km urban squares were surveyed across Birmingham and in 2019 another 100 were surveyed in north Wales. In 2020, another 99 1 km urban squares were surveyed in Hull, East Yorkshire and Hartlepool, Co. Durham (nee Cleveland), in northeast England. The 2020 squares were selected from the overall stratified random sample of squares for the overall GBS in England survey, which randomly selected 15% of squares within each urban stratum within each county. The squares selected for the 2020 DAS offer coverage of all urban stratum and the additional survey area ensured representation of another part of the country. Table 1 below presents sample square totals selected in each DAS area per urban stratum.

Table 1. DAS square totals for each survey areas per stratum.

Survey	year	Ind/Sub mix	Most Urb	Sub	Ind	TOTALS
Birmingham	2018	23	18	48	11	100
North Wales	2019	17	10	49	24	100
northeast England	2020	5	8	57	29	99
TOTAL	**all**	**45**	**36**	**154**	**64**	**299**

All image analysis and validation were carried out manually by experts and used consistent methods (Woodward et al., 2020) across the three seasons of data. Each AON or IND identified by DAS was geo-referenced, accurately mapped using the software QGIS

(QGIS.org, 2020) and assigned to a 1 km square. For each square, AON and IND totals were calculated.

Not intentional by methodological design, it transpired that differences existed in the equipment used between the 2018-19 and 2020 DAS. During the 2018-19 surveys the camera captured images along 500 metre wide transects and therefore any 1 km square could potentially be fully covered by two transects (Weiß et al., 2016). However, in 2020 the bank of four cameras on board the aircraft were a generation before the 2018-19 set up and had reduced capability. Each individual camera captured images along 125 m wide transects and together all four covered a transect 500 metres wide. However, the camera arrangement created three 10 metre gaps between them. Therefore, two transects of one 1 km square with six 10-metre wide gaps would mean an un-surveyed area of 60 metres by 1000 metres, or 94% coverage, on average.

2020 aerial transects were overlaid onto maps of their 1 km survey squares to calculate un-surveyed areas. Due to high dependency on the flight path and the timing of cameras being switched on and off, 2018 and 2019 DAS transect data was also checked through the same process. The % range coverage totals of squares for each survey area and stratum by DAS is shown in Table 2; this only details coverage of squares where GBS also occurred, hence the overall total is 257, not 299. The total count for each species and unit, for every square surveyed, were then adjusted up to total urban area within each 1 km square, based on the new estimate of coverage.

Table 2. Numbers of squares with DAS coverage ranges, calculated through the mapping of each camera transect across the target survey squares that had both DAS and GBS coverage. This is broken down by stratum and survey area.

	<25%	**≥25% and <50%**	**≥50% and <75%**	**≥75%**	**TOTALS**
Birmingham (2018)					
Ind/Sub mix	0	0	0	17	**17**
Most Urb	0	0	0	15	**15**
Sub	0	0	0	37	**37**
Ind	0	0	0	11	**11**
north Wales (2019)					
Ind/Sub mix	0	1	2	11	**14**
Most Urb	0	1	0	8	**9**
Sub	0	2	0	39	**41**
Ind	0	1	5	16	**22**
northeast England (2020)					
Ind/Sub mix	0	2	1	3	**6**
Most Urb	0	1	4	3	**8**
Sub	1	4	28	23	**56**
Ind	0	2	11	8	**21**
TOTALS:	**1**	**14**	**51**	**191**	**257**

2.2 Ground-based survey (GBS) data collection and validation

GBS were carried out by both contracted and volunteer surveyors in most of the same squares surveyed by DAS and as close to the DAS survey dates as possible. 167 squares were surveyed by both GBS and DAS during the 2018 and 2019 seasons and 162 squares of these were taken forward for analysis by Woodward et al (2020); the five excluded were due to <50% accessible habitat. Of the 99 squares selected during the 2020 season, 90 were comparable squares; the nine excluded were also due to access restrictions for GBS.

The overall total of comparable squares available for analysis from all three survey seasons (2018-2020) was 257 squares.

The GBS methodology was consistent throughout all three seasons (Appendix 1). Timing of surveys aimed to coincide with the peak incubation period and anecdotal evidence suggested this was clinal across the UK so, where possible, local knowledge was sought to synchronise timings appropriately.

In summary, GBS surveyors recorded three types of gull counts, these were apparently occupied nests (AON), apparently occupied territories (AOT) and total adults on breeding habitat (IND); totals were not mutually exclusive. Totals were collated for each 1 km square and for each species. Additional information recorded included survey times and date; weather conditions; types of any gull deterrents present; and the percentage of the urban area of each square that could not be surveyed. Note this is not the percentage of roof space that the surveyor could not see, it instead refers to general areas of urban fabric within each survey square that could not be surveyed. Therefore, 'detectability' from GBS is appropriately examined by data concerning observation limitations within those portions of survey squares that were actually surveyed.

The first two seasons of data collection, collation and validation were coordinated by the BTO under contract to Defra. In 2020, data collection was coordinated by BL Ecology Ltd, under Defra contract. Surveyors in 2020 submitted records daily online on forms held on JNCC servers and daily downloads of the form were then sent to BL Ecology Ltd for validation. This daily data submission allowed for errors or potential methodology misunderstandings to be intercepted and quickly resolved.

All three breeding seasons of data were then collated and checked for any outstanding errors. The % range coverage totals for each survey area and stratum by GBS is shown in Table 3. This only details coverage of squares where DAS also occurred, hence the overall total is 257, not 299. Unlike Table 2 and DAS data, % coverage relates to the area of urban fabric with each 1 km survey square, not the entire square.

Table 3. Numbers of squares with GBS coverage and with estimated percentage ranges of urban area not surveyed. This is broken down by stratum and survey area.

	<25%	**≥25% and <50%**	**≥50% and <75%**	**≥75%**	**TOTALS**
Birmingham					
Ind/Sub mix	0	1	0	16	**17**
Most Urb	0	0	1	14	**15**
Sub	0	0	0	37	**37**
Ind	0	0	1	10	**11**
north Wales					
Ind/Sub mix	0	0	2	13	**15**
Most Urb	0	0	0	9	**9**
Sub	0	0	1	40	**41**
Ind	4	0	0	18	**22**
northeast England					
Ind/Sub mix	0	1	1	3	**5**
Most Urb	0	0	0	8	**8**
Sub	0	0	2	54	**56**
Ind	0	2	3	16	**21**
TOTALS:	**4**	**4**	**11**	**238**	**257**

2.3 Analysis

Initial exploratory analysis concluded that squares with less coverage from either or both DAS and GBS added variability and potential error in the models. To address this and avoid reducing sample sizes too much, the dataset was filtered to exclude squares with <50% coverage from either survey type. This filtering reduced the final dataset from 257 to 235 squares. Table 4 presents a breakdown of the final sets of sample squares used for the analysis with a further breakdown to indicate square totals that are inland or coastal, explained below.

Table 4. survey squares with >50% coverage by DAS and GBS in each urban stratum per survey area and broken down as either inland or coastal.

Survey	Ind/Sub mix		Most Urb		Sub		Ind		Total
	Inland	Coastal	Inland	Coastal	Inland	Coastal	Inland	Coastal	
Birmingham	16	0	15	0	37	0	11	0	**79**
north Wales	7	6	1	7	7	32	9	9	**78**
northeast England	0	3	1	6	13	38	3	14	**78**
Total	**23**	**9**	**17**	**13**	**57**	**70**	**23**	**23**	**235**

A number of statistical, graphical, modelling and verification tools that are all compatible and relate to 'R' were employed at different stages of the analysis. R is a computing language and software used for computing statistical analyses and producing graphics. The statistical software RStudio (R Core Team, 2018) was used to explore and analyse the data. The package ggplot2 (Wickham et al., 2020) was used to graphically explore and present the data collected, examples are shown as Figure 1 and 2 below. The packages MASS (Ripley et al., 2020) and DHARMa (Hartig and Lohse, 2020) were used to produce and validate the species-specific models, respectively.

Models were produced using the MASS package (Ripley et al., 2020) under the negative binomial (NB) and Poisson (pois) distributions. These were run with increasing complexity using three potential explanatory variables:

1. The natural log (+1) of the adjusted GBS counts of IND or AON. The rationale being that if birds were confirmed breeding in the square by DAS, the GBS should pick up at least one AON and/ or some IND.
2. The urban strata, these were run both as separate variables (no interaction) and as variables that could influence the relationship between the DAS and GBS counts (interaction). Strata capture broadly different roof types which likely vary in their nesting potential, therefore having influence on the number of AONs detected by DAS. As interactive variables, strata were theorised to also be important for detectability, e.g. if city centres ("Most Urb") have a greater proportion of taller buildings then GBS detectability of AON and IND may be relatively lower than perhaps compared to that of gulls nesting on lower elevation suburban ("Sub") housing.
3. The proportion of each 1 km square covered by urban fabric (PROP) was also run as both a separate variable (no interaction) as well one that could influence the relation between the DAS and GBS counts (interaction). The PROP would likely affect the number of AON detected by DAS. Simplistically, the higher the PROP the more suitable urban nesting habitat there is assumed to exist. Interaction was also explored because whilst nesting potential could increase, detectability by GBS could reduce as a related function of urban sprawl and density.

Validation steps were taken to ensure the most appropriate model distribution and explanatory variables for each species, were selected. To verify that the most appropriate variables were selected, comparisons of Akaike information criterion (AIC) and root-mean-square error (RMSE) values were used. Model distribution and assumptions were validated by comparing the outputs from the DHARMa package (Hartig and Lohse, 2020) on the residual: dispersion, zero-inflation and uniformity around the assumed distribution. Plots of residuals against the explanatory variables were also checked as part of the validation process. Additionally, log-likelihood tests using the lmtest package (Hothorn et al., 2020) were performed between the NB and pois models to ensure the most appropriate distribution for the data was used.

Once the most appropriate model was identified for each species, the GBS data were passed through the predict function (R Core Team, 2018) to produce estimated mean AON for each square and corresponding 95% confidence limits (CLs). These were used to compare the 'true' (adjusted DAS) AON count for the surveyed squares against the model predicted AON. This would give an indication of modelled validity and error.

A bootstrapped extrapolation up to the total number of squares in each survey area, and for all areas, was then conducted to produce 'true' and modelled population estimates beyond just the surveyed squares. For this exercise, squares were also categorised as either inland or coastal, defined as located less than or greater than 5 km from mean high water mark, and calculated by mapping the distance between the mean high water mark (OS OpenData, 2016) and the centre of the square. This is consistent with Woodward et al., (2020). The adjusted DAS AON were randomly sampled, with replacement, up to the cumulative number of squares in each stratum and summed, for both the whole survey area and the individual survey areas. This was repeated 999 times and the totals sorted in ascending order, the median 499^{th} was selected for the estimated AON and 2.5^{th} and 97.5^{th} percentiles (25^{th} and 975^{th}) used as the 95% CLs.

The simulate function (R Core Team, 2018) was used to predict the AON of squares from the selected model, this was run for each bootstrap allowing for the error from the model to be accounted for. These predicted AONs were then bootstrapped using the same process as performed for the DAS AON.

3 Results

3.1 Exploration of potential explanatory variables

Analysis assumed that DAS AON counts, including where extrapolated in squares with 50 - 100% DAS coverage, were the true count of AONs within each of those surveyed squares. Under this assumption, these data were explored to reveal relationships between GBS totals and urban strata.

Graphical exploration suggested DAS AON counts and GBS IND counts produce better fits (Figure 1 & 2) than GBS AON counts (Appendix 2). AIC and RMSE values for GBS AON models were significantly higher than those produced using GBS IND (Appendix 3). Log-likelihood tests suggested the negative binomial distribution was more appropriate than Poisson (Appendix 4). Previous analyses produced by Woodward *et al.* (2020), also support this finding.

Although a better fit, the amount of variation, and possibly the relationship in detectability, appears to differ between urban fabric types (strata) (Figure 1 & 2). For HG, the suburban (Sub) strata GBS IND counts generally seem generally higher than the equivalent DAS AON, but are a reasonable fit in other strata. For LB, relationships between DAS AON counts and GBS IND counts for each stratum were weaker than those shown in HG data (Figure 2), but these did not show obvious bias and although somewhat weaker, these relationships are better than those between DAS AON and GBS AON data (Appendix 2).

The other explanatory variable explored was the percentage cover of urban fabric within each 1 km square (PROP). This variable appears to have very broad representation in three of the four strata (Appendix 5) and differed in range from the other (Most Urb), which was considered relevant since differences in PROP could:

a) Influence relative detectability by GBS of nesting habitat;
b) Influence gull nesting density but have non-linear comparative GBS detectability.

Considering these potential effects, totals and differences between total numbers of DAS AON and GBS IND were both plotted against PROP. For both species, DAS AON counts were hugely variable at almost any PROP (Figure 3), and similarly variation between DAS AON and GBS IND counts were hugely variable in both over-estimating and under-estimating ‘true’ totals (Figure 4). These plots pooled all strata types.

Given all of the relationships explored, all urban strata and PROP were taken forward as potential explanatory variables for the end correction model for GBS IND counts from DAS AON counts for both HG and LB. No other explanatory variables were explored as it was surmised that any more may lead to over-fitting of the models.

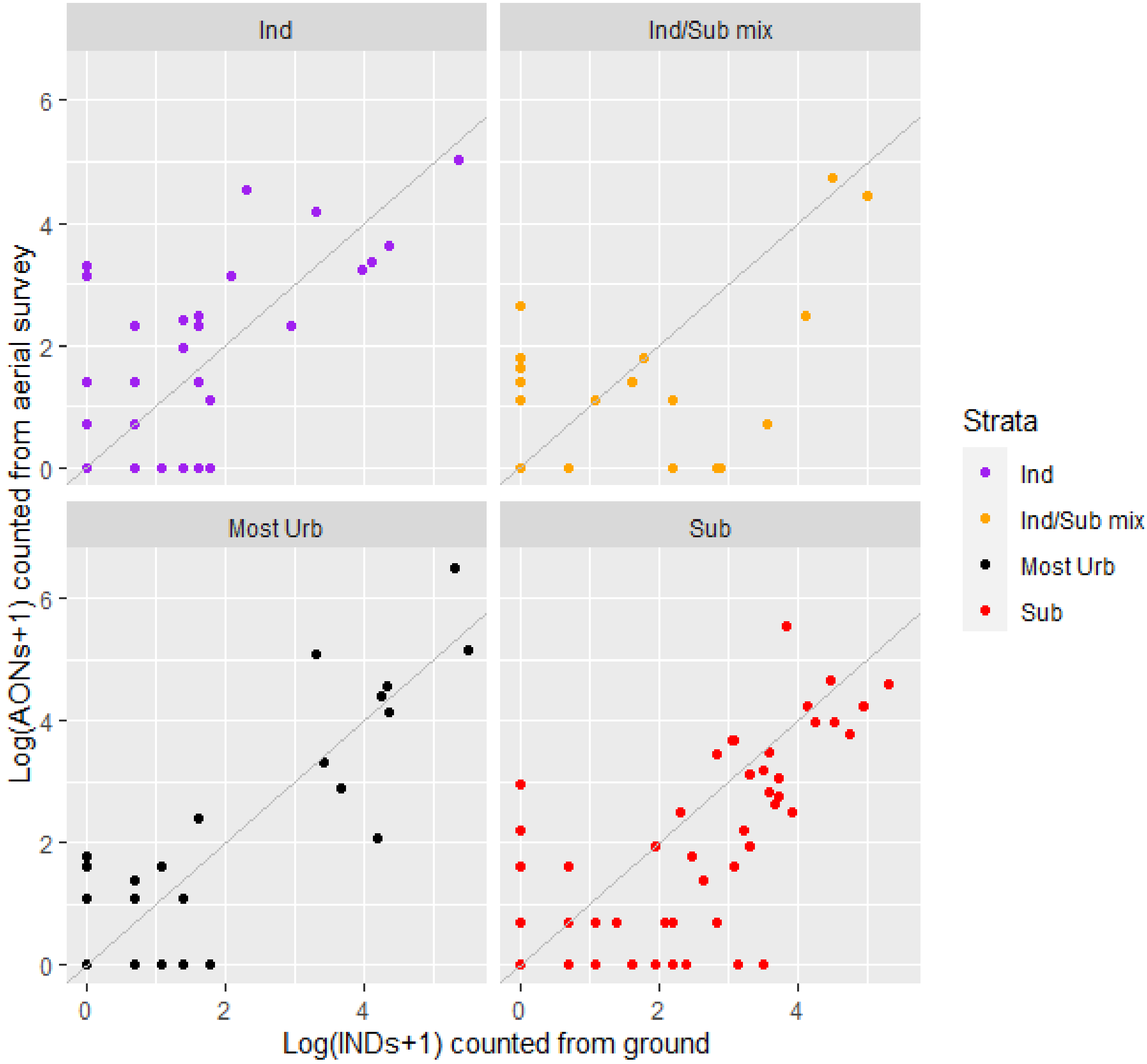

Figure 1. Scatterplot of **Herring Gull** AON counted via DAS and IND counted via GBS, split across the four urban strata, with a y=x reference line for aiding pattern observation. Both types of survey count have been transformed using the natural Log (+1). Each point represents a 1 km square surveyed in England and North Wales within the survey period 2018-2020 (inclusive). Note: there are not 235 visible dots partly due to multiple dots plotted on the 0 intersection.

.

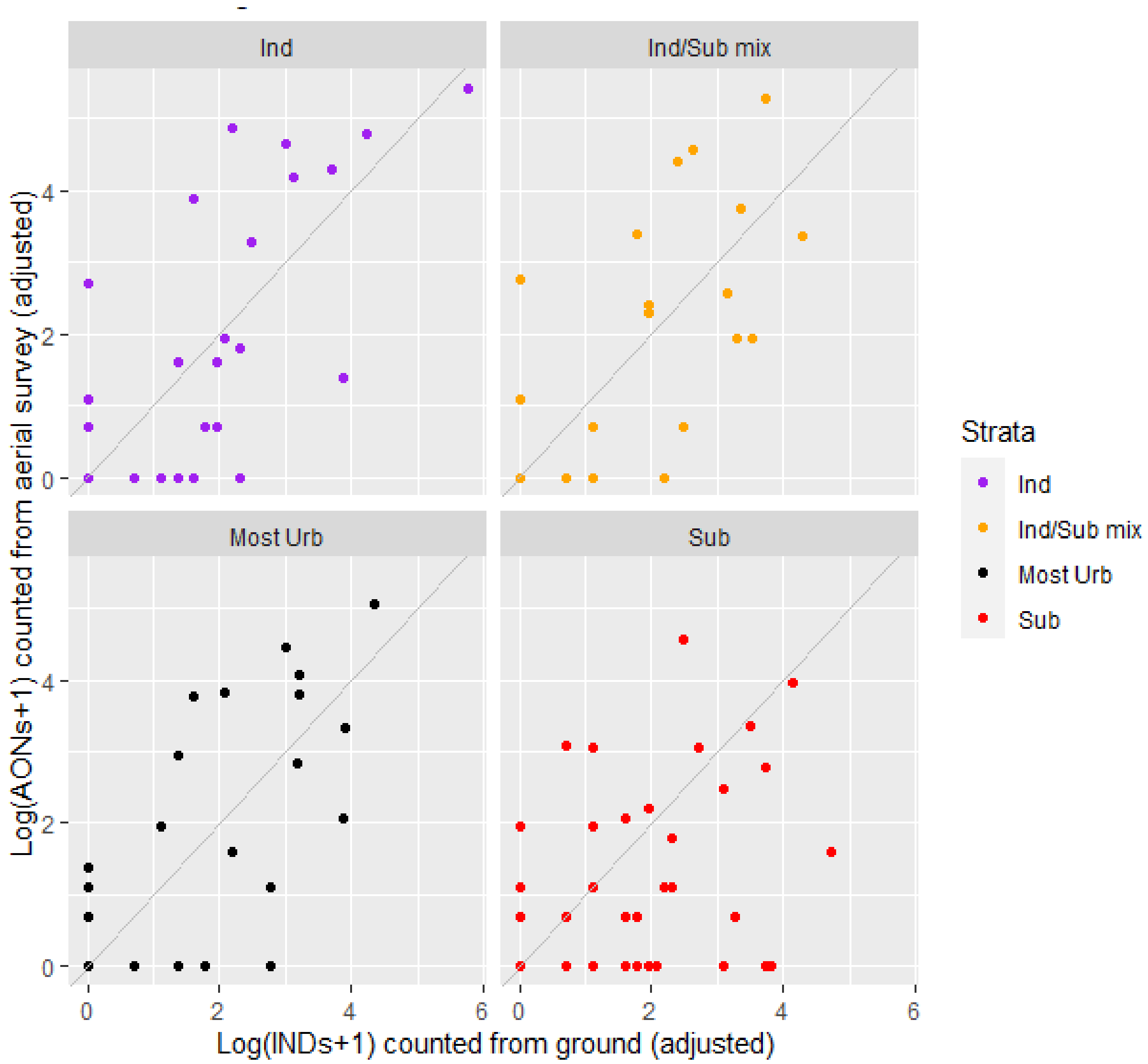

Figure 2. Scatterplot of **Lesser Black-backed Gull** AON counted via DAS and IND counted via GBS, split across the four urban strata, with a y=x reference line for aiding pattern observation. Both types of survey count have been transformed using the natural Log (+1). Each point represents a1 km square surveyed in England and North Wales within the survey period 2018-2020 (inclusive). Note: there are not 235 visible dots partly due to multiple dots plotted on the 0 intersection.

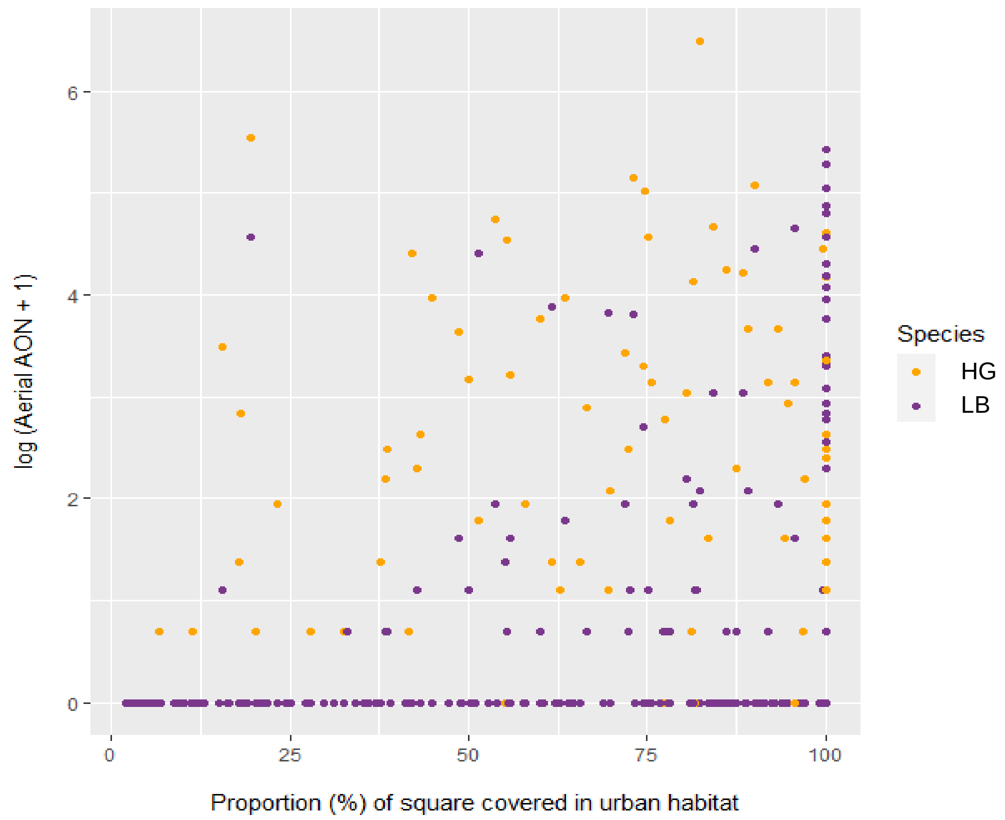

Figure 3. Scatterplot **HG** (orange) and **LB** gull (purple) DAS AON counts as a product of increasing proportion (%) of square covered in urban habitat. Counts were transformed using the natural log (+1) prior to calculating the difference

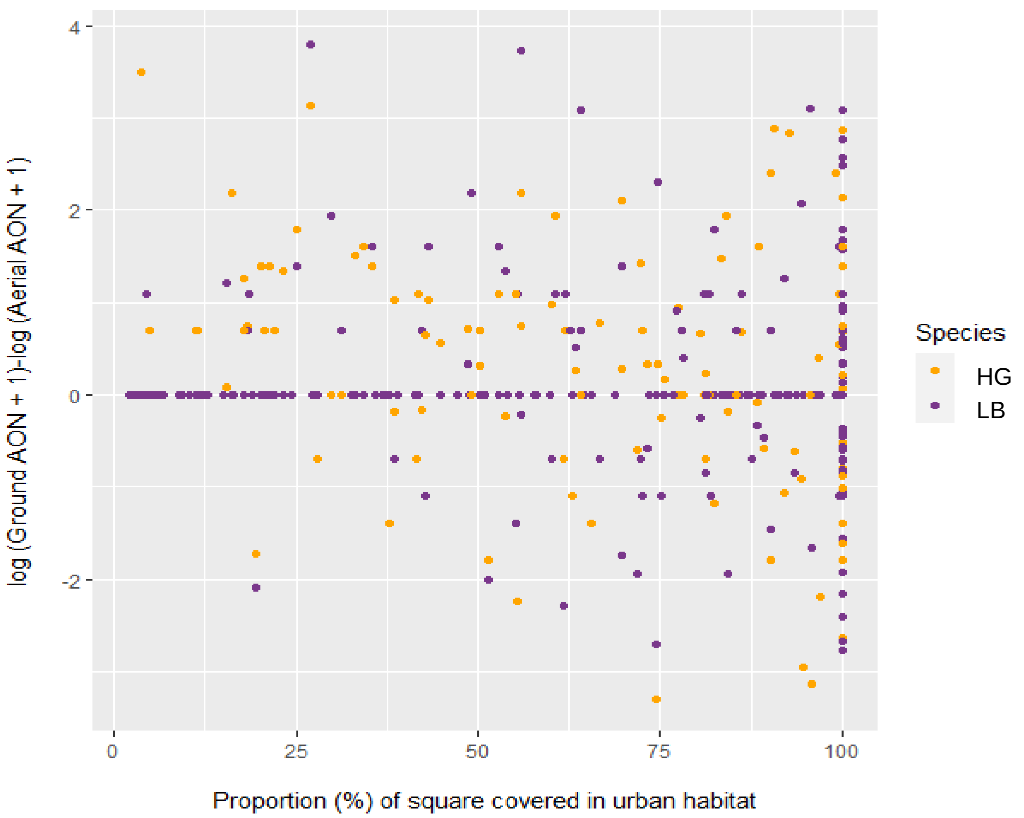

Figure 4. Scatterplot of the difference between GBS IND and DAS AON counts of **HG** (orange) and **LB** (purple) AON, as a product of increasing proportion (%) of square covered in urban habitat. Counts were transformed using the natural log (+1) prior to calculating the difference.

3.2 Data Analysis

Data analysis focussed on the hypothesis that true AON, assumed from DAS, can be sufficiently well predicted from GBS IND counts where other explanatory variables are accounted for. To model DAS AON counts as a function of the explanatory variables, negative binomial GLMs with log link functions, were fitted for each species.

Ten models were run for each species, starting from simple to progressively more complex combinations, and their goodness of fit was evaluated using AIC, the lowest being the preferred option (Burnham and Anderson, 2004). Also, differences between observed and predicted sample response were evaluated using RMSE, again the lower the value the closer the model is to that actually observed. For both species it was clear, based on the AIC values, that models containing all three explanatory variables: Log(GBS IND count +1) (discrete), Strata (categorical) and PROP (continuous), performed better than those that used just one or two variables (Table 5). The best performing models for each species contained the interaction term between the GBS IND count and PROP. However, the species differed in the overall complexity of the model with LB having better fit with the less complex model that did not contain the interaction between GBS IND and Strata, unlike HG.

Table 5. Comparison of AIC and RMSE values produced by ten negative binomial GLMs, with log link functions, using all explanatory variables. Values in bold highlight the overall best performing model for each species. Both species models were run on the same number of data points (n=235).

Model:	Herring Gull		Lesser Black-backed Gull	
	AIC	RMSE	AIC	RMSE
1. DAS AON ~ (log(GBS IND + 1))	973.4	38.2	829.9	125.5
2. DAS AON~ (log(GBS IND + 1)) + Strata	962.2	44.2	825.7	124.2
3. DAS AON~ (log(GBS IND + 1)) * Strata + (log(GBS IND + 1)) + Strata	962.6	34.3	831.03	117.0
4. DAS AON~ (log(GBS IND + 1)) + PROP	956.1	47.6	817.8	169.3
5. DAS AON~ (log(GBS IND + 1)) * PROP + (log(GBS IND + 1)) + PROP	947.5	38.7	808.9	39.5
6. DAS AON~ (log(GBS IND + 1)) + Strata + PROP	945.8	83.5	817.6	212.4
7. DAS AON~ (log(GBS IND + 1)) + PROP*Strata + PROP + Strata	948.1	83.5	818.2	384.5
8. DAS AON~ (log(GBS IND + 1))*PROP + (log(GBS IND + 1)) + PROP + Strata	927.2	64.8	**802.3**	**35.5**
9. DAS AON~ (log(GBS IND + 1))* Strata + (log(GBS IND + 1)) + Strata + PROP	940.1	85.2	820.5	109.2
10. DAS AON~ (log(GBS IND + 1))*PROP + (log(GBS IND + 1))*Strata + (log(GBS IND + 1)) + Strata + PROP	**921.5**	**40.4**	807.8	30.4

For HG, outputs (Table 6) of the preferred model (model 10) suggest that DAS AON counts differ significantly between the "Most Urban" and "Suburban" strata in comparison to those in "Industrial" strata (not explicitly noted in the table as it is the strata being tested against). Likewise, PROP also appears to have a significant effect on the counts of DAS AON. The relationship between DAS AON and GBS IND counts is also suggested to be significant in the "Most Urban" and "Sub" compared to the test category Industrial (not explicitly noted in the table as it is the test variable). In addition, this relationship also appears to be significantly influenced by the PROP. No significant dispersion was detected in the HG model residuals, nor any significant deviation away from the assumed model distribution of residuals (Kolmogorov-Smirnov test). There does, however, appear to be some weak zero-inflation in the residuals in the preferred model, but given that other residual outputs were not problematic, and noting that adding complexity with a zero-inflation element in the model would likely push the model towards over fitting, this weak zero-inflation in the residuals was deemed manageable.

Table 6. Model outputs and validation test results for the best performing model for **Herring Gull**. Results with statistical significance (p <0.05) are noted in bold. 'Interaction' relates to DAS AON with GBS IND.

DAS AON ~ (Log(GBS IND count +1))*PROP + Log(GBS IND count +1))*Strata + (Log(GBS IND + 1)) + Strata + PROP				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.254	0.583	3.866	**<0.001**
Log(GBS IND + 1)	1.698	0.288	5.893	**<0.001**
"Ind/Sub mix"	-1.131	0.598	-1.892	0.058
"Most Urb"	-2.210	0.645	-3.430	**<0.001**
"Sub"	-2.610	0.507	-3.386	**<0.001**
PROP	0.046	0.007	6.316	**<0.001**
Interaction "Ind/Sub mix"	0.113	0.287	0.397	0.691
Interaction "Most Urb"	0.681	0.266	2.556	**0.010**
Interaction "Sub"	0.631	0.232	2.722	**0.006**
Interaction PROP	-0.014	0.003	-4.231	**<0.001**
AIC	921.47			
Theta (Standard Error)	0.396 (0.061)			
RMSE	40.39			
DHARMa dispersion (p-value)	0.575 (0.264)			
Kolmogorov-Smirnov test (p-value)	0.053 (0.524)			
DHARMa Zero-inflation test (p-value)	1.090 (**0.048**)			

Table 7. Model outputs and validation test results for the best performing model for **Lesser Black-backed Gull**. Results with statistical significance (p <0.05) are noted in bold. 'Interaction' relates to DAS AON with GBS IND.

DAS AON ~ (Log(GBS IND count +1))*PROP + (Log(GBS IND + 1)) + PROP + Strata				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.967	0.670	-4.428	**<0.001**
Log(GBS IND + 1)	2.850	0.434	6.564	**<0.001**
PROP	0.042	0.008	5.068	**<0.001**
"Ind/Sub mix"	-0.443	0.524	-0.845	0.398
"Most Urb"	-0.471	0.517	-0.911	0.362
"Sub"	-1.404	0.428	-4.021	**0.001**
Interaction PROP	-0.019	0.004	-4.021	**<0.001**
AIC	802.27			
Theta (Standard Error)	0.275 (0.042)			
RMSE	35.46			
DHARMa dispersion (p-value)	0.473 (0.136)			
Kolmogorov-Smirnov test (p-value)	0.041 (0.816)			
DHARMa Zero-inflation test (p-value)	1.0314 (0.528)			

For LB (Table 7), outputs of the preferred model (model 8) broadly resemble those seen in the HG model. DAS AON counts are significantly different in the "Sub" strata compared to the test strata “Ind” (not explicitly noted in the table as it is the test variable). PROP is suggested to significantly influence the number of DAS AON counted and similarly significantly affect the relationship between DAS AON and GBS IND counts.. No dispersion or deviation from the assumed model distribution were apparent in the LB model residuals, nor was there any significant zero-inflation detected in the residuals.

Exploration of the residual graphs, with scaled residuals against the predictors, clearly show distributed values. There was very slight skew in higher counts of GBS HG IND towards higher residual values. This is likely a product of low sample size in this category.

3.3 Model performance at survey level

Data for each species were passed through the predict function (R Core Team, 2018), using the selected model to produce mean estimates and associated CLs of AON for each square. The predict function can produce estimates in the either the response or link form, the former can give negative values, which are not suitable for count data. To prevent this, the data were passed through a log link function then multiplied ‘back’ by the inverse of the link and rounded to the nearest whole number.

For HG (Table 8), the overall ‘true’ total of AON from all 235 comparable survey squares sits within the CL range for their model predicted AON. The area and stratum combinations that do not appear to be predicted well by the model were “Suburban” strata in Birmingham and the “Suburban” and “Industrial/Suburban mix” strata in the northeast England survey area. Two of these three combinations had the smallest numbers of true AON observed. These predications and CLs will expect to improve with more data.

Least confidence i.e. broadest CLs, around model predictions appeared to be in the “Most Urban” stratum for HG. Better confidence was around predicted values associated with lower DAS AON counts in the other three strata; also see Figure 5.

Table 8. Comparison of the preferred model AON predictions for **Herring Gull**, with 95% CLs in brackets, against ‘true’ DAS AON counts for each survey area, per stratum, and overall.

Survey areas (total squares)	Observed ‘true’ DAS AON	Model ‘predicted’ AON (95% CLs)
TOTAL (235)	**3,182**	**3,487 (1,265 - 10,325)**
Ind/Sub mix (31)	254	200 (69 - 681)
Most Urb (30)	1,307	1,565 (476 - 5,259)
Sub (127)	1,083	1,155 (506 - 2,644)
Ind (46)	538	567 (214 - 1,741)
Birmingham (79)	**174**	**351 (151 - 804)**
Ind/Sub mix (16)	46	90 (38 - 236)
Most Urb (15)	26	37 (15 - 77)
Sub (37)	10	62 (24 - 103)
Ind (11)	92	162 (74 - 388)
north Wales (78)	**1,755**	**2,291 (828 - 6,822)**
Ind/Sub mix (13)	205	87 (21 - 390)
Most Urb (8)	436	1,114 (339 - 3,753)
Sub (39)	1,001	890 (395 - 2,074)
Ind (18)	113	200 (73 - 605)
northeast England (78)	**1,253**	**845 (286 - 2,699)**
Ind/Sub mix (3)	3	23 (10 - 55)
Most Urb (7)	845	414 (122 - 1429)

Sub (51)	72	203 (87 - 467)
Ind (17)	333	205 (67 - 748)

Table 9. Comparison of the preferred model AON predictions for **Lesser Black-backed Gull**, with 95% CLs in brackets, against 'true' DAS AON counts for each survey area, per stratum, and overall.

Survey areas (total squares)	Observed 'true' DAS AON	Model 'predicted' AON (95% CLs)
TOTAL (235)	**2,205**	**3,069 (1,124 - 9,162)**
Ind/Sub mix (31)	536	550 (208 - 1,488)
Most Urb (30)	516	595 (251 - 1, 459)
Sub (127)	316	698 (212 - 2,716)
Ind (46)	837	1,226 (453 - 3,499)
Birmingham (79)	**1,777**	**2,451 (872 - 7,600)**
Ind/Sub mix (16)	525	419 (168 - 1,063)
Most Urb (15)	371	352 (149 - 869)
Sub (37)	126	603 (174 - 2,504)
Ind (11)	755	1,077 (381 - 3,164)
north Wales (78)	**329**	**321 (124 - 871)**
Ind/Sub mix (13)	11	126 (38 - 416)
Most Urb (8)	53	78 (31 - 188)
Sub (39)	190	79 (36 - 178)
Ind (18)	75	38 (19 - 89)
northeast England (78)	**99**	**297 (128 - 691)**
Ind/Sub mix (3)	0	5 (2 - 9)
Most Urb (7)	92	165 (71 - 402)
Sub (51)	0	16 (2 - 34)
Ind (17)	7	111 (53 - 246)

For LB (Table 9), the overall 'true' AON from all 235 comparable survey squares sits well within their predicted AON from the preferred model; and for the Birmingham and north Wales survey areas. The model, however, appeared to not work as well for the northeast England survey area, with the 'true' AON falling well outside their model predicted range, largely driven by considerable over-estimation in the "Ind" strata. Other potential problem areas in the LB model come from overestimates in the "Suburban" and "Industrial/Suburban mix" strata in Birmingham and north Wales, respectively. Conversely, the "Suburban" stratum was underestimated in north Wales.

Least confidence, i.e. broadest CLs, around model predictions appeared in the "Sub" and "Ind" strata for LG, but overall performed better for the "Most Urb" stratum; the opposite shown for HG. But like HG, better confidence was also around some predicted values associated with lower DAS AON counts, with some distinct exceptions; also see Figure 6.

It should be noted that models will always work better on the data that produced them, so the confidence around these estimates is likely better than what would be produced if another dataset were passed through it.

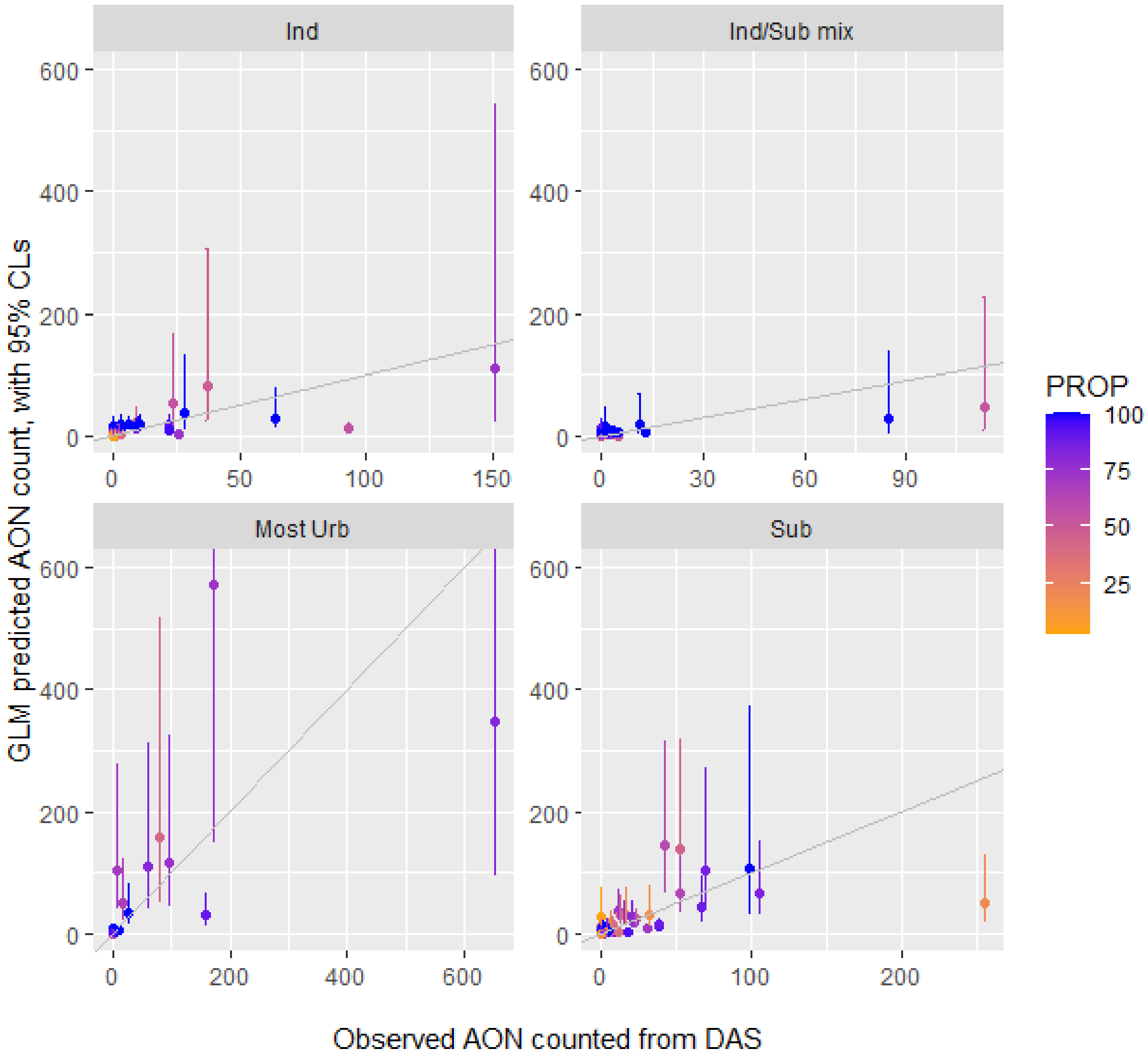

Figure 5. Scatterplot of the GLM predicted AON for **Herring Gull**, and their associated 95% CLs, against the observed DAS AON. Data points are coloured by the corresponding % urban fabric cover (PROP) for the square. Grey X=Y line added for guidance.

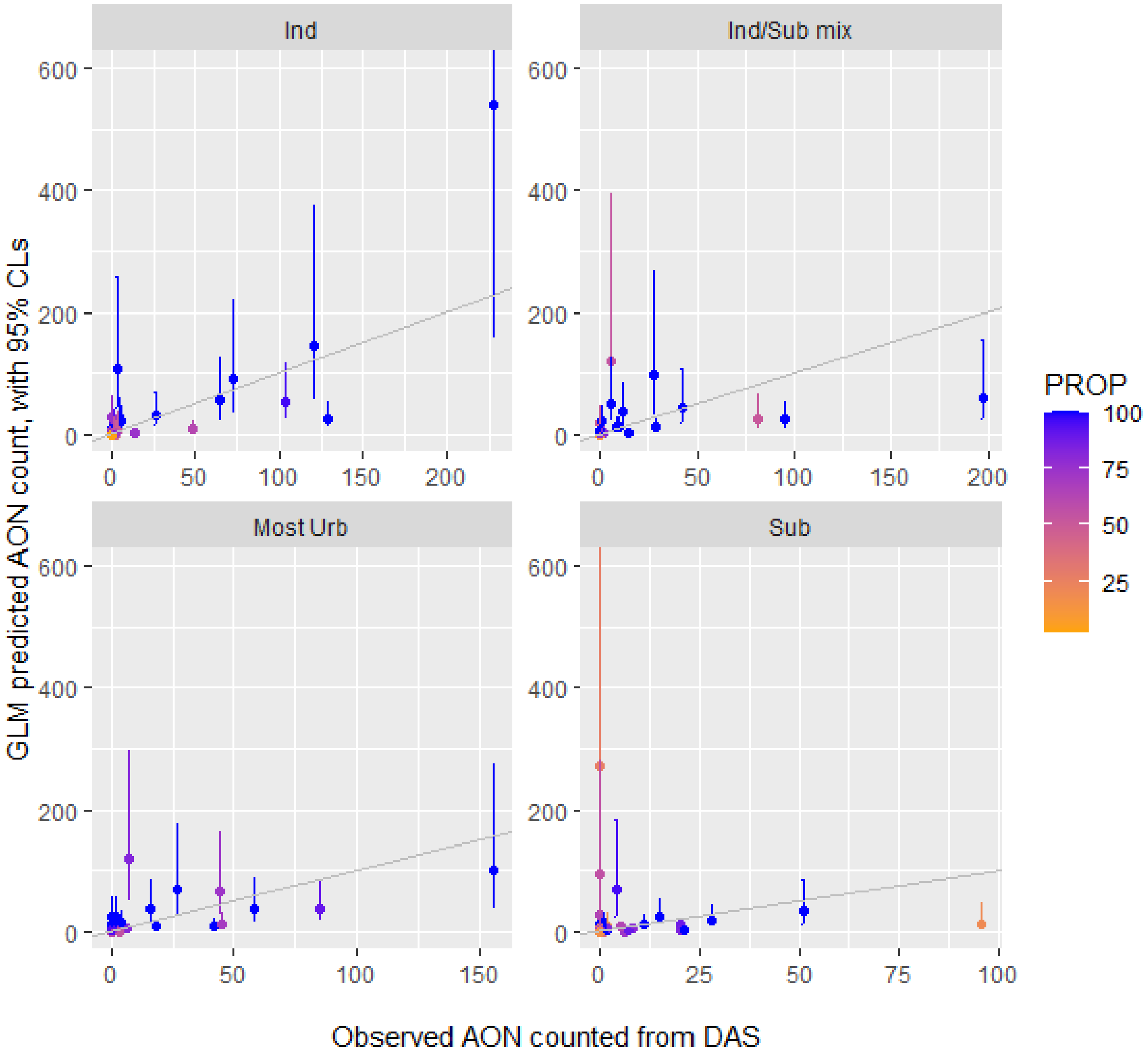

Figure 6. Scatterplot of the GLM predicted AON for **Lesser Black-backed Gull**, and their associated 95% CLs, against the observed DAS AON. Data points are coloured by the corresponding % urban fabric cover (PROP) for the square. Grey X=Y line added for guidance.

3.4 Model performance at survey area (population) level

To gain a better understand about sampling errors, in addition to the modelling error, population estimates for each of the survey areas were produced through bootstrapping. Urban strata vs. inland or coastal combinations were run separately for the whole survey area and for each survey areas, hence the sum of survey area AON estimates don't equate to the AON estimate presented for "Whole survey area" in Table 10 and Table 11.

Table 10. Bootstrapped population estimates for **Herring Gull** from the adjusted DAS AON data and the simulated predicted AON from the model (which accounts for the variation in the model). Estimates are broken down by each survey areas, stratum and inland or coastal combination. 95% CLs are represented by the 2.5th and 97.5th percentiles in the bootstrapping.

Survey areas (number of squares)	**DAS AON (95% CLs)**	**Modelled AON (95% CLs)**
Whole survey area (1,284)		
Ind/Sub mix - inland (37)	72 (35-119)	120 (33 - 391)
Ind/Sub mix - coastal (43)	975 (475 -1,515)	368 (50 - 1,870)
Most Urb - inland (41)	60 (31-95)	80 (22 - 241)
Most Urb - coastal (39)	3,740 (1,936 - 5,997)	3,412 (669 - 16,675)
Sub - inland (497)	123 (84 - 171)	611 (239 - 1,462)
Sub - coastal (379)	5,746 (4,448 - 7,119)	5,343 (2,268 - 12,296)
Ind - inland (91)	507 (348 - 672)	717 (220 - 1,869)
Ind - coastal (157)	2,802 (1,978 - 3,754)	2,108 (548 - 7,306)
All:	**14,009 (11,569-17,044)**	**14,354 (7,893 - 28,602)**
Birmingham (240)		
Ind/Sub mix - inland (29)	83 (45 - 128)	129 (30 - 420)
Most Urb - inland (20)	34 (12 - 61)	43 (8 - 149)
Sub - inland (179)	47 (24 - 79)	254 (87 - 729)
Ind - inland (12)	99 (49 - 166)	144 (25 - 507)
All:	**266 (191 - 347)**	**623 (326 - 1,249)**
north Wales (462)		
Ind/Sub mix - inland (11)	0 (0-0)	4 (0 - 60)
Ind/Sub mix - coastal (8)	272 (15 - 552)	191 (13 - 1,790)
Most Urb - inland (5)	0 (0-0)	1 (0-40)
Most Urb - coastal (10)	625 (297 - 986)	265 (23 - 1,757)
Sub - inland (194)	112 (76 - 152)	116 (15 - 642)
Sub - coastal (165)	5,092 (4,007 - 6,482)	6,183 (2,489 - 16,183)
Ind - inland (40)	155 (67 - 266)	21 (2 - 162)
Ind - coastal (29)	248 (128 - 393)	957 (273 - 4,015)
All:	**6,517 (5,294 - 8,030)**	**6,436 (2,859 - 14,438)**
northeast England (582)		
Ind/Sub mix - inland (10)	0 (0-0)	11 (0-100)
Ind/Sub mix - coastal (22)	21 (9 - 36)	135 (10 - 1,714)
Most Urb - inland (6)	0 (0-0)	5 (0-79)
Most Urb - coastal (39)	5,410 (2,884 - 8,473)	310 (36 - 2,438)
Sub - inland (124)	0 (0-0)	184 (32 - 894)
Sub - coastal (215)	398 (196 - 678)	1,995 (503 - 9,714)
Ind - inland (127)	0 (0-0)	199 (26 - 985)
Ind - coastal (39)	903 (446 - 1,480)	310 (36 - 2,438)
All:	**6,674 (3,956 - 10,403)**	**4,010 (1,235 - 17,395)**

For HG (Table 10), the overall 'true' AON estimate sits well within the 95% CLs of the modelled estimated AON, as is also the case for the north Wales and northeast England survey areas, but not in the Birmingham survey area, where there appears to have been over-estimation by the modelled AON, largely driven by the "Sub – inland" combination.

Only about half (11 of 20) of the strata vs. inland or coastal combinations of 'true' mean estimated AON across the three survey areas sit within the 95% CLs of their modelled estimated AON counterparts. The 95% CLs of DAS estimated AON are within another quarter of these (5 of 20) and the four exceptions with no overlap in either the CLs of the 'true' AON or modelled AON totals included the "Sub – inland" combination strata for Birmingham and the other three were in the northeast England survey area. This survey area was especially problematic due to mostly comprising of very small or very large 'true' AON totals. Despite some favourable sample sizes in the northeast England survey area, the absence of gulls from these was an exacerbating factor. When pooled together with data from the other survey areas, these errors were partially propagated, but despite having the largest sample of survey squares, the "Sub – inland" combination was supported by disproportionately fewer 'real' HG AON and modelling appears to have struggled with this.

Table 11. Bootstrapped population estimates for **Lesser Black-backed Gull** from the adjusted DAS AON data and the simulated predicted AON from the model (which accounts for the variation in the model). Estimates are broken down by each survey areas, stratum and inland or coastal combination. 95% CLs are represented by the 2.5th and 97.5th percentiles in the bootstrapping.

Survey area (number of squares)	**DAS AON (95% CLs)**	**Modelled AON (95% CLs)**
Whole survey area (1284)		
Ind/Sub mix - inland (37)	820 (393 - 1,465)	500 (82 - 2,129)
Ind/Sub mix - coastal (43)	53 (30 - 79)	171 (50 - 3,826)
Most Urb - inland (41)	871 (447 - 1,380)	667 (114-2,787)
Most Urb - coastal (39)	429 (169 - 751)	426 (34 - 3,382)
Sub - inland (497)	1,105 (755 - 1,475)	3,815 (892 - 18,407)
Sub - coastal (379	1,017 (632 - 1,465)	413 (126 - 1,289)
Ind - inland (91)	3,052 (2,057 - 4,270)	2,724 (450 - 17,601)
Ind/Sub mix - inland (37)	448 (234 - 725)	729 (123 - 2,621)
All:	**7,845 (6,366 - 9,417)**	**11,900 (5,421 - 31, 960)**
Birmingham (240)		
Ind/Sub mix - inland (29)	929 (459 - 1,522)	594 (104 - 2,705)
Most Urb - inland (20)	487 (185 - 900)	344 (45 - 1,723)
Sub - inland (179)	599 (355 - 889)	1,947 (434 - 11,129)
Ind - inland (12)	816 (498 - 1,306)	658 (59 - 5,063)
All:	**2,888 (2,121 - 3,739)**	**4,390 (1,631 - 15,954)**
north Wales (462)		
Ind/Sub mix - inland (11)	0 (0-0)	2 (0 - 16)
Ind/Sub mix - coastal (8)	14 (5 - 27)	30 (0 - 1,319)
Most Urb - inland (5)	0 (0-0)	0 (0-20)
Most Urb - coastal (10)	64 (8-187)	27 (0 - 806)
Sub - inland (194)	0 (0-0)	46 (0-986)
Sub - coastal (165)	955 (596 - 1,434)	296 (62 - 1,161)
Ind - inland (40)	62 (18 - 119)	10 (0 - 100)
Ind - coastal (29)	186 (46 -366)	69 (3 - 484)
All:	**1,324 (910 - 1,872)**	**820 (248 - 2,654)**
northeast England (582)		
Ind/Sub mix - inland (10)	0 (0-0)	0 (0-0)
Ind/Sub mix - coastal (22)	0 (0-0)	14 (0 - 140)

Most Urb - inland (6)	85 (0 - 255)	0 (0 - 90)
Most Urb - coastal (39)	488 (205 - 913)	476 (14 - 5,868)
Sub - inland (124)	0 (0-0)	35 (0 - 145)
Sub - coastal (215)	0 (0-0)	73 (20 - 207)
Ind - inland (127)	52 (28 - 80)	352 (0 - 2,875)
Ind - coastal (39)	15 (3 - 32)	172 (10 - 1,050)
All:	**644 (289 - 1,068)**	**1,676 (381 - 8,408)**

For LG (Table 11), the overall 'true' AON estimate sits well within the 95% CLs of the modelled estimated AON, as is also the case for all three survey areas, although CLs accommodatingly very broad. The individual urban strata vs. inland or coastal combinations also all fall within modelled estimated AON CLs, except in five of the seven combinations where the 'true' total was zero. In the remaining two, the modelled LCL was also zero. As seen with HG results, 'true' gull absence appears to confound modelled estimates.

Both species appear to have consistently larger confidence limits around estimates in northeast England where model estimates of AON were confounded by 4 (in HG) and 6 (in LB) of the eight strata vs. coastal or inland combinations samples having 'true' DAS AON as absent, or nearly absent, for which the model some instances predicted the presence of moderately sized populations. This was despite sufficiently sized squares samples for some of these combinations. Confidence in the HG model prediction for this survey area was further reduced by a disproportionately large 'true' count for the "Most Urb – coastal" stratum, which the model failed to accurately predict. However, at the whole survey scale, this and other combinations with higher HG densities were propagated rather well.

For LB overall, the "Ind/Sub mix - coastal" combination had the lowest LB density, but was sufficiently sampled. Its 'true' AON estimate mean was almost outside of the modelled AON CL range, thus emphasising the issue also identified above with HG concerning gull absence confounding modelling. The model was particularly poor at predicting the true "sub - inland" combination overall and the reason for this is not yet known.

4 Discussion

4.1 Relationships between GBS and DAS counts and model improvement

This study set out to expand our understanding about relationships between GBS and DAS counts and to determine if models produced by Woodward *et al.* (2020) could be improved with the availability of additional data and alternative or additional explanatory variables.

Basic results show some relationships between HG GBS count data and the 'true' DAS AON present in urban environments, but this was less clearly shown for LB data. These relationships improved with the inclusion of the explanatory variables, particularly the percentage cover of urban fabric within each square (PROP) and the predominant urban fabric strata present in each square; Woodward *et al.* (2020) did not consider PROP as an explanatory variable.

Like the findings in Woodward *et al.* (2020), this study saw the strongest relationship of DAS AON counts with GBS IND counts. Both studies observed better model fit with the presence of an interaction between GBS IND counts and strata for HG. Likewise, both studies also showed better fit for models excluding the interaction between GBS IND and strata for LB.

Although PROP appeared to differ with urban strata, intuitively it was a relevant variable to explore and its inclusion improved model fitness for both species. Additional improvements were observed, for both species, when PROP was added with interaction between it and the GBS IND count. This result makes sense given the higher variability seen between the GBS IND counts and the DAS AON counts as PROP increased. However, it should be noted that PROP may be acting as a proxy to something else affecting the counts, comparing the population estimates for the survey areas from this study and Woodward *et al.* (2020), may indicate this to be the case.

The most comparable survey area between the two studies, that will not have had major changes through the adjustment of the DAS count to account for coverage, is Birmingham. CLs around the modelled HG AON estimate for this area were improved in this study compared to Woodward *et al.* (2020). However, for LB this study produced wider CLs. It is difficult to compare outputs for both studies in the north Wales due to the adjustments made to the DAS AON counts and the exclusion of some squares that had insufficient DAS coverage.

The process of excluding data from poorly covered squares from both DAS and GBS in this study resulted in a dataset that had comparative total numbers of squares across the three breeding seasons. However, this unique approach did not lead to vastly different results to those in Woodward *et al* (2020), at least with respect to model performance. These similarities validate conclusions from both studies that the models used are a good fit for these data. An improvement in this study was to resolve issues of over-dispersion in the HG model residuals experienced in Woodward *et al* (2020), suggesting an improved goodness of fit for the data.

However, the model performance and robustness, when used to produce population estimates is somewhat less clear. In some instances, the new model appears to improve confidence and in others it does not. Understanding more about how PROP behaves in these models is recommended. The exclusion of squares with <50% urban coverage from GBS and DAS, which reduced sample sizes adds uncertainty that could have contributed to wider CLs produced in this study. DAS coverage was not accounted in the Woodward *et al, (*2020) study, which may be why the CLs are larger in this studies results.

4.2 Utilisation of models

Validation steps on the residuals of the models showed no concerning dispersion or deviation from assumed model distribution for either species. There appeared to be weak zero-inflation in the favoured HG model but given that other validation steps showed no significant problems this was not deemed to be an issue. Although model fit was good for both species, there were various irregularities that lowered confidence in favoured models as a means of converting GBS IND counts.

Models appear to be less effective at predicting 'true' DAS AON where these exist at lower densities for HG and LG. For HG (Table 8), in all three examples of survey area strata predications where the CL range over-estimated the true AON, these occurred where the true mean HG AON estimates were at lowest densities. For LG (Table 9), in all eight examples of poorer model predications, low density was also a theme. This reduced predictability also occurred in samples of reasonable sizes, although smaller sample sizes of urban strata were an exacerbating factor. Higher 'true' density of HG tended to produce the broadest predicted CLs whereas for LB, broad CL occurred more erratically with often broad, i.e. lower confidence, at lower densities too.

For LB, the model estimated AON was much less consistent in predicting the 'true' AON estimate. Investigation into this found there were a few high GBS IND counts in squares where there were little to no, DAS AON counts, mainly from industrial and suburban strata. Drivers of this were elusive; a possible explanation was that many of these problematic squares were surveyed by GBS 10 to 28 days prior to the DAS. Therefore, any factor that resulted in their dispersal during the intervening period may account for these discrepancies. These could also be driven by non-attending adults foraging in areas away from their nesting square or natural colony.

In each of the survey areas, for HG the closest predications and greatest confidence occurred in the "Sub" strata of north Wales, which had a relatively large sample size, a large population of true AON that occurred at a relatively high density. For LG, despite the slightly below average sized samples of "Most Urb" and "Ind/Sub mix" strata in Birmingham, these supported large numbers of true AON at relatively high densities, and the model produced the closest predications and greatest confidence there.

While under- and over-estimation from GLMs at small spatial scales were problematic, these appeared to balance out as the scale increases, in that the 'true' AON estimates and their CLs derived from the combined survey areas are all within the CLs of the modelled AON estimates.

4.3 Conclusion and recommendation for future work

It is recommended that in order to narrow CLs and increase the accuracy in population estimates through increased statistical power, additional surveys would be required. These additional surveys should be coordinated GBS and DAS that target urban strata, and areas, which are less well represented. Furthermore, targeting of known high breeding density squares is advisable. Another recommendation to improve survey design is to decrease time lag between GBS and DAS counts, thus reducing the likelihood of human disturbance events or other factors influencing intra-seasonal gull movements.

Analysis could be made more robust through the inclusion of more detailed and finer scale data about the explanatory variables. For example, more information about building height could improve investigation about implications concerning GBS detectability. As stated previously, this could be what PROP is acting as a proxy for and may not be needed if this information were available. Details about building design to explain some variations within urban strata may help if these closely relate to nesting opportunities, for instance, presence of chimney pots in the suburban strata, and presence of low angled, old asbestos roof surfaces in the industrial strata, anecdotally are habitats favoured by nesting gulls. Re-categorising urban fabric or accounting for proportions of each urban strata within each survey square are additional considerations. However, any additional variables would need to be met with either a substitution of another variable or increased sample size, otherwise there is the risk of over-fitting models.

Better coverage of each square that is surveyed by DAS and GBS is desirable. DAS can be resolved by modifying the camera array to avoid gaps. This will reduce potential error from the assumption of even distribution of gulls within the square when correcting up to 100% coverage.

This report sets out how statistical analysis and modelling were performed to produce species-specific correction models, which could be applied to survey data to produce up-to-date urban nesting HG and LB population estimates at national or larger scales. The inclusion of the extra explanatory variable PROP, in theory, provides better model fit, however, in practice, the CLs for population estimates for the survey areas are still considerably wide.

If these models are to be used as a tool to produce population estimates they should be used with care and caution. It is not recommended to use these models to report estimated AON at small spatial scales. Even at larger spatial scales, as indicated by results in this report, there remain some weaknesses in the ability of these models to always adequately predict populations where given sets of environmental variables exist, with which to then extrapolate to the desired scale. Even at national scales, if gull populations are supported by few, very large dense colonies plus uneven lower densities and regional absence of nesting gulls, these characteristics may continue to confound population estimates, without appropriate adaptations to survey design and modelling.

When communicating population estimates derived from these models, sources and expressions of errors, such as CLs, should also be transparently shared to allow prudent interpretation. The use of population estimates to inform policy decisions should be similarly caveated.

References

Balmer, D.E., Gillings, S., Caffrey, B.J., Swann, R.L., Downie, I.S., Fuller, R.J., 2013. Bird Atlas 2007-11: The Breeding and Wintering Birds of Britain and Ireland. British Trust for Ornithology.

Blight, L.K., Bertram, D.F., Kroc, E., 2019. Evaluating UAV-based techniques to census an urban-nesting gull population on Canada's Pacific coast. J. Unmanned Veh. Syst. https://doi.org/10.1139/juvs-2019-0005

Burnham, K.P., Anderson, D.R., 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociol. Methods Res. 33, 261–304. https://doi.org/10.1177/0049124104268644

Christie, K., Gilbert, S., Brown, C., Hatfield, M., Hanson, L., 2016. Unmanned aircraft systems in wildlife research: Current and future applications of a transformative technology. Front. Ecol. Environ. 14, 241–251. https://doi.org/10.1002/fee.1281

CLC, (Corine Land Cover), 2012. Version 18.5.1 [WWW Document]. URL https://www.eea.europa.eu/data-and-maps/data/external/corine-land-cover-2012 (accessed 11.20.20).

Coulson, J.C., Coulson, B.A., 2015. The accuracy of urban nesting gull censuses. Bird Study 62, 170–176. https://doi.org/10.1080/00063657.2015.1013523

Eaton, M., Aebischer, N., Brown, A., Hearn, R., Lock, L., Musgrove, A., Noble, D., Stroud, D., Gregory, R., 2015. Eaton, M., Aebischer, N., Brown, A., Hearn, R., Lock, L., Musgrove, A., Noble, D., Stroud, D. and Gregory, R., 2015. Birds of Conservation Concern 4: the population status of birds in the UK, Channel Islands and Isle of Man. Br. Birds 108, 708–746.

Hartig, F., Lohse, L., 2020. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models [WWW Document]. URL https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html (accessed 11.16.20).

Hothorn, T., Zeileis, A., Farebrother (pan.f), R.W., Cummins (pan.f), C., Millo, G., Mitchell, D., 2020. lmtest: Testing Linear Regression Models.

Mitchell, P.I., Newton, S.F., Ratcliffe, N., Dunn, T.E., 2004. Seabird Populations of Britain and Ireland. JNCC, Peterborough.

OS OpenData, 2016. Mean High Water Springs.

QGIS.org, 2020. QGIS Geographic Information System. QGIS Association.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raven, S.J., Coulson, J.C., 1997. The distribution and abundance of Larus gulls nesting on buildings in Britain and Ireland. Bird Study 44, 13–34. https://doi.org/10.1080/00063659709461035

Ripley, B., Venables, B., Bates, D.M., ca 1998), K.H. (partial port, ca 1998), A.G. (partial port, Firth, D., 2020. MASS: Support Functions and Datasets for Venables and Ripley's MASS.

Ross, K., Burton, N., Balmer, D., Humphreys, E., Austin, G., Goddard, B., Schindler-Dite, H., Rehfisch, M., 2016. Urban Breeding Gull Surveys: A Review of Methods and Options for Survey Design. https://doi.org/10.13140/RG.2.2.26828.03205

Thaxter, C., Horswill, C., Ross, K., Austin, G., Balmer, D., Burton, N., 2017. Urban Breeding Gull Surveys: A Survey Design Simulation. https://doi.org/10.13140/RG.2.2.30183.47523

Walsh, P.M., Halley, D.., Harris, M.., del Nevo, A., Sim, I.M.., Tasker, M.., 1995. Seabird monitoring handbook for Britain and Ireland: a compilation of methods for survey and monitoring of breeding seabirds.

Weiß, F., Büttger, H., BAER, J., Welcker, J., Nehls, G., 2016. Erfassung von Seevögeln und Meeressäugetieren mit dem HiDef Kamerasystem aus der Luft. Seevögel 37, 14–21.

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., RStudio, 2020. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.
Woodward, I.D., Austin, G.E., Balmer, D.E., Boersch-Supan, P., Irwin, C.G., Scott, M.S., Thaxter, C., Webb, A., Burton, N.H.K., 2020. Improving Methods for Counting Dispersed Bird Populations: Urban Gull Case Study (Science and research project report No. BE0149). Department for Environment, food and Rural Affairs.

Appendix 1: urban gull survey methodology

Urban nesting gulls have expanded their distribution over the last 15-20 years. To be able to produce population estimates and distribution maps, a new method and sample regime has been created. The method consists of surveying a sample of 1 km squares in urban areas, across the UK and Ireland. The sample is a stratified random sample based on the number urban squares, and the ratio of strata[1], found within each county. With such sampling, it is likely that some squares will not contain gulls, however, for the data to be statistically robust and not biased, these squares will still need to be surveyed **and reported on**.

The methods for the survey are **ground-based** counts of herring and Lesser Black-backed Gulls, vantage points are not necessary for this survey. If a clear sight of several roofs can be gained at the top of street or at a junction etc, there is no need to walk the entirety of the street. You only need to survey the urban areas of the square, surrounding fields etc, do not need to be included in the count. Not all birds will be able to be viewed from the ground, please do not worry if you think you have missed some, just record what you can see. The analysis of these results adjusts for the lower detection rate that occurs in urban environments.

Counts for each species will need to be reported on separately, for each 1 km square. A count for **all three** units below will need to be recorded for each species:

- **AON** = apparently occupied nest (well-constructed nest or scrape nest, either containing eggs or young, or capable of holding eggs (possibly attended by an adult) or an apparently incubating adult),
- **AOT** = apparently occupied territory (estimated by the spacing of birds or pairs on different rooftops and observations of apparent territorial behaviour, when actual nests cannot be discerned. Any AONs should also be considered a territory, so the number of AOTs will always be equal to or greater than the number of AONs.)
- **IND** = Individual adults (Count the total number of birds in full adult plumage. Individual birds should only be counted once. However, where movement occurs it will sometimes be impossible to be certain whether some birds have already been counted in which case you should use your best judgement to decide. Birds in flight can be counted if it is clear they are using rooftops in the square, but birds observed flying over the square should not be counted)

Please note these counts are cumulative i.e. the count of AOT's will be the number of AON's observed plus any extra AOT's seen, the number of individuals will be all birds seen, including those on the nests and territories. An example recording sheet is attached at the end of this document to highlight this. Proposed survey timings can also be found at the end of this document.

Also needing to be recorded:

- The survey square, date of survey, time of survey and weather. These should be noted on the count sheets in the appropriate box.
- A note of any use of gull deterrents or control measures e.g. netting on roofs will be much appreciated.
- A rough estimate of the % of the square that **could not** be accessed for survey (i.e. private land)
- Finally, use the comment section for any other observations made.

Please use this link to submit your data:

Survey timings

*These timings are only a guide, local knowledge and own judgement based on when the gulls are at peak incubation should be used.

Survey Timings	
England	Late April – Late May (Southern & inland surveys closer to the earlier period progressively getting later further North & towards the coast)
Wales	Late April – Late May (Inland colonies closer to the earlier period, coastal can be towards the latter half)
Northern Ireland	May (coastal colonies may breed closer to the latter half of the period)
Scotland	May – 1st week June (coastal colonies may breed closer to the latter half of the period)

Seabirds Count – Urban Gull Survey Form

Recorder name:	Daisy Burnell	County/ district of survey:	Cleveland
Recorder contact details: Phone:	01224 266577	email:	Daisy.burnell@jncc.gov.uk

Square grid ref	Species Code	Date (DD/MM/YY)	Start Time (24hr)	End Time (24hr)	Counts			% of square that could not be accessed	Weather			Gull deterrent present	description	Additional comments
					AON	AOT	IND		Vis.	Wind	Rain			
NZ5233	HERGU	20/05/19	10:00	11:30	10	13	16	5	1	1	2	No		90% of this square is the sea. The % not accessed is based on the urban area that actually contains urban fabric.
NZ5233	LBBGU	20/05/19	10:00	11:30	4	7	9	5	1	1	2	No		As above
NZ5133	HERGU	19/05/19	12:30	13:30	4	9	8	10	1	0	1	Yes	Netting on some rooftops	Some birds resting on water, but not included in count.
NZ5133	LBBGU	19/05/19	12:30	13:30	0	0	3	10	1	0	1	Yes	Netting on some rooftops	As above

The number of AOTs will always be either equal or more than the count of AONs. In this example 10 AONs could be discerned and 3 AOTs could be seen therefore the total AOT is 10+3=13

The number of IND is the number of all adult birds present on suitable nesting habitat. This includes those sitting on nests or associated with territories. It will almost always be equal or more than the AOT count. However, a nest counted as an AON/AOT with no adult present will mean a lower count of IND.

Zero counts are very important, do not forget to report these. If there are adults around on suitable urban nesting habitat don't forget to count them, even if there are no visible nests.

Appendix 2: relationship between DAS AON and GBS AON

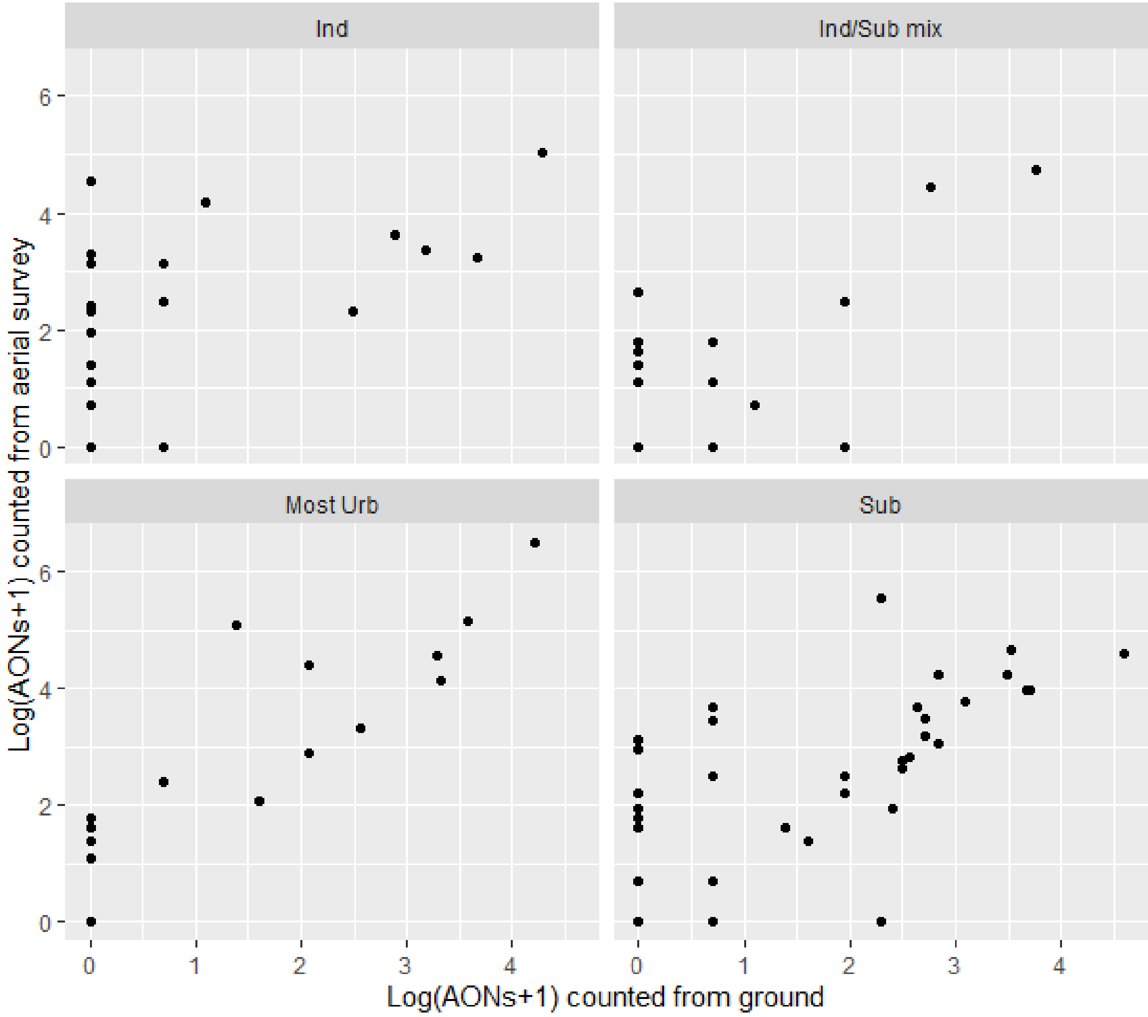

Figure 7. Comparison of **Herring Gull** AON counts (natural logged +1), from DAS and GBS across 235 survey squares.

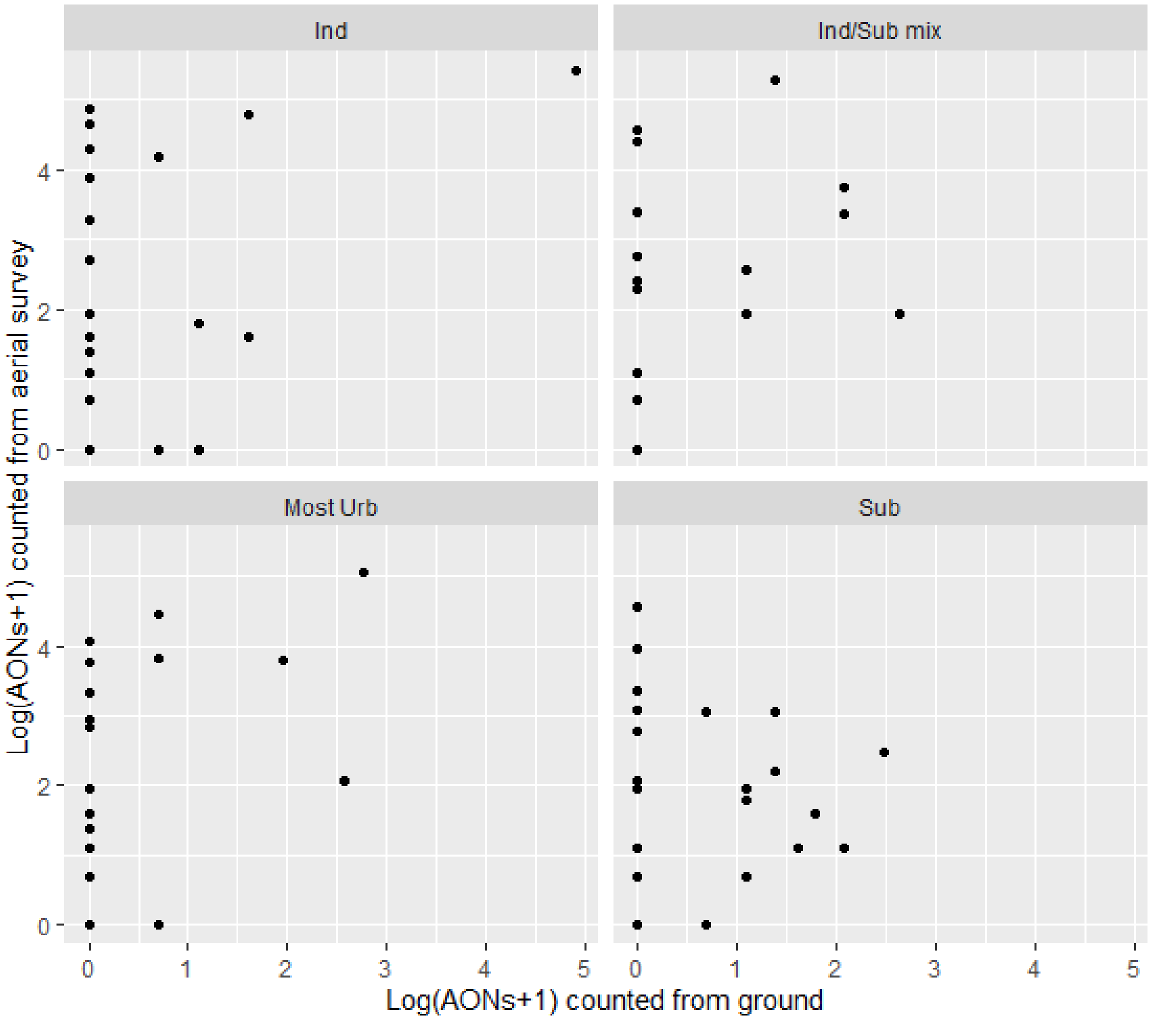

Figure 8. Comparison of **Lesser black-backed Gull** counts (natural logged +1), from DAS and GBS across 167 survey squares.

Appendix 3: AIC and RMSE output values for GBS AON and DAS AON

Table 12. AIC and RMSE values for the comparable models run using the GBS AON (adjusted counts) rather than the GBS IND counts.

	HERGU		LBBGU	
model	AIC	RMSE	AIC	RMSE
1. A_AON_adj ~ (log(GL_AON_adj + 1)	1002	51.7	910.1	52.5
2. A_AON_adj ~ (log(GL_AON_adj + 1)) + Strata	990.5	127.5	902.5	39.4
3. A_AON_adj ~ (log(GL_AON_adj + 1))*Strata	992.9	96.5	908.4	25.7
4. A_AON_adj ~ (log(GL_AON_adj + 1)) + PROP	994.3	109.5	897.9	27.3
5. A_AON_adj ~ (log(GL_AON_adj + 1))*PROP	993	45	899.5	48.5
6. A_AON_adj ~ (log(GL_AON_adj + 1)) + PROP + Strata	979.7	424.3	899	26.6
7. A_AON_adj ~ (log(GL_AON_adj + 1)) + PROP * Strata	982.6	365	887.4	104.7
8. A_AON_adj ~ (log(GL_AON_adj + 1))* PROP + Strata	973.8	193.6	900.5	45.4
9. A_AON_adj ~ (log(GL_AON_adj + 1))* Strata + PROP	977.2	782.8	904.6	24
10. A_AON_adj ~ (log(GL_AON_adj + 1))* Strata + (log(GL_AON_adj + 1))* PROP	973	107.7	906.2	27.1

Appendix 4: loglikelihood tests of negative binomial models and Poisson models

Table 13. Herring Gull outputs from a loglikelihood ratio test on the comparable Poisson model of the best fit negative binomial model.

Model	#Df	LogLik	Df	Chisq	Pr(>Chisq)
Model 1 (Poisson): DAS_AON_adj ~ (log(GBS_IND_adj + 1)) * Strata + (log(GBS_IND_adj + 1)) * PROP	10	-1955.62			
Model 2 (negbin): DAS_AON_adj ~ (log(GBS_IND_adj + 1)) * Strata + (log(GBS_IND_adj + 1)) * PROP	11	-449.73	1	3011.8	**<0.001**

Table 14. Herring Gull DHARMa outputs from the Poisson equivalent model of the selected negative binomial model.

DAS_AON_adj ~ (log(GBS_IND_adj + 1)) * Strata + (log(GBS_IND_adj + 1)) * PROP **(family = Poisson)**	
DHARMa dispersion	7.24 (<0.001)
Kolmogorov-Smirnov test	0.52 (<0.001)
DHARMa zero-inflation	5.10 (<0.001)

Table 15. Lesser Black-backed Gull outputs from a loglikelihood ratio test on the comparable Poisson model of the best fit negative binomial model.

Model	Df	LogLik	Df	Chisq	Pr(>Chisq)
Model 1 (Poisson): DAS_AON_adj ~ (log(GBS_IND_adj + 1)) * PROP + Strata	7	-1719.41			
Model 2 (negative binomial): DAS_AON_adj ~ (log(GBS_IND_adj + 1)) * PROP + Strata	8	-393.13	1	2652.6	<0.001

Table 16. Lesser Black-backed Gull DHARMa outputs from the Poisson equivalent model of the slected negative binomial model.

DAS_AON_adj ~ (log(GBS_IND_adj + 1)) * PROP + Strata **(family = Poisson)**	
DHARMa dispersion	6.61 (<0.001)
Kolmogorov-Smirnov test	0.33 (<0.001)

DHARMa zero-inflation	2.12 (<0.001)

Appendix 5: PROP ranges for four urban strata

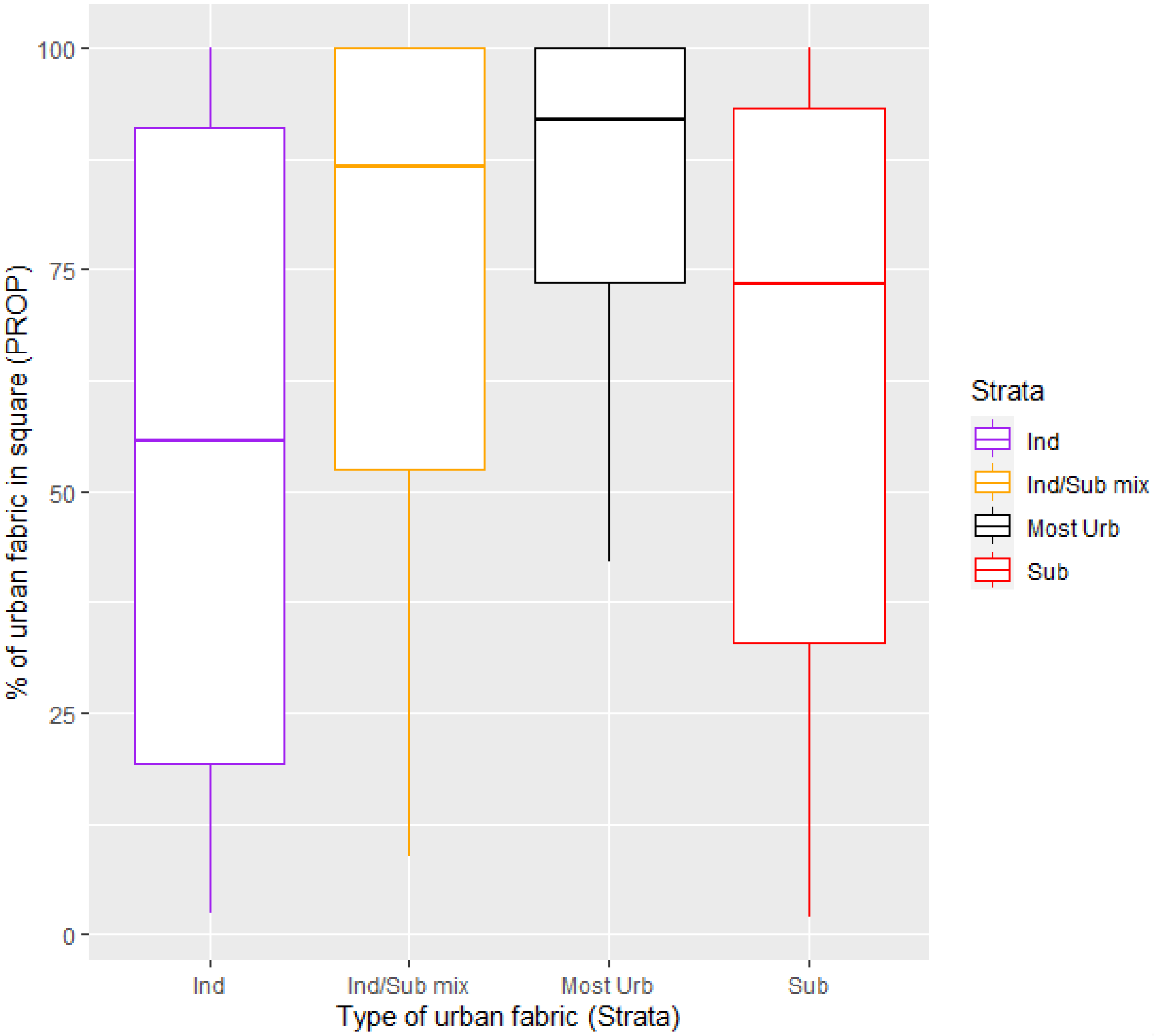

Figure 9. Boxplot of proportion (PROP) of square covered in urban fabric (%) across the four urban strata in the final data set used for analysis.

Glossary

AIC: (Akaike's Information Criterion): measures relative quality of numerous comparative statistical models for a given set of data. Specifically, it measures the amount of information lost, the less information a model loses the better the model represents the process. For a number of different models that are tested, that with the lowest AIC value can be considered to be the model that best fits the data. For models with similar AIC, but added complexity e.g. more explanatory variables or an interaction term, the cut off of a difference of two is used. If the difference is less than 2 the less complex model is favoured.

AON: Apparently Occupied Nest; a well-constructed nest or scrape nest, either containing eggs or young, or capable of holding eggs (possibly attended by an adult) or an apparently incubating adult (Walsh et al., 1995).

AOT: Apparently Occupied Territory; estimated by the spacing of adult gulls or pairs on different rooftops and observations of apparent territorial behaviour, when actual nests cannot be discerned. Any AONs are considered a territory, so the number of AOTs will always be equal to or greater than the number of AONs (Walsh et al., 1995; Woodward et al., 2020).

Bootstrapping: A method to estimate the variance or confidence interval of a predicted value by random resampling with replacement from the observed data.

Confidence limits: a range of plausible values for an unknown parameter (for example, the mean, or in the case of this study the mean estimated AON). The interval has an associated confidence level such that the true parameter value is in the proposed range. All confidence limits shown in this report are 95% confidence limits.

Extrapolation: calculations involved with estimating an entire population from a sample, e.g. a number of observations made within a survey area (e.g. AON) and multiplying this up to predict the total number of AON across a wider geographical area. This method therefore makes the strong assumption that observations within the sample survey area are representative of the wider area.

GLM (Generalised Linear Model): A generalization of ordinary linear regression that allows for response variables that have a distribution other than a normal distribution (e.g. count data which are discrete and bounded by zero). Such models are regularly used with count data.

IND: individual adults; the total number of birds in full adult plumage. Individual birds should only be counted once. Birds in flight only to be counted if it is clear they are using rooftops in the survey square, but birds observed just flying over the survey square are not counted. [Note: Any individual adult can also be categorised as an AON (i.e. an incubating adult) or one or both of a pair attending an AOT, so the IND figure will nearly always be higher than AOT, which itself is equal to or higher than AON. Each single AOT and AON can be represented by one parent bird, both parent birds, or by any number of chicks of that brood].

Negative binomial distribution: A statistical distribution for discrete (bounded by 0) data, such as count data or success and fail data. It's the mean and variance do not need to equal, which can be seen as over-dispersion in other distributions such as the Poisson distribution, a common occurrence in abundance data.

Non-linearity: where relationship between two variables is said to be non-linear, i.e. not 'straight'. There are a number of methods to enable modelling of non-linear data including transforming one or both variables (see log-transformation), using a generalised linear model, or using a generalised additive model.

Overdispersion: if a set of values has a variance larger than that expected under the modelling assumptions applied to that data.

Poisson distribution: often used as a starting point for analysing count data when fitting GLM models (see GLM) and assumes that the mean and variance are equal; however, biological data often fail to meet this assumption. (see also 'overdispersion' and 'negative binomial distribution').

Residual: a measure of how far away a data point is from the estimated regression line of a model.

RMSE: Root Mean Squared Error; a measure of the average absolute difference between the observed data values and the estimated values of a statistical model. In general, when two models using the same dataset are compared, the model with the lower RMSE is considered the better fitting model.

Significant: in statistical terms, a result that is very unlikely to occur given a pre-specified null hypothesis. The significance level used in this report is 95% or 0.05 or a 1 in 20 chance.

Skew: a non-symmetric distribution around the average. If a dataset is positively skewed, values larger than the mode (the most commonly observed value) of the distribution are more common than values lower than the mode.

Zero-inflation: a dataset that contains more zero values than expected under the distribution used to model it (see Poisson distribution and Negative binomial distribution).