

# **Uniwersytet Wrocławski**

## **Wydział Nauk Biologicznych**

**Daniel O'Connell**

290762

### **Praca licencjacka**

**Czy zależności ekologiczne są podobnie szacowane z użyciem danych o występowaniu i liczebności gatunku?**

Are ecological relationships similarly estimated with data on abundance and occurrence of a species?

Praca wykonana pod kierunkiem  
dr Grzegorza Neubauera  
w Pracowni Biologii Lasu  
Instytutu Biologii Środowiskowej

Wrocław, 2019

# Spis treści

1. Wstęp.....	3
2. Materiał i metody.....	5
2.1. Prace terenowe.....	5
2.2. Dane.....	6
2.3. Analiza.....	7
2.3. Modelowanie.....	9
2.3.1 Modele liniowe.....	9
2.3.2. GLM.....	10
2.3.3. Modele hierarchiczne.....	11
2.3.4. Modele mieszanek.....	13
2.3.5. Modele występowania.....	13
2.3.6. Ograniczenia zastosowanych modeli.....	14
2.3.7. Założenia stosowanych modeli.....	15
2.4. Analizowane gatunki.....	18
2.5. Kowariaty (zmienne) siedliskowe.....	19
2.6. Kowariaty (zmienne) opisujące proces obserwacji.....	20
2.7. Forma przedstawienia wyników.....	21
2.7.1 Korelacje.....	21
3. Wyniki.....	23
3.2. Sosnówka <i>Periparus ater</i> .....	23
3.2. Sikora uboga <i>Poecile palustris</i> .....	26
3.4. Muchotłówka żałobna <i>Ficedula hypoleuca</i> .....	29
3.5. Czubatka <i>Lophophanes cristatus</i> .....	32
3.6. Siniak <i>Columba oenas</i> .....	35
3.7. Raniuszek <i>Aegithalos caudatus</i> .....	38
4. Dyskusja.....	41
5. Wnioski.....	44
6. Bibliografia.....	45
7. Abstract.....	47
7.1. Streszczenie.....	47
7.2. English summary.....	48
Załącznik (ryciny 8-13).....	49

## 1. Wstęp

Podstawowym problemem ekologii jest badanie zależności opisujących rozmieszczenie i liczebność populacji organizmów (Krebs 2009, Begon 1986). Rozmieszczenie i liczebność kształtowane są przez zmienne środowiskowe, charakterystyczne dla obszarów zamieszkiwanych (bądź nie) przez populację. Gatunki nie będą występowały na obszarach, na których nie są w stanie przeżyć lub/i się rozmnożyć (co gwarantuje stabilność zasiedlenia). Można także spodziewać się większej liczebności gatunku na obszarach wysokiej jakości (oferujących zasoby w pełni zapewniające przeżycie i rozród), można więc do pewnego stopnia przyjąć, że liczebność jest uniwersalnym wskaźnikiem jakości obszaru dla danego gatunku (Rosenzweig 1991). Z powyższego można założyć, że mając zbiór stanowisk różniących się między sobą zarówno liczebnością populacji, jak i warunkami środowiskowymi, można określić jakie czynniki środowiska mają największy wpływ na jakość obszaru dla danego gatunku. Do opisanie wpływu czynników środowiskowych służą modele statystyczne, a w przypadku tej pracy, modele hierarchiczne.

Termin „modele hierarchiczne” nie jest jednoznaczny nawet w odniesieniu do narzędzi statystycznych używanych w ekologii i można spotkać się z różnymi jego znaczeniami (Gelman i Hill 2007, Kéry i Schaub 2012, Kéry i Royle 2016). W niniejszej pracy zastosowano ten termin do nazwania modeli, składających się z dwóch submodeli, połączonych jedną funkcją wiarygodności: jednej opisującej wykrywalność, i drugiej opisującej liczebność (bądź występowanie). Liczebność lub występowanie modelowane są jako funkcje cech siedliska, takich jak np. wiek drzewostanu, rzeźba terenu, które zwykle się nie zmieniają w trakcie badań (choć modele wykorzystujące zmienne o zróżnicowaniu między latami są

również dostępne), ale są różne między stanowiskami. Wykrywalność natomiast może być w naturalny sposób modelowana jako funkcja kowariat zmiennych w czasie (np. godzina czy data prowadzenia obserwacji, pogoda), które będą się różniły w zależności od tego, kiedy prowadzone są obserwacje (na tym samym stanowisku). Pozwala to na generowanie opartych na wynikach modeli uproszczeń dotyczących prawdziwej (tzn. nieobciążonej niepełną wykrywalnością) liczebności / występowania gatunku, np. „Rudzik występuje częściej w wilgotnych, cienistych i starszych lasach”.

W mojej pracy odpowiedziałem na pytanie czy zależności ekologiczne (w postaci relacji cecha siedliska - występowanie) mogą być skutecznie charakteryzowane z użyciem modeli wykorzystujących binarne dane o obecności gatunku.

Przewidywane prawdopodobieństwo występowania porównałem z liczebnością, przewidywaną w oparciu o inne modele - tzw. modele mieszanek, wykorzystujące dane dokładniejsze - liczebność. Gatunki na których przeprowadziłem analizę i porównanie są ptakami leśnymi, często trudno wykrywalnymi i łatwymi do przeoczenia, dlatego zazwyczaj ciężko jest określić z całkowitą pewnością liczebność populacji lokalnej (zamieszkującej dany obszar). Żeby zminimalizować ten problem, często alternatywnym stosowanym parametrem populacyjnym jest występowanie (ang. *occupancy*) (Hui 2009, Freckleton 2006, Gaston 2000).

Rozpowszechnienie jest zmienną binarną, która określa czy na danym obszarze występuje gatunek czy nie. Warto jednak zauważyć, że jest ono pochodną liczebności. Celem tej pracy jest sprawdzenie, czy można z podobną skutecznością ustalić parametry określające jakość obszarów, korzystając z informacji o rozpowszechnieniu zamiast liczebności.

## **2. Materiał i metody**

Dane do badań pochodzą z projektu pt. *Wpływ prowadzonej gospodarki leśnej na populacje wybranych gatunków ptaków interioru leśnego w lasach nizinnych Polski*, finansowanego przez Dyрекcję Generalną Lasów Państwowych. Projekt został wykonany w latach 2015-2018 przez konsorcjum trzech podmiotów: Muzeum i Instytutu Zoologii PAN, Ogólnopolskiego Towarzystwa Ochrony Ptaków i Instytutu Badawczego Leśnictwa, zawiązane na potrzeby jego realizacji. Dane uzyskane w trakcie badań terenowych projektu są wykorzystane w pracy zgodnie z zapisami umowy między Skarbem Państwa - DG LP a ww. konsorcjum (umowa OR.271.3.12.2015 z dnia 18 maja 2015).

### **2.1. Prace terenowe**

Liczenia ptaków wykonano na 300 powierzchniach próbnych (wydzieleniach LP) w lasach nizinnych Polski. W pierwszym sezonie badań, wiosną 2016, każda powierzchnia została skontrolowana czterokrotnie w okresie 1 kwietnia-31 maja, co około 2 tygodnie. W pracach terenowych wzięło udział 61 obserwatorów ptaków, którzy łącznie wykonali 1200 kontroli terenowych, spędzając w terenie 90 118 minut (około 1501 godzin). Przeciętnie, pojedyncza kontrola na jednej powierzchni trwała około 75 minut, a długość jej trwania była uzależniona przede wszystkim od wielkości powierzchni próbnej (zakres 5-35,6 ha). Przeciętna miara wysiłku terenowego wyrażona czasem poświęconym na spenetrowanie jednostki powierzchni 1 ha wydzielenia wynosiła około 9 minut / 1 ha i była zbliżona do zalecanej w metodzie kartograficznej. Dokładniejszy opis metodyki zawierają sprawozdania z realizacji projektu dostępne w wyszukiwarce tematów badawczych

finansowanych przez DG LP pod adresem [https://tbr.lasy.gov.pl/apex/f?p=102:3::NO::P3\\_TEMAT:3902](https://tbr.lasy.gov.pl/apex/f?p=102:3::NO::P3_TEMAT:3902).

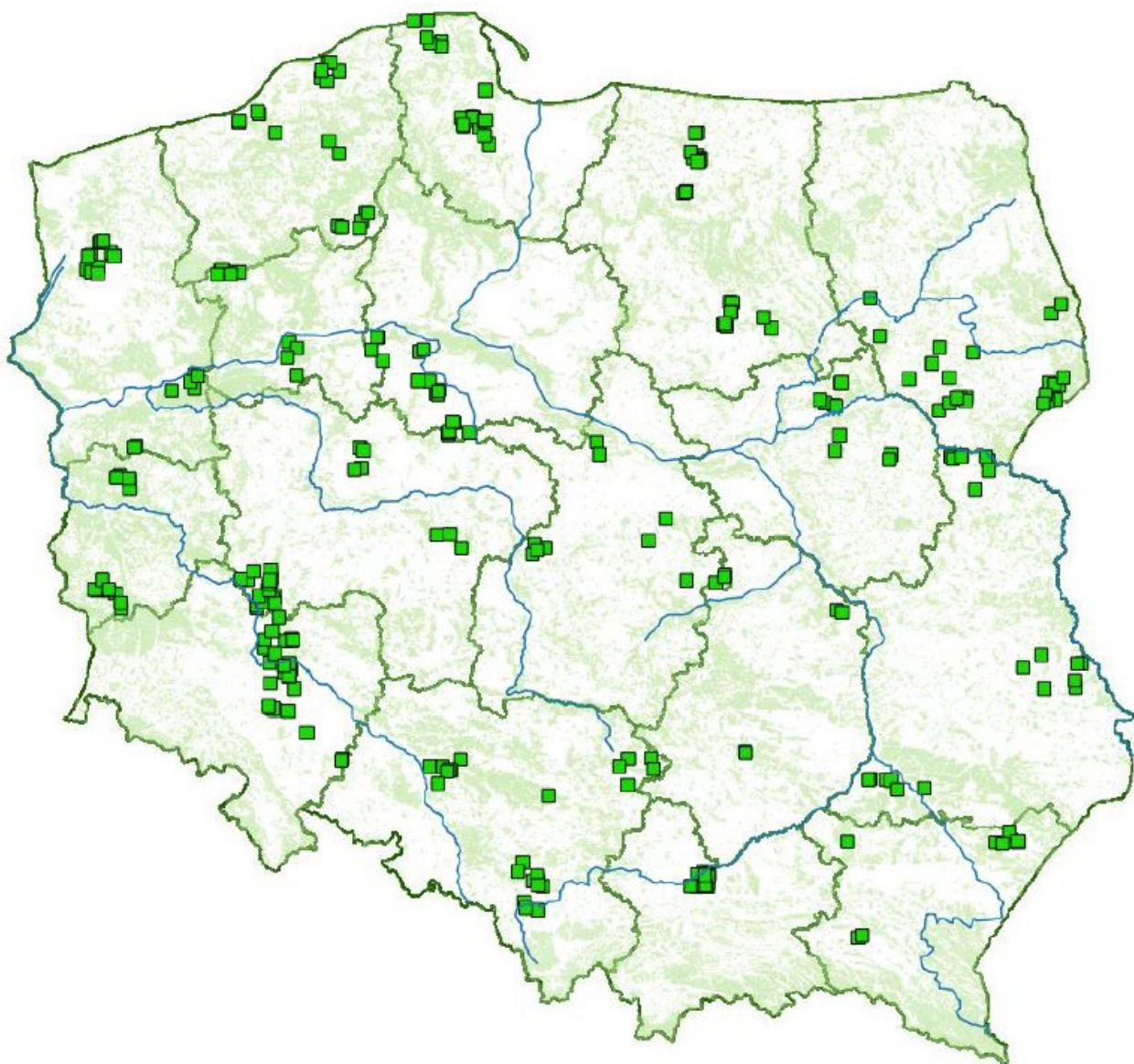
## 2.2. Dane

Dane są zapisane w postaci pliku csv, w którym każdy wiersz zawiera informacje o stanowisku (np. powierzchnia, wilgotność, dominująca roślinność), informacje odnośnie samej obserwacji (np. data obserwacji, czas w terenie) oraz liczbę zaobserwowanych ptaków, w podziale na gatunki i nr obserwacji; w analizach wykorzystałem wyłącznie stwierdzenia śpiewających samców, poza raniuszkiem *Aegithalos caudatus*, dla którego użyto wszystkich osobników (osobniki nieśpiewające, pary itd.). Wiele zmiennych siedliskowych było na tyle szczegółowych, że wносиły niewiele do modelu i zostały pominięte. Niektóre kolumny zostały zagregowane bądź znormalizowane, jak np. WILG (średnia ważona, gdzie jako wartości podano klasy wilgotności, a jako wagi obszar zajęty przez daną klasę) czy wysiłek (czas spędzony na powierzchni podzielony przez jej obszar w ha, dający w efekcie miarę wysiłku terenowego, traktowaną jako kowariatę procesu obserwacji). Dokładniejszy opis wraz z danymi można znaleźć w sprawozdaniach na witrynie wspomnianej powyżej.

W sumie zebrano dane dla 51 gatunków ptaków, jednak do analizy wykorzystano tylko sześć gatunków: dwa względnie pospolite - sosnówka (wykryta na 67% stanowisk) i sikora uboga (60%), dwa średnio pospolite - czubatka (35%) i muchołówka żałoba (35%) oraz dwa rzadkie gatunki - siniak (23%) i raniuszek (15%).

## 2.3. Analiza

Analiza została przeprowadzona z zastosowaniem modeli hierarchicznych. Modele hierarchiczne można zdefiniować jako zbiór prostszych modeli powiązanych z sobą warunkowo (Kéry i Royle 2016). Modele budujące model hierarchiczny, zastosowane w tej pracy, to uogólnione modele liniowe (*generalized linear model* – GLM), korzystające z funkcji logitowej (dla występowania) lub logarytmicznej (dla liczebności) jako funkcji wiążącej (*link function*). Dane zostały przygotowane do analizy i przeanalizowane w środowisku R wersja 3.3.4 (R Core Team 2018) w bibliotece *unmarked* (Fiske, Chandler 2011).



Ryc. 1. Rozmieszczenie powierzchni próbnych, z których pochodzą dane z liczeń ptaków wykonane w ramach projektu i wykorzystane w niniejszej pracy



## 2.3. Modelowanie

### 2.3.1 Modele liniowe

Ogólny model liniowy (*general linear model*) jest rozszerzeniem regresji liniowej. Regresja liniowa zakłada, że zależność pomiędzy zmienną objaśnianą a objaśniającą jest zależnością liniową, którą można wyrazić w postaci równania:

$$Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2} + \dots + B_p X_{ip} + \varepsilon_i \text{ dla } x = 1, \dots, n,$$
 gdzie  $Y_i$  to oczekiwana wartość dla obserwacji  $i$ -tej,  $X_{i1}, X_{i2}, \dots, X_{ip}$  to wartości zmiennych zależnych dla  $i$ -tej obserwacji,  $\varepsilon_i$  to błąd (reszta, różnica między wartością obserwowaną a przewidywaną przez model) dla obserwacji  $i$ -tej, a  $B_0, B_1, \dots, B_p$  to współczynniki określające zależności między poszczególnymi zmiennymi niezależnymi a zmienną zależną. Analiza regresji liniowej ma na celu określenie wartości współczynników regresji dla każdej zmiennej niezależnej, gdzie liniowość polega na tym, że współczynniki są liniowe (występują tylko jako mnożniki zmiennych). Przyjmuje się, że błędy nie są z sobą skorelowane i wykazują wielowymiarowy rozkład normalny.

Regresja liniowa pozwala zbadać procesy stochastyczne, które zwracają wartość skalarną, np. „Ile ptaków zaobserwuje badacz, wiedząc ile czasu spędził w terenie”, które można by (przykładowo) oszacować za pomocą dwóch zmiennych niezależnych: czas pobytu w terenie oraz doświadczenie obserwatora. Zakładając, że współczynniki są prawidłowo oszacowane, można oszacować liczbę zaobserwowanych ptaków, podstawiając do wzoru odpowiednie wartości obu zmiennych niezależnych (predyktorów). Regresję można także zastosować „odwrotnie”: w celu znalezienia modelu, który najlepiej opisuje rzeczywistość – np. odpowiadając na pytanie „Wiedząc ile ptaków zaobserwowali różni badacze, ile czasu spędzili w terenie, jakie mają doświadczenie itp., to jaki zbiór czynników

najlepiej wytłumaczy zmienność w liczbie zaobserwowanych ptaków?”. W tym celu należałoby sporządzić różne modele opisujące dane, a następnie zobaczyć, który model daje najmniejsze różnice między przewidzianą a rzeczywistą liczbą zaobserwowanych ptaków.

### 2.3.2. GLM

Ogólne modele liniowe (*general linear models*) rozszerzają regresję liniową, by można było określić wektor zmiennych zależnych. Intuicyjnie można to rozumieć jako  $m$  osobnych regresji liniowych, wyrażonych jednym wzorem

$$Y_{ij} = B_{0j} + B_{1j}X_{i1} + B_{2j}X_{i2} + \dots + B_{pj}X_{ip} + \varepsilon_{ij}, \text{ gdzie } j = 1, \dots, m.$$

Ograniczeniem ogólnych modeli liniowych jest ich założenie rozkładu normalnego zmiennych zależnych. Funkcje linowe charakteryzują się nieograniczoną dziedziną, tj. mogą przyjmować argumenty z zakresu  $(-\infty; \infty)$ , co się przekłada na podobny zbiór wartości. Jest to problematyczne w przypadku modeli, które w rzeczywistości mają ograniczone zbiory wartości. Dla tej pracy jest to szczególnie ważne, jako że liczebność jest ograniczona do niskich liczb naturalnych, a rozpowszechnienie (rozumiane i modelowane jako prawdopodobieństwo występowania) może przyjmować wartości tylko między 0 lub 1. Rozwiązaniem są uogólnione modele liniowe (*generalized linear model* – GLM) (Nelder 1972), które rozszerzają linowe modele o możliwość opisywania rozkładów innych niż normalny.

GLM opisują liniowy efekt współczynników nie jako spodziewany wynik „wprost”, lecz jako przekształcenie oczekiwanej wartości. Przekształcenie jest dokonywane za pomocą funkcji wiążącej (*link function*), przetwarzającej wartości z rozkładu danych na rozkład normalny. GLM składa się z trzech komponentów (Kéry i Royle 2016):

- Losowy składnik wyniku – statystyczny rozkład  $f$  z parametrami  $\theta$

$$y_i \sim f(\theta)$$

- Predyktor liniowy (*linear predictor*), który jest modelem liniowym

$$\eta_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_p X_{pi}$$

- Funkcja wiązania  $g$ , która opisuje jak średnia  $E(y) = \mu_i$  zależy od predyktora

$$\text{liniowego } g(E(y)) = g(\mu_i) = \eta_i$$

Jako funkcję wiążącą można zastosować dowolną funkcję, która odpowiednio przekształci wartości, jednak dla pewnych rozkładów istnieją już przyjęte i ugruntowane funkcje, znane jako kanoniczne funkcje wiążące. W przypadku tej pracy, rozkłady brane pod uwagę to rozkład Poissona (dla liczebności) oraz rozkład dwumianowy (dla występowania), dla których kanoniczne funkcje to odpowiednio logarytm naturalny oraz funkcja logitowa. Dla rozkładu normalnego, funkcją wiążącą jest funkcja tożsamościowa, czyli  $g(\mu_i) = \mu_i$  co w efekcie daje ogólny model liniowy.

### 2.3.3. Modele hierarchiczne

GLM pozwalają opisać wiele zmiennych, mających różne rozkłady, jednak zakładają, że poszczególne obserwacje są od siebie niezależne. To założenie często nie jest prawdą, np. można się spodziewać, że różne obserwacje na tej samej powierzchni będą bardziej podobne do siebie niż obserwacje między różnymi powierzchniami. Różnice między powierzchniami będą wynikały z różnic w opisywanym procesie ekologicznym, odpowiedzialnym za rzeczywisty stan populacji. Przy krótkich odstępach między liczeniami (np. wykonywanymi w ciągu jednej wiosny, jak w przypadku danych używanych w niniejszej pracy) można przyjąć, że stan populacji jest stały, zatem różnice w liczebności/występowaniu

osobników na tej samej powierzchni będą wynikały z różnic w wykrywalności, a nie ze zmiany rzeczywistego stanu populacji. Niepełną wykrywalność można zdefiniować jako prawdopodobieństwo wykrycia osobnika przy wizycie na powierzchni badawczej mniejsze niż 1. Skoro prawdopodobieństwo wykrycia osobnika nie jest równe 1 (100%), można to także przełożyć na liczebność i stwierdzić, że liczba zaobserwowanych osobników będzie mniejsza lub równa rzeczywistej liczebności populacji. Z tego wynika, że dane (występowanie bądź liczebność) zebrane w terenie będą miały niższe wartości niż powinny mieć. Sposobem na zwiększenie prawdopodobieństwa otrzymania prawdziwej wartości występowania czy liczebności są wielokrotne kontrole na każdej powierzchni, które nie są bardzo oddalone w czasie i dla których można przyjąć założenie o zamknięciu populacji w danym okresie.

Model opisujący populację można zatem rozbić na dwie warstwy. Jedna z nich opisuje rzeczywisty proces ekologiczny jako funkcję cech siedliska, które nie będą się różniły między obserwacjami (np. wiek drzewostanu, rzeźba terenu prawdopodobnie nie zmienia się istotnie w ciągu jednej wiosny), a druga modeluje wykrywalność jako funkcję cech (kowariat) zmiennych w czasie, które mogą być różne między obserwacjami (np. godzina czy data obserwacji, pogoda). Tym właśnie są modele hierarchiczne, pozwalające rozdzielić zależności ekologiczne (określające rzeczywisty stan populacji) od zaburzeń wynikających z niedoskonałości obserwacji.

### 2.3.4. Modele mieszanek

Do modelowania liczebności zastosowano hierarchiczne modele mieszanek (*N-mixture models, binomial mixture models*) (Royle 2004). Modele te korzystają z wyników zliczeń zreplikowanych w czasie i przestrzeni, tj. każda powierzchnia badawcza została skontrolowana co najmniej dwukrotnie podczas badań. W tych modelach zaobserwowana liczebność ( $\lambda$ ) jest parametrem opisującym stan populacji i jest modelowana za pomocą rozkładu Poissona, natomiast proces obserwacji jest opisany za pomocą prawdopodobieństwa wykrycia ( $p$ ) i modelowany rozkładem dwumianowym. Matematycznie można to zapisać jako:

$$\begin{aligned} N_i &\sim \text{Poisson}(\lambda_i) \\ y_{ij} | N_i &\sim \text{Binomial}(N_i, p) \\ \text{log}(\lambda_i) &= B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_n X_n \end{aligned}$$

gdzie  $N_i$  oznacza rzeczywistą liczebność na powierzchni  $i$ ,  $y_{ij}$  oznacza liczbę osobników wykrytych na powierzchni  $i$  podczas kontroli  $j$ , a  $p$  to prawdopodobieństwo wykrycia osobnika.  $B_i$  oraz  $X_{ij}$  to odpowiednio współczynniki i zmienne niezależne w regresji (GLM) opisującej rzeczywisty stan populacji.

Modele mieszanek można uznać za bardziej dokładne niż modele występowania, ponieważ korzystają z danych o liczebności, które zawierają więcej informacji niż samo występowanie (Kéry i Royle 2016).

### 2.3.5. Modele występowania

Obecność (rozpowszechnienie, występowanie) była modelowana za pomocą modeli występowania (*site-occupancy models*) (Tyre 2003, MacKenzie 2004). Podobnie jak modele mieszanek, modele występowania korzystają z wyników zliczeń zreplikowanych w czasie i przestrzeni, a proces obserwacji jest opisany za pomocą prawdopodobieństwa wykrycia  $p$ , przy czym, w odróżnieniu od modeli

mieszanek, jest to prawdopodobieństwo wykrycia gatunku na powierzchni, nie prawdopodobieństwo wykrycia osobnika. W modelach występowania parametrem opisującym populację jest prawdopodobieństwo wystąpienia  $\Psi$  opisane za pomocą rozkładu Bernoulliego. Matematycznie można zapisać jako:

$$\begin{aligned} z_i &\sim \text{Bernoulli}(\psi) \\ y_{ij} | z_i &\sim \text{Binomial}(z_i, p) \\ \text{logit}(\psi_i) &= B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_n X_i \end{aligned}$$

gdzie  $z_i$  oznacza rzeczywiste występowanie gatunku na powierzchni  $i$ ,  $y_{ij}$  określa czy wykryto gatunki na powierzchni  $i$  podczas kontroli  $j$ , a  $p$  to prawdopodobieństwo wykrycia gatunku.  $B_i$  oraz  $X_{ij}$  to odpowiednio współczynniki i zmienne niezależne GLM opisujące rzeczywisty stan populacji.

### 2.3.6. Ograniczenia zastosowanych modeli

Na liczebność i występowanie populacji będą miały wpływ także czynniki takie jak zależności międzygatunkowe (konkurencja, drapieżnictwo), które są dynamiczne i potencjalnie bardzo skomplikowane. Prostym przykładem jest zależność między wybiórczością osobników a zagęszczeniem. Gdy zagęszczenie jest niskie, wiele gatunków wykazuje zdecydowane preferencje siedliskowe, natomiast gdy zwiększa się zagęszczenie, preferencje te zanikają, co prowadzi do zajmowania siedlisk nieoptymalnych (Rosenzweig 1991). Konkurencja międzygatunkowa także może mieć wpływ, ponieważ wiele gatunków będzie unikało powierzchni zajętych przez gatunki zajmujące podobne nisze ekologiczne (Jones 2014, Grether 2013) lub obszary o zwiększonej aktywności drapieżników (Pintar 2018, Liebezeit 2002, Resutarits i Royle 2018). Są także przypadki, gdy pewne gatunki preferują stanowiska zasiedlone przez inne gatunki, np. jako rodzaj ochrony przed

drapieżnikami (Eriksson 1982). Ze względu na poziom komplikacji, zależności te zostały pominięte w tej pracy.

Zastosowane modele zakładają, że parametry przestrzenne nie są dynamiczne (np. typ lasu, rodzaj podłoża), co jest uproszczeniem, jednakże w skali czasowej badań (ok. 8 tygodni) można przyjąć, że założenia te są spełnione. Kolejnym uproszczeniem modeli jest założenie, że populacja jest zamknięta (brak migracji, rozrodu czy śmiertelności). Jeżeli badania są przeprowadzone w dość krótkim okresie, podczas lęgów, gdy ptaki nie migrują i są osiadłe w swoich terytoriach oraz jeżeli gatunek nie wykazuje się wysoką śmiertelnością, można przyjąć, że rzeczywista populacja będzie względnie stała. Większym problemem, w przypadku ruchliwych gatunków, jest niedopasowanie powierzchni badawczej do terytorium ptaka. Wtedy podczas kontroli ptak może być znacznie oddalony od badacza, w wyniku czego zostanie pominięty. Zjawisko to jest znane jako czasowa emigracja (*temporary emigration*), w wyniku czego populacja jest permanentnie otwarta w sensie geograficznym (Neubauer 2016, Resetarits i Royle 2018). By uwzględnić to zjawisko, modele można rozszerzyć o kolejną warstwę odpowiadającą za i szacującą tzw. osiągalność lub dostępność do wykrycia (*availability*). Ze względów złożoności, czasowa emigracja nie była brana pod uwagę w tej pracy.

### **2.3.7. Założenia stosowanych modeli**

Zastosowane modele hierarchiczne mają pewne ogólne założenia co do danych (Kéry 2016), mianowicie:

1. Założenie zamknięcia populacji (*closure assumption*) (Royle 2004) zakłada brak zmian liczebności w wyniku migracji, rozrodu oraz śmiertelności podczas badań. Ponieważ badania prowadzono w okresie dwóch miesięcy

wiosennych, większość ptaków (włączając gatunki objęte analizą) przebywa w tym okresie na swoich terytoriach i nie migruje. Jednak geograficzne (przestrzenne) zamknięcie populacji nie będzie całkowicie spełnione, szczególnie, że badane powierzchnie próbne nie pokrywają się ściśle z terytorium ptaków: część terytoriów znajduje się poza granicami powierzchni, ale osobniki z tych terytoriów mogą (okresowo) przebywać w granicach powierzchni i zostać wykryte przez obserwatora, chociaż centra ich aktywności i nawet większość terytorium mogą leżeć poza powierzchnią. Dlatego oszacowania z modeli mieszanek odnoszą się zazwyczaj do powierzchni nieco większej niż faktycznie objęta kontrolami terenowymi. Powierzchnia ta jest nieznana, co uniemożliwia wyliczenie zagęszczenia. Sposobem na uwzględnienie tego faktu może być zastosowanie innego modelu o dodatkowej warstwie (dodatkowym submodelu; jest to więc model z trzema warstwami) odpowiadającym za i szacującym tzw. osiągalność lub dostępność do wykrycia (*availability*), która jest modelowana jako dopełnienie tzw. okresowej emigracji (*temporary emigration*) do jedyńki (Chandler i in. 2011). Wielkość ta odzwierciedla prawdopodobieństwo, że osobnik jest „osiągalny” lub „dostępny” – czyli, że przebywa w części powierzchni objętej kontrolą i np. śpiewa (jest to również produkt kilku składowych, których dekompozycja w jednym modelu powiodła się dopiero wiosną 2019 r. – patrz Hostetter i in. 2019). Prawdopodobieństwo to można też interpretować jako przeciętną proporcję terytorium, znajdującą się w granicach powierzchni badawczej. Model ten wymaga danych zebranych w protokole *robust design*, gdzie zezwala się na otwarcie populacji w okresach pomiędzy powtórzonymi jej pomiarami.



2. Brak błędów typu *false-positives* (wykrycie gatunku / osobnika w sytuacji kiedy faktycznie on tam nie występuje). Błąd ten pojawia się, gdy osobnik zostanie źle zidentyfikowany lub policzony wielokrotnie. Jest to szczególnie problematyczne w przypadku modeli występowania (Royle i Link 2006). W celach minimalizacji tego typu błędów, można odrzucić wszystkie wątpliwe obserwacje. Częstotliwość błędów tego typu w dużej mierze zależy od doświadczenia obserwatora – mniej doświadczony badacz częściej pomyli gatunki lub policzy te same osobniki jako kolejne.
3. Niezależność wykrycia zakłada, że wykrycia osobników są niezależne. Oznacza to, że prawdopodobieństwo występowania osobników na jakimś obszarze jest niezależne od występowania osobników na innym obszarze. Ponadto zakłada się, że prawdopodobieństwo wykrycia osobników na tym samym obszarze, ale w różnych terminach, nie jest od siebie zależne (tzn. fakt wykrycia osobnika / gatunku podczas pierwszej kontroli nie ma wpływu na fakt wykrycia / braku wykrycia osobnika / gatunku w trakcie drugiej kontroli).
4. Jednakowe prawdopodobieństwo wykrycia osobników wew. powierzchni. Modele te próbują wytłumaczyć różnice w wykrywalności osobników między powierzchniami jako wypadkową warunków środowiskowych. Zatem osobnicze różnice w wykrywalności są całkiem pomijane, co może mieć istotny wpływ, jeżeli różnice te będą jednak istotne. To założenie jest również problematyczne, ponieważ wykrywalność często spada wraz z odległością od obserwatora, co częściowo można skorygować przez penetrację każdej części powierzchni próbnej z mniej-więcej jednakowym wysiłkiem (mierzonym czasem poświęconym na skontrolowanie 1 ha).

5. Założenie parametryczności modeli. Modele hierarchiczne próbują wytłumaczyć zmienność w odpowiedzi za pomocą zbioru parametrów dla różnych powierzchni. Z tego wynika, że źle dobrane parametry, dane mające losowy rozkład lub skomplikowane zależności między parametrami spowodują, że modele będą słabo opisywały dane wejściowe.

## 2.4. Analizowane gatunki

Do analizy wybrano 6 przykładowych gatunków ptaków, gdzie kryterium wyboru była ich frekwencja, wyrażona jako procent powierzchni, na których stwierdzono dany gatunek. Utworzono trzy podgrupy – pospolite (występowanie na 60–70% powierzchni), średnio pospolite (45–60% powierzchni) oraz rzadkie (30–45% powierzchni). Dolna granica została ustanowiona na 30%, ponieważ w przypadku ptaków jeszcze rzadszych, liczebność na badanych powierzchniach wynosiła zero lub jeden, co czyni liczebność nieodróżnialną od występowania. Jako pospolite gatunki wybrano sosnowkę *Periparus ater* (wykryta na 67% stanowisk) i sikorę ubogą *Poecile palustris* (60%), jako średnio pospolite czubatkę *Lophophanes cristatus* (35%) i muchołówkę żałobną *Ficedula hypoleuca* (35%), a jako rzadkie siniaka *Columba oenas* (23%) oraz raniuszka *Aegithalos caudatus* (14%).

Dla każdego gatunku dopasowano kilka modeli mieszanek i wybrano najlepszy wg kryterium informacyjnego Akaike (AIC, (Burnham i Anderson 2002)), które szacuje jakość każdego modelu względem innych modeli. Wartość AIC jest miarą utraty informacji, więc im niższa wartość, tym lepiej model odzwierciedla rzeczywistość. Miara AIC bierze pod uwagę także liczbę parametrów – dla dwóch

modeli o takiej samej mocy przewidywania, model z większą liczbą parametrów będzie miał wyższą wartość AIC.

Gdy już dopasowano modele mieszanek (dla liczebności), skonstruowano model występowania na podstawie parametrów z najlepszego modelu mieszanek i porównano jak dobrze opisuje dane.

We wszystkich dopasowanych modelach mieszanek i modelach występowania, submodel dla wykrywalności jest identyczny i zawiera dwie kowariaty (patrz niżej), co umożliwia w miarę plastyczne zamodelowanie prawdopodobieństwa wykrycia, a jednocześnie ogranicza liczbę koniecznych obliczeń i porównań.

## **2.5. Kowariaty (zmienne) siedliskowe**

Jako że model statystyczny ma odzwierciedlać prawdziwy proces ekologiczny, dobór kowariat jest kwestią kluczową. Z jednej strony nie może być ich za mało, bo np. model, który bierze pod uwagę tylko powierzchnię, będzie odzwierciedlał rzeczywistość (bo będzie więcej ptaków w całej Polsce niż tylko na Giewoncie), ale słabo tłumaczy zależności ekologiczne. Z drugiej strony nie należy wprowadzać zbyt dużej liczby parametrów, aby nie wpaść w pułapkę nadmiernego dopasowania (*overfitting*) modelu – wtedy model będzie bardzo dobrze odzwierciedlał konkretne badane dane, ale tylko ten konkretny zestaw i wszelkie generalizacje na jego podstawie będą wątpliwe.

Podobnie należy uważać z samymi parametrami. Należy dobrać je by miały sens z punktu widzenia biologicznego.

Przy tworzeniu modeli wybrano zmienne z tabeli 1, jako potencjalnie mające największy wpływ na liczebność i występowanie analizowanych gatunków ptaków.

Tabela 1. Zmienne opisujące powierzchnie próbne (*site covariates*) wykorzystane jako predyktory występowania lub liczebności ptaków w modelach hierarchicznych użytych w niniejszej pracy.

Nazwa zmiennej (skrót)	Opis
AREA	Powierzchnia w hektarach; zmienna ciągła (ha)
WIEK	Średni wieku drzewostanu; zmienna ciągła (lata)
gat.pan	Panujący gatunek drzewa; czynnik ze zbioru {JW, BRZ, JS, SO, DB, ŚW, GB, OL, BK}
gat.pan.wiek	Wiek panującego gatunku drzewa; zmienna ciągła (lata)
pLISd	Proporcja udział drzew liściastych w drzewostanie; zmienna ciągła <0; 1>
pLISp	Proporcja udział drzew liściastych w podszycie; zmienna ciągła <0; 1>
pIGLd	Proporcja udział drzew iglastych w drzewostanie; zmienna ciągła <0; 1>
pIGLp	Proporcja udział drzew iglastych w podszycie; zmienna ciągła <0; 1>
WILG	Średnia ważona powierzchnią klasa wilgotności powierzchni próbnej; zmienna ciągła <0; 1>
d_richness	Liczba gatunków drzew w warstwie drzewostanu; liczba naturalna (dyskretna) <0; 13>
p_richness	Liczba gatunków drzew i krzewów w warstwie podszytu; liczba naturalna (dyskretna) <0; 7>
forest_type_overall	Typ lasu; czynnik z zbioru {bór, łęg, grąd, ols}
Rich16	Liczba gatunków ptaków stwierdzonych na powierzchni w 2016; liczba naturalna (dyskretna) <5; 47>
X84	Współrzędna E; zmienna ciągła <14,77; 23,72>
Y84	Współrzędna N; zmienna ciągła <49,27; 54,79>

## 2.6. Kowariaty (zmienne) opisujące proces obserwacji

Jako zmienne opisujące proces obserwacji wykorzystano tylko datę wykonania kontroli (data juliańska, zmienna day) i wysiłek obserwatora (czas na 1 ha w min; zmienna effort). Zmienne te były wykorzystywane przez wszystkie modele.

## 2.7. Forma przedstawienia wyników

Dla każdego ptaka przedstawiono tabelę z 10 najlepszymi modelami, posortowane wg ich wartości AIC (im niższa wartość, tym lepszy model - pierwszy ma najwyższe poparcie). Następnie dla najlepszego modelu przedstawiono wkład poszczególnych współczynników, a także pokazano jak wygląda rozkład współczynników dla modelu występowania, który skonstruowano z tymi samymi współczynnikami. Poniżej tabel znajdują się wykresy przedstawiające korelacje między prognozami pochodzącymi z modeli. W załączniku zamieszczono wykresy (ryciny 8-13) przedstawiające zbieżność przewidywań (oczekiwanej liczebności lub oczekiwanego prawdopodobieństwa występowania) obu modeli (po lewej stronie dla modelu liczebności, po prawej dla modelu występowania), aby pokazać jak dokładnie poszczególne współczynniki opisują zmienność między stanowiskami, a także jak bardzo modele liczebności i występowania różnią się między sobą.

### 2.7.1 Korelacje

Dla każdego gatunku utworzono cztery wykresy korelacji. Na każdym wykresie, w tytule, podano wartość współczynnika kierunkowego funkcji korelacji, obliczonej za pomocą funkcji **cor.test**. Czerwoną linią pokazano funkcję  $f(x)=x$  by uwidocznić jakość korelacji (jeżeli występowanie odpowiada liczebności, to funkcja korelacji powinna być równa funkcji tożsamościowej). Korelacje przedstawiono jako:

- a) korelacja między wartościami oszacowanych współczynników modeli liczebności (oś X) i modeli występowania (oś Y).
- b) korelacja między prognozowaną liczebnością wg modelu liczebności a prognozowanym występowaniem na podstawie modelu występowania. W

przypadku dużych liczebności korelacja ta jest funkcją logarytmiczną, zatem dla tego wykresu pominięto wykres funkcji tożsamościowej.

c) korelacja między prognozowaną liczebnością wg modelu liczebności a prognozowanym występowaniem na podstawie modelu występowania, na skali funkcji wiążącej. Prognozowana liczebność została przekształcona funkcją logarytmiczną, natomiast prognozowane występowanie funkcją logitową.

d) za pomocą funkcji **lambda2psi**, z biblioteki *unmarked*, zamieniono prognozowane liczebności z modelu mieszanek na prognozy występowania, a następnie te wartości porównano z prognozowanym występowaniem na podstawie modelu występowania

Tabela 2 przedstawia wartości poszczególnych rodzajów korelacji. Ptaki są posortowane wg częstotliwości występowania – częstość występowania to procent powierzchni badawczych, na których zaobserwowano dany gatunek. W przypadku funkcji logarytmicznej, prognozowane występowanie przekształcono funkcją logitową by porównać prognozy przekształcone funkcją wiążącą.

Tabela 2. Korelacje prognozowanych liczebności do prognozowanego występowania, gdzie prognozowana liczebność przekształcona odpowiednią funkcją. W kolumnie „Częstość występowania” podano rozpowszechnienie (% powierzchni badawczych ze stwierdzonym gatunkiem). W przypadku funkcji logarytmicznej, prognozowane występowanie przekształcono funkcją logitową.

Ptak	Częstość występowania	Funkcja		
		Tożsamościowa	Logarytmiczna	<b>lambda2psi</b>
<i>Periparus ater</i>	67%	0.663	0.907	0.898
<i>Poecile palustris</i>	60%	0.699	0.967	0.954
<i>Ficedula hypoleuca</i>	35%	0.915	0.988	0.983
<i>Lophophanes cristatus</i>	35%	0.795	0.941	0.921
<i>Columba oenas</i>	23%	0.918	0.984	0.980
<i>Aegithalos caudatus</i>	15%	0.970	0.986	0.980

### 3. Wyniki

#### 3.2. Sosnówka *Periparus ater*

Sosnówka była często obserwowana (67% badanych powierzchni). Model, który najlepiej opisuje dane to (~p wykrycia, ~liczebność): ~day+effort, ~x84\*y84 + AREA + WIEK + pLISd + pLISp + pIGLp + d\_richness + p\_richness + Rich16 (AIC = 1957.3). Porównanie dopasowania różnych modeli przedstawiono w Tabeli 2, a współczynniki z najlepszego modelu - w Tabeli 3. W Tabeli 4 przedstawiono współczynniki parametrów modelu występowania odpowiadającego najlepszemu modelowi mieszanek. Rycina 2 przedstawia korelacje między prognozami pochodzących z modeli, natomiast Rycina 8 (załącznik) przedstawia zbieżność poszczególnych współczynników najlepszego modelu.

Tabela 3. Modele liczebności dopasowane do danych z liczeń sosnówki. Dla każdego modelu podano liczbę parametrów siedliskowych w modelu (k), wartość kryterium informacyjnego Akaike (AIC), różnicę między modelem o największym poparciu a danym modelem ( $\Delta AIC$ ) oraz wagę modelu ( $\omega AIC$ ).

Model		k	AIC	$\Delta AIC$	$\omega AIC$
Liczebność	wykrywalność				
1 x84*y84 + AREA + WIEK + pLISd + pLISp + pIGLp + d_richness + p_richness + Rich16	day + effort	9	1957.32	0.00	0.994
2 x84*y84 + AREA + WIEK + pLISd + d_richness + p_richness + Rich16	day + effort	7	1967.47	10.15	0.006
3 x84*y84 + AREA + WIEK + pIGLd	day + effort	4	1985.72	28.40	0.0
4 x84*y84 + AREA + WIEK + pLISd	day + effort	4	1987.23	29.91	0.0
5 x84*y84 + AREA + WIEK + dstan_igl	day + effort	4	1990.89	33.58	0.0
6 x84*y84 + AREA + WIEK + dstan_lisc	day + effort	4	1993.49	36.17	0.0
7 x84*y84 + AREA + WIEK + WILG + forest_type_overall	day + effort	5	2049.83	92.51	0.0
8 x84*y84 + AREA + WIEK + zwarcie_sdi	day + effort	4	2065.28	107.97	0.0
9 x84*y84 + AREA + WIEK + pIGLp	day + effort	4	2065.45	108.13	0.0
10 x84*y84 + AREA + WIEK + podszyt_igl	day + effort	4	2071.46	114.15	0.0

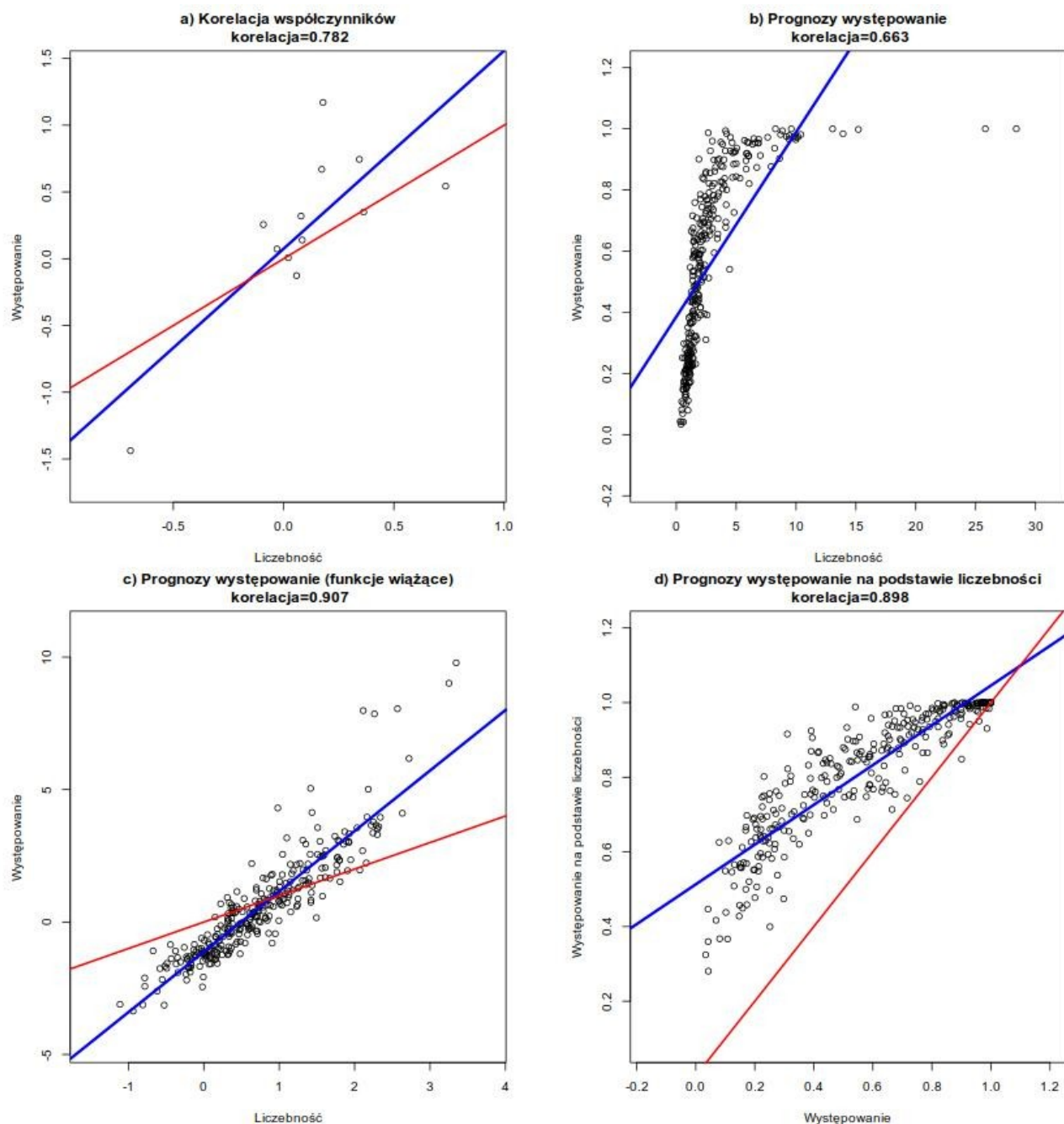
Tabela 4. Współczynniki parametrów najlepszego modelu mieszanek (submodel z rozkładem Poissona, skala logarytmiczna), objaśniającego zmienność liczebności sosnówki. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	0.065	0.058	1.118	0.2630
y84	0.291	0.058	5.067	0.0000
AREA	0.115	0.055	2.099	0.0358
WIEK	0.085	0.069	1.240	0.2150
pLISd	-0.627	0.063	-10.030	0.0000
pLISp	-0.050	0.079	-0.631	0.5280
pIGLp	0.127	0.042	3.041	0.0024
d_richness	-0.051	0.059	-0.860	0.3900
p_richness	0.063	0.074	0.854	0.3930
Rich16	0.048	0.010	4.739	0.0000
x84:y84	-0.043	0.072	-0.599	0.5490

Tabela 5. Współczynniki parametrów modelu występowania (submodel z rozkładem dwumianowym, skala logitowa) odpowiadającego najlepszemu modelowi mieszanek, objaśniającemu zmienność liczebności sosnówki. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	0.669	0.195	3.438	0.0006
y84	0.744	0.171	4.354	0.0000
AREA	0.140	0.164	0.852	0.3940
WIEK	0.320	0.183	1.750	0.0801
pLISd	-1.438	0.249	-5.776	0.0000
pLISp	0.256	0.218	1.175	0.2400
pIGLp	1.170	0.448	2.613	0.0090
d_richness	0.073	0.163	0.448	0.6540
p_richness	-0.126	0.212	-0.597	0.5500
Rich16	0.053	0.029	1.845	0.0651
x84:y84	0.007	0.198	0.037	0.9700





Ryc 2. Korelacje a) oszacowań współczynników najlepszego modelu liczebności i odpowiadającego mu modelu występowania, b) prognozowanych liczebności i prognozowanego występowania – skala liniowa, c) prognozowanych liczebności i prognozowanego występowania – przekształcone funkcjami wiązania (odpowiednio logarytmiczna i logitowa), d) relacja prognozowanego wystąpienia do prognozowanej liczebności zamieniona na występowanie za pomocą funkcji  $\lambda 2\psi$  dla sosnowki. Czerwoną linią przedstawiono funkcję  $f(x) = x$ .

### 3.2. Sikora uboga *Poecile palustris*

Sikora uboga była względnie często obserwowana (60% powierzchni badanych).

Model, który najlepiej opisuje jej liczebność to ~day+effort, x84\*y84 + AREA + WIEK + pLISd + pLISp + pIGLp + d\_richness + p\_richness + Rich16 (AIC = 1090.8).

Porównanie dopasowania różnych modeli przedstawiono w tabeli 6, a dla najlepszego modelu widać wpływ poszczególnych parametrów w tabeli 7. W tabeli 8 przedstawiono współczynniki parametrów modelu występowania odpowiadającego najlepszemu modelowi mieszanek. Rycina 3 przedstawia korelacje między prognozami pochodzących z modeli, natomiast Rycina 9 (załącznik) przedstawia zbieżność poszczególnych współczynników najlepszego modelu.

Tabela 6. Modele liczebności dopasowane do danych z liczeń sikory ubogiej. Dla każdego modelu podano liczbę parametrów siedliskowych w modelu (k), wartość kryterium informacyjnego Akaike (AIC), różnicę między modelem o największym poparciu a danym modelem ( $\Delta AIC$ ) oraz wagę modelu ( $\omega AIC$ ).

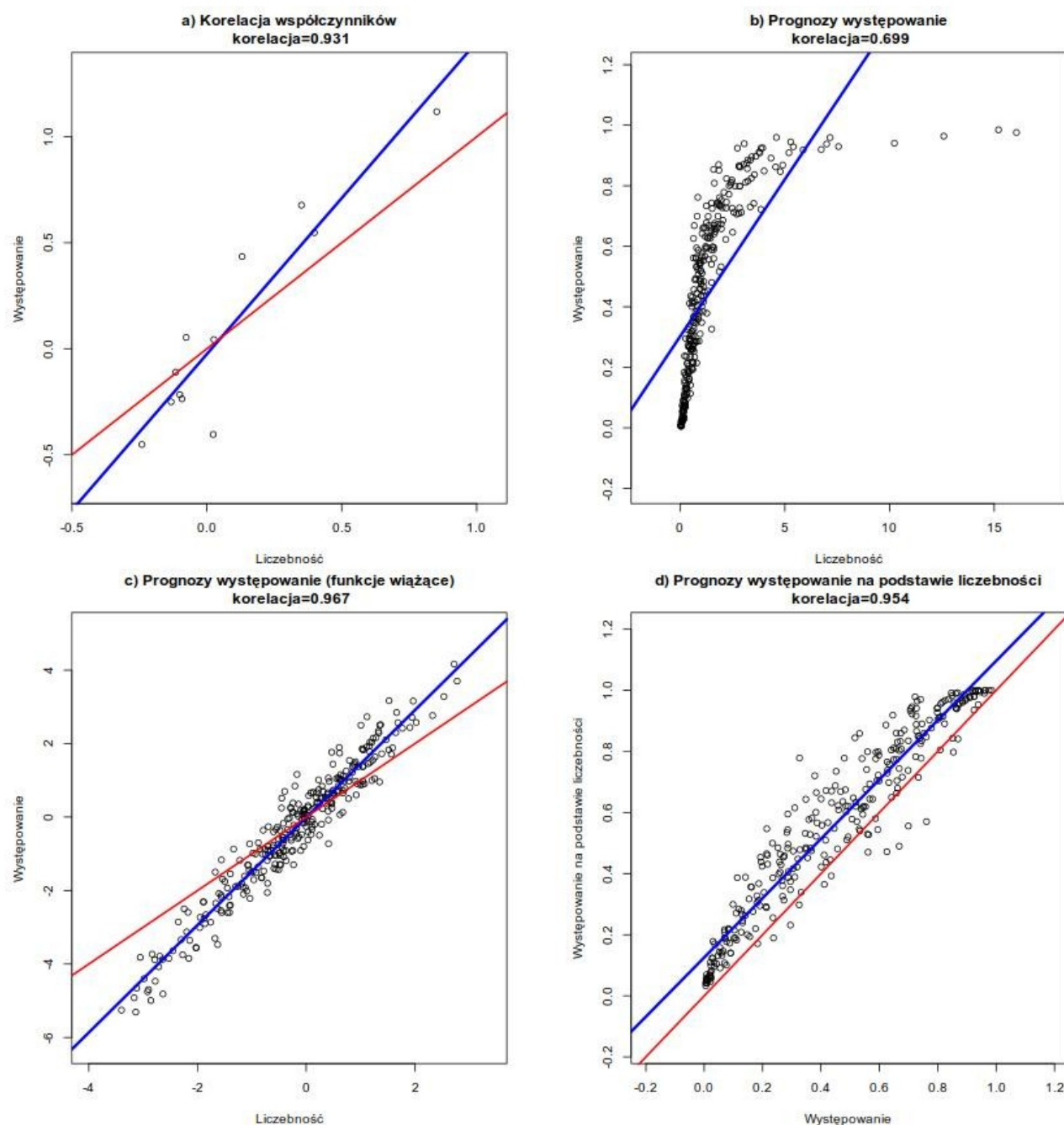
Model		k	AIC	$\Delta AIC$	$\omega AIC$
Liczebność	wykrywalność				
1 x84*y84 + AREA + WIEK + pLISd + pLISp + pIGLp + d_richness + p_richness + Rich16	day + effort	9	1090.85	0.0	0.999
2 x84*y84 + AREA + WIEK + pLISd + d_richness + p_richness + Rich16	day + effort	7	1105.82	14.98	0.001
3 x84*y84 + AREA + WIEK + pLISp	day + effort	4	1199.74	108.89	0.0
4 x84*y84 + AREA + WIEK + WILG + forest_type_overall	day + effort	5	1204.61	113.76	0.0
5 x84*y84 + AREA + WIEK + podszyt_lisc	day + effort	4	1211.16	120.31	0.0
6 x84*y84 + AREA + WIEK + WILG	day + effort	4	1211.7	120.85	0.0
7 x84*y84 + AREA + WIEK + dstan_lisc	day + effort	4	1228.64	137.8	0.0
8 x84*y84 + AREA + WIEK + pLISd	day + effort	4	1229.88	139.03	0.0
9 x84*y84 + AREA + WIEK + dstan_igl	day + effort	4	1230.52	139.67	0.0
10 x84*y84 + AREA + WIEK + p_richness	day + effort	4	1231.27	140.42	0.0

Tabela 7. Współczynniki parametrów najlepszego modelu mieszanek (submodel z rozkładem Poissona, skala logarytmiczna), objaśniającego zmienność liczebności sikory ubogiej. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	-0.123	0.078	-1.572	0.1160
y84	0.028	0.089	0.316	0.7520
AREA	-0.095	0.073	-1.298	0.1940
WIEK	-0.095	0.092	-1.029	0.3030
pLISd	0.329	0.095	3.477	0.0005
pLISp	0.402	0.105	3.818	0.0001
pIGLp	-0.118	0.165	-0.717	0.4730
d_richness	-0.157	0.083	-1.882	0.0598
p_richness	0.146	0.107	1.356	0.1750
Rich16	0.131	0.014	9.115	0.0000
x84:y84	0.037	0.098	0.376	0.7070

Tabela 8. Współczynniki parametrów modelu występowania (submodel z rozkładem dwumianowym, skala logitowa) odpowiadającego najlepszemu modelowi mieszanek, objaśniającego zmienność liczebności sikory ubogiej. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	0.053	0.204	0.262	0.7930
y84	0.043	0.179	0.241	0.8090
AREA	-0.111	0.186	-0.595	0.5520
WIEK	-0.217	0.202	-1.073	0.2830
pLISd	0.677	0.222	3.047	0.0023
pLISp	0.548	0.275	1.991	0.0465
pIGLp	-0.250	0.244	-1.027	0.3050
d_richness	-0.236	0.195	-1.210	0.2260
p_richness	0.435	0.261	1.666	0.0958
Rich16	0.170	0.041	4.131	0.0000
x84:y84	-0.404	0.244	-1.653	0.0984



Ryc 3. Korelacje a) oszacowań współczynników najlepszego modelu liczebności i odpowiadający model występowania, b) prognozowanych liczebności i prognozowanego występowania – skala liniowa, c) prognozowanych liczebności i prognozowanego występowania – przekształcone funkcjami wiążącymi (odpowiednio logarytmiczna i logitowa), d) relacja prognozowanego występowania do prognozowanej liczebności zamieniona na występowanie za pomocą funkcji  $\lambda_{2\psi}$  dla sikory ubogiej. Czerwoną linią przedstawiono funkcję  $f(x) = x$ .

### 3.4. Muchołówka żałobna *Ficedula hypoleuca*

Muchołówka żałobna była względnie rzadko obserwowana (35% powierzchni badanych). Model, który najlepiej opisuje jej liczebność to  $\sim \text{day} + \text{effort}$ ,  $\sim x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{podszyt\_lisc}$  (AIC = 593.1). Porównanie dopasowania różnych modeli przedstawiono w tabeli 9, a współczynniki z najlepszego modelu w tabeli 10. W tabeli 11 przedstawiono współczynniki parametrów modelu występowania odpowiadającego najlepszemu modelowi mieszanek. Jako że muchołówki żałobne pojawiają się w lasach pod koniec kwietnia, w modelach pominięto pierwsze dwie obserwacje, przypadające na pierwsze dwa tygodnie kwietnia. Rycina 4 przedstawia korelacje między prognozami pochodzących z modeli, natomiast Rycina 10 (załącznik) przedstawia zbieżność poszczególnych współczynników najlepszego modelu.

Tabela 9. Modele liczebności dopasowane do danych z liczeń muchołówki żałobnej. Dla każdego modelu podano liczbę parametrów siedliskowych w modelu (k), wartość kryterium informacyjnego Akaike (AIC), różnicę między modelem o największym poparciu a danym modelem ( $\Delta\text{AIC}$ ) oraz wagę modelu ( $\omega\text{AIC}$ ).

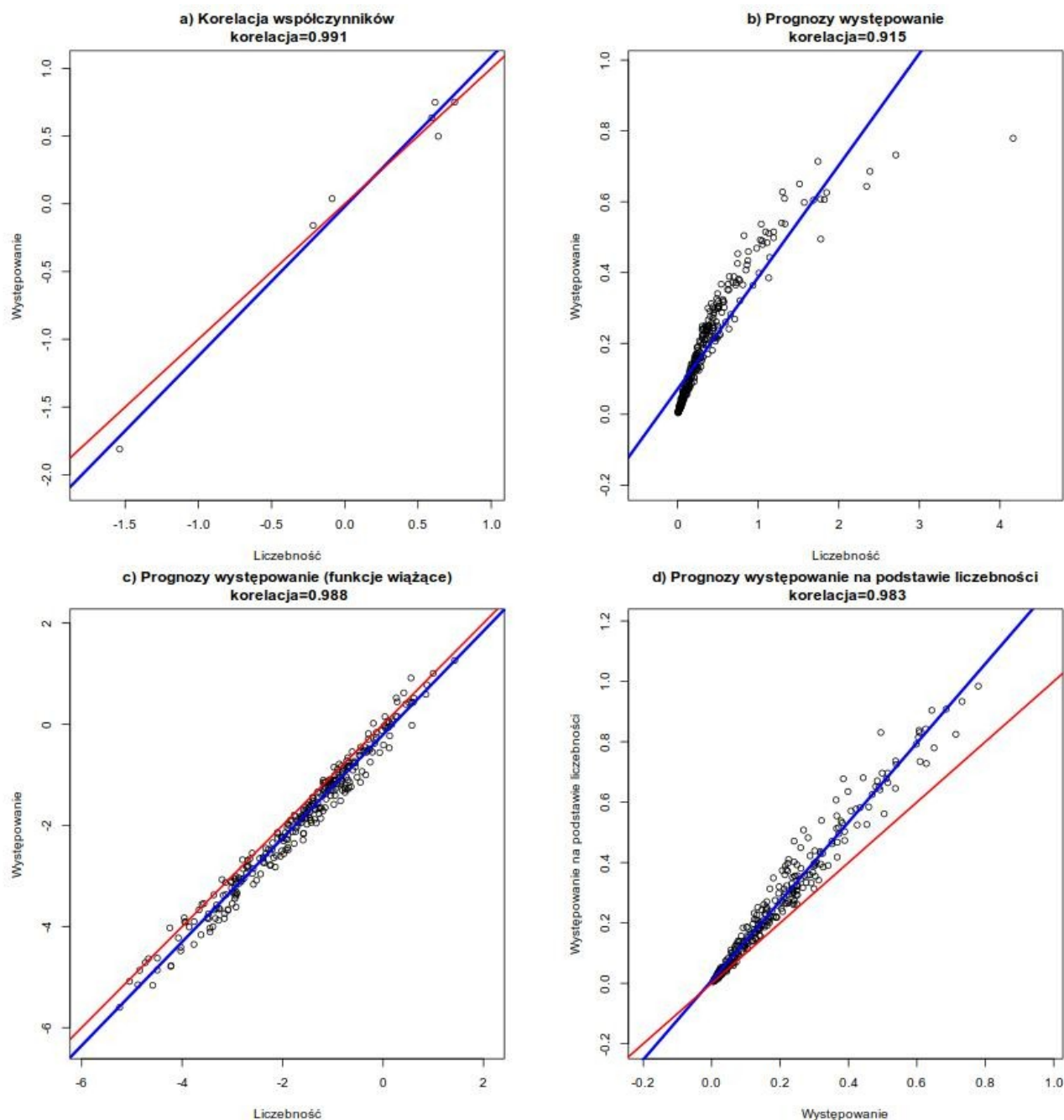
Model		k	AIC	$\Delta\text{AIC}$	$\omega\text{AIC}$
Liczebność	wykrywalność				
1 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{podszyt\_lisc}$	day + effort	4	593.14	0.0	0.477
2 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{pLISp}$	day + effort	4	595.22	2.07	0.169
3 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{p\_richness}$	day + effort	4	595.71	2.57	0.132
4 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{pLISd} + \text{d\_richness} + \text{p\_richness} + \text{Rich16}$	day + effort	7	595.96	2.81	0.117
5 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{pLISd} + \text{pLISp} + \text{pIGLp} + \text{d\_richness} + \text{p\_richness} + \text{Rich16}$	day + effort	9	598.1	4.96	0.04
6 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{podszyt\_igl}$	day + effort	4	601.67	8.53	0.007
7 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{pIGLp}$	day + effort	4	601.8	8.65	0.006
8 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{pIGLd}$	day + effort	4	601.88	8.74	0.006
9 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{dstan\_igl}$	day + effort	4	602.08	8.94	0.005
10 $x_{84} * y_{84} + \text{AREA} + \text{WIEK} + \text{pLISd}$	day + effort	4	602.0	8.95	0.005
0		9			

Tabela 10. Współczynniki parametrów najlepszego modelu mieszanek (submodel z rozkładem Poissona, skala logarytmiczna), objaśniającego zmienność liczebności muchołówki żałobnej. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	0.011	0.126	0.089	0.9290
y84	0.444	0.148	2.992	0.0028
AREA	-0.058	0.151	-0.387	0.6980
WIEK	0.625	0.138	4.539	0.0000
podszyt_lisc	0.393	0.140	2.817	0.0049
x84:y84	0.667	0.208	3.203	0.0014

Tabela 11. Współczynniki parametrów modelu występowania (submodel z rozkładem dwumianowym, skala logitowa) odpowiadającego najlepszemu modelu mieszanek objaśniającego zmienność liczebności muchołówki żałobnej. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	-0.011	0.205	-0.052	0.9580
y84	0.572	0.223	2.569	0.0102
AREA	0.023	0.227	0.101	0.9200
WIEK	0.568	0.213	2.670	0.0076
podszyt_lisc	0.571	0.234	2.442	0.0146
x84:y84	0.739	0.291	2.541	0.0111



Ryc 4. Korelacje a) oszacowań współczynników najlepszego modelu liczebności i odpowiadający model występowania, b) prognozowanych liczebności i prognozowanego występowania – skala liniowa, c) prognozowanych liczebności i prognozowanego występowania – przekształcone funkcjami wiązania (odpowiednio logarytmiczna i logitowa), d) relacja prognozowanego wystąpienia do prognozowanej liczebności zamieniona na występowanie za pomocą funkcji lambda2psi dla muchówki żałobnej. Czerwoną linią przedstawiono funkcję  $f(x) = x$ .

### 3.5. Czubatka *Lophophanes cristatus*

Czubatka była względnie rzadko zaobserwowana (35% powierzchni badanych).

Model który najlepiej opisuje jej liczebność to  $\sim \text{day} + \text{effort} \times \text{x84} \times \text{y84} + \text{AREA} + \text{WIEK} + \text{pLISd} + \text{pLISp} + \text{pIGLp} + \text{d\_richness} + \text{p\_richness} + \text{Rich16}$  (AIC = 804). Porównanie dopasowania różnych modeli przedstawiono w tabeli 12, a dla najlepszego modelu widać wpływ poszczególnych parametrów w tabeli 13. W tabeli 14 przedstawiono współczynniki parametrów modelu występowania odpowiadającego najlepszemu modelowi mieszanek. Rycina 5 przedstawia korelacje między prognozami pochodzących z modeli, natomiast Rycina 11 (załącznik) przedstawia zbieżność poszczególnych współczynników najlepszego modelu.

Tabela 12. Modele liczebności dopasowane do danych z liczeń czubatki. Dla każdego modelu podano liczbę parametrów siedliskowych w modelu (k), wartość kryterium informacyjnego Akaike (AIC), różnicę między modelem o największym poparciu a danym modelem ( $\Delta\text{AIC}$ ) oraz wagę modelu ( $\omega\text{AIC}$ ).

Model		k	AIC	$\Delta\text{AIC}$	$\omega\text{AIC}$
Liczebność	wykrywalność				
1 x84*y84 + AREA + WIEK + pLISd + pLISp + pIGLp + d_richness + p_richness + Rich16	day + effort	9	804.01	0.0	0.954
2 x84*y84 + AREA + WIEK + pLISd + d_richness + p_richness + Rich16	day + effort	7	810.27	6.26	0.042
3 x84*y84 + AREA + WIEK + dstan_lisc	day + effort	4	817.36	13.36	0.001
4 x84*y84 + AREA + WIEK + WILG + forest_type_overall	day + effort	5	817.43	13.42	0.001
5 x84*y84 + AREA + WIEK + dstan_igl	day + effort	4	817.69	13.68	0.001
6 x84*y84 + AREA + WIEK + pLISd	day + effort	4	819.17	15.16	0.0
7 x84*y84 + AREA + WIEK + pIGLd	day + effort	4	821.38	17.37	0.0
8 x84*y84 + AREA + WIEK + gat.pan	day + effort	4	829.77	25.77	0.0
9 x84*y84 + AREA + WIEK + gat.pan.udzial	day + effort	4	887.05	83.04	0.0
10 x84*y84 + AREA + WIEK + d_richness	day + effort	4	892.2	88.28	0.0

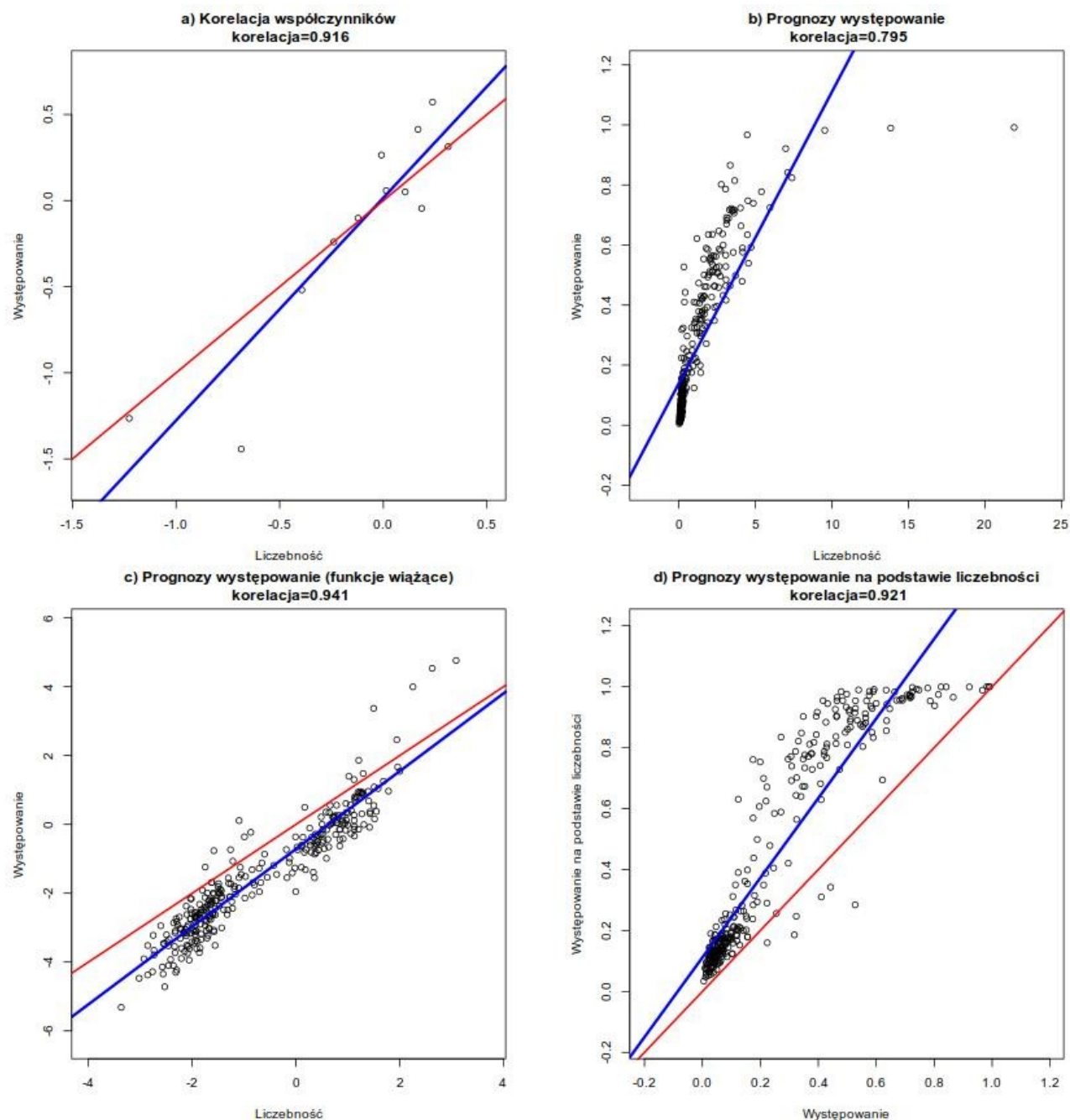


Tabela 13. Współczynniki parametrów najlepszego modelu mieszanek (submodel z rozkładem Poissona, skala logarytmiczna), objaśniającego zmienność liczebności czubatki. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	-0.198	0.112	-1.766	0.0774
y84	-0.026	0.109	-0.241	0.8090
AREA	0.278	0.111	2.506	0.0122
WIEK	0.083	0.140	0.591	0.5540
pLISd	-1.142	0.156	-7.309	0.0000
pLISp	0.205	0.137	1.491	0.1360
pIGLp	0.214	0.062	3.469	0.0005
d_richness	-0.384	0.124	-3.089	0.0020
p_richness	-0.151	0.132	-1.140	0.2540
Rich16	-0.004	0.017	-0.237	0.8130
x84:y84	0.015	0.140	0.104	0.9170

Tabela 14. Współczynniki parametrów modelu występowania (submodel z rozkładem dwumianowym, skala logitowa) odpowiadającego najlepszemu modelu mieszanek objaśniającego zmienność liczebności czubatki. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	-0.102	0.200	-0.511	0.6090
y84	0.264	0.201	1.313	0.1890
AREA	0.313	0.191	1.642	0.1010
WIEK	0.050	0.221	0.226	0.8220
pLISd	-1.265	0.224	-5.645	0.0000
pLISp	-0.045	0.255	-0.178	0.8590
pIGLp	0.572	0.236	2.428	0.0152
d_richness	-0.519	0.211	-2.460	0.0139
p_richness	-0.240	0.239	-1.005	0.3150
Rich16	0.009	0.033	0.263	0.7920
x84:y84	0.414	0.260	1.592	0.1110



Ryc 5. Korelacje a) oszacowań współczynników najlepszego modelu liczebności i odpowiadający model występowania, b) prognozowanych liczebności i prognozowanego występowania – skala liniowa, c) prognozowanych liczebności i prognozowanego występowania – przekształcone funkcjami wiązania (odpowiednio logarytmiczna i logitowa), d) relacja prognozowanego wystąpienia do prognozowanej liczebności zamieniona na występowanie za pomocą funkcji lambda2psi dla czubatk. Czerwoną linią przedstawiono funkcję  $f(x) = x$ .

### 3.6. Siniak *Columba oenas*

Siniak był rzadko zaobserwowany (23% powierzchni badanych). Model który najlepiej opisuje jego liczebność to  $\sim \text{day} + \text{effort} \times \text{x84} \times \text{y84} + \text{AREA} + \text{WIEK} + \text{pLISd} + \text{pLISp} + \text{pIGLp} + \text{d\_richness} + \text{p\_richness} + \text{Rich16}$  (AIC = 614.8). Porównanie dopasowania różnych modeli przedstawiono w tabeli 15, a dla najlepszego modelu widać wpływ poszczególnych parametrów w tabeli 16. W tabeli 17 przedstawiono współczynniki parametrów modelu występowania odpowiadającego najlepszemu modelowi mieszanek. Rycina 6 przedstawia korelacje między prognozami pochodzących z modeli, natomiast Rycina 12 (załącznik) przedstawia zbieżność poszczególnych współczynników najlepszego modelu.

Tabela 15. Modele liczebności dopasowane do danych z liczeń siniaka. Dla każdego modelu podano liczbę parametrów siedliskowych w modelu (k), wartość kryterium informacyjnego Akaike (AIC), różnicę między modelem o największym poparciu a danym modelem ( $\Delta\text{AIC}$ ) oraz wagę modelu ( $\omega\text{AIC}$ ).

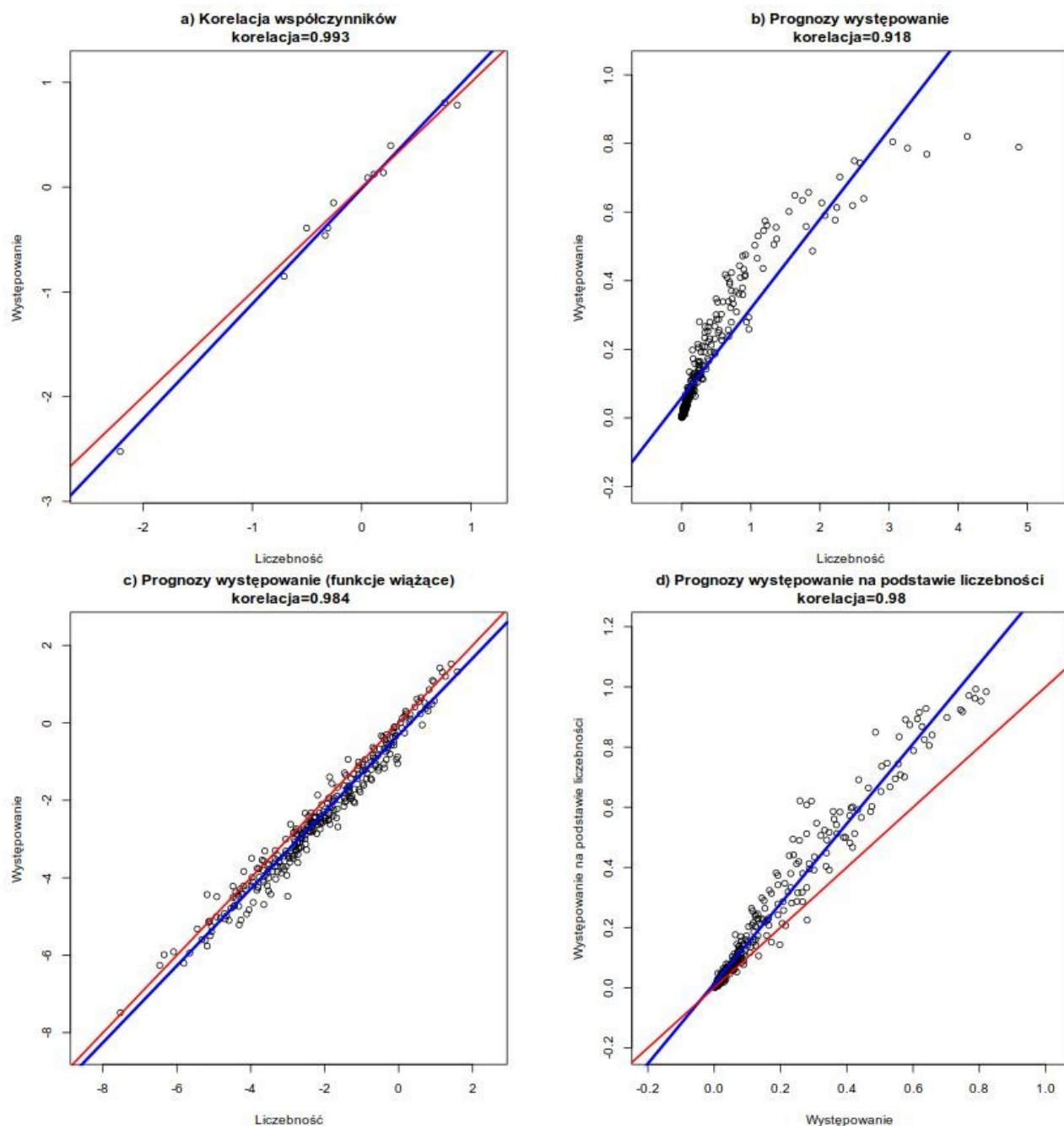
Model		k	AIC	$\Delta\text{AIC}$	$\omega\text{AIC}$
Liczebność	wykrywalność				
1 x84*y84 + AREA + WIEK + pLISd + pLISp + pIGLp + d_richness + p_richness + Rich16	day + effort	9	614.87	0.0	0.762
2 x84*y84 + AREA + WIEK + pLISd + d_richness + p_richness + Rich16	day + effort	7	617.21	2.34	0.237
3 x84*y84 + AREA + WIEK + p_richness	day + effort	4	630.36	15.49	0.0
4 x84*y84 + AREA + WIEK + gat.pan.udzial	day + effort	4	634.1	19.23	0.0
5 x84*y84 + AREA + WIEK + WILG + forest_type_overall	day + effort	5	635.97	21.1	0.0
6 x84*y84 + AREA + WIEK + pIGLp	day + effort	4	637.05	22.18	0.0
7 x84*y84 + AREA + WIEK + podszyt_igl	day + effort	4	637.76	22.89	0.0
8 x84*y84 + AREA + WIEK + pIGLd	day + effort	4	638.54	23.67	0.0
9 x84*y84 + AREA + WIEK + pLISd	day + effort	4	638.84	23.97	0.0
10 x84*y84 + AREA + WIEK + dstan_lisc	day + effort	4	639.4	24.62	0.0
		9			

Tabela 16. Współczynniki parametrów najlepszego modelu mieszanek (submodel z rozkładem Poissona, skala logarytmiczna), objaśniającego zmienność liczebności siniaka. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	-0.221	0.136	-1.619	0.1050
y84	0.290	0.132	2.197	0.0280
AREA	-0.316	0.141	-2.241	0.0250
WIEK	0.643	0.153	4.205	0.0000
pLISd	0.140	0.178	0.786	0.4320
pLISp	0.319	0.157	2.035	0.0419
pIGLp	-0.344	0.445	-0.774	0.4390
d_richness	0.046	0.125	0.371	0.7110
p_richness	-0.813	0.190	-4.276	0.0000
Rich16	0.092	0.025	3.658	0.0003
x84:y84	-0.143	0.159	-0.897	0.3700

Tabela 17. Współczynniki parametrów modelu występowania (submodel z rozkładem dwumianowym, skala logitowa) odpowiadającego najlepszemu modelu mieszanek objaśniającego zmienność liczebności siniaka. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	-0.150	0.205	-0.730	0.4650
y84	0.396	0.197	2.012	0.0442
AREA	-0.461	0.229	-2.008	0.0446
WIEK	0.783	0.234	3.349	0.0008
pLISd	0.124	0.246	0.503	0.6150
pLISp	0.138	0.266	0.519	0.6030
pIGLp	-0.391	0.484	-0.808	0.4190
d_richness	0.089	0.199	0.449	0.6530
p_richness	-0.851	0.293	-2.906	0.0037
Rich16	0.122	0.041	3.021	0.0025
x84:y84	-0.390	0.234	-1.663	0.0962



Ryc 6. Korelacje a) oszacowań współczynników najlepszego modelu liczebności i odpowiadający model występowania, b) prognozowanych liczebności i prognozowanego występowania – skala liniowa, c) prognozowanych liczebności i prognozowanego występowania – przekształcone funkcjami wiązania (odpowiednio logarytmiczna i logitowa), d) relacja prognozowanego wystąpienia do prognozowanej liczebności zamieniona na występowanie za pomocą funkcji lambda2psi dla siniaka. Czerwoną linią przedstawiono funkcję  $f(x) = x$ .

### 3.7. Raniuszek *Aegithalos caudatus*

Raniuszek był rzadko zaobserwowany (15% powierzchni badanych). Model który najlepiej opisuje jego liczebność to  $\sim \text{day} + \text{effort} \times \text{x84} \times \text{y84} + \text{AREA} + \text{WIEK} + \text{pLISd} + \text{d\_richness} + \text{p\_richness} + \text{Rich16}$  (AIC = 241.56). Porównanie dopasowania różnych modeli przedstawiono w tabeli 18, a dla najlepszego modelu widać wpływ poszczególnych parametrów w tabeli 19. W tabeli 20 przedstawiono współczynniki parametrów modelu występowania odpowiadającego najlepszemu modelowi mieszanek. Rycina 7 przedstawia korelacje między prognozami pochodzących z modeli, natomiast Rycina 13 (załącznik) przedstawia zbieżność poszczególnych współczynników najlepszego modelu.

Tabela 18. Modele liczebności dopasowane do danych z liczeń raniuszka. Dla każdego modelu podano liczbę parametrów siedliskowych w modelu (k), wartość kryterium informacyjnego Akaike (AIC), różnicę między modelem o największym poparciu a danym modelem ( $\Delta\text{AIC}$ ) oraz wagę modelu ( $\omega\text{AIC}$ ).

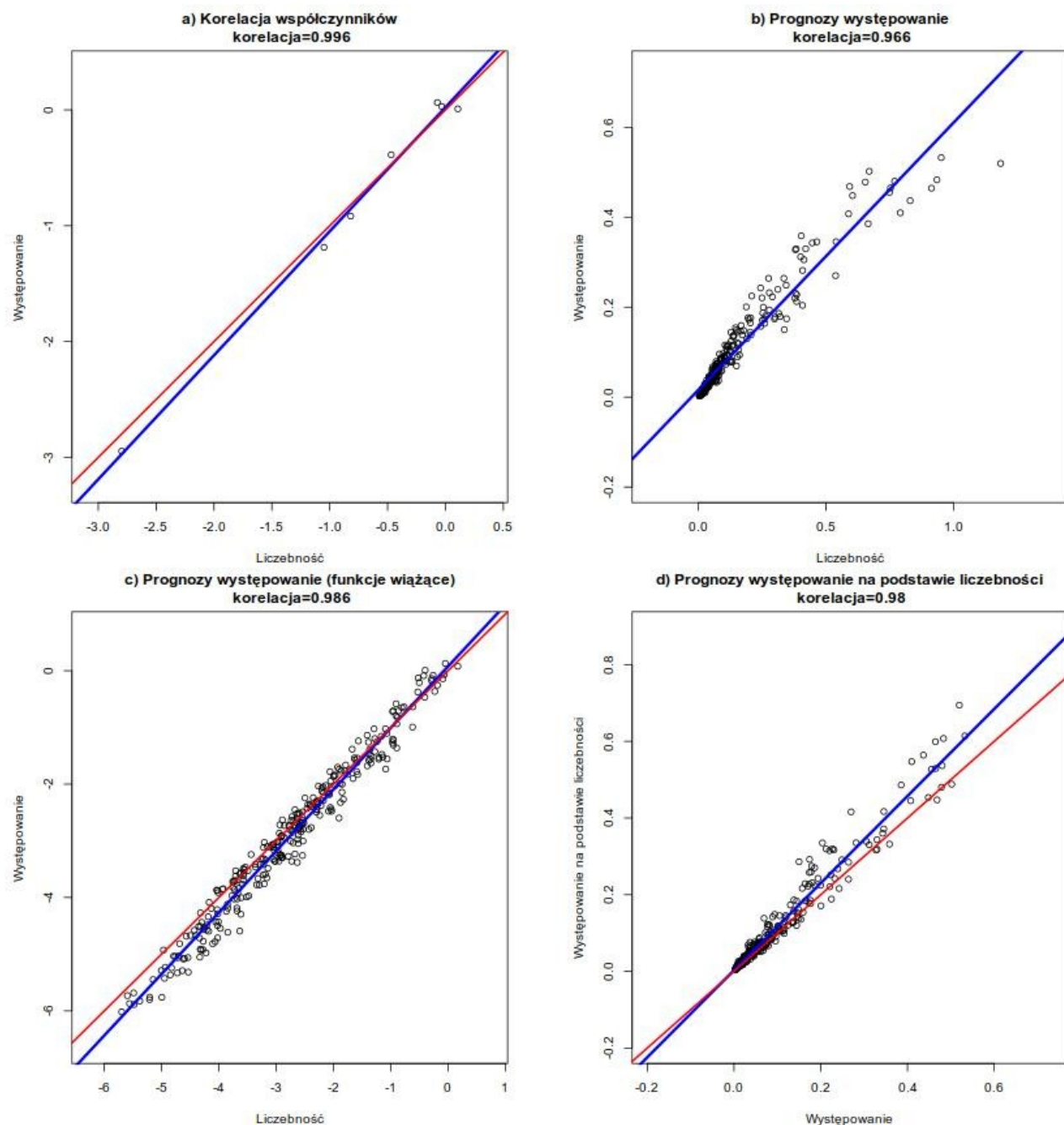
Model		k	AIC	$\Delta\text{AIC}$	$\omega\text{AIC}$
Liczebność	wykrywalność				
1 x84*y84 + AREA + WIEK + pLISd + d_richness + p_richness + Rich16	day + effort	7	241.56	0.0	0.687
2 x84*y84 + AREA + WIEK + pLISd + pLISp + pIGLp + d_richness + p_richness + Rich16	day + effort	9	245.02	3.45	0.122
3 x84*y84 + AREA + WIEK + WILG + forest_type_overall	day + effort	5	246.72	5.16	0.052
4 x84*y84 + AREA + WIEK + WILG	day + effort	4	248.09	6.53	0.026
5 x84*y84 + AREA + WIEK + pLISd	day + effort	4	248.41	6.84	0.022
6 x84*y84 + AREA + WIEK + pIGLd	day + effort	4	248.57	7.01	0.021
7 x84*y84 + AREA + WIEK + pLISp	day + effort	4	249.9	8.33	0.011
8 x84*y84 + AREA + WIEK + p_richness	day + effort	4	250.23	8.66	0.009
9 x84*y84 + AREA + WIEK + dstan_lisc	day + effort	4	250.38	8.81	0.008
10 x84*y84 + AREA + WIEK + dstan_igl	day + effort	4	250.5	8.93	0.008

Tabela 19. Współczynniki parametrów najlepszego modelu mieszanek (submodel z rozkładem Poissona, skala logarytmiczna), objaśniającego zmienność liczebności raniuszka. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	0.344	0.203	1.695	0.0901
y84	0.047	0.239	0.195	0.8452
AREA	-0.234	0.226	-1.033	0.3017
WIEK	-0.073	0.231	-0.315	0.7531
pLISd	0.354	0.253	1.401	0.1614
d_richness	-0.080	0.212	-0.376	0.7073
p_richness	0.253	0.238	1.062	0.2880
Rich16	0.120	0.039	3.111	0.0019
x84:y84	-0.018	0.256	-0.069	0.9449

Tabela 20. Współczynniki parametrów modelu występowania (submodel z rozkładem dwumianowym, skala logitowa) odpowiadającego najlepszemu modelowi mieszanek objaśniającego zmienność liczebności raniuszka. Podano wartości estymatorów wraz z ich błędami standardowymi, wartość statystyki Z i prawdopodobieństwo uzyskania większej wartości parametru niż podana w kolumnie „wartość” ( $P > |Z|$ )

Parametr	Wartość	Błąd std.	Z	$P(>  Z )$
x84	24.570	11.980	2.050	0.0403
y84	23.030	-	-	-
AREA	37.410	-	-	-
WIEK	14.290	3.680	3.890	0.0001
pLISd	7.420	-	-	-
d_richness	9.060	7.460	1.210	0.2245
p_richness	20.350	-	-	-
Rich16	3.770	2.380	1.580	0.1131
x84:y84	-18.190	-	-	-



Ryc 7. Korelacje a) oszacowań współczynników najlepszego modelu liczebności i odpowiadający model występowania, b) prognozowanych liczebności i prognozowanego występowania – skala liniowa, c) prognozowanych liczebności i prognozowanego występowania – przekształcone funkcjami wiązania (odpowiednio logarytmiczna i logitowa), d) relacja prognozowanego wystąpienia do prognozowanej liczebności zamieniona na występowanie za pomocą funkcji lambda2psi dla raniuszka. Czerwoną linią przedstawiono funkcję  $f(x) = x$ .



## 4. Dyskusja

U większości gatunków modelem o najwyższym poparciu był ten o parametrach „x84\*y84 + AREA + WIEK + pLISd + pLISp + pIGLp + d\_richness + p\_richness + Rich16”, zatem dla tych danych najlepszym wskaźnikiem liczebności gatunku (a więc także występowania) jest różnorodność roślinności na danej powierzchni badawczej. Jest to szczególnie ciekawe, ponieważ wskaźnik AIC preferuje prostsze modele, a wybrany model zawiera znacznie więcej współczynników niż pozostałe. Należałoby zobaczyć jak modele alternatywne sprawdzą się na innych danych, gdyż możliwym wytłumaczeniem skuteczności tego modelu jest nadmierne dopasowanie (*overfitting*). Z drugiej strony, łatwo zrozumieć, że dla leśnych gatunków ptaków to właśnie struktura i skład roślinności – drzew w najwyższej warstwie (charakteryzowanych przez zmienną WIEK, pLISd - udział drzewostanu liściastego oraz d\_richness – bogactwo gatunkowe drzewostanu) i krzewów (zmienne pLISp, pIGLp i p\_richness, opisujące proporcje podszytu gatunków liściastych, iglastych oraz bogactwo gatunkowe warstwy podszytu) w decydujący sposób wpływają na ich liczebność i występowanie.

Sposób porównywania modeli liczebności i występowania był dość prosty. Zamiast tworzenia modelu występowania z tymi samymi współczynnikami, co najlepszy model liczebności, można utworzyć różne modele występowania i porównać najlepszy model liczebności z najlepszym modelem występowania. Jednak to podejście niekoniecznie sprawdzałoby jak dobrze dane o występowaniu pokrywają się z danymi o liczebności – bardziej porównywałoby czy te same współczynniki środowiskowe najlepiej opisują liczebność i występowanie. Można również porównywać różne modele zbiorczo – zamiast porównywać ze sobą jedynie najlepsze modele, można zastosować miary odległości by sprawdzić jak bardzo

różnią się między sobą zbiory modeli liczebności i występowania pod względem wartości AIC – należałoby wybrać pulę kombinacji współczynników środowiskowych, na ich podstawie utworzyć modele liczebności i występowania, posortować modele w obrębie typów (osobno dla liczebności i występowania), a następnie za pomocą miar odległości zmierzyć jak bardzo te dwa zbiory różnią się między sobą. Ponieważ ważna jest nie tylko kolejność, ale też i różnica AIC między modelami, należałoby zastosować miary odległości biorąc te różnice pod uwagę (Kumar 2010).

Z miar korelacji wynika, że liczebność jest skorelowana z występowaniem, co nie powinno dziwić. Korelacja ta jest tym silniejsza, im mniejsza jest liczebność na poszczególnych powierzchniach, czego można się spodziewać, ponieważ gdy liczebność wynosi najwyżej 1, staje się nieodróżnialna od występowania. Przy bardziej pospolitych ptakach jak np. sikora uboga czy sosnówka, znacznie większa korelacja jest między występowaniem a liczebnością, gdy przekształca się liczebność na skalę logarytmiczną. Wynika to z faktu, że wtedy wartości liczebności są znacznie bliższe 1. Jeszcze bliższe korelacje można uzyskać stosując funkcje wiążące do skalowania (logarytmiczna dla liczebności oraz logitowa dla występowania) – w przypadku badanych ptaków, najgorsza korelacja wynosiła 0.907 w przypadku najbardziej pospolitego ptaka (sosnówka). Ciekawym przypadkiem jest czubatka, która, choć na występuje na małej powierzchni (35%), ma szacowane liczebności bardziej zbliżone do pospolitszych ptaków. Można to zjawisko wytłumaczyć tym, że preferuje lasy z przewagą drzew iglastych, a wśród badanych przeważały lasy liściaste – jako bory zaklasyfikowano jedynie 23% powierzchni. Podobnie zbieżnie są wykresy przedstawiające wpływ poszczególnych zmiennych siedliskowych na modeli (załącznik) - w większości przypadków zmienne są podobnie szacowane dla obu modeli (występowania i liczebności).

Zamiana prognozowanych liczebności na występowanie jest dość skuteczna, i choć wyniki były bardziej zbliżone dla mniej licznych gatunków, wartość korelacji wynosiła zawsze blisko lub powyżej 0.9 (0.898 w przypadku sosnówki). Oznacza to, że preferencje siedliskowe analizowanych – i zapewne innych – gatunków można z powodzeniem przybliżać w oparciu o proste do uzyskania dane binarne, mówiące o występowaniu gatunku w danym miejscu. Wniosek ten ma o tyle duże znaczenie praktyczne, że policzenie osobników (samców, terytoriów, gniazd) jest najczęściej trudne, a uzyskanie danych o samej tylko obecności – łatwiejsze.

Zastosowane metody nie będą skuteczne w przypadku bardzo pospolitych gatunków. Jeżeli jakiś gatunek występuje na prawie wszystkich powierzchniach, dane o występowaniu będą mało użyteczne w określeniu zależności ekologicznych, jako że modele nie będą w stanie wykryć różnic między powierzchniami. W takim przypadku jedynym rozwiązaniem będzie zebranie i użycie danych o liczebności. Podobnie, w przypadku bardzo rzadko występujących ptaków, modele nie będą w stanie znaleźć zależności, więc modele wynikowe będą bezużyteczne. Dla gatunków występujących na mniej niż 15% powierzchni niektóre modele nie konwergowały, prawdopodobnie z powodu zbyt małej ilości danych. Przypuszczalnie można przyjąć frekwencję na poziomie ok. 10–15% jako dolną granicę sensowności tworzenia modeli hierarchicznych, przy czym należałoby dokładniej zbadać gdzie leżą górne i dolne granice skuteczności modeli. Zagadnienia te najlepiej zbadać z użyciem symulacji (w których „prawda” jest znana – są to dane wsadowe).

## 5. Wnioski

Z porównań modeli liczebności i występowania można wywnioskować, że w przypadku gatunków względnie pospolitych (w tej pracy wykorzystano gatunki zaobserwowane na 15%–70% powierzchni) dane o liczebności dość dobrze przekładają się na występowanie i vice versa. Zatem w celu wskazania ważnych dla gatunku cech siedliskowych, kształtujących występowanie i liczebność, można z podobną skutecznością wykorzystać dane o występowaniu, jak i dane o liczebności. Wniosek ten pozwala na znaczne ułatwienie badań terenowych, ponieważ znacznie prostsze jest ustalenie występowania gatunku na danej powierzchni niż określenie jego liczebności. Mimo pozornego ubóstwa w informację danych binarnych, takie proste dane dostarczają istotnych informacji ekologicznych.

## 6. Bibliografia

- Begon M.**, Harper J. L. , Townsend C. R. **1986**. Ecology: individuals, populations and communities. Blackwell Scientific Publications, Oxford.
- Burnham K.P.**, Anderson D.R. **2002**. Model Selection and Inference: A Practical Information-Theoretic Approach. 2nd Edition, Springer-Verlag, New York.
- Chandler R. B.**, Royle J. A., King D. **2011**. Inference about Density and Temporary Emigration in Unmarked Populations. *Ecology*, 92: 1429-1435.
- Eriksson M. O. G.**, Götmark F. **1982**. Habitat Selection: Do Passerines Nest in Association with Lapwings *Vanellus Vanellus* as Defence against Predators? *Ornis Scandinavica*, 13: 189-192.
- Freckleton R. P.**, Noble D., Webb, T. J. **2006**. Distributions of Habitat Suitability and the Abundance-Occupancy Relationship. *The American Naturalist*, 167(2): 260–275.
- Gaston K.J.**, Blackburn T.M., Greenwood J.J.D., Gregory R.D., Quinn R.M., Lawton J.H. **2000**. Abundance-occupancy relationships. *Journal of Applied Ecology*, 37(suppl. 1): 39–59.
- Grether G. F.**, Anderson C. N., Drury J. P., Kirschel A. N. G., Losin N., Okamoto K., Peiman K. S. **2013**. The evolutionary consequences of interspecific aggression. *Annals of the New York Academy of Sciences*, 1289(1): 48–68.
- Horne, B. V. 1983**. Density as a Misleading Indicator of Habitat Quality. *The Journal of Wildlife Management*, 47(4): 893-901.
- Hostetter, N. J.**, Gardner B., Sillett T. S., Pollock K. H., Simons T. R. **2019**. An integrated model decomposing the components of detection probability and abundance in unmarked populations. *Ecosphere* 10(3):
- Hui C.**, McGeoch M.A., Reyers B., le Roux P.C., Greve M. Chown S.L. **2009**. Extrapolating population size from the occupancy-abundance relationship and the scaling pattern of occupancy. *Ecological Applications*, 19: 2038–2048.
- Jones J. A.**, Harris M. R., Siefferman L. **2014**. Physical habitat quality and interspecific competition interact to influence territory settlement and reproductive success in a cavity nesting bird. *Frontiers in Ecology and Evolution*, 2.
- Kéry M.**, Royle A.J. **2016**. Applied Hierarchical Modeling in Ecology. Analysis of distribution, abundance and species richness in R and BUGS. Vol. 1. Prelude and Static Models. Academic Press.
- Krebs, CJ, 2009**, Ecology; the Experimental Analysis of Distribution and Abundance, sixth ed. Benjamin Cummings, San Francisco.
- Kumar R.**, Vassilvitskii S. **2010**. Generalized distances between rankings. Proceedings of the 19th International Conference on World Wide Web - WWW '10.

**Liebezeit J. R.**, George T. L.. **2002**. Nest Predators, Nest-Site Selection, and Nesting Success of the Dusky Flycatcher in a Managed Ponderosa Pine Forest, *The Condor*, 104(3): 507–517.

**MacKenzie D. I.**, Nichols J. D.. **2004**. Occupancy as a surrogate for abundance estimation. *Animal Biodiversity and Conservation* 27: 461– 467.

**Nelder J. A.**, Wedderburn R. W. M. **1972**. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3): 370.

**Neubauer G.**, Sikora A. **2016**. Ocena zagęszczenia i liczebności muchołówki małej *Ficedula parva* w lasach trójmiejskich z wykorzystaniem powtarzanych nasłuchów z punktów. *Ornis Polonica*. 57: 169-186.

**Pintar M. R.**, Bohenek J. R., Eveland L. L., Resetarits W. J. **2018**. Colonization across gradients of risk and reward: Nutrients and predators generate species-specific responses among aquatic insects. *Functional Ecology*, 32(6): 1589–1598.

**R Core Team.** **2018**. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

**Fiske I.**, Chandler R. **2011**. unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, 43(10): 1-23.

**Resetarits W. J.**, Bohenek J. R., Breech T., Pintar M. R. **2018**. Predation risk and patch size jointly determine perceived patch quality in ovipositing treefrogs, *Hyla chrysoscelis*. *Ecology*, 99(3): 661–669.

**Rosenzweig, M. L.** **1991**. Habitat Selection and Population Interactions: The Search for Mechanism. *The American Naturalist*, 137: 5–28.

**Royle, J. A.** **2004**. N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. *Biometrics*, 60(1): 108–115.

**Royle J.A.**, Link W.A. **2006**. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, 87: 835– 841.

**Tyre A.**, Tenhumberg J., B., Field S. A., Niejalke D., Parris K., Possingham H. **2003**. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications* 13: 1790– 1801.

**Webb T. J.**, Freckleton R. P., Gaston, K. J. **2011**. Characterizing abundance-occupancy relationships: there is no artefact. *Global Ecology and Biogeography*, 21(9): 952–957

Wpływ prowadzonej gospodarki leśnej na populacje wybranych gatunków ptaków interioru leśnego w lasach nizinnych Polski

<https://otop.org.pl/naszeprojekty/liczymy/projekt-ptaki-lesne/>

## **7. Abstract**

### **7.1. Streszczenie**

Celem pracy było sprawdzenia czy zależności ekologiczne mogą być podobnie szacowane z użyciem danych o występowaniu i liczebności gatunku. Analiza została dokonana na zbiorze danych pochodzących z 300 leśnych powierzchni badawczych (między 5 – 36ha). Każdy obszar został skontrolowany przez doświadczonych obserwatorów cztery razy między kwietniem i majem 2016 roku, średnio co dwa tygodnie, za każdym razem przez około 100 minut, w celu zliczenia liczebności zaobserwowanych gatunków ptaków. Następnie wybrano sześć gatunków, i każdego z nich utworzono model hierarchiczny (Kéry and Royle 2016) który najlepiej opisywał dane o liczebności dla danego gatunku. Na podstawie współczynników powstałych modeli, skonstruowano odpowiadające model występowania, po czym każda para modeli została porównana.

Z porównań modeli liczebności i występowania można było wywnioskować, że w przypadku gatunków względnie pospolitych (w tej pracy wykorzystano gatunki zaobserwowane na 15%–70% powierzchni) dane o liczebności dość dobrze przekładają się na występowanie i vice versa. Zatem w celu wskazania ważnych dla gatunku cech siedliskowych, kształtujących występowanie i liczebność, można z podobną skutecznością wykorzystać dane o występowaniu, jak i dane o liczebności. Wniosek ten pozwala na znaczne ułatwienie badań terenowych, ponieważ znacznie prostsze jest ustalenie występowania gatunku na danej powierzchni niż określenie jego liczebności. Mimo pozornego ubóstwa w informację danych binarnych, takie proste dane dostarczają istotnych informacji ekologicznych.

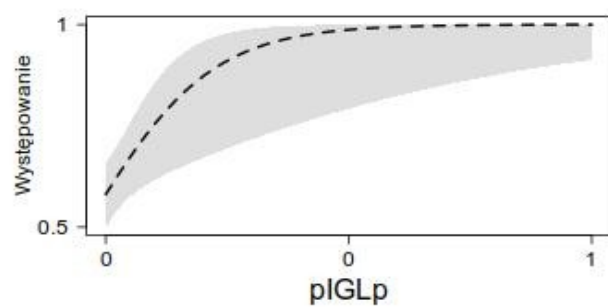
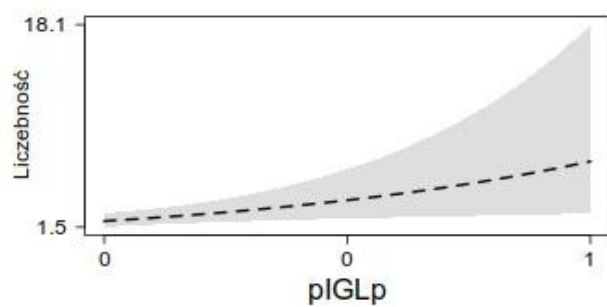
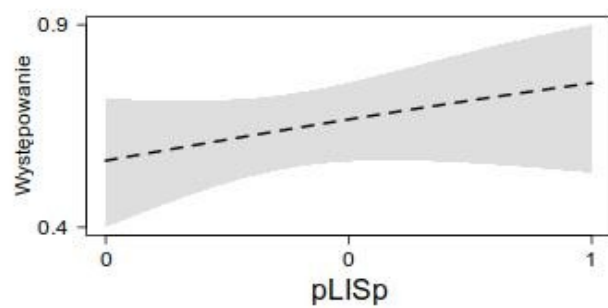
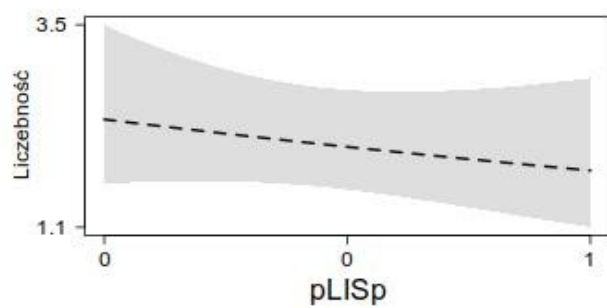
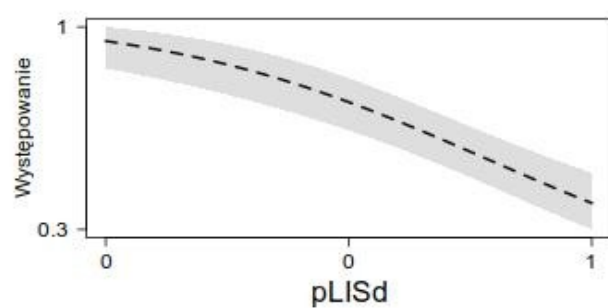
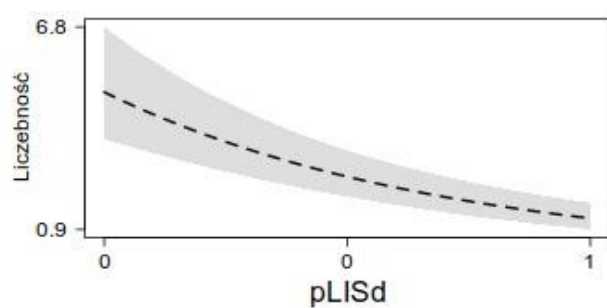
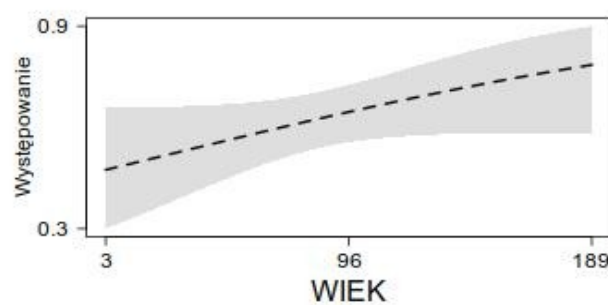
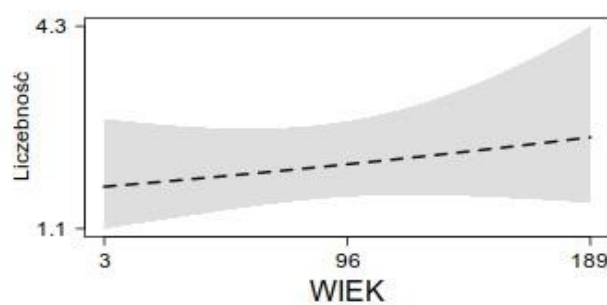
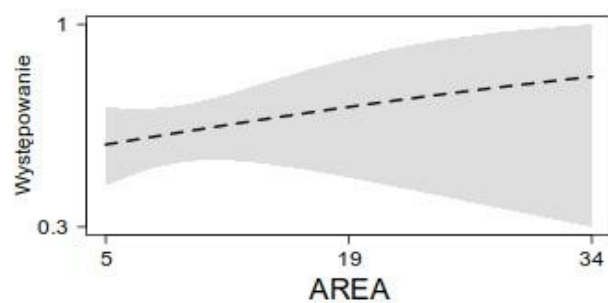
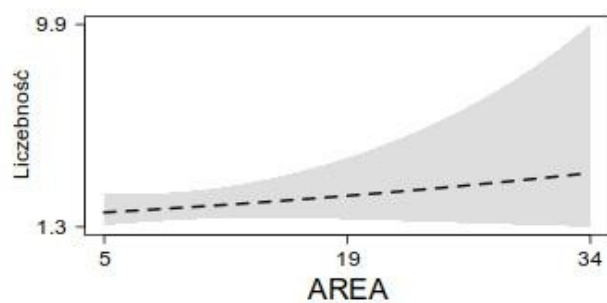
## **7.2. English summary**

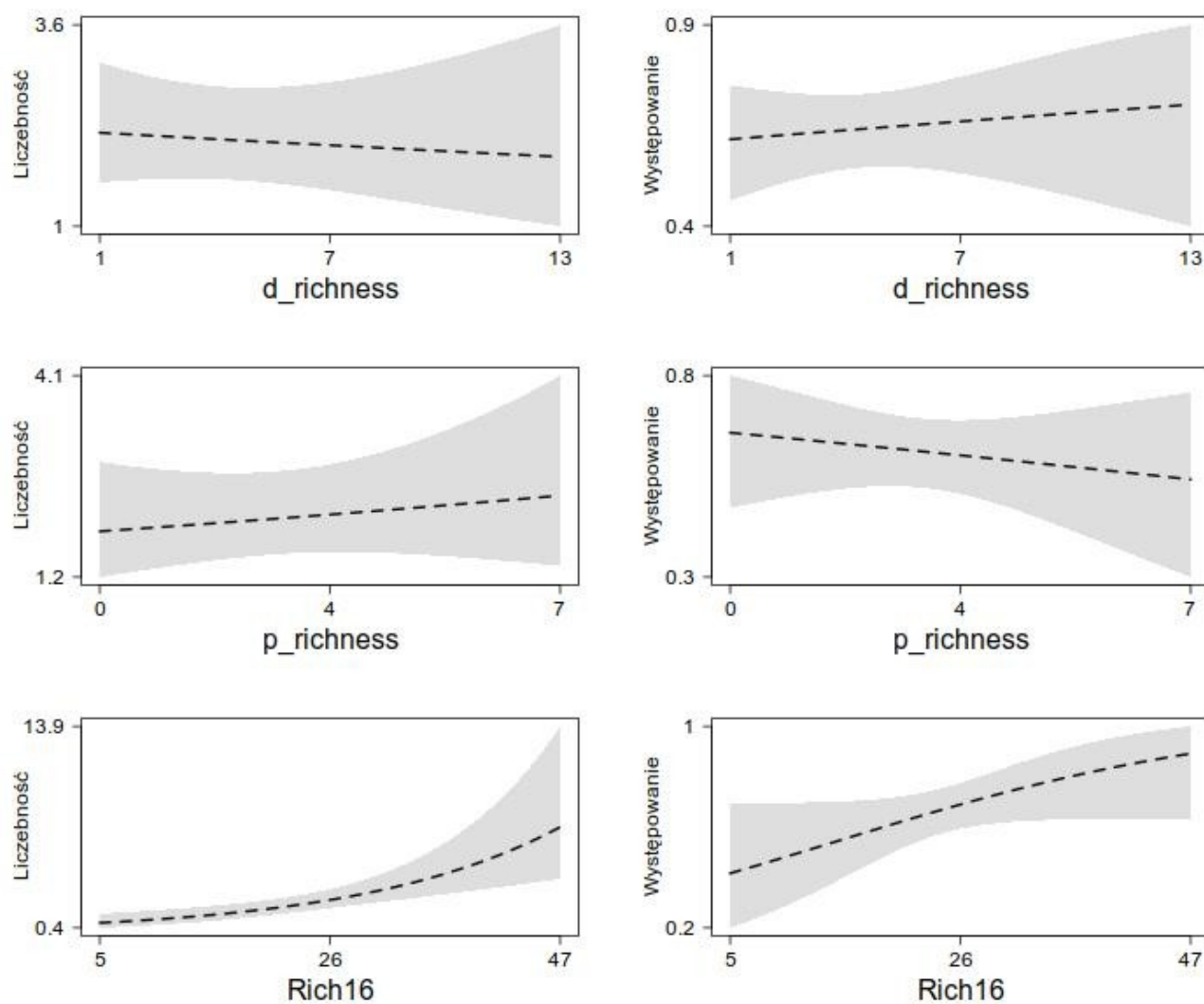
The goal of this thesis was to determine whether ecological relationships can be similarly estimated with data on the abundance and occurrence of a species. This was done on a data set of 300 forest areas (between 5 – 36ha). Each area was visited by experienced observers four times in April and May 2016, approximately every two weeks, for around 100 minutes each time, in order to count the numbers of observed bird species. Six species were then chosen, for each of which a hierarchical model (Kéry and Royle 2016) that best fitted its abundance data was constructed. On the basis of the covariants of the resulting models, corresponding occurrence models were constructed and the two (per species) models were compared.

Comparing the abundance and occurrence models it was possible to conclude that in the case of species that are moderately common (in this case defined as observed on 15%-70% of all tested areas), abundance data is easily transformed to occurrence and vice versa. Therefore, in order to describe species-specific environmental factors that effect both abundance and occurrence, both measurements can be used interchangeably with similar results. This conclusion can greatly simplify field measurements, as it's much easier to detect the presence of a species in a given area than to quantify its abundance. Despite the seemingly paucity of binary data, they can provide significant ecological information.

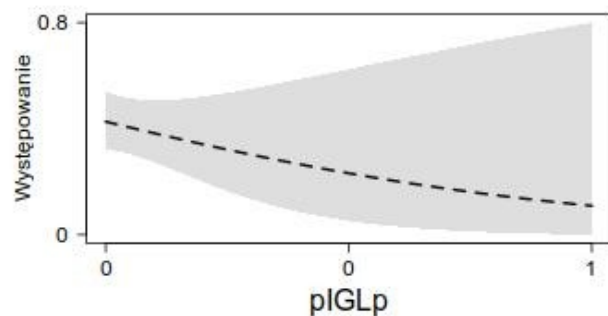
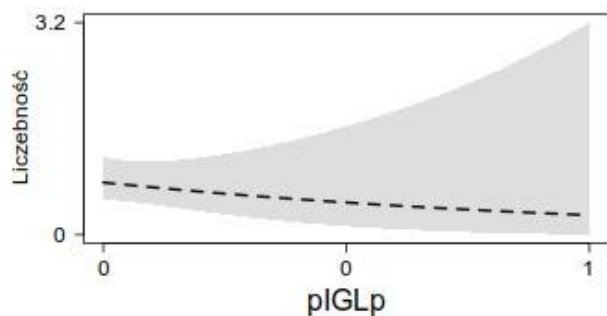
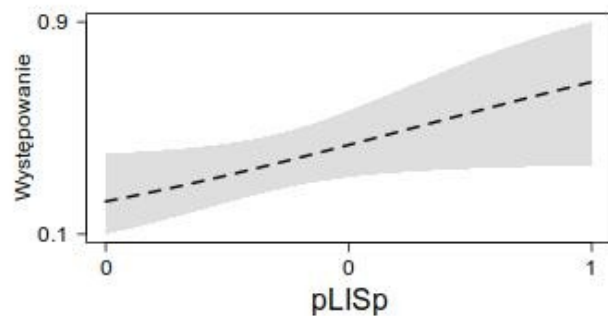
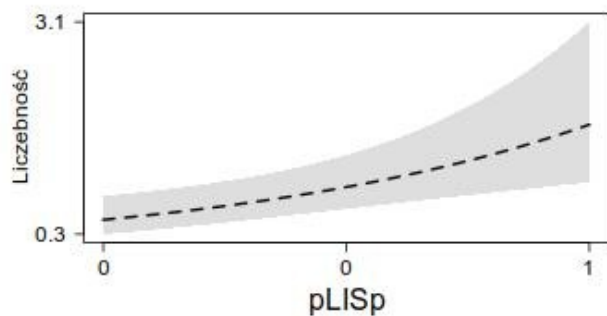
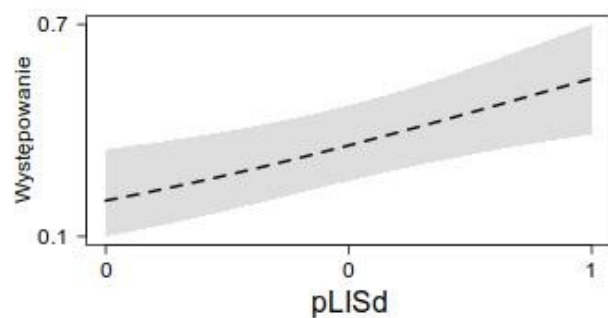
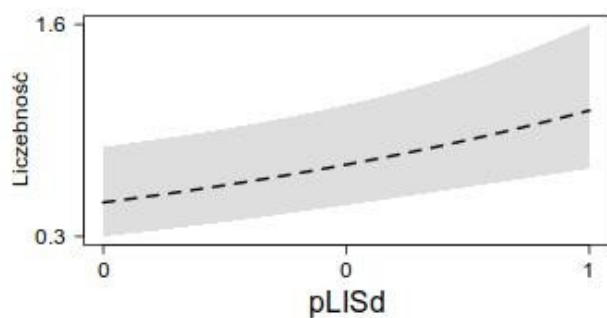
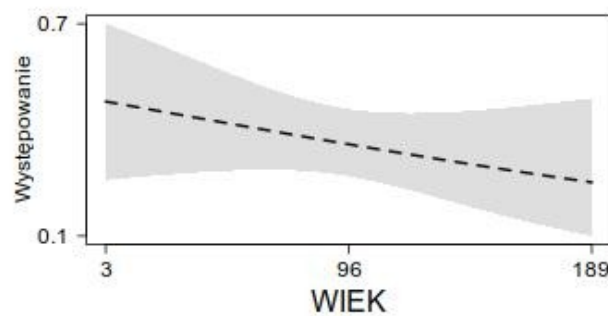
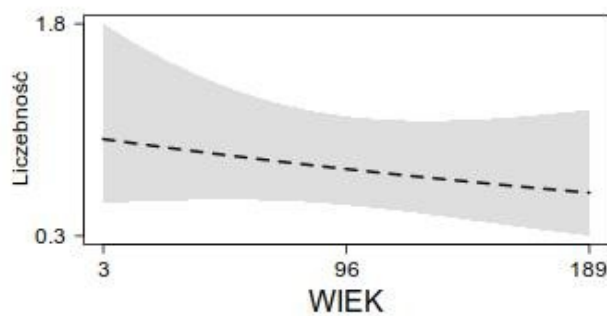
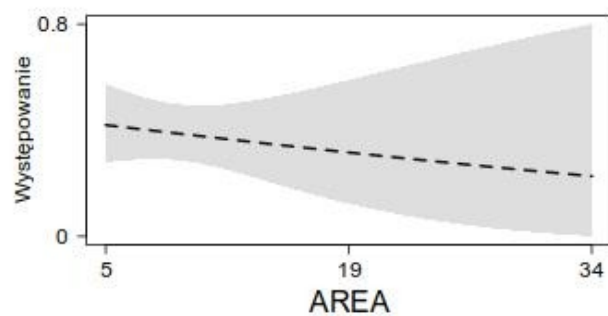
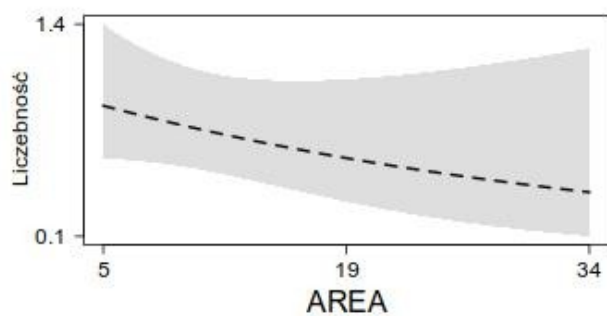


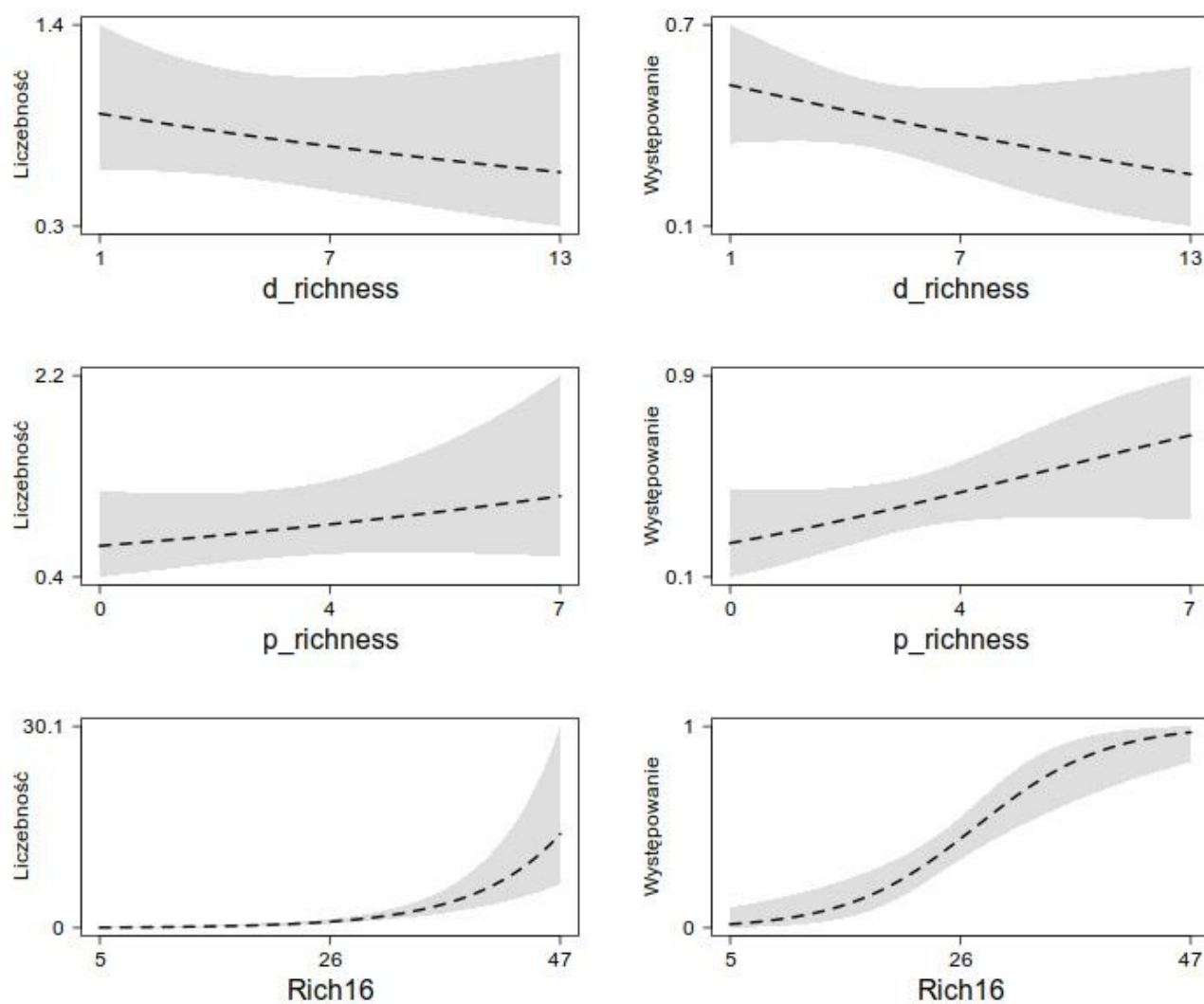
## Załącznik (ryciny 8-13)



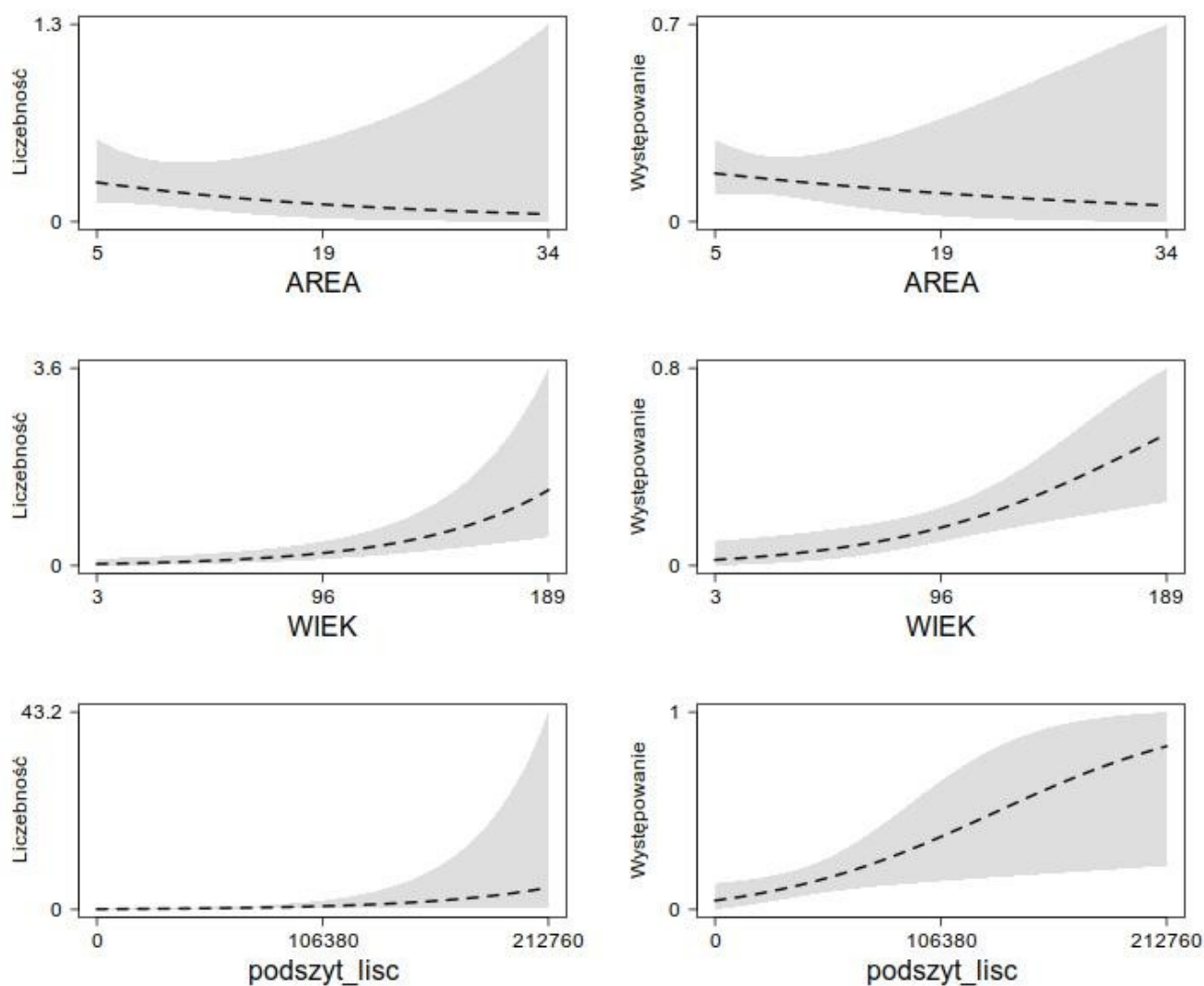


Ryc. 8. Wpływ poszczególnych zmiennych siedliskowych na liczebność (lewa kolumna) i występowanie (prawa kolumna) sosnówki.

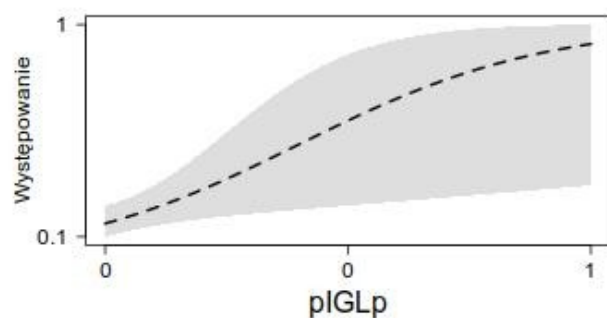
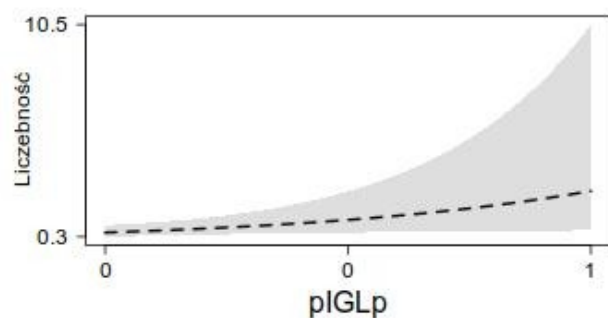
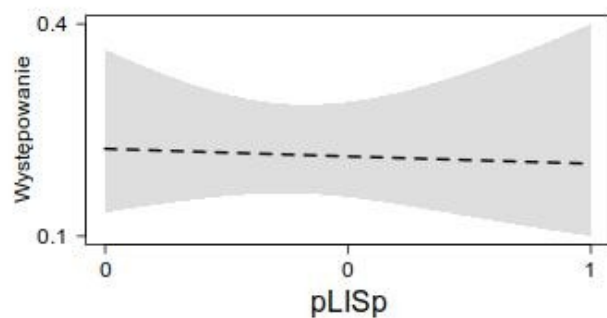
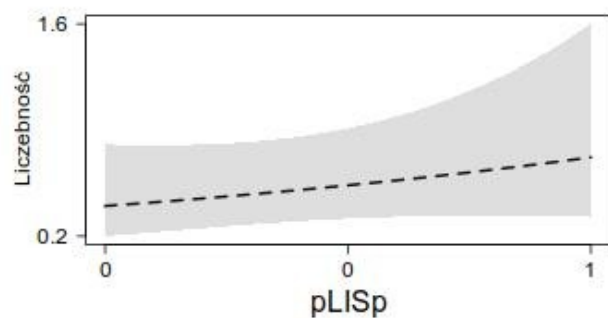
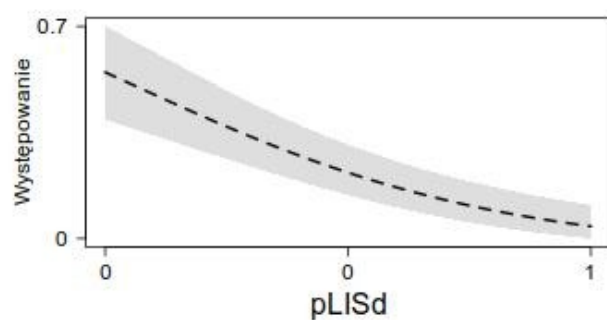
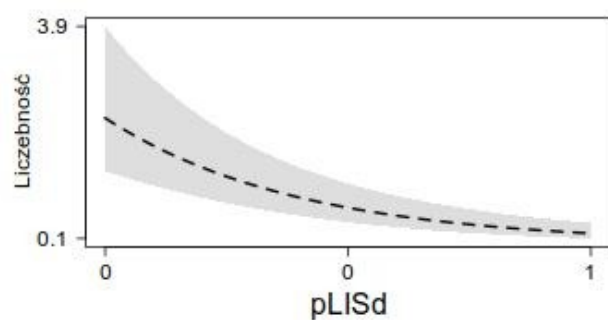
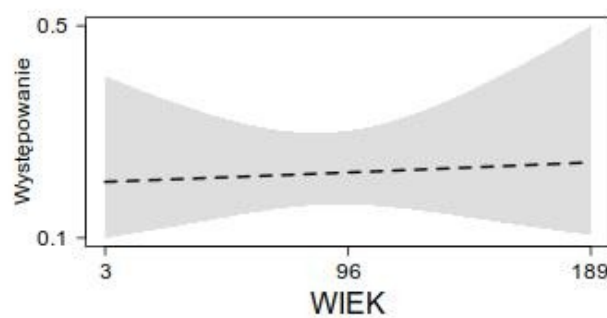
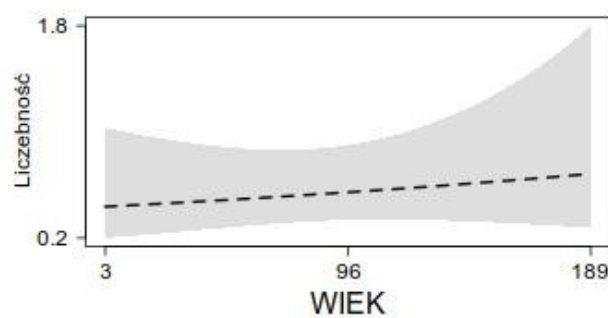
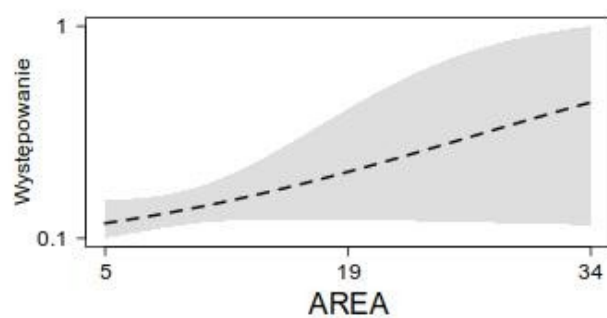
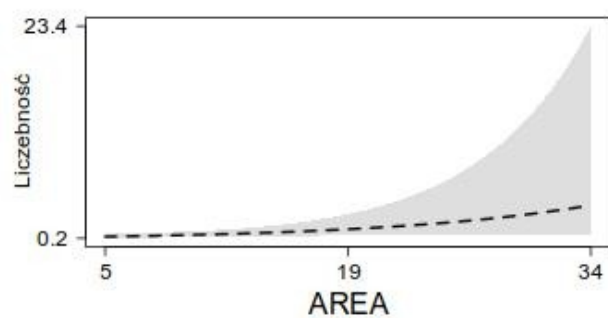


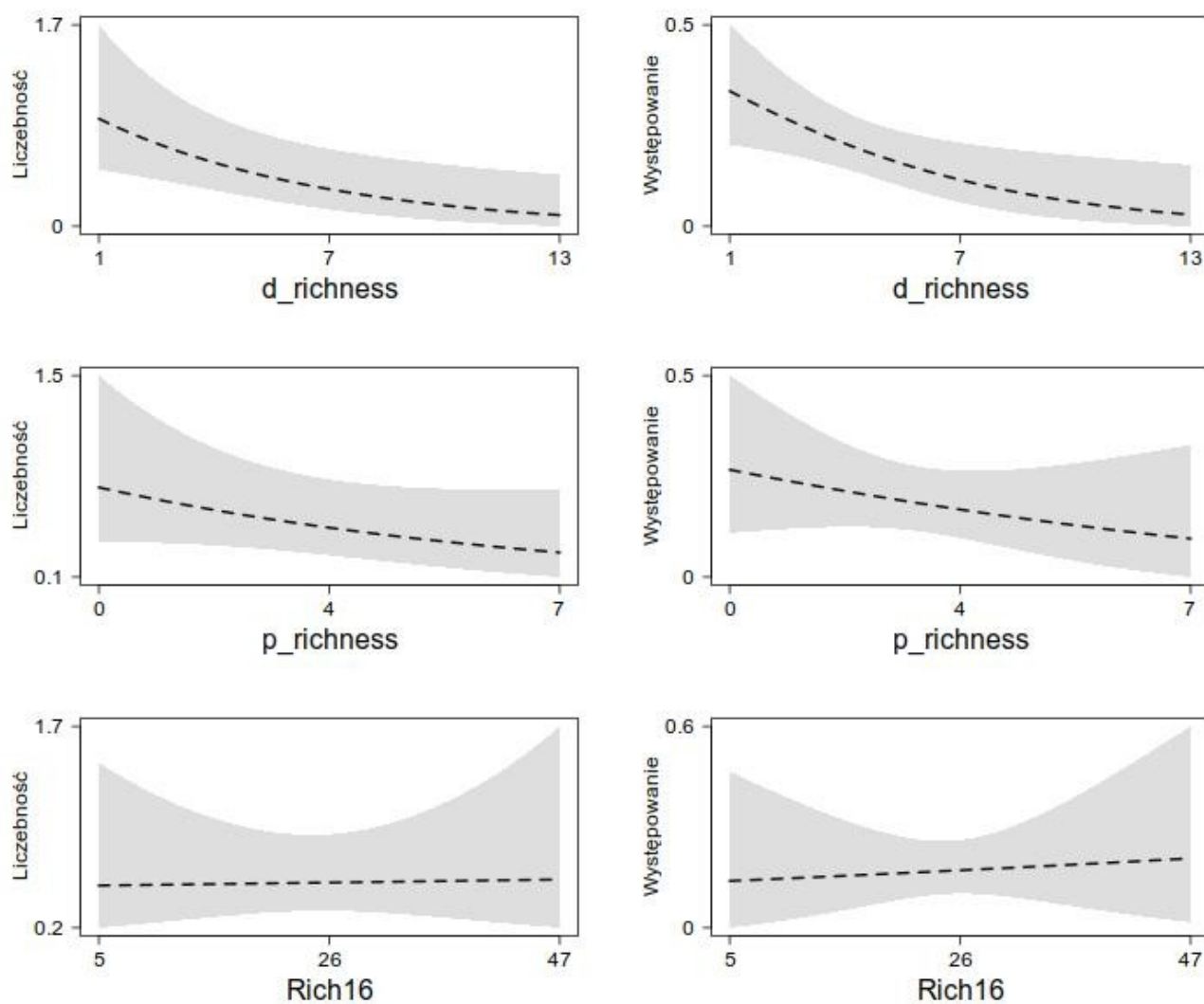


Ryc. 9. Wpływ poszczególnych zmiennych siedliskowych na liczebność (lewa kolumna) i występowanie (prawa kolumna) sikory ubogiej.

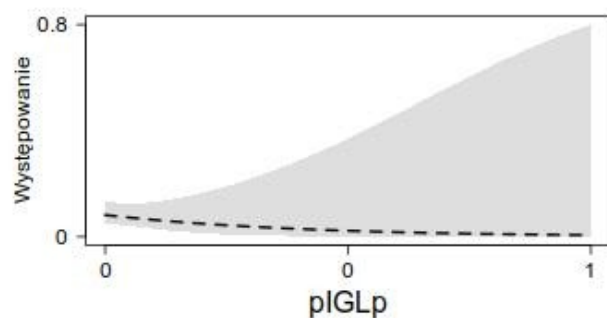
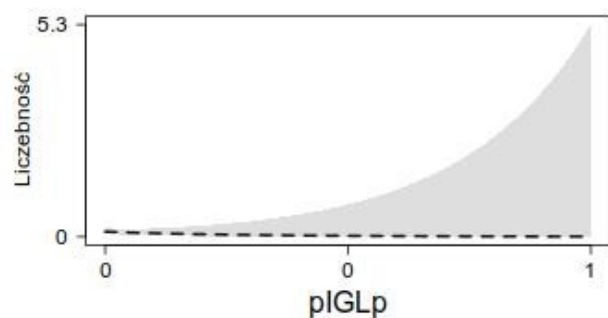
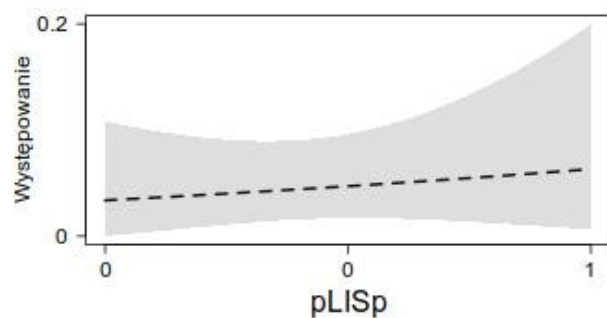
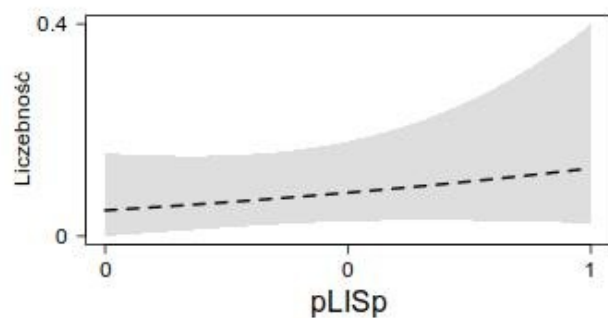
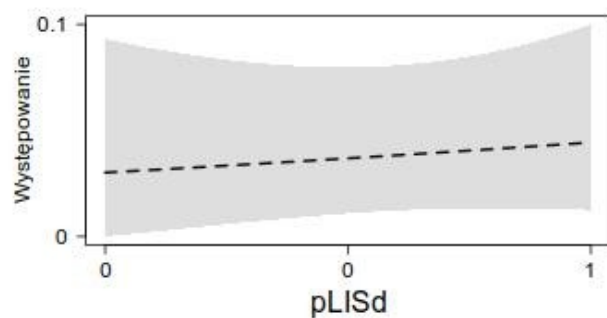
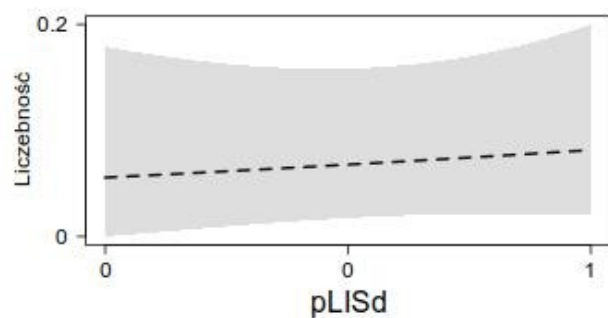
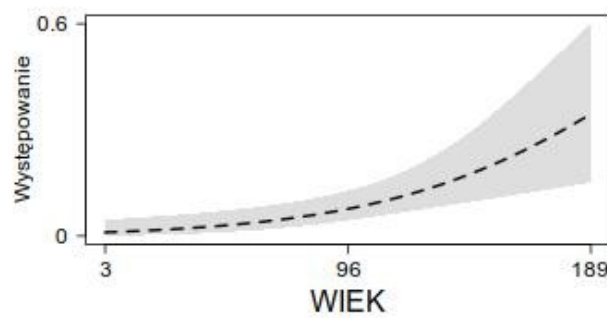
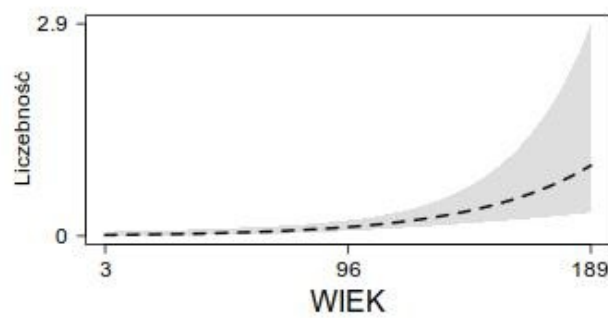
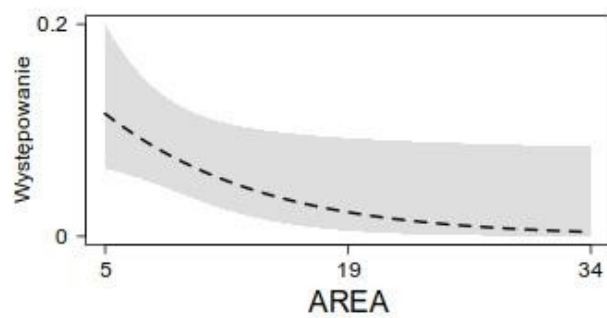
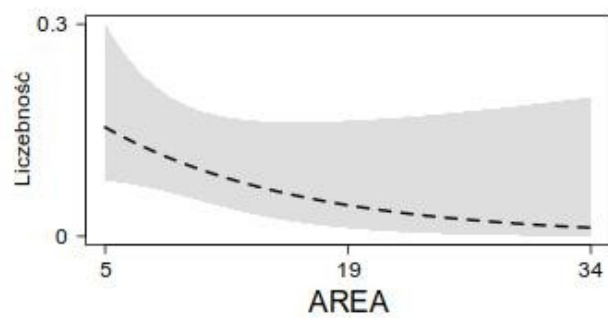


Ryc. 10. Wpływ poszczególnych zmiennych siedliskowych na liczebność (lewa kolumna) i występowanie (prawa kolumna) muchotłówki żałobnej.

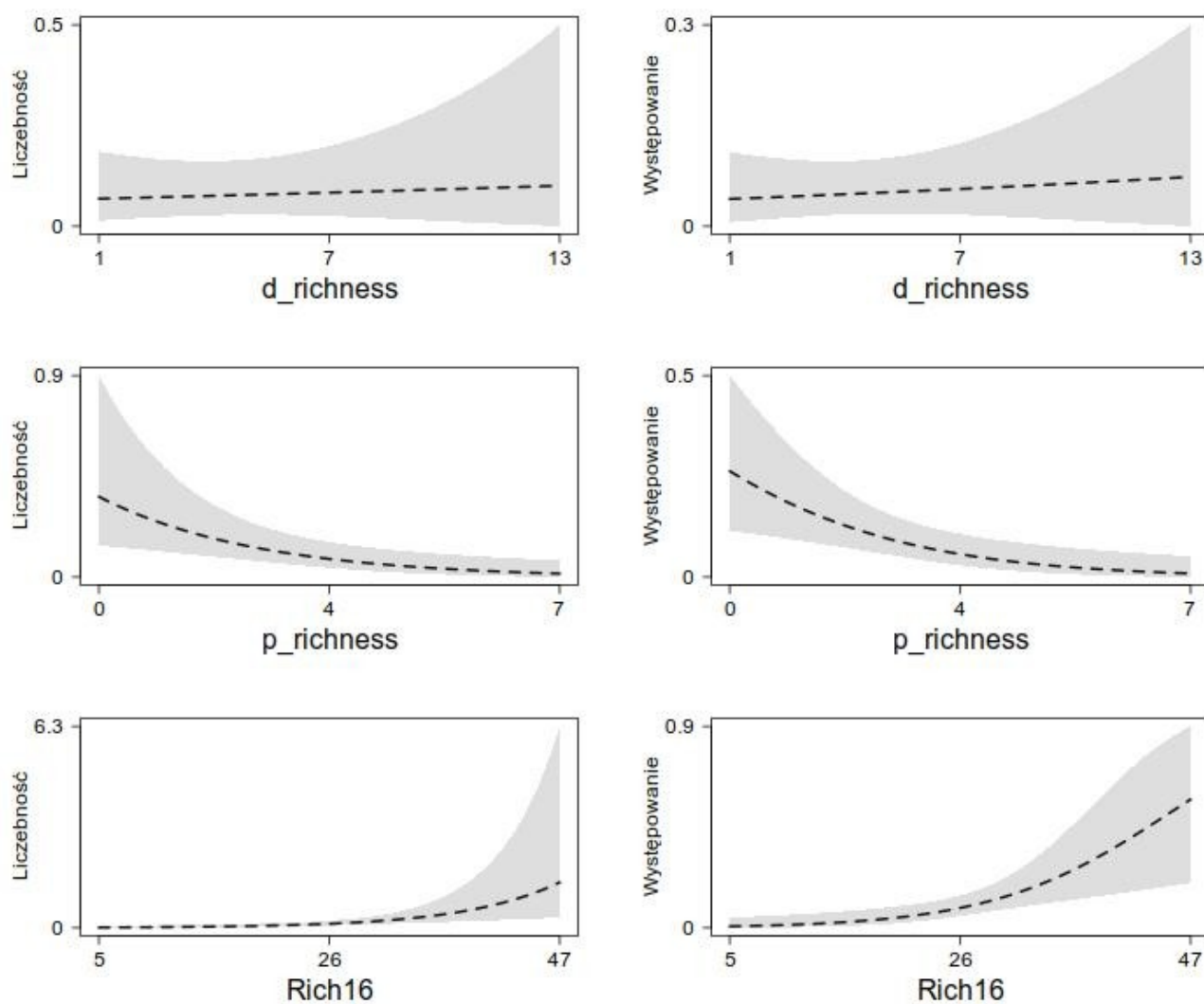




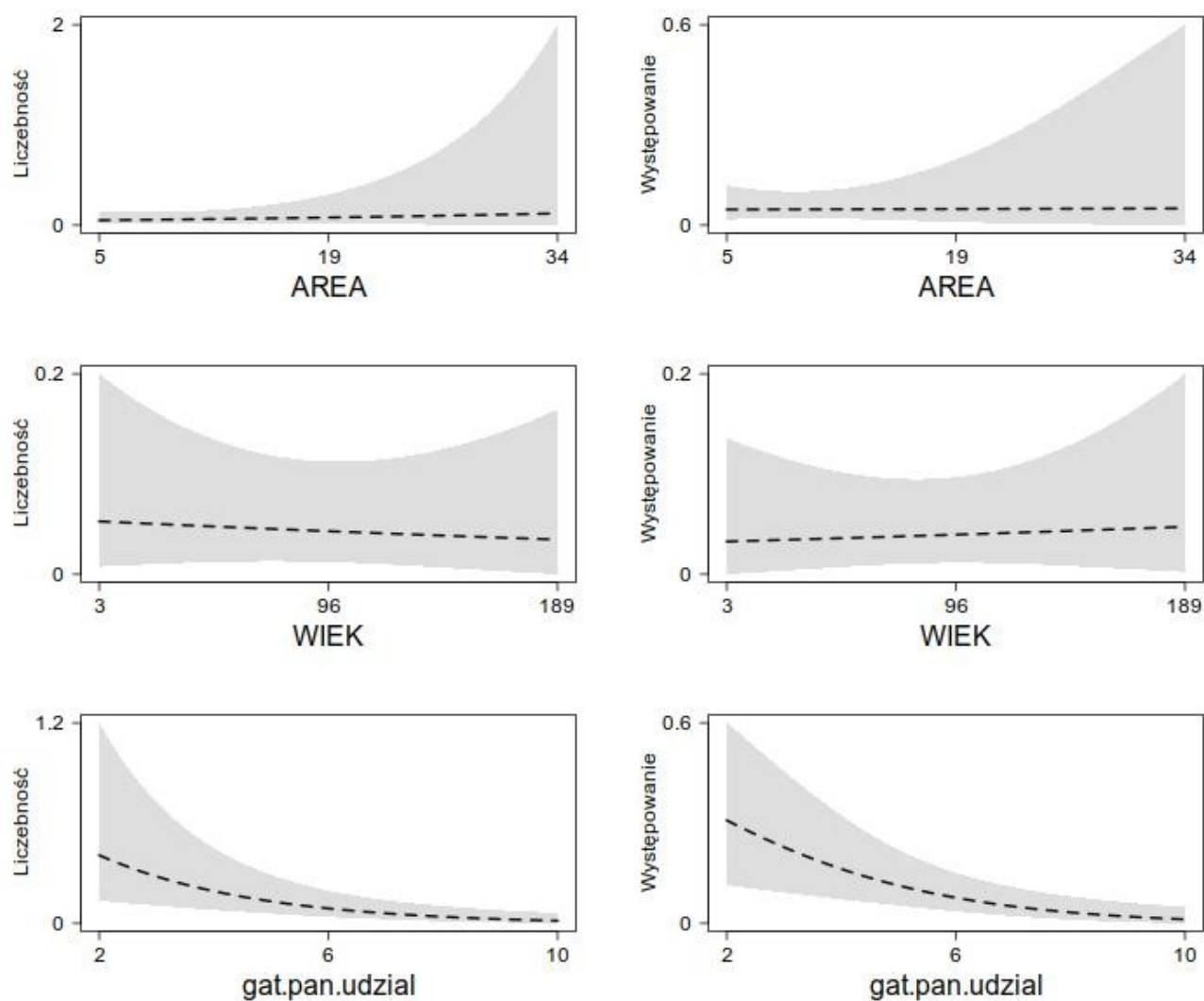
Ryc. 11. Wpływ poszczególnych zmiennych siedliskowych na liczebność (lewa kolumna) i występowanie (prawa kolumna) czubatki.







Ryc. 12. Wpływ poszczególnych zmiennych siedliskowych na liczebność (lewa kolumna) i występowanie (prawa kolumna) siniaka.



Ryc. 13. Wpływ poszczególnych zmiennych siedliskowych na liczebność (lewa kolumna) i występowanie (prawa kolumna) raniuszka.