MASTERARBEIT | MASTER'S THESIS

Titel | Title

Influences of environmental changes on the stopover behavior and physiological condition of migratory birds at the Lake Neusiedl

verfasst von | submitted by

Lisa Müllner BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 879
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Naturschutz und Biodiversitätsmanagement
Betreut von | Supervisor:	Univ.-Prof. Dott. Leonida Fusani MPhil PhD

Zusammenfassung

Diese Masterarbeit untersuchte einen historischen Datensatz (1974–2023) der Biologischen Station Illmitz und verglich die Daten mit denen der AURING-Vogelberingungsstation Hohenau-Ringelsdorf. Untersucht wurde, wie Schwankungen des Wasserstands die Aufenthaltsdauer und Populationsgröße von Zugvögeln während dem Herbstzug an der Station Illmitz beeinflussen und welche Korrelation zwischen den beiden Stationen hinsichtlich der Variation der Populationsgröße und der Unterschiede im Fettanteil festgestellt werden können. Die Themen wurden mit Hilfe von Spearman-Rangkorrelationen, generalisierte lineare Modelle mit negativ binominal Verteilung, Korrelationen mit wiederholten Messungen und Mann-Whitney-U-Tests untersucht. Die Ergebnisse zeigten, dass Schwankungen des Wasserstands keinen Einfluss auf die Dauer des Zwischenstopps oder auf die Populationsgröße hat. Die Ergebnisse des Stationsvergleichs deuteten darauf hin, dass Schwankungen in der Populationsgröße an einer Station keine Schwankungen an der anderen Station vorhersagen, aber sie können in ähnlicher Weise variieren. Beim Vergleich der einzelnen Jahre und der unterschiedlichen Fettwerte zwischen den Stationen ist eher eine artspezifische Reaktion als ein einheitlicher Standorteffekt zu erkennen. Die historischen Daten aus Illmitz sowie der Vergleich dieser beiden Lebensräume wurden auf diese Weise zum ersten Mal ausgewertet; die Ergebnisse tragen dazu bei, eine Wissenslücke in den Daten aus Europas größten Vogelschutzgebieten und künstlichen Industriekulturlandschaften mit Naturmanagement zu füllen.

Schlüsselwörter: Herbstzug, Beringungsstationen, Pegel, Populationsgröße, Stopover, Fett

Abstract

This thesis investigated a historical dataset (1974-2023) from the biological station Illmitz and compared the data with those from the AURING bird ringing station Hohenau-Ringelsdorf. Examined topics were how fluctuations of water level influence the stopover duration and population size of migratory birds during autumn migration at the station Illmitz, and what correlation can be found between the two stations on the variation of population size and the differences in fat score. The topics were approached using Spearman's rank correlations, generalized linear models with a negative binomial distribution, Repeated measures correlations, and Mann-Whitney U-tests. The main results are that water level changes do not affect stopover duration. Water levels and population size do not correlate with each other. Results of the station comparison suggested that fluctuations in population trend at one station do not predict fluctuations in the other, but they may vary in a similar way. The comparison of data on fat scores at both stations supports a species-specific variation for each year. The historical data from Illmitz, as well as the comparison of these two habitats, have never been evaluated in this way before; the findings help fill a knowledge gap in data from Europe's largest bird-protected areas and man-made industrial cultural landscape with nature management.

Key words: Autumn migration, bird ringing stations, water level, population size, stopover duration, fat score

Acknowledgment

A huge acknowledgment goes to the AURING Bird Ringing Station Managers Heinrich Froetscher, Bernhard Paces, Benjamin Seaman, Matthias Schmidt, Martin Rössler, Flora Bittermann, Barbara Waringer, and Karin Donnerbaum for collecting the data (together with volunteers), allowing me to work with it, and brainstorming with me on possible research questions. I also want to thank Ivan Maggini for his patience, valuable feedback, and for always taking the time to answer all of my questions. Furthermore, I want to thank my husband, Daniel, for listening to my thought processes in difficult times, offering advice, and having my back with all Python-related questions. Last but not least, I want to acknowledge my friends and family for always cheering me on. And especially Miriam for the endless library writing session, her valuable insights, feedback, and motivational conversations.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Migratory birds in a changing world

Today's environment is being molded by the loss of biodiversity. Recent trends move towards a homogeneous ecosystem, with a reduced avifauna, containing a few species that increase in abundance (Petras and Vrezec [2022]). Since 1980, a decline of 560-620 million birds, which represents 17-19 percent, has been estimated for the native avifauna of the European Union. However, biodiversity loss varies across habitats, with losses primarily observed in birds associated with agricultural and grassland habitats (Burns et al. [2021]). Additionally, Afro-Palaearctic migrants have been experiencing a persistent and often severe decline in population for several decades. This trend has been more negative for trans-Saharan migrants than for short-distance migrants or stationary species (Sanderson et al. [2006], Berthold et al. [1998], Burns et al. [2021], Newton [2004]). Several factors contribute to these declines. For diet specialists such as insectivores, evidence of declines in insect populations is alarming and suggests a close connection to the negative trend in bird communities dominated by that diet (Bowler et al. [2019]). A further problem is the potential of phenological mismatch between peak food availability and breeding time. In the northern hemisphere, recent rapid climate change has been linked to substantial shifts in phenology for both plants and animals, resulting in earlier optimal reproductive timing. However, not all species appear to have adjusted their breeding schedules in the same way, potentially resulting in a mismatch between reproduction and peak food availability. The degree of mismatch may vary among different species, and therefore might be greater in those experiencing population declines (Møller et al. [2008]). Migratory species may be less adaptable to alterations in resource availability at breeding grounds than resident birds (Sanderson et al. [2006]). Short-distance migratory birds show higher productivity and flexible behavior, therefore presenting a greater chance of necessary adaptation, due to the fast-changing environment, compared to trans-Saharan migrating birds. Another factor that is contributing to the decline of migratory birds is the cumulative exposure, since the wintering and breeding areas of long-distance migrants are ecologically and geographically separated from each other. Unlike resident bird species, there are more ecological and environmental factors that may influence the birds during migration (Petras and Vrezec [2022], Newton [2004]), like the stopover site (Cimprich et al. [2005]). Migration routes, wintering grounds, and breeding grounds in spring can influence the population of migrating birds. If unsuitable conditions happen at just one stage, it may create a bottleneck that could decrease the population numbers (Newton [2004]). If migration is delayed at stopover sites due to slow fattening rates, there appears to be a domino effect. Birds arrive later at the following site, where food supplies may have already been depleted. During spring migration, this domino effect could also predict a late arrival at breeding grounds and therefore a decreasing chance for breeding (Newton

[1997], Berthold [2002]). Defining the cause of any population decline is not easy, since many factors or a combination of factors may be at work across diverse guilds and species. There is a general negative trend in population numbers of European species that winter in the Afro-tropical areas in exposed habitats prone to drought. The decline is due to habitat degradation, use of pesticides and notable reduction of rainfall, which is associated with less food supplies (Sanderson et al. [2006], Newton [2004]). While most studies show that declines in migratory bird populations often happen at breeding or wintering grounds, there is growing evidence that the conditions at stopover sites may also play an important role (Newton [2004]).

1.2. Migration ecology: stopover, fueling and decision making

Millions of songbirds that breed in the Palaearctic migrate annually to Afro-tropical areas in the winter (Bairlein [1995]). The migratory journey can be described as alternating between two stages. The flight, which covers distance and consumes energy, and the stopover periods, during which energy (fuel) is stored for the next flight segment. Due to trade-offs in flight performance and fuel deposition, different stopover and flight strategies have evolved (Alerstam and Hedenstrom [1998]). In migratory birds, energy for flight is accumulated at stopover sites, where individuals decide when to depart for the next flight stage. These departure decisions determine stopover duration and fuel availability for subsequent flights. Together with the rate of fuel deposition, stopover duration is a key factor shaping migration speed and ultimately influences the success of migration (Schaub et al. [2008], Schaub and Jenni [2001]). The accessibility of food appears to be a highly important driver of stopover duration (Buler et al. [2017]). Which reflects on the fact that habitat quality is one of the most important factors influencing stopover success (Mccabe and Olsen [2015]). While population declines in migratory birds have often been explained by conditions on breeding or wintering grounds, processes occurring at stopover sites along migration routes may also be important. Although used only briefly, the large numbers of birds that are passing through can lead to a high density of migrants and intense competition. Hence, a rapid depletion of food resources can occur. The quality of a stopover site is therefore not only shaped by food availability but also by predation risk, disturbance, and other threats (Newton [2004]). For migration to be successful, birds need to store energy in the form of fat for migratory flights. Especially for the trans-Saharan journey, an appropriate amount of fuel is necessary before the journey (Bairlein [1995]). For scientific comparison, a fat score, that serves as a critical measure of its energy reserves, between 0-8 has been defined (Kaiser [1993]). Through this scale, researchers have been provided with a visual proxy of the amount of stored fat available to fuel the migratory journey (Kaiser [1993], Biebach et al. [1986], Bairlein [1985]). Fat score influences the stopover decision; birds arriving at a site with low amounts of fat normally stay longer to replenish their stores, whereas birds with high fat scores may rest at a site or depart again immediately (Biebach et al. [1986], Cherry [1982], Maggini et al. [2020]). Studies showed that in some species the rate of fattening differs between sites or even years, depending on local food sources, or higher density of migrating birds (Cherry [1982], Moore and Yong [1991], Kelly et al. [2002], Whalen and Watts [2002]).

1.3. Wetlands as stopover habitats

Human action in the form of habitat loss or re-naturalization of areas can influence the bird numbers accordingly (Newton [2004]). To design a proper conservation strategy, it is of utmost importance to identify the mechanisms through which environmental changes influence population size and distribution (Mclean et al. [2016]). Previous studies have indicated that inter-annual variation in population size and adult survival can be correlated with fluctuations in rainfall, which may be linked to changes in winter food availability (Newton [2004]). In this context, water level may represent a meaningful predictor, as it is likely to reflect hydrological conditions associated with variation in resource availability (Nemeth and Dvorak [2022]). Wetlands, particularly reed-dominated habitats, have been identified as important stopover sites for many passerine birds during migration. These habitats provide food resources, shelter, and suitable micro-climatic conditions that are critical for refueling during migration. (Arizaga et al. [2013], Trnka and Prokop [2006], Bozó et al. [2018]). Several migratory species like the European Robin (*Erithacus rubecula*) and the Common Chiffchaff (*Phylloscopus collybita*) can be found in scrubby habitats at the beginning of migration season, but increasingly shift to reed beds at the end of migration, between October and November (Bairlein [1983]). Wetland habitats are also of considerable importance for species that are not strictly specialized in them, meaning that wetland loss can negatively affect wetland specialists and habitat generalists equally (Petras and Vrezec [2022]).

1.4. Ringing stations in east Austria: biological station Illmitz and AURING bird ringing station Hohenau-Ringelsdorf

No other taxon is probably better known and studied than the avifauna (Gregory and Strien [2010]). Bird ringing is one of the tools used to understand the full significance and complexity of bird migration (Alerstam and Hedenstrom [1998]), population size, and geographical trends (Petras and Vrezec [2022]). A suitable study area for ringing stations must follow certain characteristics: relatively stable vegetation, habitat diversity that supports birds with different ecological requirements, and long-term existence assurance (Berthold et al. [1998]). In this thesis, the main focus is on data from the biological station Illmitz (ILLM), as well as on comparisons with data from the bird ringing station Hohenau-Ringelsdorf (HOH), which are approximately 110km apart. ILLM (48.58 N, 16.91 E) is located on the eastern shore of Lake Neusiedl, which lies in the middle of the national parks Lake Neusiedl/Fertö in Austria/Hungary. Lake Neusiedl is one of the largest endorheic steppe lakes in Central Europe, and it contains the largest bird-protection area of Europe. Distinct key habitat features make this area highly specialized and appealing to birds. Lake Neusiedl is surrounded with an extensive reed belt, which forms one of the largest continuous reed habitats in Europe (Pretzer et al. [2017]) and is considered to be of upmost importance for the conservation of species depending on this ecosystem (Nemeth and Dvorak [2022]). Furthermore, the Illmitz area contains saline ponds (Dvorak et al. [2012]), wet meadows, sand steppes (Zulka et al. [2024]), and vineyards (Karner-Ranner et al. [2008]).

HOH (47.77 N, 16.76 E) is located in the March-Thaya-Auen, close to the border of Lower Austria and Slovakia (Zuna-Kratky et al. [2022]). The habitat in this area is unique because it is an industrial cultural landscape that developed from the former sedimentation basins of the sugar factory of Hohenau (Zuna-Kratky et al. [2022, 2023]). Hence, this habitat holds extremely special key features like man-made basins that contain nutrient-rich ruderal flora (Rössler and Zuna-Kratky [1995]). HOH carries out habitat management by focusing on maintaining various successional stages, ranging from open mudflats and shallow water for waders to dense shrubs and elderberry bushes for songbirds (Hillbrand [2004]). The station also actively implements water management at the basins due to the lack of natural flooding from the March River. This helps preserve reed beds and creates open water habitats for breeding and migrating waterbirds (Zuna-Kratky et al. [2022]). In summary, these two stations and their habitats show different landscapes, vegetation, and water types (see Table 1.1).

Table 1.1.: Key differences between HOH and ILLM

feature	HOH	ILLM
water	man-made basins, water management	steppe lake, saline ponds
primary vegetation	ruderal vegetation, mixed-scrub	reed belts, wet meadows
landscape matrix	intensive agriculture	wetland, sand steppes

1.5. Motivation, research questions and hypothesis

ILLM has collected data on ringed birds during autumn migration since the 1970s. It represents a rare, long-term dataset over a time frame of 49 years, however not all years have been digitized yet. 24 sampling years can be used to address questions regarding themes such as stopover duration, fat score, and population size in relation to water level. Data collection at HOH began in 1994. The lack of recaptures between the stations indicates that they are not used by the same individuals during one migratory season. This observation is useful to address the question whether the two habitats may be selected by migrating individuals preferentially.

1.5.1. Biological station Illmitz - Lake Neusiedl

In recent years, studies have focused on the lower water levels of Lake Neusiedl in terms of climatic influences on the lakes future. Factors like rainfall and temperature have been studied, and showed that the current water level is influenced not only by rain and temperature, but also by the water level of the previous month (Hackl and Ledolter [2023]). The main topic of this thesis are birds during autumn migration. I examined correlations between water level, stopover duration, and population size of migratory birds. This leads to the first two research questions and hypotheses.

Research question and hypothesis 1

How does the water level of Lake Neusiedl influence the stopover duration of the European Reed Warbler - *Acrocephalus scirpaceus* and the Sedge Warbler - *Acrocephalus schoenobaenus* at the biological station Illmitz over the time period of almost 49 years (1974-2023)?

It can be predicted that fluctuations of the water level at Lake Neusiedl will influence the stopover duration of these migratory reed species. Lower water levels are predicted to correlate with shorter stopover times and reduced abundance due to decreased food availability or habitat suitability (Newton [2004], Morrison et al. [2010], Payevsky [2006]).

Research question and hypothesis 2

What is the relationship between the water level at the biological station Illmitz (Lake Neusiedl) and the population size, reflected by the number of captured migratory birds?

Multidecadal ringing and atlas work show widespread declines in Afro-Palaearctic migrants and a gradual homogenization of European avifaunas (Burns et al. [2021], Petras and Vrezec [2022], Morrison et al. [2010]). Lower or more variable water levels at Lake Neusiedl in recent decades (Hackl and Ledolter [2023]) may have reduced habitat breadth and food resources during autumn migration, disproportionately affecting migratory species. Therefore, it can be predicted that higher water levels show a positive correlation with the population size of migratory birds in general, while lower water levels show a negative correlation (Newton [2004], Bieringer et al. [2013]).

1.5.2. A comparison of two bird ringing stations in east Austria

The differences in the habitats of ILLM and HOH (see Table 1.1), as well as the stations locations allow for an interesting comparison. The chosen parameters are population trends and fat score during stopover. Therefore, two more research questions and hypotheses have been drafted.

Research question and hypothesis 3

In what way do the population trends, reflected by the number of captured migratory birds, at the biological station Illmitz and the AURING Hohenau-Ringelsdorf bird ringing station vary?

Regional-scale drivers, such as climate change, and local-scale processes, such as habitat change, suggest interacting patterns that produce trend variation in populations (Morrison et al. [2010]). Therefore, there are two different predictions: If the variation in numbers follows the same pattern in both stations,the observed population trends are more likely to be global. If the pattern varies differently in the stations, the trends may be due to local differences in habitat suitability.

 1. Introduction

Research question and hypothesis 4

Which differences can be observed between the migratory birds at the biological station Illmitz and the Hohenau-Ringelsdorf bird ringing station in terms of fat score?

The selection of stopover sites of migratory birds is closely related to the environmental conditions of these habitats. In particular, food resources influence the choice of stopover site (Schaub and Jenni [2000]). The available fat reserves of individual birds influence their decisions about where to land during migration. Birds with low fat reserves need to refuel, whereas birds with high amounts of fat do not need high food availability at the stopover site (Biebach et al. [1986]). This suggests that it is vital for migratory birds with low fat scores to select a stopover site with sufficient available food. Both the biological station Illmitz, which was designated in 1982 as a Ramsar Site (Dick [1987]), and the bird ringing station Hohenau, which applies water management (Zuna-Kratky et al. [2022]), provide viable stopover and fueling options (Schaub and Jenni [2000], Nemeth and Dvorak [2022]). Fat score has two possible predictions: if one station provides a better habitat than the other, more birds with lower fat scores are expected to be caught at this site. If the habitats are of similar quality for the birds, both stations should show equal fat distribution.

2. Methods

2.1. General data information

I conducted this thesis with data from the Austrian Ornithological Center, which coordinates ringing data collection at ILLM and HOH. The original data for this work was collected daily during the autumn migration periods (July to October) between the years 1974-2023 in ILLM and 1994-2023 in HOH. Birds were captured using mist nets, which were checked hourly from sunrise to sunset. On days with rain, strong winds, or extreme heat, the nets were closed. At both stations, standard measurement procedures, including measurements of body mass and fat score (only after 1994), were used on the captured birds. To indicate the capture history for an individual bird, the trapping status (r-type) was recorded. If an individual was caught without a ring, it has been in the net for the first time and became the status "e" in the r-type column. If the same bird, in the same year, has been caught again, it became the status "w" for recapture. However, if a bird with a ring has been caught the first time in a new season, it received a "k", which means first recapture in a new migratory season (Bairlein [1995]).

2.2. Data cleaning

For the data cleaning process I used the Python programming language in the Google Colab online platform. The original data was divided into 31 different .xls or .xlsx files, and in total consisted of 452764 rows before cleaning and filtering. First, standardization was required, as not all files used the same date representation or formatting. To ensure consistency when combining data from multiple sources, I defined a fixed list of column names that all Excel files had to follow. I then specified the data type for each column. Since the files also differed in how the species names were documented (German, scientific name, or short scientific form), I created a species mapping to avoid losing data in the combined file. As a last step, I combined the files into a single pandas DataFrame. To ensure data quality, I first assessed missing rows and invalid dates and then went through a filtering process. For this thesis, I selected mist nets that have been in stable use over all years, and removed capture data from all extra added nets in various years. To filter out possible breeding bird individuals, I kept only data from the months during autumn migration (July to October). In addition, if a bird had a stopover duration longer than two weeks, that individual was filtered out, too. After the main data cleaning was completed, I tested the code on a test file that included all the rules it should execute on the original data. This ensured that no data was filtered by mistake. The cleaned dataset now contains 241606 rows and 31 columns. Out of these 31 columns, eight are being used in the analysis (station, ring number, r-type, species, date of capture, time of capture, fat, water level). With the new cleaned dataset I then checked which years had enough data to be defined as complete data ready for testing. Years that had less months than the chosen four months (July-October) were skipped (see Data cleaning pipeline B.1 in appendix: Python Code). After this last step of data cleaning, 24 years, out of the possible time span of 49 years, met all criteria for the analysis. 20 years in this time frame are not yet digitalized (1984-1998, 1994-2003, 2012-2016), and 6 years did not have enough data to meet the defined criteria (2004-2008, 20011). For each research question, adapted to the requirements of the analysis, additional data cleaning was performed, which is documented in the corresponding sections.

2.3. Data analysis

I conducted the statistical analysis in Google Colab with Python. Throught the whole thesis the statistical significance was assessed at $p < 0.05$. Since only the short scientific name for the species was used in tests, figures and tables, I created a table that will serve as a reference source (see Table 2.1).

Table 2.1.: Reference source for all short scientific names of species used in this thesis

short scientific name	scientific name	english name
ACRSCI	*Acrocephalus scirpaceus*	European Reed Warbler
ACRSCH	*Acrocephalus schoenobaenus*	Sedge Warbler
ACRARU	*Acrocephalus arundinaceus*	Great Reed Warbler
ACRMEL	*Acrocephalus melanopogon*	Moustached Warbler
EMBSCH	*Emberiza schoeniclus*	Common Reed Bunting
ERIRUB	*Erithacus rubecula*	European Robin
HIRRUS	*Hirundo rustica*	Barn Swallow
PARCAE	*Cyanistes caeruleus*	Eurasian Blue Tit
PHYCOL	*Phylloscopus collybita*	Common Chiffchaff
SYLATR	*Sylvia atricapilla*	Eurasian Blackcap

2.3.1. Biological station Illmitz - Lake Neusiedl

Research question 1

Recapture numbers across all years with complete data defined the years that were used in the analysis (24 years: 1974-1983, 1989-1993, 2009-2010, 2017-2023). Species had to fulfill the following conditions: > 1000 recaptures over all years and represented at least once in all 24 years. ACRSCI and ACRSCH met these requirements (see Table 2.2).

First, I calculated the minimum stopover duration, which is the time between the first and last capture of an individual at a site (Bairlein [1985]). Then I performed a Shapiro-Wilk test over all years with the original data and with log-transformed data. For additional data points, I also

Table 2.2.: Numbers of recaptured migratory birds over all years. Green > 30, yellow 10 - 30, red < 5 recaptures.

species	total_unique	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1989	1990	1991	1992	1993	2009	2010	2017	2018	2019	2020	2021	2022	2023
ACRSCI	3982	256	213	348	171	88	134	167	217	354	275	188	207	66	320	180	167	94	65	103	70	45	77	131	46
ACRSCH	1491	45	57	268	92	58	48	58	69	78	138	96	39	28	158	64	14	15	1	46	5	9	25	47	33
PARCAE	742	10	12	5	7	8	4	6	18	0	32	91	42	19	56	53	17	71	89	15	27	21	44	34	62
ACRARU	395	34	31	44	37	22	10	12	22	16	13	22	10	5	31	10	13	18	2	2	3	8	8	20	5
ACRMEL	187	23	12	9	20	2	6	10	1	46	16	8	5	2	16	3	2	0	1	0	4	0	1	0	0
EMBSCH	180	18	9	10	31	4	5	7	15	10	4	18	6	6	5	11	1	3	1	1	2	0	5	6	2
LOCLUS	122	4	14	6	7	2	12	1	1	17	2	9	3	7	8	5	2	3	2	5	3	4	1	4	0
PHYCOL	61	3	0	0	1	0	0	1	0	0	3	0	0	0	3	1	1	1	3	6	2	4	3	22	7
LUSSVE	45	9	1	4	3	1	1	1	7	4	3	0	3	1	0	0	0	0	0	1	1	1	0	4	0
PANBIA	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	5	0	3	1	2	1	10	1
ALCATT	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	2	8	8	6	4	2	0
ERIRUB	29	0	1	1	2	0	0	1	2	2	0	1	2	0	0	0	0	1	1	2	0	1	1	4	7
LANCOL	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	2	0	0	2	8	2	7
ACRPAL	26	2	0	19	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	2	0
REMPEN	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	1	4	2	0	1	1	0	1
SYLCOM	21	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	1	4	2	0	6	4
TROTRO	9	0	0	0	2	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2	0	2	0	1
JYNTOR	9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	4	1
PARMAJ	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7	0	0
MOTALB	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	0	0	0	2	0	0
SYLATR	6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	3	0	0	0	1
PHYTRO	6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	2
HIRRUS	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	3	0	0
PRUMOD	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	1
SYLCUR	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	1
AEGCAU	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
STUVUL	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
SAXRUB	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
IXOMIN	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
MOTFLA	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
SYLNIS	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
FICHYP	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
LUSMEG	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
PASMON	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

ran the test separately for each year and for the monthly median. To test the linear relationship between each species monthly median stopover duration and water level, I used scatterplots. As a final analysis, I performed a Spearman's rank correlation.

Research question 2

To determine the relationship between water level and the population size of captured migratory birds at ILLM, I used the dataset spanning 24 years (1974-1983, 1989-1993, 2009-2010, 2017-2023). For this research question, I selected species that met the following two conditions. Firstly, they were present in all selected years. Secondly, the total number of captures (recaptures in the same season were not counted) across all 24 years was > 2000 (see Table 2.3). The selected focus species were: ACRSCI, ACRSCH, PARCAE, EMBSCH, ACRARU, ACRMEL, PHYCOL,and ERIRUB.

Table 2.3.: Number of captured migratory birds at ILLM, without recapture data, over 24 years (1974-1983, 1989-1993, 2009-2010, 2017-2023).

species	population size
ACRSCI	51245
ACRSCH	29204
PARCAE	14838
EMBSCH	11468
ACRARU	5691
ACRMEL	5573
PHYCOL	3320
ERIRUB	3090

As a first step, I performed a Shapiro-Wilk test. Based on the results, I then conducted a

Spearman's rank correlation and, for further data exploration, a generalized linear model (GLM) with a Poisson distribution. Due to an overdispersion, I decided to use a negative binomial GLM as a robust alternative for count data with overdispersion. I plotted the GLM results for each species for a better visual interpretation with the original data points: the GLM predicted mean and a confidence interval of 95 percent.

2.3.2. Bird ringing station comparison

Research question 3

To examine how the population trend of captured migratory birds varies at ILLM and HOH, only years with data available at both stations were included (2009-2010, 2017-2023). For this research question I selected species that were found at both stations in all 9 years and for which the total number of captures across all years (recaptures were not counted) was > 500 (see Figure 2.1). The selected focus species were: ACRSCI, ACRSCH, PARCAE, EMBSCH, ACRARU, PHYCOL, SYLATR, and HIRRUS.

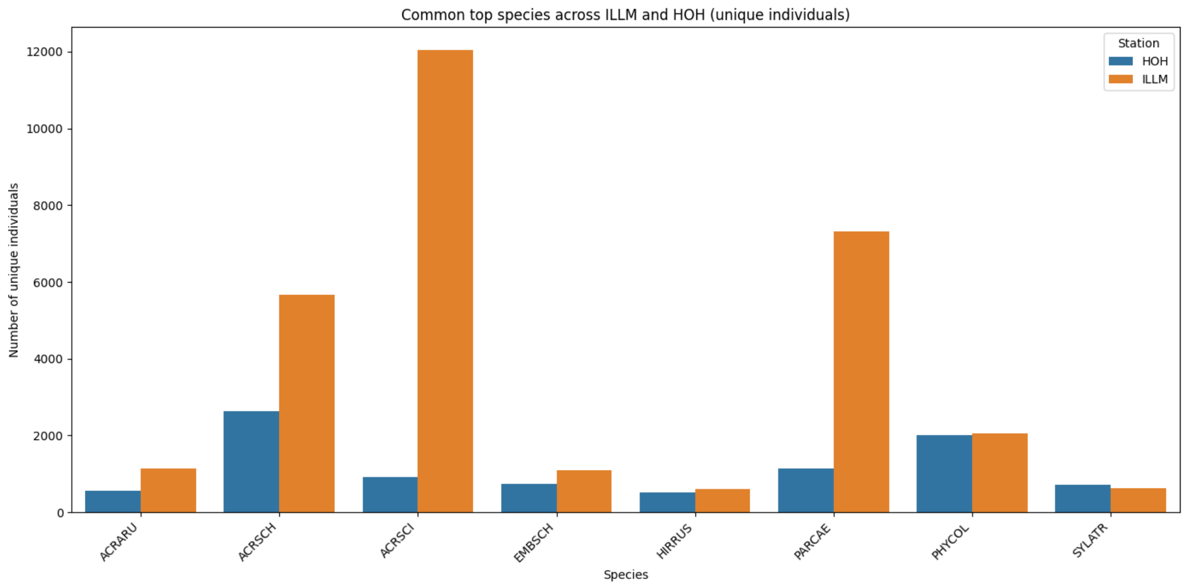

Figure 2.1.: Common top species at ILLM and HOH

First, I performed a Shapiro-Wilk test. Based on the results, I conducted a Spearman's rank correlation across all years. Since I was only able to test the correlation for nine years, there were also only nine data points for each species. However, yearly aggregations tend to smooth out possible monthly patterns. Hence, I decided to also perform a Spearman's rank correlation on monthly data for more data exploration, a higher statistical power, and an easier detection of underlying correlations between the sites. The number of data points for monthly correlations varied between 21 - 36 points depending on the species. This is due to the fact, that some species were present at both stations in all years, but not in every month (see Table 2.4).

 2. Methods

Table 2.4.: Spearman's rank correlation with monthly data

species	data points
ACRSCI	36
ACRSCH	36
PARCAE	29
EMBSCH	35
ACRARU	29
PHYCOL	34
SYLATR	32
HIRRUS	21

For visual interpretation, I conducted time series plots from the original data. Since the length of the nets being used for catching birds differs between the stations (49 m in Illmitz, 144 m in Hohenau), I created the time series plots with secondary y-axis to accurately visualize the proportion of birds caught. This serves as a standardization of the data, with the positive effect that the original number of unique individuals is visible in the plot for better evaluation. As a final step, I calculated a Repeated measures correlation (RMC) for the annual data to assess whether a trend exists. The RMC accounts for the trend of each species over the years and tests if one data point follows another, but it does not determine its' nature.

Research question 4

To test whether there is a difference in the fat scores of captured migratory birds at ILLM and HOH, only years with available data at both stations were chosen (2009-2010, 2017-2023). Since ILLM collects data every day and HOH only approximately four days a week, I chose to only include days were both stations have documented birds. This data cleaning step ensured a more accurate comparison. Furthermore, only species with > 500 captures over all nine years were selected (recaptures were not counted). The focus species were: ACRSCI, ACRSCH, PARCAE, EMBSCH, ACRARU, and PHYCOL. The two stations also use different accuracy in measuring fat scores (ILLM to unit, HOH to the half unit). I equalized them by following the definition given by Kaiser and grouped each species depending on it's fat score into the main fat classes between 0-8 (Kaiser [1993]). To address differences in fat score between the stations, I created a plot for visual interpretation over all years. Because the net length for catching birds varies across stations, as in research question 3, I created the figure with a secondary y-axis. The two y-axis are presenting the proportion of the caught birds and their fat scores. As a last step, I chose a Mann-Whitney U-Test, om account of the fact that the data was discrete and ordinal. I performed the test on the combined dataset and for each year separately to assess potential inter-annual variability. For better visual interpretation I then created time series plots with the results of the mean fat score for each year and species.

3. Results

3.1. Biological station Illmitz - Lake Neusiedl

3.1.1. Research question 1:

The Shapiro-Wilk test over all years with the original data and with log-transformed data did not have a normal distribution. The results of the Shapiro-Wilk test for each year separate and for data with the monthly median, also revealed no normal distribution (see Table A.1, Table A.2, and Table A.3 in appendix). The scatterplots illustrated a simple cloud, and no linear relationship for both species (see Figure 3.1, and Figure 3.2).

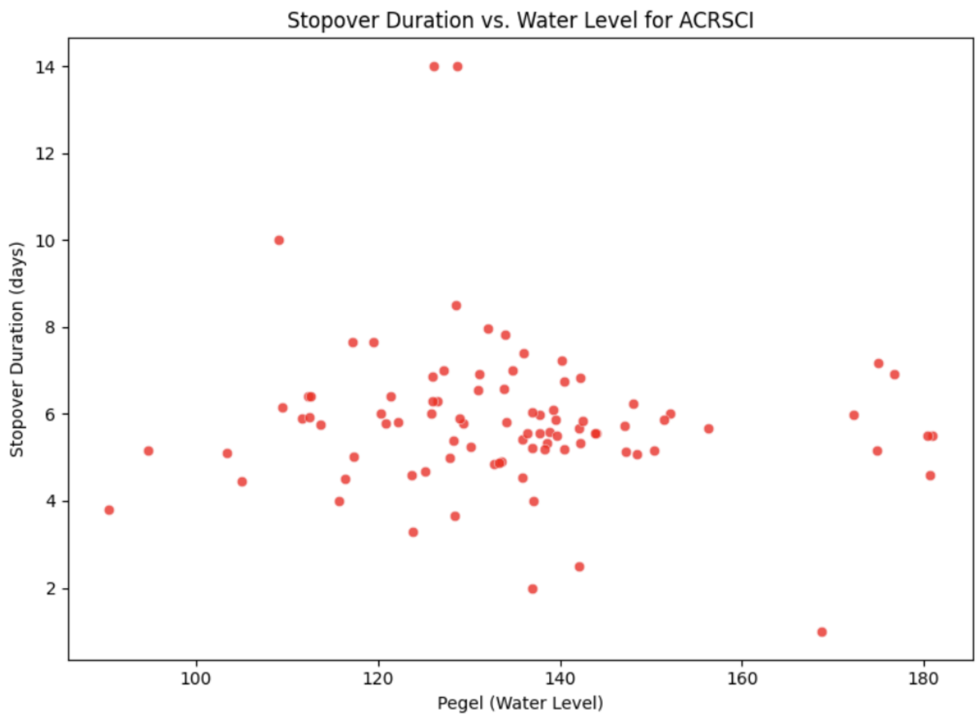

Figure 3.1.: Scatterplot: stopover duration vs. water level for ACRSCI

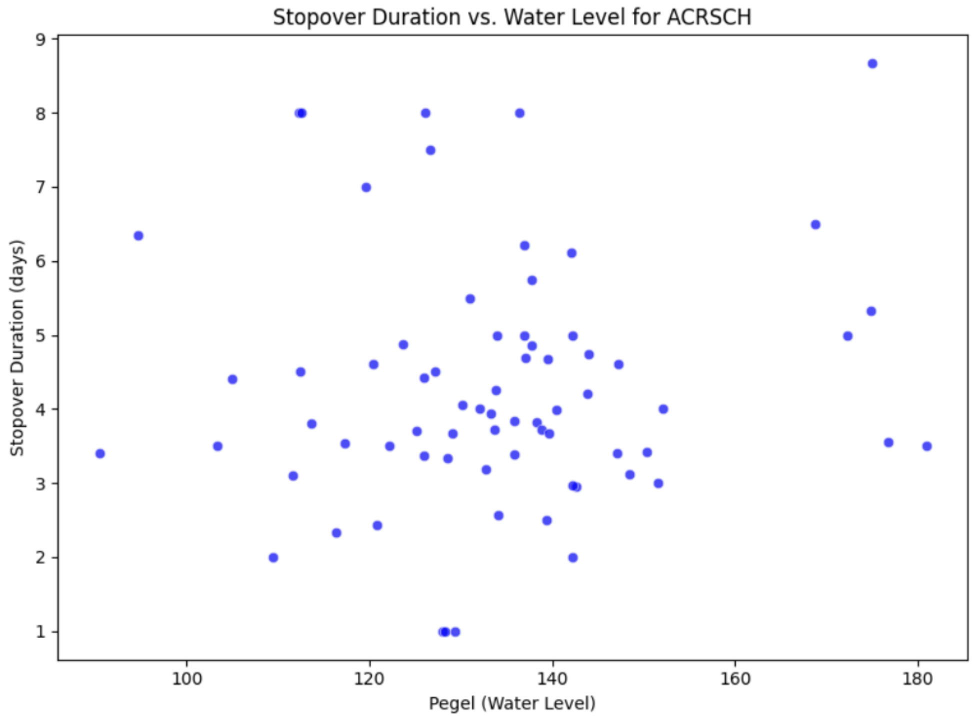

Figure 3.2.: Scatterplot: stopover duration vs. water level for ACRSCH

The results of the Spearman's rank correlation demonstrated a negative correlation for ACRSCI and a positive correlation for ACRSCH. Neither correlation was statistically significant (see Table 3.1).

Table 3.1.: Spearman's rank correlation for ACRSCI and ACRSCH

species	correlation	p-value
ACRSCI	-0.0723	0.5007
ACRSCH	0.0564	0.6430

3.1.2. Research question 2:

The Shapiro-Wilk test results for population size per species showed no normal distribution (see Table A.4 in appendix). The Spearman's rank correlation revealed a minimal negative correlation for PHYCOL and positive correlations for the remaining species. However, the correlation was for neither species statistically significant (see Table 3.2).

Table 3.2.: Spearman's rank correlation results

species	correlation	p-value
ACRSCI	0.073913	0.731419
ACRSCH	0.163478	0.445297
PARCAE	0.012174	0.954977
EMBSCH	0.070450	0.743580
ACRMEL	0.306220	0.145583
ACRARU	0.200870	0.346621
PHYCOL	-0.021314	0.921256
ERIRUB	0.271798	0.198862

The generalized linear model (GLM), fitted with a Poisson distribution, showed statistically significant p-values (see Table A.5 in appendix). However, the calculated dispersion parameters were significantly greater than 1 for each species (see Table 3.3).

Table 3.3.: Dispersion parameter results for GLM models with Poisson distribution

species	dispersion parameter
ACRSCI	598.4492
ACRSCH	568.9788
PARCAE	452.7756
EMBSCH	346.6350
ACRMEL	145.4744
ACRARU	117.5476
PHYCOL	64.1445
ERIRUB	39.9335

Therefore, a negative binominal GLM was performed. The results showed, similar to the Spearman's rank correlation, that no species displayed significant p-values (see Table 3.4).

Table 3.4.: Generalized linear model with negative binomial distribution

species	coefficient	p-value
ACRSCI	-0.0015	0.895
ACRSCH	-0.0027	0.815
PARCAE	0.0053	0.648
EMBSCH	-0.0135	0.247
ACRMEL	0.0131	0.261
ACRARU	0.0007	0.950
PHYCOL	-0.0057	0.622
ERIRUB	0.0119	0.306

When comparing the GLM plots of the Poisson distribution and the negative binomial distribution, the first analysis clearly showed an overdispersion, with many data points outside the very narrow confidence interval (see Figure A.1 in appendix). The negative binomial GLM plots revealed wider confidence intervals, which better reflect the high variability observed in the count data. The species ACRSCI, ACRSCH, and EMBSCH showed a negative relationship with the yearly average water level. Whereas the prediction curve for PHYCOL remained only slightly negative. PARCAE, ACRMEL, ACRARU, and ERIRUB mostly retained a positive relationship with water level (see Figure 3.3).

 3. Results

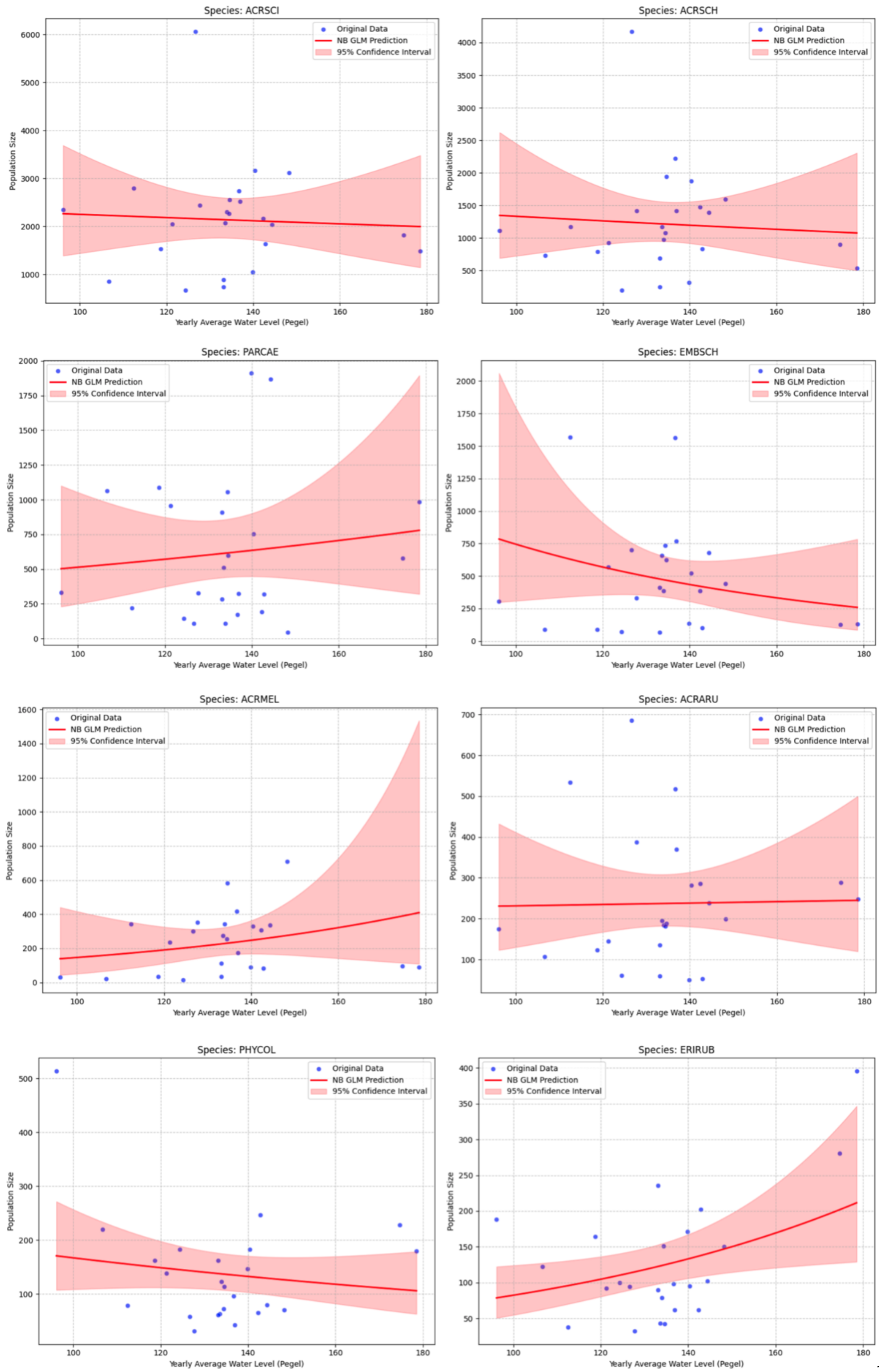

Figure 3.3.: Plotted negative binomial distribution GLM - predictions with 95 percent confidence intervals vs. actual data for each top species. Y-axis: population size, x-axis: yearly average water level

3.2. Bird ringing station comparison

3.2.1. Research question 3:

The results of the Shapiro-Wilk test showed that some, but not all species, were normally distributed (see Table A.6 in appendix). However since not all were normally distributed I chose to do a Spearman's rank correlation for better comparison. The Spearman's rank correlation with yearly data revealed positive and negative correlations for the number of captured species at both sites. ACRARU, ACRSCI, ACRSCH, and PHYCOL had a moderate positive correlation. Whereas EMBSCH and PARCAE displayed negative correlations. SYLATR and HIRRUS were close to 0 and showed no tendency for correlation at both stations. All species were not statistically significant (see Table 3.5).

Table 3.5.: Spearman's rank correlation: yearly results

species	correlation coefficient	p-value
ACRARU	0.583333	0.099186
ACRSCI	0.500000	0.170471
ACRSCH	0.483333	0.187470
EMBSCH	-0.334731	0.378595
PHYCOL	0.234312	0.543967
PARCAE	-0.166667	0.668231
SYLATR	0.083333	0.831214
HIRRUS	-0.016737	0.965913

Table 3.6.: Spearman's rank correlation: monthly results

species	correlation coefficient	p-value
ACRARU	0.644454	0.000161
ACRSCI	0.702699	0.000002
ACRSCH	0.790382	0.000000
EMBSCH	0.240247	0.164480
PHYCOL	0.752015	0.000000
PARCAE	0.640671	0.000181
SYLATR	0.230473	0.204425
HIRRUS	-0.042000	0.856553

When looking at the Spearman's rank correlation with monthly data, HIRRUS was the only species with a negative correlation coefficient. All other species (ACRARU, ACRSCI, ACRSCH, EMBSCH, PHYCOL, PARCAE, and SYLATR) reviled a positive correlation between the number of captures at both sites. Five species (ACRARU, ACRSCI, ACRSCH, PHYCOL, and PARCAE) were statistically significant (see Table 3.6). The time series plots of the Spearman's rank correlations clearly highlight the results of the positive and negative correlations of yearly data(see Figure 3.4).

The results of the RMS revealed a low but slightly positive correlation, indicating a minimal trend. However, the results are not statistically significant (see Table 3.7).

Table 3.7.: Repeated measures correlation

correlation coefficient	dof	p-value	CI 95 percent	power
0.100842	63	0.424127	[-0.15, 0.34]	0.126053

 3. Results

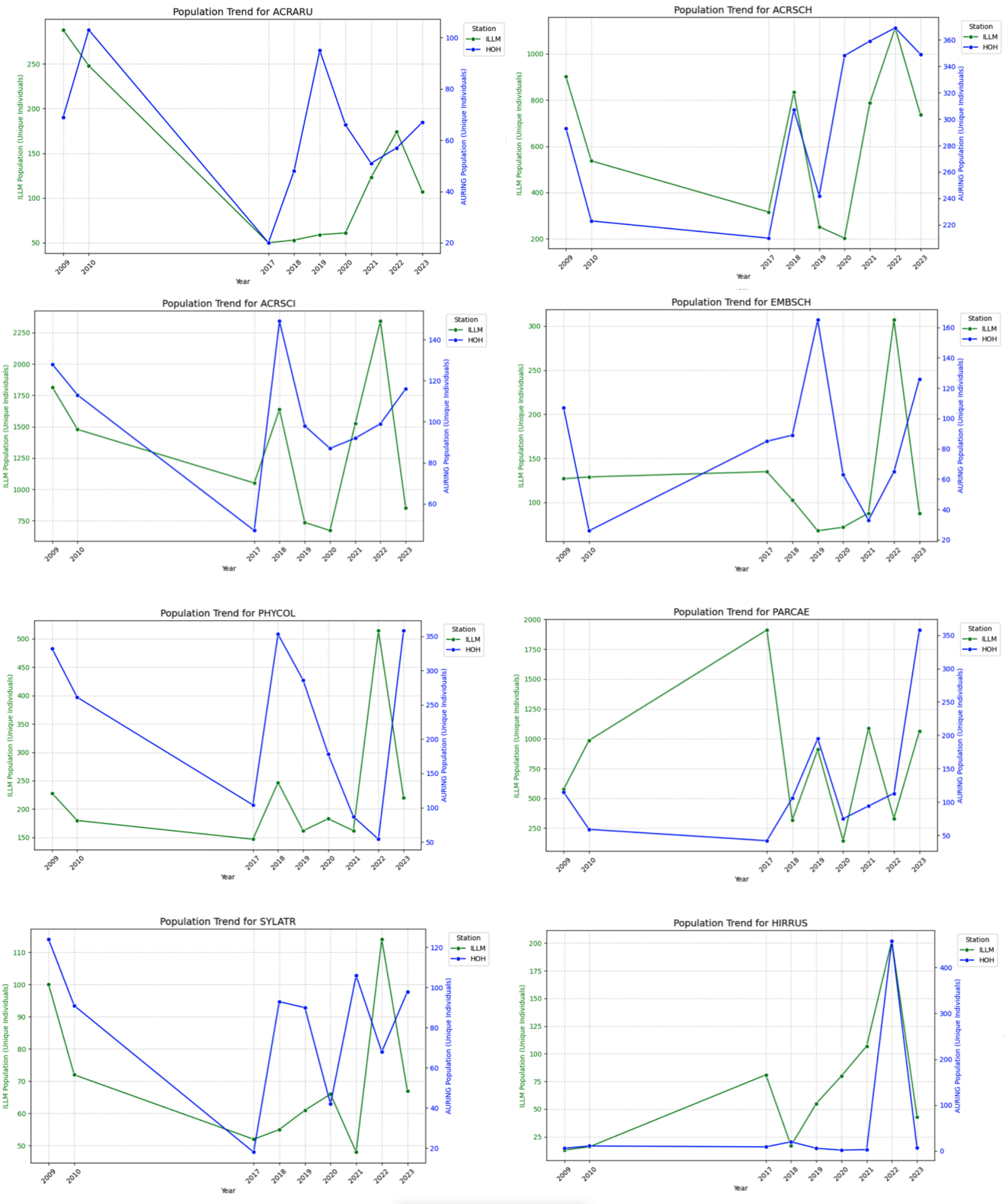

Figure 3.4.: Yearly population of ILLM (green) and HOH (blue). The secondary y-axis visualizes the proportion of birds caught at the two stations with different net lengths. The x-axis shows data from the years 2009-2010 and 2017-2023.

3.2.2. Research question 4:

The bar chart revealed that the overall distribution of fat scores across the two stations was similar. In the fat categories 0, 2 and 3 the ILLM had more individuals than HOH,. However, HOH had more birds of fat category 1 than ILLM. Categories 4, 5, 6, 7 and 8 appear to have similar numbers of birds (see Figure 3.5).

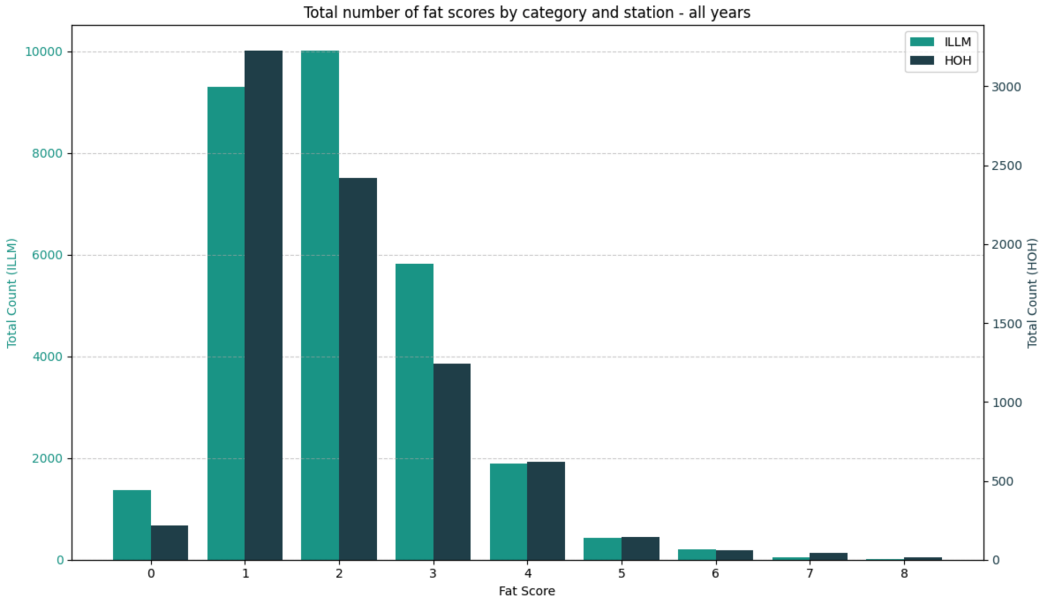

Figure 3.5.: Total number of fat scores by category and station

The Mann-Whitney U-test across all years showed statistically significant difference in fat scores between the stations for most species (ACRSCH, ACRSCI, PARCAE, EMBSCH and PHYCOL). Only ACRARU had no statistically significant difference in fat scores between the two stations when tested with all years (see Table 3.8).

Table 3.8.: Mann-Whitney U-test for all years: p-values

species	p-value
ACRARU	0.0752
ACRSCI	0.0176
ACRSCH	0.0000
EMBSCH	0.0000
PHYCOL	0.0000

Results of the Mann-Whitney U-test for each year were displayed in a cross table. Overall the results were different to the Mann-Whitney U-test. A year-to-year variability for every species was made visible. However 2017 and 2018 are the only years where all species, except ACRARU, showed significant differences in fat scores (ACRSCH, ACRSCI, EMBSCH, PARCAE, and PHYCOL) (see Table 3.9).

Table 3.9.: Mann-Whitney U-test for each year: year-to-year variations displaying p-values

years	ACRARU	ACRSCH	ACRSCI	EMBSCH	PARCAE	PHYCOL
2009	0.0149	0.0449	0.0000	0.7097	0.0000	0.5488
2010	0.1754	0.0016	0.0002	0.8913	0.2957	0.0000
2017	0.0881	0.0000	0.0000	0.0000	0.0000	0.0011
2018	0.0642	0.0172	0.0036	0.0007	0.0000	0.0000
2019	0.1334	0.1280	0.8615	0.0000	0.0027	0.3302
2020	0.6755	0.0000	0.2710	0.0468	0.1156	0.0294
2021	0.0082	0.0000	0.0865	0.6671	0.0000	0.0022
2022	0.1473	0.0000	0.0000	0.1376	0.0029	0.0512
2023	0.0106	0.0824	0.4344	0.1684	0.4801	0.0070

The time series plots for mean fat score visualized which site had higher and smaller scores, with red dots indicating that a year was statistically significant (see Figure 3.6 and 3.7). ACRARU had three statistically significant years and revealed lower fat score means in ILLM in the years 2021 and 2023, and in HOH in 2009. ACRSCH had seven statistically significant years and displayed more lower fat score means in HOH in the years 2018, 2020, 2021, and 2022. ILLM showed lower statistically significant mean fat scores in 2009, 2010, and 2017. ACRSCH had five statistically significant years, with lower fat data in ILLM in 2009, 2010, and 2017, and in HOH in 2018 and 2022 (see Figure 3.6). EMBSCH displayed all four statistically significant years with smaller scores at ILLM in the years 2017, 2018, 2019, and 2020. PARCAE had six statistically significant years and revealed lower fat score means in HOH in the years 2009, 2018, 2019, 2021, and 2022, and at ILLM in 2017. PHYCOL revealed six statistically significant years and displayed more lower fat score means in HOH in the years 2018, 2020, 2021, and 2023. ILLM showed lower mean fat scores in 2010 and 2017 (see Figure 3.7).

 3. Results

Figure 3.6.: Yearly mean fat scores for all years in detail - species: ACRARU, ACRSCH, ACRSCI. ILLM (green), HOH (purple), red dot shows statistically significant differences

 3. Results

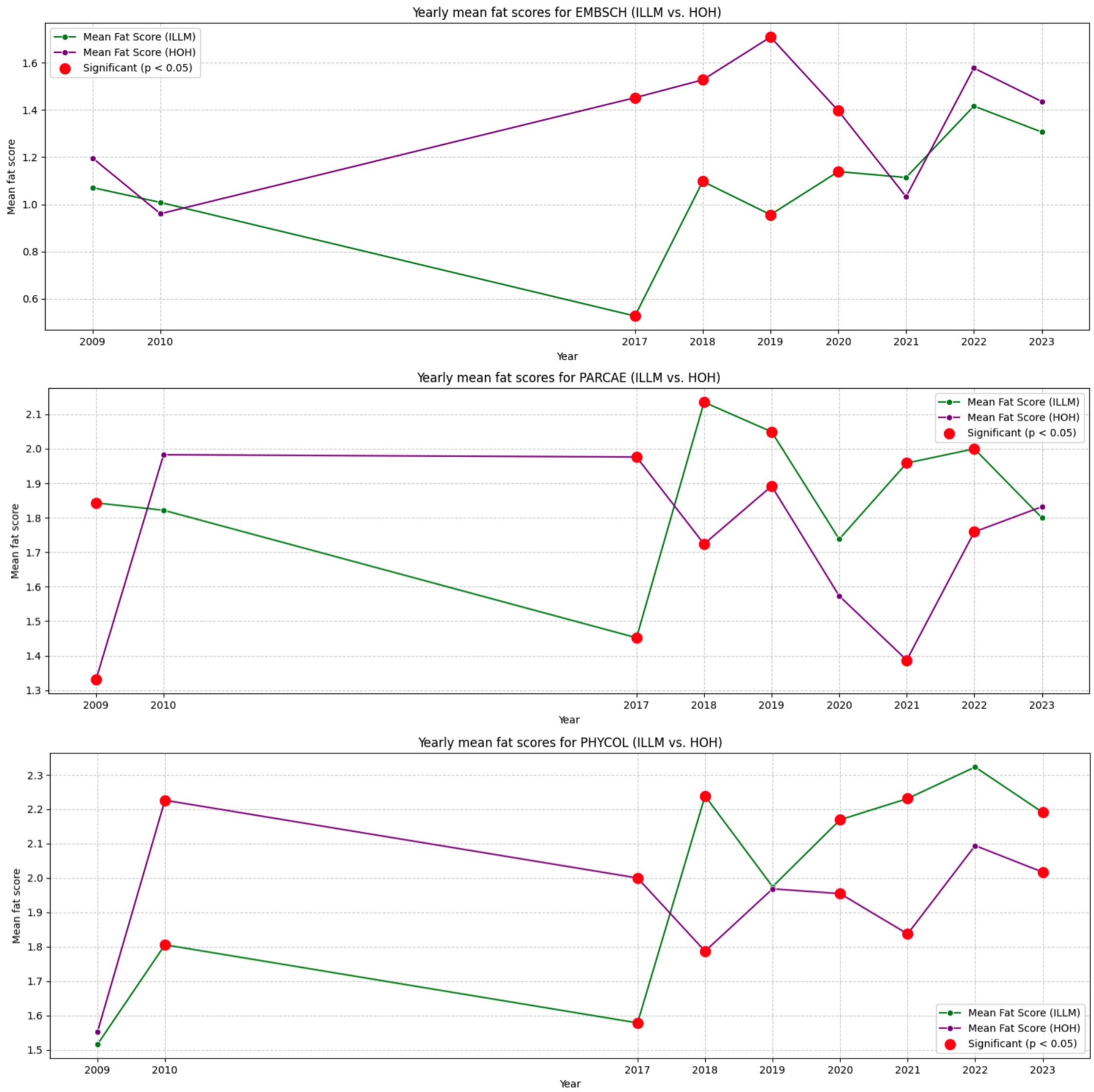

Figure 3.7.: Yearly mean fat scores for all years in detail - species: EMBSCH, PARCAE, PHYCOL. ILLM (green), HOH (purple), red dot shows statistically significant differences

4. Discussion

4.1. Biological station Illmitz: water level, stopover duration and population size

Due to the fact that wetlands have been identified as important stopover sites (Arizaga et al. [2013], Bozó et al. [2018]), the possible influence of water level on the stopover duration was intriguing. However, the results revealed no statistically significant monotonic relationship between the stopover duration of European Reed Warblers and Sedge Warblers and the water level of Lake Neusiedl. Other studies have shown that the decision-based process of stopover duration was influenced by the positive or negative fuel deposition rate and by predation risk (Schaub et al. [2008]). Nevertheless, the habitat of this ringing station is, due to the endorheic steppe lakes and extensive reed belt, extremely appealing to birds (Pretzer et al. [2017]). Combined with the research of the last decade on the decline of reed harvesting and the formation in old reed areas (Nemeth and Dvorak [2022]), for future studies, it would be interesting to investigate the correlation between the changing reed belt and stopover duration. Reed Warblers are known to be flexible foragers who can easily extend their foraging areas beyond reed beds, as well as dry reeds. Whereas Sedge Warblers were known to prey specifically on plum aphids (Bibby and Green [1981]). A paper in 2000 suggested that they are more flexible in their diet selection than previously thought (Chernetsov and Manukyan [2000]). The adaptability of these reed species may explain why their minimum stopover duration did not correlate with water level. Another reason for this result could be a general low capture probability (Schaub et al. [2001]). It should be taken into account that the daily capture probability of tape-lured Reed Warblers may not exceed 25 percent (**?**). Therefore, the real number of individuals and the stopover duration were probably higher, which may have influenced the result. Additionally, it is important to note that I was only able to calculate the minimum stopover duration. A more accurate and optimal method to calculate the stopover duration would have been the Cormack-Jolley-Seber model (CJS), which is a type of capture/recapture analysis and works with probabilities. However, the CJS model needs high recapture numbers, which this dataset did not provide. Therefore, the best approximation remained with the calculation of the minimum stopover duration. For future testing, I suggest digitizing the remaining analog data to obtain a larger data set. With additional historical data, it would be interesting to investigate and compare the results of the stopover duration method with those calculated by the CJS model. The CJS model should provide a more accurate estimate of the time a bird spends at a site (Schaub et al. [2001]).

Bird ringing data can define species population size and trends (Petras and Vrezec [2022]). Therefore, the data was used to answer the question of the relationship between the water level and population size at ILLM. The Spearman's rank correlation was for neither species statistically significant. The results of the GLM with Poisson distribution showed significant p-values. However, due to an overdispersion, which can appear when dealing with count data, like population size, it can happen that p-values appear smaller and more significant than they actually are. The high results of the dispersion parameters confirmed this possibility. When I tested the data with a negative binomial GLM, the confidence intervals around the prediction line turned out wider, but the p-values were still not statistically significant. However, the fact that the actual data points were better enveloped by the wider intervals was a clear sign that the data variability was better represented with a negative binomial GLM. The GLM plots displayed population size trends. Species with an upward curve (PARCEA, ACRMEL, ERIRUB, and ACRARU) revealed an increasing trend when the water level rose. Plots with a downward curve (ACRSCI, ACRSCH, EMBSCH) suggested the opposite. The trend for PHYCOL remained more or less flat, which brings me to the conclusion that the water level was not a strong predictor for its species population size. However, the wide confidence intervals present in all plots suggested that the true population size may vary within a broad range, even though it generally increased or decreased. The results did not fully support the hypothesis that higher water levels always reveal a positive correlation, and lower water levels a negative correlation with the population size of migratory birds. For future analysis, it would be interesting to involve additional environmental factors that might explain the wide variability observed in some species.

4.2. Bird ringing station comparison: population size and fat score

When testing the number of captured species (population size) at both sites with yearly data, the Spearman's rank correlation was not statistically significant. However, when using monthly data, five out of the eight species revealed a statistically significant p-value. The different results may have several reasons connected to the data itself and the nature of bird population size. When working with the mean of yearly data, seasonal patterns tend to smooth out. Choosing to use monthly data provides more data points for each species and a more detailed picture of possible patterns. This can lead to a higher statistical power, making it simpler to detect true underlying correlations, if they exist. Patterns like the arrival, peak presence, and departure of birds at ringing stations are better captured with monthly data. To sum it up, yearly counts can reflect broader trends in the number of captured species and the reflected population size, whereas monthly counts may reveal more detailed information about possible correlations. In this thesis, only the monthly Spearman's rank correlation revealed statistically significant p-values for five species (ACRARU, ACRSCI, ACRSCH, PHYCOL, and PARCAE) out of eight. All of the five species also had a positive correlation coefficient. This indicates that the observed population size trends of these five migratory bird species are more likely to be global. When testing the yearly data set with RMC, it revealed a low positive correlation result, only indicating a minimal, not statistically significant linear trend. This means that at a yearly perspective, species fluctuation in one station does not predict species fluctuation in the other station. Nonetheless, the low statistical power of the test warns to be cautious with interpretations. It does not mean that there is a definite absence of any relationship. With caution, it can be suggested that the results support the hypothesis that trend fluctuations on the population size of the focus species (with the exception of HIRRUS, SYLATR, and EMBSCH) at these ringing stations are positively correlated and therefore vary in the same way. For future investigation and a more profound statement, I suggest testing this hypothesis with different focus species and increasing the number of sites to see whether the changes occur everywhere in the same way.

The distribution of the total number of fat sores in all nine years between ILLM and HOH seemed similar. However, looking closer, the fat scores 0-3 had the highest count number of individuals and revealed a different total count number depending on the ringing station. The Mann-Whitney U-test over all years showed general trends in fat scores from caught migratory birds between ILLM and HOH. However, only the data from the first capture was analyzed. Hence, the results may reflect possible differences, like choice for the stopover site due to fat score, in a better way. The results showed statistically significant differences between ILLM and HOH for all species, except ACRARU. When tested with yearly data, more insights into year-to-year variation have been revealed. The difference in fat score, that have been visible between the sites over all years, were not constant over time and can be seen in the cross tabulation displaying the year-to-year variation of p-values. Due to yearly fluctuations in food availability, predation risk, weather, or other environmental factors (Buler et al. [2017], Newton [2004]) at each site, the optimal stopover habitat for a specific species may change. Keeping that in mind, the results for yearly data may suggest species-specific response rather than an uniform site effect (see Table A.8 and Table A.9 in appendix). When analyzing the time series plots, 2018 seemed to be a year where a shift between the ringing station sites happened. Previously, the species ACRSCH, ACRSCI, and PHYCOL displayed lower mean fat values at ILLM. From 2018 onward, low mean fat values with statistical significance only occur at HOH. This suggested that since 2018, the area of HOH may have provided a better habitat for these three species. For future studies, I suggest testing this data additionally with environmental data to get an idea of the possible changes that happened in 2018. Furthermore, it would be interesting to see if new collected data since 2023 still reveal similar results, or if another pattern shift happened.

An important point, I feel the need to address for this thesis, is the concept of shifting baseline syndrome (SBS). It explains a collective failure to understand the full magnitude of environmental changes, due to the fact that each new generation defines its expectations for healthy ecosystems based on its own initial observations. Therefore, earlier declines are being discounted (e.g., Alleway et al. [2023]). Sadly, the threat of SBS may already be present within the knowledge about avifauna, due to the fact that the oldest digitized bird ringing data of Illmitz are form the year 1972, where the data collection was initiated through the “Mettnau-Reit-Illmitz” program (Berthold et al. [1998]) and most studies use the beginning of the 1980s as a baseline (Reif et al. [2021]). In this study, I was unable to use all available historical bird-ringing data from the biological station Illmitz, as not all of it has been digitized yet. This adds an even stronger shifting baseline effect than is already present. For future investigations using historical Illmitz data, I highly recommend investing time and resources in digitizing the missing years and species. At this stage, only 24 years were representative and fully digitized out of the possible long-term dataset of 49 years (1974-2023). Therefore, all results of this thesis should be interpreted with an added shifting baseline effect in mind.

5. Conclusion

This thesis set out to examine the data of migratory birds during autumn migration at two different ringing stations. Firstly, I investigated the correlations between water level and stopover duration, and between water level and population size. Secondly, the two stations, ILLM and HOH, were compared for population size trend variation and differences in fat score.

Contrary to my expectations, the water level fluctuations of Lake Neusiedl did not influence the stopover duration of European Reed Warblers and Sedge Warblers. However, a possible driver for stopover duration, besides fuel deposition rate and predation risk (Schaub et al. [2008]), could also be the changing reed belt in that area (Nemeth and Dvorak [2022]). Therefore, it would be interesting to investigate the correlation between changes in the reed belt and stopover duration in future studies.

Every test I used to analyze the relationship between the water level at ILLM and the population size of migratory birds resulted in non-statistically significant p-values. The plots of the negative binomial GLM showed wide confidence intervals, highlighting the plausible range of population sizes while acknowledging ecological variability. Still, the hypothesis that higher water levels reveal positive correlations and lower water levels negative correlations with the population size of migratory birds was rejected. For future analysis, I suggest including additional environmental or biological factors to better understand the wide variability observed in some species.

Overall, these results highlight that water level did not influence stopover duration or population size. These findings inspire deeper investigation into possible correlations with other parameters. However, certain constraints should be acknowledged when analyzing the current ILLM data set, such as the minimum stopover, rather than using a CJS model.

When comparing the yearly data from the two bird ringing stations, there was no statistically significant population trend (number of captured birds) between them. The findings of the monthly analysis showed that ACRARU, ACRSCI, ACRSCH, PHYCOL, and PARCAE had statistically significant positive correlations. Suggesting that the trend variation in population size followed the same pattern at both stations. Therefore, the prediction that observed population trends, that vary in the same way, are more likely to be global, may have been proven true for these five migratory bird species. However, the Repeated measures correlation results challenge the new plausible predictions. The RMC revealed only a minimal linear trend, without being statistically significant, suggesting that species fluctuations at one station did not, in fact, predict fluctuations at the other station. Nevertheless, this does not mean that there is no other kind of relationship present. With caution, it can be suggested that the number of captured birds for ACRARU, ACRSCI, ACRSCH, PHYCOL, and PARCAE may vary in a similar way at both stations.

Comparing the stations based on the fat score of caught migratory birds visually revealed an overall similar distribution. However, the fat score with the most birds in it was 1-3. When

 5. Conclusion

testing with data that combined all years, the results showed that nearly all species (ACRSCH, EMBSCH, PARCAE, PHYCOL, and ACRSCI) differ in fat scores between ILLM and HOH. However, yearly data revealed more insights into year-to-year variations, denying that the difference in fat score numbers between the stations was constant over time. In conclusion, yearly data may suggest species-specific response rather than a uniform site effect. The species-specific responses for each year were made visible in the time series plots. What stands out was the year 2018, revealing a pattern shift between the sites. From 2018 onward, the species ACRSCH, ACRSCI, and PHYCOL only had low mean fat values, with statistical significance occurring at HOH. This may be an indicator that, for these three species, HOH from then on provided a better habitat to refuel during their stopover.

In summary, both the population size and the fat score tended to reveal clearer results when using the yearly data. Year-to-year variations provided deeper insight into the data than tests that used data from all years combined.

References

Tina Petras and Al Vrezec. Long-Term Ringing Data on Migrating Passerines Reveal Overall Avian Decline in Europe. *Diversity*, 14(11):905, October 2022. ISSN 1424-2818. doi: 10.3390/d14110905. URL `https://www.mdpi.com/1424-2818/14/11/905`.

Fiona Burns, Mark A. Eaton, Ian J. Burfield, Alena Klvaňová, Eva Šilarová, Anna Staneva, and Richard D. Gregory. Abundance decline in the avifauna of the European Union reveals cross-continental similarities in biodiversity change. *Ecology and Evolution*, 11(23): 16647–16660, December 2021. ISSN 2045-7758, 2045-7758. doi: 10.1002/ece3.8282. URL `https://onlinelibrary.wiley.com/doi/10.1002/ece3.8282`.

Fiona J. Sanderson, Paul F. Donald, Deborah J. Pain, Ian J. Burfield, and Frans P.J. Van Bommel. Long-term population declines in Afro-Palearctic migrant birds. *Biological Conservation*, 131(1):93–105, July 2006. ISSN 00063207. doi: 10.1016/j.biocon.2006.02.008. URL `https://linkinghub.elsevier.com/retrieve/pii/S000632070600070X`.

P. Berthold, W. Fiedler, R. Schlenker, and U. Querner. 25-Year Study of the Population Development of Central European Songbirds: A General Decline, Most Evident in Long-Distance Migrants. *Naturwissenschaften*, 85(7):350–353, July 1998. ISSN 0028-1042, 1432-1904. doi: 10.1007/s001140050514. URL `http://link.springer.com/10.1007/s001140050514`.

Ian Newton. Population limitation in migrants. *Ibis*, 146(2):197–226, 2004. ISSN 1474-919X. doi: 10.1111/j.1474-919X.2004.00293.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1474-919X.2004.00293.x`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1474-919X.2004.00293.x.

Diana E. Bowler, Henning Heldbjerg, Anthony D. Fox, Maaike De Jong, and Katrin Böhning-Gaese. Long-term declines of European insectivorous bird populations and potential causes. *Conservation Biology*, 33(5):1120–1130, October 2019. ISSN 0888-8892, 1523-1739. doi: 10.1111/cobi.13307. URL `https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.13307`.

Anders Pape Møller, Diego Rubolini, and Esa Lehikoinen. Populations of migratory bird species that did not show a phenological response to climate change are declining. *Proceedings of the National Academy of Sciences*, 105(42):16195–16200, October 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0803825105. URL `https://pnas.org/doi/full/10.1073/pnas.0803825105`.

David A. Cimprich, Mark S. Woodrey, and Frank R. Moore. Passerine migrants respond to variation in predation risk during stopover. *Animal Behaviour*, 69(5):1173–

1179, May 2005. ISSN 00033472. doi: 10.1016/j.anbehav.2004.07.021. URL https://linkinghub.elsevier.com/retrieve/pii/S000334720500031X.

Franz Bairlein. European-African songbird migration network Man- ual of field methods. Revised edition. *ESF-Network Field Instructions*, page 32, 1995.

Thomas Alerstam and Anders Hedenstrom. The Development of Bird Migration Theory. *Journal of Avian Biology*, 29(4):343, December 1998. ISSN 09088857. doi: 10.2307/3677155. URL https://www.jstor.org/stable/3677155?origin=crossref.

Michael Schaub, Lukas Jenni, and Franz Bairlein. Fuel stores, fuel accumulation, and the decision to depart from a migration stopover site. *Behavioral Ecology*, 19(3): 657–666, 2008. ISSN 1465-7279, 1045-2249. doi: 10.1093/beheco/arn023. URL https://academic.oup.com/beheco/article-lookup/doi/10.1093/beheco/arn023.

Michael Schaub and Lukas Jenni. Stopover durations of three warbler species along their autumn migration route. *Oecologia*, 128(2):217–227, July 2001. ISSN 0029-8549, 1432-1939. doi: 10.1007/s004420100654. URL http://link.springer.com/10.1007/s004420100654.

Jeffrey J. Buler, Rebecca J. Lyon, Jaclyn A. Smolinsky, Theodore J. Zenzal, and Frank R. Moore. Body mass and wing shape explain variability in broad-scale bird species distributions of migratory passerines along an ecological barrier during stopover. *Oecologia*, 185(2):205–212, October 2017. ISSN 0029-8549, 1432-1939. doi: 10.1007/s00442-017-3936-y. URL http://link.springer.com/10.1007/s00442-017-3936-y.

Jennifer Mccabe and Brian Olsen. Tradeoffs between predation risk and fruit resources shape habitat use of landbirds during autumn migration. *The Auk*, 132:903–913, October 2015. doi: 10.1642/AUK-14-213.1.

Andreas Kaiser. A New Multi-Category Classification of Subcutaneous Fat Deposits of Songbirds (Una Nueva Clasificación, con Multi-categorías, para los Depósitos de Grasa en Aves Canoras). 64(2):246–255, 1993. URL https://www.jstor.org/stable/4513807.

H. Biebach, W. Friedrich, and G. Heine. Interaction of bodymass, fat, foraging and stopover period in trans-Sahara migrating passerine birds. *Oecologia*, 69(3):370–379, June 1986. ISSN 0029-8549, 1432-1939. doi: 10.1007/BF00377059. URL http://link.springer.com/10.1007/BF00377059.

Franz Bairlein. Body weights and fat deposition of Palaearctic passerine migrants in the central Sahara. *Oecologia*, 66(1):141–146, April 1985. ISSN 0029-8549, 1432-1939. doi: 10.1007/BF00378566. URL http://link.springer.com/10.1007/BF00378566.

Jeffrey D. Cherry. Fat Deposition and Length of Stopover of Migrant White-Crowned Sparrows. *The Auk*, 99(4):725–732, 1982. ISSN 1938-4254. doi: 10.1093/auk/99.4.725. URL https://doi.org/10.1093/auk/99.4.725.

Ivan Maggini, Marta Trez, Massimiliano Cardinale, and Leonida Fusani. Stopover dynamics of 12 passerine migrant species in a small Mediterranean island during spring migration. *Journal of Ornithology*, 161(3):793–802, July 2020. ISSN 2193-7192, 2193-7206. doi: 10.1007/s10336-020-01768-7. URL `http://link.springer.com/10.1007/s10336-020-01768-7`.

Frank R. Moore and Wang Yong. Evidence of food-based competition among passerine migrants during stopover. *Behavioral Ecology and Sociobiology*, 28 (2):85–90, February 1991. ISSN 1432-0762. doi: 10.1007/BF00180984. URL `https://doi.org/10.1007/BF00180984`.

Jeffrey F. Kelly, Linda S. DeLay, and Deborah M. Finch. Density-Dependent Mass Gain by Wilson's Warblers during Stopover. *The Auk*, 119(1):210–213, 2002. ISSN 0004-8038. doi: 10.2307/4090024. URL `https://www.jstor.org/stable/4090024`.

David M. Whalen and Bryan D. Watts. Annual Migration Density and Stopover Patterns of Northern Saw-whet Owls (Aegolius acadicus). *The Auk*, 119(4):1154–1161, October 2002. ISSN 1938-4254. doi: 10.1093/auk/119.4.1154. URL `https://doi.org/10.1093/auk/119.4.1154`.

Nina Mclean, Callum Lawson, Dave Leech, and Martijn van de Pol. Predicting when climate-driven phenotypic change affects population dynamics. *Ecology letters*, 19, April 2016. doi: 10.1111/ele.12599.

Erwin Nemeth and Michael Dvorak. Reed die-back and conservation of small reed birds at Lake Neusiedl, Austria. *Journal of Ornithology*, 163(3):683–693, July 2022. ISSN 2193-7192, 2193-7206. doi: 10.1007/s10336-022-01961-w. URL `https://link.springer.com/10.1007/s10336-022-01961-w`.

Juan Arizaga, Miren Andueza, and Ibon Tamayo. Spatial Behaviour and Habitat use of First-Year Bluethroats *Luscinia svecica* Stopping over at Coastal Marshes during the Autumn Migration Period. *Acta Ornithologica*, 48(1):17–25, June 2013. ISSN 0001-6454, 1734-8471. doi: 10.3161/000164513X669964. URL `http://www.bioone.org/doi/abs/10.3161/000164513X669964`.

Alfréd Trnka and Pavol Prokop. Reedbed structure and habitat preference of reed passerines during the post-breeding period. *Biologia*, 61(2):225–230, April 2006. ISSN 0006-3088, 1336-9563. doi: 10.2478/s11756-006-0034-8. URL `https://link.springer.com/10.2478/s11756-006-0034-8`.

László Bozó, Wieland Heim, and Tibor Csörgő. Habitat use by Siberian warbler species at a stopover site in Far Eastern Russia. *Ringing & Migration*, 33(1): 31–35, January 2018. ISSN 0307-8698. doi: 10.1080/03078698.2018.1446889. URL `https://doi.org/10.1080/03078698.2018.1446889`. _eprint: https://doi.org/10.1080/03078698.2018.1446889.

Franz Bairlein. Habitat Selection and Associations of Species in European Passerine Birds during Southward, Post-Breeding Migrations. *Ornis Scandinavica*, 14:239, September 1983. doi: 10.2307/3676157.

Richard D. Gregory and Arco Van Strien. Wild Bird Indicators: Using Composite Population Trends of Birds as Measures of Environmental Health. *Ornithological Science*, 9(1):3–22, June 2010. ISSN 1347-0558. doi: 10.2326/osj.9.3. URL `http://www.bioone.org/doi/abs/10.2326/osj.9.3`.

Carina Pretzer, Irina S. Druzhinina, Carmen Amaro, Eva Benediktsdóttir, Ingela Hedenström, Dominique Hervio-Heath, Steliana Huhulescu, Franciska M. Schets, Andreas H. Farnleitner, and Alexander K. T. Kirschner. High genetic diversity of Vibrio cholerae in the European lake Neusiedler See is associated with intensive recombination in the reed habitat and the long-distance transfer of strains. *Environmental Microbiology*, 19(1):328–344, 2017. ISSN 1462-2920. doi: 10.1111/1462-2920.13612. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.13612`. _eprint: https://enviromicro-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/1462-2920.13612.

Michael Dvorak, Johannes Laber, Andreas Ranner, and Alois Lang. ARTENLISTE Die Vögel des Neusiedler See-Gebiets, July 2012.

Klaus Peter Zulka, Katharina Huchler, Bernhard Schön, Thomas Wrbka, Helmut Kudrnovsky, and Stefan Schindler. Österreichische Hotspots der Biodiversität. Technical Report REP-0945, Umweltbundesamt, 2024. URL `https://www.umweltbundesamt.at/`.

Eva Karner-Ranner, Alfred Grüll, and Andreas Ranner. Monitoring von Kulturlandvögeln im Nationalpark Neusiedler See – Seewinkel als Grundlage für Managementmaßnahmen. 49, 2008.

Thomas Zuna-Kratky, Benjamin Seaman, Matthias Schmidt, Bernhard Paces, Martin Rössler, and Ute Nüsken. Biologische Station Hohenau-Ringelsdorf Jahresbericht 2022, 2022.

Thomas Zuna-Kratky, Benjamin Seaman, Matthias Schmidt, Karin Donnerbaum, and Ute Nüsken. Biologische Station Hohenau-Ringelsdorf Jahresbericht 2023, 2023.

Martin Rössler and Thomas Zuna-Kratky. Bericht über die Fangsaison 1994 der Vogelberingungsstation Hohenau/March, January 1995.

Martina Hillbrand. *Nachbrutzeitliche Verweildauer, Wegzugphänologie und Ortstreue der Rohrammer (Emberiza schoeniclus) in nitrophilen Ruderalfluren der March-Auen*. PhD thesis, May 2004.

Peter Hackl and Johannes Ledolter. A Statistical Analysis of the Water Levels at Lake Neusiedl. *Austrian Journal of Statistics*, 52(1):87–100, March 2023. ISSN 1026-597X. doi: 10.17713/ajs.v52i1.1444. URL `https://www.ajs.or.at/index.php/ajs/article/view/1444`.

Catriona A. Morrison, Robert A. Robinson, Jacquie A. Clark, and Jennifer A. Gill. Spatial and temporal variation in population trends in a long-distance migratory bird. *Diversity and Distributions*, 16(4):620–627, July 2010. ISSN 1366-9516, 1472-4642. doi: 10.1111/j.1472-4642.2010.00663.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1472-4642.2010.00663.x.

V. A. Payevsky. Mechanisms of population dynamics in trans-Saharan migrant birds: A review. *Entomological Review*, 86(1):S82–S94, January 2006. ISSN 1555-6689. doi: 10.1134/S001387380610006X. URL https://doi.org/10.1134/S001387380610006X.

Georg Bieringer, Michael Dvorak, Eva Kranker-Ranner, Bernhard Kohler, Johannes Laber, Erwin Nemeth, Georg Rauer, and Beate Wendelin. Ornithologisches Monitoring im Nationalpark Neusiedler See- Seewinkel, 2013.

M. Schaub and L. Jenni. Fuel deposition of three passerine bird species along the migration route. *Oecologia*, 122(3):306–317, February 2000. ISSN 0029-8549, 1432-1939. doi: 10.1007/s004420050036. URL http://link.springer.com/10.1007/s004420050036.

Gerald Dick. The significance of the Lake Neusiedl area of Austria for migrating geese. 38:19–27, 1987.

C. J. Bibby and R. E. Green. Autumn Migration Strategies of Reed and Sedge Warblers. *Ornis Scandinavica (Scandinavian Journal of Ornithology)*, 12(1):1–12, 1981. ISSN 0030-5693. doi: 10.2307/3675898. URL https://www.jstor.org/stable/3675898.

Nikita Chernetsov and Andranik Manukyan. Foraging strategy of the Sedge Warbler (Acrocephalus schoenobaenus) on migration. 40:189–197, 2000.

Michael Schaub, Roger Pradel, Lukas Jenni, and Jean-Dominique Lebreton. MIGRATING BIRDS STOP OVER LONGER THAN USUALLY THOUGHT: AN IMPROVED CAPTURE–RECAPTURE ANALYSIS. *Ecology*, 82(3):852–859, March 2001. ISSN 0012-9658. doi: 10.1890/0012-9658(2001)082[0852:MBSOLT]2.0.CO;2. URL http://doi.wiley.com/10.1890/0012-9658(2001)082[0852:MBSOLT]2.0.CO;2.

Heidi K. Alleway, Emily S. Klein, Liz Cameron, Kristina Douglass, Ishtar Govia, Cornelia Guell, Michelle Lim, Libby Robin, and Ruth H. Thurstan. The shifting baseline syndrome as a connective concept for more informed and just responses to global environmental change. *People and Nature*, 5(3):885–896, June 2023. ISSN 2575-8314, 2575-8314. doi: 10.1002/pan3.10473. URL https://besjournals.onlinelibrary.wiley.com/doi/10.1002/pan3.10473.

Jiří Reif, Filip Szarvas, and Karel Šťastný. 'Tell me where the birds have gone' – Reconstructing historical influence of major environmental drivers on bird populations from memories of ornithologists of an older generation. *Ecological Indicators*, 129: 107909, October 2021. ISSN 1470160X. doi: 10.1016/j.ecolind.2021.107909. URL https://linkinghub.elsevier.com/retrieve/pii/S1470160X21005744.

A. Appendix

Table A.1.: Shapiro-Wilk test results for all years with original data and log transformed data

species	data type	observations	test statistic	p-value
ACRSCI	original data	4079	0.9362	0.0000
ACRSCH	original data	1509	0.8475	0.0000
ACRSCI	log transformed data	3982	0.9535	0.0000
ACRSCH	log transformed data	1491	0.9384	0.0000

 A. Appendix

Table A.2.: Shapiro-Wilk test results for each year seperate

species	year	observations	test statistic	p-value
ACRSCI	1974	256	0.9371	0.0000
ACRSCI	1975	213	0.9191	0.0000
ACRSCI	1976	348	0.9112	0.0000
ACRSCI	1977	171	0.9309	0.0000
ACRSCI	1978	88	0.9084	0.0000
ACRSCI	1979	134	0.9614	0.0008
ACRSCI	1980	167	0.9297	0.0000
ACRSCI	1981	217	0.9240	0.0000
ACRSCI	1982	354	0.9421	0.0000
ACRSCI	1983	275	0.9524	0.0000
ACRSCI	1989	188	0.9300	0.0000
ACRSCI	1990	207	0.9304	0.0000
ACRSCI	1991	66	0.9452	0.0057
ACRSCI	1992	320	0.9247	0.0000
ACRSCI	1993	180	0.9428	0.0000
ACRSCI	2009	167	0.9418	0.0000
ACRSCI	2010	94	0.9256	0.0000
ACRSCI	2017	65	0.9199	0.0004
ACRSCI	2018	103	0.9100	0.0000
ACRSCI	2019	70	0.9519	0.0091
ACRSCI	2020	45	0.9500	0.0506
ACRSCI	2021	77	0.9373	0.0009
ACRSCI	2022	131	0.9170	0.0000
ACRSCI	2023	46	0.9481	0.0395
ACRSCH	1974	45	0.7764	0.0000
ACRSCH	1975	57	0.8207	0.0000
ACRSCH	1976	268	0.8215	0.0000
ACRSCH	1977	92	0.9117	0.0000
ACRSCH	1978	58	0.8696	0.0000
ACRSCH	1979	48	0.8472	0.0000
ACRSCH	1980	58	0.8731	0.0000
ACRSCH	1981	69	0.8933	0.0000
ACRSCH	1982	78	0.8393	0.0000
ACRSCH	1989	96	0.8811	0.0000
ACRSCH	1990	39	0.9216	0.0098
ACRSCH	1991	28	0.8828	0.0046
ACRSCH	1992	158	0.8422	0.0000
ACRSCH	1993	64	0.9123	0.0002
ACRSCH	2010	15	0.9138	0.1549
ACRSCH	2019	5	0.9345	0.6275
ACRSCH	2020	9	0.8264	0.0407
ACRSCH	2021	25	0.7186	0.0000
ACRSCH	2022	47	0.8083	0.0000
ACRSCH	2023	33	0.8575	0.0005

 A. Appendix

Table A.3.: Shapiro-Wilk test for 'stopover duration(days)' and 'Water Level' per species over all months (median):

species	column	observations	test statistic	p-value
ACRSCI	stopover duration(days)	89	0.8100	0.0000
ACRSCI	Water Level	89	0.9473	0.0012
ACRSCH	stopover duration(days)	70	0.9409	0.0025
ACRSCH	Water Level	70	0.9562	0.0155

Table A.4.: Shapiro-Wilk test results for population size per species

species	test statistic	p-value
ACRSCI	0.8423	0.0016
ACRSCH	0.8265	0.0008
PARCAE	0.8601	0.0034
EMBSCH	0.8235	0.0007
ACRMEL	0.9053	0.0279
ACRARU	0.8864	0.0112
PHYCOL	0.7830	0.0002
ERIRUB	0.8672	0.0046

Table A.5.: Generalized linear model with Poisson distribution

species	coefficient	p-value
ACRSCI	-0.0013	0.000
ACRSCH	-0.0020	0.000
PARCAE	0.0055	0.000
EMBSCH	-0.0082	0.000
ACRMEL	0.0061	0.000
ACRARU	0.0007	0.336
PHYCOL	-0.0106	0.000
ERIRUB	0.0180	0.000

 A. Appendix

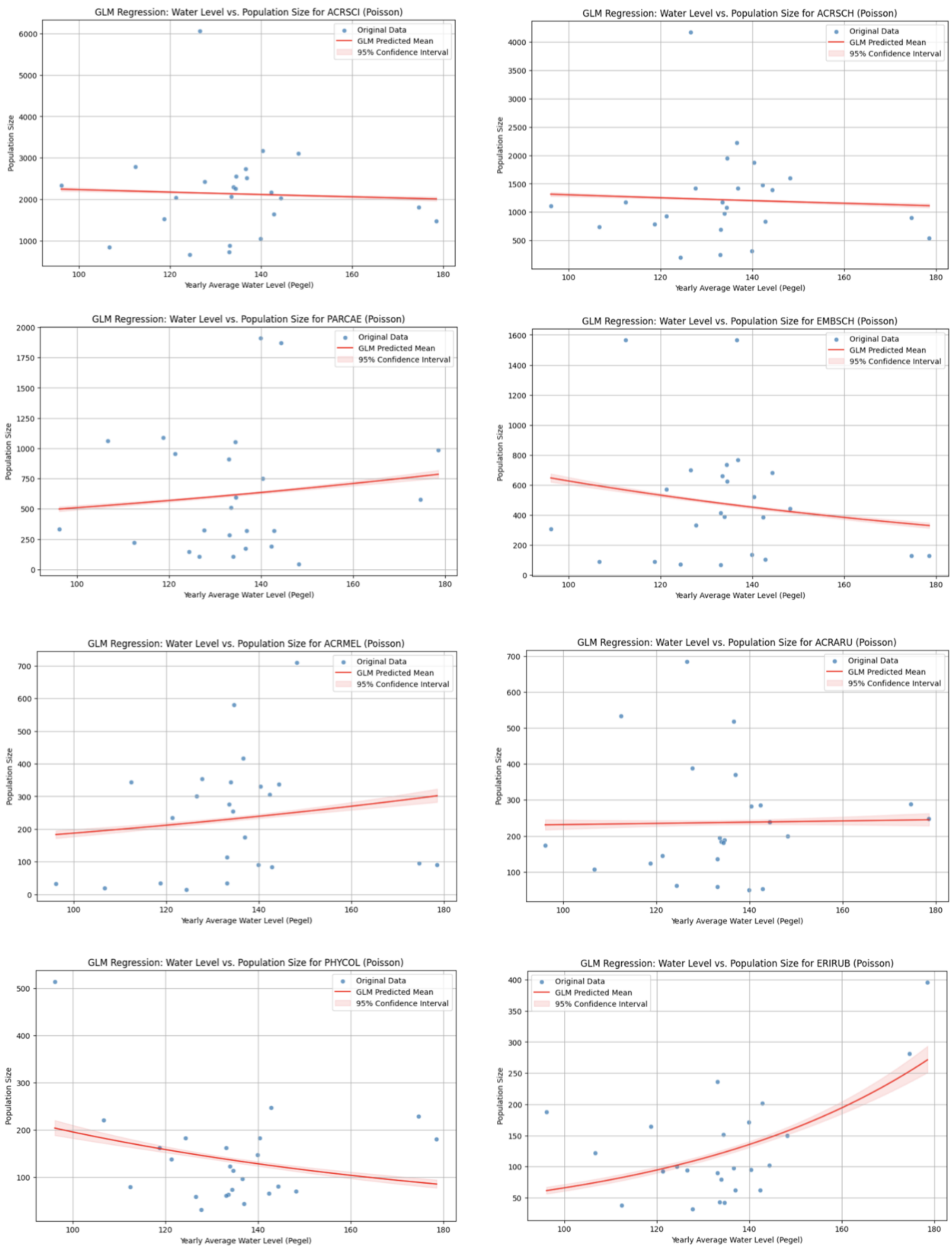

Figure A.1.: Plotted general linear model regression Poisson distribution - for each species

Table A.6.: Shapiro-Wilk test results for yearly unique individuals

species	station	test statistic	p-value
ACRARU	HOH	0.9551	0.7456
ACRARU	ILLM	0.8474	0.0698
ACRSCH	HOH	0.8851	0.1775
ACRSCH	ILLM	0.9319	0.4999
ACRSCI	HOH	0.9635	0.8338
ACRSCI	ILLM	0.9398	0.5798
EMBSCH	HOH	0.9667	0.8653
EMBSCH	ILLM	0.7036	0.0016
HIRRUS	HOH	0.4203	0.0000
HIRRUS	ILLM	0.8522	0.0787
PARCAE	HOH	0.7734	0.0101
PARCAE	ILLM	0.9190	0.3837
PHYCOL	HOH	0.8891	0.1951
PHYCOL	ILLM	0.6664	0.0006
SYLATR	HOH	0.9134	0.3403
SYLATR	ILLM	0.8565	0.0878

Table A.7.: Mann-Whitney U-Test results for all years

species	u-statistic	p-value	mean fat (ILLM)	mean fat (HOH)
ACRARU	308458.00	0.0752	2.37	2.49
ACRSCH	8358001.00	0.0000	2.29	2.07
ACRSCI	5277814.00	0.0176	2.02	2.25
PARCAE	4259861.00	0.0020	1.79	1.73
EMBSCH	340116.00	0.0000	1.12	1.45
PHYCOL	2210015.00	0.0000	2.06	1.90

 A. Appendix

Table A.8.: Mann-Whitney U-Test results for each year (2009-2018)

species	year	u-statistic	p-value	mean fat (ILLM)	mean fat (HOH)
ACRARU	2009	11758.5	0.0149	2.55	2.13
ACRSCH	2009	120549.5	0.0449	2.06	2.39
ACRSCI	2009	85875.5	0.0000	1.75	2.48
PARCAE	2009	42910.0	0.0000	1.84	1.33
EMBSCH	2009	6628.0	0.7097	1.07	1.20
PHYCOL	2009	38176.0	0.5488	1.52	1.55
ACRARU	2010	13731.5	0.1754	2.60	2.40
ACRSCH	2010	49644.0	0.0016	1.86	2.28
ACRSCI	2010	62718.5	0.0002	1.61	2.28
PARCAE	2010	22095.5	0.2957	1.82	1.98
EMBSCH	2010	1536.5	0.8913	1.01	0.96
PHYCOL	2010	16345.5	0.0000	1.81	2.23
ACRARU	2017	375.0	0.0881	1.92	2.45
ACRSCH	2017	21099.0	0.0000	1.57	2.43
ACRSCI	2017	12985.0	0.0000	1.44	2.74
PARCAE	2017	25100.0	0.0000	1.45	1.98
EMBSCH	2017	2183.0	0.0000	0.53	1.45
PHYCOL	2017	5685.0	0.0011	1.58	2.00
ACRARU	2018	1008.5	0.0642	2.13	2.62
ACRSCH	2018	138959.5	0.0172	2.41	2.27
ACRSCI	2018	138701.5	0.0036	2.36	2.16
PARCAE	2018	21703.5	0.0000	2.14	1.72
EMBSCH	2018	3448.0	0.0007	1.10	1.53
PHYCOL	2018	54035.5	0.0000	2.24	1.79

 A. Appendix

Table A.9.: Mann-Whitney U-Test results for each year (2019-2023)

species	year	u-statistic	p-value	mean fat (ILLM)	mean fat (HOH)
ACRARU	2019	2280.0	0.1334	2.17	2.48
ACRSCH	2019	31847.0	0.1280	1.94	1.85
ACRSCI	2019	35695.0	0.8615	1.87	1.93
PARCAE	2019	97824.5	0.0027	2.05	1.89
EMBSCH	2019	3479.5	0.0000	0.96	1.71
PHYCOL	2019	23858.5	0.3302	1.97	1.97
ACRARU	2020	1897.0	0.6755	2.30	2.38
ACRSCH	2020	43838.0	0.0000	2.19	1.79
ACRSCI	2020	27212.5	0.2710	1.89	2.22
PARCAE	2020	6076.0	0.1156	1.74	1.57
EMBSCH	2020	1886.0	0.0468	1.14	1.40
PHYCOL	2020	18153.0	0.0294	2.17	1.95
ACRARU	2021	2278.0	0.0082	2.06	2.46
ACRSCH	2021	189555.5	0.0000	2.69	1.86
ACRSCI	2021	76024.5	0.0865	2.22	2.23
PARCAE	2021	72609.5	0.0000	1.96	1.39
EMBSCH	2021	1428.0	0.6671	1.11	1.03
PHYCOL	2021	8439.0	0.0022	2.23	1.84
ACRARU	2022	4029.0	0.1473	2.22	2.60
ACRSCH	2022	254381.0	0.0000	2.64	1.87
ACRSCI	2022	141786.0	0.0000	2.37	1.91
PARCAE	2022	21580.0	0.0029	2.00	1.76
EMBSCH	2022	8811.5	0.1376	1.42	1.58
PHYCOL	2022	15729.0	0.0512	2.32	2.09
ACRARU	2023	2592.0	0.0106	2.45	3.02
ACRSCH	2023	134172.0	0.0824	2.23	2.14
ACRSCI	2023	46981.5	0.4344	2.24	2.49
PARCAE	2023	181642.0	0.4801	1.80	1.83
EMBSCH	2023	5041.5	0.1684	1.31	1.44
PHYCOL	2023	43326.5	0.0070	2.19	2.02

B. Appendix: Python Code

B.1. Data cleaning pipeline

```python
"""# Helper functions"""

def fillMissingArt(df):
    print("--- Empty species rows before fill", df[df['species'].isnull()].shape[0])
    print("--- Empty Art rows before fill", df[df['Art'].isnull()].shape[0])

    # Load aktuell_species-art-mapping.csv
    species_art_mapping = pd.read_csv('/content/drive/MyDrive/lisa masterarbeit
    ↪  daten/aktuell_species-art-mapping_mit_umlauten.csv', delimiter=";",
    ↪  encoding='ISO-8859-1')

    # Create dictionaries for easy lookup
    species_to_art = dict(zip(species_art_mapping['species'],
    ↪  species_art_mapping['Art']))
    art_to_species = dict(zip(species_art_mapping['Art'],
    ↪  species_art_mapping['species']))

    # Function to fill missing values
    def fill_art_species(row):
        if pd.isna(row['Art']) and not pd.isna(row['species']):
            row['Art'] = species_to_art.get(row['species'], row['Art'])
        elif pd.isna(row['species']) and not pd.isna(row['Art']):
            row['species'] = art_to_species.get(row['Art'], row['species'])

        # replace invalid Art values with mapping
        if row['Art'] in [0, 111111]:
            if row['species'] in species_to_art:
                row['Art'] = species_to_art[row['species']]
            else:
                return None  # will be removed later
        return row

    # Apply the function to each row
    df = df.apply(fill_art_species, axis=1)

    print("--- Empty species rows after fill", df[df['species'].isnull()].shape[0])
    print("--- Empty Art rows after fill", df[df['Art'].isnull()].shape[0])

    # remove rows where "Art" and species is empty
    df = df[~df['Art'].isnull()]
    df = df[~df['species'].isnull()]

```

 B. Appendix: Python Code

```
    print("--- Empty species rows after remove", df[df['species'].isnull()].shape[0])
    print("--- Empty Art rows after remove", df[df['Art'].isnull()].shape[0])

    # print df length
    print("df length after mapping fill", len(df))

    return df

def fill_blank_rtypes_by_rules(df):

    # Ensure relevant columns are present
    if 'ring_nr' not in df.columns or 'date' not in df.columns or 'rtype' not in df.columns:
        raise ValueError("The file must contain 'ring_nr', 'date', and 'rtype' columns.")

    # Convert date to year format
    df['date'] = pd.to_datetime(df['date'], errors='coerce')
    df['year'] = df['date'].dt.year

    # Initialize rtype column
    df['rtype'] = df['rtype'].fillna('')

    # Dictionary to store the first captures
    first_captures = {}

    # Apply the rule
    for index, row in df.iterrows():
        ring_nr = row['ring_nr']
        year = row['year']

        if ring_nr not in first_captures:
            # If the bird has never been captured
            first_captures[ring_nr] = year
            df.at[index, 'rtype'] = 'e'
        else:
            if year == first_captures[ring_nr]:
                # If it's the same year as the first capture
                df.at[index, 'rtype'] = 'w'
            else:
                # If it's a new year and the bird has been captured before
                first_captures[ring_nr] = year
                df.at[index, 'rtype'] = 'ew'
                # Set all subsequent captures in the same year to 'w'
                df.loc[(df['ring_nr'] == ring_nr) & (df['year'] == year) & (df['rtype'] == ''), 'rtype'] = 'w'
    print("fill_blank_rtypes_by_rules done")
    return df

def cleanup_rtype(df):
    # Helper function to process individual DataFrames (ILLM and non-ILLM)
    def process_df(sub_df):
        # fill blanks with rules
        sub_df = fill_blank_rtypes_by_rules(sub_df)
```

 B. Appendix: Python Code

```python

        # fill blank hours with 0
        sub_df['time_hour'] = sub_df['time_hour'].fillna(0)

        # Ensure date is datetime
        sub_df['date'] = pd.to_datetime(sub_df['date'])
        # Add year column for convenience
        sub_df['year'] = sub_df['date'].dt.year

        # Split the data into 'e' and 'w' types for easier processing
        df_e = sub_df[sub_df['rtype'] == 'e']
        df_w = sub_df[sub_df['rtype'] == 'w']

        # Function to process each group of the same ring_nr and year
        def process_group(group):
            if 'e' in group['rtype'].values:
                e_entry = group[group['rtype'] == 'e']
                w_entries = group[group['rtype'] == 'w']

                if not w_entries.empty:
                    # Keep only the last 'w' entry of the year
                    last_w_entry = w_entries.loc[w_entries.groupby(['date'])['time_hour'].idxmax()].iloc[-1:]

                    # If the last 'w' entry is on the same day as an 'e' entry, discard it
                    if last_w_entry['date'].values[0] == e_entry['date'].values[0]:
                        return e_entry

                    return pd.concat([e_entry, last_w_entry])
                else:
                    return e_entry
            else:
                # If there are only 'w' entries, keep the first and last 'w'
                first_w_entry = group.iloc[0:1]
                last_w_entry = group.iloc[-1:]

                # Rename the first 'w' to 'ew'
                first_w_entry.loc[first_w_entry.index, 'rtype'] = 'ew'

                # Discard if the last 'w' entry is on the same day as the first
                if first_w_entry['date'].values[0] == last_w_entry['date'].values[0]:
                    return first_w_entry

                return pd.concat([first_w_entry, last_w_entry])

        # Apply the processing function to each group of ring_nr and year
        grouped = sub_df.groupby(['ring_nr', 'year'])
        result_df_list = []

        total_groups = len(grouped)
        processed_groups = 0

        for name, group in grouped:
```

 B. Appendix: Python Code

```
            processed_group = process_group(group)
            result_df_list.append(processed_group)

            # Update and print progress
            processed_groups += 1
            progress = (processed_groups / total_groups) * 100
            if progress % 10 == 0:  # Print only at 10% intervals
                print(f"Processing progress: {progress:.2f}% ({processed_groups}/{total_groups} groups processed) - {pd.Timestamp.now()}")

        result_df = pd.concat(result_df_list).reset_index(drop=True)
        # drop year column before returning
        result_df = result_df.drop(columns=['year'])
        return result_df

    # Split the DataFrame into two based on the station column
    df_illm = df[df['station'] == 'ILLM']
    df_non_illm = df[df['station'] != 'ILLM']

    # Process each DataFrame separately
    processed_df_illm = process_df(df_illm)
    processed_df_non_illm = process_df(df_non_illm)

    # Combine the two DataFrames again
    final_df = pd.concat([processed_df_illm, processed_df_non_illm]).reset_index(drop=True)

    print("cleanup_rtype done")
    return final_df

def removeWiederfang(df):

  # First, ensure the 'date' column is correctly converted to datetime
  df['date'] = pd.to_datetime(df['date'], errors='coerce')  # Converts non-convertible values to NaT (Not a Time)

  # Check the datatype of 'date' after conversion
  print(df['date'].dtype)  # This should output 'datetime64[ns]'

  # If the output is not 'datetime64[ns]', further investigation into the data values is required

  # Assuming the date conversion is now correct, continue with your operations
  df['year'] = df['date'].dt.year
  df = df.sort_values(by=['ring_nr', 'year', 'date'])

  # Calculate first and last capture dates within each year for each ring_nr
  capture_dates = df.groupby(['ring_nr', 'year'])['date'].agg(['min', 'max'])

  # Calculate the difference in days
  capture_dates['days_diff'] = (capture_dates['max'] - capture_dates['min']).dt.days
  capture_dates = capture_dates.reset_index()
```

 B. Appendix: Python Code

```python

  # Merge this back to the original dataframe
  df = df.merge(capture_dates[['ring_nr', 'year', 'days_diff']], on=['ring_nr', 'year'], how='left')
  df['days_diff'] = df.groupby(['ring_nr', 'year'])['days_diff'].transform('first')

  # Flag and filter operations
  df['flag'] = df['days_diff'] > 14
  stationary_birds = df[df['flag']]['ring_nr'].unique()
  filtered_df = df[~df['ring_nr'].isin(stationary_birds)]

  # remove columns
  filtered_df = filtered_df.drop(columns=['year', 'days_diff', 'flag'])

  return filtered_df

def clean_df(combined_df):
  print("----------------------")
  print("Cleaning dataframe started - ", len(combined_df))
  print("----------------------")

  # remove empty date rows and make them datetime values
  combined_df = combined_df[~combined_df['date'].isnull()]
  combined_df['date'] = pd.to_datetime(combined_df['date'], errors='coerce')
  print("remove empty date rows. Length:", len(combined_df))

  # Remove unwanted years
  ranges = [
      (pd.Timestamp('1974-01-01'), pd.Timestamp('1983-12-31')),
      (pd.Timestamp('1989-01-01'), pd.Timestamp('1993-12-31')),
      (pd.Timestamp('2009-01-01'), pd.Timestamp('2010-12-31')),
      (pd.Timestamp('2017-01-01'), pd.Timestamp('2023-12-31'))
  ]

  # remove rows where ring_nr is empty
  combined_df = combined_df[~combined_df['ring_nr'].isnull()]
  print("remove rows where ring_nr is empty. Length:", len(combined_df))

  def in_date_ranges(date, ranges):
      # Convert `date` to a pandas Timestamp
      date = pd.to_datetime(date)
      for start, end in ranges:
          if start <= date <= end:
              return True
      return False

  # Filter unwated years for ILMM only
  def filter_unwanted_years(row, ranges):
      if row['station'] == "ILLM":
          return in_date_ranges(row['date'], ranges)
      return True

  # Apply the filtering function
```

 B. Appendix: Python Code

```python
combined_df = combined_df[combined_df.apply(lambda row: filter_unwanted_years(row, ranges), axis=1)]
print("keep specific ranges for station ILLM. Length: ", len(combined_df))

# Remove if station column contains "Infozentrum Nationalpark"
combined_df = combined_df[~combined_df['station'].str.contains('Infozentrum Nationalpark')]
print("remove if station column contains 'Infozentrum Nationalpark' Length:", len(combined_df))

# Rename station value from "ILL" to "ILLM"
combined_df.loc[:, 'station'] = combined_df['station'].replace('ILL', 'ILLM')
combined_df.loc[:, 'station'] = combined_df['station'].replace('Biologische Station Illmitz', 'ILLM')
print("rename station value from 'ILL' to 'ILLM' Length:", len(combined_df))

# Remove row if species or art column contains 0 or 111111
combined_df = combined_df[~combined_df['species'].isin(['0', '111111'])]
combined_df = combined_df[~combined_df['species'].isin([0, 111111])]
combined_df = combined_df[~combined_df['Art'].isin(['0', '111111'])]
combined_df = combined_df[~combined_df['Art'].isin([0, 111111])]
print("remove row if species or Art column contains '0' or '111111' Length:", len(combined_df))

# Remove row if date column is empty
combined_df = combined_df[~combined_df['date'].isnull()]
print("remove row if date column is empty Length:", len(combined_df))

# If time_hour contains values like this: 12:45, keep the 12 and move the 45 into the time_minute column
combined_df['time_hour'] = combined_df['time_hour'].astype(str)
combined_df['time_minute'] = combined_df['time_hour'].str.split(':').str[1]
combined_df['time_hour'] = combined_df['time_hour'].str.split(':').str[0]
print("split hour into hour and minute")

# Convert time_hour and time_minute back to int, keeping NaN where applicable
combined_df['time_hour'] = pd.to_numeric(combined_df['time_hour'], errors='coerce').astype('Int64')
combined_df['time_minute'] = pd.to_numeric(combined_df['time_minute'], errors='coerce').astype('Int64')
print("convert hour and minute to int")

# lowercase the rtype column
combined_df['rtype'] = combined_df['rtype'].str.lower()
print("lowercase the rtype column")

# if k in rtype replace it with e
combined_df['rtype'] = combined_df['rtype'].str.replace('k', 'e')
print("replace k with e")

# uppercase species column
combined_df['species'] = combined_df['species'].str.upper()
print("uppercase species column")
```

 B. Appendix: Python Code

```python

# Remove rows where the "net" column has these values: A, B, C, IZ, K, KE, PR, R, R1, R2, R3, R4, R5, RE
combined_df = combined_df[~combined_df['net'].isin(['A', 'B', 'C', 'IZ', 'K', 'KE', 'PR', 'R', 'R1', 'R2', 'R3', 'R4', 'R5', 'RE'])]
print("remove invalid rows from net column. Length:", len(combined_df))

# remove rows where net column is empty
combined_df = combined_df[~(combined_df['net'].isnull())]
print("remove rows where net column is empty", len(combined_df))

# make net str with leading zeros to int
combined_df['net'] = combined_df['net'].astype(int)

 # Add year column for convenience
combined_df['year'] = combined_df['date'].dt.year

# remove rows where station is ILLM, years are between 1974 and 1993 and net is between 15 and 22
# 1973 and 1994 must be used for the between function
combined_df = combined_df[~((combined_df['year'].between(1973, 1994)) & (combined_df['station'] == 'ILLM') & (combined_df['net'].between(15, 22)))]
print("remove rows where station is ILLM, years are between 1974 and 1993 and net is between 15 and 22", len(combined_df))

# remove rows where station is ILLM, years from 2007 to today and net is 5 or higher
combined_df = combined_df[~((combined_df['year'].between(2006, pd.Timestamp.now().year)) & (combined_df['station'] == 'ILLM') & (combined_df['net'] >= 5))]
print("remove rows where station is ILLM, years from 2007 to today and net is 5 or higher", len(combined_df))

# remove 'nm' values all columns but not within longer strings
columns = [col for col in combined_df.columns if col in ["feather_le", "winglength", "fat", "muscle", "tarsus"]]
try:
  for col in columns:
      combined_df[col] = combined_df[col].str.replace('nm', '')
      combined_df[col] = pd.to_numeric(combined_df[col], errors='coerce')
except Exception as e:
    print(f"Error converting column {col}: {e}")
print("remove 'nm' values all columns but not within longer strings", len(combined_df))

# Remove row if age contains A, 7, 9
combined_df = combined_df[~combined_df['age'].isin(['A', '7', '9'])]
print("remove wrong age values", len(combined_df))

# replace age 5 and 6 with 4
combined_df['age'] = combined_df['age'].replace(['5', '6'], '4')
print("replace age 5 and 6 with 4", len(combined_df))

```

 B. Appendix: Python Code

```python
# In ringer_1 column, ensure two letters. If there are longer names, take the first letters and combine them
def shorten_name(name):
    if pd.isna(name):
        return name
    elif len(name) <= 2:
        return name
    else:
        shortened = "".join([part[0] for part in name.split()])
        return shortened[:2]

# fill empty art or species rows
combined_df = fillMissingArt(combined_df)
print("fill missing art or species rows")

# Apply the function to the 'ringer_1' column
combined_df.loc[:, 'ringer_1'] = combined_df['ringer_1'].apply(shorten_name)
print("rename ringers to two letter names")

# remove if date is empty
combined_df = combined_df[~combined_df['date'].isnull()]
print("remove if date is empty", len(combined_df))

# remove all dates that are not july, august, september, october in any year
combined_df['date'] = pd.to_datetime(combined_df['date'], format='%Y-%m-%d')
combined_df = combined_df[combined_df['date'].dt.month.isin([7, 8, 9, 10])]
print("remove all dates that are not july, august, september in any year", len(combined_df))

# remove rows where the same ring_rn appears multiple times in a period longer than 14 days
combined_df = removeWiederfang(combined_df)
print("remove rows where the same ring_rn appears multiple times in a period longer than 14 days", len(combined_df))

# cleanup rtype
combined_df = cleanup_rtype(combined_df)
print("cleanup rtype", len(combined_df))

# Enforce dtypes
dtype_enforcements = {
    'time_hour': 'Int64', 'time_minute': 'Int64', 'age': 'Int64',
    'sex': 'Int64', 'moult_bf_i': 'Int64', 'feather_le': 'float',
    'winglength': 'float', 'fat': 'float', 'muscle': 'float',
    'weight': 'float', 'tarsus': 'float', 'ringer_1': 'str',
    'pentade': 'float'
}

for col, dtype in dtype_enforcements.items():
    combined_df[col] = combined_df[col].astype(dtype, errors='ignore')
print("enforce types")
```

 B. Appendix: Python Code

```
    print("----------------------")
    print("Cleaning dataframe done")
    print("----------------------")

    return combined_df

def analyze_df(df):
    # Display basic information about the dataframe
    print("Basic Information about the DataFrame:")
    df.info()

    # Display the first few rows of the dataframe
    print("\nFirst Few Rows of the DataFrame:")
    print(df.head())

    # Display summary statistics of the dataframe
    print("\nSummary Statistics of the DataFrame:")
    print(df.describe(include='all'))

    # Check for missing values
    missing_values = df.isnull().sum()
    missing_percentage = (missing_values / len(df)) * 100
    missing_data = pd.DataFrame({'Missing Values': missing_values, 'Percentage':
    ↪  missing_percentage})

    print("\nMissing Data in the DataFrame:")
    print(missing_data)

    # Display data types of each column
    data_types = df.dtypes
    print("\nData Types of Each Column:")
    print(data_types)

    # Display unique values for each column
    unique_values = {col: df[col].unique() for col in df.columns}

    print("\nUnique Values in Each Column (First 10 Values):")
    for col, values in unique_values.items():
        print(f"Column: {col}")
        print(f"Unique Values: {values[:10]}")  # Display only the first 10 unique
        ↪  values for brevity
        print("\n")

def create_master_file(df):
    current_time = datetime.now().strftime('%Y-%m-%d-%H-%M')
    output_filename = f'master_file-{current_time}.csv'
    master_csv_path = f'/content/drive/MyDrive/lisa masterarbeit
    ↪  daten/output/{output_filename}'
    df.to_csv(master_csv_path, index=False, sep=';')
    print(f"Master CSV file created successfully! - master_file-{current_time}.csv")

import pandas as pd
import os
from datetime import datetime
```

 B. Appendix: Python Code

```python

# Define expected columns and data types
columns = [
    'station', 'ring_nr', 'rtype', 'euring', 'species', 'Art', 'date', 'time_hour', 'time_minute',
    'net', 'age', 'sex', 'moult_bf_i', 'feather_le', 'winglength', 'fat', 'muscle',
    'weight', 'tarsus', 'ringer_1', 'pentade', 'Luftfeuchte_Tmittel', 'Lufttemperatur_Tmittel',
    'Wasserstand', 'Windgeschindigkeit_Max', 'Windgeschwindigkeit_Tmittel', 'Windrichtung'
]

dtypes = {
    'station': 'str', 'ring_nr': 'str', 'rtype': 'str', 'euring': 'str', 'species': 'str', 'Art': 'str',
    'date': 'str', 'time_hour': 'str', 'time_minute': 'str', 'net': 'int', 'age': 'str',
    'sex': 'str', 'moult_bf_i': 'str', 'feather_le': 'str', 'winglength': 'str',
    'fat': 'str', 'muscle': 'str', 'weight': 'str', 'tarsus': 'str', 'ringer_1': 'str',
    'pentade': 'str', 'Luftfeuchte_Tmittel': 'float', 'Lufttemperatur_Tmittel': 'float',
    'Wasserstand': 'float', 'Windgeschindigkeit_Max': 'float', 'Windgeschwindigkeit_Tmittel': 'float',
    'Windrichtung': 'float'
}

def combine_all_bird_excel_files(columns, dtypes):
    """
    Processes all Excel files in a predefined directory, standardizes their structure,
    and combines them into a single CSV file.
    """
    # set dtype for net to str in here (later in cleaning it should be int64)
    dtypes['net'] = 'str'

    # Define the directory paths
    directory_path = '/content/drive/MyDrive/lisa masterarbeit daten/birds/standardized/'
    output_path = '/content/drive/MyDrive/lisa masterarbeit daten/output/combined_file_standardized.csv'

    # Check if the directory exists
    if not os.path.exists(directory_path):
        raise FileNotFoundError(f"The directory {directory_path} does not exist.")

    dataframes = []
    total_files = len([f for f in os.listdir(directory_path) if f.endswith('.xlsx')])
    processed_files = 0

    for filename in os.listdir(directory_path):
        if filename.endswith('.xlsx'):
            file_path = os.path.join(directory_path, filename)
```

 B. Appendix: Python Code

```

        try:
            xls = pd.ExcelFile(file_path)
            for sheet_name in xls.sheet_names:
                df = pd.read_excel(xls, sheet_name=sheet_name, engine='openpyxl', dtype=dtypes)
                df['sheet_name'] = sheet_name  # Add sheet name column

                print(f"Processing {filename} - {sheet_name}")

                # Replace ";" with " " and "," with "."
                df = df.replace(';', ' ', regex=True)
                df = df.replace(',', '.', regex=True)

                # Drop duplicates and rows with all missing values
                df = df.drop_duplicates()
                df = df.dropna(how='all')

                # Handle missing columns
                missing_cols = [col for col in columns if col not in df.columns]
                missing_cols_str = ' '.join(missing_cols)

                df['origin_file'] = filename
                df['missing_columns'] = missing_cols_str

                for col in columns:
                    if col not in df.columns:
                        df.loc[:, col] = pd.NA

                # Standardize date format
                if 'date' in df.columns:
                    try:
                        df = df[~df['date'].isnull()]
                        df['date'] = pd.to_datetime(df['date'], errors='coerce')
                        print(f"Errors after date conversion: {df['date'].isna().sum()}")
                    except Exception as e:
                        print(f"Error converting dates in file {filename}: {e}")

                # Reorganize columns
                df = df[['origin_file', 'sheet_name'] + columns + ['missing_columns']]
                dataframes.append(df)

        except Exception as e:
            print(f"Error reading file {filename}: {e}")
            continue

        processed_files += 1
        print(f"------ Processed file {processed_files}/{total_files}: {filename}")
        print("")

```

 B. Appendix: Python Code

```python
    # Combine all processed dataframes
    combined_df = pd.concat(dataframes, ignore_index=True)

    # Save combined dataframe to CSV
    combined_df.to_csv(output_path, index=False, sep=';')
    print("----------------------")
    print("Combined dataframe created.")
    print("Length:", len(combined_df))
    print("Saved to:", output_path)
    print("----------------------")
    return combined_df

"""# Combine files"""

# Only call the function if the combined file does not already exist
combined_file_path = '/content/drive/MyDrive/lisa masterarbeit daten/output/combined_file.csv'

if not os.path.exists(combined_file_path):
    print("Combined file does not exist. Starting processing...")
    combined_df = combine_all_bird_excel_files(columns, dtypes)

else:
    print("Combined file already exists at:", combined_file_path)
    combined_df = pd.read_csv(combined_file_path, sep=';', dtype=dtypes)
    print("Loaded from file")

combined_df = combine_all_bird_excel_files(columns, dtypes)

cleaned_df = clean_df(combined_df)

# Remove duplicates without considering the origin_file and missing_columns columns, but keep the first occurrence
cleaned_df = cleaned_df.drop_duplicates(subset=[col for col in columns if col not in ['origin_file', 'missing_columns', 'sheet_name']], keep='first')
print("remove full duplicates", len(cleaned_df))
cleaned_df.reset_index(drop=True, inplace=True)
print("reset indexes", len(cleaned_df))

create_master_file(cleaned_df)
```