

GLM zur Analyse von Einflussfaktoren auf Zuentfernungen

Anhang zum Artikel:

Jahreslebensraum, Zugphänologie und Ansiedlungsverhalten ostdeutscher Kormorane *Phalacrocorax carbo sinensis* nach Beringungsdaten aus dem Zeitraum 2009 bis 2020

Carla Behringer, Klaus-Dieter Feige, Ulrike Schlägel & Christof Herrmann

1 GLM mit kategorialer Herkunftsregion

1.1 Visualisierung der Daten

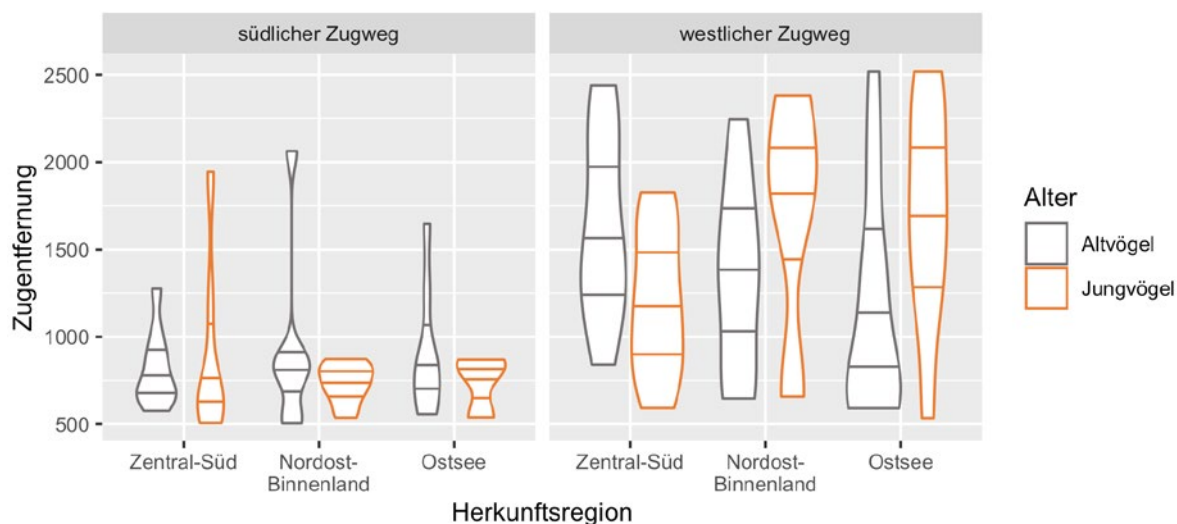


Abb. A1: Zuentfernung nach Herkunftsregion, für unterschiedliche Zugwege und geteilt nach Alter. Der Violin-Plot gibt die Verteilung der Daten an (etwas detaillierter als ein Boxplot). Horizontale Linien sind, von unten nach oben, die Quantile 0, 0,25, 0,5 (Median), 0,75, 1 der Daten.

Der Plot legt nahe, dass es einen Unterschied zwischen dem südlichen und dem westlichen Zugweg gibt. Außerdem gibt es möglicherweise einen Effekt der Herkunftsregion (zumindest auf dem westlichen Zugweg), der aber auch vom Alter beeinflusst wird. Um dies zu prüfen, passen wir ein GLM an die Daten an, in dem wir alle Interaktionen zulassen.

1.2 Das Modell

Wir benutzen ein GLM (generalisiertes lineares Modell) mit Gamma-Verteilung und Identitäts- Link-Funktion. D.h. wir modellieren einen linearen Zusammenhang zwischen der mittleren Zuentfernung und den drei erklärenden Variablen Alter, Herkunftsregion und Zugweg. Wir nehmen für die abhängige Variable "Zuentfernung" eine Gamma-Verteilung an, da die beobachteten Zuentfernungen in den meisten Gruppen/Untergruppen asymmetrisch verteilt sind (rechtsschief; siehe Daten-Visualisierung). Für die Herkunftsregion nehmen wir als Referenzkategorie die Region Zentral-Süd, weil sich bei dieser ein ausgeglichenes Verhältnis zwischen der West- und Süd-Route gezeigt hat.

```
mod <- df %>% glm(Entfernung ~ Zugweg * Herkunft * Alter,  
  data = .,  
  family = Gamma(link = „identity“))
```

Tab. A1: Ergebnis des GLMs. Hervorgehoben sind Effekte mit $p < 0,1$. Schätzwert und Standardfehler sind zur besseren Lesbarkeit gerundet auf die nächste ganze Zahl. Zugentfernung in einzelnen Faktorkombinationen ergeben sich durch Zusammenrechnen der Schätzwerte aller relevanten Effekte, woraus wiederum Unterschiede zwischen Faktorkombinationen berechnet werden können.

Term	Schätzwert	Std.-Fehler	t-Wert	p-Wert
(Intercept)	815	114	7,14	<0,001
Zug_West	813	221	3,67	<0,001
Herkunft_Ostsee	90	176	0,51	0,610
Herkunft_NOBinnen	108	185	0,58	0,562
Alter_Jung	69	151	0,46	0,647
Zug_West x Herkunft_Ostsee	-565	276	-2,04	0,043
Zug_West x Herkunft_NOBinnen	-392	300	-1,30	0,195
Zug_West x Alter_Jung	-494	280	-1,77	0,080
Herkunft_Ostsee x Alter_Jung	-236	270	-0,87	0,384
Herkunft_NOBinnen x Alter_Jung	-260	237	-1,10	0,275
Zug_West x Herkunft_Ostsee x Alter_Jung	1.218	396	3,07	0,003
Zug_West x Herkunft_NOBinnen x Alter_Jung	1.040	441	2,35	0,020

Das Ergebnis zeigt:

- Vögel der Westroute haben prinzipiell längere Zugentfernungen
- Unterschiede nach Herkunftsregion und Alter zeigen sich nur auf der West-Route und zwar in gegenseitiger Wechselbeziehung:
 - Bei den Altvögeln ziehen Individuen der Ostsee deutlich weniger weit als Individuen aus der Zentral-Süd Region (der Effekt für Altvögel der Region Nordost-Binnen ist hingegen nicht signifikant). Bei den Jungvögeln hingegen ziehen Individuen der Herkunftsregionen Ostsee und Nordost-Binnen weiter als Individuen aus der Zentral-Süd Region.
 - Gleichzeitig führt dies dazu, dass bei Vögeln aus den Herkunftsregionen Ostsee und Nordost-Binnen Jungvögel deutlich weiter ziehen als Altvögel. Hingegen bei Vögeln aus der Region Zentral-Süd ziehen junge Vögel weniger weit als Altvögel (dieser Unterschied ist jedoch nicht sehr deutlich mit einem p-Wert von 0,08.)

1.3 Modell Diagnostik

Wir überprüfen unsere Modellannahmen (insb. die Gamma-Verteilung) mithilfe von Residuen-Diagnostik, bereit gestellt durch das R Paket ``DHARMA'' (Hartig 2020).

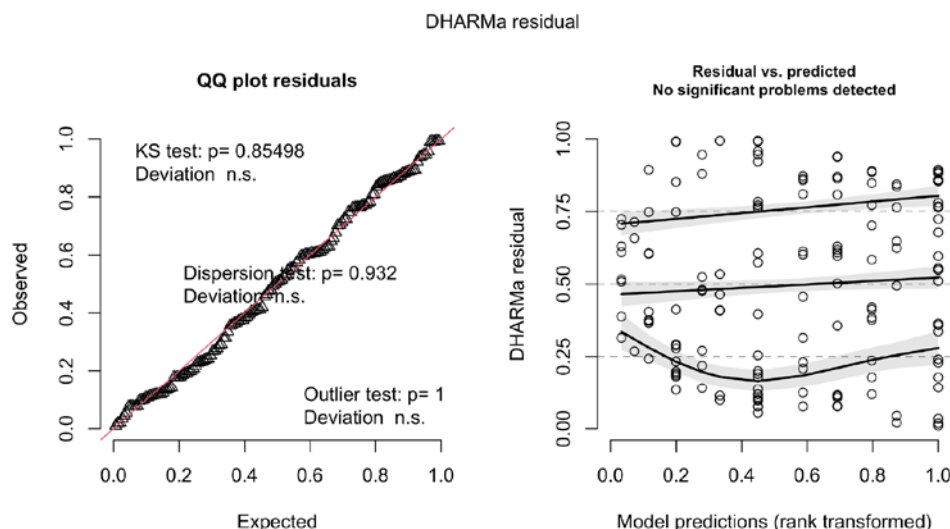


Abb. A2: Links: QQ-Plot der Residuen. Rechts: Scatter-Plot der Residuen, geplottet gegen gefittete (d.h. laut Modell vorhergesagte, Rang-transformierte) Werte.

Laut Residuen-Diagnostik sind keine Probleme sichtbar. Der QQ-Plot (Abb. A2 links) zeigt keine Abweichungen der beobachteten Verteilung von der erwarteten, d. h. die Annahme der Gamma Verteilung ist passend. Die visuelle Darstellung von Modellvorhersagen und den Daten zeigt, dass es auf dem südlichen Zugweg einzelne Zuentfernungen gibt, die größer sind als die Mehrheit (Abb. 14 im Hauptartikel). Der Scatter-Plot der Residuen (Abb. A2 rechts) zeigt allerdings weder starke Muster noch Ausreißer, so dass diese Werte noch im Rahmen der Modellannahmen (Gamma-Verteilung für Zuentfernungen) liegen.

1.4 Alternative Modelle

Die kleinen p-Werte der dreifach-Interaktionseffekte legen nahe, dass einfachere Modelle der Situation nicht vollständig gerecht werden können. Wir können dennoch gegenchecken, ob ein einfacheres, rein additives Modell die Daten ähnlich gut beschreibt. Dazu verwenden wir einen Likelihood-Ratio-Test:

```
## Likelihood ratio test
##
## Model 1: Entfernung ~ Zugweg * Herkunft * Alter
## Model 2: Entfernung ~ Alter + Zugweg + Herkunft
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   13 -1136.4
## 2    6 -1145.8 -7 18.713  0.009137 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Der Test ergibt einen besseren Fit des komplexen Modells gegenüber dem additiven Modell ($p = 0,009$). Da es allerdings grundsätzlich problematisch ist, Modell-Wahl mit einer Interpretation von t-Tests einzelner Effekte zu kombinieren, sehen wir von einer umfangreichen Modell-Wahl ab und belassen es bei der obigen Interpretation des vollen Modells.

2 GLM mit kontinuierlichem Hochwert des Herkunftsortes

In der obigen Analyse hat sich gezeigt, dass auf dem westlichen Zugweg Jungvögel, die von der Ostsee oder dem nordöstlichen Binnenland stammen, tendenziell weitere Zuentfernungen zurücklegen als Jungvögel des zentralen und südlichen Ostdeutschland. Dieser Effekt ist größer für Individuen der Ostsee als Individuen des nordöstlichen Binnenlandes. Um ein mögliches Nord-Süd-Gefälle im Einfluss der Herkunftsregion genauer zu untersuchen, haben wir ein Modell angepasst, in dem wir anstelle der kategorialen Variable "Herkunftsregion" als kontinuierliche Variable den Hochwert (in Metern, UTM Zone 33) der genauen Herkunftsorte benutzen.

```
mod.hochwert <- df %>%
  glm(Entfernung ~ Zugweg * Alter * Hochwert.shift,
      data=.,
      family=Gamma(link="identity"))
```

Tab. A2: Ergebnis des GLMs mit kontinuierlichem Hochwert. Hervorgehoben sind Effekte mit $p < 0,1$.

Term	Schätzwert	Std.-Fehler	t-Wert	p-Wert
(Intercept)	781,507	132,760	5,89	<0,001
Zug_West	897,014	229,440	3,91	<0,001
Alter_Jung	116,532	173,661	0,67	0,503
Hochwert_Herkunft	0,000	0,001	0,79	0,428
Zug_West x Alter_Jung	-645,657	304,355	-2,12	0,036
Zug_West x Hochwert_Herkunft	-0,002	0,001	-2,25	0,026
Alter_Jung x Hochwert_Herkunft	-0,001	0,001	-1,28	0,202
Zug_West x Alter_Jung x Hochwert_Herkunft	0,004	0,001	3,43	<0,001

Um die Ergebnisse (Effektgrößen) des Modells besser lesbar zu machen, wurde der Hochwert hier um den minimal vorkommenden Wert verschoben (südlichster Herkunftsort) und später zur graphischen Darstellung wieder zurück transformiert (Abb. 15 im Hauptartikel). Die Transformation wirkt sich nur auf die Effektgrößen des Intercepts und der additiven Terme aus, so dass der geschätzte, allgemeine Intercept die mittlere Zugentfernung eines Altvogels mit Herkunft vom südlichsten Herkunftsort und südlichem Zugweg darstellt. Die Effektgrößen der Interaktionsterme mit dem Hochwert sind weiterhin sehr klein, da die Hochwerte große numerische Werte haben (6-stellig).

Die Ergebnisse untermauern die vorherigen. Vögel des westlichen Zugweges legen per se weitere Entfernungen zurück, während die Effekte von Alter und Herkunft in Wechselbeziehung stehen und nur auf dem westlichen Zugweg deutlich werden. Hier zeigt sich, dass auf dem westlichen Zugweg Jungvögel mit südlicherer Herkunft weniger weit ziehen als Jungvögel mit weiter nördlich gelegener Herkunft (positive Effektgröße von 0,004 mit p-Wert < 0,001). Bei Altvögeln ist der Effekt umgekehrt (negative Effektgröße von -0,002 mit p-Wert 0,026).

Wir überprüfen auch hier unsere Modellannahmen wie beim obigen Modell. Auch hier zeigt die Modell-Diagnostik keine Auffälligkeiten, so dass wir davon ausgehen, dass unsere Modellannahmen in Ordnung waren.

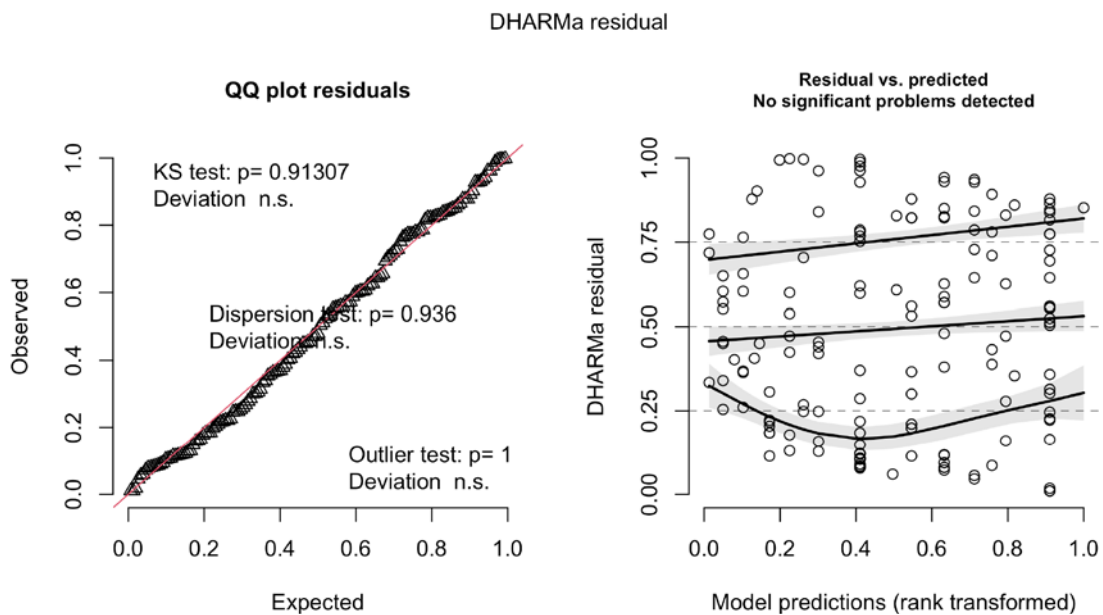


Abb. 3: Links: QQ-Plot der Residuen. Rechts: Scatter-Plot der Residuen, geplottet gegen gefittete (d.h. laut Modell vorhergesagte, Rang-transformierte) Werte.

```
sessionInfo()
## R version 4.2.1 (2022-06-23)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur ... 10.16
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
##
```

```

## other attached packages:
## [1] kableExtra_1.3.4 broom_1.0.0      lmtest_0.9-40      zoo_1.8-10
## [5] DHARMA_0.4.5      ggeffects_1.1.3    ggplot2_3.3.6      viridis_0.6.2
## [9] viridisLite_0.4.0 here_1.0.1          tidyr_1.2.0        flextable_0.8.3
## [13] officer_0.4.4     knitr_1.39         dplyr_1.0.9
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.9         svglite_2.1.0      lattice_0.20-45    assertthat_0.2.1
## [5] rprojroot_2.0.3    digest_0.6.29      utf8_1.2.2         R6_2.5.1
## [9] backports_1.4.1    evaluate_0.16      httr_1.4.3         pillar_1.8.0
## [13] gdtools_0.2.4      rlang_1.0.4        uuid_1.1-0         rstudioapi_0.13
## [17] minqa_1.2.4        data.table_1.14.2 nloptr_2.0.3       Matrix_1.4-1
## [21] rmarkdown_2.14     splines_4.2.1      webshot_0.5.3     lme4_1.1-30
## [25] stringr_1.4.0      munsell_0.5.0      compiler_4.2.1     xfun_0.32
## [29] pkgconfig_2.0.3    systemfonts_1.0.4 base64enc_0.1-3    htmltools_0.5.3
## [33] tidyselect_1.1.2   tibble_3.1.8       gridExtra_2.3      fansi_1.0.3
## [37] withr_2.5.0        MASS_7.3-58.1      grid_4.2.1         nlme_3.1-159
## [41] gtable_0.3.0       lifecycle_1.0.1    DBI_1.1.3          formatR_1.12
## [45] magrittr_2.0.3     scales_1.2.0       zip_2.2.0          cli_3.3.0
## [49] stringi_1.7.8      xml2_1.3.3         generics_0.1.3     vctrs_0.4.1
## [53] boot_1.3-28        tools_4.2.1        glue_1.6.2         purrr_0.3.4
## [57] fastmap_1.1.0      yaml_2.3.5         colorspace_2.0-3   rvest_1.0.2

```