

# Lineare Modelle für Funktionale Antwortvariablen

Simon Müller

Institut für Stochastik und Anwendungen  
Universität Stuttgart

4. April 2012

# Gliederung

- 1 Einleitung
- 2 Funktionale Varianzanalyse
- 3 Lineares Funktionales Modell mit Funktionaler Beobachtung
- 4 Eigene Forschung

# Gliederung

## 1 Einleitung

Was sind funktionale Daten?

Woran sind wir interessiert?

Worin liegen die Herausforderungen?

Referenzen

## 2 Funktionale Varianzanalyse

Datensatz: Vogelpopulation

Das Modell

Resultat

## 3 Lineares Funktionales Modell mit Funktionaler Beobachtung

Datensatz: Sterblichkeitsraten schwedischer Frauen

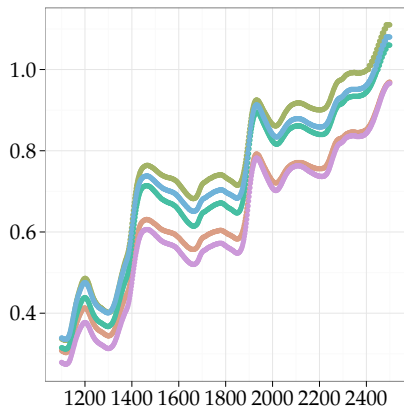
Das Modell

Resultat

## 4 Eigene Forschung

# Was sind funktionale Daten?

**Beispiel:** NIR-Spektrum von Weizen,  
100 Kurven mit jeweils 701 Beobachtungen



Eigenschaften der Daten?

- ▶ Hohe Anzahl an Beobachtungen im Vgl. zur Stichprobengröße
- ▶ Ähnlichkeit des Kurvenverlaufs, Daten sind collinear.
- ▶ Glatte Daten, aber komplexer zugrundeliegender Prozess

# Was sind funktionale Daten?

Im Allgemeinen geht man davon aus, dass

$$y_{ij} = x_i(t_{ij}) + \varepsilon_{ij},$$

dabei sind die  $t_{ij}$  diskrete Abtastungen einer kontinuierlichen Variable  $t$  (Zeit, Frequenz, usw.) und  $x_i(t)$  die funktionalen Daten.

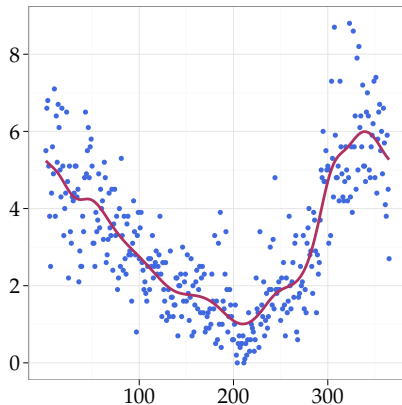
Daten erhält man z.B. aus:

- ▶ Spektralen Messungen (z.B. Nahrungsmittelindustrie, Astronomie),
- ▶ Elektrischen Messungen (z.B. EKG, EEG),
- ▶ Audio-Messungen (z.B. Lungengeräusche, Sprache).

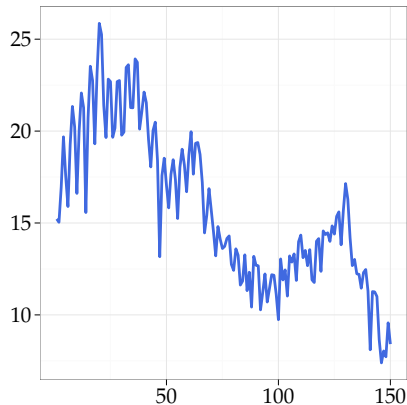
Interpretation des Rauschens  $\varepsilon_{ij}$  hängt von der Art der Daten ab.

# Ausblick: Glättung

## Niederschlag in Vancouver



## log-Periodogramm zum Laut aa



# Woran sind wir interessiert?

- ▶ Deskriptive Darstellung von Funktionen
  - ▶ Erwartungswert, Median
  - ▶ Varianz
  - ▶ Covarianz
- ▶ Beziehung funktionaler Daten zu
  - ▶ Antwortvariablen (z.B. Fettgehalt, Wort)
  - ▶ Antwortfunktionen (z.B. Niederschlag pro Tag)
  - ▶ den anderen Beobachtungen (z.B. Sterblichkeitsraten)
- ▶ Beziehung zu den Ableitungen der funktionalen Daten (z.B. Dynamik des Schreibens)
- ▶ Zeitereignisse in den Funktionen (z.B. Registrierung Handschriftdaten)

# Worin liegen die Herausforderungen?

- ▶ Schätzung der funktionalen Daten  $x_i(t)$  anhand diskreter und verrauschter Daten
- ▶ Numerische Repräsentation unendlich-dimensionaler Objekte
- ▶ Beschreibung statistischer Zusammenhänge zwischen unendlich-dimensionalen Objekten
- ▶ Darstellung der Variation in unendlich-dimensionalen Räumen



# Referenzen (alle von Springer)

## Parametrische funktionale Datenanalyse:

- ▶ Ramsay, Hooker, and Graves, *Functional Data Analysis with R and MATLAB*, 2009
- ▶ Ramsay and Silverman, *Functional Data Analysis*, 2005
- ▶ Ramsay and Silverman, *Applied Functional Data Analysis*, 2002

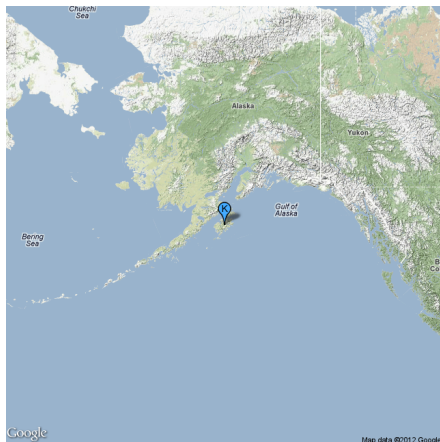
## Nichtparametrische funktionale Datenanalyse:

- ▶ Ferraty and Vieu, *Nonparametric Functional Data Analysis*, 2007

# Gliederung

- ① Einleitung
  - Was sind funktionale Daten?
  - Woran sind wir interessiert?
  - Worin liegen die Herausforderungen?
  - Referenzen
- ② Funktionale Varianzanalyse
  - Datensatz: Vogelpopulation
  - Das Modell
  - Resultat
- ③ Lineares Funktionales Modell mit Funktionaler Beobachtung
  - Datensatz: Sterblichkeitsraten schwedischer Frauen
  - Das Modell
  - Resultat
- ④ Eigene Forschung

# Datensatz: Vogelpopulation

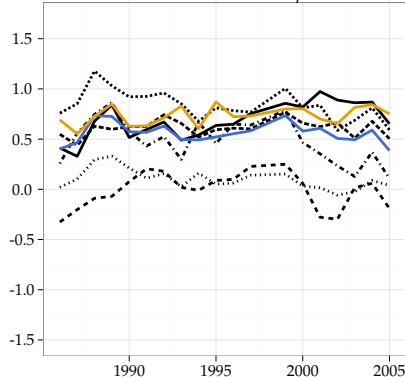


- ▶ Die Anzahl von 15 Vogelarten wurde jährlich seit 1986 im Winter in einer Reihe von Buchten und festen Gebieten auf den Kodiak Inseln (Alaska) von der *Kodiak National Wildlife Refuge* gezählt.
- ▶ In der nachfolgenden Analyse sind die Daten
  - ▶ auf die 2 Buchten Ugalik und Uyak beschränkt und
  - ▶ es werden 13 von den 15 Vogelarten betrachtet.

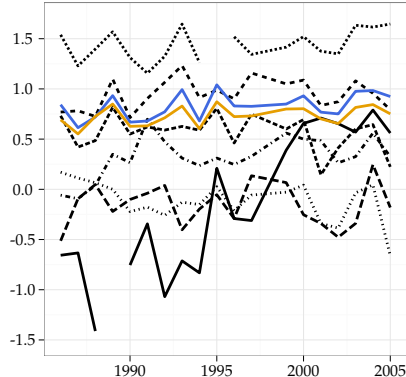
# Fragestellung

Es soll untersucht werden, ob und welchen Einfluss die Ernährung auf die Population der Vogelarten hat.

Faktorstufe: Krustentier/Weichtier



Faktorstufe: Fisch



# Beschreibung

- ▶ Es wird der Einfluss der Ernährung auf den zeitlichen Trend der logarithmierten gemittelten Anzahl von Vögeln analysiert.
- ▶ Die Vogelarten werden in 2 Faktorstufen ( $= \{\textit{Krusten- bzw. Weichtiere; Fische}\}$ ) eingeteilt
  - ▶ Die Stichprobe zur FS *Krustentiere* enthält  $n_1 = 6$  Elemente
  - ▶ Die Stichprobe zur FS *Fische* enthält  $n_2 = 7$  Elemente
- ▶ Zeitraum: 1986 – 2005 (ohne 1998)

# Das Modell

Wir betrachten folgendes Modell:

$$y_{ijk}(t) = \mu(t) + (-1)^i \alpha(t) + \beta_{ij}(t) + \varepsilon_{ijk}(t)$$

- ▶  $i = 1, 2$  indiziert die Nahrungsquelle
- ▶  $j = 1, \dots, n_i$  indiziert die Vogelart in der jeweiligen Nahrungsgruppe
- ▶  $k = 1, 2$  indiziert die Bucht
- ▶ Der funktionale Parameter  $\mu(t)$  bezeichnet den mittleren globalen Trend über alle Vogelarten.

# Das Modell

Wir betrachten folgendes Modell:

$$y_{ijk}(t) = \mu(t) + (-1)^i \alpha(t) + \beta_{ij}(t) + \varepsilon_{ijk}(t)$$

- ▶ Der Parameter  $\alpha(t)$  bezeichnet den Trend über die mittlere Differenz zwischen den Krusten- bzw. Weichtier und den Fisch essenden Vogelarten.
- ▶  $\alpha(t)$  wird mit 1 multipliziert, wenn die Beobachtung zu einer Krusten- bzw. Weichtier essenden Vogelart gehört und mit  $-1$  bei einer Fisch essenden.
- ▶ Der Parameter  $\beta_{ij}(t)$  bezeichnet den Trend, welcher die Abweichungen von  $\mu(t)$  darstellt, für jede Vogelart, die zu einer der Nahrungsgruppen gehört.

# Das Modell

$$y_{ijk}(t) = \mu(t) + (-1)^i \alpha(t) + \beta_{ij}(t) + \varepsilon_{ijk}(t)$$

- ▶ Dies wird für alle  $t$  erreicht durch die Nebenbedingungen

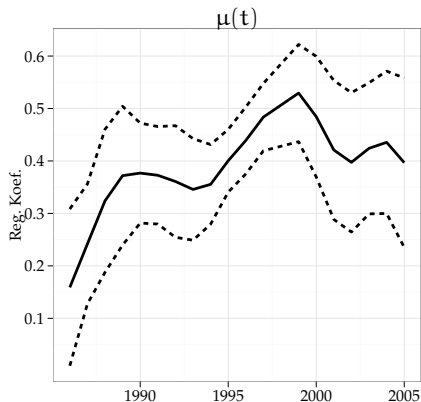
$$\sum_{j=1}^{n_1} \beta_{1j}(t) = 0 \text{ und } \sum_{j=1}^{n_2} \beta_{2j}(t) = 0.$$

- ▶ Der Parameter  $\varepsilon_{ijk}(t)$  bezeichnet die Residuen.
- ▶ Kein Einfluss: Bucht und der Interaktion von Bucht \* Nahrung
- ▶ Dies ergibt 28 Gleichungen: 2 Blöcke der 13 Vogelarten und die beiden Nebenbedingungen.

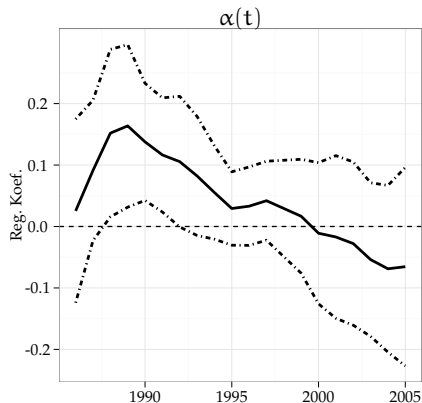


# Resultat

Zeitlicher Trend  $\mu(t)$  über alle Vogelarten und Nahrungsquellen



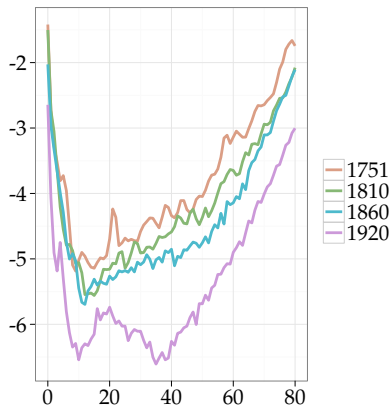
Zeitlicher Trend  $\alpha(t)$  über die mittlere Differenz



# Gliederung

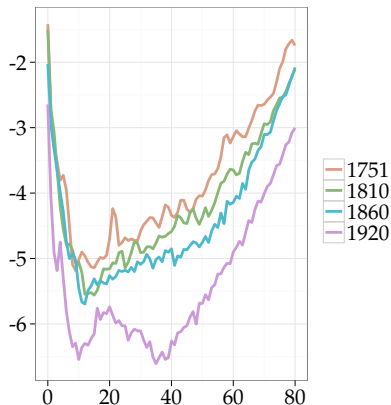
- ① Einleitung
  - Was sind funktionale Daten?
  - Woran sind wir interessiert?
  - Worin liegen die Herausforderungen?
  - Referenzen
- ② Funktionale Varianzanalyse
  - Datensatz: Vogelpopulation
  - Das Modell
  - Resultat
- ③ Lineares Funktionales Modell mit Funktionaler Beobachtung
  - Datensatz: Sterblichkeitsraten schwedischer Frauen
  - Das Modell
  - Resultat
- ④ Eigene Forschung

# Datensatz: Sterblichkeitsraten schwedischer Frauen



- ▶ Logarithmierte Hazardrate schwedischer Frauen im Alter von 0 bis 80 Jahren.
- ▶ Berechnet aus Daten von: [www.mortality.org](http://www.mortality.org)
- ▶ Zeitraum Datensatz: 1757 – 1900

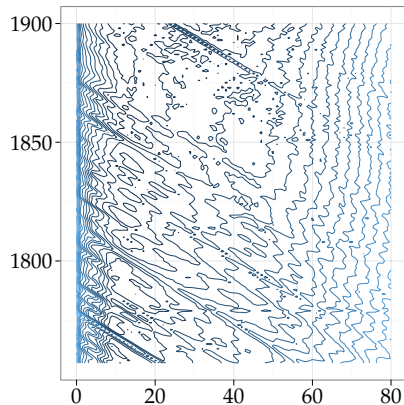
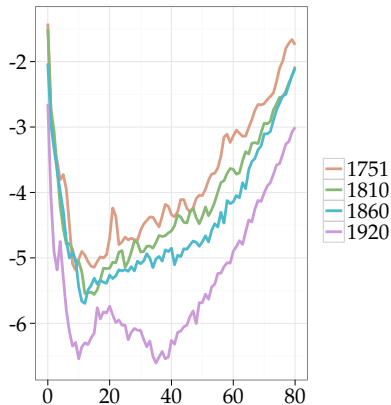
# Datensatz: Sterblichkeitsraten schwedischer Frauen



Sterbewahrscheinlichkeit ist

- ▶ im Säuglingsalter besonders hoch,
- ▶ hat ein Minimum im Teenageralter
- ▶ und nimmt mit dem Alter zu.
- ▶ Mit zunehmender Gesundheit nimmt die Sterbewahrscheinlichkeit ab.
- ▶ Lokale Ausschläge entsprechen z.B. Kriegen, Krankheiten, usw.

# Datensatz: Sterblichkeitsraten schwedischer Frauen



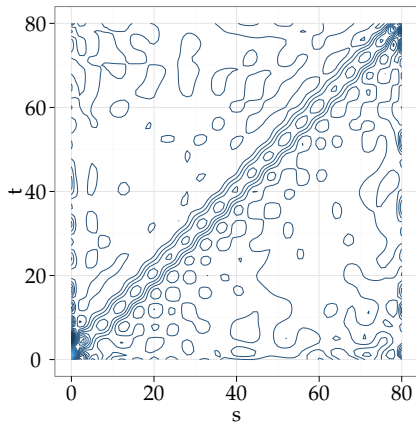
# Das Modell

Wir passen folgendes Auto-regressives Modell an die Daten an:

$$x_{i+1}(t) = \beta_0(t) + \int_{\Omega_t} x_i(s) \beta_1(s, t) ds + \varepsilon_i(t)$$

- ▶  $x_i(t)$  logarithmierte Hazardrate für die Jahre  $i = 1757, \dots, 1900$  zum Alter  $t \in [0, 80]$
- ▶ Integrationsgebiet  $\Omega_t = [0, 80]$
- ▶  $\varepsilon_i(t)$  Rauschen

# Resultat



- ▶  $\beta_0(t) \approx 0$
- ▶  $x_{i+1}(t) \approx \int x_i(s) \beta_1(s, t) ds$
- ▶ Die Koeffizienten haben einen Einfluss auf der Diagonalen,  $\beta_1(s-1, s)$
- ▶ Sterblichkeitsrate des aktuellen Jahres steht in Beziehung zu der im vorigen Jahr

# Eigene bisherige Forschung

- ▶ Nichtparametrische Regression:  $Y = m(X) + \varepsilon$ 
  - ▶ für funktionale und
  - ▶  $\alpha$ -mischende Daten.
  - ▶ Schätzer: Nadaray-Watson Kernschätzer
- ▶ Bandweitenwahl
  - ▶ k-Nächste Nachbarn Kernschätzer
  - ▶ Bootstrapping
- ▶ Gleichmäßige Konvergenzaussagen



**Vielen Dank für Ihre  
Aufmerksamkeit!**