Projekt LVS-IR-Taubenstein

Projektpartner: Sascha Filimon, Roman Ossner
Gruppenbetreuer: Dr. André Klima
Projektgruppe: Alexander Fogus, Lea Vanheyden, Zorana Spasojević

9. Mai 2020

Ludwig Maximilians Universität

Inhaltsverzeichnis

1.Hintergrund

- Konfliktsituation zwischen Mensch und Natur bzw. Tierreich im Alpengebiet
  ↳ Vorallem Wildtiere stark betroffen
- Kooperation des Departments für Geographie an der LMU, Lawinencamp Bayern, Gebietsbetreuer Mangfallgebirge, Alpenregion Tegernsee/Schliersee und dem Deutschen Alpenverein München (DAV)
- Aktionstag
- Spezielle Untersuchungen am Spitzingsee
  ↳ Territorium von Wildtieren
  ↳ Beliebte Gegend für Sportler
- Untersuchung über die Mitnahme von LVS-Geräten anhand von Checkpoints und manueller Datenerhebung
  ↳ Ziel: Das Verhalten der Besucher analysieren

1.Hintergrund

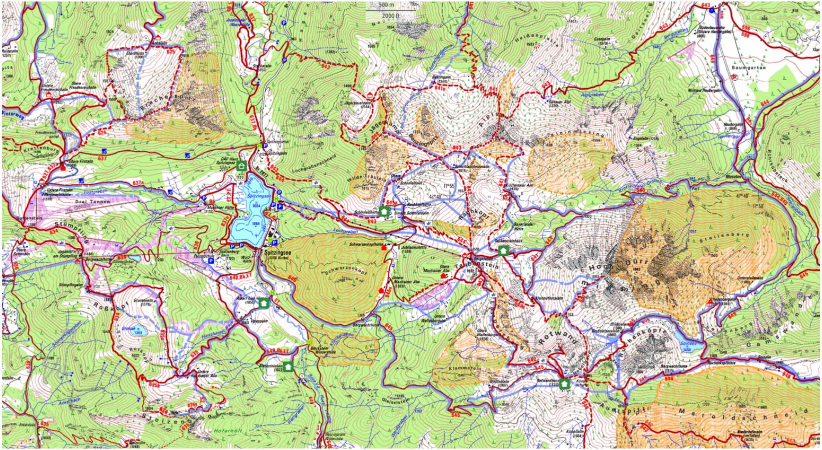

Abbildung 1: Untersuchungsraum: Spitzingsee

2.Datengrundlage

- Untersuchungsgegenstand: Wintersportler (vorrangig Skitourengänger & Schneeschuhgeher)
- Untersuchungszeitraum der Checkpoints Wintersaison 18/19
  ↳ Genauer Zeitraum: 25.12.2018 - 13.04.2019
- Checkpoints an zwei Routen (Nord- und Südseite) erfassen:
  ↳ Messungen insgesamt: 37216
  ↳ Personen mit LVS-Gerät: 8305
  ↳ Personen ohne LVS-Gerät: 28911
- Untersuchungszeitraum der manuellen Erhebung am 27.02 und 28.02.2019

2.Datengrundlage

Zielvariable:

- Anteil der Personen mit und ohne LVS-Gerät

Kovariablen:

- Datum
- Tag der Woche
- Feiertag
- Schneehöhe (in cm)
- Temperatur (in $^\circ C$)
- Sonneneinstrahlung (in W/m^2)
- Lawinenwarnstufe
- Uhrzeit der Messung

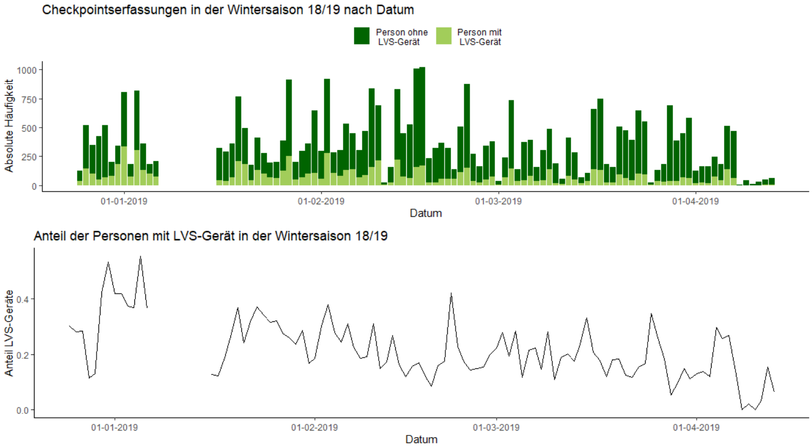

Abbildung 2: Absolute Häufigkeit und Anteil der Personen mit und ohne LVS-Gerät nach Datum

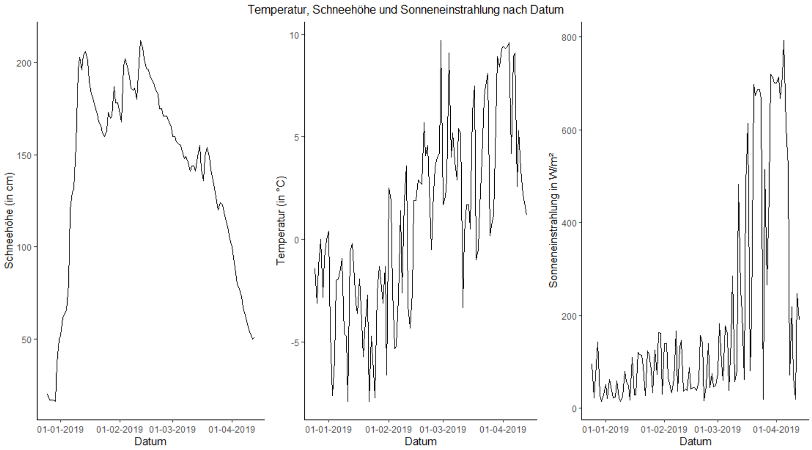

Abbildung 3: Temperatur, Schneehöhe und Sonneneinstrahlung nach Datum

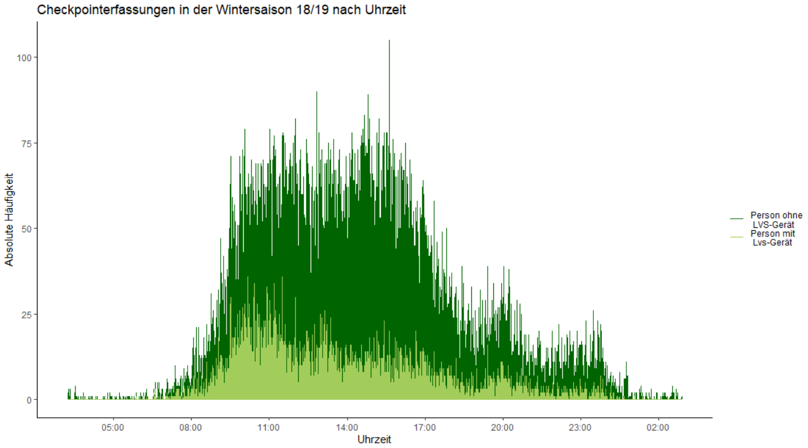

Abbildung 4: Absolute Häufigkeit der Personen mit und ohne LVS-Gerät nach Uhrzeit

Umcodierung Variablen vllt.

Modell

$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \epsilon_i$

- Erwartungswert einer Zufallsvariable wird durch Linearkombination von Kovariablen beschrieben
- Problem: Oft unzureichend, da auch nicht lineare Einflüsse auf Zielvariable wirken können

Additives Modell

Standardmodell der additiven Regression:

$$y_i = \underbrace{f_1(z_{i1}) + \ldots + f_q(z_{iq})}_{\text{nicht-parametrische Effekte}} + \underbrace{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}}_{\text{parametrische Effekte}} + \epsilon_i$$

- Stellt ein nicht-parametrisches Regressionsmodell dar
- Vorteil: Neben linearen Effekten können auch nicht-lineare Einflüsse von Kovariablen auf Zielvariable modelliert werden

Penalisierte Splines basierend auf B-Splines

- Funktion f durch Spline modellieren
- Konstruktion des Penalisierten Splines durch B-Spline Basisfunktion und Strafterm

Penalisierte Splines basierend auf B-Splines

B-Spline Basisfunktion

(a) Basisfunktion aus Polynomstücke des gewünschten Grades
    ↳ werden an den Knoten stetig zusammengesetzt

(b) Basisfunktion skalieren

(c) Skalierte Basisfunktionen addieren

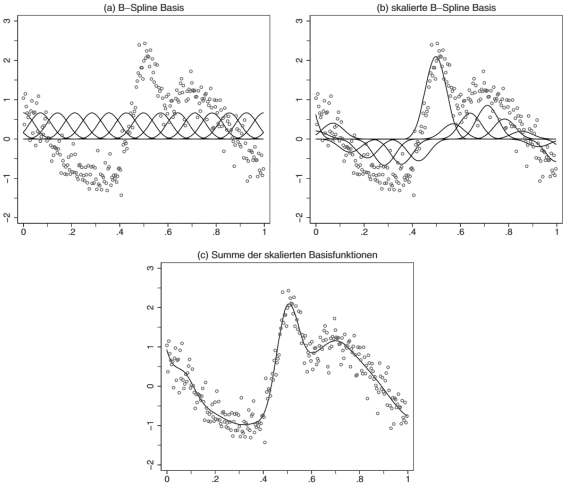

Abbildung 5: Visualisierung von B-Spline Basisfunktion anhand von fiktiven Daten

Penalisierte Splines basierend auf B-Splines

Strafterm

- Ziel: *Rauen* Schätzfunktionen entgegenwirken, d.h. zu starke Anpassung der Daten verhindern
- Bestrafung durch Glättungsparameter
- Für Glättungsparameter gegen ∞, annähernd lineare Schätzfunktion

Zyklische P-Splines

- Funktion der Variable Wochentag mit Hilfe von zyklischem P-Spline modelliert
- Werte am Ende einer Woche zusammenhängend zu den Werten am Anfang der Woche
- Beispiel: Sonntag und Montag

Generalisiertes additives Modell

- Keine normalverteilte Zielvariable
- Zielvariable nimmt typischerweise Verteilung aus Exponentialfamilie an
  ↳ Entweder normal-, binomial-, poisson- oder gammaverteilt

Generalisiertes additives Modell

Zielgröße

Beobachtete Zielgröße ist Binomial-verteilt $y_i|x_i \sim B(1, \pi_i)$:

$$y_i = \begin{cases} 1 & \text{,wenn Person mit LVS-Gerät identifiziert wird} \\ 0 & \text{,wenn Person ohne LVS-Gerät identifiziert wird.} \end{cases}$$

Erwartungswert

Für den Erwartungswert der Zielvariable gilt:

$$E(y_i) = \pi_i = \frac{exp(\eta_i)}{1+exp(\eta_i)} = h(\eta_i)$$

Logit-Link

$$\log(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$$

Generalisiertes additives Modell

Additiver Prädiktor:

Der Erwartungswert μ der Zielvariable y wird mit dem additiven Prädiktor:

$$\eta_i = f_{1,2}(Uhrzeit_i, Datum_i)^* + f_3(Lawinenwarnstufe_i) + f_4(Wochentag_i) + f_5(Temperatur\_Residuen_i)^{**} + f_6(Sonneneinstrahlung\_Resiuden_i)^{**} + f_7(Schneehoehe\_Residuen_i)^{**} + \beta_0 + \beta_1(Ferientag_i)$$

*Bemerkung**: Modell kann durch Interaktionseffekt ergänzt werden
↳ Nicht-lineare Interaktion zwischen Datum und Uhrzeit

*Bemerkung***: Smooth-Funktion für Residuen der Kovariablen Temperatur, Sonneneinstrahlung und Schnehöhe, da ansonsten
↳ *Concurvity*, stellt nicht-lineare Form der Kollinearität dar

Generalisiertes additives Modell

Bzgl. Concurvity umcodierte Variablen

Gruppierte Daten

- Bisher: Betrachtung von Individualdaten
- Weitere Möglichkeit: Gruppierte Daten
  ↳ Nach identische Zeilen der Kovariablen-Datenmatrix gruppieren (z.B. Datum)
- Problem: Überdispersion, d.h. Unterschätzung der Varianz
  ↳ Lösung: Dispersionsparameter (ϕ)

Einführung eines Dispersionsparameters in Varianzformel

$$Var(y_i|\mathbf{x}_i) = \underbrace{\phi}_{\text{Dispersionsparameter}} \quad \underbrace{\pi_i(1-\pi_i)}_{\text{Varianz}}$$

Generalisierte additive (gemischte) Modell

- Aufbau wie generalisiertes additives Modell
- Miteinbeziehung von zufälligen Effekten
- Benötigt um Problem der Autokorrelation zu lösen
  ↳ Korrelation einer Funktion mit sich selbst zu einem früheren Zeitpunkt

Ergebnisse

Zwei Modelle:

- Datumsmodell
  - ↳ Gruppierte Daten pro Tag
  - ↳ Uhrzeit der Messung nicht beachtet
- Zeitmodell
  - ↳ Gruppierte Daten pro Minute
  - ↳ Uhrzeit der Messung beachtet
  - ↳ Aber: Messungen nur einmal am Tag $\rightarrow$ niedrige Devianz

Ergebnisse

Datumsmodell

- plot Datumsmodell
- klar machen welche Signifikant sind
- außerdem intercept und deviance explained

Zeitmodell

- plot Zeitmodell
- klar machen welche Signifikant sind
- außerdem intercept und deviance explained

Messfehleranalyse

- Wissen aus Untersuchungen: Checkpointdaten haben Messfehler
- Neuen Daten zur Saison 19/20
  ↳ Checkpointdaten: vollständig
  ↳ Manuell erhobene Daten: nur 2 Tage
- Überlegung: Mögliche Szenarien für Messfehler erstellen und mögliche Änderungen im Modell prüfen

Messfehleranalyse

Plots zur deskriptiven Analyse

Messfehlerszenarien

4 verschiedene Szenarien:

- Generelle Unterschätzung von 20%
- Unterschätzung nach Gruppengröße
- Nächtliche Überschätzung
- Unterschätzung bei niedrigen Temperaturen

Messfehlerszenarien

Smooth-Vergleich zum ersten Szenario, stufenweise

Signifikanzvergleich zum ersten Szenario, stufenweise

B-Spline Basisfunktion

$$f(z) = \sum_{j=1}^{d} \gamma_j B_j(z), \; wobei \; d = m + l - 1$$

$$B_j^l(z) = \frac{z - k_j}{k_{j+l} - k_j} B_j^{l-1}(z) + \frac{k_{j+l+1} - z}{k_{j+l+1} - k_j + 1} B_{j+1}^{l-1}(z).$$

- Basisfunktion aus $(l + 1)$ Polynomstücke des gewünschten Grades (l)
  ↳ werden an den Knoten $(l - 1 - mal)$ stetig zusammengesetzt
- Basisfunktion $(B_j(z))$ mit dem Kleinste-Quadrate-Schätzer $\hat{\gamma}_j$ skalieren
- Skalierte Basisfunktionen addieren

Strafterm

$$\lambda \int (f''(z))^2 dz$$

- Ziel: *Rauen* Schätzfunktionen entgegenwirken, d.h. zu starke Anpassung der Daten verhindern
- Bestrafung durch Glättungsparameter λ
- Für $\lambda \to \infty$, annähernd lineare Schätzfunktion