

Projekt LVS-IR-Taubenstein

Projektpartner: Sascha Filimon, Roman Ossner

Gruppenbetreuer: Dr. André Klima

Projektgruppe: Alexander Fogus, Lea Vanheyden, Zorana Spasojević

25. Mai 2020

Ludwig Maximilians Universität



1. Hintergrund
2. Datengrundlage
3. Modell
4. Messfehler
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



1. Hintergrund
2. Datengrundlage
3. Modell
4. Messfehler
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



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
 - ↳ „Natürlich auf Tour“



1.Hintergrund

- Untersuchungsgebiet: 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

1.Hintergrund



Abbildung 2: Checkpoints an der Nord- und Südseite



1. Hintergrund
2. Datengrundlage
3. Modell
4. Messfehler
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



2.Datengrundlage

- Untersuchungsgegenstand: Wintersportler (vorrangig Skitourenzügler & 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 ohne LVS-Gerät	28911
Personen mit LVS-Gerät	8305

- Untersuchungszeitraum der manuellen Erhebung am 27.02.- und 28.02.2020



Zielvariable:

- Anteil der Personen mit LVS-Gerät

Kovariablen:

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



Zielvariable:

- Anteil der Personen mit LVS-Gerät

Kovariablen:

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



2.Datengrundlage

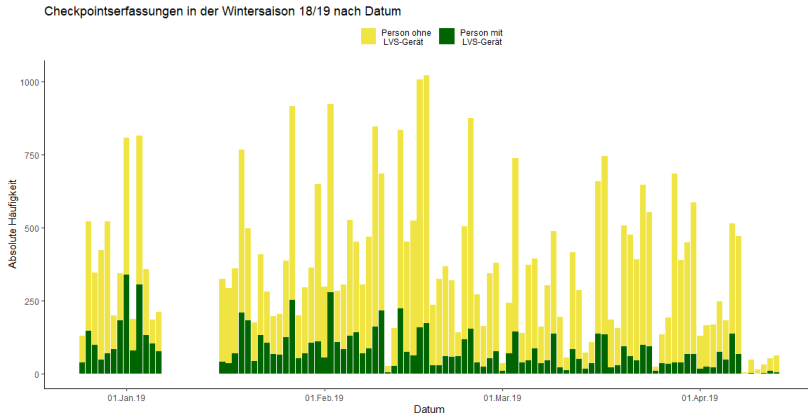


Abbildung 3: Absolute Häufigkeit der Personen mit und ohne LVS-Gerät nach Datum



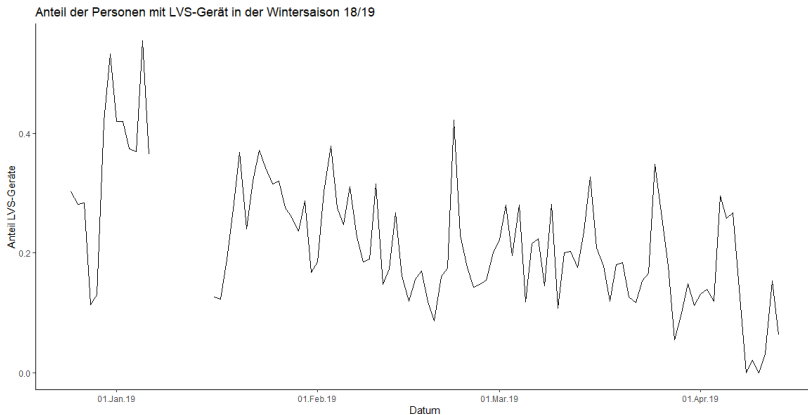


Abbildung 4: Anteil der Personen mit LVS-Gerät nach Datum



Zielvariable:

- Anteil der Personen mit LVS-Gerät

Kovariablen:

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



2.Datengrundlage

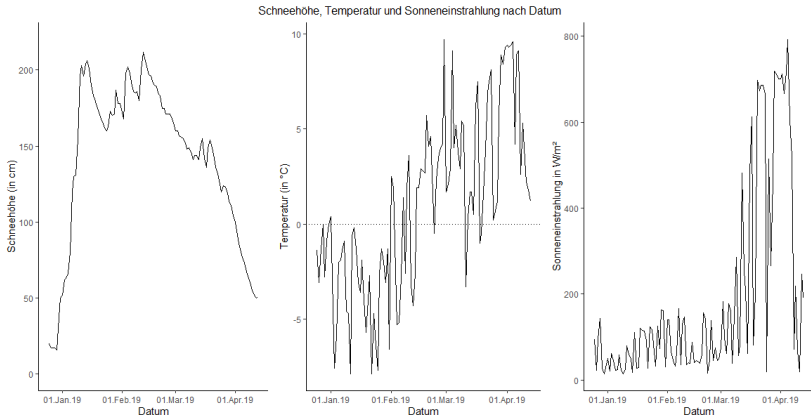


Abbildung 5: Schneehöhe, Temperatur und Sonneneinstrahlung nach Datum



Zielvariable:

- Anteil der Personen mit LVS-Gerät

Kovariablen:

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



2. Datengrundlage

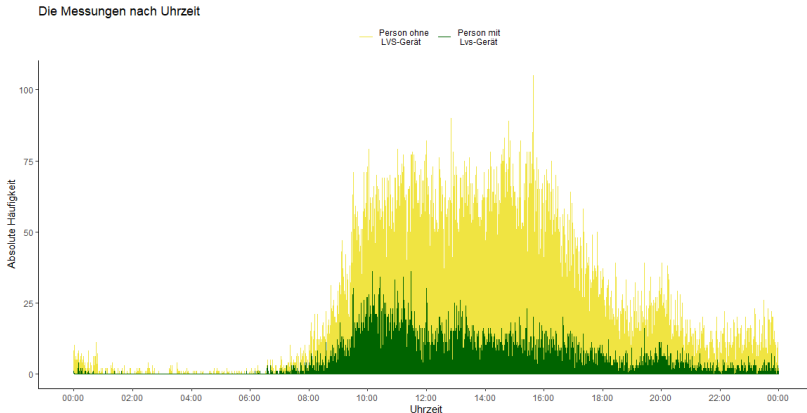


Abbildung 6: Absolute Häufigkeit der Personen mit und ohne LVS-Gerät nach Uhrzeit



1. Hintergrund
2. Datengrundlage
3. Modell
 - 3.1 Theorie
 - 3.2 Schwierigkeiten
 - 3.3 Ergebnisse
4. Messfehler
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



1. Hintergrund
2. Datengrundlage
3. Modell
 - 3.1 Theorie
 - 3.2 Schwierigkeiten
 - 3.3 Ergebnisse
4. Messfehler
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



3.1 Theorie/Lineares Regressionsmodell

Modell

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

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

(Quelle: Fahrmeir, Ludwig, et al. Regression, 2007: S.20)



3.1 Theorie/Generalisiertes lineares Regressionsmodell

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

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

(Quelle: Fahrmeir, Ludwig, et al. Regression, 2007: S.192)



3.1 Theorie/Additives Modell

Standardmodell der additiven Regression:

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

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

(Quelle: Fahrmeir, Ludwig, et al. Regression, 2007: S.401)



3.1 Theorie/Penalisierte Splines basierend auf B-Splines

- Funktion f durch Spline modellieren
- Verwendete Splines:
 - ↳ Penalisierte Splines
 - ↳ Zyklische P-Splines
 - ↳ Thin-Plate-Plines
- Konstruktion des Penalisierten Splines durch B-Spline Basisfunktion und Strafterm

(Quelle: Fahrmeir, Ludwig, et al. Regression, 2007: S.306 f.)



3.1 Theorie/B-Spline Basisfunktion

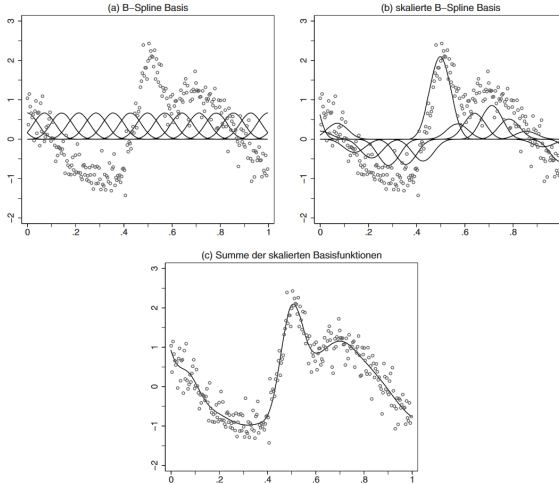


Abbildung 7: Visualisierung von B-Spline Basisfunktion anhand von fiktiven Daten (Quelle: Fahrmeir, Ludwig, et al. Regression, 2007: S.307)^{24/54}



3.1 Theorie/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

$$f(z) = \sum_{j=1}^d \gamma_j B_j(z)$$

(Quelle: Fahrmeir, Ludwig, et al. Regression, 2007: S.306 f.)



3.1 Theorie/Penalisierte Splines basierend auf B-Splines

Strafterm

- Ziel: *Rauen* Schätzfunktionen entgegenwirken, d.h. zu starke Anpassung der Daten verhindern
- Bestrafung durch Glättungsparameter
- Für Glättungsparameter = 0 keine Penalisierung
- Für Glättungsparameter gegen ∞ , annähernd lineare Schätzfunktion

(Quelle: Fahrmeir, Ludwig, et al. Regression, 2007: S.306 ff.)



3.1 Theorie/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

(Quelle: Wood, Simon N. Generalized additive models: an introduction with R., 2017: S. 202f. und S.206)



3.1 Theorie/Thin-Plate-Splines

- Eignet sich sehr gut zur Schätzung einer glatten Funktion mit mehreren Variablen
- Hier: Zwei Dimensionale Smooth Funktion
↳ Datum und Uhrzeit
- Knotenpositionen oder Basisfunktionen auswählen nicht notwendig

Thin-Plate-Spline

Glattheit, wenn x zweidimensional ist, durch Minimierung der Funktion von f :

$$||y - f||^2 + \lambda J_{md}(f)$$

(Quelle: Wood, Simon N. Generalized additive models: an introduction with R., 2017: S. 215ff.)



3.1 Theorie/Generalisiertes additives Modell

Additiver Prädiktor:

$$\eta_i = \beta_0 + \beta_1(\text{Ferientag}_i) + f_{1,2}(\text{Uhrzeit}_i, \text{Datum}_i)^* + f_3(\text{Lawinenwarnstufe}_i) + f_4(\text{Wochentag}_i) + f_5(\text{Temperatur}_i) + f_6(\text{Anteil Max Sonneneinstrahlung}_i)^{**} + f_7(\text{Neuschnee}_i)^{**}$$

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

Bemerkung^{**}: Transformation der Kovariablen Schneehöhe und Sonneneinstrahlung, da ansonsten Concurvity-Problem
↳ *Concurvity*, stellt nicht-lineare Form der Kollinearität dar

(Quelle: Fahrmeir, Ludwig, et al. Regression, 2007: S.401 ff.)



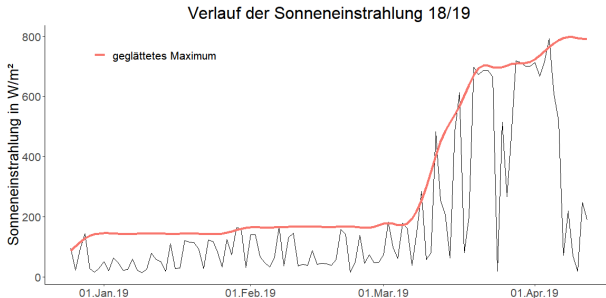
1. Hintergrund
2. Datengrundlage
3. Modell
 - 3.1 Theorie
 - 3.2 Schwierigkeiten
 - 3.3 Ergebnisse
4. Messfehler
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



3.2 Schwierigkeiten/Concurvity

Umcodierung der Variablen mit hohem Maß an Concurvity:

- Neuschnee in cm (statt Schneehöhe)
- Anteil an der maximalen Sonneneinstrahlung (statt Sonneneinstrahlung)



3.2 Schwierigkeiten/Überdispersion

- 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

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

Dispersionsparameter

$$\phi = \frac{\text{Devianz}}{\text{Freiheitsgrade der Residuen}}$$



3.2 Schwierigkeiten/Zeitreihe

- Problem: Autokorrelation der Uhrzeit
 - ↳ Korrelation einer Funktion mit sich selbst zu einem früheren Zeitpunkt

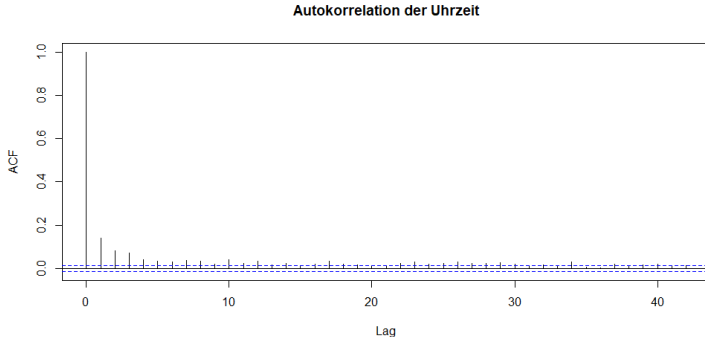


Abbildung 9: Plot zur Autokorrelation der Uhrzeit



3.2 Schwierigkeiten/Zeitreihe

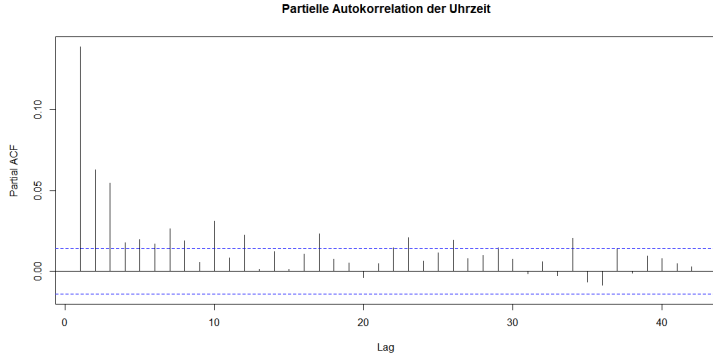


Abbildung 10: Plot zur Partiellen Autokorrelation der Uhrzeit



1. Hintergrund
2. Datengrundlage
3. Modell
 - 3.1 Theorie
 - 3.2 Schwierigkeiten
 - 3.3 Ergebnisse
4. Messfehler
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



3.3 Ergebnisse

Zwei Modelle:

- Datumsmodell
 - ↳ Gruppierte Daten pro Tag
 - ↳ Uhrzeit der Messung nicht beachtet
- Zeitmodell
 - ↳ Gruppierte Daten pro Minute
 - ↳ Uhrzeit der Messung beachtet
 - ↳ Aber: Viele Variablen nur einmal am Tag erhoben



3.3 Ergebnisse/Datumsmodell

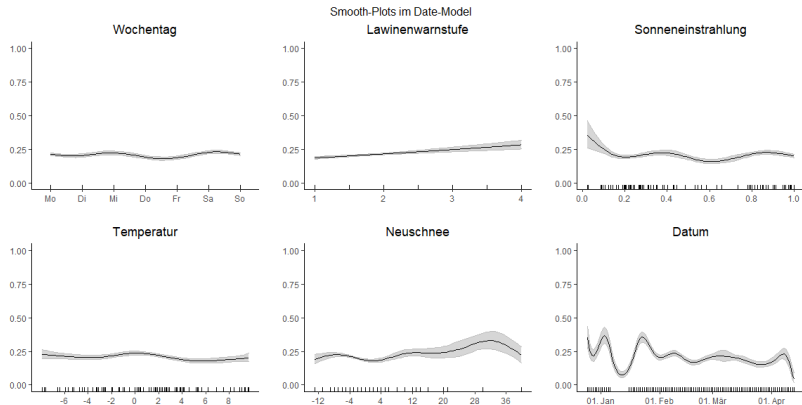


Abbildung 11: Smooth-Plots für das Datumsmodell

Kein Feiertag: 0.208; Feiertag: 0.225



3.3 Ergebnisse/Zeitmodell

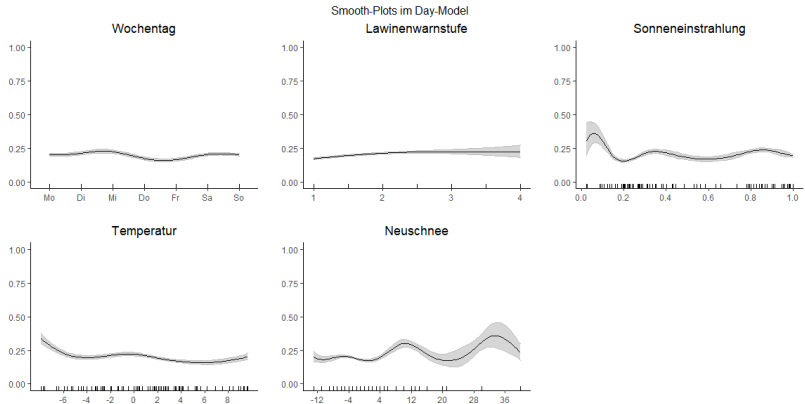


Abbildung 12: Smooth-Plots für das Zeitmodell

Kein Feiertag: 0.198; Feiertag: 0.232



3.3 Ergebnisse/Zeitmodell

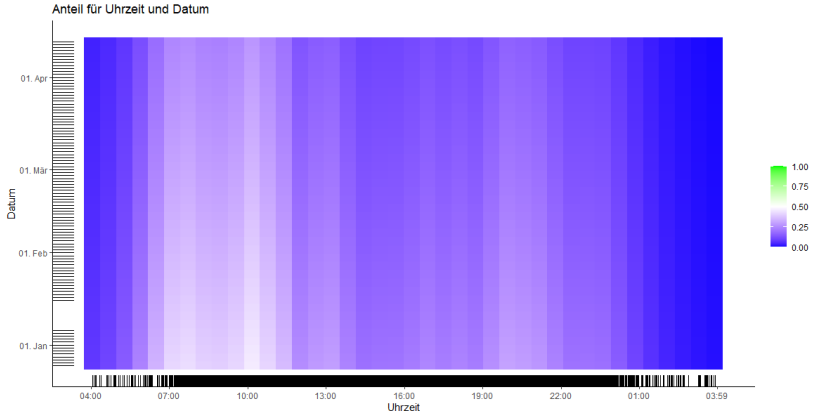


Abbildung 13: Smoothfunktion für die Uhrzeit und das Datum im Zeitmodell



1. Hintergrund
2. Datengrundlage
3. Modell
4. Messfehler
 - 4.1 Messfehleranalyse
 - 4.2 Messfehlerszenarien
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



1. Hintergrund
2. Datengrundlage
3. Modell
4. Messfehler
 - 4.1 Messfehleranalyse
 - 4.2 Messfehlerszenarien
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



4.1 Messfehleranalyse

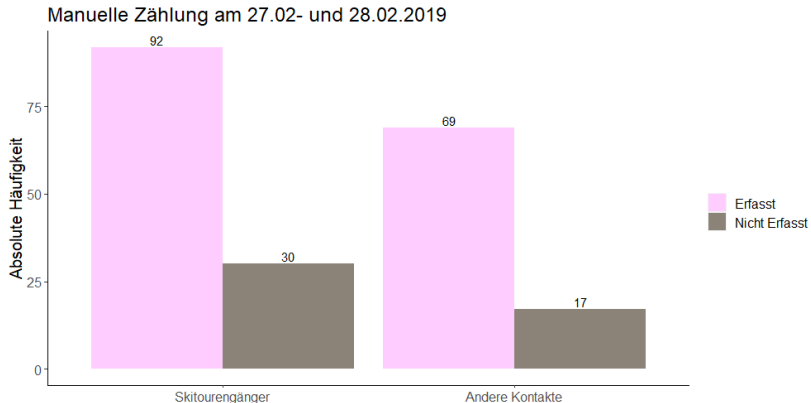


Abbildung 14: Manuelle (Studenten) Zählung am 27.02 und 28.02.2019



4.1 Messfehleranalyse

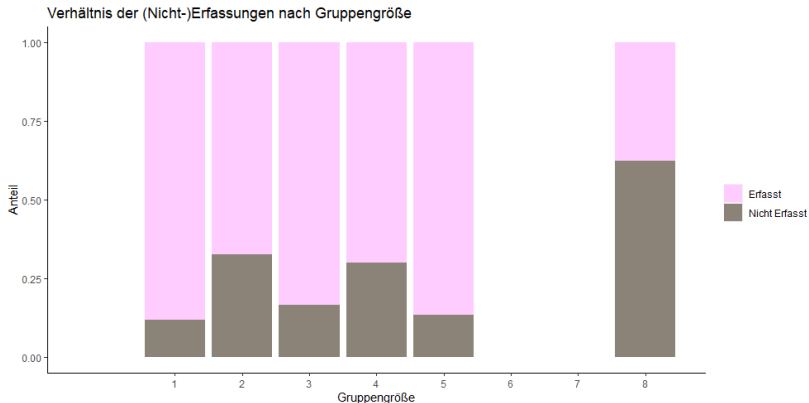


Abbildung 15: Verhältnis der (Nicht-)Erfassungen nach Gruppengröße



1. Hintergrund
2. Datengrundlage
3. Modell
4. Messfehler
 - 4.1 Messfehleranalyse
 - 4.2 Messfehlerszenarien
5. Fazit
6. Ausblick
7. Literaturverzeichnis
8. Anhang



4.2 Messfehlerszenarien

- Überlegung: Mögliche Szenarien für Messfehler erstellen und mögliche Änderungen im Modell prüfen
- 4 verschiedene Szenarien:
 - ↳ Generelle Unterschätzung von 20%
 - ↳ Unterschätzung nach Gruppengröße
 - ↳ Nächtliche Überschätzung
 - ↳ Unterschätzung bei niedrigen Temperaturen



4.2 Messfehlerszenarien

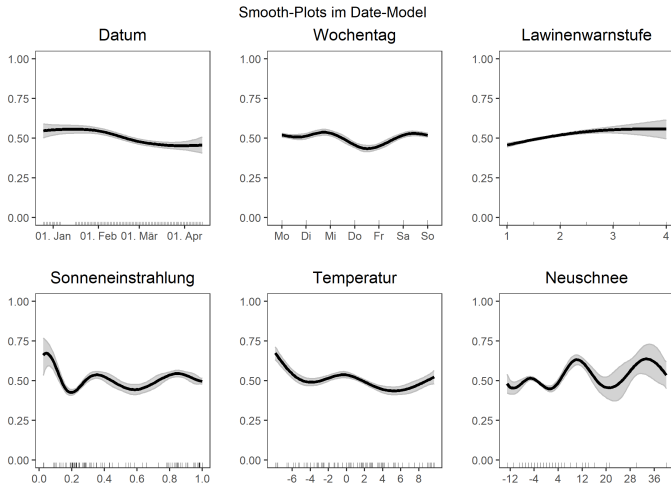


Abbildung 16: Smooth-Plots für das Datummodell



4.2 Messfehlerszenarien

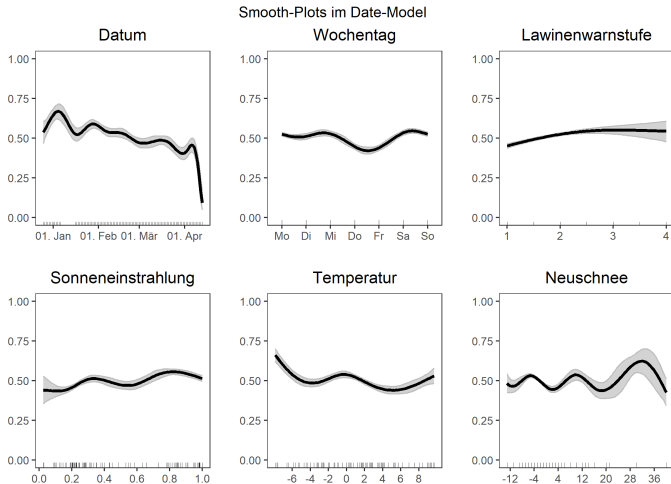


Abbildung 17: Smooth-Plots für das Datumsmodell (Szenario 1)



4.2 Messfehlerszenarien

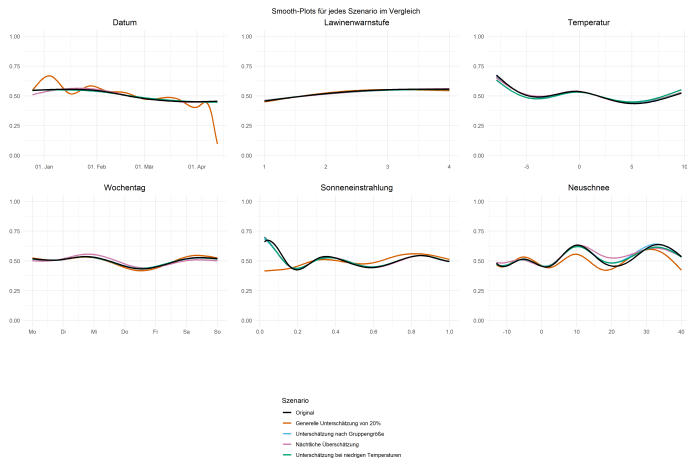


Abbildung 18: Smooth-Plots für das Datumsmodell (für jedes Szenario)



4.2 Messfehlerszenarien

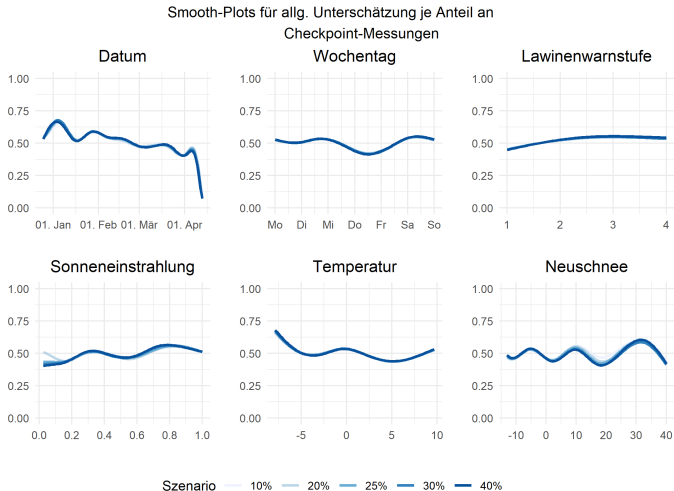


Abbildung 19: Smooth-Plots für das Datumsmodell (für jedes Szenario)



4.2 Messfehlerszenarien

Tabelle 1: Add caption

Anteil hinzugefügter Messungen	10%			20%			25%			30%			40%		
	Koeffizient	p-Wert		Koeffizient	p-Wert		Koeffizient	p-Wert		Koeffizient	p-Wert		Koeffizient	p-Wert	
(Intercept)	0.178	<2e-16	***	0.156	<2e-16	***	0.147	<2e-16	***	0.141	<2e-16	***	0.129	<2e-16	***
Ferientag	0.559	0.16		0.549	0.48		0.555	0.41		0.547	0.49		0.547	0.49	
s(Datum)		0.06	.		0.17			0.18			0.10	.		0.04	*
s(Lawinengefahr)		0.01	*		0.13			0.09	.		0.08	.		0.07	.
s(Wochentag)		0.04	*		0.02	*		0.02	*		0.01	*		0.01	**
s(Temperatur)		0.04	*		0.19			0.19			0.19			0.15	
s(Sonneneinstrahlung)		0.16			0.30			0.24			0.24			0.20	
s(Neuschnee)		0.18			0.51			0.49			0.37			0.43	



1. Hintergrund
2. Datengrundlage
3. Modell
 - 3.1 Concurvity
 - 3.2 Überdispersion
 - 3.3 Autokorrelation
4. Ergebnisse
 - 4.1 Messfehleranalyse
 - 4.2 Messfehlerszenarien
- 5. Fazit**
6. Ausblick
7. Literaturverzeichnis
8. Anhang



7.Literaturverzeichnis

- Fahrmeir, L., T. Kneib, and S. Lang. "Regression. Modelle, Methoden und Anwendungen. 2007."
- Wood, Simon N. Generalized additive models: an introduction with R. CRC press, 2017.



8.Anhang/Penalisierte Splines basierend auf B-Splines

B-Spline Basisfunktion

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

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

- 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 $\lambda \rightarrow \infty$, annähernd lineare Schätzfunktion

