ulm university universität ulm

Geostatistische Modellierung von räumlich aufgelösten Schadenhöhen

Prof. Dr. Evgeny Spodarev, Institut für Stochastik

ESRI, München, 24. Mai 2011

Naturkatastrophen

Hundertjahreshochwasser, 2002

Wintersturm “Kyrill”, 2007

Überblick

- 1D-Analyse von Risiken
- Räumliche Datenanalyse
- Räumliche Darstellung der Datensätze
- Modellierung: Geoadditive und räumliche Regression
- Räumliche Modellierung von Extremschäden
- Räumlich aufgelöste Prämien

Kriterien für die Auswahl des Datensatzes

1. Anzahl der Realisierungen pro Jahr
2. Homogenität
3. Verfügbarkeit der Daten über den gesamten Zeitraum
   - Analyse einer Zeitreihe
   - zeitliche Trends im Datensatz
4. Gruppierung der Daten:
   PLZ-Gebiete $\leftrightarrow$ einzelne Risikoadressen

Anzahl der Realisierungen - Unterscheidung nach Schadenursachen

Anzahl der Schäden zwischen 1996 und 2006, Sparte Sturm

	1996	1997	1999	2000	2001
Hochwasser	9	10	421	373	80
Sturm	2.568	4.204	4.535	7.233	3.108

	2002	2003	2004	2005	2006
Hochwasser	6.612	409	409	2.351	1.601
Sturm	19.217	12.418	14.795	8.877	5.013

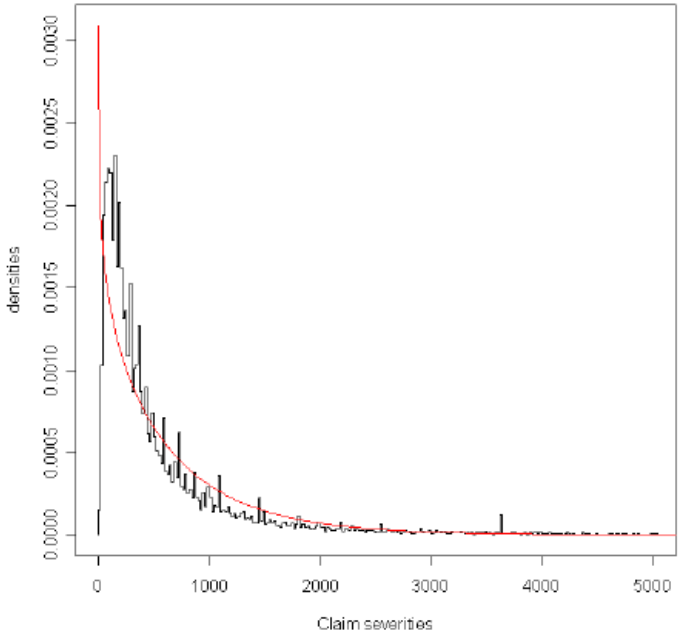
$\Rightarrow$ Einschränkung der Analyse auf Sturmschäden

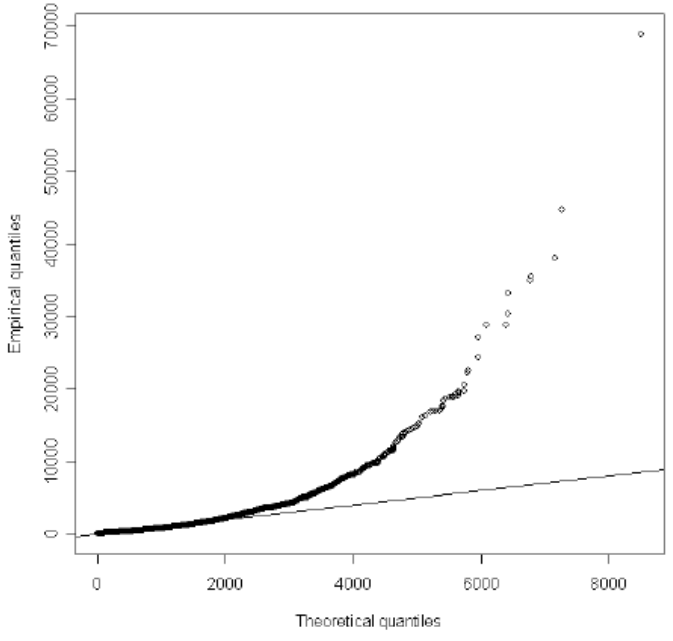
Ausgewählter Datensatz

- Sparte Sturm
- Tarifhauptgruppe Eigenheim
- Schadenursache Sturm
- Gruppierung nach PLZ-Gebieten
- Zeitraum: 10 Jahre

Durchschnittliche Schadenzahlungen

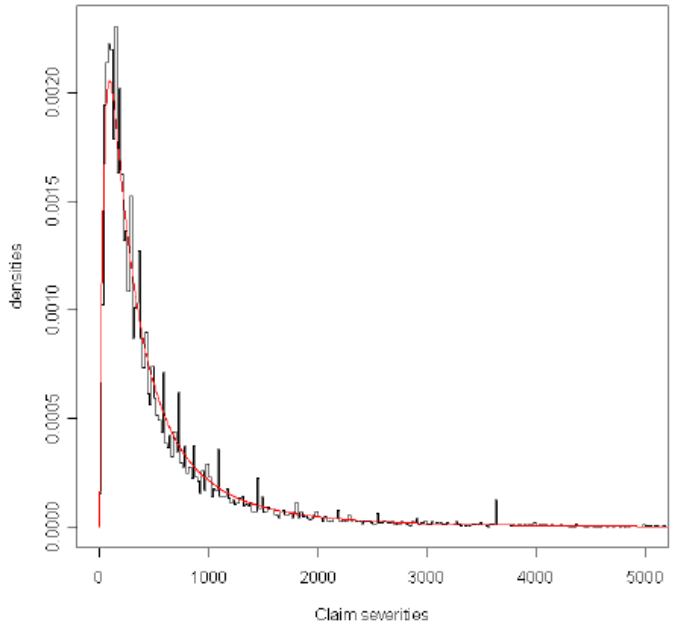
Anpassen einer Gammaverteilung an die Daten

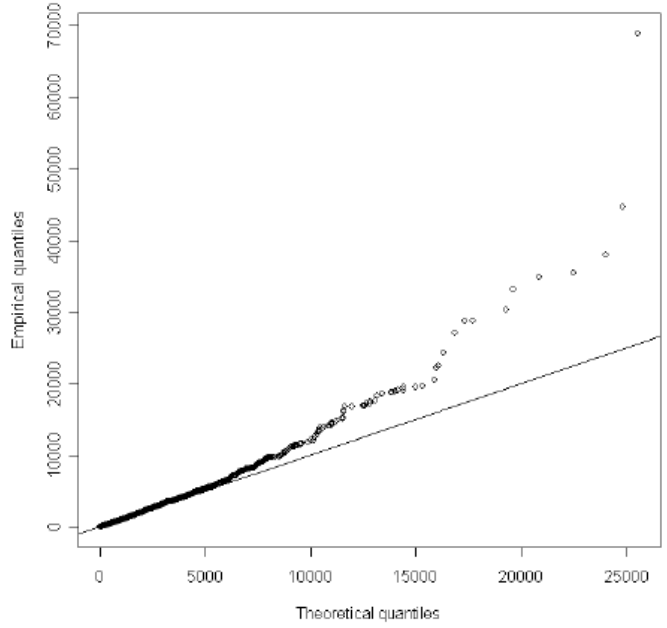




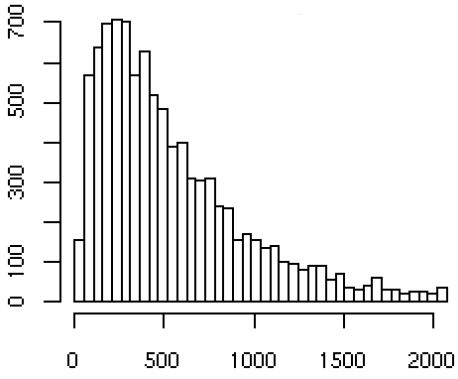
Durchschnittlichen Schadenzahlungen

Anpassen einer Lognormal-Verteilung an die Daten

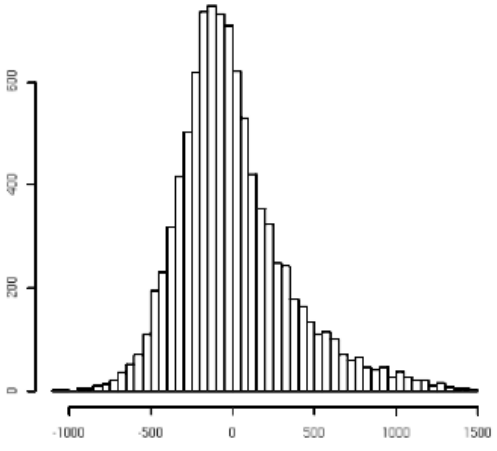




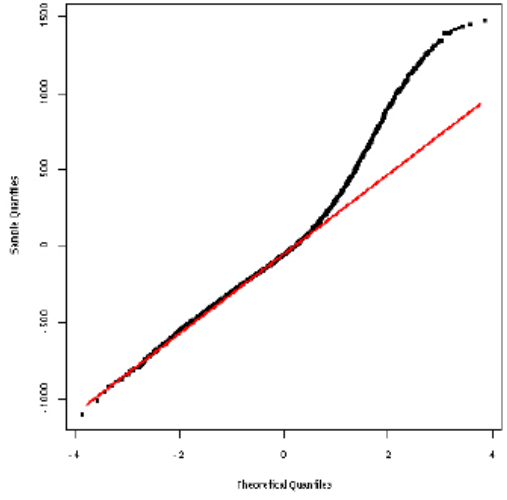
Abweichungen der durchschnittlichen Schadenzahlungen vom Mittelwert



Histogramm der durchschnittlichen Schadenzahlungen



Histogramm der Abweichungen



Q-Q Plot der Abweichungen

- Durchschnittliche Schadenzahlungen: Logstabile Verteilung
- Abweichungen: schiefe Verteilung mit schweren Tails

Anpassung einer α-stabilen Verteilung an die Daten

Parameterschätzung im eindimensionalen Fall:

α	β	σ	μ
1.3562	0.2796	234.286	6.7787

Methoden:

- Maximum-Likelihood-Methode (J. Nolan (2001))
- Sample characteristic-Methode (S. Kogon, D. Williams (1998))
- Quantilmethode (J.H. McCulloch (1986))

Räumliche Verteilung der Postleitzahlgebiete

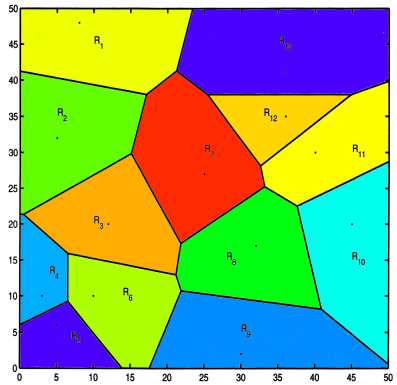
Zentren der 2053 Postleitzahlgebiete

Kartographische Darstellung der Risiken

Extrapolationsmethoden:

- Zufällig eingefärbte Mosaike
- Kriging
- Radiale Methoden
- Geoadditive Regressionsmodelle
- Partition Models
- Splines
- ...

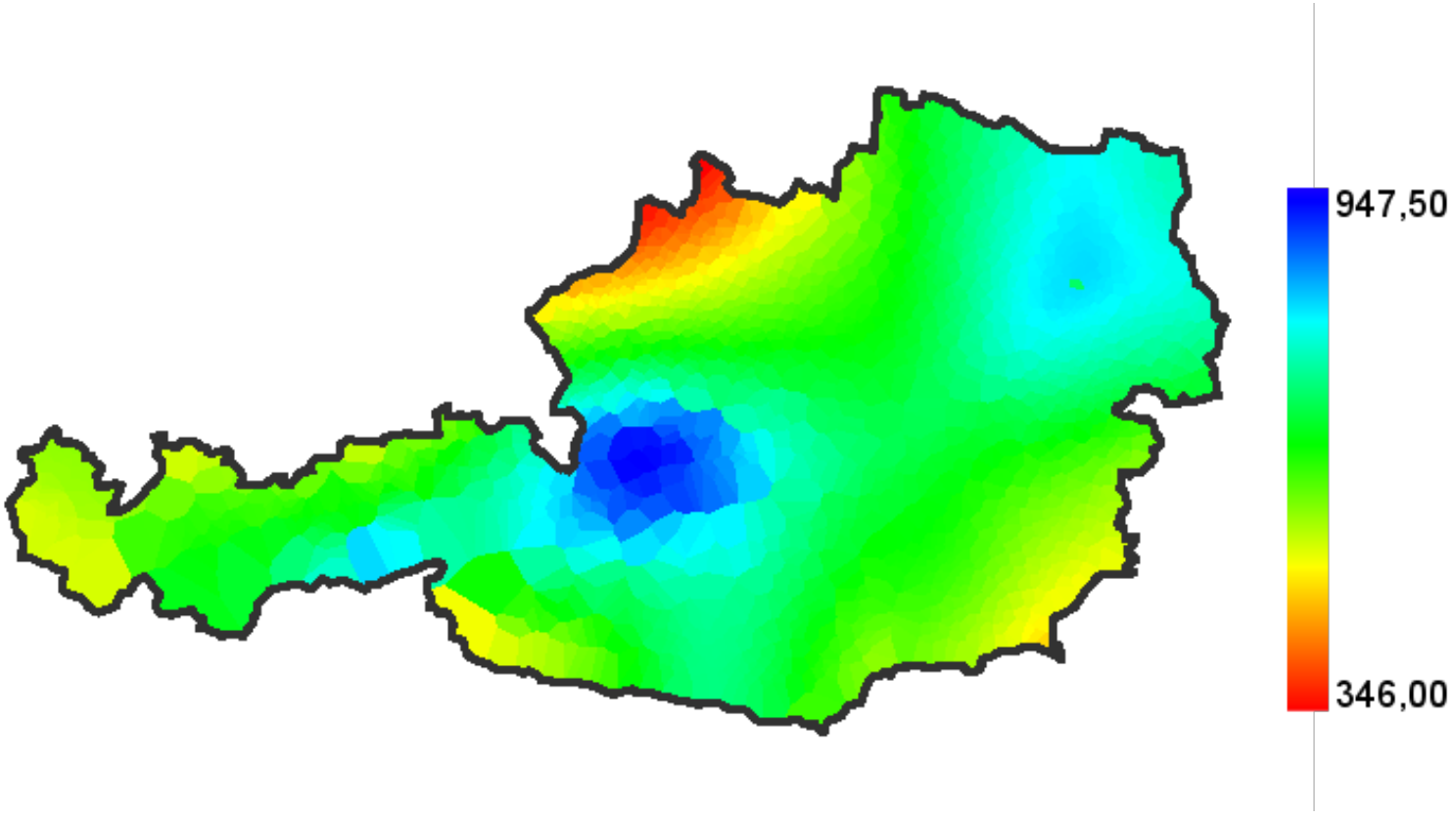
Zufällig eingefärbte Mosaike



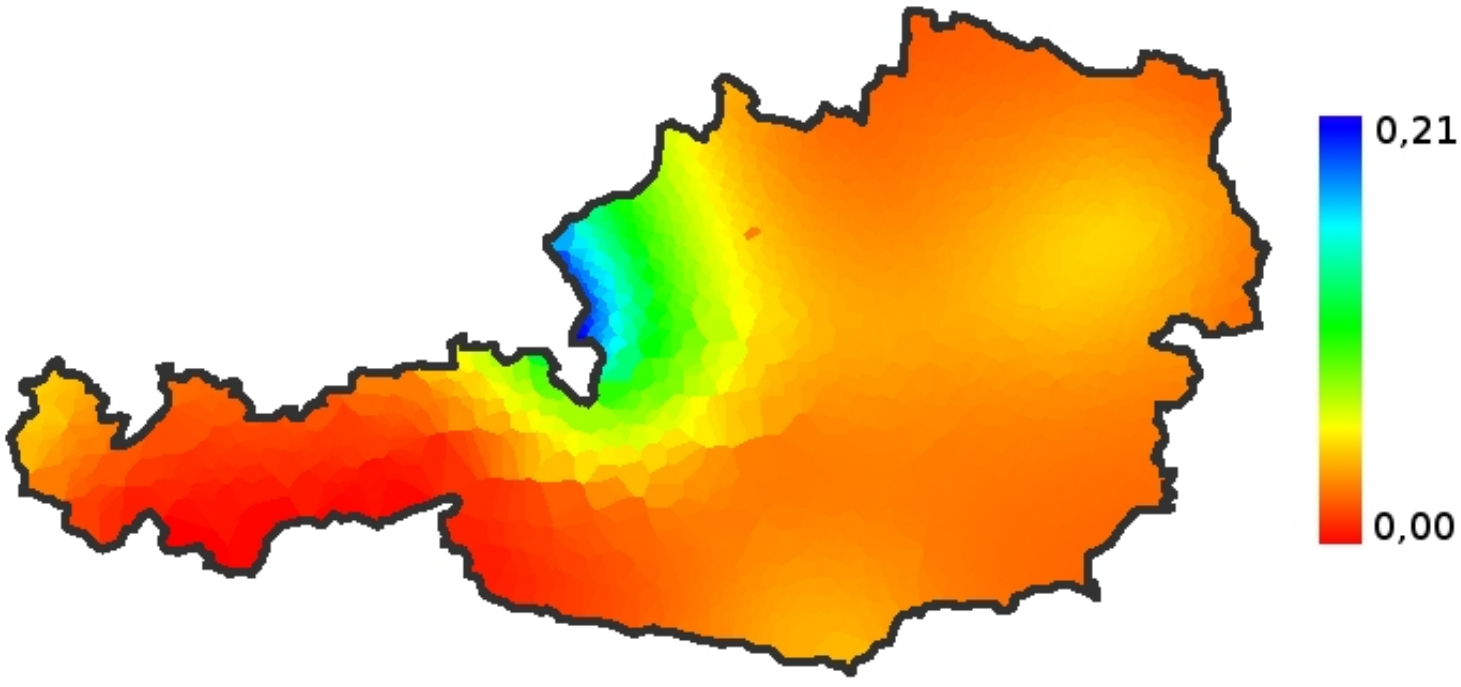
Voronoi-Tessellation

- Zerlegung des Beobachtungsgebietes in Regionen, die durch eine vorgegebene Menge an Punkten (Zentren) bestimmt werden.
- Jede Region wird durch genau ein Zentrum bestimmt und umfasst alle Punkte des Beobachtungsgebietes, die näher an dem Zentrum der Region liegen als an jedem anderen Zentrum.

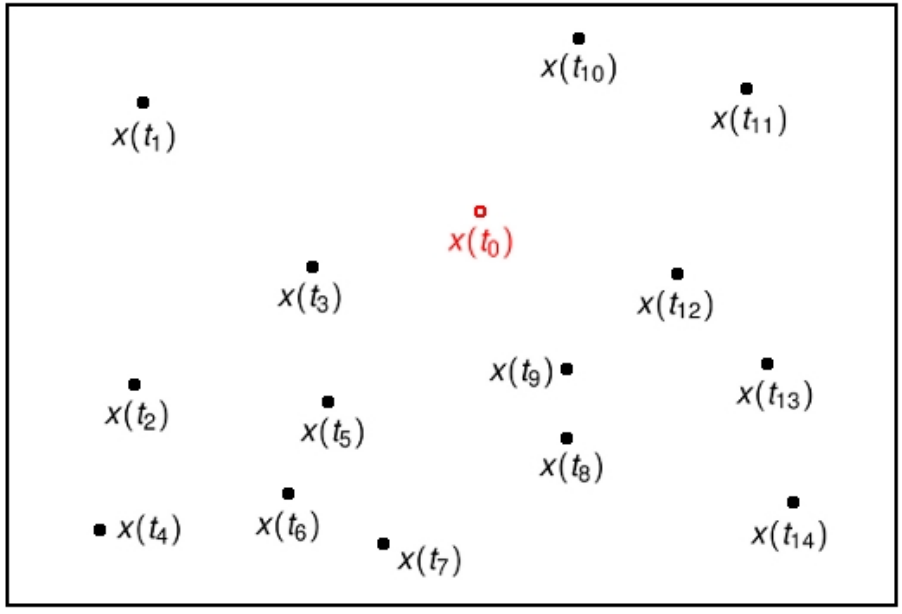
Durchschnittliche Schadenzahlungen der Sturmschäden



Schadenfrequenzen der Sturmschäden



Kriging

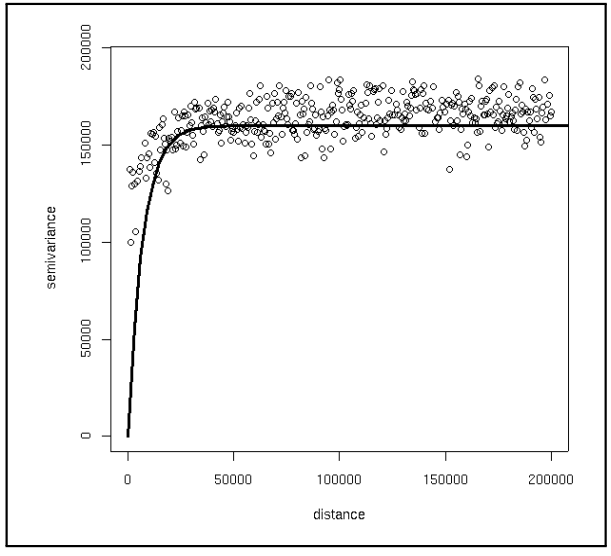


Bestimmung von $x(t_0)$ als gewichtete Summe der Werte $x(t_1), ..., x(t_{14})$:

$$x(t_0) = \sum_{i=1}^{14} w_i x(t_i), \quad \sum_{i=1}^{14} w_i = 1$$

Kriging

Beispiel-Variogramm der durchschnittlichen Schadenzahlungen



Schäden innerhalb von 50km weisen ähnlichere durchschnittliche Schadenzahlungen auf.

Klassisches lineares Regressionsmodell

Zielgröße: y
Kovariablen: $x_1, ..., x_k$
Störgröße: ε

Multivariates lineares Regressionsmodell:

$$y = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k}_{\text{linearer Prädiktor}} + \varepsilon \quad \text{mit}$$

$$\mathbb{E}(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

Beispiel:

$$\textit{Schadenhöhe} = \beta_0 + \beta_1 \textit{Vers.summe} + \beta_2 \textit{Bausubstanz} + \varepsilon$$

Geoadditives Regressionsmodell

Erweiterung des multivariaten linearen Regressionsmodells:

- Verknüpfung von $\mathbb{E}(y)$ mit dem linearen Prädiktor durch eine Linkfunktion g.
- Verallgemeinerung des linearen Prädiktors zu einem geoadditiven Prädiktor.

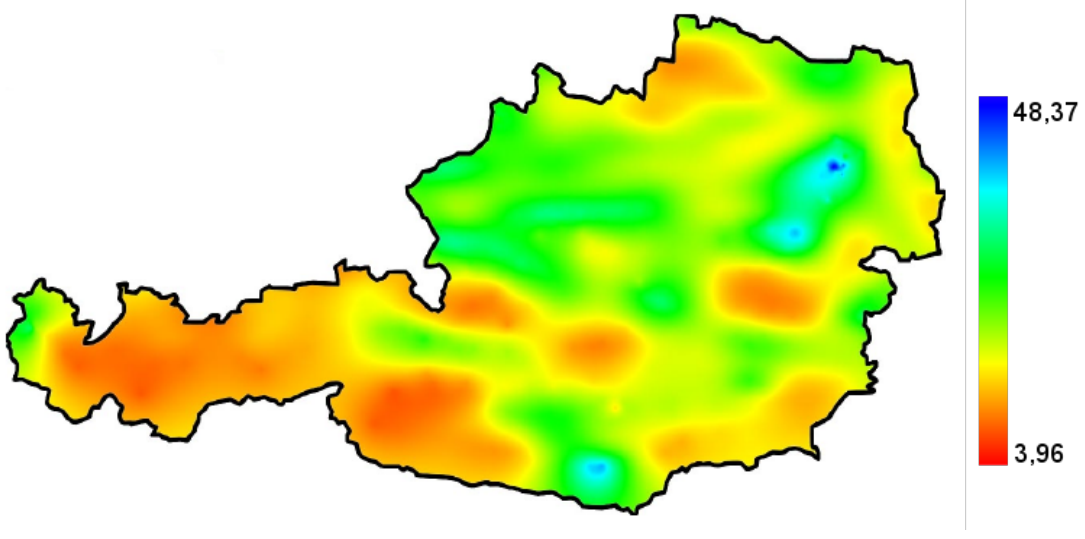
$$\begin{aligned} g(\mathbb{E}(y)) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \\ &\quad + f_1(z_1) + \ldots + f_p(z_p) + f_{geo}(s) \end{aligned}$$

Beispiel:

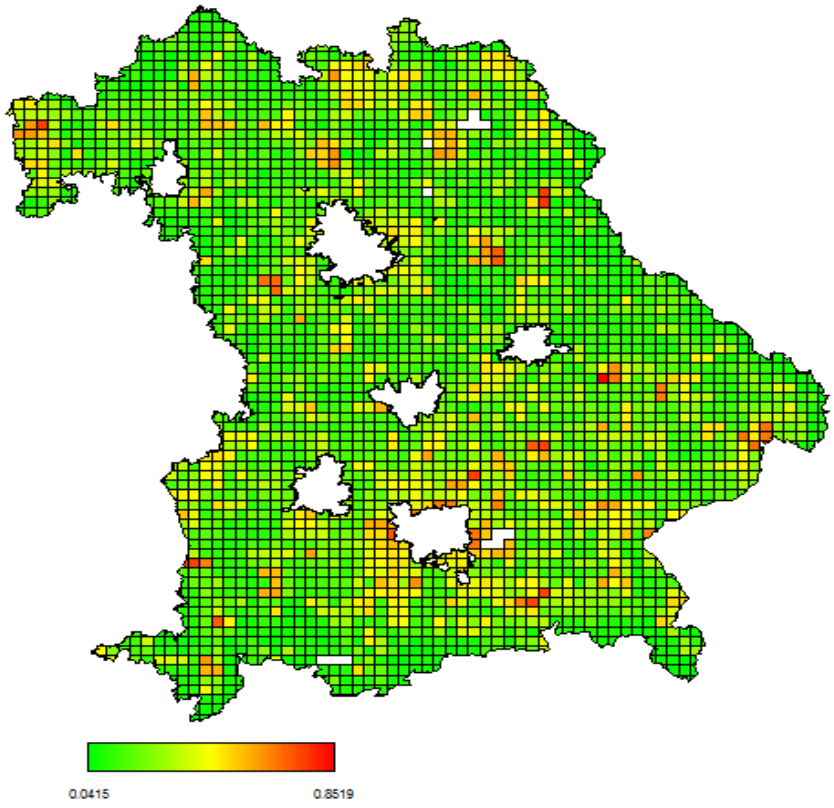
$$\begin{aligned} \log(\mathbb{E}(\mathit{Schadenfreq.})) &= \beta_0 + \beta_1 \mathit{Vers.summe} + \beta_2 \mathit{Bausubstanz} \\ &\quad + \mathit{Kriging}(\mathit{Risikoadresse}) \end{aligned}$$

Anwendung auf die Sturmschäden

- Verteilungsannahmen:
  - Schadenfrequenz: Poissonverteilung
  - Schadenhöhe: Lognormal-Verteilung
- Verwendung der natürlichen Linkfunktion für g



Vorhersage der Kriminalitätssituation in Bayern



Wahrscheinlichkeiten des Einbruchs in April, ermittelt über Jahre 2004-2010

Räumliche Regression

Zielgröße: y
Kovariablen: $x_1, ..., x_k$
Störgröße: $\varepsilon(t_1, t_2)$ an der Risikoadresse mit Koordinaten (t_1, t_2)

$$y = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k}_{\text{linearer Prädiktor}} + \varepsilon(t_1, t_2),$$

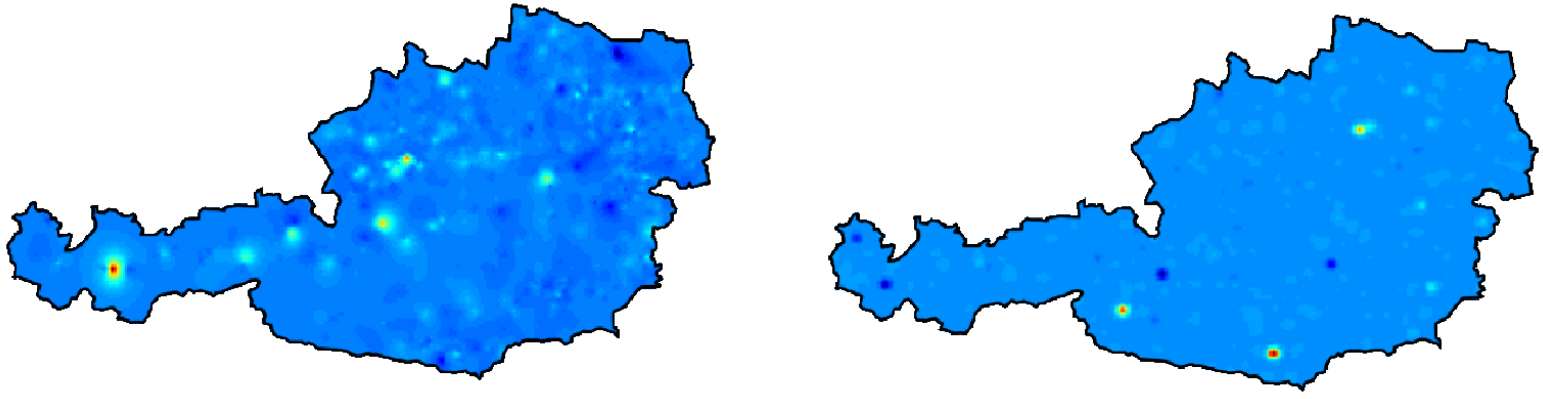
wobei $\{\varepsilon(t_1, t_2)\}$ ein Zufallsfeld ist.

Anwendung des Modells auf die Sturmdaten

- Verwendung des räumlichen Regressionsmodells für die Schadenhöhen
- Verwendung eines α-stabilen Zufallsfeldes für die Residuen
- An die Schadenfrequenzen wird pro PLZ-Gebiet eine Poissonverteilung angepasst.

Zufallsfeld der Störgrößen

Simulationsergebnisse:



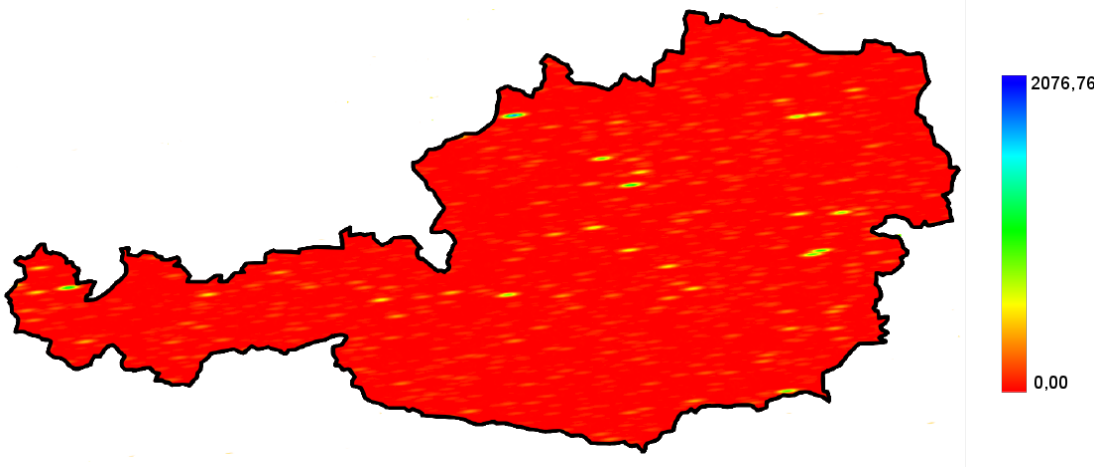
Gekrigte Abweichungen von Mittelwertschäden (Jahr 2000) (links) und eine Realisierung des angepassten α-stabilen Zufallsfeldes (rechts).

Extremwerttheorie

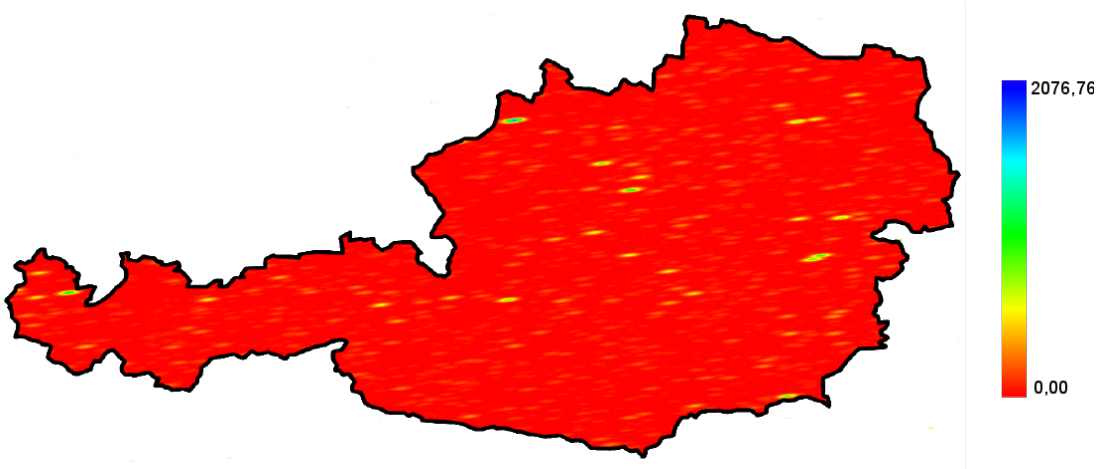
- Zentrales Resultat: Für das Maximum einer Stichprobe sind im Wesentlichen nur drei Grenzverteilungen möglich (Gumbel, Fréchet, Weibull)
  $\longrightarrow$ Extremwertverteilungen bzw. max-stabile Verteilungen
- Gemeinsame Betrachtung der Maxima von Schadenzahlungen, Schadenfrequenzen o.ä. an den PLZ-Zentren
  $\longrightarrow$ Multivariate Extremwertverteilungen
- Gemeinsame Betrachtung der Maxima von Schadenzahlungen, Schadenfrequenzen o.ä. in ganz Österreich
  $\longrightarrow$ max-stabile Zufallsfelder

Max-stabile Zufallsfelder

- Anpassung eines max-stabilen Zufallsfeldes an die Maxima der Schadenzahlungen pro PLZ-Gebiet.
- Eine Realisierung des max-stabilen Zufallsfeldes:



Max-stabile Zufallsfelder



- Ellipsenförmige Konturlinien → Stürme ziehen von Westen nach Osten
- Abhängigkeiten zwischen den Maxima:

  erste Hauptachse: 6,2 km
  zweite Hauptachse: 1,5 km

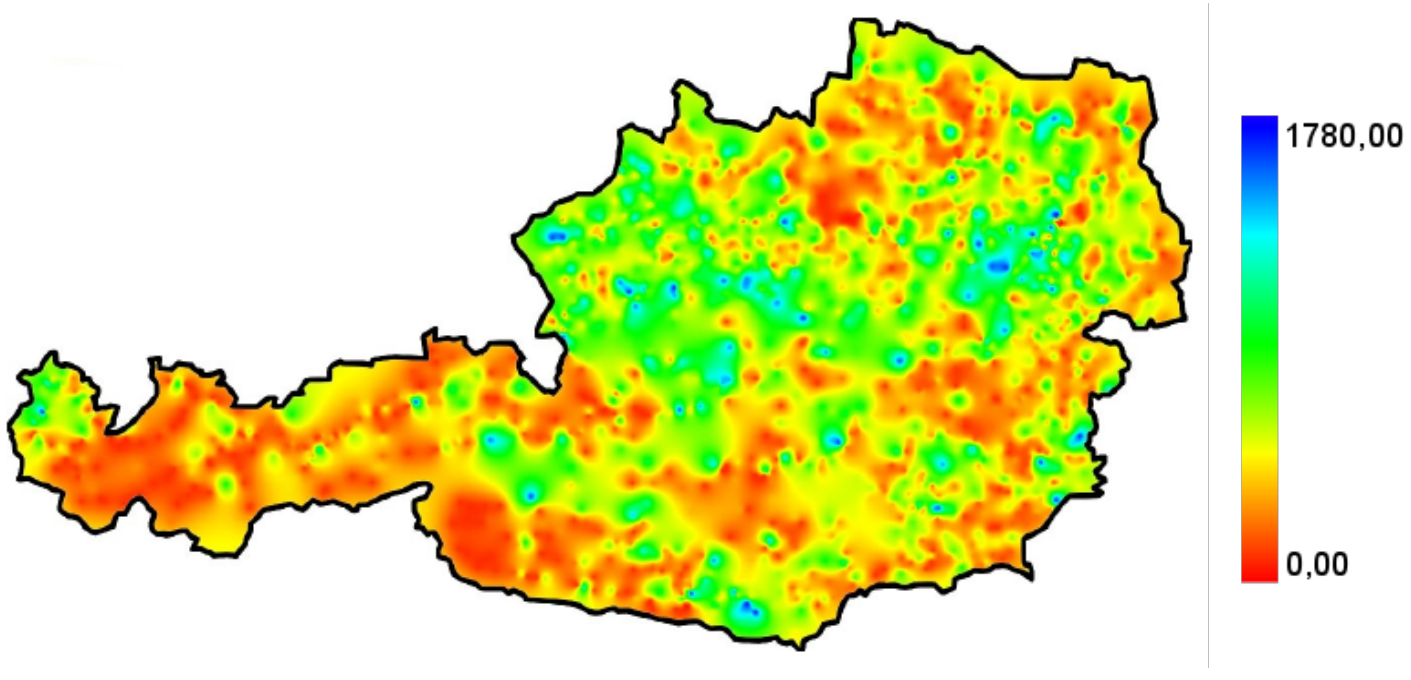
Räumlich aufgelöste Prämien nach Erwartungswertprinzip

Erwartungswertprinzip:

N	: Zufällige Anzahl der Schäden im Portfolio
U_i	: Einzelschäden in einem Portfolio ($i = 1, ..., N$) unabhängig und identisch verteilt
$X = \sum_{i=1}^{N} U_i$	: Gesamtschaden des Portfolios
$\mathbb{E}(X) = \mathbb{E}(U_1)\mathbb{E}(N)$	: Erwarteter Gesamtschaden des Portfolios
$\Pi(X) = \mathbb{E}(X)$	: Prämie nach dem Erwartungswertprinzip
Schätzung ($\hat{\Pi}(X)$)	: durchschnittl. Schadenhöhe $\cdot$ Schadenfrequenz
	$\hat{\mu}_U \cdot \hat{\mu}_N$

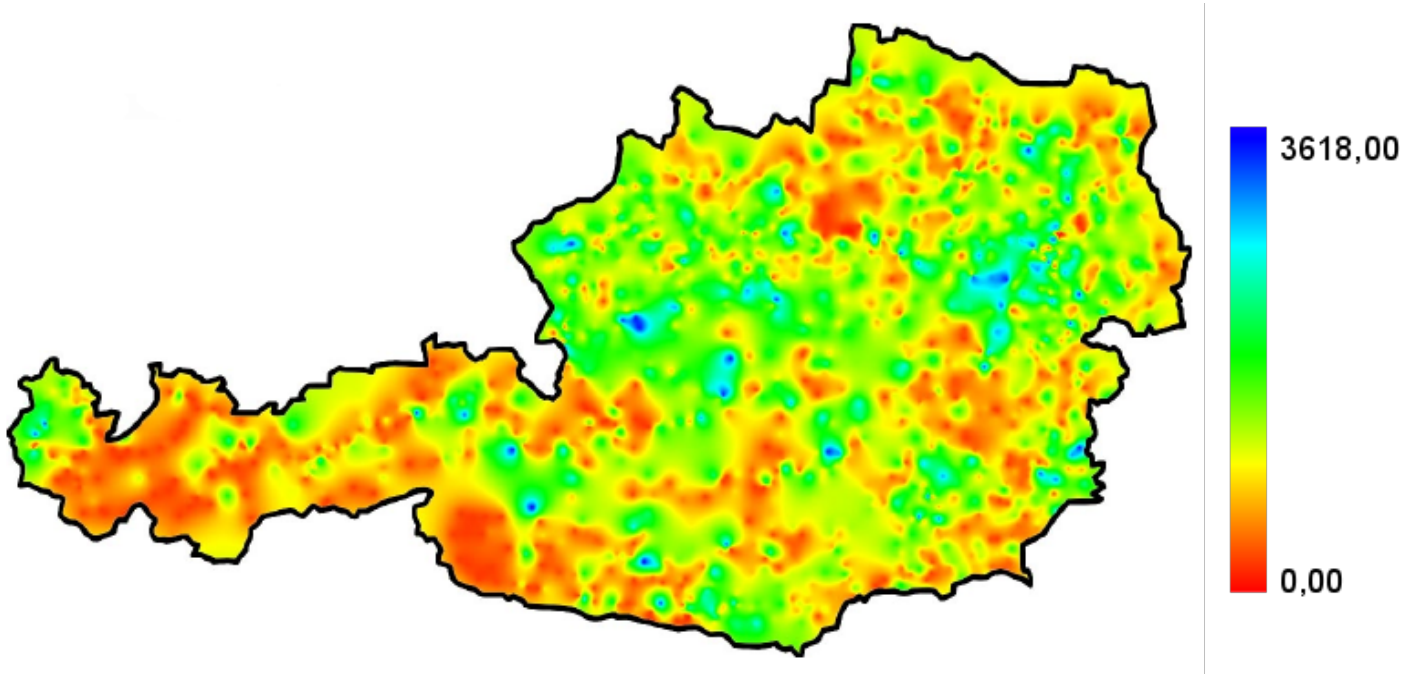
Räumlich aufgelöste Prämien nach Erwartungswertprinzip

$\Pi(X_i) = \mathbb{E}(X_i)$ für jedes Postleitzahlgebiet $i = 1, ..., 2053$



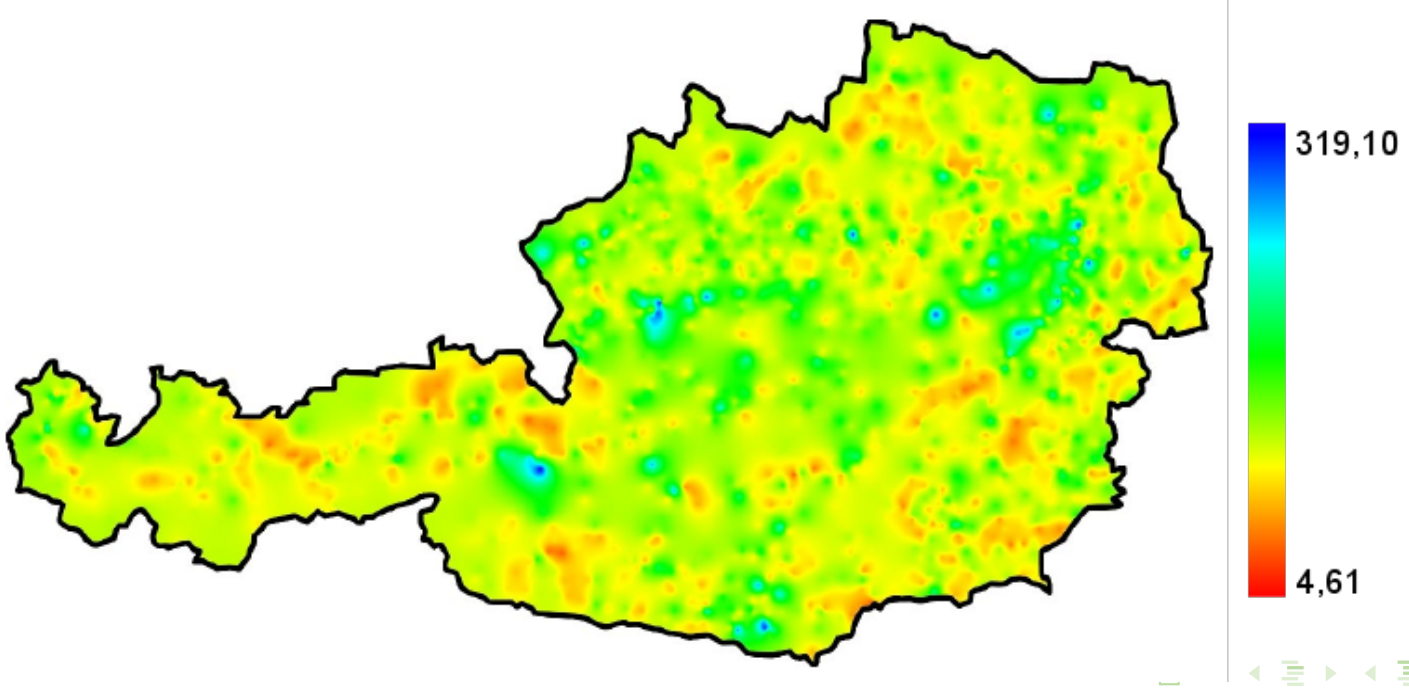
Räumlich aufgelöste Prämien nach Standardabweichungsprinzip

$$\Pi(X_i) = \mathbb{E}(X_i) + a\sqrt{\mathrm{Var}(X_i)} \quad (a = 0,25)$$



Räumlich aufgelöste Credibility-Prämien

$$
\begin{aligned}
\tilde{\mu}_U^i &= \alpha_i \hat{\mu}_U^i + (1-\alpha_i)\hat{\mu}_U \quad \text{für jedes PLZ-Gebiet } i \\
\tilde{\mu}_N^i &= \beta_i \hat{\mu}_N^i + (1-\beta_i)\hat{\mu}_N \quad \text{für jedes PLZ-Gebiet } i \\
\hat{\Pi}(X_i) &= \tilde{\mu}_U^i \cdot \tilde{\mu}_N^i
\end{aligned}
$$



Berechnung des Risikozuschlags

- Forderung: Die Wahrscheinlichkeit, dass die Prämien zur Deckung des Gesamtschadens in einem Jahr ausreichen, soll 99% sein.
- Umsetzung: Erzeugung von genügend vielen Realisierungen des Modells und Bestimmung des 99%-Quantils des Gesamtschadens.
- Berechnung der Gesamtnettoprämie und des Risikozuschlags.

Vergleich des Risikozuschlags

- Mittlerer Risikozuschlag bei den regionalisierten Prämien nach dem Standardabweichungsprinzip ($\Pi(X) = \mathbb{E}(X) + a\sqrt{\mathrm{Var}(X)} \quad (a = 0,25)$):

  39,60
- Risikozuschlag für das präsentierte Modell ($\mathbb{P}(\Pi(X) \geq X) = 0,99$):

  20,37

Zusammenfassung

- Flexibilität des Modells durch die Aufnahme von Kovariablen gegeben.
- Verteilungsannahmen (α-stabile Verteilung vs. Lognormal-Verteilung/Gammaverteilung) passen besser zu den Daten.
- Berücksichtigung von vorhandenen räumlichen Abhängigkeiten in den Residuen.
- Möglichkeit, einen Risikozuschlag basierend auf Ruintheorie zu berechnen.
- Berücksichtigung des aktuellen Bestands bei der Prämienkalkulation.

Vielen Dank für die Aufmerksamkeit!