Statistical analysis of weather-related property insurance claims

Christian Rohrbeck[1]
c.rohrbeck@lancaster.ac.uk

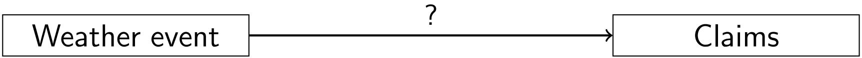
Joint work with Emma Eastoe, Arnoldo Frigessi and Jonathan Tawn

Department of Mathematics and Statistics & Data Science Institute

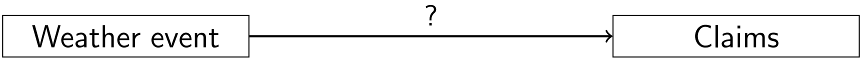
July 16, 2018

[1]Beneficiary of an AXA Research Fund postdoctoral grant

Motivation



Motivation



Hazards:

- Severe rainfall
- Thunderstorm

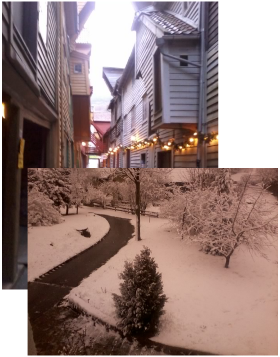
Motivation

Weather event → ? → Claims

Hazards:

- Severe rainfall
- Thunderstorm
- Snow-melt

Events occur rarely and differ in length



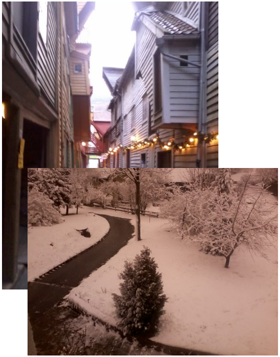
Motivation

Weather event → ? → Claims

Hazards:

- Severe rainfall
- Thunderstorm
- Snow-melt

Events occur rarely and differ in length



Risks:

- Localized flooding
- Sewage back-flow
- Blocked pipes

Lag in recording process

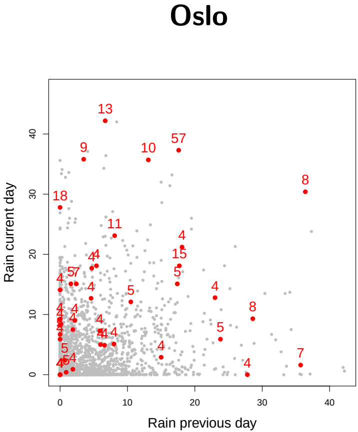
Data I

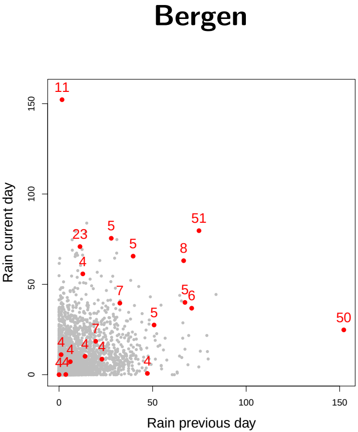
Daily records for Norwegian municipalities for 1997-2006 on

- Reported number of water-related claims N
- Amount of precipitation R
- Amount of snow S
- Surface run-off D
- Mean-temperature T

We aim to model N in dependence on $\mathbf{X} = (R, S, D, T)$.

Data II


Oslo
Rain current day
Rain previous day


Bergen
Rain current day
Rain previous day

Previous research

Scheel et al. (2013) propose a model of the form

$$\mathbb{P}(N = n \mid \mathbf{X}) = \begin{cases} \alpha(\mathbf{X}) & \text{if } n = 0, \\ [1 - \alpha(\mathbf{X})]\,\mathbb{P}(Y = n \mid \mathbf{X}, Y > 0) & \text{if } n > 0, \end{cases}$$

where Y is a Poisson random variable.

Previous research

Scheel et al. (2013) propose a model of the form

$$\mathbb{P}(N = n \mid \mathbf{X}) = \begin{cases} \alpha(\mathbf{X}) & \text{if } n = 0, \\ [1 - \alpha(\mathbf{X})]\,\mathbb{P}(Y = n \mid \mathbf{X}, Y > 0) & \text{if } n > 0, \end{cases}$$

where Y is a Poisson random variable.

But Table 2 in Scheel et al. (2013) shows that their model underpredicts the highest claim numbers:

Period	*Results for Oslo*			*Results for Bergen*		
	Median	*95% prediction interval*	*Observed* ΣN_{kt}	*Median*	*95% prediction interval*	*Observed* ΣN_{kt}
(a)	4	(0, 14)	11	3	(0, 8)	7
	4	(1, 11)	11	3	(0, 7)	7
	3	(0, 8)	8	2	(0, 6)	6
	3	(0, 7)	7	2	(0, 7)	6

Our approach

Rohrbeck et al. (2018) extend Scheel et al. (2013) by

- Introducing a more flexible model using methodology from extreme value theory to handle the highest numbers of claims,

Our approach

Rohrbeck et al. (2018) extend Scheel et al. (2013) by

- Introducing a more flexible model using methodology from extreme value theory to handle the highest numbers of claims,
- Deriving additional predictors to incorporate the spatial and temporal behaviour of the rainfall and snow-melt,

Our approach

Rohrbeck et al. (2018) extend Scheel et al. (2013) by

- Introducing a more flexible model using methodology from extreme value theory to handle the highest numbers of claims,
- Deriving additional predictors to incorporate the spatial and temporal behaviour of the rainfall and snow-melt,
- Proposing a clustering algorithm to merge days whose claims are probably related to the same severe weather event.

In the following, we will focus on the third point.

Clustering algorithm I - Concept

Motivation

- Potential lag in the recording process
- Event may effect claim dynamics on several days

Clustering algorithm I - Concept

Motivation

- Potential lag in the recording process
- Event may effect claim dynamics on several days

Trigger

- Heavy rain $R_t > c$
- Snow-melt $S_{t-1} - S_t > 0$

Clustering algorithm I - Concept

Motivation
- Potential lag in the recording process
- Event may effect claim dynamics on several days

Trigger
- Heavy rain $R_t > c$
- Snow-melt $S_{t-1} - S_t > 0$

Cluster end
- Small change in surface run-off $D_t - D_{t-1} \leq d$
- No snow left on the ground $S_t = 0$

Clustering algorithm I - Concept

Motivation

- Potential lag in the recording process
- Event may effect claim dynamics on several days

Trigger

- Heavy rain $R_t > c$
- Snow-melt $S_{t-1} - S_t > 0$

Cluster end

- Small change in surface run-off $D_t - D_{t-1} \leq d$
- No snow left on the ground $S_t = 0$

Predictors

- Aggregated snow-melt $\Delta\tilde{S}$
- Aggregated rainfall R_{Σ}
- Maximum daily rainfall $R_{\max}$

Clustering algorithm II - Example

Original Data

N	D	S	T	R_t
N_1	0.4	20.4	-8.3	11.2
N_2	0.4	31.6	-2.8	3.5
N_3	0.7	28.1	2.0	0.0
N_4	1.3	18.8	3.1	1.0
N_5	2.0	8.8	3.3	9.0
N_6	2.4	4.6	1.6	2.0
N_7	2.4	4.6	-0.1	1.9

Clustered Data

$\tilde{N}$	$\Delta\tilde{S}$	R_Σ	$R_{\max}$

Clustering algorithm II - Example

Original Data

N	D	S	T	R_t
N_1	0.4	20.4	-8.3	11.2
N_2	0.4	31.6	-2.8	3.5
N_3	0.7	28.1	2.0	0.0
N_4	1.3	18.8	3.1	1.0
N_5	2.0	8.8	3.3	9.0
N_6	2.4	4.6	1.6	2.0
N_7	2.4	4.6	-0.1	1.9

Clustered Data

$\tilde{N}$	$\Delta\tilde{S}$	R_Σ	$R_{\max}$
N_1	0.0	0.0	0.0

Clustering algorithm II - Example

Original Data

N	D	S	T	R_t
N_1	0.4	20.4	-8.3	11.2
N_2	0.4	31.6	-2.8	3.5
N_3	0.7	28.1	2.0	0.0
N_4	1.3	18.8	3.1	1.0
N_5	2.0	8.8	3.3	9.0
N_6	2.4	4.6	1.6	2.0
N_7	2.4	4.6	-0.1	1.9

Clustered Data

$\tilde{N}$	$\Delta \tilde{S}$	R_Σ	$R_{\max}$
N_1	0.0	0.0	0.0
N_2	0.0	0.0	0.0

Clustering algorithm II - Example

Original Data

N	D	S	T	R_t
N_1	0.4	20.4	-8.3	11.2
N_2	0.4	31.6	-2.8	3.5
N_3	0.7	28.1	2.0	0.0
N_4	1.3	18.8	3.1	1.0
N_5	2.0	8.8	3.3	9.0
N_6	2.4	4.6	1.6	2.0
N_7	2.4	4.6	-0.1	1.9

Clustered Data

$\tilde{N}$	$\Delta\tilde{S}$	R_Σ	$R_{\max}$
N_1	0.0	0.0	0.0
N_2	0.0	0.0	0.0

Clustering algorithm II - Example

Original Data

N	D	S	T	R
N_1	0.4	20.4	-8.3	11.2
N_2	0.4	31.6	-2.8	3.5
N_3	0.7	28.1	2.0	0.0
N_4	1.3	18.8	3.1	1.0
N_5	2.0	8.8	3.3	9.0
N_6	2.4	4.6	1.6	2.0
N_7	2.4	4.6	-0.1	1.9

Clustered Data

$\tilde{N}$	$\Delta\tilde{S}$	R_{Σ}	$R_{\max}$
N_1	0.0	0.0	0.0
N_2	0.0	0.0	0.0
$\sum_{i=3}^{7} N_i$	27.0	3.0	9.0

Clustering algorithm III - Results

We set the thresholds c and d in the clustering algorithm as

$$c = q_{0.8}\left(R_t \mid R_t > 0\right)$$

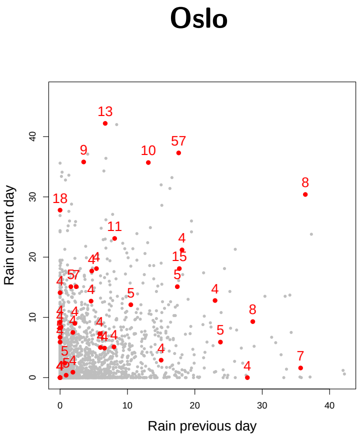
and

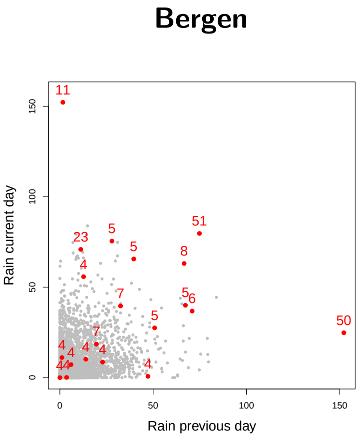
$$d = q_{0.8}\left(D_t - D_{t-1}\right).$$

This gave the following frequency of cluster lengths:

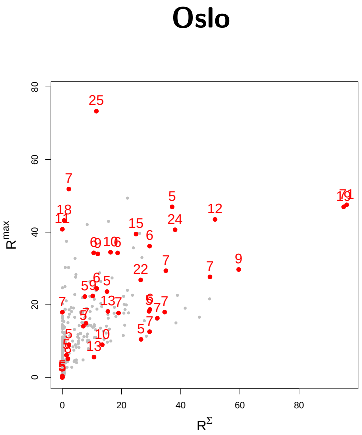
Cluster length in days	1	2	3	4	5	6	> 6
Oslo	2091	254	57	98	43	23	17
Bærum	2453	105	43	92	46	19	18
Bergen	1868	340	55	131	39	23	11

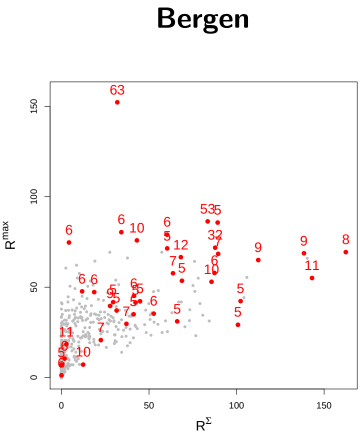
Clustering algorithm IV - Results





Clustering algorithm IV - Results





Statistical model

Model $N \mid (\mathbf{X}, N > 0)$ via a two-component mixture

$$N \mid (\mathbf{X}, N > 0) \sim \begin{cases} Y \mid (\mathbf{X}, Y > 0) & \text{with probability } p, \\ Z \mid Z > 0 & \text{with probability } 1 - p. \end{cases}.$$

Statistical model

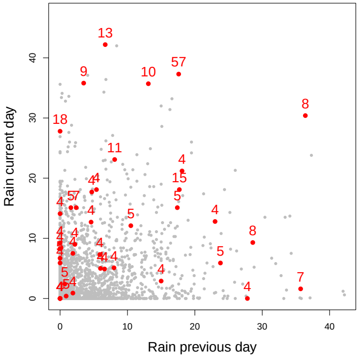
Model $N \mid (\mathbf{X}, N > 0)$ via a two-component mixture

$$N \mid (\mathbf{X}, N > 0) \sim \begin{cases} Y \mid (\mathbf{X}, Y > 0) & \text{with probability } p, \\ Z \mid Z > 0 & \text{with probability } 1 - p. \end{cases}.$$

- Y corresponds to claims related to the observed rainfall and snow-melt
- Z represents claims due to unobserved processes or a high lag.
- Both Y and Z are Poisson distributed, but we replace the tail of Y with a distribution used in extreme value analysis.

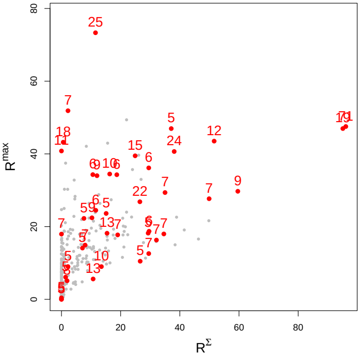
Results - Oslo

Original Data



Data plot

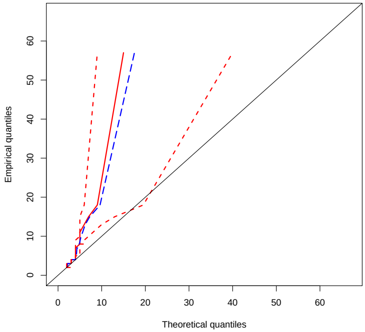
Clustered Data



Data plot

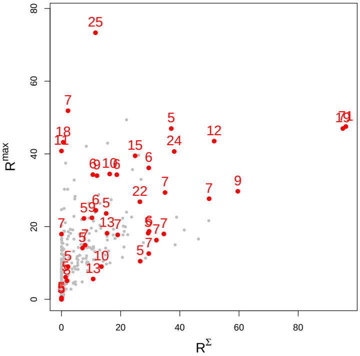
Results - Oslo

Original Data



QQ plot

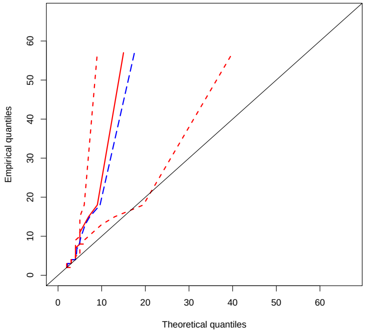
Clustered Data



Data plot

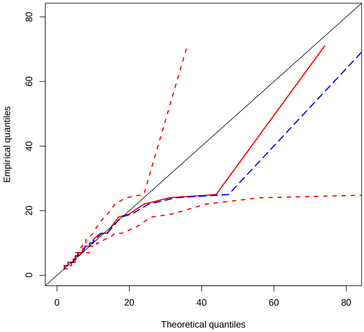
Results - Oslo

Original Data



QQ plot

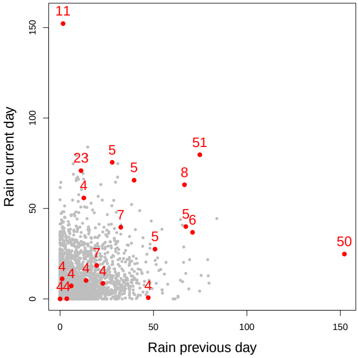
Clustered Data



QQ plot

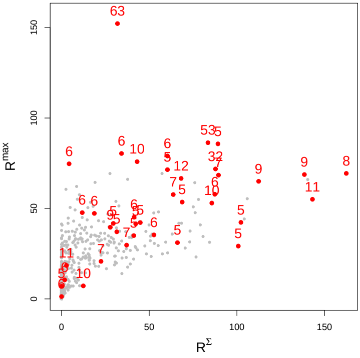
Results - Bergen

Original Data



Data plot

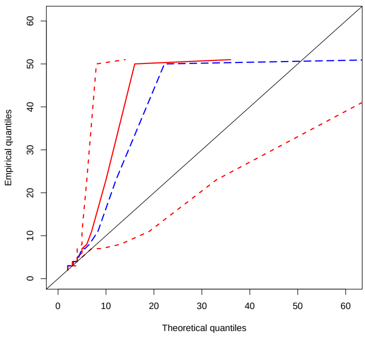
Clustered Data



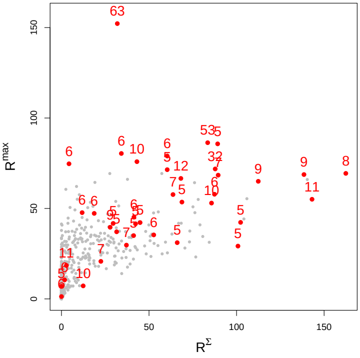
Data plot

Results - Bergen

Original Data | **Clustered Data**



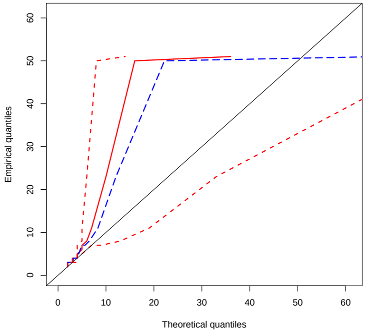
QQ plot



Data plot

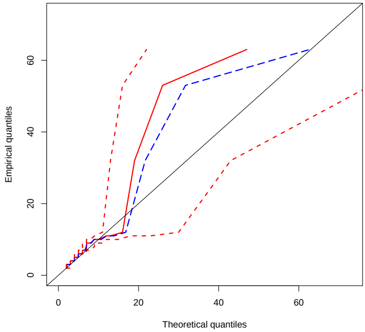
Results - Bergen

Original Data



QQ plot

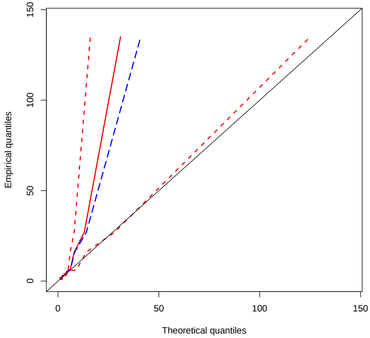
Clustered Data



QQ plot

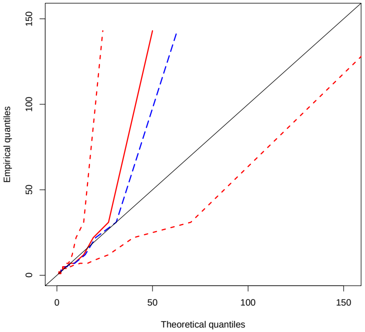
Results - Bærum

Original Data



QQ plot

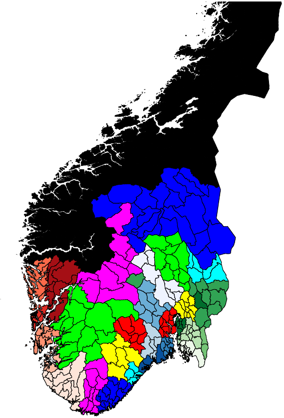
Clustered Data



QQ plot

Current Research

- We want to consider all 430 municipalities.
- **But:** Days with a higher number of claims are very rare for rural municipalities.
- **Idea:** Share statistical information across municipalities.
- **First step:** Detect clusters of municipalities with similar severe rainfall pattern.



References

- Rohrbeck, C., Eastoe, E. F., Frigessi, A., and Tawn, J.A. (2018). Extreme-value modelling of water-related insurance claims. *Annals of Applied Statistics*, 12(1):246-282.
- Rohrbeck, C. and Tawn, J.A. (2018). Spatial clustering of extremal behaviour for hydrological variables. *In preparation.*
- Scheel, I., Ferkingstad, E., Frigessi, A., Haug, O., Hinnerichsen, M., and Meze-Hausken, E. (2013). A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims. *Journal of the Royal Statistical Society: Series C*, 62(1):85-100.

Thank you