Présentation des sujets de recherche de la thèse

A.Mornet[1,◇] - T.Opitz[2] - M. Luzi[◇] - S. Loisel[1] - P.Leveillard[◇]

[1] ISFA - University Claude Bernard Lyon 1
[2] University Montpellier 2
[◇] IARD - Allianz - Paris

WE à Montpellier ▷ *Novembre* 2014

Intro - Les thèmes

- **Création d'un indice tempête**

- **Kilométrage et sexe en assurance auto**

What is a storm ? - Example

A storm over a 550 000 km^2 territory like France can last several days. The disaster claim can also be more or less inaccurate. Therefore, the "day" reference is not the best.

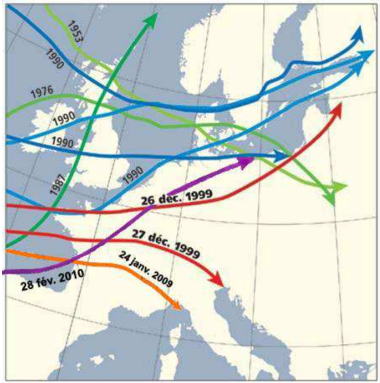


Figure: Storm trajectory over France

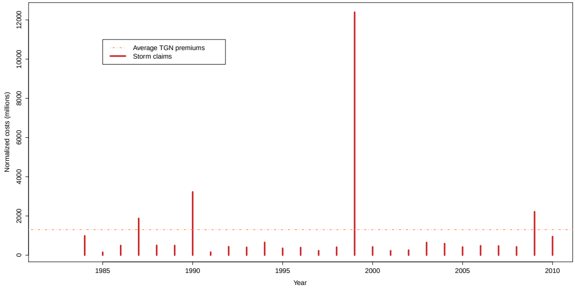
Storm	Date
Lothar	Dec. 25-27/1999
Martin	Dec. 26-27/1999
Klaus	Jan. 23-25/2009
Xynthia	Feb. 26-01/2010

Insurance issues - Intensity and Volatility

From a Market point of view, let us consider a few figures that show the volatility and height of costs caused by storms (*in billion euros*) :

	Annual premiums	Annual claims
All property damages	16	11
Storm guarantee	1.3	?
Lothar and Martin current value	-	12

- 1999 : Same level as 2 years of claims. Storms volatility and intensity have major impacts on the balance of insurance companies' results.
- We have 30 years of historical data, but we want to know Lothar return period (estimations : 50 - 100 years).



Insurance issues - Normalizing Data

In order to work with historical data, we must be able to update their costs.

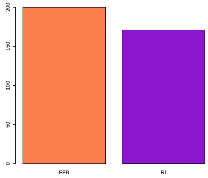


Figure: Update with FFB and RI rates

- Updating insurance data is difficult :
  - FFB rate for private individuals and RI rate for business concerns
  - Relative weight of segments private individuals / business
  - Spreading rate of storm guarantee
  - Growth of real estate in France
- For a 1982 storm the difference between FFB and RI update reach 18%
- If we add the progress of the exposures, the values 2012 of year 1999 increase by 21%.

Storm insurance data are insufficient and hardly reliable. What are the possible solutions ?

Insurance data - 1970-2012

Date	IC	IG	Events
12/25/1999	85	100	Lothar
12/27/1999	43	51	Martin
11/07/1982	10	25	Nov 82
01/23/2009	19	20	Klaus
02/03/1990	14	19	Herta
10/15/1987	6	10	87J
11/30/1976	4	10	Nov 76
07/18/1983	4	10	Jul 83
01/25/1990	6	9	Daria
02/27/2010	7	8	Xynthia
02/26/1990	5	7	Vivian
07/11/1984	3	7	Jul 84
12/15/1990	5	7	Dec 90
08/18/1971	2	6	
02/11/1972	2	6	
01/15/1992	3	4	Jan 92
07/27/2005	3	3	Jul 05
12/17/2004	3	3	Dec 04
05/25/2009	3	3	
07/15/1993	2	3	

List of major storms since the 70s (*according to M. Luzi's works* [7]).

The table shows two kinds of approach :

- IC : inflation referring only to costs
- **IG** : inflation taking into account costs, rising number of parks and of spreading of guarantees

With a broader period, we can work with more substantial events.
Normalizing costs implies some subjectivity due to assumptions.

Meteorological data - Stations

- We use the recordings of 130 Meteorological Stations. If we go back to 1973, fewer stations are available, but their distribution is comparatively homogenous.
- Stations are located very precisely thanks to their GPS coordinates, but it is different for damages that are distributed all over the department.

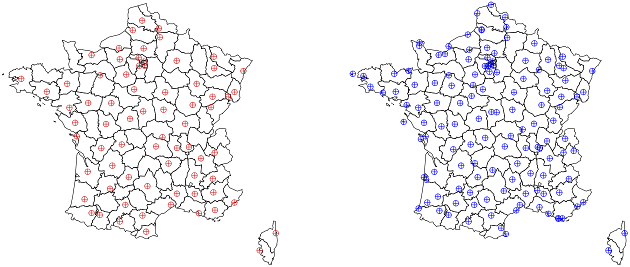


Figure: Location of active stations : 1973 - 1998 (red) and 1998 - 2012 (blue)

Meteorological data - 2 examples

Station	Min.	Qu 0.25	Qu 0.5	Moy.	Qu 0.75	Qu 0.9	Qu 0.99	Max.	Var.
(34)	6.80	35.60	50.40	53.33	64.40	82.4	111.2	183.20	395.04
(08)	6.80	28.80	35.60	34.76	43.20	50.4	68.4	104.00	169.29

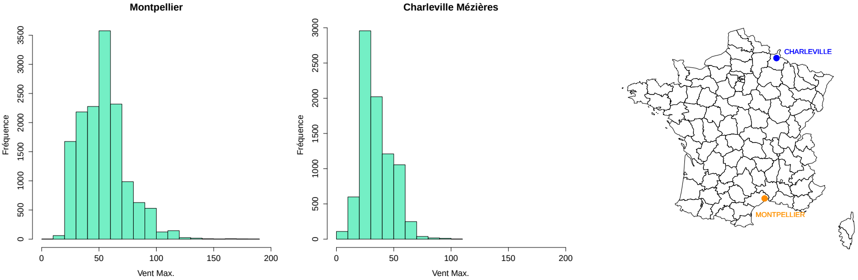


Figure: Maximum wind speeds in Montpellier and Charleville Mézières between 1973 and 2012

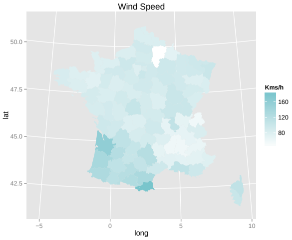
They are very different according to the different parts of the French territory taken into consideration, even through a long period of observation. So different areas = different wind sensitivity

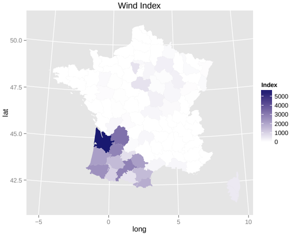
Index construction - Wind index

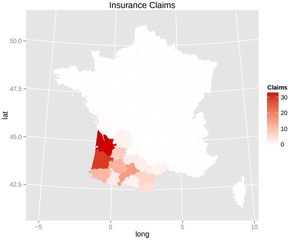
In different areas, the same wind speed will not cause the same damage. Let's define I_w^j the **Wind index** on date j :

$$I_w^j(k) = \left([w^j(k) - w_{98}(k)]_+\right)^\alpha \qquad (1)$$

- $w^j(k)$: Max wind speed at date j in the station k
- w_{98} : 98% quantile for the whole period's speeds







Index construction - Storm index

Storm costs depend on :

- The wind index on date j, $I_w^j(k)$
- The number of risks balanced by the size of population $H(k)$
- The size of the damaged area : geographic aggregation
- The duration of the storm : daily aggregation
- The number of active stations on date j, N_a^j

Storm index :

$$I_S = \sum_{j \in J} \sum_{k} \frac{I_w^j(k) * H(k)}{N_a^j} \qquad (2)$$

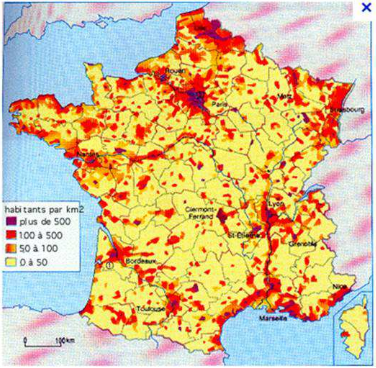


Figure: French population

Comparisons - Department level

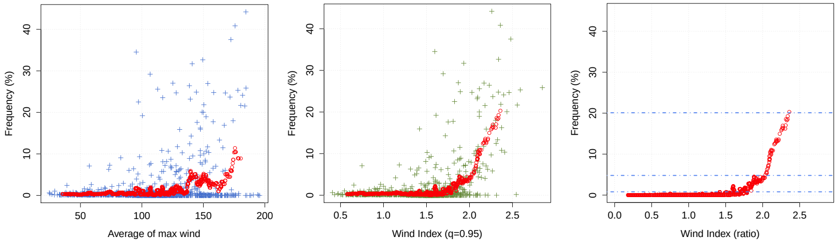
Optimization of the claim criteria

Lines	C/P (2012 euros)	average cost (2012 euros)	frequency
Global	0.137	-0.062	0.164

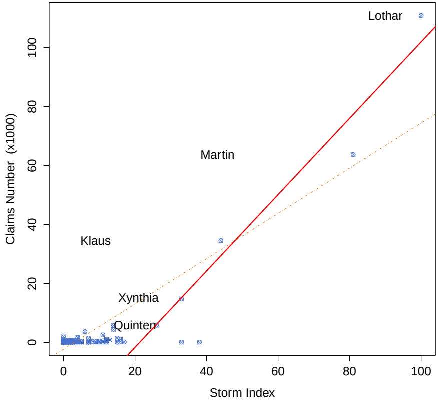
Optimization of the wind criteria

Average speed	Average of maximum speed	Maximum of maximum speed
0.090	0.164	0.146

Moving average



Comparisons - Event scale



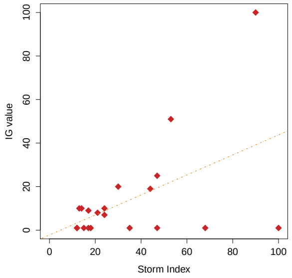
- We estimate the number of claims N due to the main events : Lothar, Martin, Klaus and Xynthia

$$N = 1.2 * I_T - 28.6$$

- For *small* events we obtain weak correlations, but their impact on portfolio is not significant.

Comparisons - Cubic storm index (CI) : values

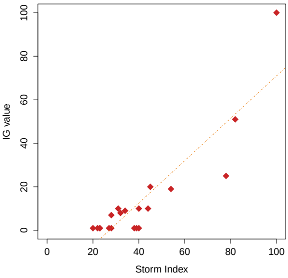
Storm	date	rank(IG)	CI_S	IG
	12/17/1986	>20	100	1
Lothar	12/25/1999	1	90	100
	01/24/1978	>20	68	1
Martin	12/27/1999	2	53	51
	12/01/1984	>20	47	1
Nov 82	11/07/1982	3	47	25
Herta	03/02/1990	5	44	19
	02/03/2002	>20	35	1
Klaus	01/23/2009	4	30	20
Jul 83	07/13/1983	8	24	10
Vivian	02/26/1990	11	24	7
Xynthia	02/27/2010	10	21	8
	02/02/1986	>20	18	1
	08/02/1984	>20	17	1
Daria	01/25/1990	9	17	9
	02/13/1976	>20	15	1
87J	10/15/1987	6	14	10
Nov 76	11/30/1976	7	13	10
	08/20/1976	>20	12	1



- When $\alpha \geq 3$: the order is perturbed
- Localised Storms with high wind speed are overestimated
- *Dec 86 : 2 speeds $>$ 175 kms/h*

Comparisons - Storm index (I_S) : values

Storm	date	rank(IG)	I_S	IG
Lothar	12/25/1999	1	100	100
Martin	12/27/1999	2	82	51
Nov 82	11/07/1982	3	78	25
Herta	02/03/1990	5	54	19
Klaus	01/23/2009	4	45	20
Jul 83	07/13/1983	8	44	10
Nov 76	11/30/1976	7	40	10
	12/17/1986	>20	40	1
	02/07/1984	>20	39	1
	02/03/2002	>20	38	1
Daria	01/25/1990	9	34	9
Xynthia	02/27/2010	10	32	8
87J	10/15/1987	6	31	10
Vivian	02/26/1990	11	28	7
	12/01/1984	>20	28	1
Quinten	02/09/2009	>20	27	1
	01/24/1978	>20	23	1



- With $\alpha = 2$ better order but the scale is different from insurance data.

Comparisons - New parameter

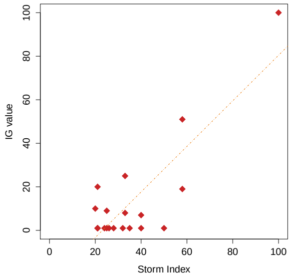
In our case the gaps between the three biggest storms double. This peculiarity is not an isolated case and finds itself for example in the classification of hurricanes proposed by Pielke et al. [11].
In this approach, we have added a new parameter β in order to check the scale of storms.

$$I_S = \exp\left(\beta \sum_{j \in J} \sum_{k} \frac{I_w^j(k) * G(k)}{N_a^j}\right) \tag{3}$$

- The wind index $I_w(k)$
- The number of risks balanced by Global portfolio $G(k)$
- The size of the damaged area : geographic agregation
- The duration of the storm J : daily agregation
- The number of active stations : N_a

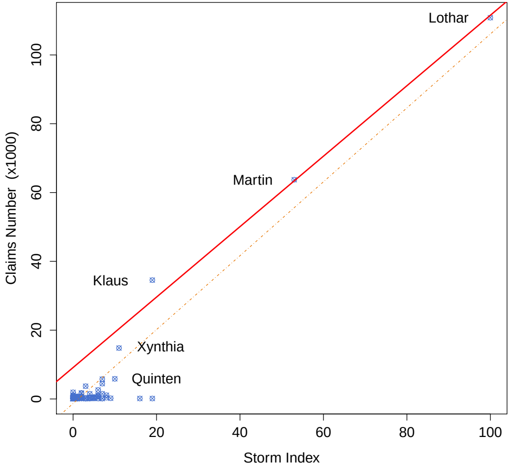
Comparisons - Storm index : values

Storm	date	rank(IG)	I_S	IG
Lothar	12/25/1999	1	100	100
Martin	12/27/1999	2	53	51
Nov 82	11/07/1982	3	37	25
Herta	02/03/1990	5	24	19
Klaus	01/23/2009	4	19	20
Jul 83	07/13/1983	8	16	10
	02/03/2002	>20	14	1
	11/22/1984	>20	14	1
Jul 84	02/07/1984	12	12	7
Daria	01/25/1990	9	12	9
Nov 76	11/30/1976	7	12	10
Xynthia	02/27/2010	10	11	8
87J	10/15/1987	6	10	10
Vivian	02/26/1990	11	10	7
Quinten	02/09/2009	>20	10	1
	12/01/1984	>20	9	1
	01/22/1988	>20	8	1
	01/24/1978	>20	8	1
	05/10/1987	>20	7	1



- With this new index the order and the scale are more coherent with insurance data.

Comparisons - Storm index : plot

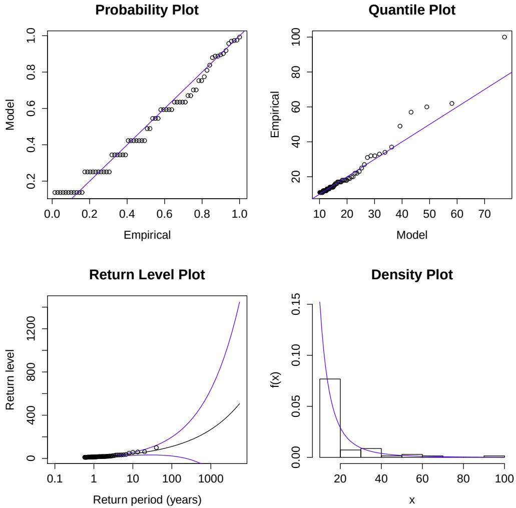


- The relation between this new index and the main events becomes still more obvious :

$$N = I_S + 9.1$$

- It shows the relevancy of our index, which may be useful in anticipation models
- Additional variables (duration of the gusts - building type - ...) could improve the results

Models - Storm index



We use the generalized Pareto distribution (GPD).

$$H(z) = 1 - \exp\left(-\frac{z}{\sigma}\right)$$

or

$$H(z) = 1 - \left(1 + \frac{\xi z}{\sigma}\right)^{\frac{-1}{\xi}}$$

Several thresholds are checked. The index is higher than threshold $u = 10$ in 70 checks i.e. 1.7 times a year in the period. Package `extRemes` - `R` - Nelder-Mead

Models $u = 10$

Return period : it appears that a storm as strong as Lothar takes place every 75 years. Yet on a longer length of time, accuracy decreases

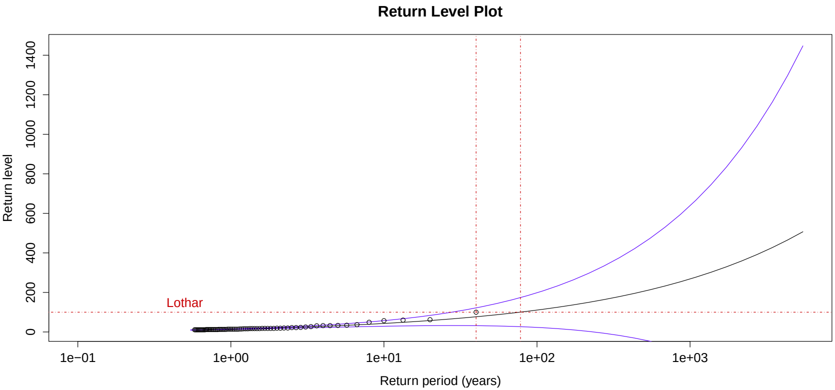


Figure: Storm Return period

Models $u = 20$

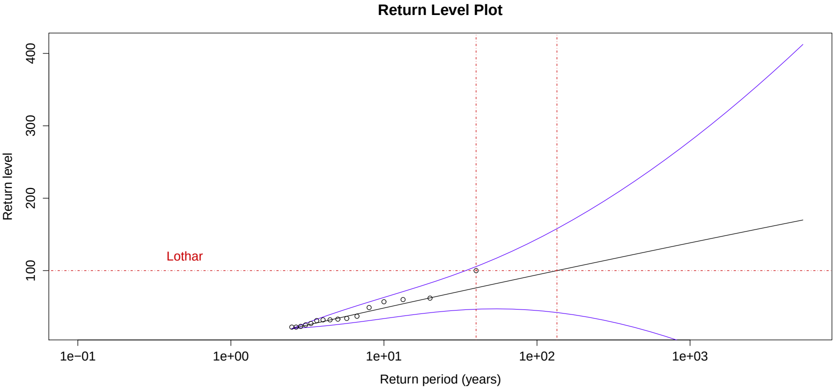


Figure: Storm Return period

By moving the threshold u of the model between 10 and 20, the return period of our time of observation (Lothar) increases to a 135 year period

Models $u = 10$ - Without Lothar

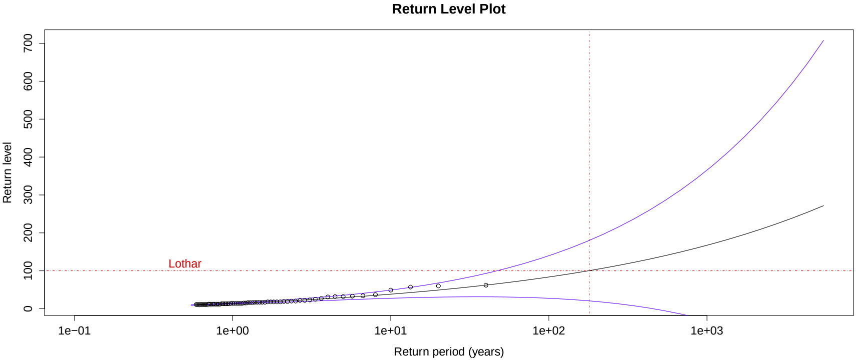


Figure: Storm Return period

We have a Lothar-sized storm every 180 year or about. Without knowing Lothar, you multiply by 2.4 the return period of an event of such importance. Without knowing Martin, we get a return period of 100 years for Lothar.

ASSURANCE AUTO

La cour de justice de l'Union Européenne a décidée de mettre fin le 21 décembre 2012 à la dérogation sur les tarifs et prestation unisexe.
Pour les assureurs, la variable sexe intervient en particulier chez les novices (0 à 3ans d'assurance). On s'intéresse donc à cette population et à son évolution en terme de sinistralité, d'abord en fonction du kilométrage parcouru, ensuite à travers la construction d'un tarif pur avec et sans sexe, enfin en utilisant les techniques de datamining et notamment les arbres de décision.

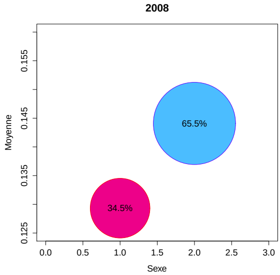


Figure: Moyenne de sinistres hommes et femmes novices en 2008

Caractérisation de la partition Homme-Femme
CART : modèle (2)

Sur les 21 variables descriptives retenues pour caractériser le sexe, le programme n'en retient que 11.

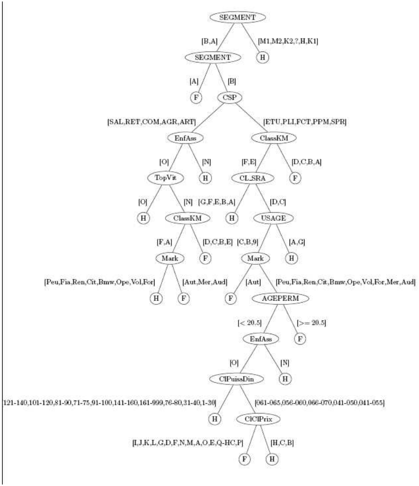


Figure: Arbre de décision

Le modèle est obtenu sur une base de 250 000 polices.
Il est testé sur un échantillon de 125 000 polices.

Résultats sur échantillon test

	SINRC0	SINRC1	SINBDG
Modèles novices sans Sexe			
obs	3148	4124	3583
Pred	3180.36	4124.67	3512.6
Diff. Relative	1.028%	0.016%	1.965%
Modèles novices avec Sexe			
obs	3148	4124	3583
Pred	3180.34	4126.66	3512.12
Diff. Relative	1.027%	0.065%	1.978%

La variable sexe n'apporte pas d'amélioration à l'estimation.

Driver's experience
Approche dynamique (1)

- A : Moins d'un an pour parcourir 12 000 km
- B : Moins de deux ans pour parcourir 12 000 km
- C : Moins de trois ans pour parcourir 12 000 km
- D : Plus de trois ans pour parcourir 12 000 km

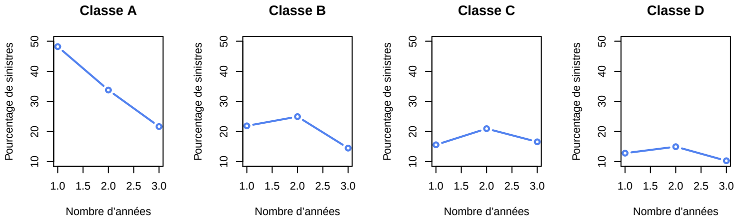


Figure: Évolution de la sinistralité moyenne annuelle entre 2008 et 2010

Lorsque l'on considère la sinistralité par km parcouru, la diminution de la fréquence/Km les résultats changent.

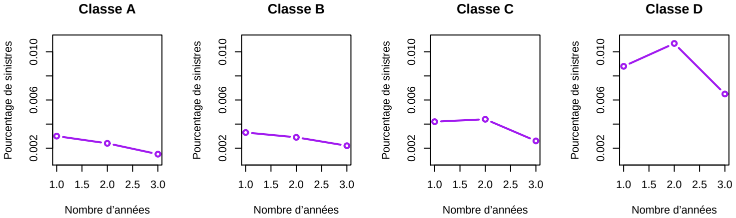


Figure: Évolution de la sinistralité moyenne annuelle par kilomètre entre 2008 et 2010

Some challenges for the future

Algorithme de clustering (k-médoids, aussi appelé PAM = partitioning around medoids) - La distance : corrélation extrêmale entre les vitesses de vent.

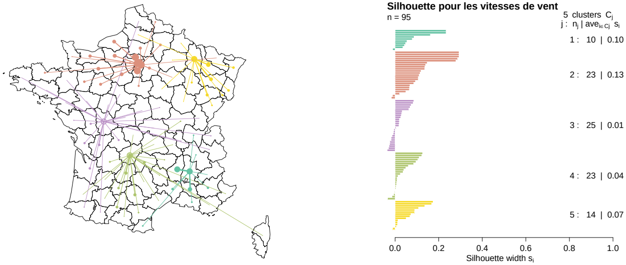


Figure: Silhouette et partition de la France selon les vitesses de vent.

More **insurance** data from the past could make it possible to work on a **broader scale**. **New meteorological informations** such as the duration of the gusts could improve the index

- To study the sensitivity and volatility of the storm costs in the Solvency II context

Bibliography

A. M. Chandler, E. J. W. Jones and M. H. Patel,
Property Loss Estimation for Wind and Earthquake Perils
Risk Analysis, Vol. 21, No. 2, 2001.

M. Klawa and U. Ulbrich
A model for the estimation of storm losses and the identification of severe winter storms in Germany
Natural Hazards and Earth System Sciences 3 : 725-732, 2003.

D. Cooley, P. Naveau, P. Poncet
Variograms for spatial max-stable random fields.
Dependence in probability and statistics, 2006.

P. Embrechts, C. Klüppelberg, T. Mikosch,
Modelling extremal events for insurance and finance
Springer, 4eme édition 2008.

A. Frigessi
Weather events and insurance claims : a Bayesian spatial variable selection model.
Statistics for innovation, 2009.

S.Hochrainer-Stigler, G.Pflug
Risk Management against Extremes in a Changing Environment.
International Institute for Applied Systems Analysis, 2012.

M. Luzi
Travaux sur les historiques Tempêtes
Directeur IARD - Allianz, 2012.

Association française de l'assurance : AFA
Tempêtes, grêle et neige : Résultats 2010
FFSA and GEMA, 2012.

I. Scheel, et al,
A Bayesian hierarchical model with spatial variable selection : the effect of weather on insurance claims
University of Oslo and Norwegian Computing Center, 2013.

Scarrott, Carl and MacDonald, Anna,
A review of extreme value threshold estimation and uncertainty quantification
REVSTATs, Vol. 10, No. 1, 2012.

Pielke, R.A.Jr. and Gratz, J. and Landsea, C.W. and Collins, D. and Saunders, M.A. and Musulin, R.,
Normalized Hurricane Damage in the United States : 1900-2005
NATURAL HAZARDS REVIEW, Vol. 9, No. 1, 2008.

Michel DENUIT, Stefan LANG
Non-life rate-making with Bayesian GAMs.
Insurance : Mathematics and Economics, 35 627-647, 2004.

Allen Greenberg, AICP
Designing Pay-Per-Mile Auto Insurance Regulatory Incentives Using the NHTSA Light Truck CAFE Rule as a Model.
http ://www.vtpi.org/07-3457.pdf, 2006.

Dawn L. Massie, Paul E. Green, Kenneth L. Campbell
Crash involvement rates by driver gender and the role of average annual mileage.
Accid. Anal. and Prev., Vol 29, No 5, pp 675-685, 1997.