



**HAL**  
open science

## Index for predicting insurance claims from wind storms with an application in France

Alexandre Mornet, Thomas Opitz, Michel Luzi, Stéphane Loisel

► **To cite this version:**

Alexandre Mornet, Thomas Opitz, Michel Luzi, Stéphane Loisel. Index for predicting insurance claims from wind storms with an application in France. *Risk Analysis*, 2015, 35 (11), pp.2029-2056. 10.1111/risa.12395 . hal-01081758

**HAL Id: hal-01081758**

**<https://hal.science/hal-01081758>**

Submitted on 10 Nov 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Construction of an Index that links Wind Speeds and Strong Claim Rate of Insurers after a Storm in France

A.Mornet\*

T.Opitz†

M.Luzi‡

S.Loisel§

July 2014

---

\* Université de Lyon, Université Claude Bernard Lyon 1, Institut de Science Financière et d'Assurances, 50 Avenue Tony Garnier, Lyon F-69007, France and ALLIANZ, Coeur Défense, 82 Esplanade du Général de Gaulle, Courbevoie F-92400, France.

† Université Montpellier 2, Institut de Mathématiques et de Modélisation de Montpellier, 2 Place Eugène Bataillon, Montpellier F-34095, France

‡ Ancien Directeur actuariat non vie et Directeur recherche et développement chez Allianz France, Membre qualifié de l'Institut des Actuaire, 132, rue du Président Wilson, Levallois-Perret F-92300, France

§ Université de Lyon, Université Claude Bernard Lyon 1, Institut de Science Financière et d'Assurances, 50 Avenue Tony Garnier, Lyon F-69007, France

### **Abstract**

For insurance companies, wind storms represent a main source of volatility, leading to potentially huge aggregated claim amounts. In this article, we compare different constructions of a storm index allowing us to assess the economic impact of storms on an insurance portfolio by exploiting information from historical wind speed data. Contrary to historical insurance portfolio data, meteorological variables can be considered as stationary between years and are easily available with long observation records; hence, they represent a valuable source of additional information for insurers if the relation between observations of claims and wind speeds can be revealed. Since standard correlation measures between raw wind speeds and insurance claims are weak, a storm index focusing on high wind speeds can afford better information. This method has been used on the German territory by Klawns and Ulbrich [14] and gave good results for yearly aggregated claims. Using historical meteorological and insurance data, we assess the consistency of the proposed indices construction and we test their sensitivity to their various parameters and weights. Moreover, we are able to place the major insurance events since 1998 on a broader horizon of 40+ years. Our approach provides a meteorological justification for calculating the return periods of extreme storm-related insurance events whose magnitude has rarely been reached.

**Keywords.** Storm Index, Wind Speed, Insurance, Extreme Value Theory, Extreme Dependence

# 1 Introduction

For property damage insurance, natural events are the main source of volatility. Historical data covering a 30-year period show that, among these events, storms cause more than half of the costs. This represents a challenging problem for insurers that have to decide on pricing, reinsurance and capital requirements. Classical techniques for the prediction of future claim amounts are based solely on companies' information (exposed portfolios and damage), which leads to some typical problems. The insurers keep in a detailed way only the recent history (15/20 years). So, the amount of information makes it more difficult to get reliable results, more particularly for the most extreme events that are also the least frequent ones.

As we work on historical data, it is necessary to normalize them, which introduces a distortion of the database. This distortion increases with the number of factors that must be taken into account (property values, growth of real estate, spreading rate of storm guarantee).

Managing extreme risks [12] is an issue involving the whole society and not only insurers. When we model extreme events using only damage data ([8], [13]), results show a strong dependence on model hypotheses which are difficult to verify in practice. In our context, adding meteorological data to insurance data is one possible solution to improve the quality of predictions and the fundamental understanding of how storms generate damage. Several authors suggest comparing the amount of damages with meteorological data. [22] apply a model using raw meteorological variables to predict the number of claims per day in Norway. [23] seek to predict the number of casualties of storms and typhoons in the USA. For a more accurate assessment of damages, [2] consider the calculation of a damage function for storms and earthquakes, whereas [10] and [5] create an indicator of extreme climate events on a large scale. Other studies have tried to find an empirical link between the wind speeds and the vulnerability to storms in order to calculate a damage function [11]. Approaches similar to ours have been proposed, leading to the construction of storm indices that indicate the major events in terms of insurance costs and that are based on wind speeds as well as geographical and demographic variables. For the German territory, Klawns and Ulbrich [14] obtain good results for yearly aggregated claims with data from the relatively small number of 23 meteorological stations. In [14], [17] and [6], damage is assumed to be proportional to wind speed to a power of 3. This choice is justified by physical reasoning, with the cube of wind speed being proportional to wind energy. Other authors [7, 18] propose to link damage to an exponential of wind speeds and include information about wind direction. Prettenhalder and al. [18] also take into account the vulnerability of a building. For the construction of our index, we operate on a fine spatial and temporal resolution, using daily data from 130 weather stations covering the territory of France. We study constructions based either on powers or on exponential expressions of wind speed. Parameter like the exponent are considered as free, so we can adjust them to maximize the correspondence between the storm index value and damage for the major storms in the last decades.

For several decades, the insurance industry has covered storm damages. The sector has access to historical data structured according to various aspects (chronological, geographical, costs, frequencies, ...). However, as pointed out above, one must be aware of the limitations and problems of using such data, especially when we deal with the most extreme events.

Our work aims at creating a simple but useful index to assess the economic impact of storms. As we can rely on meteorological data that are linked to storms, we try to find solutions to make use of that kind of information. Meteorological variables such as wind speed present the advantage of being easily available with long observation records over a dense network of weather stations. A major part of observed wind speeds are small or moderately large and cause claim amounts that are negligible compared to damages caused by the most extreme wind speeds. We focus on extreme wind speeds to build a storm index that adds useful information to our research. To ensure the validity of the index we must verify the solidity of the conditions. We test the coherence of the index by comparing it to the meteorological and insurance history. We bring out its sensitivity with respect to its parameters and weighting schemes. Following multivariate Extreme Value Theory (EVT) approach [8], [3], we show that our storm index presents strong tail dependence with actual claim frequency, which is highly desirable for risk management purposes. Moreover, the index approach enables us to analyze the major events on the broadest possible observation period.

Our paper is organized as follows: first, we recall different definitions for a storm. Then, we present the economic issues and we describe the available insurance and meteorological data to try to tackle

them. In the following section we explain step by step the construction of the wind index and the storm index. The last part is dedicated to comparisons between claims and different index. From these comparisons, we define the best adapted Formula and we study the index evolution during the most extreme and expensive events for the insurer.

## 2 What is a storm?

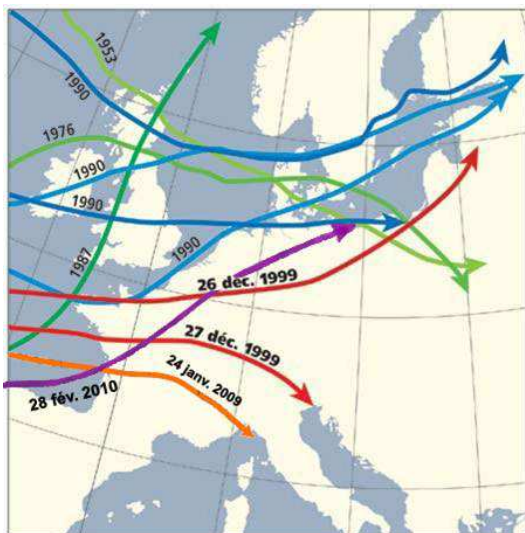
### 2.1 Meteorological point of view

Can we give a single simple definition of a storm event? A precise description of the phenomenon is hard to formulate. From a meteorological point of view, storms are considered as atmospheric disturbances that involve pressure and wind variations that can bring about risks of property damage. At sea, it means an atmospheric depression that brings about an average 90 kph wind speed, i.e. force 10 or more on the Beaufort scale (from 0 to 12). On land, a storm is a depression that brings about gusts of wind that cause damage and heavy precipitations. Atlantic storms that reach France before most European countries are due to two concomitant causes: the frequent lasting jet stream at high altitudes born above Newfoundland that blows at impressive speeds (200 kph) above Northern Europe and the midday decrease of earth temperature around the globe from the Equator to the North. For Meteo France a storm is declared when more than 5% of the weather stations record wind speed above 100 kph during 3 consecutive days.

### 2.2 Insurance point of view

From an insurance standpoint, the claims registered in storm guarantee range from damages due to the slightest wind breath to those caused by large-scale phenomena (strong intensity, large affected territorial area, period which could exceed 3 days). For example, in 1999, Lothar hit the Northern half of France, with wind speed above 140 kph, between 25 and 27 December.

Note that the insurance of storms in France follows a regulatory framework . The 25th June 1990 law rules that the storm guarantee is compulsory for all risks that are covered for fire. It entails that all the insured risks are necessarily covered without adverse selection. According to the rules of the market, hail and snow guarantees are associated with storm guarantee.



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

Figure 1: Storm trajectory over France

Generally storm data are available on a daily basis. However a storm over a 550,000  $km^2$  territory like France can last several days as illustrated on Figure 1. The main storms since the 50s are represented. Four of them crossed France: Lothar and Martin in 1999, Klaus in 2009 and Xynthia in 2010. All of them lasted more than one day.

If the major storms take time to cross the country, the disaster declaration can also be more or less inaccurate, especially by absent people at the time of the facts. For these reasons, the treaties of reinsurance plan generally to group in the same event the claims declared for a period from 2 to 3 consecutive days. Lothar and Martin constitute a particular case because these two storms took

place in the same period by one day but on different areas. We could then think of grouping them in one cluster, but for the insurers it consists in two different events. This definition of the disaster is important in particular for the reinsurance which also works by event. In our article these two storms are treated separately.

Combining the meteorological and insurance approaches, we can conclude that a storm event is characterized by the wind intensity, the geographic spread of the damaged area and its duration that may exceed one day.

### 3 Insurance issues

In personal and commercial insurance, wind property damages are covered through TGN (Tempête, Grêle et Neige : Storm, Hail and Snow) guarantee. A study published by AFA (Association Française de l'Assurance : French Insurance Association) [9] in February 2012 shows the volatility of yearly costs related to wind on a national scale. The current costs vary between 70 million and 7 billion euros. Since 1984, storms represent 87.5% of the claims of the TGN guarantee. With a total of about 12 billion, the 2012 cost of damage caused by Lothar and Martin storms (year 1999) is at the same level as the average annual cost for the whole Property coverage. Regarding premiums, the sum of all property damages amounts to 16 billion among which 1.3 billion are dedicated to the TGN guarantee, about 8 %.

In order to compare costs during a long period, normalization is key to the validity of results. We choose here to convert all the costs into 2012 euro value. This stage of data treatments is generally underdeveloped. We nevertheless underline that depending on the used hypotheses, the results vary appreciably. For example, according to the choice of the construction index (FFB or RI<sup>1</sup>), the values 2012 of year 1982 vary by 18%. If we add the progress of the exposures (growth of the housing stock), the values 2012 of year 1999 increase by 21%.

Though some approximations cannot be avoided, we try to be as exhaustive as possible by taking into account the evolution of FFB rate for private individuals and RI rate for business concerns, the relative weight of segments private individuals / business in the portfolio, as well as the spreading rate of storm guarantee that only became compulsory in July 1990 and the growth of real estate in France.

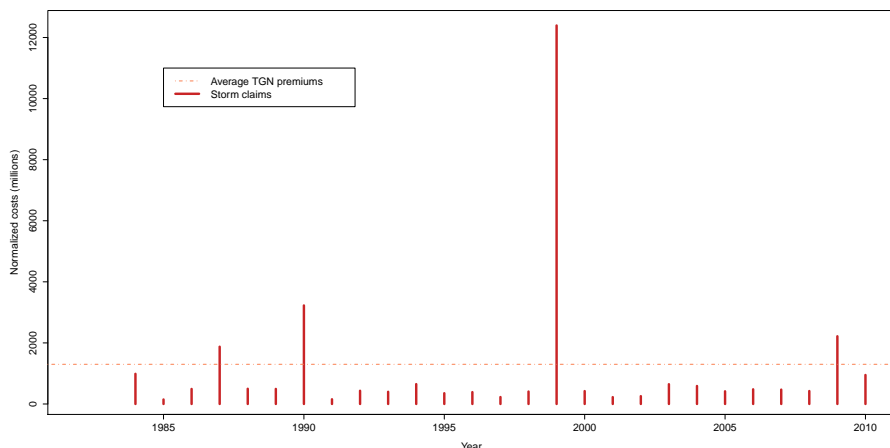


Figure 2: Normalized annual costs for storm guarantee between 1984 and 2010

Figure 2 refers to the 2012 annualized costs related to storms since 1984 (the figures come from FFSA's report: French Federation of Insurance Companies) . It shows both the volatility and the severity of the costs. We notice that the great majority of the years have values lower than the TGN annual average premium (dotted orange line: 1.3 billion). Only 4 years exceed this level, among which year 1999 which is very appreciably higher. We point out that the burden of the year 1999 represents more than 43% of the whole period costs. The return period of events of the same or even greater importance has considerable influence on results. For instance, if one spreads the impact on Lothar and Martin over 50 or 100 years, the annual TGN cost tends to rise substantially.

We have to differentiate here the notions of annual costs and costs by event. The quality and accuracy of these data are different. Except for the major storms, the statistics are not followed by event at Market level. Consequently, the use of data from an insurance company brings additional details.

<sup>1</sup>FFB (Fédération Française du Bâtiment : French Building Federation, 3.93% / year) and RI (Risques Industriels : Industrial Risks, 3.38% / year) are two French building discount rates

## 4 Insurance data

### 4.1 1998-2012

#### 4.1.1 Database description

We use the Allianz France data base. This portfolio contains all the individual claims between 1998 and 2012. It is a base of 520,000 claims, among which 310,000 for personal lines and the rest for commercial lines. For every claim, we know the amount, the date of occurrence and the place of the claim (department <sup>2</sup>, even municipality).

#### 4.1.2 Comparaison with Market

If the use of the data of a single company can skew the results, we notice that the annual loads are strongly correlated to those of the market. Also, if we study the geographical distribution of the portfolio, we notice that the distribution is close to that of the French population. Even if the results are not totally identical everywhere (parts between 4% and 14% according to departments, for an average value close to 9%), they are not very different from those of the market.

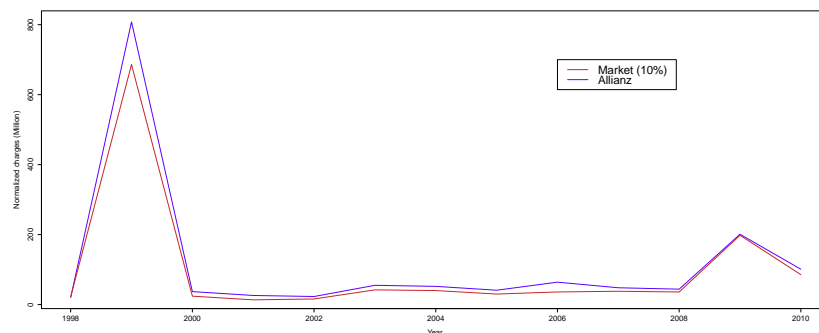


Figure 3: Comparison between Allianz annual costs and those of the market

Figure 3 compares Allianz' wind property damage costs (in blue) with 10% of the market (in red). It reports that the Allianz' loads are very highly related (0.99 %) to those of the market.

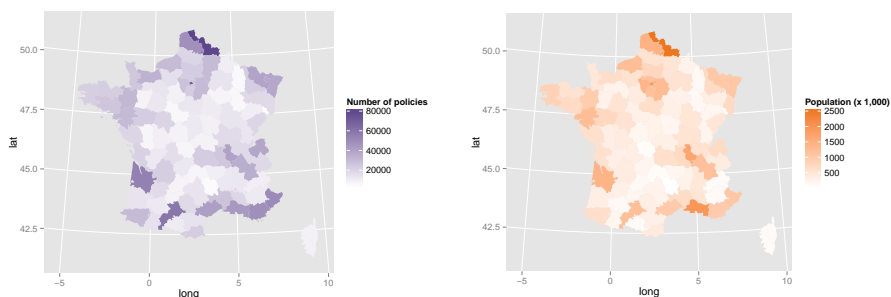


Figure 4: Distribution of Allianz portfolio and French population in 2009

Figure 4 shows the distribution of the number of contracts in each department in Allianz portfolio and the French population. On the left map we notice great geographical differences in the representation of policyholders. For instance, there are less than 5,000 contracts in some departments, whereas there are more than 50,000 contracts in some other ones. However, the comparison with the second map makes it possible to relativize differences which are generally in proportion with population. In fact,

<sup>2</sup>In metropolitan France, the territory is subdivided into 95 departments which correspond to administrative areas.



the most populated areas are also those where we find the greatest number of contracts, and vice-versa.

Thus, it is possible to rely on Allianz data for a study of damage on a national level.

### 4.1.3 Database analysis

The distribution of individual claims brings an essential information about the scale and the nature of the damages caused by storms. Table I shows the various quantiles as well as the average and the maximum of the 430,000 positive individual claims <sup>3</sup> registered by Allianz during the period from 1998 to 2012. The values are normalized in euros 2012. This distribution is very uneven with a large number of small claims and only around thirty exceptional disasters exceeding one million euros. The average individual claim amount approximately 4,600 euros <sup>4</sup>. Half of the claims are lower than 1,700 euros and 99% of them are lower than 45,000 euros. Yet, the maximum claim takes a much higher value by reaching more than 5 million euros.

In this portfolio, the personal lines represent 260,000 positive claims out a total of 430,000. By looking closer at the values, we notice that the personal lines and the commercial lines have not the same claim distribution. The damages sustained by professionals are superior in every point than those of private individuals. The average claim is 2.3 times higher for the professionals and the difference increases with a 5 times higher maximum claim.

Quantiles	2012 values	commercial lines	personal lines
Min.	1	1	1
Qu 0.25	677	749	640
Qu 0.5	1 667	2 005	1 542
Moy.	4 581	7 032	2 993
Qu 0.75	3 879	5 062	3 279
Qu 0.9	7 865	11 983	6 263
Qu 0.99	45 184	82 682	24 050
Qu 0.999	244 621	455 144	73 982
Max.	> 5 M	> 5 M	> 1 M

Table I: Distribution of individual claims (1998-2012)

The period from 1998 to 2012 spans 5479 days. Only 958 days (17.5%) are without claim. But less than half of the days record more than 5 claims. Less than 1% of the days record more than 500 claims. On the most damaged day the number of claims reached 200,000. The distribution in numbers and costs is quite uneven.

The results presented below focus on all the days when Allianz recorded at least one claim with a strictly positive value.

Quantiles	2012 values	Number
Min.	3	1
Qu 0.25	1 921	2
Qu 0.5	6 557	5
Moy.	434 800	116
Qu 0.75	25 630	12
Qu 0.9	103 160	39
Qu 0.99	1 708 116	494
Qu 0.999	35 007 641	14 056
Max.	915 000 000	196 000

Table II: Distribution of daily damages (1998-2012)

<sup>3</sup>We choose to retain only the positive claims so that our results are not influenced by the non positive claims (more than 16%) which can be bound to cases without continuation, amounts lower than the franchise or other unknown reasons.

<sup>4</sup>The average cost by retaining all claims (positive and non positive) would rather be 3,500 euros.

Table II shows the distribution of daily damage over the 1998 - 2012 periods. Costs are normalized in euros 2012. In fact, generally, most of the stricken days bring relatively light costs. In one quarter of the days, costs do not exceed 2000 euros and in one half, they are under 7000 euros. Only 10% of the days are over 100,000 euros. Yet, the average claim reaches over 430,000 euros. Daily costs can literally rocket during exceptional events such as Lothar and Martin in 1999 or Xynthia in 2010. We then get 78 days over 1 million euros, 10 days over 10 million euros, and a maximum 915 million euros (Dec. 26th, 1999). If we consider the number of events, the result is nearly identical: 90% of impacted days register fewer than 40 claims and half of the days register fewer than 5 claims. The parts under 10,000 euros represent 2/3 of the whole claims. The top part (over 10 million) seems to be very unrequent.

Thus we generally deal with damage due to the wind but which are too localized to be called storms. According to the first results it brings to study the data from the standpoint of events. Once the data will be grouped in events, we obtain approximately 130 significant events during the period. Table III shows the distribution of the main storms that crossed France since 1998. The 4 most important ones are named and the others are grouped in the last column of the Table. Claims recorded during storms Lothar and Martin in December, 1999 are relatively higher at any point than those of more recent events like Klaus or Xynthia. The average claim varies between 6,000 euros for Lothar and 2,700 euros for Xynthia. At first sight the individual claims of storms having caused globally most damages are higher than the others.

Quantiles	Lothar	Martin	Klaus	Xynthia	Others
Min.	1	1	1	1	1
Qu 0.25	820	1 000	646	476	677
Qu 0.5	2 141	2 318	1 485	1 075	1 667
Moy.	5 968	5 190	3 388	2 707	4 549
Qu 0.75	4 850	4 940	2 717	2 212	3 878
Qu 0.9	9 884	9 124	6 245	5 070	7 865
Qu 0.99	62 748	46 672	31 830	25 746	45 164
Qu 0.999	355 300	230 379	143 271	111 763	243 642
Max.	> 5 M	> 2 M	> 2 M	> 1 M	> 2 M

Table III: Distribution of main event damages (1998-2012)

From these data, it is impossible to provide a detailed explanation for these variations. However we can bring background hypotheses. The cost of the building repairs can increase during a strong demand. The normalization choices, given that two main events took place in the beginning of period, the size and the location of the damaged area are not known and could nevertheless strongly influence the results.

## 4.2 1970-2012

If it is not possible to get back older historical data with the same level of detail, it is however possible to have information relative to the major events. This information is generally available in the departments responsible for reinsurance.

These data present certain advantages. It is possible to identify the dates when significant events occurred. We also have an information about the level of the damages supported for every event, what allows us to have a certain hierarchy. With these data which could go back up on several decades, it is possible to specify better the return periods of rather rare events.

On the other hand, it is not possible to work on a detailed level, on an individual claim level or on a department level. Besides, considering the age of information held or the length of time for which it may be kept, the problems of updating are even more acute. Very often, the studies use only an updating on the amounts of the claims, based on a building index. In fact, the subject is more complex and the gaps from methods can lead to very different results.

Table IV presents the normalized cost of storms since the seventies (according to Luzi's work [15]). Therefore, we propose two kinds of approaches: **Act IC** corresponds to an inflation only based on costs, **Act IG** corresponds to an inflation that takes into account costs, extensions of parks and

guarantee spreading.

Date	Act IC	Act 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	

Table IV: List of major storms since the 70s

Thanks to this data recovery, we have about twenty major events over the period from 1970 to 2012, while we limited ourselves to 7 events over the period from 1998 to 2012.

The normalization choice modifies appreciably the results, especially those from the oldest exercises. Even the ranking of the events can be impacted. Throughout this paper, we use the **Act IG** index which offers the advantage to take into account the inflation of costs, the extensions of the building parks and the guarantee spreading.

In spite of the care taken to value the claims, these data will of course never reach the level of precision of a purely physical measure.

## 5 Meteorological data

### 5.1 Stations

The use of meteorological data is quite natural in the context of a storm study in order to complete insufficient insurance information. Those data are available on a individual day basis and with a relatively long history (since 1973). We use the National Climatic Data Center (NCDC) which gets information from Météo France for France. It is an American website which offers a free and unlimited access. From time to time, we noticed some differences with data produced by Météo France.

From 1973, it is possible to have recordings on numerous stations. Before this date, the information is poorer. From 1998, weather station network allows us to obtain at least one source of information for each department.

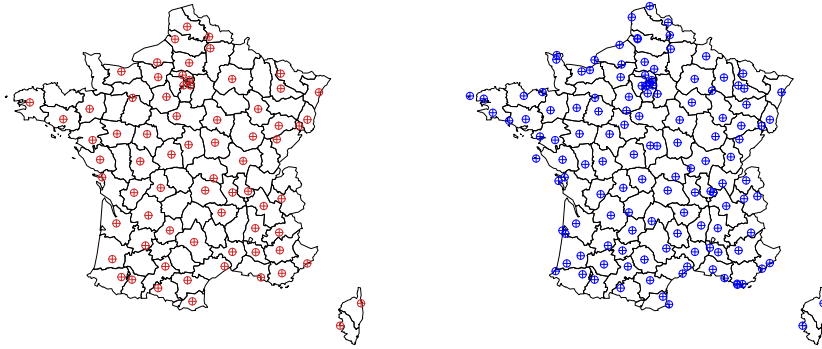


Figure 5: Location of active stations : 1973 - 1998 (red) and 1998 - 2012 (blue)

We work with on average 90 stations between 1973 and 1998, then up to 130 stations between 1998 and 2012 (Figure 5). If every station is very precisely located, the claims can occur everywhere in the department. The distance between the station and certain buildings can be furthermore of 100 km.

### 5.2 Wind speed

On the National Climatic Data Center (NCDC) we have collected different kind of information about the wind speed: average wind speed, maximum sustained wind speed and maximum wind gust. The air flows generally irregularly pulling a strong variability of the wind in direction and in force. This is why the meteorologists measure the instantaneous wind which varies ceaselessly and the average wind calculated over a period of 10 minutes. When the instantaneous wind speed exceeds that of the average wind of more than 10 knots (18 kph) the meteorologists speak about gust.

We convert all data into kph. For every station, we establish a distribution of max wind speed per day. It leads us to highlight very different situations from one station to the other. In order to account for these differences, we show in Table V and Figure 6 how speeds are distributed in the towns of Montpellier (from department 34) and Charleville-Mézières (from department 08).

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

Table V: Daily maximum wind speed in Montpellier and Charleville Mézières between 1973 and 2012

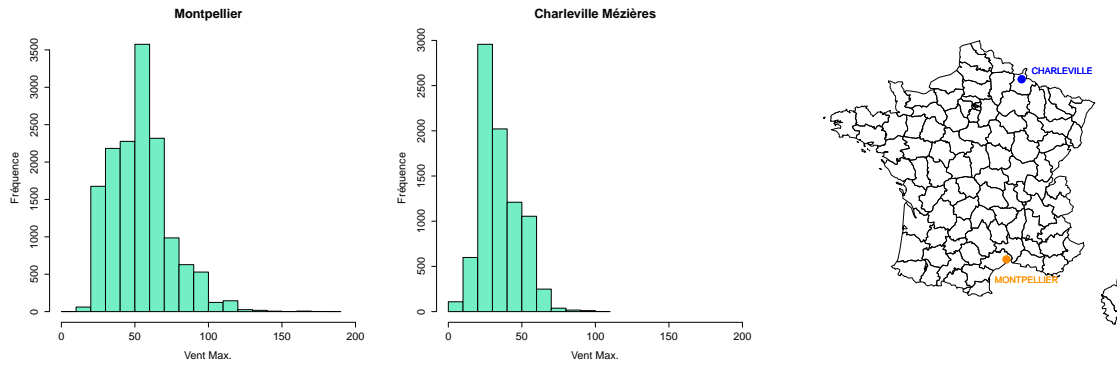


Figure 6: Daily maximum wind speed in Montpellier and Charleville Mézières between 1973 and 2012

Sometimes they are over 200 kph in some stations, while others hardly reach 100 kph. In Montpellier the highest recorded speed is 183.2 kph, 10% of the recordings are over 82 kph and 1% over 111.2 kph. Though it is relatively protected by the nearby hills, the station faces Mistral and Tramontane winds. On the contrary, wind speeds in the Charleville-Mézières stations are far lower. The peak speed is only 104 kph and only 1% of the recordings are over 68.4 kph. Consequently an unusual wind speed will not always be a sufficient criterion. It will also be necessary to take into account the scale of the phenomenon, and not to overestimate the individual overtakings (as we shall see it in the Subsection 7.4.9, page 31).

More generally, these differences between the various zones can be observed all over the French territory. Figure 7 shows the average wind speeds and the 99% speed quantile in each of the 95 departments. Generally speaking, coasts record wind average higher than in the center of the country. Thanks to the 99% quantile, the highest speed in each department can be observed. Then the discrepancies are very much increased with a 60 kph minimum and a 160 kph one. This 100 kph difference shows the lack of homogeneity in the distribution of wind speeds in France. So it seems that some departments have developed a behavior of adaptation to the wind. It may have an influence on the local ways of building and of the reactions of the insured about storm risks. For instance, house roofs will be stronger if they are often exposed to gusts over 100 kph. Our study will take these disparities into account by observing wind-speeds according their local distribution.

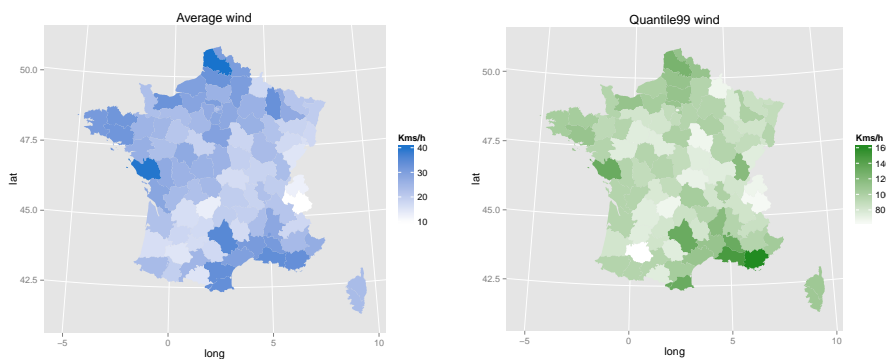


Figure 7: Average wind speed and 99%-quantile in France between 1973 and 2012

## 6 Index construction

### 6.1 Wind Index

Properly speaking, there is no windless day, but in most cases the wind-speed is so low that it does not cause any damage. On the other hand, the periods corresponding to a storm are much less numerous.

For our study, we thus have to spot the periods corresponding to the chosen definition for storms: an event presenting strong wind speeds or high claims, on a relatively large territory, on a period from 1 to 3 days. This definition remains unclear, but it already allows us to eliminate one very large number of days without significant effects. Moreover the used criteria are not necessarily identical depending on the studied base. Nevertheless, for the main events, which are the target of our study, a correspondence is possible from a base to the other one.

On the insurance side, if we are able to connect a number of claims and a cost, it is necessary to find a correspondence from meteorological data. So we can determine a rule of comparison. It is thus necessary to create an index which translates the intensity of every storm. This index must rather simply be in line with the numbers of claims or with the costs. In an intuitive way, we imagine that this index takes into account wind speeds from every station and that the event takes into account results from all the stations, during 1 to 3 consecutive days.

How to define an intensity at the level of the station? We imagine to use the wind speed as main information. But which function can we choose? Several kinds of Formula have been tested, as well as numerous parameters. In our study, we hold the following function:

$$I_w^j(k) = ([w^j(k) - w_q(k)]_+)^{\alpha} \quad (1)$$

where  $w^j(k)$  is the maximum wind speed at date  $j$  in the station  $k$  and  $w_q$  is the upper  $q\%$  quantile for the whole period's speeds. We only keep  $(1 - q)\%$  of the highest speeds in every station in relation with their own distribution. In the following section, different quantile values are tested. By adding the exponent  $\alpha$  to this index, we increase or decrease the influence of the most extreme wind speeds. This method can surely be improved, but we essentially wish to see if we can improve the results of our model by using such an index.

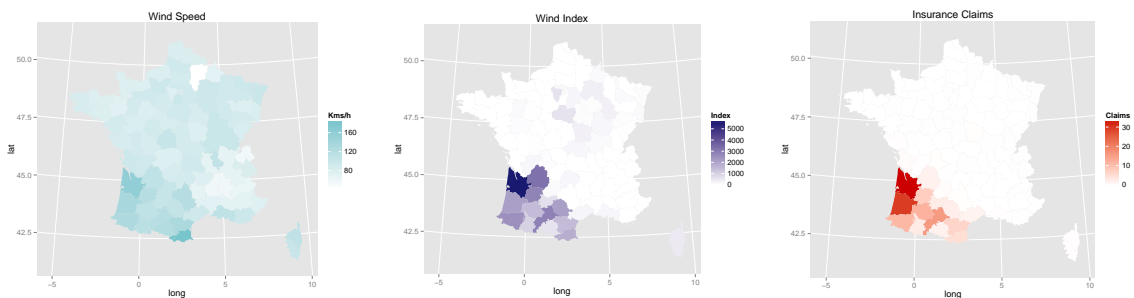


Figure 8: 3 illustrations of the KLAUS Storm

The first test is a comparison, during a storm, between the maximum wind speeds and the wind index in each department and over 2 or 3 days, to try and find some similarity with the damages registered by Allianz. Figure 8 shows the maximum wind speeds, the wind index and the normalized costs of damages (in million euros) during the Klaus storm between Jan 23rd and 25th 2009. Klaus is a severe windstorm that struck the South West of France. From the wind speeds, the most impacted zone can be delimited but with perfectible accuracy. When you study the 2nd map with the wind index in every department, you can see that it is better connected with claims, so that zones that are little impacted by the storm can be excluded. With this graphic approach, the correlation between winds and their frequency is more obvious.

Other Formulae for the wind index have been already proposed. M. Klawa and U. Ulbrich [14] use a cubic index based on a ratio between the maximum wind speed of the day and a quantile of wind in the station  $CI_w$ :

$$CI_w(k) = \left( \frac{w(k)}{w_{98}(k)} - 1 \right)^3 \quad (2)$$

A regional wind loss relation is also proposed by M.G. Donat et al.[6]:

$$\text{loss ratio}(k) = A(k) \left( \left[ \frac{w(k)}{w_{98}(k)} - 1 \right]_+ \right)^3 + B(k) \quad (3)$$

The different parameters are tested later, by comparing the results we get with insurance data.

## 6.2 Storm index

We have already mentioned that strong wind speeds are an essential cause of damages, but they are not enough to define a storm. From a geographical point of view, every station corresponds to a different exposure area. Several stations cover very populated departments with numerous insured risks. Others, on the contrary, cover more desert zones. It is thus necessary to balance the results of every station by a variable corresponding at best to the local exposure. Theoretically, it would be necessary to know the data that is specific to each event. Failing that, we can, at least, use the geographical distribution of the population <sup>5</sup>.

It is also necessary to notice that in certain cases, several storms affect the country on the same day. It was particularly the case on December 26th, 1999 with Lothar and Martin. It is then necessary to distinguish departments touched by the different events. It is not always easy. So, the first step of the construction of our storm index consist in multiplying the wind index on date  $j$ ,  $I_w^j(k)$  and the [number of risks](#) balanced by the size of population  $H(k)$  in the station  $k$ .

$$I_w^j(k) * H(k)$$

Then the [size](#) of the damaged area must be considered. So we apply a geographic aggregation of all the weighed wind speeds. We obtain a first Formula of the storm index on date  $j$ ,  $I_S^j$ .

$$I_S^j = \sum_k I_w^j(k) * H(k)$$

The definition of this index has been established empirically and could be improved. However, it is based on established standards of damage-assessment. Here again, the choice of variables and stations will be adapted according to insurance data.

From a temporal point of view, the lasting days consideration allows to imagine several solutions. At this stage of the construction of the index we have to choose between two methods of aggregation of the values, the sum or the maximum. The sum offers the advantage to preserve all the information and the maximum focuses on the main day of the storm event. We finally decide to keep the sum by considering that the trajectory of the storm can bring about strong values of wind on several days in different places. This choice was then supported by the results obtained in term of correlation (Subsection 7.4, page 25). In that way, we take into account both the course of the storm that can stretch over a few days and the frequent delays in damage claims. We define  $I_S$  as

$$I_S = \sum_{j \in J} \sum_k \frac{I_w^j(k) * H(k)}{N_a^j}, \quad (4)$$

where in each station  $I_w^j(k)$  is the wind index on date  $j$  and the [number of risks](#) is balanced by the size of population  $H(k)$ . We also take into account the [size](#) of the damaged area (geographic aggregation), the [duration](#) of the storm  $J$  (daily aggregation) and the number of active stations on date  $j$ ,  $N_a^j$ .

For reasons of confidentiality, the results are established by assigning the value 100 to the most important event. Thus we point out the most striking events in the period according to our index. One advantage of the storm index is its relatively long period of observation. In fact, we can use reports of wind speeds since 1973 in one station at least in each department. This first result is

<sup>5</sup>Other weighings are possible as the number of contracts of every department (Subsection 7.4.5, page 27). It would be interesting to know also the number of all the buildings classified by categories (individual houses, buildings, etc.) but we did not obtain this information.

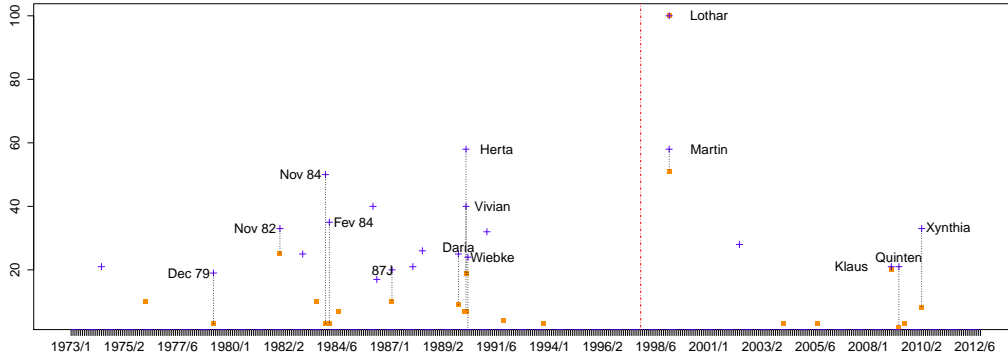


Figure 9: Storm index in connection with population between 1970 and 2012

encouraging (Figure 9). The blue cross represent the storm index values and the orange squares represent insurer classification. All the data picked out by the index are those of known storms (Meteo France history). Lothar storm in December 1999 is remarkable with an index that is almost twice higher than Martin's. As we expected, this broadening of the observation period makes it possible to remark other storms in the seventies, eighties, and nineties. Yet the ordering of events in decreasing values does not express exactly the impact they may have had in terms of damage and insurers' costs. For instance, Xynthia storm is clearly above Klaus, though the latter cost twice as much. Herta storm gets more or less the same index as Martin though it was not so important. At this step, the link between our index and insurance results is not obvious. We find unmatched points in both categories of information. We will see (Figure 25 and 28) to what extent optimization of the used factors and the variable choices improve correlations.



## 7 Comparisons

### 7.1 Objectives

We compare results stemming from meteorological data with observed results from insurance. Our purpose is to study the correlations between the diverse parameters, to measure the relevance of the correlations, to select the most appropriate parameters. If these comparisons are relatively classic, it is necessary to underline that none of these results is perfect.

### 7.2 Some issues

Concerning claims, the issues come from the updating and from the geographical representativeness of a limited portfolio. We work here on the Allianz' database combining the personal and the global lines at the department level. We use the claims values normalized and balanced by sum of premiums (loss ratio). On the other hand, the numbers of claims are not normalized but balanced by the number of policies to obtain claim frequencies. After the treatments of updating and by referring to portfolios insured every year, we can hold more normalized values allowing to make comparable the results between departments. So the 3 types of insurance information about claims are: the normalized loss ratio, the average normalized claim cost and the frequency of claims.

Concerning the wind, the issues come from the low granularity of stations and relevance of the information (daily average speed, maximum speed, gust). We can have several stations on the same department and some departments without station. For departments with several stations, the recordings can be very different from a place to the other on the same day. Generally, the average speed of a station tends to smooth observations and on the contrary the only maximum gust tends to exaggerate the phenomena [20]. We tried to determine an intermediate solution by coupling the data of several stations for every department. We thus retain at least 3 stations for every department. Every station is balanced according to its closeness with the most urbanized zones. In case the number of stations is insufficient, we also use the data of the nearby department stations. We test here 4 types of wind speed measures: the weighted department stations average of daily average wind speed, the maximum on department stations of daily average wind speed, the weighted department stations average of daily maximum wind speed and the maximum on department stations of daily maximum wind speed.

EVT intuition suggests to use maximum of maxima as averaging out extremes is usually a bad idea. However, for our problem, we shall see that the best solution is the third one, as the storm causes many claims if it is strong and widespread. Our comparison concerns events. Yet the scale of day is not always relevant to realize the comparisons. Even, at the departmental level, the disasters relative to the same event can be declared on several days. For the wind, a storm can take several days to cross France, or stagnate more or less time. We cannot focus on comparisons only. We thus grouped certain days between them to correspond better to the events.

### 7.3 Wind Speed and claims at department level

This step is essential to determine the most suited information at the station level. We have detailed information over 15 years (5,480 days) and on 95 departments (more than 500,000 references). If the volume is very important, we also notice that a very big part of these days are without wind and without claim. We thus have to limit our works in the significant days for our study. Several solutions were tested and finally, we choose to select the days which correspond to **the upper 3% of wind speed or the upper 3% of claims**. By not knowing the best criterion for every element, the selection is the widest for every parameter. When we combine the selected days it represents **around 35,000 data**.

#### 7.3.1 Optimization of the claim criteria

Table VI shows Pearson's correlations between the wind speed and the values and the frequency of claims. We are aware that linear correlation is not always the best tool to analyse correlation. We use it at this stage to get a rough idea of the quality of an index. Later on, we carry out a more in depth analysis, focusing in particular on correlations of extremes. For raw wind speed, we observe a positive trend with total cost as well as with frequency. Nevertheless, the correlation does not exceed

17%. In the case of average costs the correlation is negative and weak (about 5%). By comparing the results of the global and personal lines, we can consider that the claim criterion does not bring fundamental difference. We thus focus for the rest of the study on the global portfolio.

Lines	loss ratio (2012 euros)	average cost (2012 euros)	frequency
Global	0.137	-0.062	0.164
Personal	0.139	-0.047	0.161

Table VI: Correlation with average of maximum wind speed

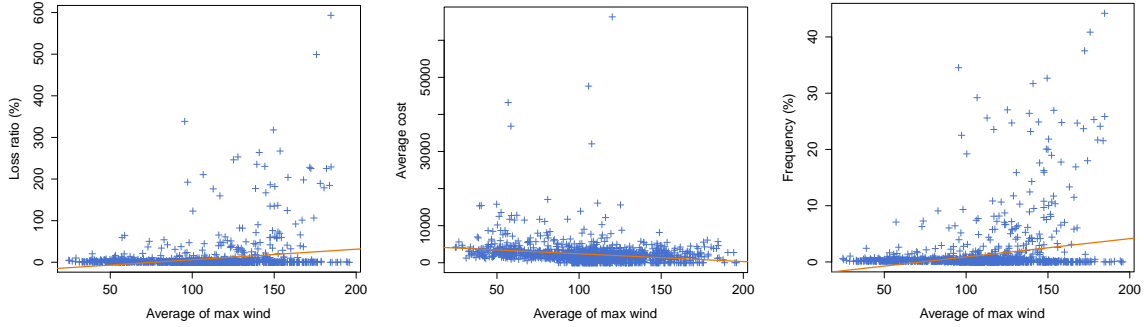


Figure 10: Correlation between claims and average of maximum wind speed

On Figure 10, the strongest claims are located in the upper right corner concerning loss ratio and frequency. We notice a certain progress of claim rates according to wind speed. Nevertheless, we observe many points along the x-axis, which is indicative of low loads or a low frequency of claims though the wind speed was high. We notice that the claim loss ratio and frequency remain very low under a certain wind threshold.

For average costs, the main claims seem to take place during days with a low wind speed. No clear relation appears with wind speed. The average costs are not thus a good information source for our work.

As the results of frequencies and loss ratio are rather close, we shall favor the frequencies as the claim criterion to optimize the criteria of the wind.

### 7.3.2 Optimization of the wind criteria

In a general way, the results obtained on the basis of average winds are unsatisfactory. We obtain better results from the maximum wind speed.

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

Table VII: Correlation with 3 types of estimation of the wind speed

On Table VII we compare the correlation obtained with 3 different ways to estimate the wind speed in a department. Among the various options to express the wind speed in every department, the best solution is to consider a weighted average of the maximal speeds (best correlation with more than 16%). On the left-hand graph (11) : average of winds, the highest claim frequencies are gathered in the central part and not to the right with the fastest winds. The right-hand graph corresponding at maximum speed on all the weather stations of the department improves the relation wind/claim but remain unsatisfactory. The middle graph corresponding to the average of the maximal speeds balanced on the department offers the best distribution of points.

Several graphs show a no loss ratio for high speeds of wind. Would there be an effect concerning the habituation in the wind? To test this phenomenon, we looked for Formulae allowing to take into account this effect. Several solutions were envisaged. First, we try a difference between wind speed

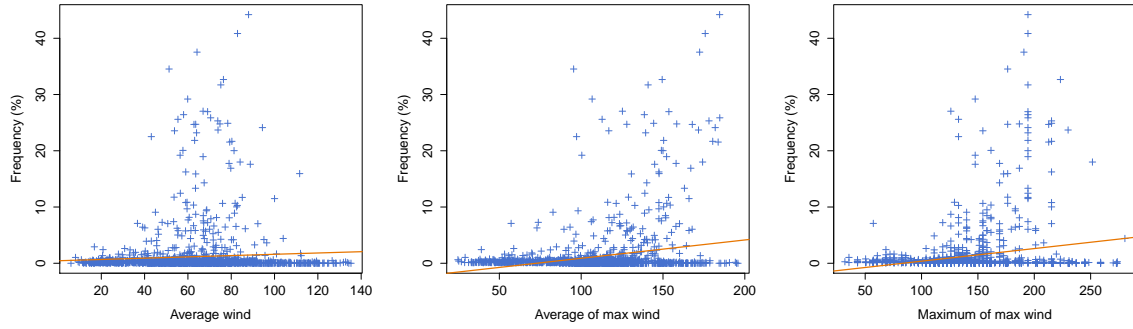


Figure 11: Correlation between claims and 3 types of estimation of the wind speed

and a quantile:  $(w - w_q)^\alpha$ , then, we compare the obtained results with a ratio between wind speed and a quantile:  $\left(\frac{w}{w_q}\right)^\alpha$ . As we use quantiles as thresholds the local reference is established from a rather rarely reached speed in the department. We can then test if the results are sensitive to the different threshold (5%, 3%, 2% ...).

Furthermore, we are also going to show that working under the shape of moving averages allows to present results of a lesser volatility.

### 7.3.3 Moving average principle

A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. The threshold between short-term and long-term depends on the application, and the parameters of the moving average will be set accordingly. From the ratio index, we can then observe the moving averages to underline the improvements brought by the weighted average of the maximal wind speed. Here we chose to calculate a moving average centered on 12 events.

Denote by  $(x_{(k)}, y_{(k)})$  the points where the first coordinate is ranked from the smallest to the largest one and  $y_{(k)}$  is associated to  $x_{(k)}$ . The red points are obtained thanks to a moving averaging principle with the Formula:

$$\left(x_{(k)}, \frac{1}{25} \sum_{i=k-12}^{k+12} y_{(i)}\right). \quad (5)$$

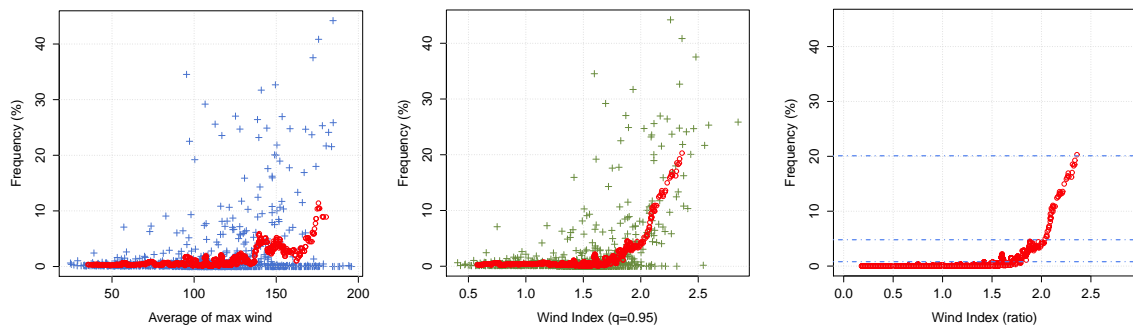


Figure 12: Moving average between claim frequency and wind ratio

On Figure 12 we can compare the scattered plot from the weighted average of maximum wind speed (blue dots), the ratio wind index (green dots) and the corresponding moving average (red circles) calculated from the relation with claim frequencies. This presentation shows clearly the sensitive improvement brought by introduction of the notion of habituation in the wind. On the right-hand graph which represents only the averages, we can distinguish three kinds of data. The first part groups a very large number of the observations, with moderated wind speeds and low levels of claims. The intermediate part (about 1400 points) contains some stronger wind speeds and higher claim rates.

The top part, concerning not much case (about 65 points), presents a stronger accentuation of the loss ratio with regard to the increase of the wind speed.

From these results, we choose to concentrate on **the most striking 1500 events** of the period which constitutes the great majority of important claims. For the wind speeds we shall use the weighted average of the maximums. For the claims we shall select the frequencies and loss ratio, and we shall abandon the average costs which give poor results. It is now necessary to study several subjects. Which function presents the best adjustment between the loss ratio and the wind? Does the use of an exponent reduce the spread? What extrapolation can we imagine (at the most, frequency cannot exceed 100%)? Do the extreme values (strong loss ratio with low wind speed or worthless loss ratio with high wind speed) present common causes (type of department, habituation in the wind...)?

### 7.3.4 Wind index Formulae

Now, we test various Formulae of the wind index at the level of the department and during the 1,500 most important events of the period 1998-2012 from insurance and meteorological points of view. We begin with a difference between the maximum wind speed  $w$  and a quantile  $w_q$  (like in our wind index 1):  $(w - w_q)^\alpha$ . We test quantiles between 95% and 99.5%.

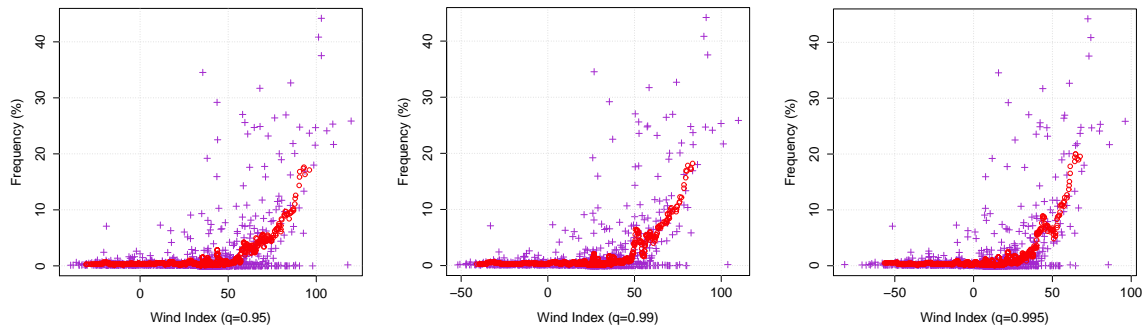


Figure 13: Correlation between claim frequency and wind index (difference)

On Figure 13, we present the scattered plot (purple dots) and the moving average (red circles) calculated between the claim frequency and the difference index. From left to right, the quantiles raise from 95% to 99% and 99.5%. For all parameter choices the use of an index improves the results with regard to raw wind speeds. The distribution of big claims agrees better with strong values of the wind index. However visually the three graphs are rather similar. This shows that the choice of the quantile has no strong influence on the quality of the results.

$\alpha$	$q_{95}$	$q_{97}$	$q_{98}$	$q_{99}$	$q_{99.5}$
1	0.349	0.358	0.362	0.362	0.375
2	0.545	0.581	0.600	0.600	0.526
3	0.597	0.622	0.633	<b>0.633</b>	0.569

Table VIII: Correlation between wind index (difference) and claim frequency

Tables VIII shows the different correlations between claims and the difference index. We try various upper quantiles (from 95% to 99.5%) as threshold of wind speed. We also test values between 1 and 3 for the exponent  $\alpha$ . The best link is obtained with  $q_{99}$  and  $\alpha = 3$  with a correlation over 63%. It is necessary to note that the Pearson correlation values are linked to the data scale. Consequently, the increase of the correlation when  $\alpha \geq 2$  could be partially due to the strong values taken by the index. We thus have only weak differences depending on quantile choices. We can just notice that the correlation increases when the exponent increases and that the most appropriate exponent seems to be  $\alpha = 3$ . For other variations no trend is clearly recognizable.

In the following, we compare the claims with a ratio between wind speed  $w$  and a quantile  $w_q$  (like in Formula 2):  $\left(\frac{w}{w_q}\right)^\alpha$ .

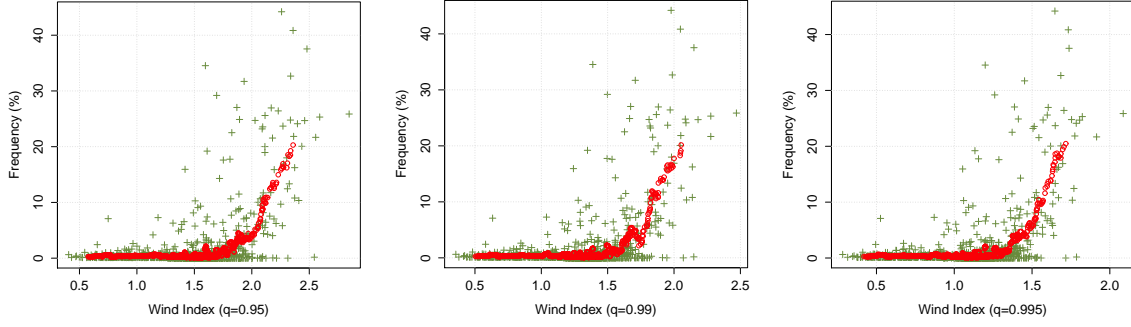


Figure 14: Correlation between claims and wind index (ratio)

Figure 14 shows the correlations between claims and the ratio index. As for wind speeds, the comparisons are made with claim number balanced by the number of contracts in each department. We observe the scatter plot (green dots) and the moving average (red circles) calculated between the frequency and the ratio index. From left to right, the quantiles raise from 95% to 99% and 99.5%. As previously the three graphs corresponding to three greatneses of quantiles are rather similar. The strongest correlation is near 50%.

$\alpha$	$q_{95}$	$q_{97}$	$q_{98}$	$q_{99}$	$q_{99.5}$
1	0.383	0.386	0.386	0.386	0.387
1.5	0.431	0.433	0.434	0.434	0.436
2	0.476	0.479	0.479	0.479	0.482
2.5	0.517	0.520	0.521	0.521	0.524
3	0.553	0.556	0.557	0.557	<b>0.561</b>

Table IX: Correlation between wind index (ratio) and claim frequency

Here also, correlations are better with the claim frequencies (between 38 and 56%) than with their values. In either case, the strongest relation is obtained by the Formula including a ratio associated with upper 99.5% quantile and an exponent  $\alpha = 3$ . This result is slightly different from that of Klawa et al. [14] which recommends a threshold at upper 98% quantile. Nevertheless, if we look all the values of the Table, changes between quantiles are not yet significant. These differences are not sufficient to determine clearly the best Formula, in particular considering the other sources of uncertainties. Furthermore, even if the correlations are higher with a cubic index than with the raw wind speeds, they remain around 50%.

We can conclude from these various tests that the only notable improvement comes from the consideration of the adaptation to the specific wind speeds in every department via the use of a quantile. The correlation seems also to increase with the exponent. Other adjustments and parameterizations remain marginal considering all the inaccuracies which accompany at the same time the meteorological measures and the insurance results.

### 7.3.5 Spearman's correlation and wind

The results presented so far are those of Pearson's classical correlation. We use here the same approach as previously but by comparing this time not anymore the values but the ranks of the various variables. The obtained correlation is called Spearman's correlation and graphs correspond to the empirical copulas. The empirical copula was defined in Deheuvels [4] as

$$C_n(u) = F_n (F_{n1}^{-1}(u_1), F_{n2}^{-1}(u_2))$$

, where  $F_n$  is the empirical joint cumulative distribution function of the  $n$ -sample and  $F_{nj}^{-1}$  is the marginal quantile function of the  $j$ th coordinate sample. The empirical copula is invariant under monotone increasing transformations of the marginals, so it depends on the data only through the ranks.

As previously, we begin with raw wind speeds. In each department we consider the average of weighted maximum wind speed (the most appropriate measure from previous Subsections). We compare the speed ranks with that of the claims expressed in loss ratio, average cost and frequency. The correlations presented in Table X, all the values are strongly negative (between -28 and -14%). So, if we only consider these Figures, the wind factor seems to influence the severity of storms for the insurers in a negative way, what seems illogical. We have to observe the empirical copulas to understand these negative correlations.

Events nb	loss ratio (2012 euros)	average cost (2012 euros)	frequency
1500	-0.196	-0.277	-0.149
150	-0.221	-0.134	-0.228

Table X: Spearman correlation with wind speed

In Figures 15, 16, 17 we show rank plots of a sub-dataset composed of couples whose at least one coordinate is in the 3% largest values. This explains that we have four quadrants in most of those figures: the bottom left-hand corner is usually less populated as it corresponds to small values for both coordinates. The two rectangles (bottom right and top left) contain points for which one coordinate is extreme but the other is not: they enable one somehow to quantify cases where we have complete mismatch between average max wind speed and the  $y$  coordinate. The top right-hand quadrant corresponds to points contained in the sub-dataset for which both coordinates are part of their the 0.3% largest values. This quadrant usually features some survival Clayton shape for loss ratio and claim frequency. With this approach, we both have a view on the probability to miss large losses with a low index (top left-hand rectangle) or to strongly overestimate the consequences of large wind speeds (bottom right-hand rectangle), as well as on the desired correlation for extreme values. Therefore, the correlations computed on this sub-dataset and presented in Tables X, XI, XII correspond to a summary of negative correlation coming from the two rectangles and reinforced by this somewhat unfavorable sub-dataset choice and of positive correlation of extremes exhibited in the top right-hand quadrant.

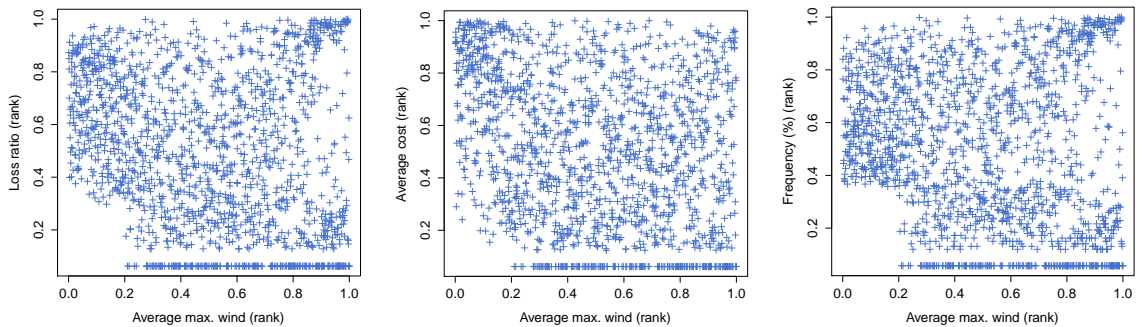


Figure 15: Copula between claims and wind speed

On Figure 15, we observe the empirical copulas between the weighed average of maximum wind speed and loss ratio, average cost and frequency of claims. For average costs (middle graph) the

point concentrations in the upper left and lower right parts of the square corresponds to the negative correlation that we have already met previously. On the rest of the square the distribution of points is rather uniform and it remains difficult to establish a link between average cost for the insurers and the wind speed.

For loss ratio and frequency (left and right graph respectively) the distribution of points is rather similar. The most marked zone in term of point density is located at the top right corner, what corresponds well to the most extreme events for both variables. However the rest of the square is also filled relatively uniformly. It is thus difficult at this step to identify a copula corresponding to this distribution.

Then, we compare claim ranks to those of a difference between the maximum wind speed  $w$  and several wind quantiles  $w_q$ . To obtain the values of the index, we use the formula  $\text{rank}(w - w_q)^\alpha$ .

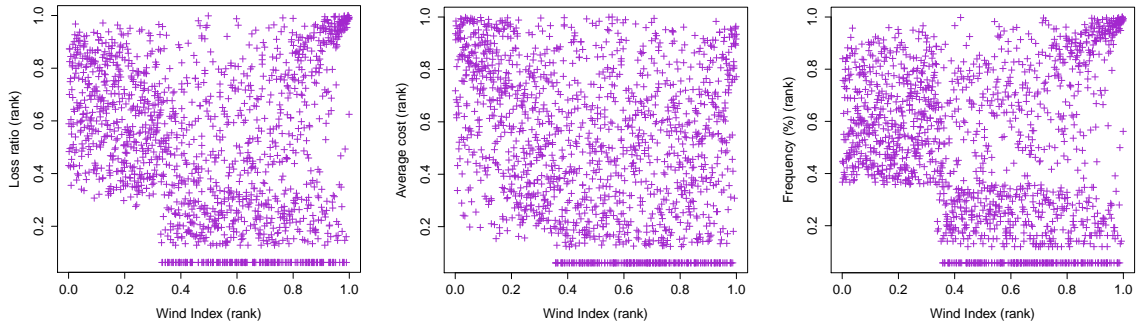


Figure 16: Copula between claims and wind index (difference)

Figure 16 presents the empirical copulas obtained by comparing the ranks of the index with those of the claims. We denote a better point concentration along the diagonal for the first one and the last graph corresponding to loss ratio and to frequency. The values from the top right corner (corresponding to the major claims) are very concentrated. Graphically, the shape of these scatter plot gets closer to functions of the survival Clayton copula or Gumbel copula. They are asymptotically dependent copulas, so this shape goes along with the notion of extreme dependence. Nevertheless, zones along x-axis and y-axis still present a significant density of points. This explains the negative values of the obtained correlations.

For the average cost the point distribution remains uniform on the main part of the square with a concentration on the top left corner. No relation is thus visible. We present the value of the corresponding Spearman's correlation in the next Table (XI).

Events nb	loss ratio (2012 euros)	average cost (2012 euros)	frequency
1500	-0.155	-0.264	-0.097
150	0.106	0.146	0.114

Table XI: Spearman correlation between wind index (difference) and claims (rank)

The values are still negative (between -9% and -26%). It would be necessary to concentrate only on the most extreme events to obtain again a positive correlation. This will be detailed in the next Subsection on tail dependance index.

We compare finally claim ranks with a ratio between wind speed and quantiles. The index formula become  $\text{rank}\left(\frac{w}{w_q}\right)^\alpha$ .

Figure 17 shows the empirical copulas between claims and ratio index. Visually three graphs have the same look as for difference index. The distribution is asymptotically dependant for loss ratio and frequency (first and last graphs). But a significant number of points is still situated along x-axis and y-axis.

Once again the copula of the average costs concentrates on the top-left corner but the distribution of points on the rest of the square is almost uniform.

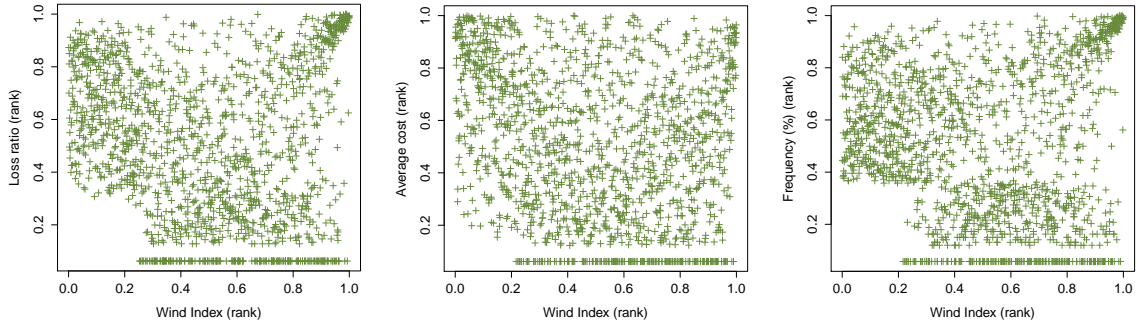


Figure 17: Copula between claims and wind index (ratio)

Events nb	loss ratio (2012 euros)	average cost (2012 euros)	frequency
1500	-0.082	-0.213	-0.013
150	0.350	0.346	0.368

Table XII: Spearman correlation between wind index (ratio) and claims (rank)

Table XII presents the Spearman's correlation between the ratio index and claims. These correlations are also negative but decrease when the quantiles increase contrary to the previous index based on a difference. However the value of the correlation has here either no direct meaning. It just reflects the malfunction of the index if we do not focus on the extreme values. With this approach we confirm the existence of a relation between claims and wind index without being able to choose in a categorical way the best parameters.

We see here the limits of the improvements brought by a Formula. For a better understanding of all the important claims from a meteorological index, other criteria are doubtless necessary (as the nature of the buildings and houses, the duration of wind gusts). The precision of the data can also be improved if we had direct sources like Météo France.

We obtain nevertheless rather satisfactory results and the gaps observed at a local level can cure with a global observation of the phenomenon. We shall see these improvements in the next section focused on main events thanks to the tail dependence index and from the section 7.1.5 dedicated to the storm index.

### 7.3.6 Tail dependence index

The standard way of measuring the correlation for the risk management is to zoom in on extremes by looking at the tail dependence index to the right. It is the limit when epsilon tends towards 0 of the conditional probability that the second coordinate is in the  $\epsilon$  % bigger values (of the second coordinates), given that first coordinate is in the  $\epsilon$  % bigger values (of the first coordinates). Mathematically, for two random variables  $X$  and  $Y$  with distribution function  $F_X$  and  $F_Y$ , the right tail dependence index is defined by:

$$\lambda_\epsilon = \lim_{\epsilon \searrow 1} (\mathbf{P}(Y > F_Y^{-1}(\epsilon) | F_X^{-1}(\epsilon))) \quad (6)$$

One advantage of this measure is that it is symmetric between the two variables. We look at the same time at the cases where we do not predict a storm with a strong wind index, and the cases where we have a storm while the wind index was low.

The theory says that there is some strong tail dependence on the right if the limit is strictly positive. Variables are asymptotically independent if the limit is zero. We thus estimate empirically the indices corresponding to various thresholds. We choose to observe the conditional probability that the loss ratio is in the  $\epsilon$  bigger values given that the index is also in the  $\epsilon$  bigger values.

Figure 18 shows the evolution of the dependence index for different numbers of peak exceedances going from 1000 to 1. The higher the threshold, the closer we get to extreme events. From left to



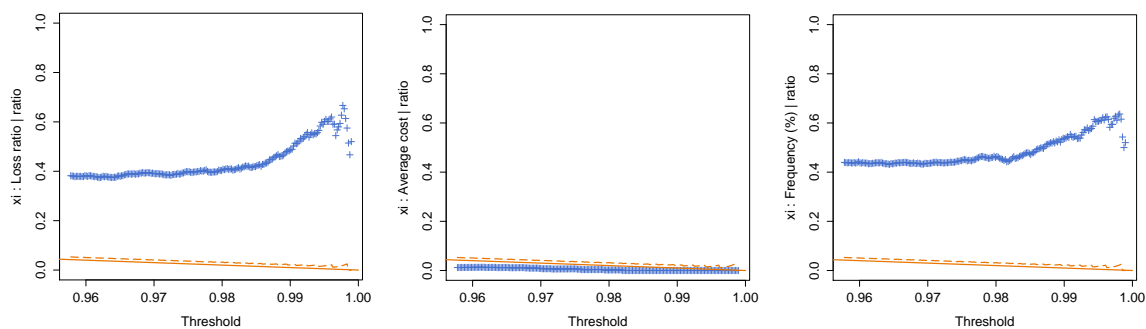


Figure 18: Empirical tail dependence index for loss ratio, average cost, frequency, and wind index ratio

right we can see the dependence between the frequency, the average cost, the total cost and the ratio wind index. The full line represents the asymptotic independence with in dotted lines a confidence interval. Points situated over the dotted lines denote the existence of extreme dependence between claims and wind index. These graphs (we look at the limit of the blue curve when we aim towards 1, because the threshold depends on  $1 - \epsilon$ , what is standard) confirm the quality of the results for the frequency and the total cost (tail dependence index between 0,35 and 0,4). It also confirms that the average cost is not correlated in extremes to the indication wind index (it was a priori only few or not globally correlated thus it is logical) because the limit is 0 for the curve.

It is interesting to look also at the conditional probability that the second coordinate is in the  $k.\epsilon$  % bigger values (of the second coordinates), given that the first coordinate is in the  $\epsilon$  % bigger values (of the first coordinates), and conversely. It allows us to extend the perspective and better understand the variation of the index. Here we observe the evolution of the conditional probability that the frequency is in the  $k.\epsilon$  % bigger frequencies, given that the index is among the  $\epsilon$  % bigger index. We choose  $k \in \{2, 3, 5\}$  and  $\epsilon$  in a way that  $\epsilon$  % corresponds to 1000 points.

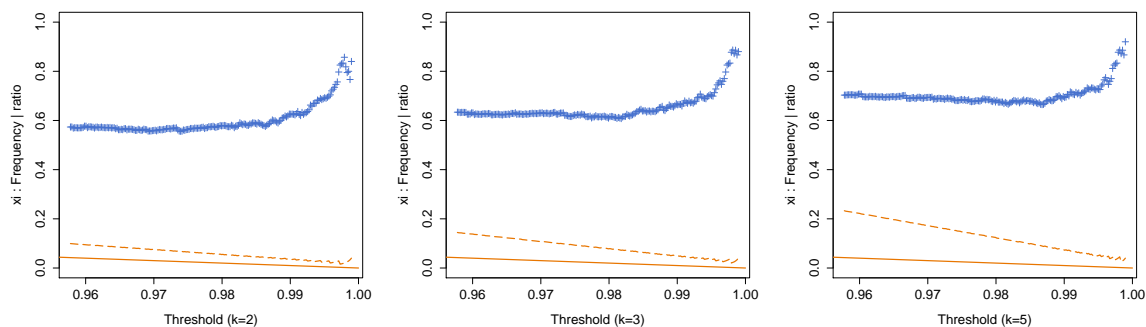


Figure 19: Empirical tail dependence index for claims frequency and wind index ratio for  $k \in \{2, 3, 5\}$  (from left to right)

The analysis carried out in 7.3.5 could give the impression that correlation is quite low for the largest values. Figure 18 shows that strong correlation of extremes is really present for loss ratio and claim frequency, and completely absent for average cost. We see that the conditional probability to exceed  $(1 - \epsilon)$ -VaR for one variable given that the other one exceeds its  $(1 - \epsilon)$ -VaR increases in  $(1 - \epsilon)$  when  $1 - \epsilon \geq 98\%$ . The right tail dependence index is the limit and is approximately 60%, which is quite high. Figures 15, 16, 17 exhibited many points in the bottom right-hand and top left-hand rectangles, which could give the impression that when one coordinate is extreme, the other one might be extreme or anything. Figure 19 shows that the conditional probability to exceed 97.5%-VaR for one variable given that the other one exceeds its 99.5%-VaR is approximately 80%.

Overall, even if one cannot avoid false alarms or failure to represent a storm with the index, the desirable property of strong correlation of extremes between the index and frequency (or loss ratio) is satisfied, with high values of right strong tail dependence index. This shows that our index seems suitable for risk management purposes, where extremes matter.

## 7.4 Relation between wind and claims at event scale

### 7.4.1 On the basis of the events from 1998 to 2012

Here the comparison concerns only the significant events by taking into account all the stations affected by the same storm. We work with the sums of the local claim values and numbers which are used to match the storm index  $I_S$  defined in the previous section (4).

Concerning the damages, we work with the normalized **total cost** and with the **number** of claims. We want to focus on the main storms. We choose here to retain only the 200 main events. So, we lay the emphasis on the days over 150,000 euros for costs and 60 for daily claim numbers. The major events therefore have no risk of being eliminated with such a threshold.

Concerning the wind, we use the **weighted average of maximum speeds** in each department. The results which follow are obtained by building this index from the upper 99% quantile and our Formula of the wind index (1). In order to select the main events, we have chosen a threshold of 20 for the index (between 0 and 100).

We do not consider anymore the days taken one by one but groups of days corresponding to the storm events. The selection of the events was made by **associating the days when the index or the claims exceed their respective thresholds**. The grouping of the days is done on a case-by-case basis by spotting the highest values over a period from 1 to 4 days around an event. On Figure 20 we can observe the repartition of the 200 main events.

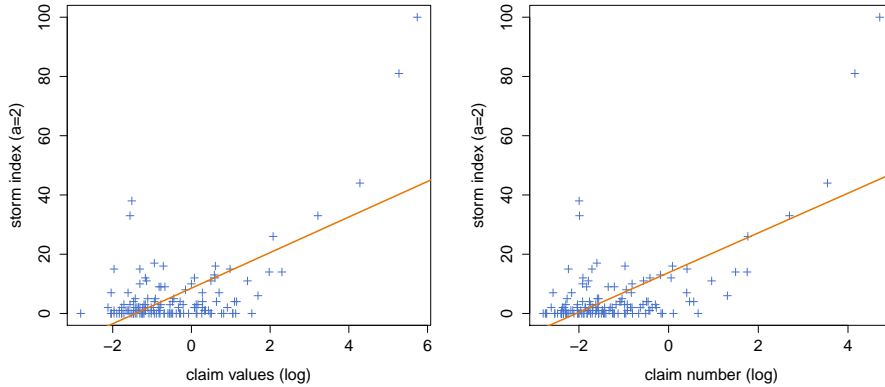


Figure 20: Correlation between claims and storm index during the 200 main events

We calculate a storm index agglomerated on each event. Two approaches are considered. The first one consists of keeping only the maximum value of the index for each event and the second one uses the sum of the index values for each event.

Values	Total cost	Number
Maximum	0.708	0.734
Sum	0.712	0.745

Table XIII: Correlation with index maximum and sum - Events

This method (Table XIII) gives better correlations than those of the daily approach. We obtain the highest values by compiling the storm indices of each episode. This approach by episodes corresponds more to the economic reality than the daily approach. It also allows us to reflect more clearly a linear trend between the major storms and the highest indices. For the less important events we can see graphically that the relationship with the index is not as good. But our study focuses on the major events which are at the origin of the most important damages.

For example, we can estimate fairly accurately the numbers of claims ( $N$ ) bred by Lothar, Martin, Klaus and Xynthia on the Allianz portfolios via the relationship:

$$N = 1.2 * I_S - 28.6$$

On Figure 21, we have shown in an orange dotted line the linear regression on the whole set of values and in red on the 4 main events. We shall see later on that this relationship can still be improved.

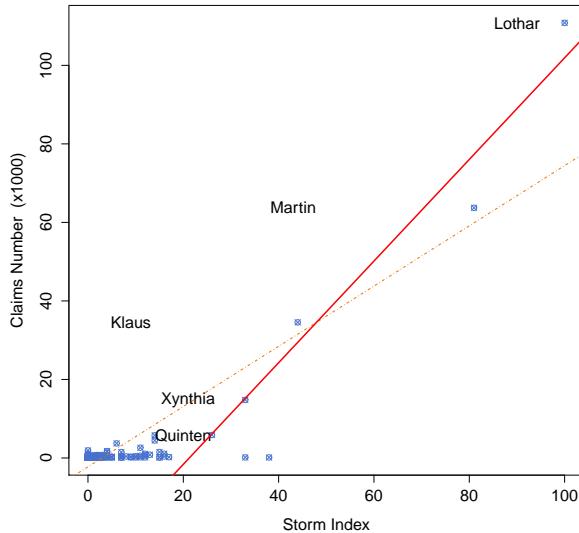


Figure 21: Linear regression between the number of damages and the Storm index

The relation between claims and storms cannot be controlled from a single formula. There are several slices of events among which the causes and the consequences are very different. For our part only the high slice constituted by about twenty major events has a real importance on the variability of the insurer results. It will thus be important to define a threshold from which the number of claims must follow a law depending on the storm index. From a statistical perspective, the threshold is loosely defined such that the population tail can be well approximated by an extreme value model, obtaining a balance between the bias due to the asymptotic tail approximation and parameter estimation uncertainty [21]. The decisions concerning the choice of the threshold and the number of events which we hold can be made according to various methods. Graphical diagnostics like the mean residual life plot, the Hill estimator, the threshold stability plot or usual distribution fit diagnostics, are the most used. But other solutions are also available like the rules of thumb (square root rule or empirical driven rule), the computational approaches (resampling method for estimation of the optimal tale fraction (Hall) or bootstrap procedure for tail index estimation) or the mixture models (parametric, semiparametric and nonparametric bulk models).

#### 7.4.2 Spearman's Correlation and storm

With Pearson's correlation the strong values have more weight than the other ones. For the insurer the major events are also the main source of damage. This bias is therefore not necessarily detrimental to our model but it can be avoided by using a correlation based on rankings and independent of the spread of values. We present here the approach by episodes with the Spearman's correlation (Table XIV) as an element of comparison.

Total cost (rank)	Number (rank)
0.407	0.396

Table XIV: Spearman's rank Correlation with index sum - Events

The weakest correlations obtained with this method are mainly due to the events of small size (both in term of costs and index) which are the most common. We nevertheless obtain a Spearman correlation around 40% which implies a link between the rankings of storms according to the index and

the insurers. Figure 22 shows empirical copula. On the 2 graphs, only the values from the top right corner (corresponding to the major events) are concentrated. As in the case of wind index, the shape of these scatter plot gets closer to asymptotically dependent copula (like survival Clayton copula or Gumbel copula).

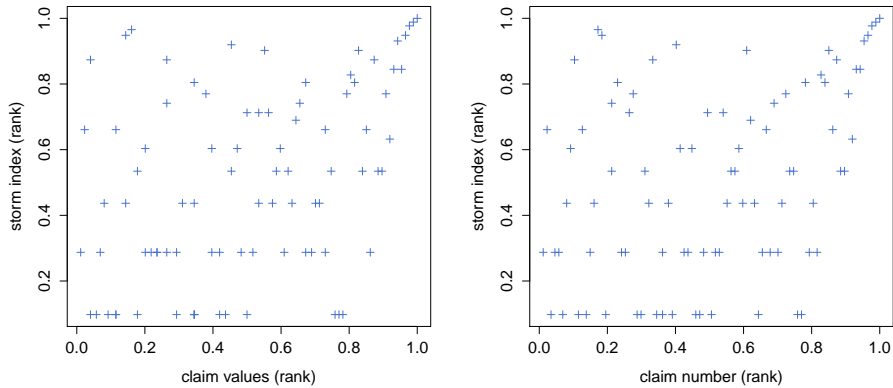


Figure 22: Spearman's rank Correlation between normalized claims and storm index in France by event

### 7.4.3 On the basis of the events from 1970 to 2012

We take into consideration the list of major storms prepared by Luzi [15] that we now confront with different formulae of storm index. As we have seen previously the results are very sensitive to the methods used. We have therefore opted for a comparison based on both the values and on the rankings of major storms in the form of summary tables associated with a graph of empirical copula. We tested successively different weights, different quantiles and different approaches for the wind index to obtain the strongest relationship between storms and index.

For the weighting of the wind index, we first try demographic information like the population of each department or the density. Then we use Allianz' portfolio information like the number of contracts in the individual and in the global lines. For the quantiles, we notice that small values give unsatisfactory results and we decide to test different values between 95% and 99.5%. The wind index Formulae presented in Subsection 6.1 are again optimized according to the events on a national scale. The presented results are shown here by a method of optimization of the absolute differences between the index and the normalized costs. The 1998-2012 period is used to calibrate the parameters values. Then the Formula is tested on the whole period (1973-2012).

Optimisation method on the main Storm costs (1998-2012):

```
> val <- log(sum(abs(costs - C * S.Index)))
> fit <- optim(par=...)
```

### 7.4.4 Population Weighting

We started the comparisons with the index obtained via a weighing by the population of each department. In the following table, we have only kept the most important 20 storms according to the index  $I_S$  in the descending order. For each of these storms, the classification of insurers appears in the third column. If the storm is not part of the 20 most important ones in term of cost of insurance, we denote >20 in the place of the classification. The graph associated with each table allows the comparison of the values of the storm index with the normalized values of their cost.

With these weights, storm Lothar stands out clearly with a value of nearly twice greater than Martin which appears second. But the rest of the results is much less conclusive with values and a classification very different from that of insurance.

### 7.4.5 Individual Portfolio weighing

Storm	date	rank (IG)	$I_S$	IG
Lothar	12/25/1999	1	100	100
Martin	12/27/1999	2	58	51
Herta	02/03/1990	5	58	19
	11/22/1984	>20	50	1
Vivian	02/26/1990	11	40	7
	12/17/1986	>20	40	1
	02/07/1984	>20	35	1
Nov 82	11/07/1982	3	33	25
Xynthia	02/27/2010	10	33	8
	12/08/1991	>20	32	1
	02/03/2002	>20	28	1
	03/24/1988	>20	26	1
Daria	01/25/1990	9	25	9
	11/26/1983	>20	25	1
Wiebke	03/28/1990	>20	24	1
	01/22/1988	>20	21	1
Klaus	01/23/2009	4	21	20
	02/10/1974	>20	21	1
Quinten	02/09/2009	>20	21	1
87J	10/15/1987	6	20	10

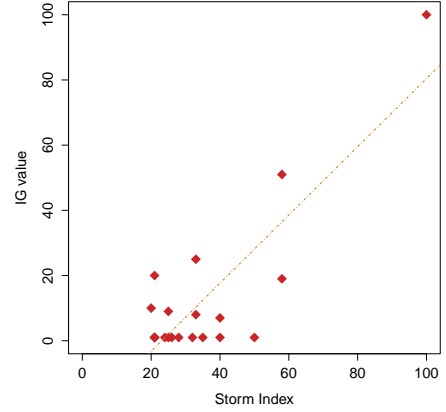


Figure 23: Storm classification with population weights

To improve the classification of indices and to show the impact of the storms for insurers, we now take into consideration the portfolio Allianz. We then incorporate the number of insured people  $A(k)$  in each department to the weights of the index. At this step, the index formula is

$$I_{S_2} = \sum_{j \in J} \sum_k \frac{I_v^j(k) * A(k)}{N_a^j}. \quad (7)$$

Storm	date	rank(IG)	$I_{S_2}$	IG
Lothar	12/25/1999	1	100	100
Martin	12/27/1999	2	83	51
Herta	02/03/1990	5	60	19
	11/22/1984	>20	60	1
Nov 82	11/07/1982	3	48	25
	12/17/1986	>20	48	1
Vivian	02/26/1990	11	44	7
	12/08/1991	>20	40	1
	02/07/1984	>20	39	1
Xynthia	02/27/2010	10	37	8
Klaus	01/23/2009	4	33	20
	03/24/1988	>20	29	1
Daria	01/25/1990	9	29	9
87J	10/15/1987	6	28	10
	11/26/1983	>20	28	1
	02/03/2002	>20	27	1
Wiebke	03/28/1990	>20	24	1
Quinten	02/09/2009	>20	23	1
	01/22/1988	>20	23	1
	02/10/1974	>20	22	1

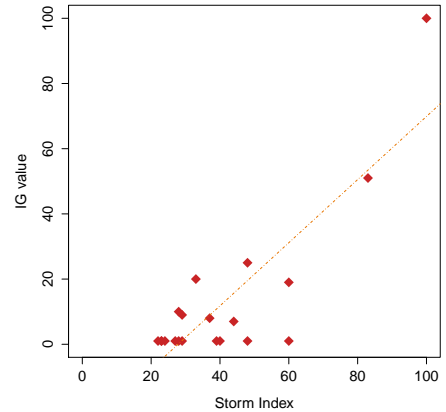


Figure 24: Storm classification with portfolio weights

With these new weights, storm Martin stands out of the rest of the values to be closer to Lothar. In the classification Klaus goes from the 17th to the 11th position before Daria and Wiebke, but stays nevertheless behind Xynthia. By focusing on the values that constitute the index during major events, we realize that wind speeds above 100 kph entail significant overruns and strongly influence the value of the index, which depends on their squares. The choice of stations used to represent each department therefore plays an important role. Indeed, it is possible to overestimate or underestimate the index of a storm if the measurement of the selected station shows a significant gap in relation to the other stations in that same department. We would then favour a non

representative station. The sub index assessment of Klaus could be due to the position of stations in relation to the trajectory of the storm. For Lothar for example this speed of 100 kph is exceeded 72 times on 83 active stations for Martin: 57/82, for Xynthia: 60/87 and for Klaus only 41/89. To verify this hypothesis, a denser meshing of stations during major events will provide more details.

#### 7.4.6 Individual Portfolio weighing (130 stations)

On the NCDC’s website the readings of 130 stations located throughout the hexagon have been collected. So the accuracy of wind speeds is improved. If we consider Klaus again, focusing on the Gironde, we now have not only Bordeaux but two other readings at Cap Ferret and Cazaux. On January 23, 2009, during the passing of the storm on the department the bordelaise station showed a maximum speed of 172.8 kph. With the other two stations we learn that the maximum speed in Cazaux was of 140.4 kph but especially that an exceptionally high reading has been recorded in the station of Cap Ferret with 251.6 kph. This important difference confirms the usefulness of a database giving the widest possible scope in particular during extreme events. We then calculate a new index  $I_{S_3}$  based on the 130 stations.

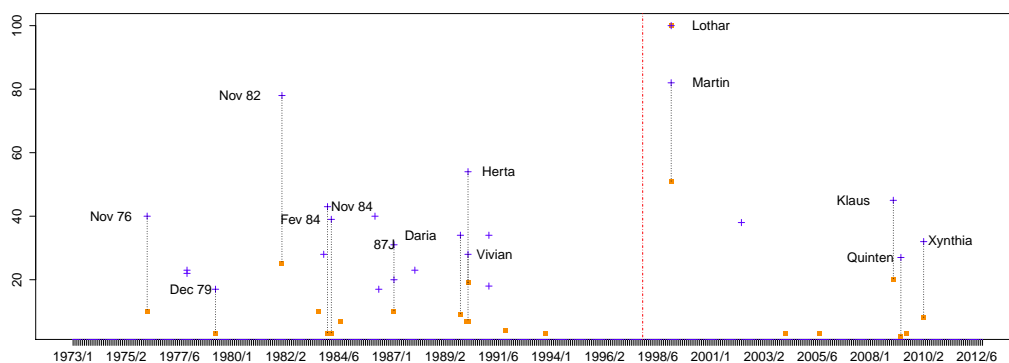


Figure 25: Storm index with 130 stations between 1973 and 2012

Storm	date	rank(IG)	$I_{S_3}$	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
	01/22/1988	>20	23	1
	11/01/1978	>20	22	1
	12/08/1991	>20	20	1

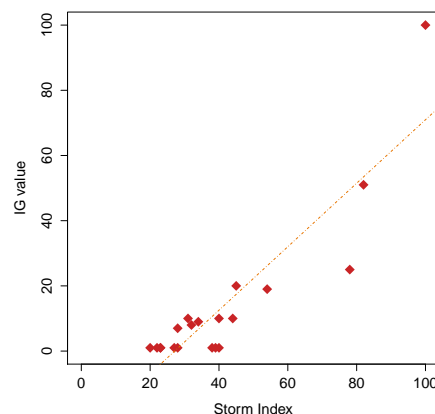


Figure 26: Storm classification with portfolio weights

With this new information, Lothar remains the main storm, still followed by Martin. The storm of early November 1982 comes inserted before Herta and thus appears to the ranking of major events. This storm has swept across western Europe causing catastrophic destruction in some thirty departments and many victims according to Météo France that ranked it among the most violent of these

past thirty years. Furthermore the storm Klaus comes now in 5th position and shows especially an index much higher than that of Xynthia according to the insurance results . The storms of February 1990 will also be recorded as well known events of our period of observation. Daria overpasses Xynthia and Vivian.

#### 7.4.7 Global Portfolio weighing (130 stations)

In this section, we have used the whole portfolio (both for individuals and professionals) to weigh the storm index. At this step, the formula reference is  $I_{S_4}$ . With this overall weighing, the results are substantially the same as for that of the individual's portfolio.

This comparison shows that it is useful to use more stations. With  $I_S$  and  $I_{S_2}$  certain events such as storms of November 82 or Klaus are under-estimated and some among the most costly were not even as that of November 76. For  $I_{S_3}$  and  $I_{S_4}$ , the most complete updating (**Act IG**) offers the most fairly estimated points. The 5 main events of the period are the same with an inversion between Klaus and Herta but relative because the gap is very low in the two approaches. The classification is then more difficult to compare but the order of arrival of Daria, Xynthia and Vivian remains the same.

#### 7.4.8 Portfolio weighing : Cubic Index

We perform here the same comparison but with the cubic index  $CI_S$  (2) proposed by Klawa et al. This index showed itself slightly more effective than ours during the local approach (department by department) but from now on, it is not the case anymore. In Table, storms are less well classified than previously. In particular, two minor events for the insurers, the storms of December 86 and January 78 are inserted between Lothar and Martin. During the 1986 storm we notice two stations with wind speeds over 175 kms/h but few damaged departments in Allianz' historical data. This storm was too local to cause important damages. We thus have poorer results with this index at a global level.

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
	10/10/1987	>20	12	1

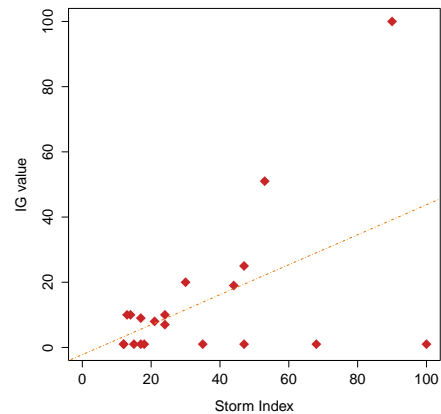


Figure 27: Storm classification with cubic index and portofolio weights

When the wind index exponent  $\alpha$  is greater or equal to 3, the order of the storm index values is very perturbed. A possible explanation is that localized storms with high wind speed are overestimated.

From the meteorological point of view it is surely possible to optimize the parameters of the storm index to obtain a better representativeness of observed events. Not all parameters are known, the duration of gusts of wind for example could refine some values. However, it must be borne in mind that, given the complexity of the phenomena at stake, a totally indisputable model will be difficult

to obtain.

From the economic point of view it is possible to improve both the accuracy and the length of this type of historical background based on more comprehensive data from the market. The results of the reinsurance could for example throw another light on this subject.

The obtained hierarchy between the major storms of the period is satisfactory. The gaps and the orders of magnitude can be improved. We propose to modify the index so as to amplify the differences. The idea here is to take the exponential of the sum of the indices of wind recorded during a storm episode in order to highlight the extra cost which seems to go with the most destructive events. In addition, it would be interesting to be able to change the weight given to the number of stations because an event can be overestimated if it takes place in an area well represented by the stations while their total number remains relatively low.

#### 7.4.9 New parameter

At this stage, we have added a new parameter in order to check the scale of storms. This adaptation of the formula may seem specific to our database but in our opinion it responds to a more general problem. We can indeed observe a gap between the most important costs and the rest of the values. 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. [16]. In their study, the worst hurricane recorded between 1900 and 2005 is the one of Miami in 1926 and its value is twice as large as the second hurricane of the ranking. Our new formula associates the exponential function with a new parameter  $\beta$ . The new index formula becomes

$$I_{S_5} = \exp \left( \beta \sum_{j \in J} \sum_k \frac{I_w^j(k) * G(k)}{N_a^j} \right), \quad (8)$$

where in each station  $I_w^j(k)$  is the wind index on date  $j$  and the **number of risks** is balanced by the size of Allianz global portofolio  $G(k)$ . We also take into account the **size** of the damaged area (geographic aggregation), the **duration** of the storm  $J$  (daily aggregation) and the number of active stations on date  $j$ ,  $N_a^j$ .

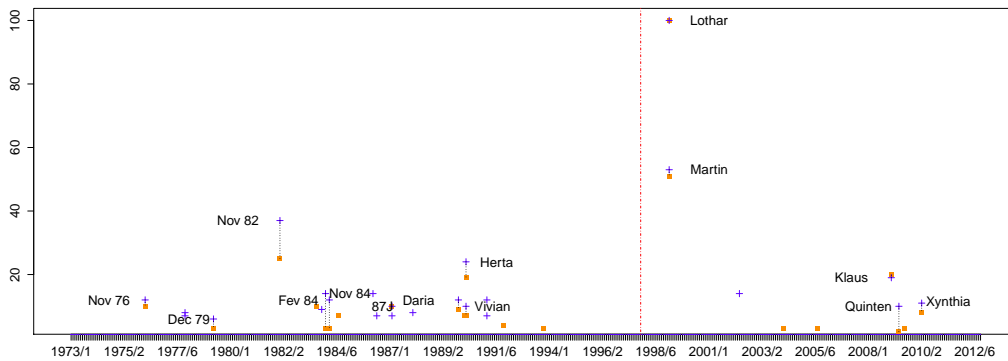


Figure 28: Storm index with 130 stations between 1973 and 2012

With this latter result, we find a much more important parallel with the insurance data from the insurance. Storm Lothar now clearly appears with an index almost twice as large as that of Martin. The following differences also correspond to the discrepancies of the normalized costs. Graphically, we obtain a much better alignment than previously for the costs, the frequencies and number of claims. The points corresponding to the major storms are closer to the linear regression.

The relationship between this new index and the loss experience of the 4 major events of the 1998-2012 period becomes even more direct with a slope equal to 1:

$$N = I_{S_5} + 9.1.$$



Storm	date	rank(IG)	$I_{S_5}$	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
	12/08/1991	>20	5	1

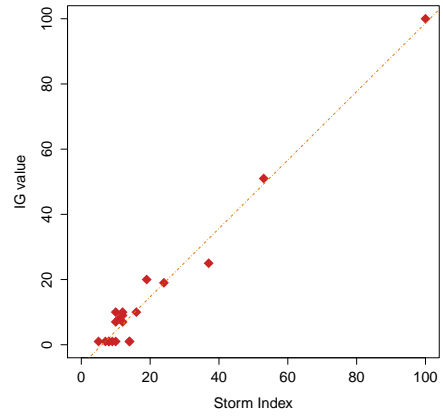


Figure 29: Storm classification with portfolio weights and  $\beta$  parameter

On Figure 21, we have shown in an orange dotted line the linear regression between the set of values and in red between the 4 main events. The very strong correlations and this linear relationship allow us to support the relevance of our storm index.

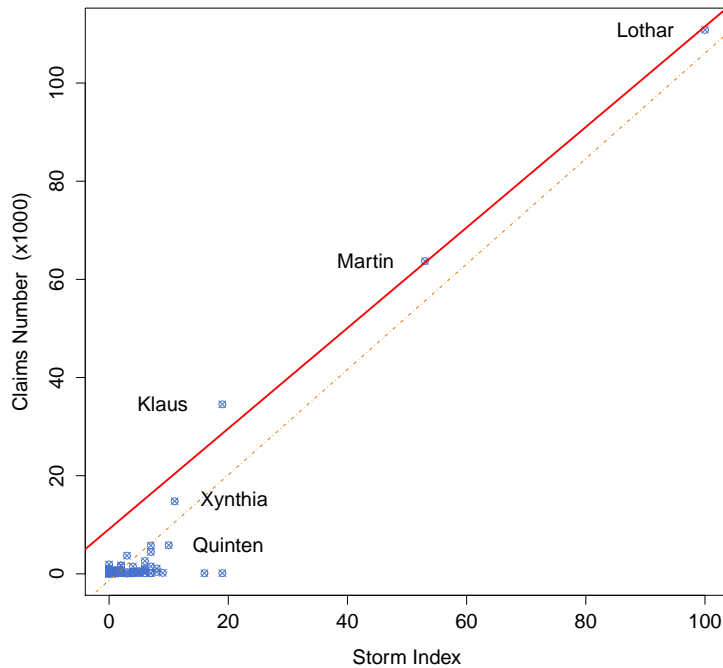


Figure 30: Linear regression between the number of damages and the Storm index

## 7.5 Wrap up

Thanks to these comparisons, we have shown how parameters and variables can influence the quality of a wind index and a storm index. Concerning the claim criteria (Subsection 7.3.1), it can be stated that the approach according to the average costs is doomed to failure (Nevertheless, it seems to be the basis of all the storm software). The frequency of claims offers the strongest relation with wind speed. The use of other claim criteria than frequency (loss ratio ; individual / global lines) does not lead to fundamentally different results.

In our search for functions of damage at the level of the department, we have highlighted that it is better to use a wind criterion based on a weighted average over each department of the maximum daily speeds (Subsection 7.3.2). This stage requires a real work of data processing to select stations with regard to departments. In Subsection 7.3.3, we introduce the concept of moving average to clarify the plots and focus on the main events.

Then we show how the modulation according to the local wind habituation brings a sensitive improvement to the correlations between wind and claims (Subsection 7.3.4). We also find that the use of the Formula (difference (1)/ ratio (2); thresholds) does not lead to fundamentally different results and we denote that the relation is not linear, or then, it is necessary to pass by segments.

In Subsection 7.3.5, we complete the first comparisons based on Pearson's correlation by using the Spearman's correlation based on variable ranks. We notice that it is very difficult to obtain a robust relation between the wind index and the claims, for the small and medium events. We have to focus on the main events to observe a clear dependance. So we decide to calculate the tail dependance index (Subsection 7.3.6) and we obtain good results which prove that there exist a strong extreme dependance.

Finally, we present the optimization of the storm index at a global level (Subsection 7.4.1 to 7.4.9). The successive models are based on the period 1998-2012 and we try to track down the main events registered since the 70s. The Formula (4) overestimates small events (very localized but with strong wind gust) when it is associated with the cubic index. So, at a national scale, our wind index allow us to improve the relation between claims and storm index. By using a new parameter and the exponential function, we finally obtain the last storm index Formula (8). This Formula reflects both the order and the scale of the main storms that made landfall in France since the 70s.

## 8 Conclusions

It is possible to construct a storm index consistent with the results of insurers, from the daily maximum speed of the wind. Several approaches and formulae have been necessary to find almost the same classification and the same differences as that of insurers which depend on the frequencies and the normalized claims in the time. We divided our approach into two stages. At first local comparisons on the scale of the department which showed the necessity of taking into account the wind habituation but also the difficulty in finding the best parameters. Then comparisons on a national scale and by event. In this case, the parameters and departure assumptions have a strong influence on the value of the index and then on its evolution in time. The different results from one model to another show this sensitivity in particular during the evaluation of the most extreme storms.

Our index has been built by successive steps based on the optimal use of meteorological data at our disposal on the one hand and the experience and advice of the insurers in this area [15] on the other hand. Several leads could make it possible to improve the accuracy of this index, such as the knowledge of the durations of the wind gusts, a denser meshing of meteorological stations available accompanied by a better information on the location of claims. Nevertheless from the results of this research the use of a storm index brings a additional and useful information in the understanding of extreme events.

Our target is to study in a more precise way the sensibility of costs and return periods of a major event on a scale comparable or superior to that of Lothar in France. This subject will be dealt with in an article to come with a modeling directly based on the values of the storm index according to extreme value theory ([1] and [19]). The variability of the results to the assumptions adopted will be tested within the framework of the European statutory reform of the insurance, Solvency II.

## References

- [1] E. Brodin and H. Rootzén. Univariate and bivariate GPD methods for predicting extreme wind storm losses. *Insurance Math. Econom.*, 44(3):345–356, 2009.
- [2] A.M. Chandler, E.J.W. Jones, and M.H. Patel. Property loss estimation for wind and earthquake perils. *Risk Analysis*, 21(2):235–249, 2001.
- [3] S. Coles. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2001.
- [4] P. Deheuvels. La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d’indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, 65(6):274–292, 1979.
- [5] P.M. Della-Marta, H. Mathis, C. Frei, M.A. Liniger, J. Kleinn, and C. Appenzeller. The return period of wind storms over europe. *Int. J. Climatol*, 29:437–459, 2009.
- [6] M.G. Donat, T. Pardowitz, G.C. Leckebusch, U. Ulbrich, and O. Burghoff. High resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over germany. *Natural Hazards and Earth System Sciences*, 11(10):2821–2833, 2011.
- [7] C. Dorland, R.S.J. Tol, and J.P. Palutikof. Vulnerability of the netherlands and northwest europe to storm damage under climate change. *Climatic Change*, 43(3):513–535, 1999.
- [8] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.
- [9] Association française de l’assurance. Tempêtes, grêle et neige : Résultats 2010. Technical report, 2012.
- [10] M.J. Heaton, M. Katzfuss, S. Ramachandar, K. Pedings, E. Gilleland, E. Mannshardt-Shamseldin, and R.L. Smith. Spatio-temporal models for large-scale indicators of extreme weather. *Environmetrics*, 22(3):294–303, 2011.
- [11] P. Heneka, T. Hofherr, B. Ruck, and C. Kottmeier. Winter storm risk of residential structures model development and application to the german state of baden-wuerttemberg. *Nat. Hazards Earth Syst. Sci.*, 6:721–733, 2006.
- [12] S. Hochrainer-Stigler and G. Pflug. Risk management against extremes in a changing environment. *Environmetrics*, 23(8):663–672, 2012.
- [13] R.W. Katz and B.G. Brown. Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, 21:289–302, 1992.
- [14] 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.
- [15] M. Luzi. Travaux sur les historiques tempêtes. *Directeur IARD - Allianz*, 2013.
- [16] R.A.Jr. Pielke, J. Gratz, C.W. Landsea, D. Collins, M.A. Saunders, and R. Musulin. Normalized hurricane damage in the united states: 1900-2005. *Natural Hazards Review*, 9(1):29–42, 2008.
- [17] J.G. Pinto, E.L. Fröhlich, G.C. Leckebusch, and U. Ulbrich. Changing european storm loss potentials under modified climate conditions according to ensemble simulations of the echam5/mpiom1 gcm. *Natural Hazards and Earth System Science*, 7:165–175, 2007.
- [18] F. Prettenhaler, H. Albrecher, J. Köberl, and D. Kortschak. Risk and insurability of storm damages to residential buildings in austria. *The Geneva Papers*, 37:340–364, 2012.
- [19] H. Rootzén and N. Tajvidi. Can losses caused by wind storms be predicted from meteorological observations? *Scand. Actuar. J.*, (2):162–175, 2001.
- [20] C. Sacré, J.M. Moisselin, M. Sabre, J.P. Flori, and B. Dubuisson. A new statistical approach to extreme wind speeds in france. *J. Wind Eng. Ind. Aerodyn.*, 95:1415–1423, 2007.
- [21] C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *Revstats*, 10(1):33–60, 2012.
- [22] I. Scheel, E. Ferkingstad, A. Frigessi, O. Haug, M. Hinnerichsen, and E. Meze-Hausken. A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, 62(1):85–100, 2013.

- [23] S. Zahran, D. Tavani, and S. Weiler. Daily variation in natural disaster casualties: Information flows, safety, and opportunity costs in tornado versus hurricane strikes. *Risk Analysis*, 33(7):1265–1280, 2013.