Université Claude Bernard – Lyon 1

INSTITUT DE SCIENCE FINANCIERE ET D'ASSURANCES

Mémoire présenté devant l'Institut de Science Financière et d'Assurances pour l'obtention du diplôme d'Actuaire de l'Université de Lyon

le ______________________

Par : Florent Cannelle

Titre: European windstorms, an analysis of the clustering effect

Confidentialité : ☑ NON ☐ OUI (Durée : ☐ 1 an ☐ 2 ans)

Membre du jury de l'Institut des Actuaires

Membres du jury I.S.F.A.

M. Jean Claude AUGROS
M. Alexis BIENVENÜE
Mme Diana DOROBANTU
Mme Anne EYRAUD-LOISEL
M. Jean-Paul LAURENT
M. Nicolas LEBOISNE
M. Stéphane LOISEL
Mlle Esterina MASIELLO
Mme Véronique MAUME-DESCHAMPS
M. Frédéric PLANCHET
M. François QUITTARD-PINON
Mme Béatrice REY-FOURNIER
M. Didier RULLIERE

Secrétariat

Mme Marie-Claude MOUCHON

Bibliothèque :

Mme Michèle SONNIER

Entreprise :

New Reinsurance Company
Zollikerstrasse 226, Zollikon, Schweiz

Directeur de mémoire en entreprise :
Florent Sennelier & Dr. Artur Klinger

Invité :

Autorisation de mise en ligne sur un site de diffusion de documents actuariels *(après expiration de l'éventuel délai de confidentialité)*

Signature du responsable entreprise

Signature du candidat

50 Avenue Tony Garnier
69366 Lyon Cedex 07

Résumé

Mots clés : réassurance, catastrophe naturelle, tempête, amas, processus de Poisson, modèle collectif, loi négative binomiale, RMS

Dans le domaine de l'assurance de tempêtes, la plupart des modèles de référence utilisent la loi de Poisson pour décrire le nombre annuel d'évènements. La faiblesse de cette loi réside dans le fait qu'elle considère l'arrivée de tempêtes de façon indépendante les unes des autres. Ce problème est souligné par certains experts qui ont montré que Lothar et Martin, épisodes catastrophiques de 1999, ont été formés par la même situation atmosphérique.

Ce mémoire a pour objet de proposer deux modèles alternatifs qui prennent en compte la formation de tempêtes en amas. Nous parvenons à introduire la distribution négative binomiale ainsi qu'un modèle collectif de type Poisson-binomial. Le test de ces deux modèles sur des données à échelle européenne permet de valider leur grande efficacité comparativement au modèle couramment employé.

Deux applications de tarification sont ensuite suggérées. La première propose d'appliquer une majoration de prime uniforme lorsque cette dernière est calculée par un modèle de référence prenant en compte uniquement l'hypothèse de Poisson. La seconde permet d'ajuster cette majoration spécifiquement au portefeuille d'un assureur. Il est important de noter que seul le modèle collectif est capable de faire cet ajustement de niveau régional.

Abstract

Key words: reinsurance, natural catastrophe, windstorm, cluster, Poisson process, collective risk model, negative binomial distribution, RMS

In windstorm insurance, most reference models use the Poisson distribution to describe the number of annual events. This distribution has a drawback: it considers storms to occur independently from each other. This problem has already been emphasized by experts who explained that Lothar and Martin, the most severe storms of year 1999, were formed by a same atmospheric situation.

This dissertation aims to define two alternative models that take the effect of cluster into account. The Negative Binomial distribution and a Poisson-binomial collective risk model are developed for this purpose. Testing these models with European wide data verifies their validity in comparison to the current model.

Two pricing applications are then suggested. The first one proposes to apply a uniform adjustment factor to the premium computed from a reference pricing tool; this corrects the absence of cluster assumption. The second one describes how to adjust this rate specifically to an insurer's portfolio. Nonetheless, only the collective risk model can be used for making this local adjustment.

Acknowledgement

First of all, I would like to express my gratitude to Mr. Florent Sennelier and Dr. Artur Klinger, head of New Re's Analytics Pricing and Structuring department.
Florent spent much of his time to explain the issues of natural catastrophes to me. His critical view relying on a strong professional experience contributed a lot to the quality of this dissertation.
The topic was proposed by Dr. Artur Klinger, who provided valuable inputs not only in the technical aspects but also the research directions.

In addition, I would like to thank two colleagues, Ms. Anne-Lise Caillat and Ms. Angela Wünsche. Although their work areas are not directly related to my research, their frequent assistance and guidance helped tremendously in this dissertation.

Finally, I would like to give special thanks to Mr. Stéphane Bonche for this opportunity to do my internship at New Re.

Contents

Introduction

Natural hazards modeling is one of the hottest reinsurance issue especially since 2005 Katrina. This hurricane is the costliest single event in the whole insurance history with insured losses of 65 billion US dollars. It deeply impacted insurers in the US and worldwide. Katrina is one example among many others that show there is a growing need for natural catastrophe modeling. However, the main models were unable to accurately estimate the loss for many reasons. This has emphasized the need for insurance professionals to better understand the strength and the weakness of the model.

Similarly, European reinsurers must deal with windstorm models. One of the most used, RMS' Risklink, proposes to describe annual storms occurrence as Poisson distributed. This assumption has a significant disadvantage; it considers the probability of storms independently from each other. Nonetheless, Europe was hit in 1990 and 1999 by series of storms which all occurred in a "few days" time frame. Meteorologists considered that some of them were created from the same specific meteorological conditions, in other words, the same cluster. Such an argument is enough to question the classic Poisson model.

The aim of this dissertation is to set two models that will take the cluster factor into consideration. To achieve this goal, the subject will be tackled in four successive steps.

First of all, an introducing part will draw the general frame of the study. The relationship between reinsurance and natural hazard will be presented. Then, we will focus on the basic windstorms aspects to understand the dilemma of clusters. This contains windstorms formation, the effect produced by different land profiles and an analysis of 1999's Anatol, Lothar & Martin series of storms.

Then, the two alternatives to Poisson model will be theoretically developed. We will first make an extension of the Poisson process to obtain a new annual storm distribution. Later, this distribution will be adjusted to obtain the last model.

A third part will present the data used for the analysis. The idea is to calibrate the models at a European scale before testing the validity of the new models. At this stage we will be able to define the pros and the cons of each of them.

Finally, we will see how the models can be applied at a portfolio level. This will show how reinsurers can customize them to the insurers' portfolios. The impact on expected losses and consequently the premium will hence be pointed out.

A Global Knowledge

A.1. Natural hazards in reinsurance

At its early beginnings[1] reinsurance companies were created to deal either with risky maritime travels or with great fires that could conglomerate entire cities. 1666's Great Fire of London and 1842's Hamburg destruction by conflagration are events that played a major role in the reinsurance development. During the 19th century, the industrial revolution led to massive and large scaled reinsurance system initiated in Germany. It began to spread to a European level to finally become worldwide in early 20th century.

Similarly, the San Francisco earthquake in 1906 was the reason that reinsurers started to cover natural hazards in their treaties. It is only in last twentieth century that natural hazards models started to be developed. Their emergence will be presented in part A.2. For now, let us first focus on how natural hazards can be defined in reinsurance.

A.1.1. Definitions, types and roles in reinsurance

In reinsurance, natural catastrophes are classified in the non-life, property business. They are often characterized as rare events that damage infrastructures and lead to significant insured financial losses.

Two remarks can be made:

- The impact on population is often not studied in insurance, so fatalities are not taken into account.
- The event must lead to important financial losses. A severe storm in the desert is hence not considered as natural catastrophe by reinsurers.

This last point takes its importance from the fact insurance is only set in countries where a market exists. As a consequence, the poorest countries are not being insurance provided. This was the case of Indonesia while touched by a 4.4[2] USD bn costly tsunami in December 2004. Furthermore, natural hazards models are only available for areas where insurance market is developed.

[1] First treaty discovered is dated from 1370
First reinsurance company, Cologne Re, was created in 1842
[2] From a World Bank study

Nowadays, natural hazards in their wide definition can be divided into five classes, according to their nature. These classes are extra-terrestrial, meteorological, climatologic, geological, geomorphic and hydrological hazards. However, reinsurers do not cover all of them because neither all are insurable, nor predictable. Therefore no model can be done. This is the case of meteorites for example. Most of reinsurers restrain their portfolio to:

Meteorological hazards, they contain:

- Windstorms. This general term stands for tropical storms (hurricanes, typhoons and cyclones) and extra-tropical storms. This dissertation is focused on European windstorms, which are extra-tropical storms. Further details will be given in part A.2.
- Hailstorms. They have a major impact in the motor hull business.

Geological hazards, they are made of:

- Earthquakes. Examples of vulnerable regions are the US, Japan, the Mediterranean and China.
- Volcanic eruptions. Some insurance exists for covering damage from volcanic eruptions.

Hydrological hazards, they are spread into:

- Floods. They are the consequence of sustained rainfall, snow of ice melt and make river flow beyond their banks.
- Storm surges. While storms occur, local atmospheric low pressure lifts the sea surface. Then, depending on the direction of the wind, the surge may move towards the shore or away from it.
- Tsunamis. They are tidal waves that have been generated thousands of kilometers away by earthquakes, volcanic eruptions, landslides or asteroid impact. They are usually associated with coasts which are tectonically active.

All these different types of perils have a specific weight in a reinsurer's portfolio. The next two graphs show how the most severe losses from the 1950-2009 periods are spread.

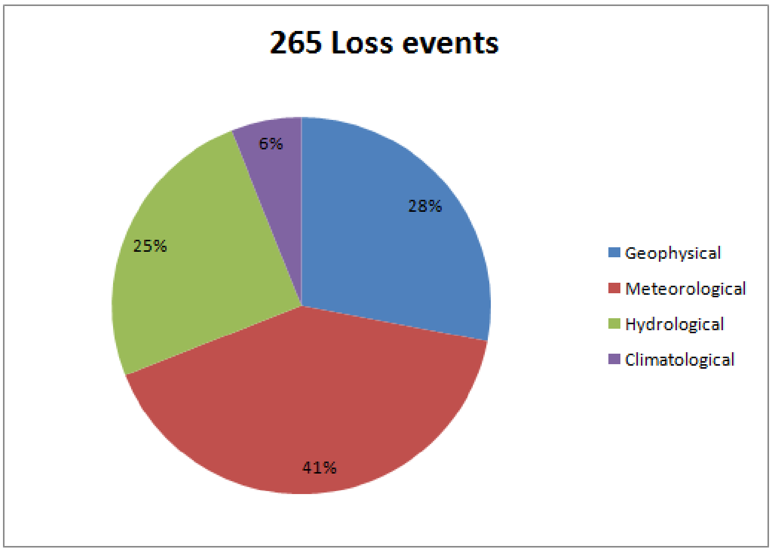


Figure A.1.a: 1950-2009, biggest losses repartition

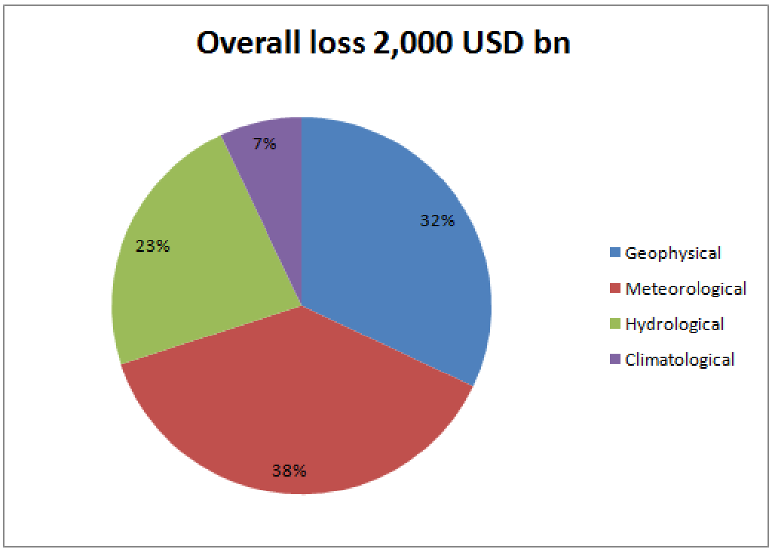


Figure A.1.b: 1950-2009, biggest overall losses

The last two graphs are almost the same. So, in the long term, the overall losses are proportional to the number of losses. This means that in average, a geophysical event seems to make as much damage as a meteorological, a hydrological or a climatologic event. However, this proportion is absent in the insured losses table:

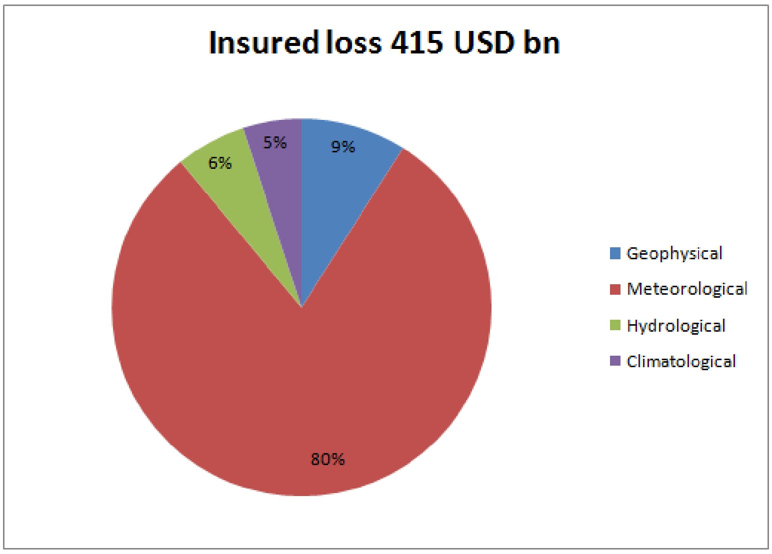


Figure A.1.c: 1950-2009, biggest insured losses

Figure A.1.c reveals the importance of meteorological events in a term of insured losses. The natural hazard marked is hence essentially based on storms. This fact will be more developed in the next part.

A.1.2. 1950-2009 40 most severe insured losses

The following table presents the 40th severe insured losses that occurred between 1950 and 2009. It is the 40th biggest taken out from the 265 events lastly presented. Their losses are 2009 indexed.

Event	Date	Country	Insured loss in 2009 USD m
Hurricane Katrina	25/08/2005	US, Gulf of Mexico	71,163
Hurricane Andrew	23/08/1992	US, Bahamas	24,479
Terror attack on WTC	11/09/2001	US	22,767
Northridge earthquake (M6.6)	17/01/1994	US	20,276
Hurricane Ike	06/09/2008	US, Caribbean	19,940
Hurricane Ivan	02/09/2004	US, Caribbean	14,642
Hurricane Wilma	19/10/2005	US, Mexico, Jamaica	13,807
Hurricane Rita	20/09/2005	US, Gulf of Mexico	11,089
Hurricane Charley	11/08/2004	US, Cuba, Jamaica	9,148
Typhoon Mireille	27/09/1991	Japan	8,899
Hurricane Hugo	15/09/1989	US, Puerto Rico	7,916
Winter storm Daria	25/01/1990	France, UK, Belgium	7,672
Winter storm Lothar	25/12/1999	Switzerland, UK, Fr	7,475
Winter storm Kyrill	18/01/2007	Germany, UK, NL	6,309
Storm and floods	15/10/1987	France, UK, NL	5,857
Hurricane Frances	26/08/2004	US, Bahamas	5,848
Winter storm Vivian	25/02/1990	Europe	5,242
Typhoon Bart	22/09/1999	Japan	5,206
Hurricane Georges	20/09/1998	US, Caribbean	4,649
Tropical storm Allison	05/06/2001	US	4,369
Hurricane Jeanne	13/09/2004	US, Caribbean, Haiti	4,321
Typhoon Songda	06/09/2004	Japan, South Korea	4,074
Hurricane Gustav	26/08/2008	US, Caribbean	3,988
Thunderstorms	02/05/2003	US	3,740
Hurricane Floyd	10/09/1999	US, Bahamas	3,637
Explosion on platform Piper Alpha	06/07/1988	UK	3,631
Hurricane Opal	01/10/1995	US, Mexico	3,530
Great Hansin earthquake (M 7.2)	17/01/1995	Japan	3,482
Winter storm Klaus	24/01/2009	France, Spain	3,372
Winter storm Martin	27/12/1999	Spain, France, SW	3,093
Blizzard, tornadoes, floods	10/03/1993	US, Canada, Mexico	2,917
Severe floods	06/08/2002	UK, Spain, Germany	2,755
Forest fires spread to urban areas	20/10/1991	US	2,680
Hail, floods, tornadoes	06/04/2001	US	2,667
Heavy rainfall, floods	25/06/2007	UK	2,575
Hurricane Isabel	18/09/2003	US, Canada	2,540
Hurricane Fran	05/09/1996	US	2,488
Winter storm Anatol	03/12/1999	Denmark, Sweden	2,454
Hurricane Iniki	11/09/1992	US, North Pacific	2,448
Hurricane Frederic	29/08/1979	US	2,361

Table A.1.d: Forty costliest storms that occurred between 1950 and 2009[3]

From table A.1.d, two observations can be made:

[3] Sigma study, Swiss Re website

- The costliest events are mostly windstorms.

Over 40 presented occurrences, 38 are classified as great natural catastrophes (9/11 terrorist act and Piper Alpha explosion are human made).

The number of storms in general, including winter-storms, tropical storms, hurricanes and typhoons represents 33 over 40 losses (82.5%). By comparing this proportion to the one introduced in figure A.1.a, it stands for the double. In other terms, the more important natural catastrophes taken into account, the bigger the proportion of storm. This can explain why meteorological events are more insured than the others[4].

Hurricane Katrina is a good indicator of the high severity of storms. The hurricane provoked more than 70 USD as if 2009 billion damages.

- Windstorms seem to occur by groups.

Beside the large number of windstorms, one interesting point is the way these occurrences are distributed. In table A.1.d, five colors represent five groups of a few days linked in time[5] events. Inside a same group, all storms stroke approximately the same areas, if we consider them at a large scale. Moreover, storms are known to have a return period of several years. It seems quite curious that an identically storm can strike a same spot few days later.

This observation is the trigger for the main development of this dissertation. Can a specific atmospheric condition lead to several storms formation? Before answering this question, let us focus on the models for natural hazards.

A.1.3. The model issues

The first NAT-CAT[6] models were created on the early nineties. The next part describes the main issues models have been dealing with. We will see that in spite of their improvement, there is still some uncertainty lying nowadays.

1) The beginnings

Natural catastrophes' modeling is a very recent topic that provoked controversy among reinsurers at its start. One of the first model makers, Karen Clarke, decided to exploit hurricane data in the early nineties[7]. For that, she combined the long term record with updated data on property exposure. The result was a simple tool. It could give the probability that an event strike a certain area, and also the inflicted loss. Despite the seriousness of her work that was reinforced by a team of engineers and scientists, insurers and reinsurers did not felt confident in her tool because some of the major potential losses were considered as over-estimated.

---

[4] See Figure A.1.c
[5] From 2 to 10 days
[6] Natural Catastrophes
[7] LEWIS M.(August 26th, 2007), *In Nature's Casino,* The New York Times

It is only on 1992, when hurricane Andrew stroke Florida with an impressive loss of 26.5 USD bn[8] that her work started to pay off. In comparison, the Lloyd's though such an event could not exceed 6 USD bn. This was the trigger for further model developments. The natural hazard market was then handled by three major companies: AIR[9], Eqecat and RMS[10]. Their work team is now composed of meteorologists, physicists, oceanographers, engineers and statisticians. Besides hurricane modeling, they spread to other risks as windstorms, floods, wildfires, tsunamis and global pandemics.

2) The lack of accuracy

The three main companies (AIR, Eqecat and RMS) all referred to historical figures to compute occurrence of perils. In fact, they did even more: they built hypothetical scenarios that have not already been observed from these historical data. This is referred as the "stochastic method". This part of "prediction" of storm track is a delicate point as sometimes, the three models do not provide the same result. A prime example is a today version of a hurricane hitting Miami. AIR estimates the losses at 80 USD bn, RMS at 106 USD bn and Eqecat at 63 USD bn. The gap is significant to put insurers in an uncomfortable position when preparing such scenarios.

Another interesting example is Katrina. This hurricane was formed over Bahamas on August 23rd 2005. The main concern during the live event was the fact the hurricane could go across New Orleans. While the hurricane had just dodged the city, the three modeling companies rushed to give their estimations. RMS claimed losses went from 10 USD bn to 25 USD bn; Eqecat called for a range between 9 USD bn and 16 USD bn and AIR had a range from 12.7 USD bn to 26.5 USD bn. These first figures were significant but not catastrophic. In addition, this under estimation also urged investors to immediately buy CAT bonds. A few hours later, Katrina made the levee break and provoked a massive flood over New Orleans. Bonds price finally plunged. One of them, Zurich's Kamp Re, dropped from 94 USD to 10 cents.

All scenarios cannot be considered because the more complex ones cannot always be foreseen by modelers. As a consequence, there are still a lot of improvements to be done. Model users must stay critical and know the tool limits.

3) The influence of scientists

Nowadays, scientists working on natural hazards modeling play also a major role on the market. The paradox is that some of their studies give contradictory results and despite this absence of unanimous agreement, results are sometimes applied and have a direct impact on insurance companies.

For instance, meteorologists had noticed an uptick in hurricane activity in the North Atlantic Basin since 1995. This uptick was found to have major contribution between July 2004 and the end of 2005. During this period, seven of the most expensive hurricanes struck the American coast. The underneath cause might be either global warming or the alternated ups and downs of

[8] billion
[9] Applied Insurance Research, companies founded by Karen Clarke
[10] Risk Management Solutions

temperatures in the North Atlantic Basin. However, the scientific community failed to agree. Depending on which scientific cause to consider, RMS first estimated a Katrina-size catastrophe to be a once-in-40-years event. Seven months later, the company estimated it as a once-in-20-years event: the risk doubled.

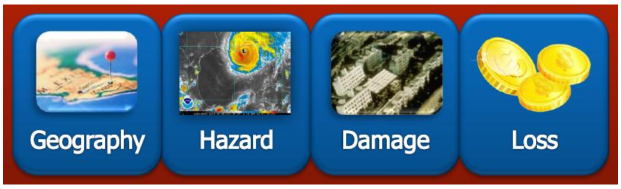
The ratings agencies rely on scientists to evaluate their exposure to natural hazards strongly impacting their solvency requirement. When scientists increased the likelihood of catastrophic storms, Standard & Poor's and Moody's demand insurance companies to raise more capital to cover their hypothetical losses. And so in addition to the more than 40 USD bn they had lost in Katrina, the insurance companies needed to raise 82 USD bn from their shareholders just to keep their investment grade rating[11].

A.1.4. A modeling tool: RMS' Risklink

In this part we will present one of the most used models: Risklink. This is a natural hazard and terrorism tool developed by RMS. In this tool, the user can enter customized inputs and choose a computation engine according the type and the location of the peril. The inputs contain location of perils, information on characteristics of buildings, insurance and reinsurance structures.

For example, the user can define a portfolio of multiple insurers having earthquakes businesses in Italy and adapt different XL[12] structures to each treaty. If he also knows details as the height of buildings, this will be important information because some of them will be more subject to damages according to the intensity and frequency of ground motion. Once he has entered this information, the user has to choose computation engines that are relative to earthquakes in Italy or more generally, earthquakes in Europe. Taking the most adapted engine gives more accuracy and saves time of calculation.

Then, the computation will be divided into four successive steps:



At the end, the results can be displayed in many forms: as an event loss table or as exceedance probability curves. The next part presents all these particular points.

[11] LEWIS M.(August 26th, 2007), *In Nature's Casino,* The New York Times
[12] Excess of Loss

1) The geographic module

It contains a database of stochastic windstorm events that represents a spectrum of likely events that can affect exposures in Europe. These are simulated events that can be used in the probabilistic loss analysis. Each event is described by its physical parameters and frequency of occurrence.

This processed is made of the following steps:

- Creation of tracks with pressure histories;

- Historical track catalog analysis
- Probability distributions calculation
- Stochastic tracks generation
- Pressure histories addition to the stochastic tracks

- Addition of two-dimensional, upper-level pressure fields and wind fields to the stochastic tracks.

2) The hazard module

It determines the wind intensity at each property location for every stochastic windstorm event that is likely to cause losses at that location. It includes the interpolation to the VRG (Variable Resolution Grid) cells and the addition of the land profile that play an important local role[13]. The cell size is related to the peril, the exposure and the hazard gradient. So the resolution of the cells is generally higher in high density areas.

[13] See part A.2.1.4 for more details

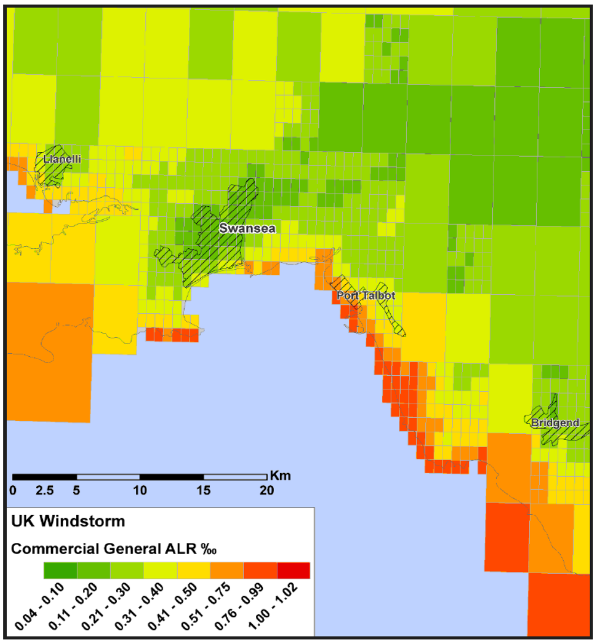


Figure: A.1.e: Grid resolution example[14]

3) The vulnerability module

It calculates the mean damage ratio and coefficient of variation to buildings, contents, and the business interruption losses. Vulnerability functions relate the level of windstorm loss to the severity of the storm at that location. Estimated damage is measured in terms of both mean damage ration and coefficient of variation to capture the uncertainty around the amount of damage incurred at any given wind speed. The damage ratio is defined as the primary uncertainty; the coefficient of variation is the secondary uncertainty.

For a given building type, the model individually evaluates the vulnerabilities of the structural and architectural components that are critical to wind resistance (roof cover, wall types) and combines them to create a vulnerability function for the entire structure. These are the results of experts' engineering studies.

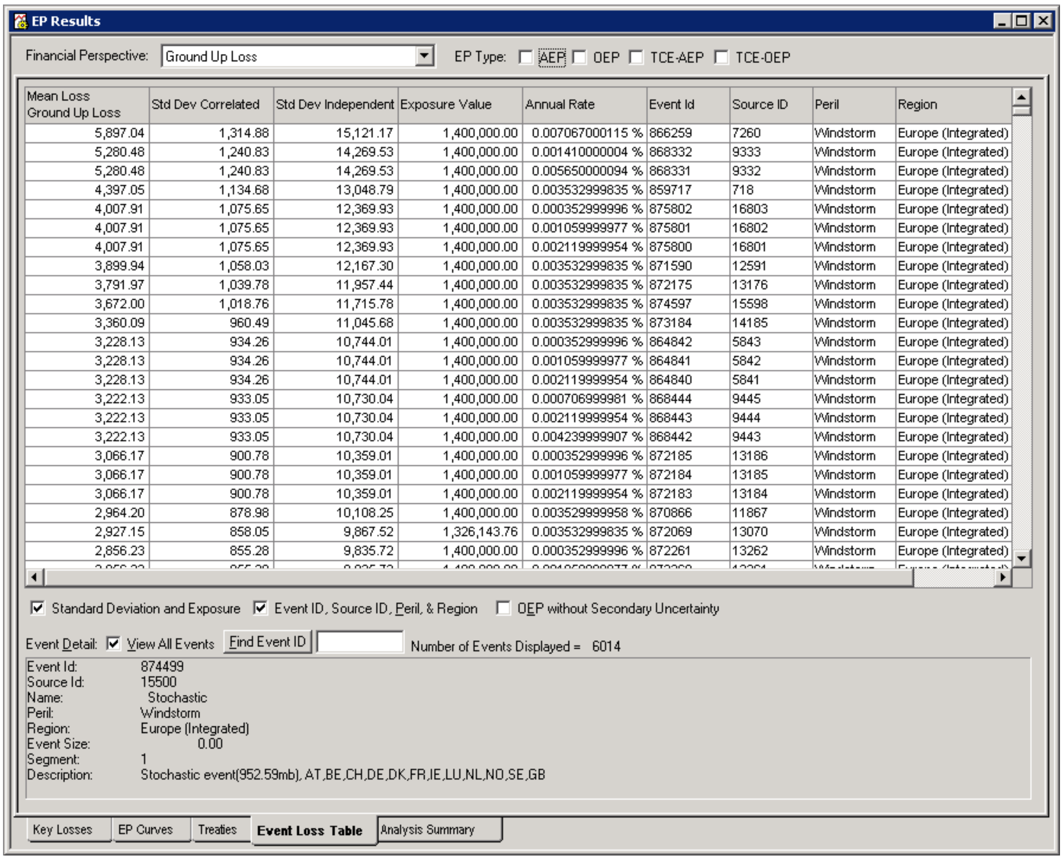
4) The financial analysis module

It calculates losses to different financial perspectives. Several treaties can be applied from both insurer and reinsurer side. The results can be obtained at different pricing steps: ground up loss, gross loss, net loss pre-cat, net loss post cat, net loss post corporate cat, reinsurance gross loss and reinsurance net loss. This can be displayed as an event loss table or as exceedance probability curves.

[14] RMS website documentation (www.rms.com)

5) The event loss table

Once all four modules have participated in computing the final result, an event loss table can be drawn, as shows figure A.1.f. This table provides a list of events that are all likely to affect the portfolio. Each line corresponds to one single event. Its occurrence is Poisson distributed; the parameter can be read in the annual rate column. The severity of each event is beta distributed; its parameters can be obtained from the mean loss and standard deviations columns. Occurrence and severity probabilities are assumed to be independent. The risks inside the same event loss are also all assumed independent.



Mean Loss Ground Up Loss	Std Dev Correlated	Std Dev Independent	Exposure Value	Annual Rate	Event Id	Source ID	Peril	Region
5,897.04	1,314.88	15,121.17	1,400,000.00	0.007067000115 %	866259	7260	Windstorm	Europe (Integrated)
5,280.48	1,240.83	14,269.53	1,400,000.00	0.001410000004 %	868332	9333	Windstorm	Europe (Integrated)
5,280.48	1,240.83	14,269.53	1,400,000.00	0.005650000094 %	868331	9332	Windstorm	Europe (Integrated)
4,397.05	1,134.68	13,048.79	1,400,000.00	0.003532999835 %	859717	718	Windstorm	Europe (Integrated)
4,007.91	1,075.65	12,369.93	1,400,000.00	0.000352999996 %	875802	16803	Windstorm	Europe (Integrated)
4,007.91	1,075.65	12,369.93	1,400,000.00	0.001059999977 %	875801	16802	Windstorm	Europe (Integrated)
4,007.91	1,075.65	12,369.93	1,400,000.00	0.002119999954 %	875800	16801	Windstorm	Europe (Integrated)
3,899.94	1,058.03	12,167.30	1,400,000.00	0.003532999835 %	871590	12591	Windstorm	Europe (Integrated)
3,791.97	1,039.78	11,957.44	1,400,000.00	0.003532999835 %	872175	13176	Windstorm	Europe (Integrated)
3,672.00	1,018.76	11,715.78	1,400,000.00	0.003532999835 %	874597	15598	Windstorm	Europe (Integrated)
3,360.09	960.49	11,045.68	1,400,000.00	0.003532999835 %	873184	14185	Windstorm	Europe (Integrated)
3,228.13	934.26	10,744.01	1,400,000.00	0.000352999996 %	864842	5843	Windstorm	Europe (Integrated)
3,228.13	934.26	10,744.01	1,400,000.00	0.001059999977 %	864841	5842	Windstorm	Europe (Integrated)
3,228.13	934.26	10,744.01	1,400,000.00	0.002119999954 %	864840	5841	Windstorm	Europe (Integrated)
3,222.13	933.05	10,730.04	1,400,000.00	0.000706999981 %	868444	9445	Windstorm	Europe (Integrated)
3,222.13	933.05	10,730.04	1,400,000.00	0.002119999954 %	868443	9444	Windstorm	Europe (Integrated)
3,222.13	933.05	10,730.04	1,400,000.00	0.004239999907 %	868442	9443	Windstorm	Europe (Integrated)
3,066.17	900.78	10,359.01	1,400,000.00	0.000352999996 %	872185	13186	Windstorm	Europe (Integrated)
3,066.17	900.78	10,359.01	1,400,000.00	0.001059999977 %	872184	13185	Windstorm	Europe (Integrated)
3,066.17	900.78	10,359.01	1,400,000.00	0.002119999954 %	872183	13184	Windstorm	Europe (Integrated)
2,964.20	878.98	10,108.25	1,400,000.00	0.003529999958 %	870866	11867	Windstorm	Europe (Integrated)
2,927.15	858.05	9,867.52	1,326,143.76	0.003532999835 %	872069	13070	Windstorm	Europe (Integrated)
2,856.23	855.28	9,835.72	1,400,000.00	0.000352999996 %	872261	13262	Windstorm	Europe (Integrated)

Figure A.1.f: Event loss table[15]

However, two columns stand for standard deviation:

- A fully positively correlated standard deviation;
- And an independent standard deviation.

Their role can be simply introduced and illustrated by the following picture:

[15] Screenshot from Risklink

The soil composition influences a lot the damages due to earthquakes. The buildings standing on a stable and energy absorbing soil will be more likely to stay vertical. The buildings laying on a solid rock will be more impacted by the earthquakes' vibration. This difference of soil composition affects the correlation between buildings. If all buildings are built on a same soil, they will be more likely to get damage the same way, so they are correlated.

For each area and each peril, a coefficient of correlation is defined in order to give for each event a correlation that is between the independent and fully correlated ones. This can be written as:

$$\sigma_{portfolio} = \sigma_{CORR} + \sigma_{IND} = w \sum \sigma_{building} + (1-w)\sqrt{\sum \sigma_{building}^2}$$

This equation directly comes from the following relationships:

- Let X_1 and X_2 be two independent random variables, then :

$$\sigma_{X_1+X_2}^{ind} = \sqrt{{\sigma_{X_1}}^2 + {\sigma_{X_2}}^2}$$

- Let X_1 and X_2 be two fully positively correlated random variables, then :

$$\sigma_{X_1+X_2}^{CORR} = \sigma_{X_1} + \sigma_{X_2}$$

6) The exceedance probability curves

Two curves can be extracted from the tool:

- An occurrence exceedance probability curve.

It represents the probability that one event from the list produces damages above a certain threshold. As all events are Poisson distributed, it considers the event can only occur once in a year.

- An aggregated exceedance probability curve.

It is derived by summation of all events from the list. It gives the probability that this sum of event exceeds a certain threshold. Because adding beta distributions does not give any simple known distribution, a numerical fast Fourier transform method is used for computing the summation. For more information, refer to the appendix.

These curves are presented the following way:

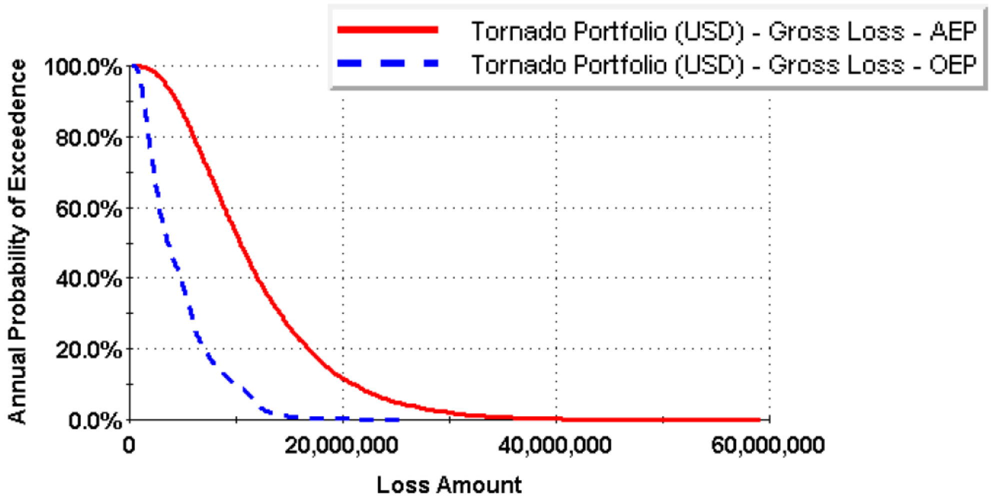


Figure A.1.g: Occurrence and aggregated probability curves[16]

These curves do not necessarily reach 100%. For instance, they can start at 16% for a local area where it is mostly possible that no event happens during a year.

These curves are important because they are the output used for pricing treaties. Furthermore, we will see in part D how to incorporate the clustering effect of storms into them.

[16] Screenshot from Risklink

A.2. Meteorological facts

This part exclusively focuses on windstorms. We will first introduce characteristics and land elements playing direct roles on the severity of storms. Then, the meteorological process that led to 1999's series of storms will be presented.

A.2.1. Windstorm characteristics

As the dissertation focuses only on European storms, the distinction between extra-tropical and tropical storm will firstly be set.

1) Extra-tropical and tropical windstorms

Several wind perils fall under the general term "windstorm". Extra-tropical storms are the ones we study. They are a subset of the 150 to 200 Atlantic depressions generated each year. While they can generate hurricane force winds, they are not classified as hurricanes, which are a subset of the 10 to 20 tropical storms generated in the Atlantic each year.

Here is a table catching the different characteristics of tropical and extra-tropical storms in the Atlantic Basin:

Type	Extra-Tropical storms	Tropical storms
Formation	Polar front	Tropical
Fuel	Temperature Contrast	Evaporating water
Damaging winds	**50-400 miles** from center	**5-100 miles** from eye
Lifetime	Highest wind c. 24 hours	Up to 10 days
Typical Wind speeds	Up to **100 mph**	Up to **140 mph**
Typical Forward-speeds	30-50 mph	10-40 mph
Land falling behavior	Can intensify	Always decay
Occurrences	**150-200** events per year	**<10** events per year
Seasonality	Winter (October-April)	Summer (July-October)

Table A.2.a: Windstorms characteristics[17]

Extra-tropical cyclones winds can cause moderate damage over a wide area of land. This area can extend as far as many hundreds of kilometers from the center of the storm, with winds penetrating hundreds or even thousands of kilometers along the storm path. Furthermore, the intensity of an extra-tropical storm is not significantly affected as the storm moves over land. As described in part A.2.1.4, variations in terrain can also have an impact on the local peak gust speeds.

[17] RMS website documentation (www.rms.con)

2) Formation

As one can read in Woo (1999):

"In middle latitudes, the large-scale variability of the weather is dominated by extra-tropical storms. From differential heating between high and low attitudes arise large-scale horizontal temperature gradients. The pole ward decrease in the temperature is usually concentrated in baroclinic zones. Subject to wave perturbations, these zones become unstable. The air flow breaks up into large vortex structure known as cyclones. Some of most active weather events are associated with instabilities which occur at fronts. These fronts are discontinuities that form in strongly baroclinic zone of the atmosphere. One of the most important is the polar front which separates polar air from tropical air. Along parts of the polar front, colder and denser polar air advances towards the equator to form a cold front; parts moving pole ward form a warm front. The temperature contrast between the warmer and colder air masses is especially large in winter; a factor which explains the greater severity of winter storms.

When warm and cold front meet, a depression forms because the two air masses whirl. With additional energy being concentrated in the low pressure area, strong air currents may be created over several days around the center of the low pressure area. An intensification of the depression would cause the isobars to become tighter and winds to increase. Wind can be very strong but rarely exceed 55 m/s (198 km/h)."

3) Intensity

Wind intensity can be measured by the Beaufort scale:

Bft	Descriptive term	Mean wind speed at 10m above surface				
		m/s	km/h	mph	knots	Kg/m2
0	Calm	0-0.2	0-1	0-1	0-1	0
1	Light air	0.3-1.5	1-5	1-3	1-3	0-0.1
2	Light breeze	1.6-3.3	6-11	4-7	4-6	0.2-0.6
3	Gentle breeze	3.4-5.4	12-19	8-12	7-10	0.7-1.8
4	Moderate breeze	5.5-7.9	20-28	13-18	11-15	1.9-3.9
5	Fresh breeze	8.0-10.7	29-38	19-24	16-21	4.0-7.2
6	Strong breeze	10.8-13.8	39-49	25-31	22-27	7.3-11.9
7	Near gale	13.9-17.1	50-61	32-38	28-33	12.0-18.3
8	Gale	17.2-20.7	62-74	39-46	34-40	18.4-26.8
9	Strong gale	20.8-24.4	75-88	47-54	41-47	26.9-37.3
10	Storm	24.5-28.4	89-102	55-63	48-55	37.4-50.5
11	Violent storm	28.5-32.6	103-117	64-72	56-63	50.6-66.5
12	Hurricane	32.7-	118-	73-	64-	66.6-

Table A.2.b: Beaufort scale[18]

[18] www.uwsp.edu

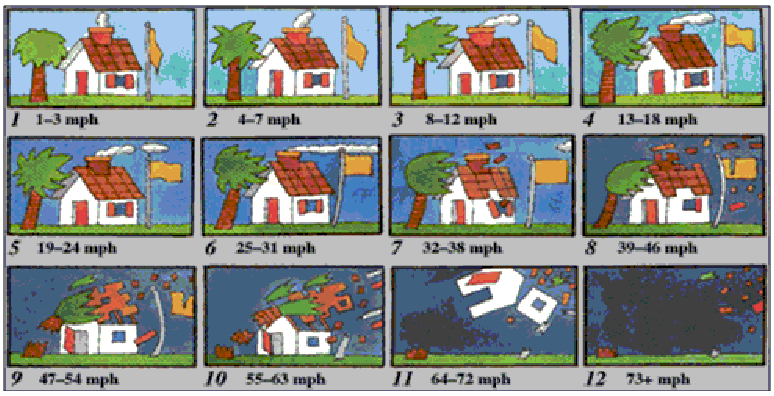


Figure A.2.c: Beaufort scale illustration[19]

The Beaufort scale, presented by Table A.2.b, gives the speed value that is recorded by a wind station at 10 meters above the surface. They hence do not reflect local gusts which are a consequence of the relief.

4) Land profile

The following graph presents three different contributions of the relief:

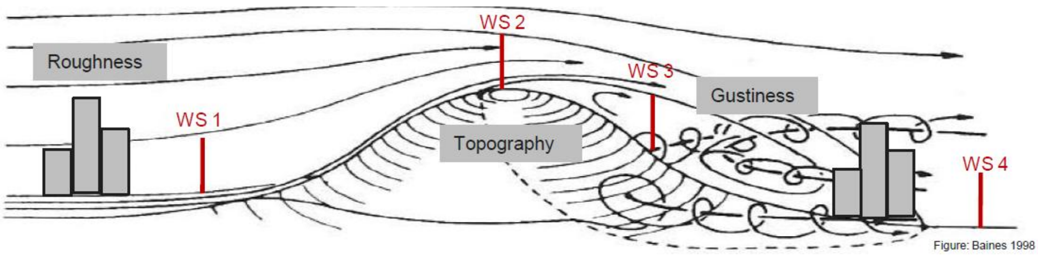


Figure A.2.d: Land wind characteristics[20]

The surface roughness describes the way that winds are slowed down by frictional effects. Frictional effects are minimal over open water but then they become increasingly important as the wind passes over land. The wind may be affected by natural roughness; the hilly and forested areas slow it down. The wind may also be affected by manmade roughness, particularly in cities where the wind is slowed as it interacts with buildings.

Models usually propose different inputs to take this parameter into account:

[19] www.ussartf.org
[20] RMS website documentation (www.rms.com)

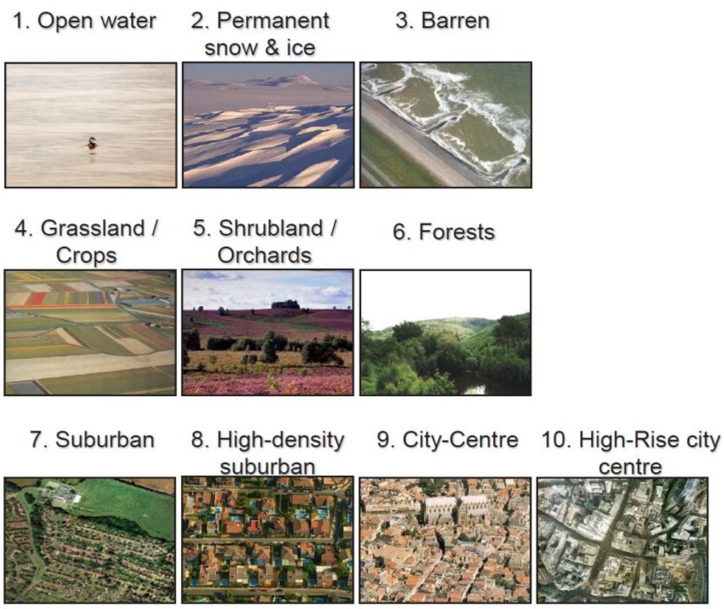


Figure A.2.e: Panel of different roughness lands[21]

Wind interacts in a complex manner with topography as it flows around large obstructions. Stronger winds are experienced at higher altitudes while at the same time, mountainous terrain reduces the strength of the wind field due to frictional effects. In the opposite, mountains can also create Venturi effect like in Switzerland where records of wind speed were established.

The principal factor of the variation in loss between storms with identical peak gust speeds is the number of intense gusts that are very close to the level of the peak gust. Higher insurance losses were experienced where the greatest numbers of gusts were close to the maximum recorded at a site, in other words, where the density of gusts was high. Gust-density is related to three parameters:

- the maximum wind speed in the storm;
- the duration of the storm;
- the storm width.

Therefore, smaller storms tend to be gustier than larger, longer lasting, diffuse storms.

A.2.2. 1999's series of storms over Europe

In December 1999 Europe was hit by three severe windstorms which, in the most severely affected countries, reached a wind force of 12 on the Beaufort[22] wind scale. Wind speeds peaks over 180 km/h were locally reached. One of the causes might be that year 1999 was the third warmest year

---

[21] RMS website documentation (www.rms.com)
[22] See Table A.2.b

of the century in central Europe, together with 1990. In December, temperatures on the European mainland were also much higher than the long-term average, with the exception of the week before Christmas (no storm occurred during that week). These high temperatures were supposedly responsible for the large scaled meteorological situation which led to the development of Anatol, Lothar, and Martin. The conditions were quite similar to the ones at the beginning of the series in 1990.

1) Anatol, 12/03/1999

Figure A.2.f: Path of Anatol[23]

The development of Anatol began over northwest of Ireland in the early morning of December 3rd. The central pressure of the low-pressure vortex dropped by more than 40 hPa in the space of only twelve hours. It reached the minimum of 952 hPa over Denmark. This made Anatol the strongest gale to hit Denmark in the 20th century. The highest wind speeds were recorded south of the low-pressure system over German Bight in the central North Sea and in southern Denmark. With gusts of 180–185 km/h record levels were registered in the long-term wind time series at some meteorological stations.

[23] GEO RISKS RESEARCH (2002), *Tempêtes d'hiver en Europe*, Munich Re

2) Lothar, 12/26/1999

Figure A.2.g: Path of Lothar[24]

On 25th December 1999 a disturbance appeared over the northeastern Atlantic on a frontal zone between cold air in the north and warm air in the south, belonging to a mighty central low. It developed into a series of intense low-pressure systems. Around noon, one of these systems, the secondary low named Lothar, had a central pressure of 995 hPa. As in the Anatol case, the development of the low-pressure system was almost explosive, involving a large pressure drop within the space of a few hours. The most noticeable change in pressure was observed at the Caen station on the French coast, where a pressure drop of 28 hPa was recorded. Such dimensions had not been observed in Europe for at least thirty years. The atmospheric pressure reached its minimum of 961 hPa over Normandy. Anatol had a minimum central pressure of 952 hPa, making Lothar a much less extreme event in meteorological terms.

Lothar's wind field over France was very similar to Herta's in February 1990. And Lothar struck a severe blow to Greater Paris on 26th December 1999, just as Herta had done on 3rd February 1990. There were also strong similarities in terms of the geographical extent of the two wind fields. The main difference between these two windstorm systems was the wind speed. With peak gusts of around 170 km/h in Paris, Lothar was 30–40% more intense than Herta, which reached about 120–130 km/h.

[24] GEO RISKS RESEARCH (2002), *Tempêtes d'hiver en Europe*, Munich Re

3) Martin, 12/27/1999

Figure A.2.h: Path of Martin[25]

By the evening of 26th December 1999 Lothar's center had shifted eastwards to Poland. But then another secondary low, Martin, detached itself on the frontal zone over the North Atlantic in almost the same way as Lothar. Its track was a little further to the south so that the regions particularly affected in France were around Bordeaux, Biarritz, and Toulouse. On 27th December 1999 Martin brought gale force winds to northern Spain, western parts of Switzerland, and Upper Italy too.

Lothar and Martin were hence created from the same atmospheric conditions in the North Atlantic frontal zone. The insurance industry, particularly in France, was confronted with catastrophic windstorm losses from two events in less than 48 hours. The development of a windstorm series over the North Atlantic with a close succession of intense low-pressure systems is not unusual given an appropriate large-scale meteorological situation. This is the reason why the formation of clusters is not to be ignored in consideration of accumulation aspects. Relying on this particular point, it seems as necessary as natural to draw models based on this clusters phenomenon.

[25] GEO RISKS RESEARCH (2002), *Tempêtes d'hiver en Europe*, Munich Re

B Cluster modeling

As described in section A.2.3, insurers had to face series of storms in 1999. This year was not the only one in which the phenomena could also be observed. 1990's Daria, Herta, Vivian and Wiebke were comparable series as presented in the following table:

Storm losses (EUR[26] million, as if 2009)							
Name	**Daria**	**Herta**	**Vivian**	**Wiebke**	**Anatol**	**Lothar**	**Martin**
Year	**1990**	**1990**	**1990**	**1990**	**1999**	**1999**	**1999**
Month	**01**	**02**	**02**	**02-03**	**12**	**12**	**12**
Day	**25/26**	**03/04**	**25/27**	**28/01**	**03/04**	**26**	**27/28**
Germany	520	260	520	520	100	650	
Austria			70	70			
Belgium	220	100	170	50			
Denmark	50		30		2,000		
Spain							50
France	260	600	90	100		4,450	2,450
United Kingdom	2,600		700	280			
Luxembourg	50	50	50	50			
Netherlands	700	100	90	30			
Switzerland			50	50		800	
TOTAL	4,400	1,110	1,820	1,180	2,250	5,900	2,500

Table B: 1990 & 1999 series of storms losses

As underlined by Table A.1.d and Table B, the assumption of memoryless inter-arrival times does not seem to be well adapted. The impact of such a simple hypothesis can be very important regarding the amount of insured damages. Moreover, some reinsurance treaties only cover second or third events. Taking the existence of cluster into account will have an impact on the FGU[27] losses[28], but this impact will be even more significant for these specific treaties.

We want to model the situation previously described in part A.2.2. To do that, we will model a situation in which a group of storms is assumed to be created from same specific meteorological conditions. The word used for it is "cluster". We will assume these clusters are Poisson distributed, not windstorms.

[26] Euro
[27] From Ground Up, direct losses without any financial structure
[28] -This point will be developed in part D

B.1. Poisson Process review

This part gives a few characteristics of the Poisson process. For further information and proofs please refer to ROSS M. (2009).

Let us briefly recall the definition of a Poisson process:

Let $N(t)$ be a counting process that represents the number of storms created from initial time value 0 to t^-, the lower limit of time t.

The classic way to define this process as Poisson is given by the following points:

- $N(0) = 0$
- Independent increments : $N(t) - N(s) \perp N(t') - N(s'),\ \forall\, s' < t' < s < t$
- Stationary increments : $N(t+h) - N(t) = N(s+h) - N(s),\ \forall\, t, s, h$

However, this definition is not the most practical for the study. We will hence use the following one. Let us define the process $\{N(t), t \geq 0\}$ as:

- $P(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h)$
- $P(N_{t+h} - N_t = 1) = \lambda h + o(h)$
- $P(N_{t+h} - N_t > 1) = o(h)$

where $o(h)$ is such that $\lim_{h \to 0} \frac{o(h)}{h} = 0$.

One direct consequence is that $N(t)$ has a Poisson distribution with rate λt. Parameter λ is known as the “intensity rate” of the process.

B.1.1. Exponential inter-arrival time property

We lastly defined the distribution of number of events, although focusing on the inter arrival time distribution is also possible. If one denotes $X(k)$ the inter-arrival time between the k^{th} and $(k+1)^{th}$ storm then $X(k) \sim Exponential(\lambda)$, we have:

$$P(X(k) \leq t) = P\left(N_{t + \sum_{i=0}^{k-1} X(i)} - N_{\sum_{i=0}^{k-1} X(i)} \geq 1\right)$$

$$= P(N_t \geq 1)$$

By using conditional probabilities and stationary increments

$$P(X(k) \leq t) = 1 - P(N_t = 0)$$

$$= 1 - e^{-\lambda t}$$

Remark: this result does not depend on k.

B.1.2. Labeling a Poisson process

Let us consider a Poisson process N with intensity rate λ. Let us also consider one can label the Poisson occurrences and can affect them a distribution which is independent from the Poisson process. Then these sub-processes that can be distinguished according to their labels are also Poisson processes.

For example, Poisson processes are commonly used in queuing systems. We can consider a server, like a hospital reception, in which patients arrive at rate λ. Now imagine the receptionist makes a distinction according to the type of injury, in order to send the patient to a specialized department. If $\frac{1}{3}$ of the patients comes because of a fracture and $\frac{2}{3}$ because of a sprain, then the number of people coming regarding their injuries can be modeled as two Poisson processes with respective rates $\frac{\lambda}{3}$ and $\frac{2\lambda}{3}$.

To prove it, let us consider r different labels for a Poisson process N. Let A_h be the event of 2 or more points in N in the time interval $(t, t+h]$, B_h be the event of exactly one points in N in the time interval $(t, t+h]$ and C_h for zero point. Let A_{ih}, B_{ih} and C_{ih} be the corresponding events for N_i. Let H_t denote the history of the processes up to time t^-:

$$H_t = \{N_i(s); 0 \leq s < t; i = 1, \dots, r\}$$

Since $P(A_h|H_t) = o(h)$ and A_{ih} is a subfield of A_h, we have:

- $P(A_{ih}|H_t) = o(h)$
- $P(B_{ih}|H_t) = p_i P(B_h|H_t) + o(h) = p_i\big(\lambda h + o(h)\big) + o(h) = p_i \lambda h + o(h)$
- $P(C_{ih}|H_t) = 1 - P(B_{ih}|H_t) - P(A_{ih}|H_t) = 1 - p_i\lambda h + o(h)$

This proves N_i, $i \in [\![1, r]\!]$, is a Poisson process with rate $p_i\lambda$.

B.1.3. Parameter estimation

Regarding a one year length period, let us assume $X \sim Poisson(\lambda)$.
By calculating the maximum likelihood estimator, one shows:

$$L_{NB}(x_1, \dots, x_n; \lambda) = \prod_{i=1}^{n} P(X_i = x_i)$$

$$l_{NB}(x_1, \ldots, x_n; \lambda) = ln\big(L_{NB}(x_1, \ldots, x_n; \lambda)\big)$$

$$= \sum_{i=1}^{n} ln\big(P(X_i = x_i)\big)$$

$$= \sum_{i=1}^{n} \{-\lambda + x_i ln(\lambda) - ln(x_i)\}$$

$$\frac{\partial}{\partial \lambda} l_{NB}(x_1, \ldots, x_n; \lambda) = 0 \Leftrightarrow \hat{\lambda}_{MLE} = \bar{x}$$

Where $\bar{x}$ is the arithmetic mean $\frac{1}{n}\sum_{i=1}^{n} x_i$.

Both Maximum Likelihood and Moment methods give the same estimator.

B.2. Negative binomial modeling

In this section, we will propose an extension to the classic Poisson Process. To do so, we state that the chance to have another storm increases with the number of storm that previously occurred during our observation frame. This can be modeled by making the intensity rate jump.

B.2.1. The jump of intensity rate approach

The idea of the extension proposed by J. Schirmer (2010) is to make the intensity rate increase arithmetically. The intensity rate jumps every time an event occurs:

- Initial value is : λ
- Intensity is $\lambda + \delta$ after the first event
- Intensity is $\lambda + k\delta$ after the kth event…

This process gives a negative binomial distribution.¨

1) Annual distribution

One wants to prove the annual storm distribution is Negative Binomial.

Let $w_{s,t}^{(k)}$ be the probability of having no event between *s* and *t.* By using the Poisson process definition introduced in part B.1.2:

$$w_{t,t+dt}^{(k)} = P(no\ claim\ between\ t\ and\ t+dt \mid k\ events\ already\ occured)$$

$$= 1 - (\lambda + k\delta)dt + o(dt)$$

Now, considering stopping time T, we want to demonstrate that for $k \in \mathbb{N}$ we have:

$$P_T(N = k) = \binom{r + k - 1}{k} e^{-\lambda T}\left(1 - e^{-\delta T}\right)^k \qquad (NB.1)$$

where $r = \frac{\lambda}{\delta}$.

The actual idea is to prove it by induction. So we use:

$$P_T(N = 0) = w_{0,T}^{(0)} = e^{-\lambda T}$$

Then

$$P_T(N = k + 1) = \int_0^T P_t(N = k)\big(1 - w_{t,t+dt}^{(k)}\big)w_{t,T-(t+dt)}^{(k+1)}$$

See appendix A1 for proof details.

2) Estimation of joint parameters

We compute the estimator of joint parameters (r, p) thanks to both maximum likelihood and moments methods.

Let us consider a n events sample. Let us now compute the likelihood function:

$$L_{NB}(x_1, \dots, x_n; p, r) = \prod_{i=1}^{n} P(X_i = x_i)$$

$$= \prod_{i=1}^{n} \binom{r+x_i-1}{x_i} p^r (1-p)^{x_i}$$

$$l_{NB}(x_1, \dots, x_n; p, r) = ln\big(L_{NB}(x_1, \dots, x_n)\big)$$

$$= \sum_{i=1}^{n} \left\{ln \binom{r+x_i-1}{x_i} + rln(p) + x_i ln(1-p)\right\}$$

$$\frac{\partial l_{NB}}{\partial p}(x) = 0 \Leftrightarrow \sum_{i=1}^{n} \left\{\frac{r}{p} - x_i \frac{1}{1-p}\right\} = 0$$

$$\Leftrightarrow \frac{nr}{p} = \frac{n\bar{x}}{1-p}$$

$$\Leftrightarrow p = \frac{r}{\bar{x}+r}$$

$$\frac{\partial l_{NB}}{\partial r}(x) = 0 \Leftrightarrow \frac{\partial}{\partial r} \sum_{i=1}^{n} \{ln\big((r + x_i - 1)!\big) - ln\big((r-1)!\big) + rln(p)\} = 0$$

$$\Leftrightarrow \frac{\partial}{\partial r} \sum_{i=1}^{n} \left\{ln\Big(\prod_{j=0}^{x_i-1}[r + j]\Big) + rln(p)\right\} = 0$$

$$\Leftrightarrow \frac{\partial}{\partial r} \sum_{i=1}^{n} \left\{\sum_{j=0}^{x_i-1}\{ln(r + j)\} + rln(p)\right\} = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} \sum_{j=0}^{x_i-1} \frac{1}{r+j} = -nln(p)$$

So the estimator $(\hat{r}, \hat{p})$ has to verify the following equations:

$$\begin{cases} ln\left(1+\dfrac{\bar{x}}{\hat{r}_{MLE}}\right) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\sum_{j=0}^{x_i-1}\dfrac{1}{\hat{r}_{MLE}+j} \\ \hat{p}_{MLE} = \dfrac{\hat{r}_{MLE}}{\bar{x}+\hat{r}_{MLE}} \end{cases}$$

These estimators can be calculated numerically using a simple solver algorithm.

Concerning the method of moments, we consider $N \sim NegativeBinomial(r,p)$. Let $m_r = \frac{1}{n}\sum_{k=1}^{n}(x_k)^r$ be the r-empirical moment.

So we have to match: $\begin{cases} m_1 = E[N] = r\frac{1-p}{p} \\ m_2 - (m_1)^2 = V(N) = r\frac{1-p}{p^2} = \frac{m_1}{p} \end{cases}$

Hence we find: $\begin{cases} \hat{p}_{MM} = \frac{m_1}{m_2-(m_1)^2} \\ \hat{r}_{MM} = \frac{m_1\hat{p}_{MM}}{1-\hat{p}_{MM}} \end{cases}$

These estimators are simpler to obtain than the MLE.

B.2.2. Collective risk model approach

Another method for creating a negative binomial distribution is to use a Collective Risk Model, more precisely, by combining a Poisson distribution with a logarithmic severity distribution. Let us prove the portfolio is negative binomial distributed. Let M be defined as follows:

$$M = \sum_{i=1}^{N} L_i \qquad (NB.2)$$

where $N \sim Poisson(\lambda)$, $L_i \sim Logarithmic(p)$ and L_i are independent r.v.s.

Using this form, the negative binomial distribution can be interpreted in the following way. The number of cluster is Poisson distributed. Then, the number of storms produced by each of these meteorological situations is logarithmically distributed. Each cluster produces at least one storm, since the logarithmic distribution is defined on $\mathbb{N}^*$.

The logarithmic distribution can be by using Taylor series:

$$\frac{1}{1-p} = \sum_{k=0}^{+\infty} p^k \,, \forall\, 0 < p < 1$$

So, by integration we have:

$$-ln(1-p) = \sum_{k=1}^{+\infty} \frac{p^k}{k} \Leftrightarrow -\sum_{k=1}^{+\infty} \frac{p^k}{k} \frac{1}{ln(1-p)} = 1$$

The logarithmic distribution can be defined as:

$$P(L = k) = \frac{-1}{ln(1-p)} \frac{p^k}{k}, \forall k \in \mathbb{N}^*$$

Remark on the logarithmic distribution:

- There is no $P(L = 0)$, this means the only possibility for the portfolio to have no storm for the whole year, is the probability that $N = 0$, which means that no storm weather condition appears because if it does, one storm will at least occur.

- Further information on this distribution can be found in the following table:

Mean	Variance	Moment Generating Function
$\frac{-p}{(1-p)ln(1-p)}$	$\frac{-p\left(p + ln(1-p)\right)}{(1-p)^2\, ln^2(1-p)}$	$\mathcal{M}_L(t) = \frac{ln(1-pe^t)}{ln(1-p)}$

Table B.2.a: Logarithmic distribution characteristics

1) Proof of models' equivalence

We want to prove that the model $(NB.2)$ can be written as a negative binomial distribution $(NB.1)$. As shown in Schirmer (2010) its moment generating function is:

$$\mathcal{M}_M(t) = m_N \circ \mathcal{M}_{L_1}(t)$$

where $m_N(z) = E[z^N]$ is the probability generating function.

By using $\mathcal{M}_L(t) = \frac{ln(1-pe^t)}{ln(1-p)}$ and $m_N(z) = e^{\lambda(z-1)}$, one has:

$$\mathcal{M}_M(t) = \left(\frac{1-p}{1-pe^t}\right)^{\frac{-\lambda}{ln(1-p)}}$$

By taking $r = \frac{-\lambda}{ln(1-p)}$ and $\tilde{p} = 1-p$ one can directly see that $\mathcal{M}_M(t)$ stands for a negative binomial moment generating function with parameters r and $\tilde{p}$.

See appendix A.2 for more details on the proof.

2) Estimation of single parameters

By using definition $(NB.2)$ we clearly can estimate the cluster parameter λ in a first time, then the storm per cluster parameter p. To do so, we must first decide how we will define the clusters of storm in the data. This will be the major point of part C.3.1. Parameter λ will hence be estimated the same way than the Poisson storm model (see part B.1.3).

Concerning parameter p, there is no direct solution to compute its estimator. However, we can prove the equation verified by the estimator always have a solution, under certain conditions. The method of moment and the maximum likelihood method give the same estimator.

Assume $L_i \sim Logarithmic(p)$. Then,

$$\frac{\partial}{\partial p} ln\left(\prod_{i=1}^{n} P_{L_i}(x_i)\right) = 0 \Leftrightarrow \frac{\partial}{\partial p}\sum_{i=1}^{n} ln\left\{\frac{-1}{ln(1-p)}\frac{p^{x_i}}{x_i}\right\} = 0$$

$$\Leftrightarrow \frac{\partial}{\partial p}\sum_{i=1}^{n}\left\{-ln\big(-ln(1-p)\big) + x_i ln(p) - ln(x_i)\right\} = 0$$

$$\Leftrightarrow \frac{\bar{x}}{p} + \frac{1}{(1-p)ln(1-p)} = 0$$

$$\Leftrightarrow \bar{x}q ln(q) - q + 1 = 0$$

where $q = 1-p$.

Does a solution always exist?
Let us study the function $f: q \longmapsto \bar{x}qln(q) - q + 1$ to see if it is possible to have $f(q) = 0$ for any $q \in]0,1[$. f is a $\mathcal{C}^\infty$-function, so its derivative exists and the function is continuous. Its derivative is given by the function $f': q \longmapsto \bar{x}ln(q) + \bar{x} - 1$

$$\forall q \in]0,1[: \begin{cases} f'(q) > 0 \\ f'(q) = 0 \\ f'(q) < 0 \end{cases} \Leftrightarrow \begin{cases} q > exp\left(\frac{1-\bar{x}}{\bar{x}}\right) \\ q = exp\left(\frac{1-\bar{x}}{\bar{x}}\right) \\ q < exp\left(\frac{1-\bar{x}}{\bar{x}}\right) \end{cases}$$

Moreover, $exp\left(\frac{1-\bar{x}}{\bar{x}}\right) \in]0,1[\Leftrightarrow \bar{x} > 1$. This is true for two reasons. First, all logarithmic values are bigger than one. So we only have to check the presence of one value that is strictly bigger than 1 to have $\bar{x} > 1$. If this is not the case, there is no point to estimate constant values as logarithmic distributed.

$$f\left(exp\left(\frac{1-\bar{x}}{\bar{x}}\right)\right) = -\bar{x}exp\left(\frac{1-\bar{x}}{\bar{x}}\right) + 1 < 0 \ \ if \ \ \bar{x} > 1$$

Let us draw the analysis table of function f :

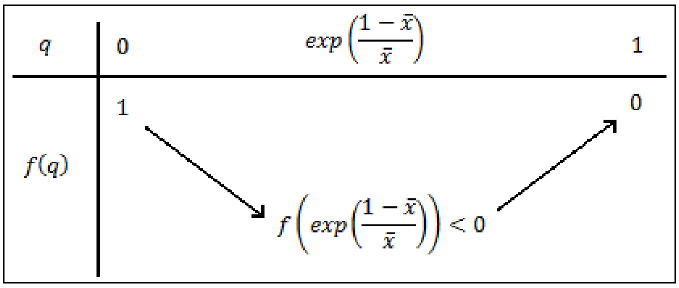


Figure B.2.b: Variation of function f

As f is a real-valued and continuous function, by using the intermediate value theorem we prove there exists one unique $q^* \in \left]0, exp\left(\frac{1-\bar{x}}{\bar{x}}\right)\right[$ such that $f(q^*) = 0$. This q^* gives us the solution of the likelihood equation $q^* = 1 - p^*$.

We proved that under the $\bar{x} > 1$ assumption, there exists a solution for the maximum likelihood equation on the interval $]0,1[$. So the numerical solver will always find a solution.

B.3. Poisson-Binomial collective risk model

The idea of this model is to assume that clusters are Poisson distributed and to change the distribution of storms per cluster. Actually, the logarithmic distribution was unsatisfying on two points:

- The number of storms in a same cluster can tend to infinity.

If one thinks of a cluster as a specific atmospheric condition, this situation is limited in time. It seems adapted to restrain the number of storms produced. In historical data, this number does not exceed 3 (as we will see in part C).

- It happens that atmospheric conditions lead to no storm formation.

This means that inside the CRM portfolio we should take into account the fact that one cluster can produce no storm.

The new CRM can hence be defined as:

$$X = \sum_{i=1}^{Y} C_i \qquad (CRM.1)$$

Where Y is the annual number of clusters, $Y \sim Poisson(\lambda)$, C_i are the numbers of storms inside the i^{th} cluster, $C_i \sim Binomial(n, p)$, and C_i are independent. n stands for the maximum number of storm per cluster.

As we just explained, the number of storm in a cluster can be zero. Nonetheless, the clusters leading to no storm will not be visible in data samples. This means we will need to work with the truncated random variable $C' = C|C > 0$ if we want to obtain estimations from this data. More details will be developed in part B.3.1.3.

B.3.1. Decomposition into Poisson sub-processes

By using the CRM model form, one can write the moment generating function as:

$$m_X(u) = m_Y \circ m_C(u)$$

$$= exp\{\lambda(\sum_{i=1}^{n} P_C(i)u^i - 1)\}$$

By using the Poisson probability generating function: $m_Y(u) = exp\{\lambda(u-1)\}$, one obtains:

$$= \prod_{i=1}^{n} exp\{\lambda P_C(i)(u^i - 1)\}$$

$m_X(u) = \prod_{i=1}^{n} m_{iX_i}(u)$

where $X_i \sim Poisson\big(\lambda P_C(i)\big)$ and, moreover, (X_i) are two-by-two independent.

Using this result, one can rewrite $(CRM.1)$ as:

$$X = \sum_{i=1}^{n} iX_i \qquad (CRM.2)$$

In this case, X_i stands for the number of clusters leading to i storms, for $i \in \mathbb{N}$. The underlying reason is that clusters producing i storms are derived from the total number of clusters which is Poisson distributed. So they can be considered as "labels" of the cluster Poisson process[29].

B.3.2. Distribution of the annual number of storms

Let us note $X^k = \sum_{i=1}^{k} iX_i$. Then, because of $(CRM.2)$, we can write $X^n = X$

Finally the distribution can be written as:

$$P_X(x) = \sum_{k_n=0}^{\left\lfloor \frac{x}{n} \right\rfloor} P_{X^{n-1}}(x - nk_n) P_{X_n}(k_n)$$

$$= \sum_{k_n=0}^{\left\lfloor \frac{x}{n} \right\rfloor} \left[\sum_{k_{n-1}=0}^{\left\lfloor \frac{x-nk_n}{n-1} \right\rfloor} P_{X^{n-2}}(x - nk_n - (n-1)k_{n-1}) P_{X_{n-1}}(k_{n-1}) \right] P_{X_n}(k_n)$$

$$= \sum_{k_n=0}^{\left\lfloor \frac{x}{n} \right\rfloor} \sum_{k_{n-1}=0}^{\left\lfloor \frac{x-nk_n}{n-1} \right\rfloor} \dots \sum_{k_3=0}^{\left\lfloor \frac{x-\sum_{i=4}^{n} ik_i}{3} \right\rfloor} \sum_{k_2=0}^{\left\lfloor \frac{x-\sum_{i=3}^{n} ik_i}{2} \right\rfloor} \left\{ \prod_{j=2}^{n} P_{X_j}\big(k_j\big)\, P_{X_1}\Big(x - \sum_{i=2}^{n} ik_i\Big) \right\}$$

The goal here is to find all the values the (X_i) can take to make the overall sum equal to x. One can start the algorithm by dividing the wanted number x by the biggest number of storms per cluster possible, which is n. This gives us information about the maximal number of n-cluster we can use in our decomposition. Then this number of n-clusters, represented by the variable k_n, can vary from 0 to the maximum, namely $\left\lfloor \frac{x}{n} \right\rfloor$. Once k_n is fixed, we do the same decomposition for the rest, that is now $x - nk_n$, and the variable k_{n-1} of $(n-1)$-clusters. k_{n-1} will vary between 0 and its maximum $\left\lfloor \frac{x-nk_n}{n-1} \right\rfloor$, etc. This process stops at k_1, the number of 1-cluster that is equal to $x - \sum_{i=2}^{n} ik_i$.

[29] See part B.1.4

When one computes all these possible decompositions in a table, we can assign to each number of i-clusters its probabilities using independency of i-storm clusters, $i \in [\![1, n]\!]$, and the fact that $X_i \sim Poisson\big(\lambda P_C(i)\big)$. Then the summation of all the decompositions gives the wanted probability $P_X(x)$.

This formula cannot be used for manual calculation. However, to understand it gives an idea of how one could implement it. A part of the appendices describe how simple it is to implement it in VBA Excel. This method is the one used for the tests made in part C.4.2.

B.3.3. Alternative definition

We already stated that the main default of the model representation $(CRM.1)$ was that clusters C_i can produce no storm. This point is important for the estimation of storm per cluster distribution (see B.3.1). The reason is that no data can be given for clusters leading to no storm. The idea is then to rewrite the model representation into an equivalent one that uses a zero truncated-binomial distribution for storms per clusters:

$$X = \sum_{i=1}^{Y'} C_i' \qquad (CRM.3)$$

where $Y' \sim Poisson(\lambda')$ with $\lambda' = \lambda P(C > 0) = \lambda[1 - (1-p)^n]$ and C_i' are truncated binomial distributed:

$P_{C'}(i) = \frac{P_C(i)}{P(C>0)} = \frac{C_n^i p^i (1-p)^{n-i}}{1-(1-p)^n}$ for $i \in [\![1, n]\!]$

Let us prove both forms are equivalent. By using the initial definition, we have:

$$\begin{aligned}
m_X(u) &= m_Y \circ m_C(u) \\
&= exp\{\lambda(\textstyle\sum_{i=0}^{n} P_C(i) u^i - 1)\} \\
&= exp\left\{\lambda P(C>0)\left(\textstyle\sum_{i=1}^{n} \frac{P_C(i)}{P(C>0)} u^i + \frac{P_C(0)}{P(C>0)} - \frac{1}{P(C>0)}\right)\right\} \\
&= exp\{\lambda'(\textstyle\sum_{i=1}^{n} P_{C'}(i) u^i - 1)\} \\
&= m_{Y'} \circ m_{C'}(u)
\end{aligned}$$

Thanks to this expression, we will be able to work the data in which clusters lead to at least 1 storm. Estimation will hence be possible and we will also obtain the equivalent parameters of model $(CRM.1)$.

B.3.4. Estimation

We will use the Poisson distribution associated with the ZT-binomial distribution CRM in this part.

1) Separated estimations of parameters

Let us consider the $(CRM.3)$ model in which all clusters produce at least 1 storm.
We consider a n-sample standing for n different years.

The same way as described in part B.1.5 we obtain the following ML estimator for cluster parameter λ':

$$\widehat{\lambda'} = \frac{\sum_{i=1}^{n} y_i'}{n}$$

Let us remind that the number of storms per cluster C_i' has a ZT-binomial(n,p) distribution. The parameter to be estimated here is p while n is previously fixed. The ML equation is given by:

$$\frac{\partial l_{C'|Y'}}{\partial p}(x) = 0 \Leftrightarrow \frac{\partial}{\partial p} ln\left(\prod_{i=1}^{y} \frac{C_n^{x_i} p^{x_i}(1-p)^{n-x_i}}{1-(1-p)^n}\right) = 0$$

$$\Leftrightarrow \frac{\partial}{\partial p} \sum_{i=1}^{y}\{x_i ln(p) + (n-x_i) ln(1-p) - ln(1-(1-p)^n)\} = 0$$

$$\Leftrightarrow \sum_{i=1}^{y}\left\{\frac{x_i}{p} - \frac{n-x_i}{1-p} - \frac{n(1-p)^{n-1}}{1-(1-p)^n}\right\} = 0$$

$$\Leftrightarrow \frac{\bar{x}}{p} - \frac{n-\bar{x}}{1-p} - \frac{n(1-p)^{n-1}}{1-(1-p)^n} = 0$$

where $\bar{x} = \frac{1}{y}\sum_{i=1}^{y} x_i$

$$\Leftrightarrow \bar{x}(1-p)(1-(1-p)^n) - (n-\bar{x})p(1-(1-p)^n) - np(1-p)^n = 0$$

$$\Leftrightarrow \bar{x}q(1-q^n) - (n-\bar{x})(1-q)(1-q^n) - n(1-q)q^n = 0$$

$$\Leftrightarrow \bar{x}q(1-q^n) - (1-q)[(n-\bar{x})(1-q^n) + nq^n] = 0$$

$$\Leftrightarrow \bar{x}q(1-q^n) - (1-q)[n-\bar{x}+\bar{x}q^n] = 0$$

$$\Leftrightarrow -\bar{x}q^n + nq + \bar{x} - n = 0$$

$$\Leftrightarrow g(q) = 0$$

where $g: q \longmapsto q^n - \frac{n}{\bar{x}}q + \frac{n}{\bar{x}} - 1$ and $q \in]0,1[$.

Does a solution always exist?
Let us study the function g to see if it is possible to have $g(q) = 0$ for any $q \in]0,1[$. g is a $\mathcal{C}^\infty$-function, so its derivative exists and the function is continuous. Its derivative is given by the function $g': q \longmapsto nq^{n-1} - \frac{n}{\bar{x}}$ and:

$$\forall q \in]0,1[: \begin{cases} g'(q) > 0 \\ g'(q) = 0 \\ g'(q) < 0 \end{cases} \Leftrightarrow \begin{cases} q > (\bar{x})^{\frac{-1}{n-1}} \\ q = (\bar{x})^{\frac{-1}{n-1}} \\ q < (\bar{x})^{\frac{-1}{n-1}} \end{cases}$$

Moreover $(\bar{x})^{\frac{-1}{n-1}} \in]0,1[\Leftrightarrow \bar{x} > 1$. This is true because all truncated values are bigger than one. We just have to check in the data that at list one value is strictly bigger than 1 to have $\bar{x} > 1$. If it is not the case, there would be no point to fit constant data with a ZT-binomial distribution.

Let us draw the analysis table of function g:

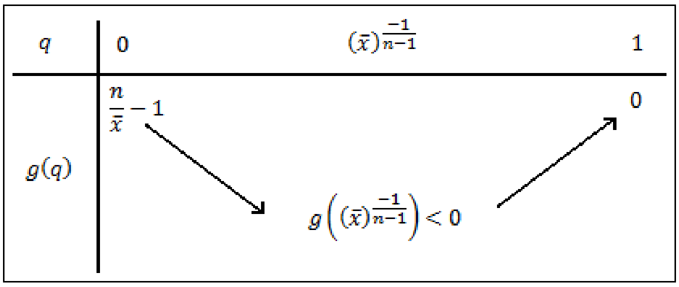


Table B.3: Variation of function g

The important point to state whether a solution exists is to determine the sign of $\frac{n}{\bar{x}} - 1$.
$\frac{n}{\bar{x}} - 1 > 0 \Leftrightarrow n > \bar{x}$ where the parameter n stands for the biggest integer the binomial(n, p) distribution can reach. It seems also natural that the empirical average $\bar{x}$ is strictly lower than n, otherwise all observed values would equal n. Again, as above, there would be no point to fit constant data to a ZT-binomial distribution.

As g is a real-valued continuous function, using the intermediate value theorem proves it exists one unique $q^* \in \left]0, (\bar{x})^{\frac{-1}{n-1}}\right[$ such that $g(q^*) = 0$. This q^* gives us the solution of the likelihood equation $q^* = 1 - p^*$

2) Estimation of joint parameters

Let us use $(CRM.1)$with fixed Binomial parameter $n = 3$:
We want to obtain an estimator by the method of moment. We first obtain:

$$\begin{aligned} E[X] &= E\big[E[X|Y]\big] \\ &= E\big[YE[C_1]\big] \\ &= E[Y]E[C_1] \\ &= 3\lambda p \end{aligned}$$

and

$$\begin{aligned} V(X) &= E[V(X|Y)] + V(E[X|Y]) \\ &= E[YV(C_1)] + V(YE[C_1]) \\ &= E[Y]V(C_1) + V(Y)E[C_1]^2 \\ &= 3\lambda p(1 + 2p) \end{aligned}$$

So $\frac{V(X)}{E[X]} = 1 + 2p \Rightarrow \hat{p}_{MM} = \frac{1}{2}\left(\frac{\sigma^2}{\bar{x}} - 1\right)$ where σ^2 is the empirical estimator of the variance.

Once $\hat{p}_{MM}$ is obtained we can compute $\hat{\lambda}_{MM} = \frac{\bar{x}}{3\hat{p}_{MM}}$

This method of estimation will be the only one used for the Poisson-Binomial CRM. We will see in part C that computing the MLE has the disadvantage to be hard to write.

C Tests of the models

In this part, we will test the Poisson and the two extension models by using European wide data. We will define whether these two models are convenient only in comparison with the usual Poisson model. To do so, several estimations and goodness of fit tests will be developed.

C.1. Data

The data is taken from the ERA-40 analysis, which is available on the European Centre for Medium-Range Weather Forecasts website: http://www.ecmwf.int/.

The ERA-40 is more precisely a re-analysis of the global atmosphere and surface conditions for 45-years, over the 1957-2002 period. Many sources of the meteorological observations were used, including radio sound, balloons, aircraft, satellites, and scatter meters. This data was run through the ECMWF computer model at a 40km resolution. As the ECMWF's computer model is one of the more highly-regarded in the field of forecasting, many scientists take its reanalysis to have similar merit. The data is stored in GRIB format. The reanalysis was done in an effort to improve the accuracy of historical weather maps and aid in a more detailed analysis of various weather systems through a period that was severely lacking in computerized data. With the data obtained from reanalyzes, many of the more modern computerized tools for analyzing storm systems can be utilized, at least in part, because of this access to a computerized simulation of the atmospheric state. Some major NAT-CAT pricing tools as RMS' Risklink build their simulation engines thanks to the historical data provided by the ERA-40.

The data can be directly downloaded from this page: http://data-portal.ecmwf.int/data/d/era40_daily/ where the user selects the wanted data by choosing:

- The time period: from 1957 to 2002;
- The hours : 0 am, 6 am, 12 am and 6 pm;
- The parameters: this includes U and V wind component[30] or even snowfall, evaporation, etc.

Once a selection is made, the use of data as a non-specialist of meteorological tool is tricky. First, it is hard to find software that can correctly read and convert the data expressed in GRIB format. Regardless of this technical aspect, how can one correctly define a storm only with wind speed components? This question does not intervene inside the actuary's field. It must first be tackled by an expert.

[30] the two space components for wind speed

This is the reason why we finally choose to build our analysis on a previous work done by Meteo Schweiz experts[31]. By using atmospheric indexes, one can observe severe storms at every wind speed peak that go beyond a certain threshold. It results a list of about 160 biggest European storms whose occurrence dates are given. We will base our tests on that list.

A table of the data can be viewed in the appendix.

C.2. Poisson Model

In this part, we want to fit the classic Poisson model that is applied in most of the NAT-CAT pricing tools. From the storm information taken from Meteo Schweiz, we draw a list of number of storms per year:

Year	Total storms	Year	Total storms	Year	Total storms	Year	Total storms
2002	3	1990	11	1978	5	1966	0
2001	0	1989	8	1977	4	1965	1
2000	4	1988	9	1976	5	1964	0
1999	4	1987	2	1975	0	1963	1
1998	5	1986	6	1974	4	1962	3
1997	6	1985	3	1973	7	1961	1
1996	1	1984	5	1972	2	1960	2
1995	3	1983	6	1971	3	1959	1
1994	4	1982	4	1970	2	1958	0
1993	7	1981	3	1969	2	1957	1
1992	5	1980	2	1968	1		
1991	4	1979	4	1967	4		

Figure C.2.a: Annual number of storm

From Figure C.2.a, we can count a total number of 158 storms that occurred within a 46 years long time period. We obtain a simple estimation of the parameter by the maximum likelihood method: the MLE is given by $\hat{\lambda} = \bar{x} = 3.4348$. We can now compare the $Poisson(3.4348\ )$ distribution with the empirical one on the next graph:

[31]DELLA-MARTA P., MATHIS H., FREI C., LINIGER M., KLEINN J., APPENZELLER C. (2009), *The return period of wind storms over Europe*, International Journal of Climatology

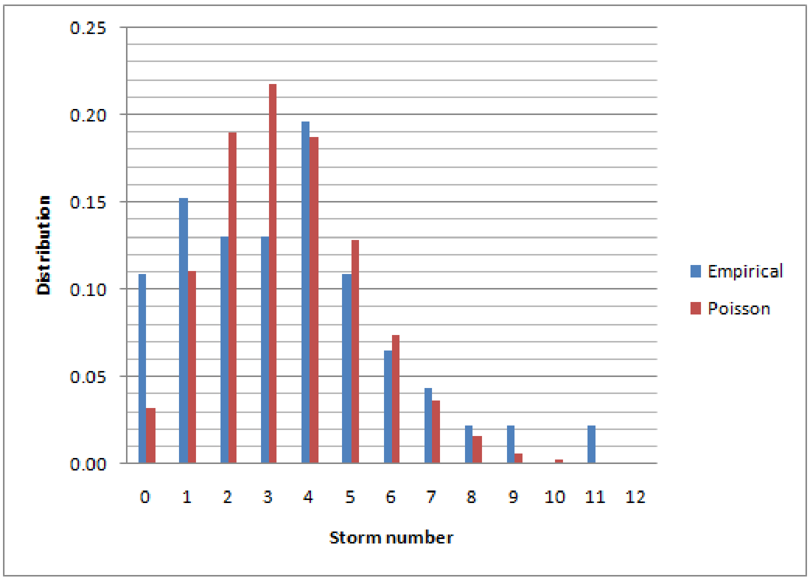


Figure C.2.b: Poisson distributed annual storms

Figure C.2.b allows us to compare the empirical and the $Poisson(3.4348\ )$ distribution for integers going from 0 to 12. The first remark we can make is that the Poisson distribution is not adapted for values going to 0 to 4. It clearly underestimates probabilities of having 0 and 1 storm and overestimates vales 2 and 3. To have a quantitative idea of this difference, we will use the Chi-square goodness of fit test for all the models we presented in part B.

To do a Chi-square goodness of fit test from a data sample, we divide this sample into subsets $i \in [\![1,r]\!]$. We use these subsets, also called classes, to compare the empirical values $(N_i)_{i\in[1,r]}$ with the theoretical ones $(np_i)_{i\in[1,r]}$. n denotes the sample size and p_i the theoretical probability.

We then define the chi-square statistic by:

$$D^2 = \sum_{i=1}^{r} \frac{(N_i - np_i)^2}{np_i}$$

The critical value for this test comes from the chi-square distribution with $r-1$ degrees of freedom. We can then obtain a p-value that characterizes the goodness of it.

For more details about the test properties, see appendix A.3.

In our Poisson model case we use the following characteristics:

- Sample size $n = 46$ (years from 1957 to 2002);
- N_i is the number of years where i storms occur, $i \in [\![0,5]\!]$;
- N_6 is the number of years where more than 6 storms occur;

- p_i is the $Poisson(3.4348)$ probability for i storms.

We obtain the following table in which we can compare numerically the difference from both distributions:

Distribution	N_i	Classes	N_i	np_i	D^2 sum terms
0	5	0	5	1.4827	**8.3439**
1	7	1	7	5.0927	0.7143
2	6	2	6	8.7462	0.8623
3	6	3	6	10.0138	**1.6088**
4	9	4	9	8.5988	0.0187
5	5	5	5	5.9070	0.1393
6	3	≥6	8	6.1589	0.5504
7	2				
8	1				
9	1				
10	0				
≥11	1				

Table C.2.c: Chi-square for Poisson distributed annual storms

We can observe that the main source of difference is the 0 value. The values that are bigger than 4 give relative small differences. By assuming $D^2 \sim \chi^2(6)$, we have the final result given by Table C.2.d:

D^2	p-value
12.2377	5.69%

Table C.2.d: p-value for Poisson distributed annual storms

The p-value is clearly too low to consider the Poisson model as adapted for windstorm modeling in Europe. Since this model is mostly used in pricing tool nowadays, this p-value will be considered as the reference for all our following tests. It seems natural that the following models will give better results as they have two parameters. They will be considered as relevant only if their p-values are really much higher than 5.69%.

C.3. Negative binomial model

C.3.1. Tests of single parameters

In this section, we compute estimations and test the distributions for each Poisson and Logarithmic parameter separately, i.e. without taking the other parameter into account. We will first focus on the cluster Poisson distribution before shedding a light on the storm logarithmic distribution.

1) Poisson distributed clusters

In this section we count the number of storm clusters per year. However, defining the way a cluster can be recognized in the data is a delicate point: our data does not contain any information on the atmospheric conditions that lead to the storms. I asked Munich Re Meteorology experts what they think about the question. Their answer is that independence between storms is assumed when there is a 3 to 5 days period in between. So I classified storms inside a same cluster when they had a maximum of 5 days in between. We can then classify the data as shows the following table:

Year	Clusters	Total storms	Year	Cluster number	Total storms
2002	3	3	1979	2	4
2001	0	0	1978	5	5
2000	3	4	1977	3	4
1999	3	4	1976	4	5
1998	4	5	1975	0	0
1997	4	6	1974	3	4
1996	1	1	1973	5	7
1995	2	3	1972	2	2
1994	3	4	1971	2	3
1993	5	7	1970	2	2
1992	4	5	1969	2	2
1991	2	4	1968	1	1
1990	6	11	1967	2	4
1989	6	8	1966	0	0
1988	7	9	1965	1	1
1987	2	2	1964	0	0
1986	4	6	1963	1	1
1985	3	3	1962	2	3
1984	5	5	1961	1	1
1983	5	6	1960	2	2
1982	3	4	1959	1	1
1981	3	3	1958	0	0
1980	2	2	1957	1	1

Table C.3.a: Annual storms & clusters numbers

From Table C.3.a we can count a number of 122 clusters versus a total number of 158 storms between 1957 and 2002. There are 24 years over 46 in which the number of annual storms is strictly larger than the number of clusters. These figures are a good indication on the relevancy of the cluster assumption.

The same way we did in part C.2 we compute the MLE for the Poisson parameter of number of clusters. The MLE is obtained by writing $\hat{\lambda} = \bar{x} = 2.6522$. We can now compare the $Poisson(2.6522)$ distribution with the empirical one on the next graph:

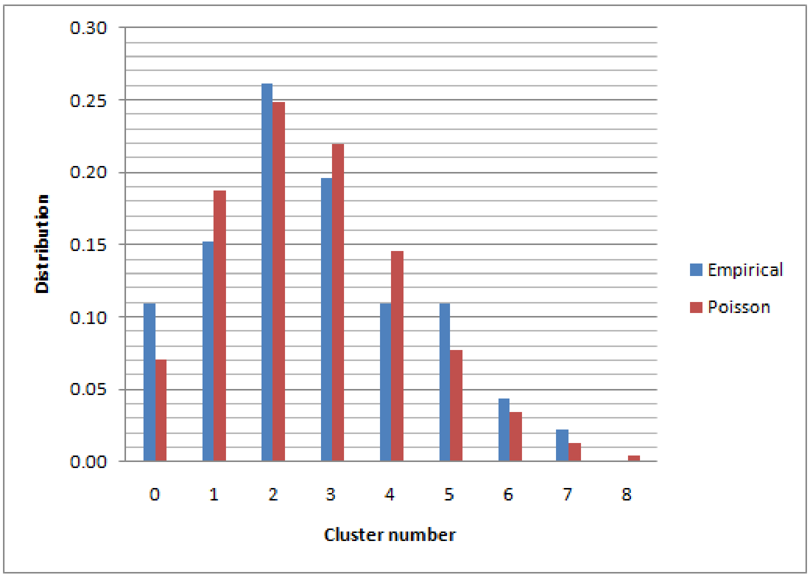


Figure C.3.b: Poisson distributed cluster

The difference between both empirical distributions is much different from the one we observed in Figure C.2.b form Poisson distributed windstorms. First, values from 0 to 3 give relatively smaller differences than previously but values larger than 4 are more impacted. The shape of both distributions seems to fit better though. Let us confirm it by using the Chi-square distribution of fit test.

In this specific case, the assumptions are:

- Sample size $n = 46$ (years from 1957 to 2002);
- N_i is the number of years where i clusters occur, $i \in [\![0,4]\!]$;
- N_5 is the number of years where more than 5 clusters occur;
- p_i is the $Poisson(2.6522)$ probability for i clusters.

Table C.3.c presents the empirical and theoretical number of cluster per class:

Distribution	N_i	Classes	N_i	np_i	D^2 sum terms
0	5	0	5	3.2429	**0.9521**
1	7	1	7	8.6007	0.2979
2	12	2	12	11.4053	0.0310
3	9	3	9	10.0830	0.1163
4	5	4	5	6.6854	0.4249
5	5	≥5	5	5.9827	0.1614
6	2				
≥7	1				

Table C.3.c: Chi-square for Poisson distributed cluster

The major source of error is still value 0. However, the differences in all other class are smaller than in Table C.2.c. By assuming $D^2 \sim \chi^2(5)$, we finally obtain the result:

D^2	p-value
1.9836	85.14%

Table C.3.d: p-value for Poisson distributed cluster

Table C.3.d proves that Poisson distributed clusters can be an acceptable assumption since its p-value reaches 85.14%. In other words, we can consider that climactic conditions that lead to storm formation happen independently from each other. Let us now verify whether the logarithmic distribution is adapted to the number of storms that are created from a same cluster.

2) Logarithmic distributed storms per cluster

The number of storms created is counted for every 122 clusters. Among these 122 clusters, 92 produced 1 storm, 24 produced 2 storms and 6 led to 3 storms. We do not observe more than 3 storms per cluster and this will surely provide a shift between the theoretical and empirical distribution. This is one disadvantage of the logarithmic distribution.

We presented in part B.2.2 that the MLE exists for this particular distribution. We although need to use a numerical solver to find it. By using the solver from Excel, we obtain $\hat{p} = 0.3914$. Figure C.3.e compares the $Logarithmic(0.3914)$ and the empirical distributions:

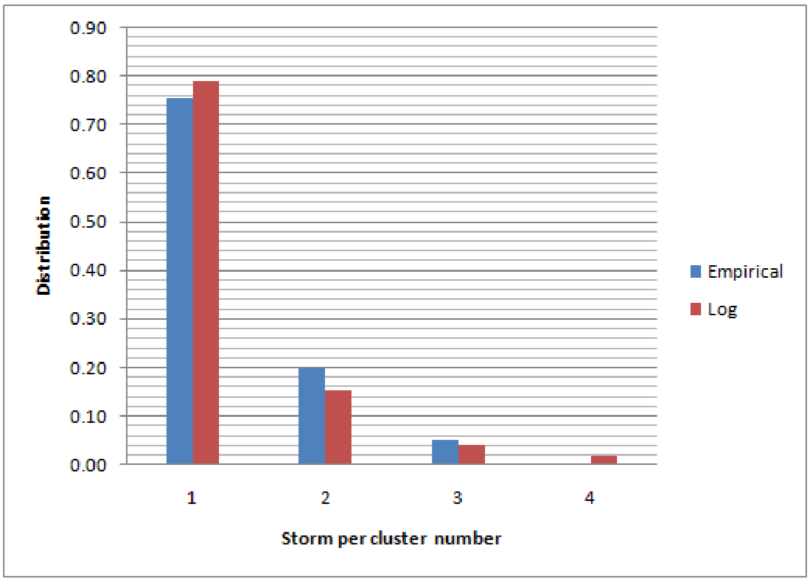


Table C.3.e: Logarithmic distributed storms per cluster

From the previous figure we can notice that the main absolute difference from both distributions does not come from the probability of having more than 4 storms per clusters, but from the probability of having 2 storms. This will be clearly underlined by the Chi-square goodness of fit table C.3.f.

The assumptions taken for this fit test are:

- Sample size $n = 122$ (number of clusters from 1957 to 2002);
- N_i is the number of clusters in which i storms occur, $i \in [\![1,3]\!]$;
- N_4 is the number of clusters where more than 4 storms occur;
- p_i is the $Logarithmic(0.3914)$ probability for i storms cluster.

Distribution	N_i	Classes	N_i	np_i	D^2 sum terms
1	92	1	92	96.1548	0.1795
2	24	2	24	18.8187	**1.4265**
≥3	6	≥3	6	7.0265	0.1500

Table C.3.f: Chi-square for logarithmic distributed storms per cluster

Probabilities of having at least 3 storms per cluster are grouped inside the same class (we must insure a minimum number of 5 storms per class). This is the reason why the difference is so small for this particular class. Class number two has a major impact on the test. By assuming $D^2 \sim \chi^2(2)$, we obtain the following result:

D^2	p-value
1.7560	41.56%

Table C.3.g: p-value for logarithmic distributed storms per cluster

The logarithmic distribution provides a medium p-value. Two aspects can be the root of such a result: either the logarithmic distribution is not convenient enough to model the annual number of storm per cluster, or, the way we decided to define a cluster in the beginning of part C.3 does not reflect the reality. At this stage, it is hard to define which the cause is. To do so, we will do the estimation and the goodness of fit test for parameters taken jointly. A shift between the estimators form both separated and joint perspectives will be an indicator of the low quality of our cluster definition.

C.3.2. Test of joint parameters

In this section, we consider the model represented by $(NB.2)$. Contrary to part C.3.1, we will estimate the parameters jointly. This means that we will make the computation directly from the distribution of the annual number to storms. There is no need to decompose storms into clusters and storms per cluster, so, the cluster definition we wrote in the beginning of part C.3.1 is no longer taken into account. To estimate both parameters, we will proceed the method of moments and then to the maximum likelihood estimation.

1) Method of Moments

By using the formula described in part B.2.1 for the method of moments, we can find $(\hat{r}_{NB}, \hat{p}_{NB}) = (4.1238, 0.5456)$. Both negative binomial and empirical distributions are shown on the following figure:

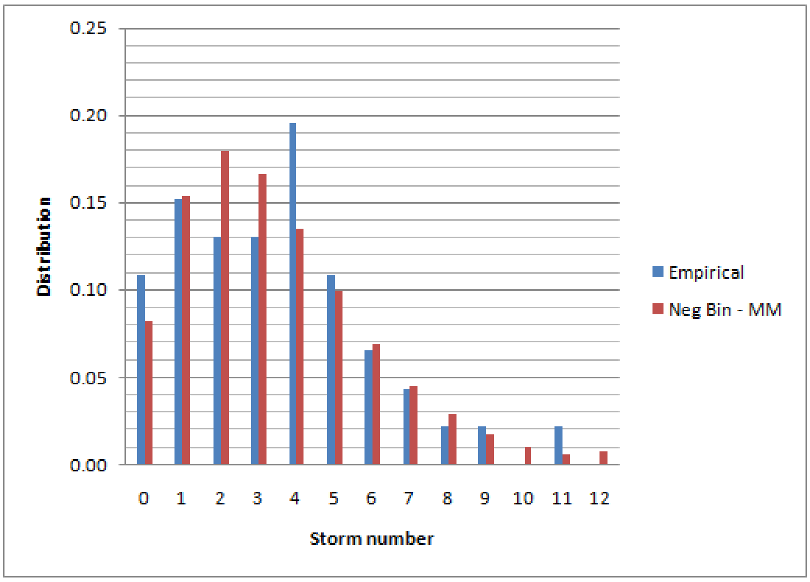


Figure C.3.h: Negative binomial distributed, annual storms

At a first glance, Figure C.3.h displays two distributions that are quite close one to each other. Differences for integers between 0 and 3 are much smaller than for the Poisson distribution at Figure C.2.b. However, the negative binomial distribution increases the dissimilarity for value 4. On the whole, the negative binomial distribution seems to fit better the empirical distribution compared to the Poisson model. Let the Chi-square goodness of fit test confirm this impression.

To make the test, we make the following hypothesizes:

- Sample size$n = 46$ (years from 1957 to 2002);
- N_i is the number of years where i storms occur, $i \in [0,5]$;
- N_6 is the number of years where more than 6 storms occur;
- p_i stands for the $NegativeBinomial(r = 4.1238, p = 0.5456)$ probability for i storms.

This leads to the following table:

Distribution	N_i	Classes	N_i	np_i	D^2 sum terms
0	5	0	5	3.7810	0.3930
1	7	1	7	7.0854	0.0010
2	6	2	6	8.2487	0.6130
3	6	3	6	7.6515	0.3565
4	9	4	9	6.1924	**1.2730**
5	5	5	5	4.5720	0.0401
6	3	≥6	8	8.4690	0.0260
7	2				
8	1				
9	1				
10	0				
≥11	1				

Table C.3.i: Chi-square for negative binomial distribution

Table C.3.i confirms a better fit for the Negative Binomial distribution. The biggest distinctness between the distributions' values comes from integer 4 where the D^2 coefficient is 1.2730. Compared to Table C.2.c, this number is even smaller than the one that concern Poisson model's class number 3. The difference of 8.3439 we had on value 0 is now reduced to 0.3930. By assuming $D^2 \sim \chi^2(6)$, Table C.2.d gives a p-value of:

D^2	p-value
2.7025	84.51%

Table C.3.j: p-value for negative binomial distribution

At this stage, it is interesting to compare the p-value of 84.51% for the annual storm distribution we both p-values of 85.14% and 41.56% we have respectively for number of clusters and number of storms per cluster. The medium p-value of 41.56% for storms per cluster has a relatively small impacted on the whole model. Joint parameters and separated parameters can provide an explanation. This point will be developed further in the sum up part C.5.

2) Maximum likelihood estimation

In this part, we develop the same points as before but by using first the maximum likelihood estimation. As presented in part B.2.1, the maximum likelihood parameters $(\hat{r}, \hat{p})$ verify the following equations:

$$\begin{cases} ln\left(1+\frac{\bar{x}}{\hat{r}}\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{x_i-1}\frac{1}{\hat{r}+j} \\ \hat{p} = \frac{\hat{r}}{\bar{x}+\hat{r}} \end{cases}$$

By solving the system numerically on Excel, we find: $(\hat{r}_{NB}, \hat{p}_{NB}) = (3.8005, 0.5253)$. The relative negative binomial distribution is compared to the empirical one on Figure C.3.k:

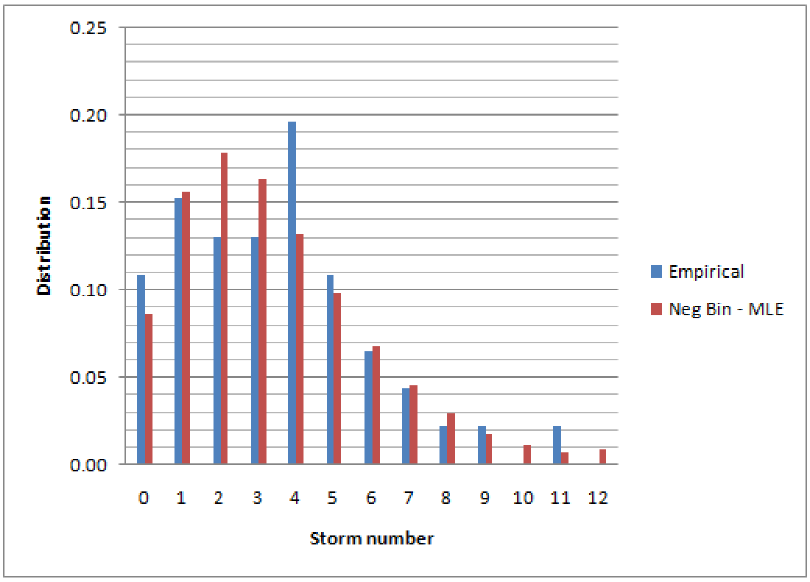


Figure C.3.k: Negative binomial distributed, annual storms

Method of moments and maximum likelihood give different estimators. The MM gives $(\hat{r}_{MM}, \hat{p}_{MM}) = (4.1238, 0.5456)$ whereas the MLE makes $(\hat{r}_{MM}, \hat{p}_{MM}) = (3.8005, 0.5253)$. However, Figure C.3.k and Figure C.3.h show very close values. The Chi-square test can confirm it:

Its characteristics are:

- Sample size $n = 46$ (years from 1957 to 2002);
- N_i is the number of years where i storms occur, $i \in [\![0,5]\!]$;
- N_6 is the number of years where more than 6 storms occur;
- p_i *is* the $NegativeBinomial(r = 3.8005, p = 0.5253)$ probability for i storms.

Distribution	N_i	Classes	N_i	np_i	D^2 sum terms
0	5	0	5	3.9818	0.2604
1	7	1	7	7.1840	0.0047
2	6	2	6	8.1859	0.5837
3	6	3	6	7.5137	0.3050
4	9	4	9	6.0643	**1.4212**
5	5	5	5	4.4913	0.0576
6	3	≥6	8	8.5789	0.0391
7	2				
8	1				
9	1				
10	0				
≥11	1				

Table C.3.l: Chi-square for negative binomial distribution

Table C.3.l gives approximately the same differences as seen for MM estimators in table C.3.i. Nevertheless, we can observe that class number 4 gives a value of 1.4212, which is bigger than table C.3.i. On the other hand, the mismatch on class 0 is reduced to 0.2604. By assuming $D^2 \sim \chi^2(6)$, we finally obtain a p-value of:

D^2	p-value
2.6716	84.88%

Table C.3.m: p-value for negative binomial distribution

As in table C.3.j, the whole model has a p-value of 84.88% despite its medium fit for logarithmic storms per cluster. Between both couple of estimators, one could prefer to choose the MLE because of its p-value (84.88% versus 84.51% in table C.3.j). Another reason for this choice is that class 0 is less impacted by dissimilarities: this probability is the most important for policies such as CAT XL treaties in reinsurance because they usually cover only one event.

C.4. Poisson-Binomial collective risk model

Before starting, we first need to decide which parameter n we should use for a $Binomial(n, p)$ distribution. We saw in the data that clusters could not produce more than 3 storms. We will then use the $Binomial(3, p)$ distribution for all the following parts.
We still consider the clusters as atmospheric situations that can lead to a group a storm with 5 days in between period. So a cluster can give no storm at all. We explained in part B.3.2 that a cluster which produces zero storm cannot be read in the data. So we introduced the alternative definition $(CRM.3)$ in which clusters are Poisson distributed with a lower rate, but they necessarily create one storm, the number of storms per cluster has a ZT-binomial distribution.

C.4.1. Tests of single parameters

As already developed in part C.3.1, we will start by calculating the model's parameters separately.
There is no need to make estimation and a test for the annual number of clusters because we still consider it as Poisson distributed. Only the number of storms per cluster will be developed in this section.

1) Poisson distributed clusters

As seen in part C.3.1, the MLE is $\widehat{\lambda'} = \bar{x} = 2.6522$.
This model has a p-value of 85.14% (see Table C.3.d for more details).

2) ZT-binomial distributed storms per cluster

As described in part B.3.4, the maximum likelihood estimator $\hat{p}$ verifies the equation:

$$\frac{\hat{q}^n - 1}{\hat{q} - 1} = \frac{n}{\bar{x}}$$

Where $\hat{p} = 1 - \hat{q}$ and n are the binomial parameters. We decided to fix $n = 3$ at the beginning of section C.4. We also saw in part B.3.4 that the above equation always has a solution. By computing it with the solver of Excel, we find $\hat{p} = 0.2484$. Figure C.4.a draws the $ZT - Logarithmic(0.2484)$ and the empirical distributions:

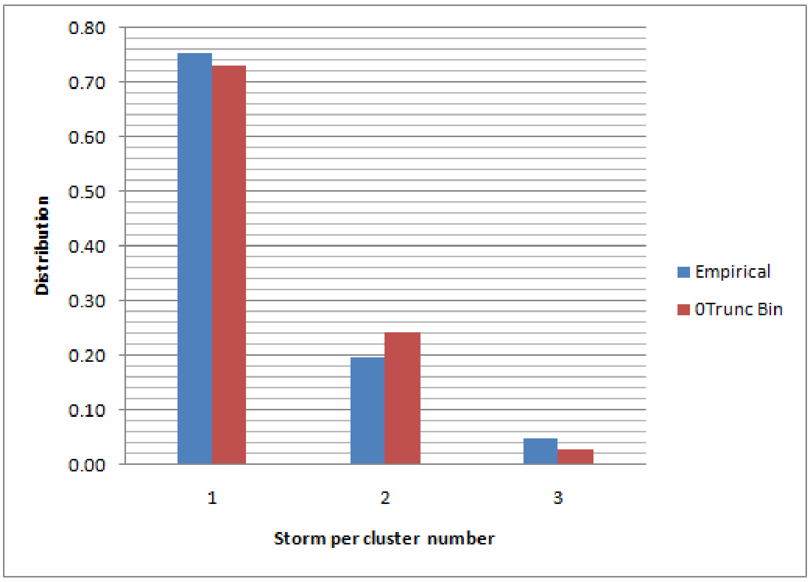


Figure C.4.a: ZT-binomial distributed storms per cluster

The zero-truncated distribution looks less adapted than the logarithmic one visible on Figure C.3.e. The relative difference is the biggest for the value 3 where the empirical distribution is twice as much as the ZT-binomial. This is confirmed by the test visible in Table C.4.b.

For this Chi-square test, we take:

- Sample size $n = 122$ (number of clusters from 1957 to 2002);
- N_i is the number of clusters in which i storms occur, $i \in [\![1,3]\!]$;
- p_i is the $ZT - Binomial\ (3, 0.2484)$ probability for i storms per cluster.

It leads to the following table:

Distribution	N_i	Classes	N_i	np_i	D^2 sum terms
1	92	1	92	89.2502	0.0847
2	24	2	24	29.4996	**1.0253**
3	6	3	6	3.2501	**2.3266**

Table C.4.b: Chi-square for ZT-binomial distributed storms per cluster

The small number of class gives each difference an important impact on the test's result. As observed in Figure C.4.a, the biggest shift comes from the third class. By assuming $D^2 \sim \chi^2(2)$, we obtain:

D^2	p-value
3.4366	17.94%

Table C.4.c: p-value for ZT-binomial distributed storms per cluster

We can remark that the p-value for the ZT-binomial distribution is much lower than the one given by Table C.3.g for the logarithmic distribution (41.56%). We cannot conclude that the Poisson-Binomial model is less adapted than the negative binomial distribution though. The reason is that our calculations are still lying on the assumptions of a cluster definition, which was made at the start of part C.3. We can only state that the negative binomial distribution is the most adapted for this particular definition of cluster. To have a wider idea on the question, we should compare the models globally i.e. by using estimations of joint parameters.

C.4.2. Joint parameters test

In this section, we consider the model given by definition $(CRM.3)$.

The goodness of fit test will be directly made on the number of storms distribution. To do so, we need write the probability for each integer from $(CRM.3)$. We know there is a maximum of 11 storms per annum in our database and that clusters can produce 1, 2 or 3 storms. The idea is to decompose for each number of storms observed per year (this number belongs to $[\![0,11]\!]$), the different type of clusters it can be made of. For example, we observe for the year 1995, 3 different storms. They can be produced either by one 3-storms cluster, one 2-storms cluster and one 1-storm cluster or three 1-storm clusters. By doing this decomposition, we can allocate probabilities to each situation. Then, we sum all the probabilities corresponding to a i storm situation. See part B.3.2 for more details.

1) Maximum Likelihood

We can first try to see is calculating the MLE is possible in this special case. For a n-sample, the log-likelihood equation is given by:

$$ln\left(\prod_{i=1}^{n} f_{N_i}(x_i)\right) = \sum_{i=1}^{n} \sum_{k_3=0}^{\left\lfloor \frac{x_i}{3} \right\rfloor} \sum_{k_3=0}^{\left\lfloor \frac{x_i-3k_2}{2} \right\rfloor} \ln\left(P_{X_3}(k_3) P_{X_2}(k_2) P_{X_1}(x_i - 3k_3 - 2k_2)\right)$$

After several lines of calculations, we can write:

$$ln\left(\prod_{i=1}^{n} f_{N_i}(x_i)\right) = \sum_{i=1}^{n} \sum_{k_3=0}^{\left\lfloor \frac{x_i}{3} \right\rfloor} \sum_{k_3=0}^{\left\lfloor \frac{x_i-3k_2}{2} \right\rfloor} \{\ln\left(c_{x_i,k_2,k_1}\right) + ln\left(f_{x_i,k_2,k_1}(\lambda)\right) + ln\left(g_{x_i,k_2,k_1}(p)\right) + ln\left(h_{x_i,k_2,k_1}(\lambda, p)\right)\}$$

where:

$$c_{x_i,k_2,k_1} = \frac{3^{x_i-3k_3-k_2}}{(x_i-3k_3-2k_2)!k_2!k_3!}$$

$$f_{x_i,k_2,k_1}(\lambda) = \lambda^{x_i-2k_3-k_2}$$

$$g_{x_i,k_2,k_1}(p) = p^{x_i}(1-p)^{2x_i-6k_3-3k_2}$$

$$h_{x_i,k_2,k_1}(\lambda, p) = e^{-\lambda(1-(1-p)^3)}$$

This problem is hard to write compared to the method of moments that is easily computable because of the CRM form. Furthermore, both methods usually give similar results. We will only focus on the MM.

2) Method of moments

We do the same analysis but using the method of moments that was previously described in part B.3.2. We find the following estimators: $(\tilde{\lambda}, \tilde{p}) = (2.7492, 0.4165)$. Figure C.4.e gives the shapes of both empirical and Poisson-Binomial CRM:

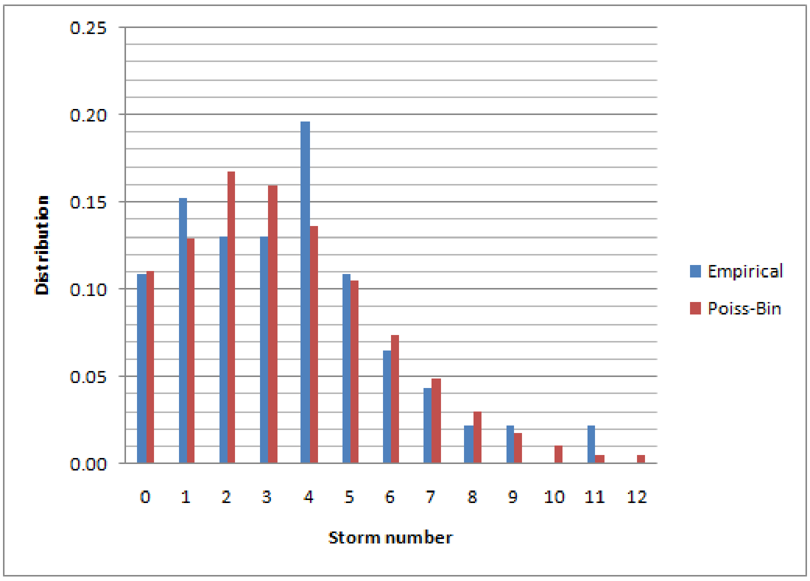


Figure C.4.e: Poisson-binomial CRM annual storms

Compared to all the previous models visible on figures C.2.b, C.3.h and C.3.k, figure C.4.e shows that the Poisson-Binomial CRM seems the most adapted to the empirical distribution. A major clue of its goodness of fit is the exactitude of probability for value 0. The biggest dissimilarity can be observed for integer 4, the same as the negative binomial distribution. Let us see how it impacts the goodness of fit test.

For the test, we take:

- Sample $n = 46$ (years from 1957 to 2002);
- N_i is the number of years where i storms occur, $i \in [\![0,5]\!]$;
- N_6 is the number of years where more than 6 storms occur;
- p_i is the probability for i storms per year.

The Chi-square's coefficients are given in the following table:

Distribution	N_i	Classes	N_i	np_i	D^2 sum terms
0	5	0	5	5.0821	0.0013
1	7	1	7	5.9441	0.1876
2	6	2	6	7.7183	0.3826
3	6	3	6	7.3262	0.2401
4	9	4	9	6.2488	**1.2113**
5	5	5	5	4.8275	0.0062
6	3	≥6	8	8.8530	0.0822
7	2				
8	1				
9	1				
10	0				
≥11	1				

Table C.4.f: chi-square for Poisson-binomial CRM annual storms

Table C.4.f clearly shows that the Poisson-Binomial distribution is the most adapted to the data. Class number 4 gives a difference value of 1.2113 which is the lowest compared to the Binomial Negative distribution (value of 1.2730 for MM and 1.4212 for MLE, see table C.3.i and C.3.l) and the Poisson distribution (value of 1.6088, see table C.2.c). With this last model, we obtain the following p-value:

D^2	p-value
2.1111	90.92%

Table C.4.g: p-value for Poisson-binomial CRM annual storms

Despite the low p-value of 17.93% that we obtained in part C.4.1 for the test of ZT-Binomial storms per cluster, the global Poisson-Binomial CRM gives the best result with a p-value of 90.92%. Part C.3 already revealed the same situation between the partial and the global model. These dissimilarities between separated and joint estimations will be the main focus of part C.5.

C.5. Sum up

In parts C.3 and C.4, we noticed an important shift between separated and joint estimators, for both Negative Binomial and Poisson-Binomial models. The role of this section is to give an explanation of this difference of tests' results.

The following table shows the previously found results for single estimators. We use these values of parameters and plug in them into the joint model. The new p-values can be read in the **left column**:

<table>
<tr><td></td><td colspan="3">Separated estimators</td><td colspan="2">p-value</td></tr>
<tr><td rowspan="3">Negative Binomial</td><td colspan="3">Maximum Likelihood</td><td>Joint Model</td><td>Indep. Model</td></tr>
<tr><td colspan="2">$r_{NB} = \frac{-\lambda_{Poiss}}{\ln(q_{log})}$</td><td>$p_{NB} = q_{log}$</td><td rowspan="2">79.87%</td><td>cluster 85.14%</td></tr>
<tr><td colspan="2">5.3402</td><td>0.6085</td><td>storm 41.56%</td></tr>
<tr><td rowspan="3">Poisson-Binomial</td><td colspan="3">Method of Moments</td><td>Joint Model</td><td>Indep. Model</td></tr>
<tr><td>λ</td><td>λ'</td><td>p</td><td rowspan="2">83.09%</td><td>cluster 85.14%</td></tr>
<tr><td>4.6091</td><td>2.6522</td><td>0.2484</td><td>storm 17.94%</td></tr>
</table>

Table C.5.a: p-values from separated estimators

Another table (table C.5.b) presents the p-values obtained with joint estimators. The joint model parameters previously computed are plugged-in into the independent model. The new p-value can be read in the **right column**:

<table>
<tr><td></td><td colspan="3">Joint estimators</td><td colspan="2">p-value</td></tr>
<tr><td rowspan="3">Negative Binomial</td><td colspan="3">Maximum Likelihood</td><td>Joint Model</td><td>Indep. Model</td></tr>
<tr><td colspan="2">r_{NB}</td><td>p_{NB}</td><td rowspan="2">84.88%</td><td>cluster 85.08%</td></tr>
<tr><td colspan="2">3.8005</td><td>0.5253</td><td>storm 29.62%</td></tr>
<tr><td rowspan="3">Poisson-Binomial</td><td colspan="3">Method of Moments</td><td>Joint Model</td><td>Indep. Model</td></tr>
<tr><td>λ</td><td>λ'</td><td>p</td><td rowspan="2">90.92%</td><td>cluster 79.20%</td></tr>
<tr><td>2.7492</td><td>2.2029</td><td>0.4165</td><td>Storm 0.00%</td></tr>
</table>

Table C.5.b: p-values from joint estimators

The role of this table is to shed a light on the difference between the estimation based on independent and joint estimations. It shows that both parameters are not independent because according to the way they are estimated (jointly or separately), their values differ. However, it is interesting to see the impact of this difference. For instance, the p-value for the Negative Binomial model is the least impacted. Both separated and joint estimations give equivalent results, despite the significant change in the r parameter.

The Poisson-Binomial CRM is more impacted by this difference of estimation. It may seem surprising this model can give a global p-value of 90% whereas using the same estimator for storm per cluster gives a 0% p-value. This questions the way we defined clusters of storms in part C.3.1. We chose to define clusters as a five days in between group of storms that come from the same geographical area.

The joint parameter estimation directly acts on the annual number of storms and does not use the definition made in part C.3.1. So, the cluster process is taken into account to provide the best fit for annual storms, but we do not know how it is actually defined inside the optimization. In other words, clustering can be considered as a latent process because we do not know how to observe it.

The joint estimation gives a smaller cluster Poisson parameter and a bigger parameter for binomial storms per cluster than the separated estimations. This means that if we change the definition made in part C.3.1 by defining clusters in such a way that they contain more storms and happen less often, this would have provided better results for separated estimation.

D Pricing applications

The aim of this section is to show how the alternative models previously presented can be applied from a pricing tool that does not consider clusters of storms. We will first present the way we can link RMS' Risklink results with the new occurrence models.

D.1. Overall impact on pricing tools

Pricing can be done thanks to specialized tools like Risklink. We will use the OEP[32] curve to integrate our own occurrence model. The OEP curve gives the probability an event exceeds a certain threshold. We will only focus on severe storms, i.e. storms whose severity goes beyond a certain threshold.

The From Ground Up premium is defined as the expectation of loss for a given peril. By considering the storm occurrence independent from the severity and all the severities of storms as independent and identically distributed r.v.s, the premium π can be written as:

$$\pi = E[X]E[N_0]$$

Where X is the annual loss taken from the OEP curve. The OEP only considers the occurrence of either 0 or 1 storm. As we want to consider the possibility of several within the same year, we define the severity distribution for one storm by taking $X|one\ event\ occurs$.
N_0 is the number of storms which has a severity bigger than the threshold. Then, premium can be rewritten as:

$$\pi = E[X]\big(1 - P(N_0 = 0)\big)\frac{E[N_0]}{1-P(N_0=0)}$$

Where $E[X]\big(1 - P(N_0 = 0)\big)$ is directly computed from the OEP curves of the CAT reference tool:

$$E[X]\big(1 - P(N_0 = 0)\big) = \int_0^{+\infty} OEP(x)dx$$

We want to integrate the possibility of several events in a year by multiplying the result by $\frac{E[N_0]}{1-P(N_0=0)}$. This term that we note $\rho(N_0)$ depends on the assumed frequency distribution i.e. Poisson, negative binomial or Poisson-binomial CRM.

---

[32] Occurrence Exceedance Probability, see part A.1.4.6

This table gives the different expressions of ρ according to the underlying distribution:

Occurrence distribution	Poisson	Negative Binomial	Poisson-Binomial CRM
ρ	$\frac{\lambda}{1-e^{-\lambda}}$	$\frac{r(1-p)}{p(1-p^r)}$	$\frac{3\lambda p}{1-e^{-\lambda[1-(1-p)^3]}}$

Table D.1.a: Expression of ρ

With the estimators found in part C, it numerically gives:

Occurrence distribution	Poisson	Negative Binomial	Poisson-Binomial CRM
ρ	3.5492	3.7603	3.8614

Table D.1.b Values of ρ

The premium π is finally given by:

$$\pi(distribution) = \left\{\int OEP_{from\ tool}\right\}\rho(distribution).$$

The premium increase due to clusters modeling at a European scale is:

Premium	From Poisson to Negative Binomial	From Poisson to Poisson-Binomial CRM
Increase	5.95%	8.80%

Table D.1.c: relative premium increases due to cluster modeling

On a FGU loss perspective, the study suggests to increase the pricing tool's premium by 5.95% to 8.80% to take the clustering effect into account. However, this makes a sense only for portfolios with a Europe wide configuration. For specific ones, like the portfolios of a localized insurer, the following part D.2 describes how to precisely adapt the cluster effect.

D.2. Impact on portfolios of insurers

We want to adapt the parameters in such a way that they take the local scale information of insurers' portfolios into account. For that, one will proceed into two successive steps. First we will make an estimation of the cluster parameter and finally end up with the logarithmic or binomial parameter.

The first idea is to keep a cluster parameter that could apply to all the insurers because we want this parameter to reflect directly the presence of atmospheric condition in the North Atlantic Basin. Then, by fixing this parameter, we could simply estimate the storm parameter for each portfolio. It will hence reveal the number of storms per cluster that impact the wanted area.

Finally, the same type of adjustment as the one we saw in part D.1 will be made to adapt the cluster effect to the OEP curve.

Three portfolios are analyzed in this part. They are taken from three insurers that we will simply call A, B and C. Inside these portfolios, we will only analyze the most severe storm i.e. the storms whose FGU losses exceed 4 million euro[33]. This is the threshold that is required for the analysis.

D.2.1. Insurer A

The list of severe storms is for the 1990 to 2009 period:

Storm name	Date	FGU losses (EUR)
Daria	25/26.01.1990	10,628,149.00
Herta	03/04.02.1990	3,931,429.62
Vivian	26/27.02.1990	7,340,844.55
Wiebke	28.02/01.03.1990	7,157,634.35
Lothar	12.26.1994	42,961,732.00
Jeanett	26/28.10.2002	15,245,110.00
	29/30.07.2005	12,143,044.00
	6.28.2006	12,896,255.00
Kyrill	18/19.01.2007	43,447,353.00
Emma	29.02.-02.03.2008	10,391,427.00
	29.05.-04.06.2008	30,498,467.00
	22.06.-24.06.2008	7,401,602.00
	24.05.-26.05.2009	11,479,414.00
	7.3.2009	6,018,835.00

Table D.2.a: Insurer A's severe storm losses

We consider Daria and Herta as being part of the same cluster. To summarize, we have:

[33] as if 2009

Nb of years	Nb of storms	Nb of cluster	Nb of 1 storm clusters	Nb of 2 storms clusters	Nb of 3 storms clusters
20	14	12	10	2	0

Table D.2.b: Sum up of Insurer A's severe storm losses

1) Poisson model

First, we try to fit a Poisson model. We obtain by the maximum likelihood method

$\hat{\lambda} = \bar{x} = \frac{14}{20} = 0.7.$

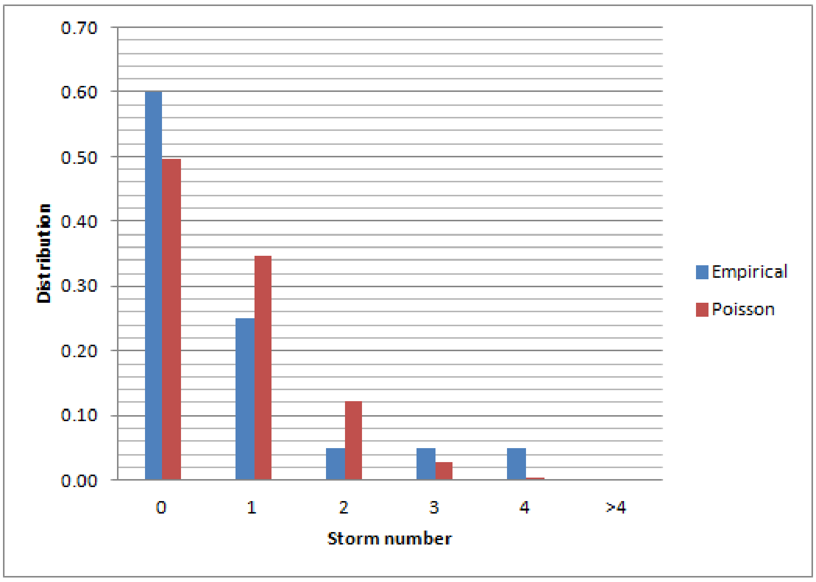


Figure D.2.c: Insurer A's Poisson distributed annual storms

The p-value is 35.50%

2) Negative Binomial model

As presented in the introduction of part D.2., we keep the European Poisson cluster estimator combined with the moment estimation that is computed directly from the CRM form:

$$\bar{x} = \hat{\lambda}_P . \frac{-\hat{p}_{LG}}{\hat{q}_{LG} ln \hat{q}_{LG}} = r_{NB} \frac{p_{NB}}{q_{NB}}$$

We search $\hat{p}_{LG}$ such that the above relationship is verified. However, as $\bar{x} < \hat{\lambda}_P$ (which means less annual storm in average in small portfolios than for the global European model), there is no solution (see part B.2.2.2.2, to know why assumptions are not verified). This means the negative binomial model is not adapted to studies of insurers' portfolios.

3) Poisson-Binomial collective risk model

As presented in the introduction part of D.2., we keep the European Poisson cluster estimation $\hat{\lambda} = 2.7492$. We estimate p as described in part B.3.4.1.1: $\hat{p}_{MM} = \frac{\bar{x}}{3\hat{\lambda}_{MM}} = 0.0849$.

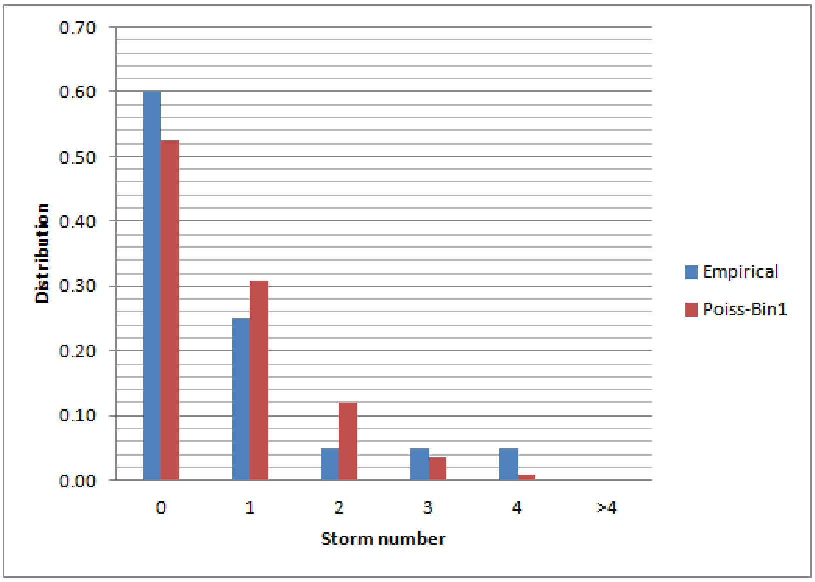


Figure D.2.d: Insurer A's storms per cluster

With a p-value of 50.80%

Another way to choose $\hat{p}_{MM}$ parameter is by minimizing the result of the Chi-square test. Then the parameter is $\hat{p}_{MM} = 0.0662$ and that gives a 100% p-value:

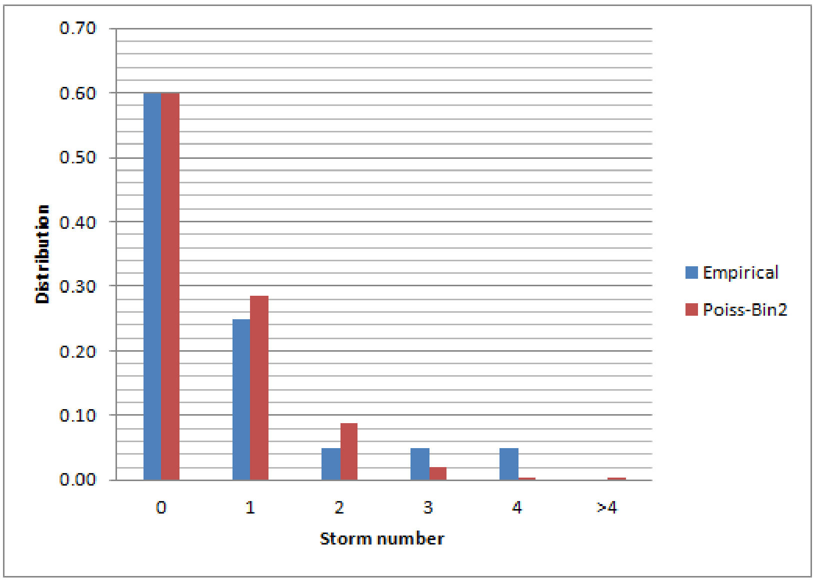


Figure D.2.e: Insurer A's storms per cluster

The classes used for this chi-square test are the number of zero storm years and the number of years containing at least one storm. One must have a minimum of five elements in each class.

By using the method described in part D.1, the premium can be obtained from the reference pricing tool by:

$$\pi_{with\ cluster} = \rho\left\{\int OEP_{from\ tool}\right\}$$

With the specific portfolio characteristics, we obtain:

$$\pi_{with\ cluster} = 1.4773\left\{\int OEP_{from\ tool}\right\}$$

D.2.2. Insurer B

The data includes the 1999 to 2009 period. The most severe storms are:

Storm name	Date	FGU losses (EUR)
Lothar	25/27.12.1999	10,229,425.01
Anna	26/28.02.2002	4,637,299.39
Jeanett	26/28.10.2002	18,110,134.84
	27/29.07.2005	5,982,008.21
Kyrill	18/20.01.2007	43,808,904.83
Lothar	25/27.05.2007	5,712,736.90
Emma	01/03.03.2008	7,339,328.76
Hilal	29/31.05.2008	22,623,765.51
EM	22/24.06.2008.	12,602,296.97
	03/05.07.2009	4,984,000.00

Table D.2.k: Insurer C's severe storm losses

To sum up, we obtain:

Nb of years	Nb of storms	Nb of cluster	Nb of 1 storm clusters	Nb of 2 storms clusters	Nb of 3 storms clusters
10	10	10	10	0	0

Table D.2.l: sum up of Insurer C's severe storm losses

1) Poisson model

First of all, we try to fit a Poisson model. The maximum likelihood estimator is $\hat{\lambda} = \bar{x} = \frac{10}{10} = 1$

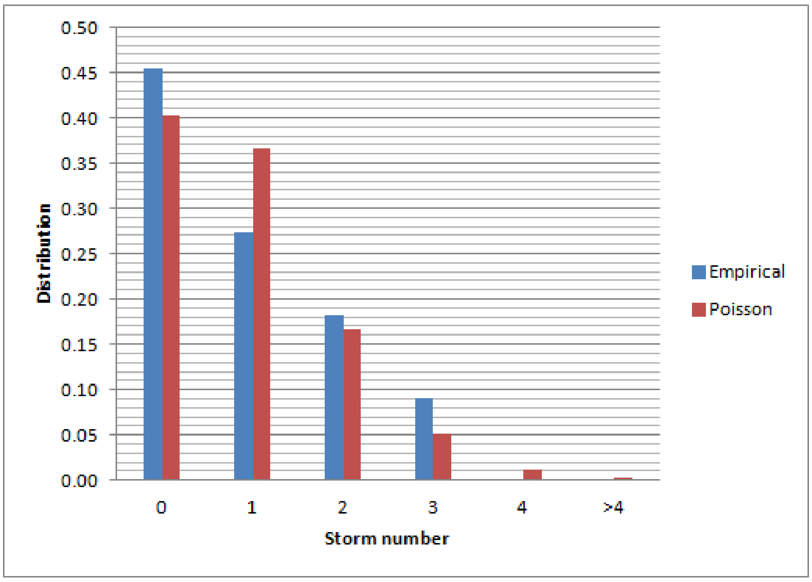


Table D.2.m: Insurer C's Poisson distributed annual storms

p-value is 72.69%

2) Negative Binomial model

As already presented in part D.2.1.2, the negative binomial model cannot be adapted to the portfolio.

3) Poisson-Binomial collective risk model

As presented in the introduction of part D.2, we keep the European Poisson cluster estimator $\hat{\lambda} = 2.7492$. We estimate p as described in part B.3.1.4.2 (by the method of moments) and we obtain $\hat{p}_{MM} = \frac{\bar{x}}{3\hat{\lambda}_{MM}} = 0.1102$.

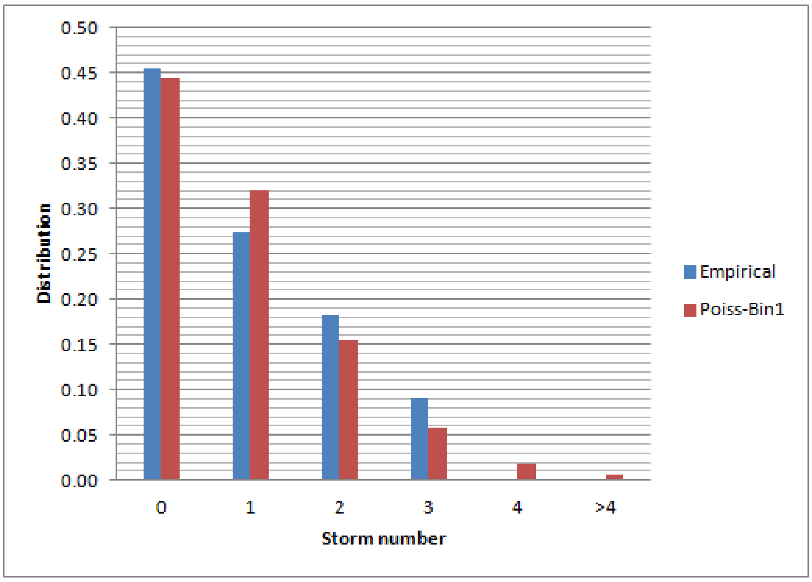


Figure D.2.n: Insurer C's storms per cluster

With a p-value of 94.24%

The minimization of the Chi-square gives $\hat{p}_{MM} = 0.1065$, with a 100% p-value:

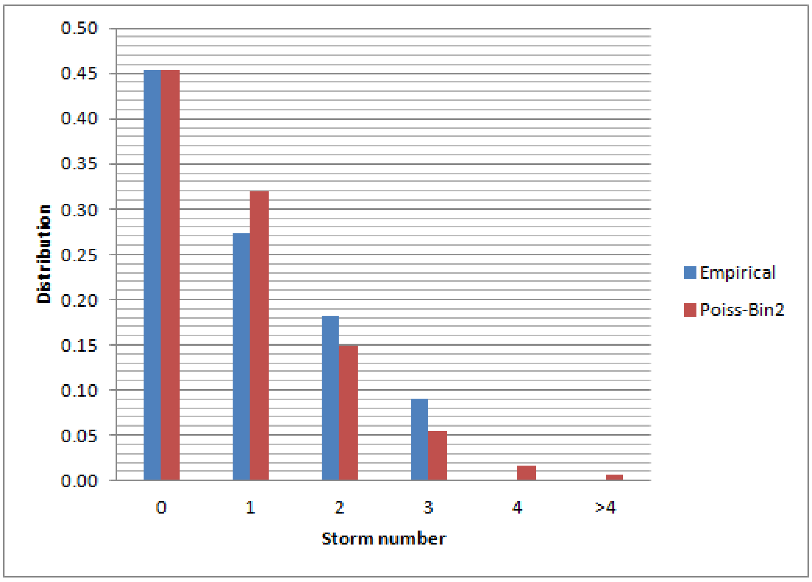


Figure D.2.o: Insurer C's storms per cluster

The classes used for this chi-square test are the number of zero storm years and the number of years containing at least one storm. One must have a minimum of five elements in each class.

As already seen in part D.2.1.3.3, the prime that takes the effect of clusters into account is given by:

$$\pi_{with\ cluster} = 1.6341\left\{\int OEP_{from\ tool}\right\}$$

Conclusion

Modeling windstorm perils is a recent and wide subject that easily goes beyond the traditional actuarial field. Their parameters have been defined by meteorological experts for a specific weather environment representative of the last decades. As any model, they have limitations simply because they will never take into account perfectly all the key variables that rule storms. Additionally, there are uncertainties about how the climate will evolve in the near future. For the purpose of pricing insurance products, it is hence crucial for the actuary to know where the models limits are.

One of the main weaknesses of the current models is that they do not consider the possible existence of clusters of storms. In this study, we stated that both the negative binomial distribution and the Poisson-binomial collective risk models seemed relevant to capture the idea of clusters. These two models gave significant results, with respect to the Chi-square goodness of fit test, with respective p-values of 85% and 91%, while the Poisson model gave 6%.

Such improvements had a considerable impact on the expected annual storm losses. On a European wide scale, we calculated an increase ranging from 5% to 8% compared to the original model assumptions. Similarly, we have shown that the Poisson-binomial collective risk model is the best to fit the geographic specificities of a single regional insurance portfolio. This model includes two parameters. One is a fixed parameter that stood for the global cluster formation in the North Atlantic. It is commonly determined for all European insurers' portfolios. The second parameter is geographically specified according to the company’s historic losses. We concluded that such a model has the advantage to be easy to be implemented and to be interpreted.

Other model improvements could be studied for windstorm modeling. One of them is to consider the North Atlantic oscillation. It is an atmospheric phenomenon that alternatively makes storms going across Northern or Southern Europe. The result of such a study would reveal a negative correlation between these two zones. At a local scale, this would allow to adjust the hazard level temporarily.

Bibliography

[1] CAILLAT AL.(2009), *An analysis of the motor reinsurance market : the ingredient of portfolio optimization*, New Re

[2] DELLA-MARTA P., MATHIS H., FREI C., LINIGER M., KLEINN J., APPENZELLER C. (2009), *The return period of wind storms over Europe*, International Journal of Climatology, volume 29, pages 437-459

[3] GEO RISKS RESEARCH (2002), *Tempêtes d'hiver en Europe*, Munich Re

[4] GEO RISKS RESEARCH (2010), *Great natural catastrophes causes and effects*, Topics Geo 2009, Munich Re

[5] LEWIS M.(August 26th, 2007), *In Nature's Casino,* The New York Times

[6] HOLLAND D. (2009), *A brief history of reinsurance*, Reinsurance News

[7] KLUGMAN S., PANJER H., WILLMOT G. (2004), *Loss models : from data to decisions, 2nd edition*, Wiley

[8] ROSS M. (2009), *Introduction to Probability Models*, Elsevier Science

[9] SCHWEPCKE A. (2004), *Reinsurance : Principles and State of the Art*, Karlsruhe

[10] SCHIRMER J. (2010) *Storm series described as a negative binomial process*, Allianz Re

[11] WOO G. (1999) *The mathematics of natural catastrophes*, Imperial College Press

Appendices

(1) Proof of (NB.1)

The following proof is taken form Schirmer (2010).

Let T be the stopping time considered for the counting process, it can be 1 for example as we are only interested in an annual time scale process, as it is our case.

Let $w_{s,t}^{(k)}$ be the probability of having no event between s and t. At this point, the Poisson process intensity has become $\lambda + k\delta$. By using the Poisson process definition 1, one can find:

$$w_{s,t}^{(k)} = P(no\ claim\ between\ s\ and\ t \mid k\ events\ already\ occured)$$
$$= e^{-(\lambda+k\delta)(t-s)}$$

Let $w_{t,t+dt}^{(k)}$ be the density function. By using the Poisson process definition 2, one can find:

$$w_{t,t+dt}^{(k)} = P(no\ claim\ between\ t\ and\ t+dt \mid k\ events\ already\ occured)$$
$$= 1 - (\lambda + k\delta)dt + o(dt)$$

So now, considering stopping time T, the probability of having no event between 0 and T is given by $P_T(N=0) = w_{0,T}^{(0)} = e^{-\lambda T}$.

Now if we do the same for $N = 1$, we have to condition on the time of occurrence, which gives us the follow equality:

$$P_T(N=1) = \int_0^T w_{0,t}^{(0)}\left(1 - w_{t,t+dt}^{(0)}\right)w_{t,T-(t+dt)}^{(1)}$$
$$= \int_0^T e^{-\lambda t}(\lambda.dt)e^{-(\lambda+\delta)(T-t)}$$
$$= \lambda e^{-(\lambda+\delta)T}\int_0^T e^{\delta t}dt$$
$$= \frac{\lambda}{\delta}e^{-\lambda T}\left(1 - e^{-\delta T}\right)$$

Now, we want to demonstrate that for $k \in \mathbb{N}$ we have:

$$P_T(N = k) = \binom{r + k - 1}{k} e^{-\lambda T} \left(1 - e^{-\delta T}\right)^k$$

where $r = \frac{\lambda}{\delta}$

Let us assume the last equation and see how one can compute $P_T(N = k + 1)$

$$P_T(N = k + 1) = \int_0^T P_t(N = k)\left(1 - w_{t,t+dt}^{(k)}\right) w_{t,T-(t+dt)}^{(k+1)}$$

$$= \int_0^T \binom{r+k-1}{k} e^{-\lambda t} \left(1 - e^{-\delta t}\right)^k (\lambda + k\delta) e^{-(\lambda+(k+1)\delta)(T-t)} dt$$

with $(\lambda + k\delta) = \delta(r + k)$

$$= \binom{r+k}{k+1} \delta(k + 1) e^{(-\lambda+(k+1)\delta)T} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \int_0^T e^{(k+1-j)\delta t} dt$$

$$= \binom{r+k}{k+1} e^{-(\lambda+(k+1)\delta)T} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \frac{k+1}{k+1-j} \left(e^{(k+1-j)\delta T} - 1\right)$$

$$= \binom{r+k}{k+1} e^{-\lambda T} \sum_{j=0}^{k+1} (-1)^j \binom{k+1}{j} \left(e^{-j\delta T} - e^{-(k+1)\delta T}\right)$$

$$= \binom{r+k}{k+1} e^{-\lambda T} \left(1 - e^{-\delta T}\right)^{k+1}.$$

If one decides to set $T = 1$:

$$P(N = k) = \binom{r+k-1}{k} e^{-\lambda} \left(1 - e^{-\delta}\right)^k,$$

and then to substitute $e^{-\lambda}$ by p:

$$P(N = k) = \binom{r+k-1}{k} e^{-\lambda} (1 - p)^k$$

$$= \binom{r+k-1}{k} e^{-r\delta} (1 - p)^k$$

$$= \binom{r+k-1}{k} p^r (1 - p)^k$$

We have now a common parameterization of the negative binomial distribution.

(2) Proof of (NB.2)

One wants to prove that the portfolio N is Negative Binomial distributed using the distribution characterization by the moment generating function.

Remember the CRM for our storm portfolio is:

$$N = \sum_{i=1}^{P} S_i$$

where $P \sim Poisson(\lambda)$, $S_i \sim Logarithmic(p)$ and S_i are independent r.v.s.
Let us compute its moment generating function:

$$\begin{aligned}
\mathcal{M}_N(t) &= E[e^{tN}] \\
&= E\left[E\left[e^{t\sum_{i=1}^{P} S_i}|P\right]\right] \\
&= E[\textstyle\prod_i^P E[e^{tS_i}|P]] \\
&= E\left[\mathcal{M}_{S_1}^P(t)\right] \\
&= m_P \circ \mathcal{M}_{S_1}(t)
\end{aligned}$$

where $m_P(z) = E[z^P]$ is the probability generating function, P a $\mathbb{N}$-measurable random variable.

The logarithmic moment generating function is: $\mathcal{M}_S(t) = \frac{ln(1-pe^t)}{ln(1-p)}$.

The Poisson probability generating function is: $m_P(z) = e^{\lambda(z-1)}$.

Combining both makes the following result:

$$\begin{aligned}
m_P \circ \mathcal{M}_{S_1}(t) &= exp\left\{\lambda\left(\frac{ln(1-pe^t)}{ln(1-p)} - 1\right)\right\} \\
&= exp\left\{\lambda \frac{ln\left(\frac{1-pe^t}{1-p}\right)}{ln(1-p)}\right\} \\
&= \left(\frac{1-p}{1-pe^t}\right)^{\frac{-\lambda}{ln(1-p)}}
\end{aligned}$$

By taking $r = \frac{-\lambda}{ln(1-p)}$ and $\tilde{p} = 1 - p$, one can directly see that $\mathcal{M}_N(t)$ stands for a negative binomial moment generating function with parameters r and $\tilde{p}$.

(3) Chi-square goodness-of-fit test

In this section we want to test whether a r.v Y is P_0 distributed. Let us assume we are given a n-sample $\{Y_i\}_{i\in[\![1,n]\!]}$ independent from Y's distribution. Let $\{C_i\}_{i\in[\![1,r]\!]}$ be a partition of $\mathbb{R}$. Each C_i is called a "class".

The statistic can be defined as:

$$\chi^2 = \sum_{i=1}^{r} \frac{n(\widehat{p_i} - p_{ni})^2}{\widehat{p_i}}$$

Where:

- n is the sample size;
- r is the number of classes of the test.
  The idea is to group the data inside these classes to make the comparison. As they form a partition of the study field, $r-1$ classes are arbitrary defined by the user. This is the degree of freedom of the test;
- $\widehat{p_i}$ is the probability the distribution P_0 falls into the i^{th} class, $\widehat{p_i} = P_0(Y \in C_i)$;
- p_{ni} is the empirical distribution test falls into the i^{th} class.

The test also exists in the following definition:

$$\chi^2 = \sum_{i=1}^{r} \frac{(E_i - O_i)^2}{E_i}$$

Where:

- $E_i = n\widehat{p_i}$ is the number of expected observation of the i^{th} class according to the model hypothesis
- $O_i = np_{ni}$ is the number of observation in the class.

The critical value for this test comes from the chi-square distribution with $r-1$ degrees of freedom. There are number of rules that have been proposed for deciding when the test is reasonably accurate. The most conservative one we used in the dissertation is that E_i must have a minimum value of 5, for each class.

(4) VBA code for decomposition of collective risk model storms

This is the implementation for the algorithm described in part B.3.1.2. We want to write the different possibilities of decomposing a number of 6 annual storms. At this stage, we consider the number of storms per cluster as ZT-binomial distributed. So, clusters can produce 1, 2 or 3 storms.

```
Sub decompCRM()
For j = 0 To 3
ActiveCell.Offset(0, j) = 0
Next

For i = 1 To 6
  For r3 = 0 To Int(i / 3)
    For r2 = 0 To Int((i - 3 * r3) / 2)
      ActiveCell.Offset(1, 0).Activate
      ActiveCell.Value = i
      ActiveCell.Offset(0, 3).Value = r3
      ActiveCell.Offset(0, 2).Value = r2
      ActiveCell.Offset(0, 1).Value = i - 3 * r3 - 2 * r
    Next
  Next
Next
End Sub
```

This code is used to fill the following table:

Storm nb	1 storm-cluster nb	2 storms-cluster nb	3 storms-cluster nb
0	0	0	0
1	1	0	0
2	2	0	0
2	0	1	0
3	3	0	0
3	1	1	0
3	0	0	1
4	4	0	0
4	2	1	0
4	0	2	0
4	1	0	1
5	5	0	0
5	3	1	0
5	1	2	0
5	2	0	1
5	0	1	1
6	6	0	0
6	4	1	0
6	2	2	0
6	0	3	0
6	3	0	1
6	1	1	1
6	0	0	2

(5) Data used for European storms (1957/2002)

The following data is taken from Meteo Schweiz website:

http://www.meteoswiss.admin.ch/web/en/research/completed_projects/nccr_climate_ii/return_period_wind_storms0.html.

Storm ID	Name	Date	Cluster group ID
1		3/9/2002	1
2	Anna	2/27/2002	2
3	Jennifer	1/29/2002	3
4		11/6/2000	4
5	Oratia	10/30/2000	5
6	Liane	1/31/2000	6
7	Kerstin	1/29/2000	6
8	Martin	12/28/1999	7
9	Lothar	12/26/1999	7
10	Anatol	12/3/1999	8
11	Lara	2/5/1999	9
12	Silke	12/27/1998	10
13	Xylia	10/28/1998	11
14	Winnie	10/24/1998	11
15	Elvira/Farah	3/5/1998	12
16	Désirée/Fanny	1/4/1998	13
17	Waltraud	4/11/1997	14
18	Sonja	3/28/1997	15
19	Heidi	2/26/1997	16
20	Gisela	2/25/1997	16
21	Daniela	2/20/1997	16
22	Ariane	2/13/1997	17
23		10/29/1996	18
24	Wilma	1/26/1995	19
25		1/21/1995	19
26	Ornella	1/11/1995	20
27	Kundry, Lydia	4/1/1994	21
28		3/14/1994	22
29	Lore	1/27/1994	23
30		1/23/1994	23
31	Victoria	12/20/1993	24
32	Quena	12/9/1993	25
33	Johanna	11/14/1993	26
34	Agnes, Barbara	1/23/1993	27
35	Walpurga	1/16/1993	28
36	Verena	1/14/1993	28
37		1/11/1993	28
38		12/2/1992	29
39	Ismene	11/26/1992	29
40	Coranna	11/12/1992	30
41		3/13/1992	31
42		1/1/1992	32
43	Nora	10/18/1991	33
44	Waltraud	1/8/1991	34
45	Undine	1/6/1991	34
46		1/3/1991	34
47		12/26/1990	35
48		11/18/1990	36
49	Wiebke	3/1/1990	37
50	Vivian	2/26/1990	37
51	Polly	2/15/1990	38

Storm ID	Name	Date	Cluster group ID
52	Ottilie	2/14/1990	38
53	Nana	2/12/1990	38
54	Judith	2/8/1990	39
55	Herta	2/4/1990	39
56		1/28/1990	40
57	Daria	1/26/1990	40
58		12/17/1989	41
59		12/14/1989	41
60		10/21/1989	42
61		4/11/1989	43
62		4/5/1989	43
63		2/26/1989	44
64		2/14/1989	45
65		1/15/1989	46
66		12/22/1988	47
67		12/19/1988	47
68		12/6/1988	48
69		11/30/1988	48
70		10/8/1988	49
71		3/1/1988	50
72		2/10/1988	51
73		2/1/1988	52
74		1/24/1988	53
75	87J	10/16/1987	54
76		3/28/1987	55
77		12/19/1986	56
78		12/15/1986	56
79		10/22/1986	57
80		3/25/1986	58
81		1/20/1986	59
82		1/14/1986	59
83		12/6/1985	60
84		11/6/1985	61
85		1/31/1985	62
86		11/23/1984	63
87		10/20/1984	64
88		2/7/1984	65
89		1/14/1984	66
90		1/3/1984	67
91		11/27/1983	68
92		10/15/1983	69
93		2/1/1983	70
94		1/18/1983	71
95		1/15/1983	71
96		1/4/1983	72
97		12/16/1982	73
98		11/8/1982	74
99		3/3/1982	75
100		3/1/1982	75
101		11/24/1981	76
102		2/4/1981	77
103		1/3/1981	78
104		10/7/1980	79
105		4/20/1980	80
105		4/20/1980	80
106		12/17/1979	81
107		12/11/1979	81
108		12/5/1979	81
109		3/30/1979	82
110		12/30/1978	83
111		11/15/1978	84

Storm ID	Name	Date	Cluster group ID
112		3/16/1978	85
113		1/12/1978	86
114		1/4/1978	87
115		12/24/1977	88
116		11/15/1977	89
117		11/12/1977	89
118		10/3/1977	90
119		12/4/1976	91
120		12/2/1976	91
121		10/16/1976	92
122		1/21/1976	93
123	Capella	1/3/1976	94
124		11/30/1974	95
125		10/20/1974	96
126		1/17/1974	97
127		1/12/1974	97
128		12/17/1973	98
129		12/14/1973	98
130		12/3/1973	99
131		11/25/1973	100
132		11/20/1973	100
133		11/13/1973	101
134		4/3/1973	102
135	Niedersachsen	11/13/1972	103
136		4/2/1972	104
137		11/21/1971	105
138		11/18/1971	105
139		1/26/1971	106
140		11/4/1970	107
141		10/21/1970	108
142		11/10/1969	109
143		2/7/1969	110
144		1/15/1968	111
145	Skane	10/17/1967	112
146		2/28/1967	113
147	A.Bermpohl	2/23/1967	113
148		2/21/1967	113
149		2/14/1965	114
150		11/19/1963	115
151	Hamburg	2/17/1962	116
152		2/12/1962	116
153		1/11/1962	117
154		12/6/1961	118
155		11/1/1960	119
156		1/20/1960	120
157		10/27/1959	121
158		11/4/1957	122

Notations

AIR	Applied Insurance Research
CRM	Collective Risk Model
EUR	Euro
FGU	From Ground Up
MLE	Maximum Likelihood Estimation/or
MM	Method of Moments
NAT-CAT	Natural Catastrophe
RMS	Risk Management Solutions
USD	US Dollar
XL	Excess of Loss
ZT	Zero Truncated