Building a hurricane risk map for continental Portugal based on loss data from hurricane Leslie

Andrea Hauser[a], Carlos Rosa[a], Rui Esteves[a],
Alexandra Moura[b], and Carlos Oliveira[b]
[a]*Companhia de Seguros Fidelidade, Portugal*
[b]*ISEG - School of Economics and Management, Universidade de Lisboa and REM-Research in Economics and Mathematics, CEMAPRE, Portugal*

December 10, 2021

Abstract

A complete model to analyse and predict future losses in the property portfolio of an insurance company due to hurricanes is proposed. A novel statistical model, in which weather data is not required, is considered. Climate data may not be reliable, or may be difficult to deal with or to obtain, hence we reconstruct the storm behaviour through the registered claims and respective losses. The model is calibrated using the loss data of the property portfolio of the insurance company Fidelidade, from hurricane Leslie, which hit the center of continental Portugal in October 2018.

Several scenarios are simulated and risk maps are built. The simulated scenarios can be used to compute risk premiums per risk class in the portfolio. These can be used to adjust the policy premiums accounting for a storm risk. The risk map of the company also depends on its portfolio, namely its exposure, providing a hurricane risk management tool for the insurance company.

Keywords— Risk; Hurricanes; Property Insurance; Regression Models

1 Introduction

Climate change and global warming are currently recognized as serious threats to human society. For instance, according to [26], the human body cannot adapt to specific temperatures and humidity stress levels. These thresholds may be reached in South and Southwest Asia in this century ([21, 14]). In the same direction, international organizations have been highlighting that climate change and global warming threaten the aim of sustainable economic and social development. On December 12, 2015, 196 countries signed the Paris agreement with the objective of keeping global warming below 2º C, in comparison with pre-industrial levels or even below 1.5º C, if possible.[1]

[1] Information regarding the Paris agreement can be found at https://unfccc.int/process-andmeetings/the-paris-agreement/the-paris-agreement

Despite the political efforts made so far, projections show that heatwaves and floods will be more intense and frequent. On the other hand, the total number of tropical cyclones will not significantly increase, although an increase in the intensity of tropical cyclones is projected ([11, 2, 20]).

The economic consequences of extreme weather events have become more and more relevant. Table 1 shows that the average costs per year with weather and climate events in the US have significantly increased in the last decades. In Europe, the average costs have mainly increased from the 1980s to the 1990s, but remained stable in the following decades.

Time period	US (billion $)	Europe (billion €)
1980-1989	18.1	6.6
1990-1999	28.9	12.3
2000-2009	54.7	13.2
2010-2019	85.8	12.4

Table 1: Average cost per year with weather and climate-related disasters in the US and Europe. Costs for the US are in billion dollars (2021 values) and costs for Europe are in billion euros (2019 values). Sources: National Oceanic and Atmospheric Administration and European Environment Agency.[2]

Part of the economic losses with climate and extreme weather events are covered by insurance companies. During the first half of 2021, natural disasters were so destructive that insurance companies had to pay out more than 41 billion dollars (predicted amount) in compensations, which is the highest value in the last 10 years[3]. In Europe, between 1980 and 2019, the total cost with climate-related and extreme weather events is estimated at 499.476 billion euros (2019 values) and, from this amount, 179.66 billion euros (2019 values) were covered by the insurance sector.[4]

In Portugal, insurance companies have paid out about 650 million euros (2019 values) in compensations between 1980 and 2019. Some of the weather and climate-related events that provoked higher losses to insurance companies in the last years are described in Table 2. We can easily see that wildfires in the center of Portugal, which occurred in October 2017, were the most costly event. The second most costly event was hurricane Leslie, in which insurance companies paid out about 101 million euros. Regarding the number of claims, the tropical hurricane Leslie was the more devastating event.

[2] Information available at https://www.eea.europa.eu/data-and-maps/indicators/direct-lossesfrom-weather-disasters-4/assessment and https://www.ncdc.noaa.gov/billions/ ([7]).
[3] News available at https://www.theguardian.com/business/2021/jul/21/2021s-extremeweather-leads-to-insurers-biggest-payout-in-10-years.
[4] Source: European Environment Agency.

Date	Type of Event	Region of Portugal	Indemnity (million €)	Claims
Jan 2014	Storm	Mainland	11.5	5544
Sept 2014	Floods	West region	1.5	489
Nov 2015	Floods	Algarve	15.5	1762
Aug 2016	Fires	Madeira	19.7	328
Jun 2017	Fires	Pedrógão	22.4	493
Oct 2017	Fires	Center	235.4	4177
Oct 2018	Leslie	Center and North	101.0	38000
Dec 2019	Elsa and Fabian	Center and North	42.0	22700

Table 2: Weather and Climate-related disasters in Portugal. Source: Portuguese Association of Insurers.[5]

According to Munich Re, tropical storms, such as hurricanes, typhoons, and cyclones are among the most costly natural hazards[6]. For instance, hurricane Katrina was the most costly natural disaster of all time for the insurance sector, with losses exceeding 60 billion dollars. Hurricane Lorenzo, which hit Azores in October 2019, caused total economic losses of 330 million euros[7]. Additionally, an increase of hurricanes of tropical origin over Western Europe during early Autumn (Aug-Oct) is expected in the future ([12]). Thus, in this work, we will present a methodology to assess the risk of insurance companies regarding this type of events in Continental Portugal.

Climate change and weather and climate-related extreme events have been pointed out as important challenges to insurance companies ([19, 5]). On the one hand, insurers have to develop new insurance products that help to mitigate these risks, particularly in developing countries where the effects of climate change are expected to be more severe ([19, 8, 1]).

On the other hand, insurers have to find novel methodologies to measure their portfolios' physical risk, considering the natural uncertainty of projections and the

[5] This information can be found in the following websites:
https://www.apseguradores.pt/Portals/0/doc/publicacoes/Revista%20APS%2001$_{P}T$%20 − %20*FINAL.pdf?ver* = 2019 − 07 − 05 − 101014 − 453
https://www.apseguradores.pt/pt/comunica%C3%A7%C3%A3o/not%C3%ADcias/2019/tempestadeleslie-sinistros
https://www.apseguradores.pt/pt/comunica%C3%A7%C3%A3o/not%C3%ADcias/2020/articleid/142/tempestade-
elsa-e-fabien-%E2%80%93-balan%C3%A7o-final-de-dados-do-setor-segurador-22-7-mil-sinistrosparticipados-com-um-custo-estimado-de-42-milh%C3%B5es-de-euros

[6] Information available at the website of Munich Re: https://www.munichre.com/en/risks/natural-disasters-losses-are-trending-upwards/hurricanes-typhoons-cyclones.html

[7] Information available at:
https://www.publico.pt/2019/10/14/sociedade/noticia/furacao-lorenzo-provocou-prejuizos-330-milhoes-euros-1889978

difficulty in diversifying the risk due to geographical correlation [4]. In the previous reference, the author explains the difficulties of covering catastrophic risks, which can only happen if a correct mutualization of risk is performed. In fact, the risk has to be spread between policyholders, reinsurance companies, and financial markets.

The economic and social impact of natural hazards has been extensively studied in the last few years (see, for instance, [6, 15, 17, 23, 24, 25], and references therein). Some authors have been using damage functions to assess the expected costs of these events ([6, 24, 25]). In this approach, one establishes a relationship between the magnitude of a natural hazard and the average damage caused on a specific item or portfolio of items ([24]). The authors of [25] compare four different damage functions to estimate the losses of winter storms. The damage functions are calibrated against the daily insurance loss data due to storms affecting the residential buildings in Germany from 1997 to 2007. In [6], the same insurance loss data and the daily maximum wind gust data from ERA-Interim reanalysis project[8] are considered, but the damage function is calibrated considering just the significant losses related to large-scale winter storms for the period 1997 to 2007.

The authors of [17] present a statistical model that predicts losses based on the variables wind speed, age of the buildings, building floor area, and appraised value of the building. To validate the model, the authors used the data set of Texas Windstorm Insurance Association regarding the claim payout records for commercial buildings after hurricane Ike. In [15], the authors present a probabilistic model, which predicts aggregated losses in the US due to tropical hurricanes. According to the authors, the methodology can also be applied to a subset of losses, for instance, the portfolio of a reinsurance company.

We contribute to the literature by proposing a complete model to analyse and predict future losses in the property portfolio of an insurance company due to hurricanes in continental Portugal. We propose a novel statistical model in which weather data is not required. Instead, we reconstruct the storm behaviour through the claims and losses registered. We calibrate the model with the loss data from hurricane Leslie of the insurance company Fidelidade. With this analysis, we can conclude that single-family houses are especially vulnerable to hurricanes. Additionally, the losses would be much larger in case a hurricane hit continental Portugal in other regions.

The paper is organized as follows: in Section 2, we describe the data set. In Section 3, we introduce the statistical model used to predict the number of claims, and the losses associated with each claim. In Section 4, we use the methodology of Section 3 to estimate the costs in case the storm landfall council is different. In Section 5, we build a risk map as well as the distribution of the estimated total cost. Finally, we conclude in Section 6.

[8] More information about the Era Interim dataset can be found at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim

2 Data

When modeling the damage caused by extreme meteorologic events, such as storms, into a specific item, such as buildings, it is natural to try to use climate data. However, the climate data of extreme weather events is extremely variable in space and time, specially when the wind is the main variable concerned as it is the case of a storm. This makes the use of climate data very challenging since it is common that the available values are averaged or extrapolated, meaning that the extreme values are not accessible. Data regarding Leslie hurricane, which occurred on October 13, 2018, is an example of the lack of reliability and accessibility of spatial climate data. The meteorological observations recorded by the Portuguese Institute for Sea and Atmosphere (IPMA) relative to that day are not easily available and can be obtained only under previous request.

Two sources of climate data, which are publicly and easily available, and widely used in the study of losses produced by meteorological events [6, 24, 25], are the ERA-Interim and ERA-5 reanalysis weather data from the ECMWF (European Center for Medium-Range Weather Forecasts). We compared the losses incurred by the property portfolio of the company due to hurricane Leslie with the ERA-5 reanalysis data for the same day. The daily maximum wind gust and the total daily precipitation are the quantities considered. The daily maximum wind gust is equal to the maximum value of the hourly 10 meter wind gust and has been computed for a grid of 142 locations, available in these databases, in the north and center of continental Portugal. The total daily precipitation instead is equal to the sum of the 1-hourly total precipitation amount and has been computed for each of those 142 locations. It was observed that the mentioned meteorological variables were not compatible with the amount of losses registered by the company. Indeed the maximum reanalysis value of the $10m$ wind gust was $88.00 km/h$, which is too low to provoke the level of losses observed, which were more than 100 million euros for the insurance sector. Also, [22] reports that a wind gust of $176 km/h$ was recorded in Figueira da Foz, which is not in line with the values from the Era-Interim and Era-5 data sets. Climate data, like other types of data, have also the problem of missing values or erroneous observations, even if automatic observatory stations are considered.

For these reasons, in this work we only consider data from the insurance company, namely, we consider information on the portfolio, and losses of the company due to the meteorologic event. In this work, we consider the property insurance portfolio of Fidelidade, which is the insurance company with the biggest market share in Portugal. The data set consists of a portfolio of 1303984 policies in 2018, from which 900917 are from councils with at least one claim due to Leslie. Table 3 displays the variables for which we have information on each policy, as well as the possible values that each of those variables can take. We define risk class as the set of all policies with the same characteristics of the variables provided by the insurance company in Table 3.

3 The Model for predicting the losses

In this section, we present our proposed model to estimate the expected loss due to a storm in a given region based only on loss data. In particular, we consider property damage. This approach is different from most works in literature, where climate data is also considered, as it is the case of [6, 24, 25] where the climate data is essential to build the damage functions that are then used to estimate losses. The construction of models that only use loss data has obvious disadvantages, but it has the great advantage that the data is reliable, unlike the climate data, as already mentioned in the previous section. We consider loss data from hurricane Leslie, which was a storm with a huge impact, see Table 2, to calibrate the model that will afterwards be used to predict the costs in different regions of continental Portugal, by estimating the costs that a similar event would have in other regions, and, ultimately, to build a storm risk map by simulating many different scenarios. In our approach, the loss data is seen as an indirect measure of the climate variables.

To build the model, we first infer from the loss data the trajectory of the storm after its landfall. By trajectory of the storm is understood as the imaginary line around which the observed claims are distributed. Thus, the trajectory is the line that passes through the affected councils, i.e. councils with claims, and it is obtained using the least-squares method. The information on the trajectory is afterwards included in the regression models of the claims number and costs. The final model allows characterizing the areas and the classes of policies, which have a higher risk for the company, in terms of expected costs, in the event of hazards similar to hurricane Leslie.

Variable	**Possible values**
Type of Property	content building
Year of construction	Level 1 if year of construction ≤ 1982 Level 2 if 1982 < year of construction ≤ 1992 Level 3 if 1992 < year of construction ≤ 2002 Level 4 if year of construction>2002
Framing of the Housing	Residential cluster Semi-detached house Other
Type of housing	Apartment Single-family house Other
Type of floor	Level 1: sub cave or ground floor or intermediate floor Level 2: last floor Not defined
Capital insured	Level 1: capital insured ≤ 80000 Level 2: 80000 capital insured ≤ 120000 Level 3: 120000 capital insured ≤ 165000 Level 4: capital insured > 165000
Region	North Center Metropolitan Area of Lisbon (MAL) Alentejo Algarve
Altitude	Level 1: altitude ≤ 90m Level 2: 90m altitude ≤ 200m Level 2: altitude > 200m
Forest area	Level 1: $LQ \leq 1.45$ Level 2: $LQ > 1.45$
Bush area	Level 1: $LQ \leq 0.19$ Level 2: $LQ > 0.19$
Urban area	Level 1: $LQ \leq 2.12$ Level 2: $2.12 < LQ \leq 3.35$ Level 3: $LQ > 3.35$

Table 3: Variables in the property damage portfolio considered. The location quotients (LQ) for variables “Forest area”, “Bush area” and “Urban area”, have been obtained from the 2018 report of the Portuguese Statistics Institute (INE - Instituto Nacional de Estatistica). The other variables are provided by the insurance company.[9]

[9] The location quotient is the share of the council with that type of territory divided by the share of Continental Portugal with that type of territory. For instance, the location quotient of forest area of a given council is the share of forest area of that council divided by the share of forest area in Continental Portugal. The full report can be consulted at:

3.1 Modeling the storm path

In this work, we consider loss data aggregated by council. However, other granularities may be used. We start by defining, for each council *i*, the cost ratio and the ratio of affected buildings, as introduced in [13], denoted respectively by CR_i and RAB_i:

$$CR_i = \frac{\text{loss in i}}{\text{tot. amount insured in i}}, \qquad RAB_i = \frac{\text{numb. of claims in i}}{\text{tot. numb. of properties insured in i}}$$

These quantities provide a measure of the impact of the storm, which are relative to the exposure of the company in each council, allowing for comparisons in different regions and epochs, according to the variability of the portfolio of the insurer.

Figure 1, left, displays the cost ratio caused by hurricane Leslie, by council. From the figure, we can infer the path of the hurricane from its landfall in Figueira da Foz, and its weakening along its path towards northeast, reflected by lower values of the cost ratio, as it moves inland.

To estimate the trajectory followed by the hurricane we consider a least square problem to find the line that best adjusts to the geographical points of the main city of each affected council, subject to the restriction that the line must pass through the landfall point, in this case, Figueira da Foz. We consider latitude and longitude measures. Hence, the least square problem with constraint to be solved is given by (1):

$$\min_{\alpha,\beta>0} \sum_{i=1}^{n} (lat_i - (\alpha + \beta\, lon_i))^2, \quad \text{s.t.} \quad \alpha = lat_k - \beta\, lon_k \tag{1}$$

where *lat* and *lon* denote latitude and longitude, and k ="Figueira da Foz". Figure 1, right, represents the geographical representation of the affected councils' main cities and the estimated line, i.e. the solution to the constraint least square problem (1).

The part of the trajectory line going from Figueira da Foz until the furthest council in the northeast of the country that reported claims, which is Bragança, has a length of approximately 260*km*, meaning that the hurricane caused losses to the company for at least 260*km* travelling inland. We do not have information available about the damages caused outside the borders of continental Portugal, but, for prudence reasons, we decide to assume that the event could produce damages for 300*km* before dissipating.

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=435668469&DESTAQUESmodo=2 of 2018 from the INE (Instituto Nacional de Estatistica).

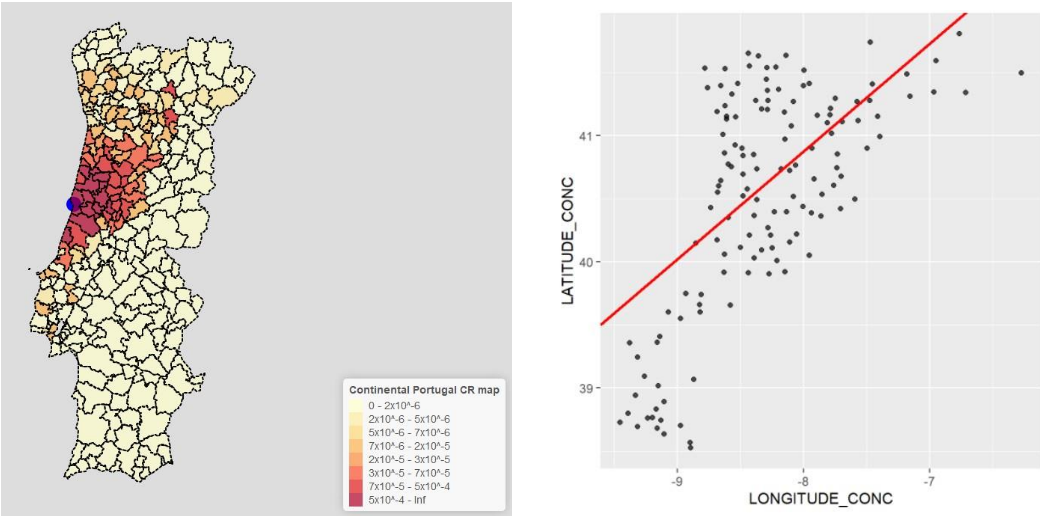

Figure 1: **Left:** Distribution of the observed cost ratios, by council, due to hurricane Leslie over continental Portugal. The blue dot refers to Figueira da Foz, the landfall point of hurricane Leslie. **Right:** Coordinates of the main cities of the councils that reported at least one claim due to hurricane Leslie and estimated trajectory of the hurricane.

Based on the estimated trajectory, we define two new variables: $dist1$, representing the distance of the insured object to the storm's landfall point, and $dist2$, representing the perpendicular distance of the insured object to the trajectory line. The use of variable $dist1$ is justified by the decrease of the observed cost ratio along the trajectory line, starting from the landfall point, while the use of variable $dist2$ is justified by the fact that the affected councils are closely distributed around the trajectory line (see Figure 1). Variables $dist1$ and $dist2$ are based on the insurer loss data and express an indirect measure of the damage. To understand how these variables relate to the observed cost ratio, CR, and ratio of affected buildings, RAB, we develop a regression tree model, representing CR and RAB for different values of $dist1$ and $dist2$ (see for instance [18, 9] for tree-based and regression tree models). We use a regression tree based on the ANOVA method [18]. The smallest number of observations allowed in a terminal node, used as stopping criteria for the size of the tree, is 20.

Figure 2 shows the two regressions obtained for CR and RAB of the councils

	$intensity\ 1$	$intensity\ 2$
level1	$dist1 < 54km$	$dist1 < 54km$
level2	$54km \leq dist1 < 78km$	$54km \leq dist1 < 78km$
level3	$78km \leq dist1 < 107km$	$78km \leq dist1 < 107km$
level4	$dist1 \geq 107km$ and $dist2 < 44km$	$dist1 \geq 107km$ and $dist2 < 27km$
level5	$dist1 \geq 107km$ and $dist2 \geq 44km$	$dist1 \geq 107km$ and $dist2 \geq 27km$

Table 4: Definition of variables *intensity*1 and *intensity*2, capturing the effect of variables *dist*1 and *dist*2 on *RAB* and *CR*, respectively.

affected by hurricane Leslie, considering variables *dist*1 and *dist*2. In this case, the regression tree algorithm has divided the space into five regions, for both *CR* and *RAB*, and the values of the response variable in each region are also reported in Figure 2, in the blue squares of the terminal nodes.

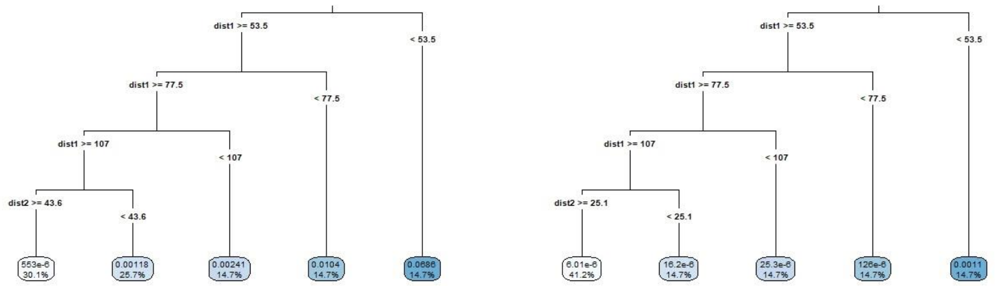

Figure 2: Partition of the space defined by variables *dist*1 and *dist*2 performed by the regression tree method applied to the observed *RAB* (left) and the observed *CR* (right) of the councils with reported claims due to hurricane Leslie. The leaves provide information on the predicted *RAB* (left) and *CR* (right), as well as the proportion of policies, in each node.

The obtained regression trees provide a mathematical description of the effect of variables *dist*1 and *dist*2, related to the storm's trajectory, on *CR* and *RAB*. We use this regression tree to build two categorical variables, denoted *intensity*1 and *intensity*2, one for *CR* and the other for *RAB*, respectively. These two categorical variables have levels defined by the splits in the regression tree, as described in Table 3. Thus, variables *intensity*1 and *intensity*2 capture the combined effect of *dist*1 and *dist*2 on the different values of *RAB* and *CR*, respectively. These are the variables that are used in the regression models for predicting claim frequency and claim amounts in light of the geographical exposure of the hurricane.

3.2 Modelling the affected councils based on the storm trajectory

First, we devise a model to classify which councils are affected by the storm event, that is, the councils with at least one claim, based on the trajectory of the storm.
The data set we use for this purpose comprises all the 278 councils of continental Portugal. Afterwards, in the next sections, we will estimate the claim frequency and claim amount on the affected councils.

To estimate the affected councils, three characteristics are considered for each council in our data set: distances *dist*1 and *dist*2 as previously explained, and a binary variable taking the values 1 or 0, if the council had at least one claim or not, respectively. For the sake of computational simplicity, we assume that the coordinates of the policies coincide with the coordinates of the respective council's main city. The distances *dist*1 and *dist*2 are thus used to predict a binary outcome, 1 or 0, if the council is affected or not, respectively. This is carried out employing a random forest model for classification based on the Breiman's random forest algorithm (see [3]).

Since the size of this data set is small, the predictive ability of the model is evaluated by means of a 10-fold cross-validation (see for instance [16] for an explanation on the k-fold cross-validation technique). The performance of the estimate is assessed through the mean and variance of the sensitivity (or positive rate) and specificity (or negative rate) of the cross-validation tests:

$$\text{sens.} = \frac{\text{num. true positives}}{\text{num. true pos. + num. false neg.}}, \ \text{spec.} = \frac{\text{num. of true negatives}}{\text{num. true neg. + num. false pos}}$$

The mean and variance of the sensitivity were 0.8593 and 0,0049, respectively, while the mean and variance of the specificity were 0.8121 and 0.0139, respectively, meaning a good predictive ability.

3.3 Modeling the claim frequency

We aim now at modeling the average number o claims for those councils that are affected by the storm. To this purpose, we only consider the affected councils, as using the whole portfolio would lead to biased estimates. The data set we are using, based on hurricane Leslie, is composed of more than 900917 policies on the affected councils. Of those, approximately 1% reported claims. In order to have a functional relation between the predicted probabilities of claims and the characteristics of the policies, a logistic regression (cf. [10]) is used to model the probability of the binary event if whether a given policy registers or not, a claim. As explanatory variables, we consider several characteristics of the policy, together with the variable *intensity*1, to account for the storm. The linear regression is represented in Equation (2). For a complete description of all the variables employed refer to Table 3.

$$\begin{aligned} log\left(\frac{p}{1-p}\right) = {} & \beta_0 + \beta_1 + \text{type of property} + \beta_2 \text{ year of construction} \\ & + \beta_3 \text{ framing of the housing} + \beta_4 \text{ type of the housing} + \beta_5 \text{ altitude} \\ & + \beta_6 \text{ type of floor} + \beta_7 \text{ forest area} + \beta_8 \text{ bush area} + \beta_9 \text{ intensity1} + \varepsilon \qquad (2) \end{aligned}$$

All the explanatory variables in our logistic regression (2) are categorical, thus each one will be represented by dummy variables (artificial variables taking the values 0 or 1), each one representing a different level of the explanatory variable (see Table 3). The estimation results are presented in Table 5.

Coefficient	Estimate	Std. Error	z value	p value	Signif. code
(Intercept)	-2.03009	0.04181	-48.553	< 2e-16	***
T.o.P Content	-2.35047	0.04111	-57.180	< 2e-16	***
Y.o.C Level 2	0.09722	0.03597	2.703	0.006875	**
Y.o.C Level 3	0.13160	0.03513	3.746	0.000180	***
Y.o.C Level 4	0.26766	0.03445	7.771	7.82e-15	***
Framing semi-det	-0.43324	0.03462	-12.514	< 2e-16	***
Framing other	-0.10034	0.04487	-2.236	0.025326	*
Type single-fam	0.67367	0.03321	20.285	< 2e-16	***
Type other	0.34408	0.05098	6.749	1.49e-11	***
Altitude Level 2	-0.34458	0.03101	-11.112	< 2e-16	***
Altitude Level 3	-0.61717	0.05267	-11.719	< 2e-16	***
intensity1 Level 2	-2.19481	0.04816	-45.577	< 2e-16	***
intensity1 Level 3	-3.37780	0.07508	-44.992	< 2e-16	***
intensity1 Level 4	-4.33357	0.11695	-37.055	< 2e-16	***
intensity1 Level 5	-5.29454	0.06855	-77.241	< 2e-16	***
T.o.F Level 2	0.25213	0.07144	3.529	0.000416	***
T.o.F ND	0.18343	0.03350	5.475	4.38e-08	***
Bush Area Level 2	-0.28077	0.02480	-11.324	< 2e-16	***
Forest Area Level 2	-0.55605	0.02629	-21.152	< 2e-16	***

Table 5: Summary of the coefficients estimated for regression (2) with the whole data set of those policies in the affected councils.

In the logistic regression (2), p represents the probability that the policy has a claim, that is $p \in (0,1)$. We need a model to decide what is the cut-off value of p above which we consider there is a claim. A randomized procedure, based on sampling outcomes 0 or 1 for each record, from a Bernoulli distribution using the probabilities estimated through the logistic regression, is adopted. To validate the adjustment of the logistic regression, we use a 10-fold cross-validation. The model quality is assessed based on its ability to predict, in the test data set, the average number of claims on a given risk class, and not for a single policy. The expected value and standard deviation of the weighted correlation[10] are respectively 0.8581 and 0.0012, and the expected value and standard deviation of the root mean squared error (RMSE) between the predicted values on the test set and the observed values on the test set are respectively 1.12 and 0.0946.

From the estimated results, we conclude that the odds of having a claim are 96% higher for single-family houses than for apartments. Also, the odds of having a claim

[10] By weighted correlation we mean the correlation between the predicted and observed values weighted by the proportion of policies in each risk class.

are almost 90% lower for the policies that are located in the area defined by the second level of the variable intensity, with respect to those located at level one. This means that the properties located at a distance inferior to 53.5*km* from the landfall point of the hurricane have 90% higher odds of having a claim, compared to those at a distance comprised between 53.5*km* and 78*km*.

3.4 Modeling the claim severity

Next, we aim at modelling the average cost of a claim, when it occurs. From our data set, regarding hurricane Leslie, the number of policies that reported a claim and thus represented a cost for the company was approximately 8500, representing approximately 1% of the whole policies of the affected councils. Remember that, for modeling purposes, only the councils classified as having at least a claim due to the storm are considered. If we consider the cost distribution relative to the whole portfolio over the affected councils, the distribution is highly right skewed, with most of its mass concentrated in 0. If instead, we consider the cost distribution relative to those policies, which reported a claim, the distribution is also highly right skewed, but there is no probability mass concentrated in 0. In the latter case, we were able to log-transform the cost distribution and observed that the log-cost distribution was well-approximated by a normal distribution. This allows us to use a multiple linear regression (MLR) model [10] to predict the average log-cost. We consider the MLR model of Equation (3). Again, the explanatory variables are characteristics of the insured property, together with a variable accounting for the storm, in this case *intensity*$_2$.

$$\log(\text{Cost} \mid \text{Cost} > 0) = \beta_0 + \beta_1 \text{ Type of Property} + \beta_2 \text{ Capital Insured} + \beta_3 \text{ Type of housing} + \beta_4 \text{ Urban area} + \beta_5 \text{ intensity2} + \varepsilon \tag{3}$$

Coefficients	Estimate	Std. Error	z value	p value	signif. code
(Intercept)	5.94432	0.03175	187.214	< 2e-16	***
T.o.P Content	-0.43345	0.04606	-9.411	< 2e-16	***
Cap. Ins Level 2	0.12030	0.03327	3.616	0.000301	***
Cap. Ins Level 3	0.30771	0.03441	8.942	< 2e-16	***
Cap. Ins Level 4	0.53860	0.03487	15.447	< 2e-16	***
Type single-fam	0.67329	0.02694	24.993	< 2e-16	***
Type other	0.48910	0.03829	12.774	< 2e-16	***
intensity2 Level 2	-0.25639	0.04639	-5.527	3.35e-08	***
intensity2 Level 3	-0.17290	0.07146	-2.419	0.015564	*
intensity2 Level 4	-0.59452	0.13755	-4.322	1.56e-05	***
intensity2 Level 5	-0.40922	0.06332	-6.463	1.08e-10	***
Urban Area Level 2	-0.20916	0.03763	-5.559	2.79e-08	***
Urban Area Level 3	-0.35671	0.02711	-13.159	< 2e-16	***

Table 6: Summary of the coefficients estimated for regression (3) with the whole data set of policies with claims in the affected councils.

The estimated values of model (3) are presented in Table 6. The prediction ability of the model is assessed on a 10-fold cross-validation. The quality measures are the following: the expected value and standard deviation of the weighted correlation are 0.9841 and 0.0173, respectively, and the expected value and standard deviation of the RMSE are 10370.34 and 6597.22. From the estimated results, it is interesting to see that the average claim cost, when there is a claim, for a single-family home is 96% higher than that of an apartment. Also, the average claim cost is 23% lower for policies located in the area defined by the second level of the variable *intensity*2, compared to those located at the first level of *intensity*2. This means that the properties located at a distance inferior to 53.5km from the landfall point of the hurricane have, approximately and on average, 23% higher claim costs compared to those located at a distance comprised between 53.5km and 78km. From Table 6 we can also see how the average claim cost tends to decrease as the concentration of urban areas in the council increases. This is expected as properties tend to be less exposed in urbanized areas.

Finally, we use regressions (2) and (3) to predict the claim counts and their costs. To evaluate the quality of the prediction, we split the portfolio of the company in the councils affected by hurricane Leslie in training and test sets, in a proportion of 60% and 40%, respectively. We test the model ten times in independent training and test sets. The expected value and standard deviation of the weighted correlation are 0.8143 and 0.0277, respectively, and the expected value and standard deviation of the RMSE are 2749.322 and 675.714, respectively.

4 Case scenarios in Portugal

In this chapter, we simulate the impact of a storm like Leslie if it landfalls in a different part of continental Portugal. For the purpose of analysing different scenarios, we have to simulate the trajectory of the storm and to chose the landfall point. Afterwards, for each simulated trajectory, we follow the methodologies described in Chapter 3. For simulating the trajectory, we need two elements that can be fixed or simulated: the storm's length and entrance angle. Since we assume that Leslie caused damages inland along 300 km, the length of the trajectory after its landfall is fixed at 300 km.

Again, for the sake of computational simplicity, we assume that both the exact point where the hurricane landfalls and the coordinates of the policies coincide with the coordinates of the respective council's main city. This implies that each policy belonging to a certain council has the same values of *dist*1 and *dist*2. Once the trajectory is drawn, the variables *dist*1 and *dist*2, and, subsequently, *intensity*1 and *intensity*2, are obtained for each policy.

In Figure 3, we present three maps of continental Portugal which highlight the cost ratio in the different councils assuming that a hurricane like Leslie reaches

Portugal in Cascais, Porto, and Faro with an entrance angle of 45, -30, and 60 degrees, respectively. The trajectory of the hurricane is also represented in the maps.

In Table 7, for each scenario, we present the number of claims, the cost, and the mean cost per claim (MCC), relative to hurricane Leslie. The MCC is computed as

$$MCC = \frac{\text{Total cost}}{\text{Total number of claims}}.$$

According to our simulations, if a hurricane like Leslie reaches Portugal in Cascais or in Porto, then the insurance company can expect a total number of claims that are approximately 3.22 and 2.61 times higher than the number of claims in Figueira da Foz due to Leslie. On the other hand, if the landfall point is Faro, then the insurance company will expect approximately half of the claims verified in Figueira da Foz. These results are expected since Cascais and Porto belong to the Lisbon Metropolitan Area and Porto Metropolitan Area, respectively, which are the regions with the biggest exposure for the company.

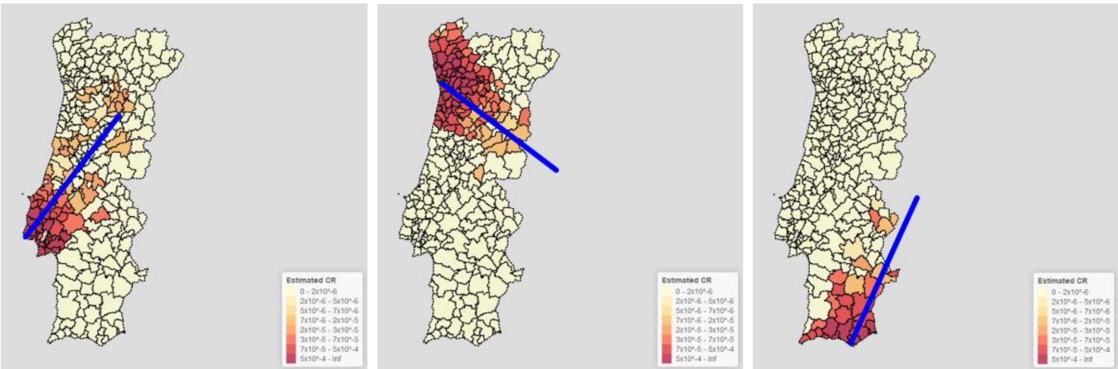

Figure 3: Cost ratio map of continental Portugal obtained for the following scenarios. Left panel - landfall point: Cascais, entrance angle - 45 degrees; Middle panel: landfall point: Porto, entrance angle - 330 degrees; landfall point: Faro, entrance angle - 60 degrees.

Landfall Point	Number of Claims	Cost	MCC
Cascais	3.22	2.08	0.65
Porto	2.61	1.99	0.76
Faro	0.51	0.45	0.87

Table 7: Cost, number of claims and MCC, relative to Figueira da Foz (benchmark), for the simulated scenarios of Cascais, Porto and Faro.

The estimated total costs in Cascais and Porto are similar and are approximately twice the total cost in Figueira da Foz. Analysing the total number of claims in Cascais and Figueira da Foz, we can expect higher costs than the ones estimated. To understand these results, we have to observe (i) the number of insured properties inside an area of 53.5 km around the landfall point, and (ii) the type of property insured in that region. The value of 53.5 km corresponds to the first level of the variables *intensity*1 and *intensity*2, and for which the coefficients in regressions (2) and (3) are the largest among all the other levels of these variables. Additionally, the variable type of housing is the one with the largest estimated coefficients. As seen before, according to the results in Tables 5 and 6, a single-family house has 96% higher odds of incurring a claim than an apartment, and the average cost for a single-family house is 96% higher than for an apartment.

Table 8 shows that the number of properties and the capital insured in an area of 53.5 km around the Cascais (resp., Porto) is 3.9 (resp., 3.1) times larger than in Figueira da Foz, which could indicate that both the number of claims and total costs should increase, similarly. However, Table 9 shows that both Cascais and Porto have a significantly smaller number of single-family houses when compared with the

Landfall Point	Portfolio size relative to the benchmark	
	Number of properties insured	Capital insured
Cascais	3.9	3.8
Porto	3.1	2.9
Faro	0.55	0.54

Table 8: Ratios between the size of the portfolio (number of properties insured or total amount of capital insured) in a radius of 54 km around the landfall point i and the number of properties insured in the same area around Figueira da Foz.

Landfall Point	Type of Housing		
	apartment	single-family house	other
Figueira da Foz	33.95 %	51.98 %	14.07 %
Cascais	69.99 %	14.94 %	15.06 %
Porto	43.96 %	40.46 %	15.58 %
Faro	56.33 %	31.52 %	12.15 %

Table 9: Concentration of the variable *Type of Housing* for those policies located in an area inferior to 53,5 km around the landfall point.

benchmark case in an area of 53.5 km around the landfall point. Since this type of house generates larger losses than the remaining types of housing, the losses increase but not as much as the number of claims. Additionally, a similar justification can be found for the fact that the estimated total cost in Porto is identical to the one in Cascais.

5 Risk map

Insurance companies with their portfolio in Portugal have experienced the impact of tropical storms in the previous years. These events might increase the premiums paid by the affected policyholders. However, the random nature of meteorological extreme events encourages insurers to look for solutions that mutualize the expected loss among other locations that could be affected in the future. In this chapter, we estimate the expected cost under different scenarios, by repeating simulations, as the ones obtained in Chapter 4, a large number of times. These scenarios are constructed by assigning different probability distributions to the (i) landfall council, (ii) the entrance angle, and (iii) the trajectory length. The following two scenarios are considered:

Scenario A: In this scenario, we assume that the councils in the coast of Portugal that are southern to Setubal are less likely to be hit by a hurricane, with a total probability of 2/6, than Setubal and the northern councils, with a total probability of 4/6. We assign equal probability to the councils in each of the two regions. The trajectory is assumed to have a fixed length of 300 km. For the West coast councils, the entrance angle is simulated according to a triangle distribution with support between -90 and 90 degrees with a mode of 45 degrees. For the South coast councils, the entrance angle is simulated according to a triangle distribution with support between 0 and 90 degrees with mode 45 degrees.

Scenario B: We now assume that the coastal councils can be hit by a hurricane according to a uniform distribution, which is with the same probability. The length of the trajectory is assumed to follow a continuous uniform distribution between 200 km and 400 km. The entrance angle also follows a continuous uniform distribution in the interval -90 degrees to 90 degrees, in West councils, and in the interval 0 degrees to 90 degrees in South councils.

In Figure 4 and Table 10, we present the distribution of the CR and the relative total cost, respectively, in continental Portugal for scenarios A and B. Despite the similarities between the two risk maps, we can easily see from Table 10 that the distribution of the total cost might be significantly different in different scenarios. Table 11 highlights those differences. We can observe that in scenario B the relative minimum total cost is significantly smaller than relative the minimum total cost in scenario A. This is due to the fact that in scenario A there is a higher likelihood that northern councils are hit by a storm. The north of Portugal is more densely populated, hence there is a higher exposure and, consequently, a higher probability of having small claim amounts. It is worth noticing that the total costs due to hurricane Leslie could be easily exceeded in both scenarios A (with probability 56%) and B (with probability 75%). In Table 12 we present the highest risk premiums in scenarios A and B.

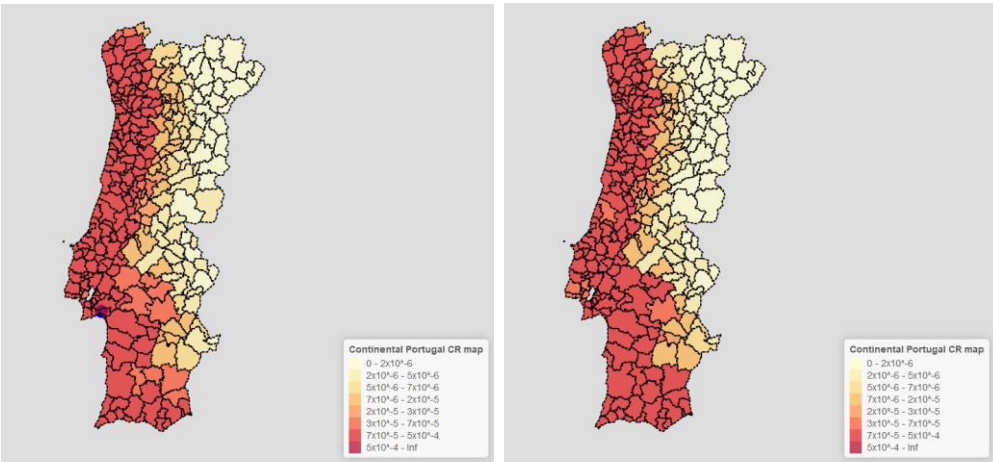

Figure 4: Distribution of the CR by council in continental Portugal due to a hurricane like Leslie in scenarios A (left) and B (right).

Relative Total Cost	Scenario A	Scenario B
TC < 0.5	30 %	12.4%
0.5 ≤ TC < 1	14.1 %	12.3 %
1 ≤ TC < 1.5	21.8 %	18.9 %
1.5 ≤ TC < 2.0	20.4 %	41.1 %
TC ≥ 2	13.7 %	15.3%

Table 10: Distribution of the Total Cost, relative to the total cost due to Leslie, for 1000 different simulated scenarios over continental Portugal.

	Scenario A	Scenario B
Min total cost	1.14	0.08
Max total cost	2.32	2.62

Table 11: Minimum and maximum values predicted for the total cost, relative to the total cost due to Leslie.

The risk premium is computed as the total cost per risk class divided by the number of policies in that class. The risk premium can be used by the insurer to adjust the policy premium to account for the storm risk. It is interesting to observe that the highest risk premium classes are the same in both scenarios. We can see that the highest risk premiums occur in Algarve for single-family houses. This is because, although Algarve is less urbanized than Lisbon or Porto regions, it is a region highly populated in the Summer due to tourism, and thus the portfolio of the insurance company in Algarve is composed of many luxury holiday single-family houses. This also explains the differences in the maximum total cost in scenarios A and B, since in scenario B there is a higher probability of a storm hitting Algarve.

Year of Construction	Framing of the Housing	Risk Premium	
		A	B
]2002,2018]	residential cluster	58,93	81.08
]2002,2018]	other	55,15	75.92
]1992,2002]	residential cluster	53,75	73.20
]1982,1992]	residential cluster	53,11	72.62
]1992,2002]	other	51,01	70.42

Table 12: Risk classes with the highest estimated risk premium. These risk classes are composed of buildings that are single-family houses, located in Algarve and with a capital insured greater than 165000€.

6 Conclusions

In this work, we simulate the impact of a storm like Leslie in the portfolio of the insurance company Fidelidade. Since climate data may not be reliable or may be difficult to deal with or obtain, we use only data on the policies of the property portfolio. The claims are seen as indirect observations of the intensity of the hurricane. The model is calibrated on data from hurricane Leslie, which hit Figueira da Foz, in continental Portugal, in October 2018. For other storm types, the model might need adjustments.

Several scenarios were simulated and risk maps were built from there. From our analysis, we can see that there is a high probability that a future event with the same intensity as Leslie will cause larger losses than Leslie. The simulated scenarios also allow computing the risk premium per risk class in the portfolio, which can be included in the policy premium calculation to account for this type of climate events. We have observed, from the simulations, that less populated regions may lead to higher losses than more urbanized areas. This depends on the exposure of the company, namely the type of property insured. Hence, this model provides a hurricane risk management tool for the insurance company.

Acknowledgments

This work was partially supported by FCT – Fundação para a Ciência e Tecnologia (Portugal), national funding through research grant UIDB/05069/2020.

References

[1] WJ Wouter Botzen. *Managing extreme climate change risks through insurance*. Cambridge University Press, 2013.

[2] Laurens M Bouwer. Observed and projected impacts from extreme weather events: implications for loss and damage. In *Loss and damage from climate change*, pages 63–82. Springer, 2019.

[3] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[4] Arthur Charpentier. Insurability of climate risks. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 33(1):91–109, 2008.

[5] Andrew F Dlugolecki. Climate change and the insurance industry. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 25(4):582–601, 2000.

[6] Markus G Donat, Tobias Pardowitz, GC Leckebusch, Uwe Ulbrich, and Olaf Burghoff. High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over germany. *Natural Hazards and Earth System Sciences*, 11(10):2821–2833, 2011.

[7] NOAA National Centers for Environmental Information (NCEI). U.s. billiondollar weather and climate disasters, 2021.

[8] Christian LE Franzke. Impacts of a changing climate on economic damages and insurance. *Economics of Disasters and Climate Change*, 1(1):95–110, 2017.

[9] Jerome H Friedman. *The elements of statistical learning: Data mining, inference, and prediction*. springer open, 2017.

[10] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.

[11] Clare M Goodess. How is the frequency, location and severity of extreme events likely to change up to 2060? *Environmental science & policy*, 27:S4–S14, 2013.

[12] Reindert J Haarsma, Wilco Hazeleger, Camiel Severijns, Hylke De Vries, Andreas Sterl, Richard Bintanja, Geert Jan Van Oldenborgh, and Henk W van den Brink. More hurricanes to hit western europe due to global warming. *Geophysical Research Letters*, 40(9):1783–1788, 2013.

[13] P. Heneka, T. Hofherr, B. Ruck, and C. Kottmeier. Winter storm risk of residential structures ndash; model development and application to the german state of baden-wu¨rttemberg. *Natural Hazards and Earth System Sciences*, 6(5):721–733, 2006.

[14] Eun-Soon Im, Jeremy S Pal, and Elfatih AB Eltahir. Deadly heat waves projected in the densely populated agricultural regions of south asia. *Science advances*, 3(8):e1603322, 2017.

[15] Thomas H Jagger, James B Elsner, and Mark A Saunders. Forecasting us insured hurricane losses. *Climate extremes and society*, 189:209, 2008.

[16] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[17] Ji-Myong Kim, Paul K Woods, Young Jun Park, Taehui Kim, and Kiyoung Son. Predicting hurricane wind damage by claim payout based on hurricane ike in texas. *Geomatics, Natural Hazards and Risk*, 7(5):1513–1525, 2016.

[18] Wei-Yin Loh. Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1):14–23, 2011.

[19] Evan Mills. Insurance in a climate of change. *Science*, 309(5737):1040–1044, 2005.

[20] Travis R Moore, H Damon Matthews, Christopher Simmons, and Martin Leduc. Quantifying changes in extreme weather events in response to warmer global temperature. *Atmosphere-ocean*, 53(4):412–425, 2015.

[21] Jeremy S Pal and Elfatih AB Eltahir. Future temperature in southwest asia projected to exceed a threshold for human adaptability. *Nature Climate Change*, 6(2):197–200, 2016.

[22] Richard J. Pasch and David P. Roberts. Tropical Cyclone Report Hurricane Leslie 23 SEPTEMBER–13 OCTOBER 2018 . Technical report, National Hurricane Center, NOAA, 03 2019.

[23] Jean-Paul Pinelli, Emil Simiu, Kurt Gurley, Chelakara Subramanian, Liang Zhang, Anne Cope, James J Filliben, and Shahid Hamid. Hurricane damage prediction model for residential structures. *Journal of Structural Engineering*, 130(11):1685–1691, 2004.

[24] B. F. Prahl, D. Rybski, M. Boettle, and J. P. Kropp. Damage functions for climate-related hazards: unification and uncertainty analysis. *Natural Hazards and Earth System Sciences*, 16(5):1189–1203, 2016.

[25] Boris F Prahl, Diego Rybski, Olaf Burghoff, and Ju¨rgen Peter Kropp. Comparison of storm damage functions and their performance. *Natural Hazards and Earth System Sciences*, 15(4):769–788, 2015.

[26] Steven C Sherwood and Matthew Huber. An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, 107(21):9552–9555, 2010.