SAFETY SCIENCE Monitor

Issue 3 2007

Article 5

VOL 11

DEPENDENCE IN THE "RISK = PROBABILITY * EFFECT" MODEL WHEN APPLIED TO TROPICAL CYCLONES MAKING LANDFALL

BOVEN, F.E. VAN,
A. Schweitzerplaats 94, 3069 GT, Rotterdam, the Netherlands, f.e.van.boven@capelleaandenijssel.nl

Key words: dependence, scenario, risk, tropical cyclones

Abstract - Risk models are used worldwide for developing safety measures in order to avoid as many victims and as much damage as possible after disasters. Even today there is no consensus about the definition of risk. The fact that the definition of risk is still being discussed raises the question when a risk model may be considered valid. In this study, the statistical dependence in the risk triplet $R = \{< s_i, p_i(\phi_i), x_i >\}$ is tested empirically on the basis of data on tropical cyclones that made landfall. The statistical hypothesis in this study involves testing the independence of the two variables scenario s_i and effect x_i. Statistical analysis on North American tropical cyclones that made landfall however, indicates that the risk model has to include dependence between scenario and effect to cover this type of event. This dependence is limited to the intensity of scenario and numbers of fatalities, the latter partitioned around low numbers (≤ 8). Besides variables such as safety measures, vulnerability and population density, dependence is an important factor for consideration in the process of reaching consensus on the unification of risk models.

INTRODUCTION

Countries around the world have developed safety policies in order to avoid victims and damage after disasters (Ale et al., 2006). In general, safety policies are based upon the concept of risk. Kaplan & Garrick (1981) defined risk as a combination of the frequency of the undesired event and the consequence of the event. Several investigators have found that risk is influenced by a number of different variables including safety measures, kinetic energy released by said event, the domino effect, vulnerability and population density (Adams et al, 2005; Dao & Peduzii, 2004; Thomalla & Schmuck, 2004; Thompson et al., 2001 Aarts & Schagen, 2006; Rappaport, 2000).

In 2000 the European Commission organised a workshop to explore the possible need for developing an internationally accepted generic ‘standard’ for risk based decision making, and to discuss the different perspectives on the subject (Kirchsteiger & Cojazzi, 2000). Sadly, five years later, in 2005, Europe remains unable to reach a consensus regarding standardizations for risk. Environmental Resources Managements (2005) reports a large variety in risk assessment methods identified in selected European countries. Environmental Resources Managements reports consequence based approaches, as well as frequency based and risk based approaches.

A valid, reliable and reproducible risk model is needed to enable its users to make adequate risk based decisions.

A valid model calls for knowledge on the statistical behaviour of variables event and consequence as defined by Kaplan and Garrick (1981). One important factor of statistical behaviour is dependence. The dependence factor in risk analysis is an active issue and methods for dealing with dependence are still very much under development (Cooke & Goossens, 2004; Dorp, 2005). Extensive knowledge on (in)dependence is necessary for obtaining confidence or prediction intervals for risk. The question to be solved in this study is whether the two variables probability (scenario) and effect are statistically independent in case of natural disasters. Disasters are limited in this study to North American tropical cyclones that made landfall.

MATERIAL AND METHOD

Over the years, several possible interpretations of the definition of risk have been presented and discussed (Gratt, 1987). Definitions can differ, depending on the scope of the study. Kaplan and Garrick (1981) defined risk as either of three possible triplets:

$$R = \{< s_i , p_i (\phi_i), x_i >\} \quad (1)$$

$$R = \{< s_i , p_i (\phi_i), \xi_i (x_i) >\} \quad (2)$$

$$R = \{< s_i , p_i (\phi_i , x_i) >\} \quad (3)$$

With:

R = risk of a disaster (or the expected number of fatalities per year)

s_i = scenario of the undesired event

p_i = probability of the scenario per year

ϕ_i = frequency of the i th scenario

x_i = effect when scenario takes place (actual number of fatalities)

ξ_i = probability of the effect of the i th scenario

This definition is often simplified to:

$$R = p_i * x_i \quad (4)$$

The probability factor has a stochastic character by definition. From a statistical viewpoint, it is more correct to consider scenario s_i as the independent random variable, rather than the parameter probability p_i. Probability p_i is a statistical parameter of the random variable scenario s_i. The effect 'x_i' is often used as a deterministic variable and calculated conditional. As far as the variable 'effect x_i' is concerned, it is clear that the number of fatalities as a result of any individual disaster cannot be estimated beforehand. However, over a longer period of time the number of fatalities may show a pattern, to which a probabilistic model may fit. In this study, the effect x_i is assumed to be random.

Beside different definitions on risk also different definitions on the variables of the used risk-triplet can be found. Kaplan and Garrick (1981) define scenario as a description of the event, "What can happen" or "What can go wrong". However, definitions of various scenario can differ. Brooker (2005), for instance, considers the elements that cause mid-air collisions not as facts that exist, but rather as judgements made about conditional futures and their consequences. Judgement, then, is about the calculus of probability of some undesired state of things that may come into existence at some future. In this study, scenario is defined as description of a hazardous event. In line with Dutch literature (Ale, 2002) hazard is defined here as "the intrinsic property of a chemical, physical, biological or psychological situation, with a potential for creating damage to human health or the environment". As scenario is considered a random variable, it can be described by statistical parameters such as intensity or strength, domino effects, frequency, space of time, location, population density threatened by the scenario. Further the period of development of the disaster – suddenly occurring or slowly progressing – is also characteristic of scenario (Becker, 2006).

For the variable effect in this study only fatalities are taken into account. Injury and economic damage are not taken into account. Fatalities can be defined in different ways. Adams et al. (2005) defines fatal as 'where deaths occur in less than thirty days as a result of the accident'. Another important consideration is whether either fatalities are related directly to the accident, or they are caused by an indirect effect of the primary event. In a study into causes and circumstances of flood disaster fatalities are defined to be only flood-related if the fatality would not have occurred without a specific flood event (Jonkman & Kelman, 2005). In this study, no specified information is available about the definitions used for fatalities. Only fatalities due to swimmers and surfers neglecting a cyclone hazard are not taken into account as they are not regarded as unintentional fatalities.

The thesis of this study is whether the two variables scenario and effect are in fact independent. If not, scenario and effect must be dependent. In that case, knowledge of the value of the first random variable would provide us some insight into the probable behaviour of the second variable (Larsen & Marx, 1986). This insight is necessary when constructing confidence intervals. In The Netherlands, risk outcomes are currently presented as point estimates. However, point estimates share a fundamental weakness: they provide no indication as to their inherent precision. What is needed is a combination of the point estimate's numerical value, including some sort of statement about its variability. This is known as confidence interval (Larsen & Marx, 1986).

Data were collected from the National Hurricane Center (NHC), Miami, USA about the North American tropical cyclones that made landfall for the years 1999-2006 (NHC, 2007). The Monthly Tropical Weather Summary gives data about the maximum wind speed and the number of fatalities during the tropical cyclone season. NHC distinguishes tropical depressions (maximum wind speed < 17 m/s), tropical storms (≥ 17 m/s) and hurricanes (> 33 m/s). Tropical depressions were removed from the sample as the effects of these depressions were assumed not to be stochastic. A remark has to be made on the definition of wind speed of tropical cyclones. In the United States the maximum wind speed is defined by the National Hurricane Center as 'maximum sustained surface winds', utilising a 1 minute average period and measured in knots (1 knot $\approx$ 1,852 km/h) (Atlantic Oceanographic and Meteorological Laboratory, 2004a). In most of the rest of the world a 10 minute average period is utilized for 'sustained wind', producing an approximately 12 % lower maximum wind speed (Atlantic Oceanographic and Meteorological Laboratory, 2004b). Depending on the chosen definition of the average period, the dataset contains a systematic error.

The tropical cyclone data provided a total of 83 observations for the period of 1999 - 2006. The tropical weather is given for the North Atlantic, the Caribbean Sea, the Gulf of Mexico and the Eastern North Pacific, from 140 degrees to 180 degrees west longitude. The null hypothesis to be tested is: X_i and Y_j are independent for $1 \leq i \leq r$ and $1 \leq j \leq c$. X is the explanatory variable for the maximum wind speed and Y the response variable for the number of fatalities. The data for tropical cyclones has been redefined so that response and explanatory variables both become categorized. More specifically because of the sample size, the data have been transformed to binary variables, creating a higher frequency of expected values. The intensity of cyclones has been categorised into two options: tropical storm or hurricane (> 64 kt or > 33 m/s). The number of fatalities has been partitioned into seven different categories.

Small numbers of samples in certain contingency tables requires the use of the Fisher Exact Test for significance. Calculations are performed with R 2.5.1 (Ihaka & Gentleman, 1996). The level of significance is taken as $\alpha = 0{,}05$. Poisson or negative binomial regression models are often used to analyse fatalities from road accidents (for example Chang, 2005). The use of a Poisson regression model instead of contingency tables will not change the final result of this present study, because the estimation of the test-statistic for generalized linear models is an equivalent process for contingency tables (after Dobson, 1994).

RESULTS

Thirty-one cyclones became tropical storms and fifty-two became hurricanes. Maximum wind speed ranged from 18 to 78 m/s with a median of 34 m/s. The number of fatalities ranged from 0 to 3000, with a median of 1. Cyclones with low wind speed but enormous consequences are relatively rare. The relative low frequencies of cyclones with low wind speed but high fatalities result in rejection of the null hypothesis, with respect to the chosen fatality partitioning (table 1).

Table 1. Exact p values for independence between intensity of tropical cyclones at landfall and the number of fatalities, the latter partitioned into seven different categories

Categorisation of fatalities	Exact p-value
0 - ≥ 1	**0.02**
≤ 1 - > 1	**0.002**
≤ 3 - > 3	**0.006**
≤ 8 - > 8	**0.03**
≤ 10 - > 10	0.054
≤ 20 - > 20	0.20
≤ 40 - > 40	0.25

Significant value in boldThe results show that in fact there is dependence between the intensity of a tropical cyclone at landfall and the number of fatalities, as long as the partitioning of number of fatalities takes place around low numbers (≤ 8).

DISCUSSION

Presented here is an empirical study of the relationship between scenario and effect using data on North American tropical cyclones that made landfall. The study involves testing whether there is independence between scenario and effect. Starting point of this study is the first risk triplet $R = \{< s_i, p_i(\phi_i), x_i >\}$.

When observing risk-characteristics of eighty-three tropical cyclones that made landfall in North America from 1999 till 2006, dependence was found between the intensity of scenario s_i and effect x_i. The dependence is limited to the intensity of the scenario and numbers of fatalities, the latter partitioned around low numbers.

When partitioning around low numbers of fatalities (≤ 8) the frequencies of cyclones with lower wind speed but enormous fatalities are relative rare, resulting in dependence. In case of partitioning around larger numbers of fatalities (> 8) all incidents with large number of fatalities become rare. In that case no dependence can be discriminated any more.

Although not all the cyclones used in this study are documented completely or unambiguously (Rappaport, 2000), and cyclones passing countries nearby were not taken into account, results are expected to be robust due to the sample size.

The strength of a hurricane is expressed in the United States by the Saffir-Simpson scale. The type of damage is theoretically associated with the Saffir-Simpson scale for hurricanes (Iman, 2005), although the damage of hurricanes can show a significant variability. Hurricane Charley illustrates the great uncertainty associated with modelling hurricane behaviour and its impact (Iman, 2005). The association found between the occurrence of hurricanes and the number of fatalities confirms the theoretical association of the type of damage with the Saffir-Simpson scale.

Rappaport (2000) studied the effects of Atlantic tropical cyclones during a 30-year period. Wind related injuries and drowning were the most important causes of death. Rappaport did not find a straightforward relationship between number of fatalities and the intensity of a tropical cyclone at landfall. However, no explanations for these results were given in Rappaport's study. Rappaport found that during the 1970-1999 period, most of the deaths occurred apparently at the time when local wind speeds were below the 64-kt threshold of hurricane intensity. With a few exceptions, the first storms of the season tended to be the most deadly, although this could be a statistical aberration. The results of Rappaport are not in accordance with results of this study and indicate that an analysis of this scenario should be studied in further detail.

In case of the tropical cyclones, further explanatory variables need to be introduced to study the influence of, for example, population density and type of building construction, which may differ within the studied area. In a descriptive study to cyclones in India, Thomalla and Schmuck (2004) concluded that 'at community level experience of the 1999 cyclone had clearly raised hazard awareness. As a result, disaster preparedness was high during the November 2002 cyclone'. This type of variable is not taken into account in this study.

The explanatory variable in this study is limited to the intensity of the tropical storm, represented by the maximum wind speed. Scenario is defined here as "the intrinsic property of a chemical, physical, biological or psychological situation, with a potential for creating damage to human health or the environment" (after Ale, 2002). The intensity of the tropical storm is assumed to be the most important statistical parameter describing the scenario. This assumption is due in part to the common practice of referring to a tropical cyclone by its most intense phase (Rappaport, 2000). However a physical phenomenon, such as a tropical storm, can be described by more parameters. Other parameters which are influencing factors in describing the potential for creating damage of a tropical storm scenario are frequency, space of time, development of the storm and location of landfall. Rappaport (2000) found also domino effects to be important.

One constraint regarding statistical power in this study is that the period of sustained high wind speed is not taken into account, whereas the frequency of maximum wind speed in one tropical cyclone is.

The frequency of the intensity of a tropical cyclone, equivalent to the probability of the cyclone is stated to be a statistical parameter of the scenario. Although frequency gives information on the number of times a tropical cyclone is likely to happen, it does not describe the intrinsic property of the physical phenomenon. By the very nature of natural disasters they are very rare but usually have enormous consequences. This suggests a linear dependency between frequency and intensity of the scenario, affecting the analysis of dependence in the risk model by means of collinearity (Myers, 1990).

For all the previously mentioned factors to a scenario (and not used in this study) the question rises whether they will be significant in a predictive model. In case of frequency of the scenario the question rises whether there is collinearity between intensity and frequency.

With regards to the model risk = scenario * effect, the dependence found in this study leads also to the question if the risk triplet can be simplified to just risk = scenario. In general this won't be possible because usually dependence is not hundred percent. But this question shows why dependence is important to consider in risk models. Currently risk outcomes are presented as point estimates. However, point estimates are not appropriate for establishing any form of inference on a future single observation (Myers, 1990). Confidence intervals are more appropriate estimations. When constructing confidence or prediction intervals for the risk of cyclones, dependence has to be taken into account. The reported dependence suggests that the risk of cyclones should be modelled by use of a copula approach or by generalized linear models. Generalized linear models are also used to model road safety.

Considering the definition of Brooker (2005) on the event, results of this study can not be generalized directly to other types of disasters. When studying risk assessment, tropical cyclones are merely one type of risk among many such as hazardous materials, nuclear installations, road safety, flooding, and earthquakes and so on. The question arises as to whether there is a relationship between scenario and effect for other types of disasters, as well as the models used to define and describe them.

Statistical analysis on North American tropical cyclones that made landfall between 1999 - 2006, indicates dependence between scenario s_i and effect x_i. The dependence is limited to the intensity of the scenario and numbers of fatalities, the latter partitioned around low numbers (≤ 8). Knowledge of the intensity of the scenario provides any insight into the probable behaviour of the effect for North American tropical cyclones. The reported dependence suggests the use of a copula approach or generalized linear models when constructing intervals for the risk of tropical cyclones. Besides factors such as safety measures, vulnerability and population density, dependence has been proven an important factor for consideration in the development or verification of reliability of a risk model in order to reach consensus on the unification of risk assessments.

REFERENCES

Aarts, L., I. van Schagen (2006,) Driving speed and the risk of road crashes: a review, Accident Anal. Prevent. 38, 215-224

Adams,J., White, M., Heywood, P. (2005), Year-round daylight saving and serious or fatal road traffic injuries in children in the north-east of England, Journal of Public Health 27 (4), 316-317

Ale, B.J.M. (2002), Risk assessments practices in The Netherlands, Safety Science 40, 105 – 126

Ale, B.J.M., P.W.M. Brighton, M. Baram (2006), Risk management, Safety Science Monitor, 10, article 2

Atlantic Oceanographic and Meteorological Laboratory (2004a), How do I convert from mph to knots (or m/s)?, Frequently asked questions, www.aoml.noaa.gov/hrd/tcfaq/A14.html

Atlantic Oceanographic and Meteorological Laboratory (2004b), What doest 'maximum sustained wind' mean? How does it relate to gusts in tropical cyclones?, Frequently asked questions, www.aoml.noaa.gov/hrd/tcfaq/D4.html

Becker, G. (2006), Comparability of risks, Safety Science Monitor, 10, article 5

Brooker, P. (2005), Reducing mid-air collision risk in controlled airspace: Lessons from hazardous incidents, Safety Science 43, 715-738

Chang, L. (2005), Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network, Safety science 43, 541–557

Cooke, R.M., Goossens, L.H.J. (2004), Expert judgement elicitation for risk assessments of critical infrastructures, Journal of Risk Research 7 (6), 643-656

Dobson, A.J. (1994), An introduction to generalized linear models, Chapman & Hall, New South Wales (Australia)

Dorp, J.R. van (2005), Statistical dependence through common risk factors: With applications in uncertainty analysis, Eur. J. Operat. Research, 161, 240 - 255

Environmental Resources Management (2005), Comparing notes: opportunities for learning and co-operation in external safety policy, Delft, The Netherlands (Reference 5129001 rapport)

Gratt, L.B. (1987), The definition of risk and associated terminology for risk analysis, In: J.J. Bonin and D.E. Stevenson (Eds), Advances in risk analysis 7, 675–680

Ihaka, R., R. Gentleman (1996), "R: a language for data analysis and graphics" J. of Computational and Graphical Statistics 5 (3), 299 – 314

Iman, R.L., M.E. Johnson, Watson, C.C. (2005), Sensitivity analysis for computer model projections of hurricane losses, Risk Analysis 25 (5) 1277-1297

Jonkman, S.N., Kelman, I. (2005), An analysis of the causes and circumstances of flood disaster deaths, Disasters 29 (1), 75-97

Kaplan, S., Garrick, B.J. (1981), On the quantitative definition of risk, Risk Analysis, 1 (1), 11–27

Kirchsteiger, C., Cojazzi, G. (2000), Promotion of technical harmonisation on risk-based decision making, Ispra (summary paper of an International Workshop organised by European Commission)

Larsen, R.J., Marx, M.L. (1986), An introduction to mathematical statistics and its applications, second edition, Prentice-Hall, New Jersey

Myers, R.H. (1990), Classical and modern regression with applications, second edition, PWS-Kent, Boston

National Hurricane Center (NHC), (2007), Monthly Tropical Cyclone Summary, USA, (www.nhc.noaa.gov/archive/), retrieved on August 21, 2007

Rappaport, E.N. (2000), Loss of life in the United States associated with recent Atlantic tropical cyclones, Bull. Amer. Meteor. Soc. 81 (9), 2065 – 2073

Thomalla, F., Schmuck, H. (2004), 'We all knew that a cyclone was coming': disaster preparedness and the cyclone of 1999 in Orissa, India, Disasters 28 (4), 373-387

Thompson, K.M., R.F. Rabouw, Cooke, R.M. (2001), The risk of groundling fatalities from unintentional airplane crashes, Risk Analysis 21 (6), 1025–1037