1

Fundamentals

Matthew Jones, Kirsten Mitchell-Wallace, Matthew Foote, and John Hillier

1.1 Overview

1.2 Catastrophes, Risk Management and Insurance
1.3 What Are Catastrophe Models?
1.4 Why Do We Need Catastrophe Models?
1.5 History of Catastrophe Models
1.6 Who Provides and Uses Catastrophe Models?
1.7 What Are Catastrophe Models Used For?
1.8 Anatomy of a Catastrophe Model
1.9 Model Input
1.10 Model Output: Metrics and Risk Measures
1.11 Statistical Basics for Catastrophe Modelling
Notes
References

1.1.1 What Is Included

This chapter contains a broad overview of the topic of catastrophe modelling, including what catastrophe models are, why they are used, their overall structure and their output. Metrics used in catastrophe modelling are presented. Basic statistical concepts required for catastrophe modelling are also included for ease of reference.

1.1.2 What Is Not Included

Detailed information on every topic is not included. This is provided in the subsequent chapters.

1.1.3 Why Read This Chapter?

This chapter aims to give the reader an introductory background to catastrophe risk management and catastrophe modelling. It is targeted primarily at those new to the subject, but should also provide a refresher to those more familiar with the discipline. Reading this chapter together

Natural Catastrophe Risk Management and Modelling: A Practitioner's Guide, First Edition.
Edited by Kirsten Mitchell-Wallace, Matthew Jones, John Hillier and Matthew Foote.

with any subsequent chapters should provide depth on the topic covered – be that the main uses of models, a discussion of the major perils, how to build a model, or how to develop a view of risk. Alternatively, this chapter can be read in isolation to provide an introduction to catastrophe risk management and modelling, from the basics of insurance to the elementary statistics required when using these models. The statistical basics are provided for completeness and reference; less mathematically-minded readers can avoid this section without compromising understanding of following chapters.

1.2 Catastrophes, Risk Management and Insurance

In its broadest sense, a **catastrophe** is something that exceeds the capability of those affected to cope with, or absorb, its effects; in the context of natural hazards the driver is an extreme event causing widespread and, usually sudden, damage or suffering. In the insurance industry, definitions of catastrophe are commonly based on an event exceeding one of a number of thresholds for loss (e.g. total economic losses, insured losses, loss of life – for an example by Swiss Re, see Table 1.1). Organizations may choose to define an event as a catastrophe if that company or the whole industry has large or unexpected losses or if significant media attention is expected. For example, the US Property Claims Service definition of a catastrophe is 'an event that causes 25 m USD or more in direct insured losses to property and affects a significant number of policyholders and insurers'.

The terms risk, peril, and hazard are often used interchangeably in conversation. However, in the context of this book, we use the following definitions:

- A **peril** is a potential cause of loss or damage such as an earthquake or windstorm.
- **Risk** is uncertainty leading to potential adverse outcomes. It is also used as shorthand for an insured object.
- **Hazard** is the danger from the peril.

Catastrophes are a risk to organizations and society. Managing this risk (**catastrophe risk management**) is the ongoing process of: (1) identifying the risk given the context of the organization or community, (2) quantifying the risk, (3) deciding what to do, given the level of risk and the **risk appetite** (i.e. how much risk an entity is willing to take) of the organization or community, and (4) monitoring the level of risk.

Table 1.1 Criteria used by Swiss Re in 2014 to determine if events were categorized as catastrophes and entered into their *Sigma* database.

Threshold	Quantity
Insured loss, maritime disasters	US$19.6 m
Insured loss, aviation	US$39.3 m
Insured loss, other losses	US$48.8 m
Total economic loss	US$97.6 m
Casualties, dead or missing	20
Casualties, injured	50
Casualties, homeless	2000

Source: Swiss Re, 2015.

The concept of **enterprise risk management (ERM)** (Sweeting, 2011) involves the preceding process, but on a holistic basis (i.e. assessing all risks together, allowing for diversifications and concentrations of risk, including risks that are easy to quantify and those that are not, e.g. reputational damage). The classic responses to risk are to reduce, avoid, transfer or retain (Sweeting, 2011). Insurance is one important mechanism to transfer risk.

Insurance is an arrangement whereby one party (the **insurer**) promises to pay another party (the **policyholder**) a sum of money in the case of a loss as a result of a specific cause. This obligation to provide compensation following a loss is called **indemnity**. A premium is charged to the policyholder to provide this service. Insurance companies provide products for individuals (**personal lines (PL)** insurance) and for corporations (**commercial lines** insurance). Insurance companies who provide money contingent on whether someone dies are called **life insurers**, whereas other insurance companies are called **non-life insurers** or **general insurers**. Companies that provide both life and general insurance are called **composite insurers**. The focus of this book is on general insurance, which can be categorized into different **lines of business**, depending on the type of assets that are being insured; for example:

- **Motor** (or **auto**) lines provide insurance for the physical car and sometimes also for the third-party liability.
- **Property** (or **Direct & Fac (D&F)**) lines provide insurance for properties, their contents and loss resulting from not being able to use the buildings because of an insured peril.
- **Marine** lines provide insurance for ships (**hull**) and the **cargo** they carry (often including goods not actually on a ship, but in transit and in warehouses; sometimes called **marine static risk**).
- **Aviation** lines provide insurance for aircraft, including third-party liability coverage.
- **Construction** (or **engineering**) lines provide insurance for building projects.
- **Liability** (or **casualty**) lines provide insurance to cover claims from third parties.

The precise origins of insurance are debated, but so long as there has been risk, people have tried to manage their individual exposure to it. The first record of insurance was the Babylonian King Hammarubi's code, an ancient tablet dating back to approximately 1750 BC. Insurance's origins were certainly in trade; Phoenicians and Greeks had similar schemes to minimize the impact of the potentially catastrophic (to any individual) loss of ships and cargo by forming a pool to spread the loss. Arrangements similar to modern marine insurance were in place by the mid-fourteenth century in Genoa. An explosion in trade in the seventeenth century led to a maritime information exchange in Lloyd's coffee shop in 1688, with the first recorded **underwriting** in 1757. After an insurance proposal was drafted, the participants all signed their names and participation on the risk underneath the proposal leading to the term 'underwriter' for those taking on the risk. Underwriting will be discussed together with insurance and **reinsurance** (the insurance of insurance companies) in Chapter 2.4. Meanwhile, in America, Benjamin Franklin founded the Philadelphia Contributorship in 1751.

The development of the (re)insurance industry, like modelling itself, has been driven by events. Large fires across Europe have been a driver for property insurance, notably the Great Fire of London in 1666. Before municipal fire-fighting facilities, insurance companies had their own services to protect the specifically marked properties of policyholders. The Hamburg Fire in 1842 precipitated the foundation of the first reinsurance company, Cologne Re. The foundation of Swiss Re has likewise been linked to the Glarus Fire of 1861. These severe events demonstrated the need for reinsurance, although many reinsurance companies were in fact founded to prevent the outflow of reinsurance premiums from local economies to foreign companies (Swiss Re, 2013). For details of the more recent evolution of insurance history, including the London Market Spiral in 1990s, see, for example, Thoyts (2010).

A fundamental concept of insurance is that **pooling risk** reduces the uncertainty in the **expected** (or average) **loss** (EL) over a specific time period. Put another way, the cost of losses, in any given year, from *a large number* of insured properties is *much more certain* than the loss cost from any *individual* property. An insurance company can, therefore, estimate the overall loss cost from a portfolio of insured properties with far more certainty than an individual could for one **policy**. In addition, for some loss scenarios (e.g. a property completely destroyed), the individual may be unable to cover the loss themselves. An insurance contract therefore provides the individual with much more certainty about the amount they will have to pay in any one year (the insurance premium, plus potentially an excess, see Section 1.9.2) as well as protection against an unaffordable loss. A more mathematical description of the rationale for insurance, in particular pooling of risk, is provided in Box 1.1.

Box 1.1 How Pooling of Risk Works

To illustrate how insurance works, consider an individual who owns a house worth £1 million, and that every year there is a 1% (0.01) chance that the house will be destroyed by an earthquake. Let us also assume that there are no other hazards and that the only outcomes are that the house is either completely intact or completely destroyed each year. The average losses expected each year, or expected loss, is relatively low compared to the value of the house (£1m $\times$ 0.01 = £10,000), however, the uncertainty in the amount of the loss is very high – either all or nothing. A measure of the uncertainty in an outcome is the **standard deviation (SD)**; see Chapters 9 and 10 of Sweeting (2011) and Section 1.11 for some basic statistics. Quantitatively, the standard deviation of the annual loss for this scenario is £99,499 (£1m $\times \sqrt{0.01 \times (1 - 0.01)}$, see Section 1.11). The **coefficient of variation** (the standard deviation divided by the expected loss cost) is 9.949.

Now consider a group of 1,000 such people with identical houses and an identical risk to earthquakes, but in different locations such that only one house can be affected by each earthquake, i.e. the risk to each house is completely independent. The expected loss per house remains at £10,000, so £10,000,000 in total across the 1,000 houses. The standard deviation for the group of independent houses is given by the square root of the sum of the individual variances and is equal to £3,146,427, giving a coefficient of variation of 0.3146, and a *standard deviation per person of £3,146*: much less than the £99,499 for the individual scenario. So although the expected loss cost per person stays the same, *the uncertainty in loss cost per person is much reduced by the pooling of risk*. This is also known as **diversification** benefit and is one fundamental reason why insurance makes sense as a concept.

An individual who wishes to avoid the potential financial ruin of an earthquake fully destroying their home can, therefore, choose to insure their house to take advantage of this pooling of risk. In doing this, they would have to pay an insurance premium that covers: (1) the expected loss, (2) a contribution to the expenses of running an insurance company, and (3) an amount to compensate the insurance company for retaining the (reduced because it is pooled) risk. This is the basis of insurance pricing discussed in Chapter 2.6.

The first point to note with the example above is that an assumption of independence of risk was introduced. Mathematically if, instead of the risks being independent, each risk was perfectly correlated so that all risks would be affected in the same way, the standard deviation of total loss would simply be the sum of the individual standard deviations; there would be no reduction in the coefficient of variation (or the SD per individual) and no advantage from pooling of risk.

Therefore, if risks are highly correlated, insurance is less economically attractive since the premium that the insurance company charges must be higher to cover the increased risk.

The second point to note is that, even when aggregating independent risks (i.e. considering their behaviour as a group), there is always some risk remaining which the insurance company must absorb. To do this, companies have **capital** (an excess of assets over their liabilities) to protect the insurance company against their *unexpected* losses, i.e. the chance that the losses in any given year could be more than expected (the average). Insurance companies will often buy insurance themselves (reinsurance) to help further protect themselves against large unexpected losses. Companies need to understand the risk they face in order to ensure they have sufficient capital or, put another way, insurance companies should only take on risk commensurate with the capital they have. Models help the companies understand their risk.

This book is about catastrophe risk management and modelling, and although in this context an earthquake completely destroying a house is quite realistic, the assumption of complete independence of risk used above is very unrealistic. In practice, the wide spatial scale of catastrophe events means that different properties can experience the same catastrophic event at the same time and so there is little independence between the risks within an event. Understanding just how independent (or correlated) the risks are, is a very important (and challenging) consideration for insurance companies, and is one of the main reasons that catastrophe models exist.

1.3 What Are Catastrophe Models?

Catastrophe models are models designed to estimate the potential loss from the extreme and wide-impact events that are termed catastrophes. The loss potential estimated by such models is usually financial.

Although various definitions of catastrophe models can be found in academic, developer, user or regulatory communities, the one used here is that provided by the United Kingdom Lloyd's Market Association (LMA, 2013):

> A catastrophe model is a computerized system that generates a robust set of simulated events and estimates the magnitude, intensity, and location of the event to determine the amount of damage and calculate the insured loss as a result of a catastrophic event such as a hurricane or an earthquake.

Catastrophe models, like all models, are abstractions of real-world processes. This determines their appropriateness and limits their use and interpretation. These models combine the science of natural hazards with engineering, socio-economic and financial processes (see Chapter 4.2). This amalgam is complex and depends on the evolving knowledge of the processes and the connections between them. The models themselves are products of specialist assumptions, which depend on scientific knowledge, often itself derived from other research or modelling sources. There are many complexities and uncertainties in such models.

However, catastrophe modelling allows the limited historical record for these losses to be extended beyond past events into the realm of what might plausibly occur and provides a framework to quantify the current risk from a catastrophe peril to a group of assets. Increasingly, catastrophe model users are required to understand the models' construction and justify each model's use in terms of appropriately representing the risk facing their particular organization (see Chapter 5).

1.4 Why Do We Need Catastrophe Models?

Understanding the risk associated with insurance is a fundamental aspect of managing an insurance or reinsurance company. This understanding ensures that the price charged is sufficient and that the capital and reinsurance protection of the company are adequate, given the risk faced.

For many types of risk, past insurance losses experienced by a company (its **claims experience**), are used to develop models to help the risk estimation. This is the mainstay of actuarial (i.e. mathematics applied to insurance) work, and is a good approach provided that: (1) sufficient claims data exist, and (2) there is a good means of bringing the historical claims to a level that reflects today's risk whether for the change in exposure or the different cost of paying the losses in today's environment (known as **on-levelling** the claims data). This is discussed more in Chapter 2.6.2.1.

For risk as a result of catastrophes, both of these criteria are problematic; **catastrophe risk**, by definition, must include the risk posed by large and unlikely events. Typically, a 10- or 20-year time series of insurance claims data is available, which is insufficient for the purpose of estimating extreme risk, such as the amount of a 1-in-a-200-year loss. However, even if a 200-year time series of losses did exist, the underlying trends (or cycles) within the data would likely make it unusable. Such trends include:

- changes in the type and location of underlying exposure, for example, the trend of increasing urbanization;
- changes in building standards, for example, wind-loading codes;
- changes in infrastructure, for example, flood defences;
- inflation, for example, increases in the cost to rebuild property;
- climate change, for example, increased sea levels or the multi-decadal cycle in North Atlantic hurricane activity.

These trends can be mitigated by using data that reflect the assets that are currently insured, together with a scientific representation of the risk posed by the peril today.

To provide this scientific representation of risk, an insurer could turn to academic work on the perils of interest and consider this, together with their insured assets, to form their risk assessment. However, this is extremely challenging and resource-intensive for two main reasons:

- *Catastrophe risk assessment involves multiple disciplines.* The insurer would need to bring together scientific experts in the perils (e.g. meteorology, seismology or hydrology, see Chapter 3) as well as engineers who can quantify the damage that these perils can cause. In addition, particular aspects of the discipline used for catastrophe risk assessment may fall outside the standard research areas of the scientific discipline or the interests and incentive structures that apply to the researchers (e.g. set by grant-awarding bodies).
- *Catastrophe risk assessment needs to include (re)insurance financial structures.* **Financial structures** within insurance policies are commonly used for the mitigation of risk (see Section 1.9.2). These can be complex and little understood outside of the (re)insurance industry. In order to do this, statisticians and actuaries are needed as well as software engineers to build a **platform** with associated data schema, so that the financial structures can be properly captured and applied.

Even with an established software platform, a high-quality catastrophe model, for a significant peril, can feasibly take around 50 person-years to build, and once built must be updated every few years. Because an insurance company will be exposed to multiple perils, a substantial team (multiple tens of people) is needed if the company is to construct their own models – even just for the main perils.

The resource-intensive nature of this process suggests that for most insurers there is a benefit in outsourcing model-building capabilities and sharing the development costs with other companies. Put another way, there is a market demand for companies that build catastrophe models. Significant catastrophe events help to crystallize this need, while the advent, and increasing prevalence, of computers have provided a mechanism for delivering models. So in the 1980s, the first catastrophe modelling companies (referred to as **vendors**) were formed.

1.5 History of Catastrophe Models

Traditionally, the beginning of catastrophe modelling is considered to be in the 1960s; pioneered by Don Friedman through research instigated by Travelers Insurance Company (Friedman, 1984). However, it was the late 1980s when the first commercially-produced model platforms were released. Grossi and Kunreuther (2005) provide a good background to the historical development of catastrophe models, and clearly highlight their evolution, identifying key developmental milestones across a range of technological, data, industrial, regulatory and disaster events, as well as the ongoing growth in model sophistication and peril coverage. New data, methods and scientific understanding have led to periods of significant model update since the late 1980s. Most recently, the advent of cloud computing and open architecture modelling has combined with the increased demand for quantitative risk models from the **disaster risk financing** (DRF) community (e.g. Ley-Borrás and Fox, 2015) to generate an upsurge in model provision, including a much wider community of model developers and users.

Figure 1.1 shows a simplified timeline of catastrophe model development milestones, including major catastrophe events, industry changes, key regulatory standardization, data and applications, commercial catastrophe modelling organizations, international academic and research initiatives, and technological milestones. Although only illustrating a small sample of activities and events, it shows how catastrophe modelling has evolved over its lifetime. The advent of computational analytics for weather forecasting and for seismic processes set the scene for the first computational models developed to estimate the risk from natural disasters. These provided the tools and data (illustrated by the 'i' symbol in Figure 1.1), particularly from government, academic and international organizations considering disaster safety and management. In addition, the development of engineering and insurance standards, including ATC in 1973 and the CRESTA organization in 1977, provided the framework around which models could integrate scientific models of hazards with engineering vulnerability data. A number of large catastrophic events (illustrated by the 'lightning' symbol in Figure 1.1), including Hurricane Hugo, and the Loma Prieta earthquake (both in 1989), highlighted the need for the insurance industry to better understand its risk to large-scale, infrequent loss events. The convergence of the capability to augment historical actuarial data with scientific hazard models and a commercial demand for quantitative risk assessment from the insurance industry, especially in the United States, led to the first commercial models from AIR and RMS in the late 1980s. In addition, the digital spatial frameworks around which hazard, exposure and vulnerability could be integrated were being developed in the **Geographic Information Systems** (GIS) community.

The cluster of large catastrophic events in the 1990s which affected the United States, Japan and Europe (Hurricane Andrew in 1992, the Kobe earthquake in 1995 and wind storms 87J and Daria in 1990) intensified the demand for additional and more sophisticated computational modelling for those territories and perils. Catastrophe modelling enabled more quantitative assessment of insurance risk management and reinsurance transactions and led to the rise of

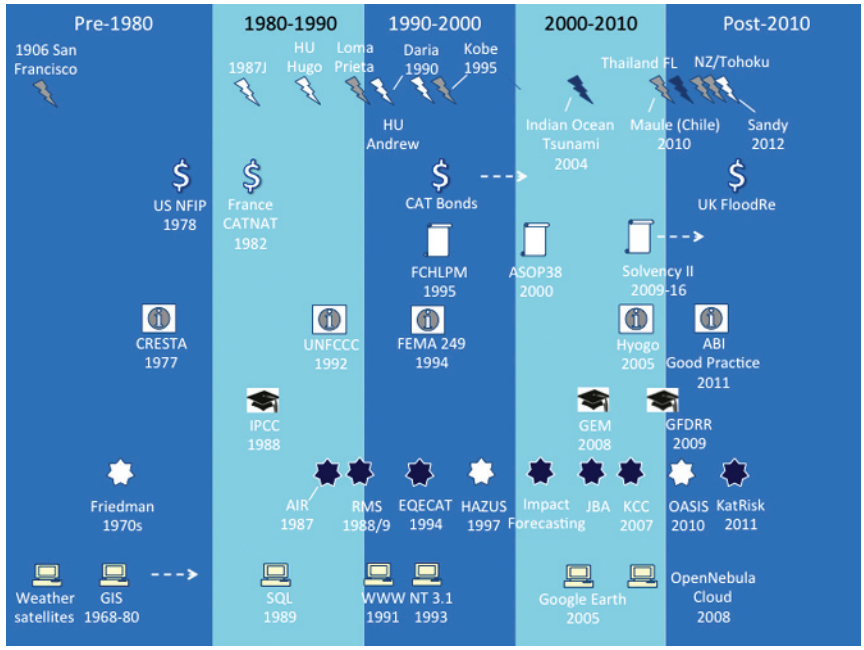


Figure 1.1 History of catastrophe modelling. Time runs left to right. From top to bottom, symbols representing activities and events are: Lightning bolt = natural hazard event; $ = financial idea or mechanism; scroll = regulation; i = data or information; star = modelling company launched; mortar board = academic contributions; computer = computational development. For acronyms, see the Glossary.

new insurance products that took advantage of the catastrophe model as a means to define both risk and price; particularly catastrophe bonds which rely on models to define the triggers and loss terms under which both the insureds and investors could operate with confidence (see Chapter 2.15). A third significant commercial modelling vendor, EQECAT, was formed in 1994, and the US FEMA organization's HAZUS multi-peril platform was first released in 1997. At this time data collection by the insurance industry began to increase in quality, granularity and coverage, for example, with the use of postcodes to collect and report on accumulations of exposure at relatively high levels of detail. The creation of high resolution digital boundaries for **CRESTA** zones, postcode and other datasets for key regions, as well as the use of **geocoding** technologies drove the increased use of electronic accumulation systems, especially for US and European insurance companies. The increase in computational power in the 1990s enabled the widespread use by insurers of database management systems such as Oracle, SQL and desktop spreadsheet systems such as Lotus and Excel. This gradually led to the shift of catastrophe models from the back office mainframe computer environment towards the actuarial and underwriting business support areas in both insurance and reinsurance. Higher resolution and more sophisticated models were made possible by increases in computational power, and the availability of higher resolution data, for example, **digital elevation models** (DEMs). This enabled other perils, such as floods, to be included for the first time in catastrophe modelling platforms.

The 2000s were characterized by the growth of the Internet for data and scientific research, and this enabled catastrophe modelling companies to further widen their engagement with insurance companies. In particular, Google Earth in 2005 (a spatial visualization platform combining satellite and aerial imagery) significantly widened the use of spatial data and

visualization by insurers at the desktop. This was a major step in the development of 'point of underwriting' quantitative risk assessment and also drove the development of higher resolution and more complete data for model construction, validation and calibration.

The internationalization of climate change concerns, most particularly in relation to the Intergovernmental Panel on Climate Change (IPCC) Third and Fourth Assessment Reports (in 2001 and 2007 respectively) also heightened insurance and reinsurance concerns and demand for catastrophe modelling, as did other events, such as the Indian Ocean tsunami in 2004 and Hurricane Katrina in 2005. However, perhaps the most important driver of catastrophe modelling development in the 2000s was the increased regulatory demand for quantitative risk management and capital rigour, including the US Actuarial Standards Board (2000), which defined actuarial standards for catastrophe risk assessment and model use, and **Solvency II** in Europe, which led to the 2011 publication of the Association of British Insurers' 'Good Practice' guidelines for catastrophe modelling (ABI, 2011).

The first half of the 2010s has seen significant changes in the catastrophe modelling community, again driven by a combination of large-scale events (in Chile, the 'Maule' earthquake (2010), the New Zealand earthquakes (2010 and 2011), the Thailand flooding (2011), the Tohoku 'Great East Japan' earthquake (2011) and Hurricane Sandy (2012)). These events led, in nearly all cases, to a reassessment of the scientific and methodological bases of the models, and to large-scale revisions of the models and their loss estimates, or the construction of new models. In some cases, these new models have been produced by organizations other than the original commercial modelling companies, including regional specialists as well as reinsurance intermediaries and governmental/academic organizations. These are illustrated in Figure 1.1 as a star (blue star denotes a commercial organization, and a white star denotes a not-for-profit organization). The role of international disaster risk management and finance groups, such as the World Bank, in driving catastrophe model use and development has also grown. This has been in parallel with the advent of the cloud as a potential computational framework for the next generation of high performance and in some cases, open access catastrophe modelling platforms, including the Oasis loss modelling framework (see Chapter 6).

The insurance industry has also evolved over the 2000s and 2010s both in terms of its requirements (regulatory and business) and in how catastrophe modelling is integrated within the business function. Increasingly, catastrophe modelling results are being used at the point of underwriting to inform decisions on price, return on capital and capacity (see Chapter 2.6). This in turn is driving advances in data quality, and in the visualization of information for effective decision making. The need for 'model completeness' and validation of the use of models to represent the view of risk (see Chapter 5) will continue to drive demand for increased model resolution and coverage, including the widening of model perils and regions across non-traditional areas.

In addition, there will continue to be growth in commercial model coverage, both geographically, and in terms of perils, in response to market demand and new scientific advances. The Key Past Events sections in Chapter 3 provide an overview of how events have driven major changes to model availability.

It is, perhaps, not surprising that major updates to models have had some impact on their interpretation and use in the (re)insurance industry. Updates have, in some cases, led to considerable challenges, including revising business plans and capital requirements. These have been amplified by the increased complexity of the models being developed and the increased choices available to users. In particular, changes in US hurricane models over the last decade, which have reflected major updates in fundamental assumptions and modelling approaches, and the provision of multiple viewpoints of event frequency, have resulted in a more careful and considered approach to the use and application of catastrophe models.

1.6 Who Provides and Uses Catastrophe Models?

Catastrophe models are used across multiple segments of the (re)insurance industry as well as, increasingly, in the disaster-risk financing communities. In (re)insurance, these models are used in businesses which take on risk themselves (insurers, reinsurers, insurance linked security investors) as well as those who advise them ((re)insurance brokers and consultants) and assess their activity (regulators and rating agencies).

Model provision and use can be explored using the concept of model agents, all of whom have specific influence in, and requirements of, the design and function of a model. Four main roles are considered: contributors, developers, analysts and assessors/overseers. Individuals may, of course, assume more than one agent role during the modelling process.

- Catastrophe model contributors are the creators of the constituent scientific, mathematical and statistical theories and methods used within the models or their components. They include primary source data providers, software and model process designers and suppliers. Contributors may not have designed or produced the components specifically for use within a catastrophe model, and instead may have developed them for other purposes (e.g. flood-risk maps).
- Catastrophe **model developers** derive model components, integrate, calibrate and validate them while creating loss modelling systems or platforms. Developers may be collaborative academic, industrial or commercial organizations producing either generic or bespoke models.
- The **model vendors**, the specialist companies who develop and license catastrophe models, are model developers; these include AIR Worldwide, Ambiental, Catrisk Solutions, Corelogic (formerly EQECAT), ERN, Impact Forecasting, JBA, KatRisk, KCC, and RMS.

  The evolution of **plug and play** open architecture model platforms and formats (i.e. the idea that a new model, or model component can be 'plugged in' to an existing model platform and quickly and easily used), including Oasis, expands the roles of developers and contributors, and future model construction and design are likely to reflect an evolution in contributor and developer roles.
- **Catastrophe risk analysts** are responsible for data entry, model selection, model operation, analysis, outputs and reporting. Catastrophe risk analysts can be considered the *primary users* of catastrophe models. The analyst is responsible for ensuring that the characteristics of the model system are tuned to be as representative as possible of the objects (hereafter referred to as **risks**) being assessed. In particular, the analyst will often be responsible for the input of **exposure data** (see Section 1.9), which are the data captured by the organization, representing the location, characteristics and value of the assets being insured. Catastrophe risk analysts often have a background in science, engineering or mathematics and may have a specialization in a hazard or engineering discipline. These backgrounds are helpful in interrogating the model components and developing a view of the suitability of the model for use within the organization (see Chapter 5). Depending on the structure of the company, there may be an overlap between the activity of the catastrophe risk analyst with the traditional roles of both the underwriting and actuarial departments. For example, catastrophe pricing can be carried out by a catastrophe risk analyst or by an underwriter or actuary. In some territories (particularly London), organizations differentiate between exposure management and catastrophe modelling, with catastrophe modelling being a subset of the broader field of exposure management. In other territories, catastrophe modelling, catastrophe management and exposure management teams often do the same thing.
- Other potential direct users of catastrophe models within a (re)insurance company may include **actuaries** who are responsible for many aspects of statistical analysis and pricing, **underwriters** who select and price risk, and capital modellers who determine the relative solvency of the (re)insurance entities. However, it is more likely that people in these roles will

use the model output rather than using the models directly. Senior management (e.g. chief risk officers, chief underwriting officers) will not usually use the models directly but will rely on **risk metrics** from these models to inform fundamental business decisions related to catastrophe risk.

Catastrophe modelling is a cross-functional activity and therefore can be viewed either as part of risk and exposure management or part of underwriting support. The organizational structure and reporting line of the catastrophe modelling, or exposure management, team can provide insight into how the team is viewed by the organization's management and how well it is integrated into the day-to-day business activities.

Catastrophe model assessors and overseers are those with responsibility to ensure best practice for the operation and use of the model, for example, catastrophe risk managers, exposure managers, risk managers and regulators. Increasingly, the growth in regulator-driven standards and associated external assessment of model design, operation and validation, particularly around the incorporation of catastrophe models into organizational **capital models** and **enterprise risk management** (ERM) systems, is developing a wider community of specialists throughout organizations requiring catastrophe model literacy. An example of this is **Solvency II** (a pan-European regulatory regime for insurers, see Chapter 2.11.4), or the US ASOP 38 (Actuarial Standards Board, 2000), both of which require model oversight and robust governance structures to understand model construction, assumptions and methods.

Models will certainly continue to be created by communities of specialists across a range of disciplines, and whether based on proprietary platforms, or open architecture systems, the developer role will continue to determine model character and applicability. Equally, all model agents, whether contributors, developers, operators or assessors, require insight into the decision processes and trade-offs necessary in model construction to enable effective and appropriate use. Put simply, all model agents must 'get inside the mind' of the model developer, to achieve successful and appropriate use of the model.

1.7 What Are Catastrophe Models Used For?

Chapter 2 provides details on the applications of catastrophe modelling. The most common use of catastrophe models is to quantify risk, allowing its transfer between parties in the insurance and reinsurance industry. Catastrophe modelling is now used in many aspects of the daily operations of those insurers, reinsurers, funds, broking houses, consultants and regulatory bodies concerned with catastrophe risk.

As much of the (re)insurance industry has a specialist vocabulary, we will not attempt to define the model uses in detail in this chapter, but rather explore them together with a discussion of the business purposes in Chapter 2. However, common questions that catastrophe model output is used to address include:

- How much should we charge for a (re)insurance policy?
- Which new business should we add?
- How profitable is it likely to be?
- How much could we typically lose and with what likelihood?
- What might a specific event cost us?
- What are the potential causes of loss, considering our business model?
- How can we best mitigate these?
- Do we have enough money set aside for specific eventualities?
- Are we operating within the constraints set by our board, our regulators and supervisory agents?

There are many different stakeholders within the (re)insurance company at multiple levels, from those choosing which policies to add to the portfolio up to the board level. Catastrophe models in the societal role of reinsurance are also briefly explored when looking at government pools (Chapter 2.13). As models are increasingly being used in the disaster-risk financing communities, some discussion of this topic is included in Chapter 2.14. Models have not been developed for all perils, but must adequately represent those that are of most concern for their users. A description of, and considerations for, modelling these perils are given in Chapter 3.

Understanding that a catastrophe model is a series of interconnected sub-models from various sources is critical to understanding how those models have been produced, should be used, and should be interpreted with respect to the specific questions they may be used to tackle. This topic is covered in more detail in Chapter 4.

Models are constructed so that they should be tested, rebuilt against the lessons learnt, and retested. On that basis, there will never be a final version of a model, simply various improved versions. This raises a doubt in the mind of the user – how wrong can a model be before it is useless, or worse still, misleading? Also, how can a continually evolving and uncertain model be used effectively for assessment and decision-making? These questions would be difficult enough if the users of models were always the same people who built them. In the financial industry, including insurance, this is generally not the case. These questions can be partially answered by developing a company-specific view of the risk; this process is described in Chapter 5.

1.8 Anatomy of a Catastrophe Model

Conceptually, most catastrophe models tend to follow a similar modular structure, reflecting the integration of the multidisciplinary geophysical, engineering and financial components, which each contribute to the overall estimate of risk. Figure 1.2 shows one representation of such a structure.

The components **hazard**, **vulnerability**, **exposure** and loss (or **financial calculation**) require validation both separately and collectively as well as integration within a computational workflow and data management framework (the platform). The following sections describe the features of each component as well as what constitutes a platform.

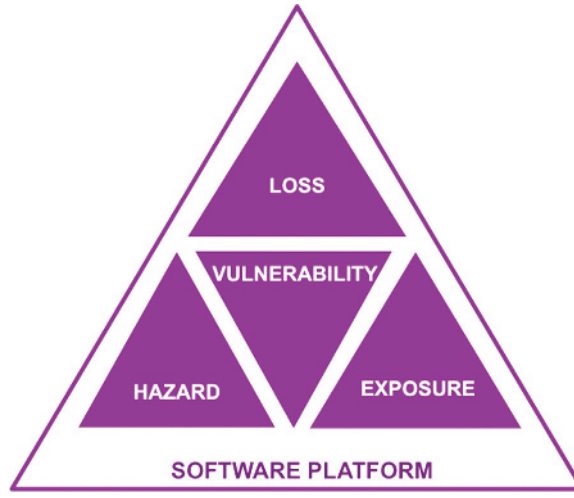
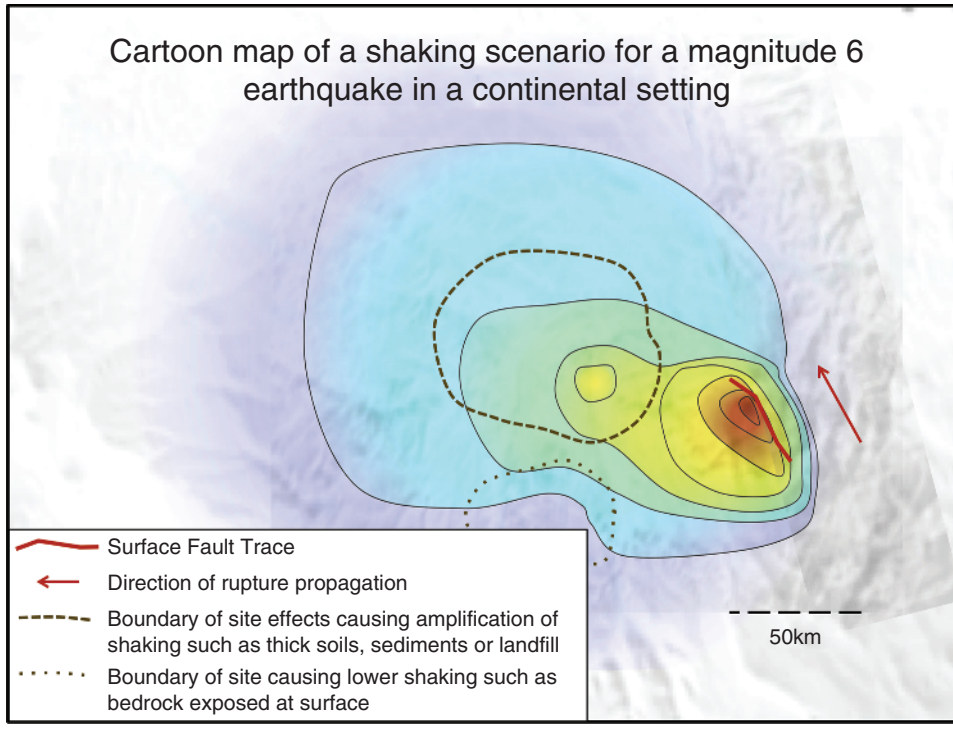


Figure 1.2 Structure of a catastrophe model showing the five key components: hazard, exposure, vulnerability, loss and a platform. Courtesy of Impact Forecasting (http://www.aon.com/impactforecasting/impact-forecasting.jsp)

1.8.1 Hazard

The hazard component reflects the extent and intensity of a peril as defined by a specific hazard metric. This often represents the hazard intensity variation across a pre-defined geospatial framework, either in a regular **raster** (grid cell-based), or in an irregular vector structure (e.g. postcode zones or points). Either way, each **event footprint** reflects the relative intensity of the hazard over the defined time period of the event. Examples include **peak ground acceleration** (PGA) for an earthquake, flood depths across a floodplain from a river breach, or **peak gust wind speeds** across a storm track. An example of such a footprint is given in Figure 1.3.



Perceived Shaking	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
Potential Damage	None	None	None	Very Light	Light	Moderate	Moderate /Heavy	Heavy	Very Heavy
Peak Acceleration (%g)	<0.17	0.17-1.4	1.4-3.9	3.9-9.2	9.2-18	18-34	34-65	65-124	>124
Peak Velocity (cm/s)	<0.1	0.1-1.1	1.1-3.4	3.4-8.1	8.1-16	16-31	31-60	60-116	>116
Instrumental Intensity	I	III-III	IV	V	VI	VII	VIII	IX	X+

Figure 1.3 The spatial footprint of shaking due to an earthquake, coloured according to various common and equivalent measures of intensity. Red line represents surface rupture. Note that in general the most intense shaking occurs where there is the greatest deformation, and decreases away from the fault. However, surface conditions affect the shaking, creating a non-uniform decrease in intensity away from the fault, such as due to thick soils west of the fault in this illustration. The scale is used in USGS ShakeMap (Figure 2.5). *Source*: USGS, 2006, adapted from Wald *et al.* (1999).

The footprints are produced via a model development process (see Figure 4.1 in Chapter 4.2) that reflects the geophysical processes operating to create the hazard as accurately as possible. The footprints are based on the particular data, assumptions and computational approaches used and taken by the model development team.

The choice of hazard metric (see Chapter 4.3.2) will also be a critical element of the hazard model, and will often be based on a generally accepted approach, although this may not be fully representative of all damage caused, for example, flood depth may not be the only causal factor for flood damage, when duration of inundation, velocity of flow or water pollution may all have an effect on damage. Instead the chosen metrics may simply be the most effectively modelled within the geophysical modelling framework. In general, most hazard-intensity factors used will be reasonable proxies, after calibration of the model.

In addition, for probabilistic models, **stochastic** event sets will also be developed, which are a collection of individual event footprints that could happen in some synthetic history (see Chapter 4.3.1). The number and range of stochastic scenarios in the overall **catalogue** of events will be designed to accurately represent the scope of the hazard in the particular model region, while recognizing computational constraints. A **frequency** or **rate of event occurrence** (i.e. how many events will happen in a given period of time) will be assigned to each event, based on a particular methodology, and will vary according to the hazard and region being modelled. See Chapter 4.3 for a more detailed narrative of hazard model construction considerations, including stochastic event construction.

1.8.2 Vulnerability

The vulnerability component is the interface between hazard, exposure and loss, and provides a means to estimate the relative damage to the asset, given a certain level of hazard. Most vulnerability models are arranged as series of damage functions, which enable look-up between hazard intensity and estimated damage as a ratio of total value. An example is shown in Figure 1.4. Damage ratios are estimated repair cost (i.e. modelled loss) as a fraction of the replacement cost of the building (i.e. insurance exposure or **sum insured**).

Construction of vulnerability functions will be dependent on available data, for example, claims data in the case of (re)insurance. For the more extreme and rare events there is usually insufficient data to construct vulnerability functions using data alone. In these cases much reliance is placed on engineering studies. There is clearly much uncertainty, especially at the higher hazard levels, and so often vulnerability functions are constructed including an estimate of this kind of uncertainty. See Chapter 4.5 for more details of the model vulnerability development process.

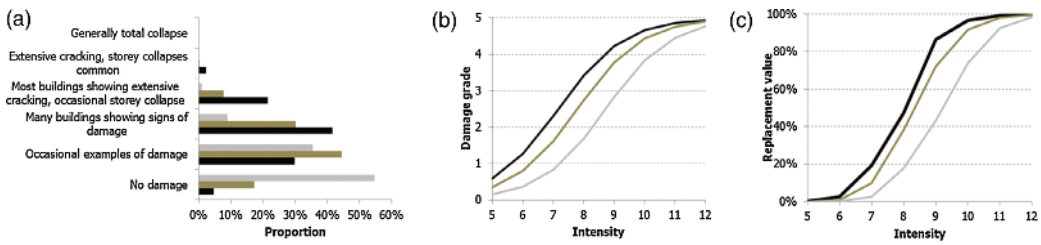


Figure 1.4 Illustrative vulnerability functions, derived according to Risk-UE methodology (Mouroux, 2006): (a) shows the distribution of damage states for a particular intensity level in a particular area for three different buildings; (b) displays the average damage status per intensity value for three different buildings in a particular area; and (c) displays the resulting replacement cost as a percentage of insured value, which can be computed once the damage grade is known.

1.8.3 Exposure

Exposure is used in two distinct ways in catastrophe models. First, exposure data for the specific objects being modelled are entered by the user into the model. This is described in Section 1.9.1. Second, a representation of industry exposure for the region covered by the model is also used in the process of building a catastrophe model. This takes the form of a database of exposure values split by area and by type of object being modelled (the primary modifiers; see Section 1.9.1.3). The values will typically be either insured or economic (i.e. insured plus non-insured) values, and may be split between buildings values, contents values and other aspects, such as business interruption values (see Section 1.9.1.2). The database will also usually contain information on insured financial structures such as deductibles and limits (see Section 1.9.2). Such a database is called an **industry exposure database** (**IED**).

The main uses of an IED are as follows:

- To enable calibration of the model, the industry exposure database can be run through the model to calculate the modelled industry losses; these can be compared to historical observations for model calibration purposes.
- To enable assumptions to be made when input exposure characteristics are missing. As discussed in Section 1.9.1.3, it is fairly common for primary modifiers to be absent in the input data. If this the case, then an aggregated or composite vulnerability curve can be formed by weighting together the individual vulnerability curves; using the IED to provide the weights.
- To enable disaggregation of coarse resolution, aggregated input data. In some cases, where limited address information is provided on the input data, the best geocoding resolution (see Section 1.9.1.1) may be too coarse for the modelled peril. The IED can be used to disaggregate the coarsely geocoded data in proportion to the geographical distribution of values contained in the IED.

As most catastrophe models have traditionally been focused on static property asset risk assessment, this tends to influence the form and scope of parameters used to represent asset inventories, although some variation is applied for specific sub-classes of assets, such as agriculture, population and vehicles. More details on the use of exposure data as part of the model development process are given in Chapter 4.4.

1.8.4 Loss and Financial Perspectives

The output from the vulnerability module of a catastrophe model is the total loss, i.e. before the application of any insurance or reinsurance financial structures. This is usually termed the **ground up loss (GUL)**. This is calculated for each location and **coverage** (buildings, contents, business interruption, see Section 1.9.1.2) for each event in the model. Whether this granularity (i.e. level of detail) of loss information is actually saved into the results database depends upon the settings when the model is run, but it will be calculated for all detailed models. Most current models do not just calculate a single loss amount for each event, location and coverage, but rather a distribution of likely loss, reflecting some of the uncertainty in the loss estimate. This is often represented by a mean and standard deviation of loss (the maximum loss is, by definition, the coverage sum insured), with some newer models also providing a full distribution of loss uncertainty (for more details, see the discussion on uncertainty in Chapter 2.16.1).

For the output to be useful to the insurer or reinsurer, the model needs to perform several further functions, specifically:

- It must use the data describing insurance financial structures (see Section 1.9.2) to *share* (or *partition*) the ground up loss between the various parties involved in the risk transfer: the insured and the insurer at a minimum, and potentially other insurers and reinsurers.

- It needs to *aggregate*, or *combine*, the loss statistics at different output resolutions. For example, location level, policy level, event level, and portfolio level (see Section 1.9.3).
- If different financial structures apply at different resolutions, the model could *back allocate* the impact of the structures applied further up in the hierarchy (e.g. portfolio or policy) down to the more detailed level (e.g. location) to allow for additivity in subsequent analysis of the detailed level losses.

These are all difficult tasks, because the range and complexity of financial structures are huge. For example, financial structures can do the following:

- Operate at different levels, from coverage level to location level to policy level to the level of the whole portfolio (for reinsurance structures).
- Be nominal (or flat) amounts or percentages of loss with minimums and maximums.
- Apply to individual events or across multiple events.
- Cover a single peril or a range of perils.
- Be interdependent: one level of financial structure can impact the next.

At the time of writing, there are no catastrophe models in production which can cope with the full range of financial structures, however, some do better than others, and this can be a distinguishing feature in model selection. The different kinds of financial structures (deductibles, limits, coinsurance) are described in Section 1.9.2. Reinsurance is discussed in Chapter 2.4.

The partitioning of modelled loss, according to the financial structures, gives rise to the need for a terminology describing how the loss is shared among different parties, or **loss perspectives** (or financial perspectives). Commonly used loss perspectives are:

- Ground up loss: The entire loss with no financial structures applied.
- **Retained** or **client loss**: The loss to the insured. Sometimes this is defined as the loss due to deductibles, more often it also includes the loss exceeding any limits.
- **Gross loss**: The loss to the insurer after limits and deductibles and co-insurance are applied, but before any forms of reinsurance.
- **Net Pre-Cat**: The gross loss with facultative and per risk reinsurance applied (see Chapter 2.4.2), but not catastrophe treaties.
- **Net Post-Cat**: The net pre-cat loss with catastrophe treaties applied.

These terms are used in most vendor models, and will often mean the same thing; however, it is important to note that the precise definition may differ from model to model, so it is always worth checking the model documentation carefully.

In practice, the most commonly used perspectives are *ground up*, *gross* and *net pre-cat*. Ground up losses are often used in model validation, and for comparison to the gross losses to quantify the modelled effect of the financial structures, particularly deductibles and limits. The net pre-cat losses are often the key output taken from the catastrophe model and used within subsequent analyses. Reinsurance that applies to a whole group of policies (treaty reinsurance; see Chapter 2.4.2) is often applied outside the catastrophe model in Asset Liability Modelling (ALM; see Chapter 2.10.6) software such as Remetrica[1] or Igloo.[2] This is because these packages offer more flexibility in applying reinsurance structures and because they are often already used for the capital modelling of a (re)insurance company, which is an important use of catastrophe model output covered in Chapter 2.10.

Whether models accurately represent the impact of financial structures depends on how well models represent uncertainty and propagate it through the respective financial structures and aggregation levels. Different vendors have different mathematical approaches to this. Some use closed form integration of a parametric distribution (such as the **Beta** distribution, see Section 1.11.2.1) fitted to the mean and standard deviation of the ground up loss at the relevant

aggregation level to partition the loss. Some use numerical convolution to aggregate and apply financial structures. More recent models often use **Monte Carlo simulation**. All models will contain an assumption about the level of uncertainty **correlation** between coverages for a given location, and between locations for a given peril. Often an assumption is made of 100% uncertainty correlation between coverages. The treatment of location uncertainty correlation varies from platform to platform: sometimes the default value is zero, sometimes it is a non-zero value that varies by peril. In some platforms the level of location correlation is a user-supplied input, in others it is 'hard-wired' into the model.

1.8.5 Platform

The four model components of hazard, exposure, vulnerability and loss are implemented and integrated together in a piece of software called a platform. In addition to integrating these components, a platform usually also provides:

- a mechanism for the user to input and validate exposure items;
- a way of converting address data to latitude and longitude (geocoding), although not all platforms provide this;
- an interface to allow the user to select run-time model options and output settings, and to initiate the model runs and monitor progress of the runs;
- a structured way of storing the input exposure and output results data;
- a method of running reports and analysing and visualizing the exposure data and output results, although not all platforms provide this.

The way a platform is constructed, the methodologies and features within a platform, and the performance of the platform are linked to the database structures that the platform is based on and the computational power available.

1.8.5.1 Computational Power and Catastrophe Models

The computational framework for a catastrophe model requires considerable data management, storage and calculation-processing capability. Catastrophe models have benefitted from a continuous evolution in hardware and systems architecture and application developments. This has enabled the granularity and sophistication of the calculations within a catastrophe model platform to steadily improve. For example, several current platforms now use Monte Carlo simulation to propagate uncertainty through the financial calculations, whereas a decade ago this would have been impossible in a business setting due to computational limitations. These limitations have been a major design consideration in all catastrophe models, affecting all elements of model construction, analytical resolution, model coverage and their use. In particular, computational limitations to 'run-time' calculations have driven many of the key design trade-offs necessary to enable effective use of the models in business environments, including statistical parameterizations, re-sampling and other optimizations, and the uncertainty inherent in these practical decisions must always be considered (see Chapter 2.16.1).

Since the development of the first commercially available catastrophe models in the late 1980s, the hardware and operating systems available to host the models have evolved continuously, determining to a large part the model versioning histories of each commercial system. Increases in processor power have driven the underlying development of increasingly sophisticated computational engines, and enabled improved data management. Early platforms were designed to operate on single processor PCs, then progressed to multiple processor/PC architectures, applying process queuing methods to enable multiple model calculations to be managed in a run-time environment.

The advent of distributed computational architectures, employing groups of machines, as well as multiple-core processors, for business applications in the early 2000s provided a means for

greater computational efficiency, with catastrophe model redesigns taking this approach in the mid-2000s. In the late 2000s and early 2010s, Microsoft High Performance Computing (HPC) clustered computational products enabled commercial vendor companies to develop bespoke server architectures to optimize their data and analytical management processes even further, with the enhanced power and computational efficiencies enabling the development of high resolution, regional flood catastrophe models, and improved model run-times for large and detailed calculations such as model sensitivity analyses (see Chapter 5.4.2).

The latest platform designs harness cloud-based systems, reducing the need for business-hosted hardware and allowing the development of models with greater user application and control. This includes the much-vaunted model transparency arising from reducing model parameterization using greater processing power. These advances have also enabled a new 'plug and play' community of model developers to grow, through the creation of 'open architecture' platforms, such as the Oasis Loss Model Framework.[3] Open architecture has the potential to enable model users to select individual model components, created around common architectural standards, to better represent their own risk, although this does result in an increased responsibility on the model user to understand and justify their own choice of model components and approaches (see Chapter 5).

1.8.5.2 Platform Database Management and Data Structures

Catastrophe model analytical functionality has been remarkably stable for most of the history of catastrophe models, with the core stochastic 'exceedance probability' calculation being applied in most models, in one form or other. This approach has therefore tended to drive most computational requirements. Around this core model requirement, the other primary computational requirement has been in the database management approach used to import, store and manage the increasing sizes of exposure, hazard and loss data produced in the computation.

Early platforms employed a range of relational database management systems (RDBMS), both bespoke, but also those produced by specialist database vendors, including Oracle, Microsoft and Bentley. Most commonly, Microsoft SQL Server has been used as the database management system underlying the commercial vendor platforms. There are limitations to the computational efficiencies possible using traditional RDBMS approaches, and this has led to the development of other data warehouse approaches, including 'data lakes' such as Microsoft Azure, and operating methods such as Hadoop, to further enhance data management for catastrophe modelling, particularly in parallel with cloud-based, open data architectures.

Additional functionality, most particularly in respect to geospatial analytics, geocoding and geo-referencing, informatics and visualization, has also evolved over time with increased demand for detailed model outputs, sensitivity analyses and other risk metrics (see, e.g. Slingsby *et al*., 2010). The most recent platforms enable seamless end-to-end data management, input, modelling and visualization, and widen the end-user community from a core of specialist catastrophe analysts to include underwriters, risk managers and others across organizations.

1.8.5.3 Exposure and Result Database Structures

All models require a standard approach to data structuring, for exposure data import (see Section 1.9) and for results reporting. These all include field structures, and data dictionaries.

As noted previously, the commercial model providers have developed their own proprietary data formats around the database management system used and their particular data structures. These provide, in some cases, a pseudo-standard for specific models, but the costs and errors associated with translation between proprietary standards, originally considered to be worthy of Intellectual Property Right (IPR) protection, have been seen as a major challenge to improving the quality and confidence of catastrophe modelling. In particular, the lack of a common and

transferable data format standard has been considered a major limitation in the desired evolution towards greater model compatibility. Although initiatives such as ACORD,[4] have driven initial attempts to create model agnostic data standards, these have only been partially successful in their aim to standardize data structures, not least due to the wide range of insurance policies, covers and data sources. Recent work by the Lloyd's Market Association to derive consistent structure frameworks for exposure data (the Exposure Data Design Project, or EDP), has considered the need for a combined process and platform approach to successfully develop a model agnostic data structure across all classes. This considered the harnessing of data formats as well as technological solutions to data integration, management and manipulation, such as data lakes, while looking to other data standards organizations, including ACORD and the Open Geospatial Consortium (OGC), for similar standards that could potentially be adopted or modified for exposure requirements.

1.9 Model Input

Catastrophe models are **exposure**-based **models** and as such provide an estimate of the risk without using the historic claims data or claims experience from the specific locations or policies being modelled. The catastrophe risk analyst should enter the following information into the models:

- Exposure details: Location information (Section 1.9.1.1), exposure values such as sum insured (Section 1.9.1.2), exposure characteristics (primary and secondary modifiers – Section 1.9.1.3), and user-defined information for classification and reporting purposes.
- Financial structure information, such as deductibles, limits and reinsurance (Section 1.9.2).
- Information about how the locations are grouped or categorized (by legal contract) into policies (Section 1.9.3).

The preceding list seems fairly tractable. However, collecting, transforming, cleaning and assessing the quality of such data always constitutes a difficult, time-consuming and costly process for a (re)insurer (see Chapter 2.5.4).

In addition to entering the preceding information, the catastrophe risk analyst will also need to decide which 'switches' to turn on in the model and which output options to select. Some of the most common switches and options are:

- whether **demand surge** is used (see Chapter 5.4.2.10);
- whether **secondary perils** are used (e.g. **storm surge**; see Chapters 2.6.3, 3.2.5 and 4.3.7);
- whether **secondary uncertainty** is used (see Chapters 2.16.1 and 4.5.1.1);
- which event set is used (e.g. long-term or warm sea surface temperature for North Atlantic hurricane, see Chapter 3.2);
- how many samples to use (for those models using Monte Carlo simulation, see Chapter 4.7.3).

The specified output options can have a significant impact on the model run-time and data storage requirements for the output data. Most models allow the user to specify options regarding which financial perspectives (Section 1.8.4) are used, what type of statistics are output (see Section 1.10), and the granularity of output statistics. For example, whether result statistics are summarized across locations, or output for each individual location. The main concern from a data space perspective is the granularity at which event loss tables or year loss tables (see Section 1.10 for a definition of these) are output, as each such table can contain tens of thousands of records, and if these are output at each location for a large number of locations (1 million would not be uncommon), the space requirements are substantial.

1.9.1 Exposure

The main exposure characteristics input to models are summarized below and then defined in detail in the subsequent sections.

- Location: Address information or coordinates.
- Sum insured: The value of the exposure.
- Primary modifiers: 'Modifiers' is the catastrophe modelling terminology for exposure characteristics that can differentiate the potential damage to the exposure from the catastrophic event (otherwise known as rating factors). Primary modifiers are those which are useful predictors of damage for most perils.
- Secondary modifiers: These are modifiers whose importance is usually peril-specific.
- User-defined information for reporting purposes: A (re)insurance company will have many reporting needs and so information must be attached to the exposures and entered into the model to ensure that these needs can be met. This includes information such as legal entity, business unit, underwriter, class of business, etc.

1.9.1.1 Location

Because natural catastrophe models are designed to evaluate the geographically correlated risk from a set of locations, it is important that information about the location of each input exposure is provided. The location of assets is the primary link between hazard, exposure and vulnerability in the model.

Most models either accept coordinate information directly (and will determine a computationally applied location via a standard coordinate and geodetic system; commonly World Geodetic System (WGS)-84[5]) or will use geocoding toolsets (geocoders) to create coordinates from supplied address information. The level of spatial (or geocoding) resolution is determined by the quality of the address information supplied and the geographical granularity of the geocoder. Coordinates will often be stored as decimal degree, bounded by +/− 90 degrees latitude and +/− 180 degrees longitude depending on the hemisphere of the location, for example:

38.889931, −77.009003;

This is the decimal degree latitude and longitude to six decimal places of the Senate Building in Washington, DC. A precision of six decimal places represents a location to an accuracy of approximately 6–10 cm in the real world, depending on the latitude of the position. Usually, a precision of three to five decimal places is an appropriate accuracy to locate an individual property within a catastrophe model given the resolution of hazard data used (typically tens of metres).

In some cases, locations are supplied with latitude-longitude coordinates, often via **GPS** or other surveyed information. Typically, however, locations are determined from address information. The address information required will be based on a political, postal or other administrative geographic framework. These may be hierarchically structured, for example, varying levels of postal code, or adopted from other administrative systems, such as CRESTA[6] zones. The level of address granularity achieved will determine the overall modelling resolution, and will also, in some models, influence the calculation of model uncertainty. The precision of such information can be high within data systems, often shown as a six-digit decimal degree, but care should be taken when interpreting the address and location data as precision does not necessarily imply an equal level of accuracy. In particular, coordinates interpreted from address data by catastrophe models, or via free-standing geocoding tools, should be carefully validated, as the assumptions made in the geocoding algorithms may introduce significant error, for

example, by selecting the wrong coordinate position for ambiguous locations with multiple potential locations.

Geocoding resolution indicates the best granularity that the geocoder believes it has been able to achieve with the address information supplied. Commonly used resolutions, in descending order of spatial accuracy are:

- Building/parcel
- Street/address
- ZIP/postcode
- City/town name
- District/parish
- CRESTA zone
- State/municipality
- Country
- Unknown

The precise resolutions returned often depend upon the country and geocoder in question; however, 'street-level' geocoding is a commonly used term that can be ambiguous. For example, street-level geocoding can mean:

- the geocode of the exact building on the street (i.e. equivalent to building level);
- an estimate of the location of the building by knowing the street and interpolating based on the street number (this is the usual definition of street level geocoding);
- the midpoint of the street, with no adjustment for building number.

It is important to clarify the precise meaning of geocoding levels, such as 'street' geocoding within the geocoder or catastrophe model that is being used.

Similarly, 'postcode'-level geocoding can have different meanings in different countries depending on the size of the postcodes. For example, in the UK, a full postcode will typically narrow the geocode down to one of 15–20 residential houses (often equivalent to 'street'-level geocoding in other countries, depending on the definition) or for a large commercial building will narrow the geocode down to the exact building. In most other countries 'postcode' resolution will not be this accurate. Further considerations related to geocoding are provided in Chapter 5.4.2.6.

1.9.1.2 Exposure Value

Capturing the correct 'value' of the exposure being assessed is a critical component of risk analysis, and notoriously difficult to achieve in a consistent and accurate way. In general terms, most exposure models will apply a financial approach to representing value. This will generally be the *100%*, or *ground up* total value, most often interpreted as the 'total rebuild' or 'reinstatement' value in monetary terms determined for calculation of loss against a vulnerability function. For other, non-property assets, value can be defined in other ways, for example, the total yield for a given crop, or population; but in most cases translation to a monetary value will be undertaken regardless of the type of asset.

In most models exposure value is required in three coverage categories: buildings, contents and **business interruption** (BI). In some regions, for example, the United States, 'other structures' such as sheds and outbuildings are entered as a fourth category of sum insured. In some models **Additional Living Expenses** (ALE) values are entered instead of BI values for personal lines exposures. Each category of exposure value will be specific to a given type of vulnerability.

The *buildings* value usually represents the total structural rebuild value, the *contents* value represents material items or other assets within but not part of the main structure, and *business*

interruption (BI) represents the estimated loss of profits resulting from closure or non-operation of the asset due to direct damage.

Some exposure models will apply predefined splits between buildings and contents, and business interruption, determined from the underlying exposure parameters pre-set by the model development process.

Care is needed to ensure the correct amounts are entered into the sum insured fields, as although this seems straightforward, there are pragmatic issues that can complicate this, as discussed below.

Building Sum Insured

In all models this should be the estimated rebuilding cost. Complications occur in personal lines insurance where the insured party (i.e. householder) can mistake the market value for the rebuilding cost, which often leads to an overstatement of sum insured. Another complication is where sum insured amounts are calculated by the insurer (**notional** sum insured) to avoid the issue of customers supplying incorrect values, and the insured is provided with a policy limit that is much higher than the estimated rebuilding cost so that the insured is comfortable that the policy meets their needs. If this limit is entered into the model as the building sum insured, an overstatement of likely modelled loss will occur; the notional sum insured is likely to be a more accurate representation of the risk.

For commercial insurance, the most likely concern is that of **under-insurance**, or **insurance-to-value** (**ITV**) issues (see Chapter 2.6.3 for more details). A further consideration is whether or not a **day-one** sum insured is used (i.e. a value with no explicit allowance for inflation during the policy and rebuilding period, but with a day-one uplift provision), or whether a **full reinstatement** sum insured is used (i.e. with an allowance for inflation already within the sum insured). This is explained and discussed further in Chapter 5.4.2.7.

Contents Sum Insured

Contents sum insured for both commercial and personal lines may be subject to underinsurance concerns. For personal lines it is common for the insurer to estimate the *notional* value required by the customer (in a similar way to the estimation of rebuilding cost) and provide higher limits than those actually required. As with buildings limits, these will provide an upper bound on the policy, not a realistic estimation of values and so using the limit figures directly as contents sum insured will overstate results. Using the notional value for modelling should prove more representative of the true risk.

Business Interruption Sum Insured

The main issues here arise around the interaction between the sum insured and the **period of indemnity** (i.e. the period of time for which loss of profit or revenue can be calculated for a business interruption loss). Some models require an annual sum insured to be input for business interruption. In some models a period of indemnity field is provided for information only and does not change the results in any way. In practice, the BI sum insured on a policy will be the maximum amount recoverable in a specified indemnity period. This indemnity period can be less than or greater than a year, depending on the specific terms of the contract. The question then arises as to how this non-annual sum insured should best be represented in a model requiring an annual sum insured. This again needs careful discussion with the model vendor to ensure the input assumptions reflect the assumptions inherent in the model. A method sometimes used is to pro-rate or scale the sum insured in order that it represents an annual amount. For example, if the policy period of indemnity is two years, the BI sum insured would be halved. In the case of increasing the modelled BI sum insured due to an actual period of indemnity that is less than a year (e.g. doubling the BI sum insured because the BI period of

indemnity is six months), then a BI limit should also be added to the model input to ensure that the actual modelled loss is never greater than the actual BI sum insured.

For personal lines policies, some models require the ALE sum insured to be entered in the BI field. These are policy specific and will usually either be a specified fixed maximum amount or some percentage of the building or contents sum insured.

1.9.1.3 Exposure Characteristics: Primary and Secondary Modifiers

The four main primary modifiers are usually:

- **Occupancy class:** The purpose for which the insured building is being used. At a high level this is usually residential, commercial, industrial or agricultural. At a more granular level the *residential* classification splits into single and multiple-occupancy, and *commercial* splits into industry trade codes (e.g. retail, manufacturing, and so on).
- **Construction type:** The material and method of construction. For example, whether the building is made from wood, masonry, reinforced concrete or steel.
- Year built: The year that the property was built.
- Building height: The height of the building; usually specified as number of floors.

Other modifiers that are sometimes classed as primary are *number of buildings* (which is commonly used where data is aggregated and sometimes used within a campus-type scenario), and *square footage* (commonly used in the United States).

The impact of primary modifiers varies by peril, but these modifiers are useful information to gather for most, if not all, perils. In most models, different values (e.g. year built is 1950 rather than 2000) of primary modifiers will cause the model to select different vulnerability curves (see Section 1.8.2). This means that the impact of a primary modifier can vary with the intensity of the hazard.

Secondary modifiers are peril-specific and very much depend upon the model being used. For example, for the wind peril, these will allow extra information to be specified about roof type or roof anchoring. For flood, these may contain modifiers around secondary defences or resilience of a property's contents. For earthquake, these may contain modifiers related to retro-fitting. Unlike primary modifiers, these do not normally have separate vulnerability curves assigned, but rather usually act as percentage scalings (i.e. multipliers) applied to the loss cost calculated by using the primary modifiers; thus, the size of the percentage scaling does not vary with the intensity of the hazard.

The extent to which exposure data is available for these modifiers depends mainly upon the peril-region in question and whether a risk engineering report is available for a specific location. In catastrophe-prone regions in the United States, it is common to obtain all the primary modifiers and several secondary modifiers. In the rest of the world it is common to obtain one or two primary modifiers (usually occupancy) and no secondary modifiers. If a risk engineering report is available (normally only for high value facilities), this will usually contain all primary and many secondary modifiers. The challenge in these cases is systemizing the process to extract the right information from the risk engineering report in an efficient and appropriate way.

Some notes on the four main primary modifiers are as follows:

- The usage, or *occupancy*, of the asset is a key modifier influencing the choice of vulnerability function within the model. This will commonly be coded using either model specific codes such as UNICEDE[7] or standard industry codes such as SIC,[8] NAICS[9] or ATC.
- Standardized *construction* classifications, such as ATC-13, originally developed for regional-level aggregate Californian earthquake risk assessment purposes, are often applied across many other territories, and at the site location level. Equally, individual catastrophe modelling

organizations will often create proprietary classifications based on their own vulnerability models, such as UNICEDE[7] or RMS construction codes.

- *Year built* or *building age* is assumed to reflect the type of building standards and regulations likely to have been applied as well as wear and tear on the building structure. Pragmatically, similar ages are often grouped together within a model (e.g. properties built in 1950–1959 may be assigned the same vulnerability curve).
- The *building height* is used by the model to determine the response of the structure to a given hazard impact, such as shaking resonance, wind speed, or flood depth. Like year built, height is often banded within the model and entered as number of floors. Care must be taken not to confuse the specific floors occupied with the total number of floors; the latter should always be used to represent the building height.
- For most territories and perils it is unusual to have insurance exposure data populated with all primary modifiers (US catastrophe-prone areas are an exception to this). Very often occupancy is present, and other primary modifiers are not. In the case of missing primary modifiers the model should contain assumptions about how to treat the exposure data, i.e. which vulnerability curve to assign. This is one reason why models will contain details of industry exposure data, as described in Section 1.8.3.

1.9.2 Financial Structure

Insurance financial structures (also commonly referred to as **policy terms**) are features designed to modify the loss payments. Financial structures can be illustrated by considering a simple buildings insurance policy (see Figure 1.5). The box represents the rebuild cost of the property, also known as its replacement cost. It is unusual for insurers to pay very small claims since they wish insurance to be priced reasonably and including many small claims with associated handling costs would make the cost of the policy uneconomical. Therefore, the most common insurance structure is the **deductible**, the amount the policyholder has to bear before they can reclaim from the policy. It can also be known as an **excess** since payments are in excess of this value or a **retention** since the losses are retained by the policyholder. Deductibles can be set as a proportion of the original policy value, a fixed monetary amount or a proportion of the loss, sometimes with a minimum and maximum value. Deductibles may vary by peril. It is common that household policies will be subject to a deductible but cover the entire rebuild cost of the home in excess of this

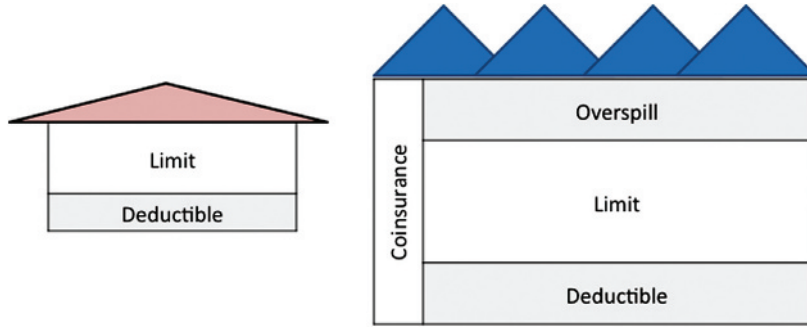


Figure 1.5 Simple representation of insurance financial structures or policy terms. See text for details. Loss is imagined as increasing upwards from the bottom of the box, i.e. small losses fall within the deductible layer. The smaller box represents a typical situation for a lower-value property. The larger box, with the blue roof, represents a common situation for a higher-value property. White boxes represent areas covered by insurance, while grey areas remain the liability of the policyholder.

value. For higher value properties, the insurer is unlikely to cover the entire rebuild cost of the property but instead only cover a specified **limit** (their limit of liability). Limits generally take the same form as deductibles. If the sum of the deductible and the limit is less than the rebuild cost, the difference will again fall to the policyholder. This difference is known as the **overspill**. There may also be insurance shared between multiple parties, known as coinsurance.

These basic structures can be applied in a variety of ways: they may differ by coverage (see Section 1.9.1.2) and also may apply at either an individual location as in the example above or across multiple locations when the structures are applied at a policy level. For instance, a chain of shops may have multiple locations with deductibles and limits at each location, but then may also have a master policy that has an overall maximum deductible so that if multiple properties are affected by the same event, the amount to be retained by the policyholder is capped. On a similar basis, there will be an overall limit to cap the payment by the insurer. This limit may only apply to particular locations, for instance, a geographical sub-set of the total, or for a specific peril, e.g. flood, and in this case it is known as a **sub-limit**.

There are variations on this theme: a **franchise deductible** is a deductible that prevents a pay-out to the policyholder until the loss reaches the level of the franchise, but then vanishes as a deductible once that level is reached. A **step policy** is a policy that pays out only a set number of specific amounts, for example, five possible levels of pay-out, each pay-out responding to a range of assessed building damage. These are common in Japan.

Policy terms vary significantly by country, and catastrophe risk analysts will need to thoroughly understand the type of policies they are modelling, by close examination of the wordings and discussion with the underwriting teams. The translation of these terms into model input depends very much on the coding schema of particular model vendors, so careful reference to the vendor documentation is also required.

Policy structures can apply to insurance and reinsurance. More detail on reinsurance policies can be found in Chapter 2.4.2.

1.9.3 Portfolio Hierarchy

In catastrophe modelling, the order of loss calculation is important, particularly when location and policy characteristics can be complex; for example, where locations may be affected by more than one set of peril-specific policy conditions, or where there are multi-territorial limits.

In general terms, the calculation hierarchy will be reflected in the structure of the data within the model database. A generic hierarchy is shown in Figure 1.6. As can be seen, location (a single asset entity, ideally with an associated location reference which can have insured value and risk characteristics (modifiers) applied to it) is the basic and most granular unit of exposure. Locations tend to aggregate to the policy (or sometimes termed **account**) level. In addition, locations will usually aggregate at varying levels of geographic hierarchy, such as ZIP/postcode, state or country. Policies may apply conditions across all locations in a single country, or across different geographic regions. In many cases, for example, policies may apply across various countries, or may exclude particular countries. In general, policies will then accumulate to the **portfolio** level. This is often defined at the insurance **business unit** level, for instance a particular property line or product may be considered as a portfolio.

Accumulation, **roll-up**, or **aggregation** are terms employed to describe the overall combination of multiple portfolios into a time-specific snapshot of exposure and modelled results, for example at a quarterly or monthly intervals. This involves the calculation of loss after each set of financial conditions are applied at each hierarchical level. This is a complex process (see Section 1.8.4) and in some models, this can create difficulties in interpretation of losses, given the challenges in calculating across multiple geographical and policy levels. Accumulation is discussed in more detail in Chapter 2.7.

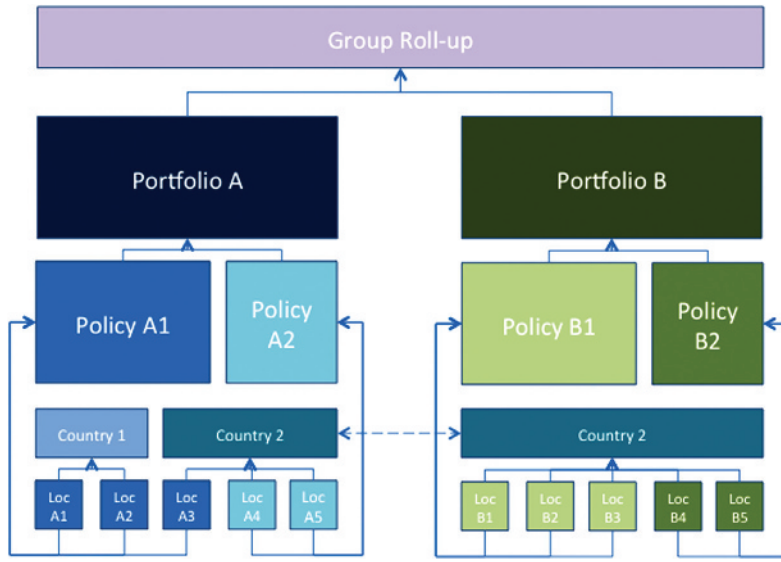


Figure 1.6 A generic insurance hierarchy. 'Loc B2' is location B2. Locations are collected together or 'aggregated' into policies, here shown in the same colour. Policies are grouped into portfolios. An insurer usually has a number of portfolios that it periodically wishes to assess in a 'roll-up'.

For some types of insurance exposure, time-varying levels of insured value, and the use of limits rather than actual sums insured can complicate this process. In these cases, it is often necessary to determine an agreed, but sub-optimal approach to the estimation of insured value at each level and apply it consistently, even though this is likely to be inaccurate.

1.10 Model Output: Metrics and Risk Measures

Catastrophe models are designed to quantify catastrophe risk, and so the fundamental output from such models is a probability distribution of loss at the appropriate aggregation level, for the relevant financial perspective. Summary statistics such as the mean and standard deviation of loss are also calculated.

As the catastrophe modelling industry has evolved, a terminology specific to this industry has also emerged and is described in this section. The distribution of the maximum loss in a year is called the **Occurrence Exceedance Probability** (OEP) distribution. The distribution of the *sum* of losses in a year is termed the **Aggregate Exceedance Probability** (AEP) distribution. These are the two main distributions obtained from catastrophe models. Catastrophe models were initially designed to answer questions such as 'What level of reinsurance should I buy?' – requiring an estimated distribution of maximum loss, and 'How much capital do I need?' – requiring an estimated distribution of the sum of losses in each year. Hence the OEP and AEP curves evolved as a standard output of most catastrophe models. These distributions and other metrics commonly output and used are described below.

1.10.1 Common Metrics

There are many metrics that are either calculated by the catastrophe modelling platform or can be calculated from the underlying model output. Some common metrics are described

below. These can be calculated for each financial perspective output from the catastrophe model.

1.10.1.1 Annual Average Loss (AAL) or Average Annual Loss

The annual expected loss. Sometimes called **Annual Mean Loss** (AML) or **pure premium**.

1.10.1.2 Standard deviation (SD) around the AAL

A measure of the volatility of loss around the AAL. It is often used in conjunction with the AAL as an input to forming technical price (see Chapter 2.6.2). However, this usually does not represent the full uncertainty in the AAL. The largest component of uncertainty missing is usually the uncertainty associated with the hazard or event generation process (the primary epistemic uncertainty, see Chapter 2.16.1 for more details).

1.10.1.3 Exceedance frequency (EF)

The annual frequency of events expected to give losses greater than a given amount. This is not usually output directly by catastrophe models, but is a useful metric and one that is readily calculable from ELTs or YLTs (see Section 1.10.5.4). Note that since this is a frequency (i.e. a count divided by a time period), not a probability, the EF can be greater than 1.0. The **return period** of a particular size of event is the reciprocal of its exceedance frequency.

1.10.1.4 Occurrence Exceedance Probability (OEP)

OEP is the probability that the *maximum event loss* in a year exceeds a given level. The occurrence return period is the reciprocal of the OEP. Another equivalent way of thinking about this is that it is the probability that *at least one* event in a year exceeds a given level. While this may not at first be obvious, it is always the case that if at least one event in a year exceeds the given level, the maximum event in that year must also have exceeded the same given level. This equality is used later on when calculating the OEP.

1.10.1.5 Aggregate Exceedance Probability (AEP)

AEP is the probability that the *sum of event losses* in a year exceeds a given level. The aggregate return period is the reciprocal of the AEP. The area under the AEP is equal to the AAL.

1.10.1.6 Value at risk (VaR)

VaR is the loss value at a specific quantile of the relevant loss distribution. For example, a 99.5% VaR based on the aggregate loss distribution would be the value at the 0.5% level (1-in-a-200-year return period) on the AEP curve.

1.10.2 Exceedance probability curve characteristics

A schematic of an AEP curve is shown in Figure 1.7. The area under the AEP curve is equivalent to the AAL.

Although the VaR from AEP and OEP curves is a widely used metric (e.g. Solvency II specifies the 99.5% VaR as the regulatory capital requirement) it should be noted that VaR is not a **coherent risk measure** as defined in Artzner *et al*. (1999); see Section 1.11.3. In particular, there is no guarantee that adding two EP curves together gives a combined EP curve that is less than or equal to the individual curves (i.e. sub-additivity is not a property of VaR). In other words, diversification is not always properly reflected when using VaR, as demonstrated in Woo (2002). Following Woo (2002), this can be illustrated by using an example of two independent portfolios (A and B), each of which contains just two events with the same event frequency (0.5%), as shown in Table 1.2 and Table 1.3.

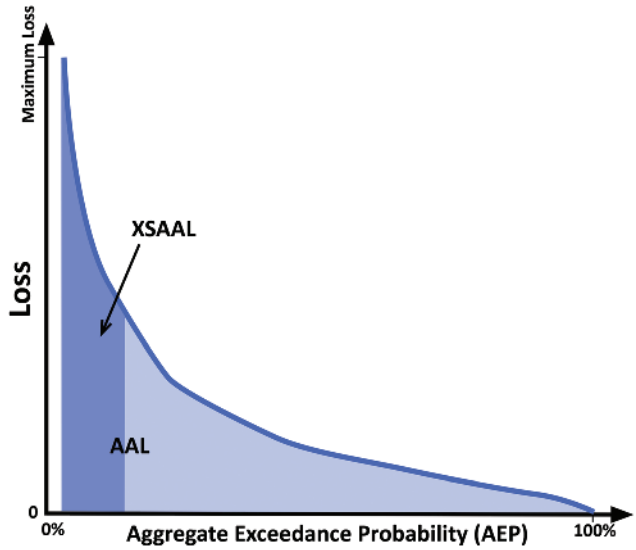


Figure 1.7 An AEP curve. The area under the whole curve (i.e. dark and light blue) is the average annual loss (AAL), and the area beyond some threshold or *excess* (dark blue) is the excess annual average loss (XSAAL). In this case, the x-axis is plotted in terms of probability, but it is also common to create AEP curves with return period in units of years (i.e. 1/probability) on the x-axis.

Table 1.2 Event loss table for a simple idealized portfolio: Portfolio A.

Loss	Rate (%)	Exceedance frequency (%)	Return Period
10,000	0.5	0.5	200 years
1,000	0.5	1.0	100 years

Table 1.3 Event loss table for a simple idealized portfolio: Portfolio B.

Loss	Rate (%)	Exceedance frequency (%)	Return period
20,000	0.5	0.5	200 years
2,000	0.5	1.0	100 years

The combined portfolio event loss table is shown in Table 1.4 and it is clear that the 99% VaR metric is 1,000 for Portfolio A, 2,000 for Portfolio B, but *10,000* for the combined portfolio. Despite the fact the portfolios are independent, combining them has not resulted in a reduced risk measure at 1 in 100 year VaR; this is counterintuitive as one would expect diversification to reduce the risk. Irrespective of this, VaR measures are commonly used in setting overall risk limits and thresholds, largely because they are easy to explain and understand. However when *allocating* such limits (e.g. capacity or capital allocation, as discussed in Chapter 2.10), it is even more important that a coherent risk measure is used. Metrics such as TVaR, and XSAAL are often used and these are described below.

Table 1.4 Event loss table for a portfolio that combines portfolios A and B (from Tables 1.1 and 1.2, respectively).

Loss	Rate (%)	Exceedance frequency (%)	Return period
20,000	0.5	0.5	200 years
10,000	0.5	1.0	100 years
2,000	0.5	1.5	67 years
1,000	0.5	2.0	50 years

1.10.3 More Advanced Metrics

1.10.3.1 Tail Value at Risk (TVaR) or Tail Conditional Expectation (TCE)

TVaR or TCE is the expected value of loss above a specific quantile of the relevant loss distribution, given that the specific quantile has been exceeded. It is, therefore, a conditional metric (unlike XSAAL, see later). It is equivalent to the area under the curve above the specified quantile divided by 1 – quantile. The value of TVaR is always greater than the VaR value at the same quantile.

1.10.3.2 *Excess VaR* (xVaR) and Excess TVaR (xTVaR)

'x' here means excess of the mean so these are the VaR and TVaR metrics, but with the mean value (AAL if the AEP distribution is being used) subtracted.

1.10.3.3 Excess Average Annual Loss (XSAAL)

XSAAL is the expected loss cost above a certain threshold (or *excess* point). It is equivalent to the area under the AEP curve above the excess point. It can be thought of as that portion of the overall AAL from events greater than the threshold. Unlike the TVaR metric, the area under the curve above the threshold is not re-normalized by dividing by (1 – threshold quantile), so the XSAAL will always be smaller than the AAL. In other words, XSAAL is *unconditional* whereas TVaR is *conditional*.

1.10.4 Event Loss Tables and Year Loss Tables

Event Loss Tables (ELTs) or **Year Loss Tables** (YLTs) (see Sections 1.10.5 and 1.10.6 for a description of these) are the fundamental outputs from catastrophe models used to calculate the metrics above. These tables are widely used as input to capital and pricing models, so it is important the practitioner is familiar with ELTs and YLTs. Models will either output an ELT or a YLT, but not usually both.

ELTs and YLTs have a significant impact on the size of the output dataset. For example, an event set could contain 40,000 events, each containing a mean and standard deviation of loss. If this is output at location level, even a relatively small portfolio of 10,000 locations can generate a large amount of output data (multiple GB). The resolution at which event set level output is turned on should be checked carefully against database size constraints and/or disk space. These tables, some methods for calculating basic metrics from them, and the advantages and disadvantages of each approach are described in Sections 1.10.5 and 1.10.6.

1.10.5 Event Loss Table (ELT)

An ELT contains the loss statistics, resulting from the input exposure, for every event in the catastrophe model (for a given peril); typically tens of thousands of events. These events may

Table 1.5 An example of a typical ELT with fictitious numbers.

Event ID (i)	Rate (r_i)	Mean loss (L_i)	Standard deviation (independent) $SDind_i$	Standard deviation (correlated) $SDcor_i$	Exposure impacted (E_i)
1	0.04	850,000	1,500,000	1,000,000	200,000,000
2	0.02	700,000	1,600,000	1,300,000	50,500,000
3	0.01	1,000,000	2,000,000	1,500,000	100,000,000
4	0.03	800,000	1,500,000	900,000	60,000,000
5	0.01	650,000	1,000,000	800,000	150,000,000

Note: See text for details of the columns.

well be a 'boiled-down' set from a larger catalogue (perhaps hundreds of thousands); the reduced number of events in the production model are chosen to preserve the important overall statistics while reducing the model run-time to acceptable levels.

The loss statistics reflect the selected output resolution and financial perspective. So, for example, an event loss table could represent the ground up loss for every event for an individual location, or the gross loss for each event summarized over a portfolio of 1 million locations. In the example shown in Table 1.5, the ELT contains just five events. In practice, the ELT will contain many thousands, or even tens of thousands, of events.

Going through each field in in turn:

- *Event ID*: This is a number or reference uniquely identifying the event. It is often used when combining multiple ELTs to ensure that the correlation between losses from the same events is preserved; in other words, the losses should be added for events with the same event ID.
- *Rate*: This is the annual frequency of occurrence of the specific event. It is not a probability, as in theory this could be greater than 1.0 (although this would be very unusual). For a given model version and Event ID, this is always the same; it does not vary according to the input exposure or financial perspective.
- *Mean loss*: This is the accumulated mean loss for the given event. It varies according to the input exposure, the accumulation level, and the financial perspective.
- *Standard deviation, independent and correlated*: The standard deviation in an ELT represents the **secondary uncertainty**, which is the uncertainty present, given that there has been an event (more details are provided in Chapters 2.16.1 and 4.5.1.1). A model will contain an assumption as to how correlated the uncertainty is, both between coverages for a given location and between locations for a given event. In order to combine standard deviations from loss distributions (e.g. when aggregating from a lower level to a higher level), it is important to know whether the distributions are correlated or not. If the distributions are 100% correlated, the standard deviations can simply be added. If the distributions are completely uncorrelated, the standard deviations are combined by taking the square root of the sum of squares. Typically the *coverage* correlation is assumed to be 100%, and so the coverage standard deviations are usually added to obtain the location standard deviation. The *location* correlation is often a non-zero quantity, and is used within a model to partition the standard deviation into a correlated part (= total location standard deviation × correlation) and an independent part (= total location standard deviation × (1-correlation)). The overall standard deviation for a given event is found by summing the independent and correlated standard deviations.

- *Exposure impacted*: Exposure impacted is the sum of the sum insured for any location potentially exposed to the event in question. It is typically used together with the mean and standard deviation of loss to parameterize a statistical distribution (e.g. Beta) in order to reflect uncertainty in subsequent calculations.

In addition to the fields mentioned above, an ELT often has a description field capturing details about each event.

The subsequent sections show how statistics can be calculated from ELTs, using the sample ELT in Table 1.5 as an example.

1.10.5.1 Limitations and Benefits of ELTs

The main form of output from some models is an ELT. Other models provide a YLT form of output instead, which is discussed in Section 1.10.6. The main drawback of an ELT approach is that it can be difficult to implement a non-parametric frequency assumption, whereas in a YLT a frequency distribution is an inherent part of the output and so parametric and non-parametric distributions (e.g. the output of a climate model) can be incorporated. Some benefits of an ELT approach are that the number of years simulated from the ELT can be changed to match user requirements, and annual rates can easily be modified if a user wishes to implement a revised view of risk.

1.10.5.2 Calculating the Mean Loss Across All Events (the AAL)

The AAL is calculated as the sum over all events of the product of the rate and mean loss. Each entry, i, in an ELT can be thought of as a **compound distribution** (a distribution resulting from combining the individual distributions of the number of losses N_i and the size of losses X_i (e.g. Kaas *et al.*, 2009) of aggregate loss in a year, S_i.

Using $E[\ ]$ to denote the expectation (or mean value) of a quantity, the mean number of events for a given entry in an ELT, $E[N_i]$, is r_i (the event rate). The mean event loss $E[X_i]$ is L_i (see Table 1.5). The mean of a compound distribution is $E[N_i] \times E[X_i]$, which is equivalent to $r_i \times L_i$ for each ELT entry. The AAL is found by summing over all entries in the ELT table:

$$AAL = \sum_i r_i \times L_i \tag{1.1}$$

1.10.5.3 Calculating the SD of Loss for One Event and Across All Events

The variance (denoted $Var[\ ]$)of a compound distribution, S_i, with frequency and severity assumed independent, is given by (e.g. Kaas *et al.*, 2009; Klugman, Panjer and Willmot, 1988):

$$Var[S_i] = E[N_i]Var[X_i] + Var(N_i)E[X_i]^2 \tag{1.2}$$

If a **Poisson** assumption is made, then both $E[N_i]$ and $Var[N_i] = r_i$ (see Section 1.11.1.1).

Referring to the columns in Table 1.5: $E[X_i]^2 = L_i^2$, and $Var[X_i] = (SDind_i + SDcor_i)^2$. This gives the following formula for the variance of an individual ELT entry i:

$$Var[S_i] = r_i(SDind_i + SDcor_i)^2 + r_i L_i^2 \tag{1.3}$$

Assuming events are independent, we can sum the variance across all ELT entries and then square root the resulting sum to obtain the standard deviation (SD) around the AAL:

$$SD = \sqrt{\sum_i r_i(SDind_i + SDcor_i)^2 + r_i L_i^2} \tag{1.4}$$

Table 1.6 Sample ELT, ordered, and with EF calculated.

Event ID (i)	Rate (r_i)	Exceedance Frequency (EF_i)	Mean Loss (L_i)	Standard Deviation (Independent) $SDind_i$	Standard Deviation (Correlated) $SDcor_i$	Exposure Impacted (E_i)
3	0.01	0.01	1,000,000	2,000,000	1,500,000	100,000,000
1	0.04	0.05	850,000	1,500,000	1,000,000	200,000,000
4	0.03	0.08	800,000	1,500,000	900,000	60,000,000
2	0.02	0.10	700,000	1,600,000	1,300,000	50,500,000
5	0.01	0.11	650,000	1,000,000	800,000	150,000,000

1.10.5.4 Calculating the Exceedance Frequency (EF) Without Secondary Uncertainty

The exceedance frequency for a given level of loss is simply the annual frequency of events greater than or equal to this level of loss. It is not a statistic normally output by catastrophe models, but it is nonetheless useful. It answers questions that OEP and AEPs cannot answer (e.g. how many losses above a certain amount can I expect in a year?), so can be useful when comparing and contrasting models with different frequency distribution or clustering algorithms in order to gain insight into the underlying model differences.

The EF is calculated by sorting the ELT in descending order of mean loss, L_i. The event rate, r_i, for the largest ELT entry is the exceedance frequency at this level of loss. The rates can then be summed, from each level up to the largest level of loss, in order to calculate the exceedance frequency for each level. This is shown in Table 1.6.

So the EF of a 700,000 loss is 0.1. Another equivalent way of expressing this is that a 700,000 event has a return period of 10 years.

1.10.5.5 Calculating the Exceedance Frequency with Secondary Uncertainty

The calculation of EF above ignored secondary uncertainty. In order to take this into account, the uncertainty distribution must be known. It is becoming more common for models to output this distribution for each event (as well as the mean event loss). However, for many models currently used, a standard deviation (perhaps separated into independent and correlated parts) as well as the maximum exposure impacted is all the uncertainty information that is provided for each event. An assumption must therefore be used about the size of loss distribution. It is fairly common for a Beta distribution to be used, to represent the size of loss normalized by the maximum exposure (or exposure impacted). The steps in calculating the EF for a particular amount of loss (A) are as follows:

1) For each ELT entry calculate the probability of the loss exceeding the amount A, using either the distribution provided by the model or a parametric assumption based on the mean and standard deviation (and potentially maximum exposure) of loss for that ELT entry. A description of the Beta distribution is given in Section 1.11.2.1.
2) For each ELT entry, multiply the probability of loss exceeding the amount A by the event rate.
3) Sum the quantity produced in (2) across all events. This is the EF with secondary uncertainty for loss amount A.

To obtain the EF for other loss exceedance levels, steps 1–3 need to be repeated for the different levels.

Table 1.7 Sample ELT with OEP calculated.

Event ID (i)	Rate (r_i)	Exceedance Frequency (EF_i)	Occurrence Exceedance Probability (OEP_i), %	Mean Loss (L_i)	Standard Deviation (independent) $SDind_i$	Standard Deviation (correlated) $SDcor_i$	Exposure Impacted (E_i)
3	0.01	0.01	0.995	1,000,000	2,000,000	1,500,000	100,000,000
1	0.04	0.05	4.877	850,000	1,500,000	1,000,000	200,000,000
4	0.03	0.08	7.688	800,000	1,500,000	900,000	60,000,000
2	0.02	0.10	9.516	700,000	1,600,000	1,300,000	50,500,000
5	0.01	0.11	10.417	650,000	1,000,000	800,000	150,000,000

1.10.5.6 Calculating the OEP

The OEP (without secondary uncertainty) can be calculated from the ELT entries, provided an event frequency distribution is known or assumed. The steps are as follows:

1) Sort the event losses in descending order and calculate the EF as shown above.
2) Use the EF at each level in the ELT, together with a frequency distribution, to calculate the probability of there being at least one event greater than the level in the ELT (which equals 1 minus the probability that there are no events). This is the OEP.

For example, if a Poisson distribution is assumed, the probability of there being at least one event greater than level i is $1 - e^{-EF_i}$ (see Section 1.11.1.1). The OEP values using the Poisson assumption are shown in Table 1.7.

The OEP figures above ignore secondary uncertainty. To calculate the OEP with secondary uncertainty, the EF with secondary uncertainty would first be calculated using the process described in Section 1.10.5.5, and the OEP would then be calculated in the manner described above, but using the EF with secondary uncertainty instead.

1.10.5.7 Calculating the AEP

Calculating the AEP is not possible in the same way as OEP, since the AEP represents the distribution of the sum of losses within a year. In order to evaluate this, the distributions from the individual ELT entries must be convolved (added together). In practice, this can either be done through Monte Carlo simulation or through the use of a technique involving discretization of the distributions, such as Panjer or FFT (e.g. Embrechts and Frei, 2009). The use of simulation is discussed in Section 1.10.5.9; the use of Panjer or FFT is beyond the scope of this book.

1.10.5.8 Correlation between ELTS

A natural question to ask is the extent of correlation between locations, policies or portfolios. If an ELT for the respective locations, policies or portfolios is available and the variance for each ELT and for the grouped (combined) ELT is known, the correlation can be calculated. Consider two ELTs, A and B, with standard deviation SD_A and SD_B respectively (calculated using Equation (1.4)). The combined ELT has a standard deviation of SD_{A+B}.

The general formula for the variance of the sum of distributions is given by:

$$Var[A + B] = Var[A] + Var[B] + 2Cov[A, B] \tag{1.5}$$

where $Cov[A, B]$ is the covariance of A and B.

The Pearson (or linear) correlation coefficient between A and B is defined by:

$$\rho_{A,B} = \frac{Cov[A,B]}{\sqrt{Var[A]Var[B]}} \tag{1.6}$$

This leads to the following formula for correlated variables, expressed in terms of standard deviations instead of variances:

$$SD_{A+B}^2 = SD_A^2 + SD_B^2 + 2\rho_{A,B}SD_ASD_B \tag{1.7}$$

Rearranging, we can see that the correlation between two ELTs, A and B, can therefore be calculated as:

$$\rho_{A,B} = \frac{SD_{A+B}^2 - SD_A^2 - SD_B^2}{2SD_ASD_B} \tag{1.8}$$

1.10.5.9 Use in Simulation: Generating YLTs

Monte Carlo simulation (e.g. Klugman, Panjer, and Willmot, 1988) can be used with ELTs to perform further calculations in order to generate statistics such as the AEP, or to calculate the impact of financial structures that operate at the level of the ELT (e.g. reinsurance on a portfolio or group of portfolios). Commercial tools may be used to do this (see Chapter 2.10.6); the process is as follows:

1) Calculate the overall ELT frequency.
2) Use this frequency, together with any other information known about the frequency distribution, to simulate the number of events each year.
3) Given the number of events in a specific year, sample the events that occur such that the chance of them occurring is in proportion to the relative frequency of the events within the ELT.
4) For each event, sample from the secondary uncertainty distribution to obtain the realization of the loss from that event, including uncertainty.
5) The event loss figures from (4) can then be used in order to calculate further statistics.

Further information on each step is as follows. In step 1, the overall ELT frequency is simply calculated as the sum of the events rates in an ELT. In step 2, if a Poisson distribution (Section 1.11.1.1) is assumed, then the only parameter that is needed is the sum of event rates. If a different distribution is used, then more information may be needed, for example a **negative binomial** will need an assumption about the standard deviation of event rate. More sophisticated algorithms can also be used, for example, a specific clustering algorithm could be implemented through a simulation framework; perhaps including a dependency between the number of events and the size of each event.

In step 3, once a number of events have been sampled from a frequency distribution, the same number of specific event entries needs to be selected from the ELT. This can be done by first normalizing the exceedance frequency for each event in the ELT by ordering the events in the ELT in descending order, calculating exceedance frequency for each event, and then dividing each event exceedance frequency by the overall sum of the event rates. For example, the normalized exceedance frequency from the sample ELT in Table 1.7 is calculated and shown in Table 1.8.

In the simulation framework, for each event sampled in a given year, a uniform distribution ranging between zero and 1 (denoted Uniform(0,1)) can be sampled, and the realization of this is used to look up the ELT entry through the normalized exceedance frequency. For example, say the frequency distribution sampling results in two events occurring in a given year, and the Uniform(0,1) distribution sampling results in 0.05 and 0.62 being sampled for each event. This

Table 1.8 Sample event loss table with normalized exceedance frequency calculated.

Event ID (i)	Rate (r_i)	Exceedance Frequency (EF_i)	Occurrence Exceedance Probability (OEP_i), %	Normalized Exceedance Frequency, %	Mean Loss (L_i)	Exposure Impacted (E_i)
3	0.01	0.01	0.995	0.01 / 0.11 = 9.1	1,000,000	100,000,000
1	0.04	0.05	4.877	0.04 / 0.11 = 45.5	850,000	200,000,000
4	0.03	0.08	7.688	0.03 / 0.11 = 72.7	800,000	60,000,000
2	0.02	0.10	9.516	0.02 / 0.11 = 90.9	700,000	50,500,000
5	0.01	0.11	10.417	0.01 / 0.11 = 100.0	650,000	150,000,000

would result in Event ID 3 ($0.05 < 9.1\%$) and Event ID 4 ($45.5\% < 0.62 < 72.7\%$). This ensures that the events are selected in proportion to their relative frequency.

The final step is to sample each realization of every event selected allowing for secondary uncertainty. If a parametric distribution (such as a Beta distribution, see Section 1.11.2.1) is used, then this can be done directly within the functionality of most simulation tools. If an **empirical distribution** (i.e. one that has not got a particular parametric form) is provided by the model, then a similar approach to the normalized exceedance frequency approach provided can be used to obtain the realization of the event.

The incorporation of uncertainty is straightforward for one ELT, but if multiple ELTs are being simulated (for a given peril), as is often the case, then the question is how to properly reflect the uncertainty correlation between ELTs. A cautious approach sometimes used (which will bias the results high) is to assume full correlation in uncertainty, and use the same sampled value of the Uniform(0,1) distribution for each ELT entry for a given realization of an event. For example, if a value of 0.75 is sampled from the uniform distribution and used to calculate the appropriate value from the Beta distribution representing the uncertainty for Event 1 in ELT A, then the same 0.75 figure is used to calculate the value from the Beta distribution for Event 1 in ELT B, and so on. This will not preserve the correlation in the same way that the grouping procedure described in Section 1.10.5.10 does. However, more sophisticated approaches are available that do preserve the correlation in uncertainty maintained by using the combination techniques described in Section 1.10.5.10. They are, however, proprietary, and the interested reader is directed to either their model vendor or the vendor of their simulation software.

The end result of the simulation process is a set of event realizations for every year, at the granularity and for the financial perspective of the ELT. These figures can then be used to derive further metrics in a very flexible way. For example, the AEP can be calculated by calculating the sum of event losses in each year. The ordering of the number of years, together with the probability of each year occurring being 1/(Number of Simulation Years) gives the AEP probability distribution. For example, if 100,000 simulation years are sampled, and ordered from high to low, then the loss value of the 1000th year represents the 100 year ($1000 \times (1/100{,}000) = 100$) AEP loss. The OEP can also be calculated in a similar way, but instead of calculating the sum of losses for each year, the maximum event loss for each year is used instead.

Financial structures can also be applied within a simulation framework, so long as they work at the same level (or a higher level) than the original ELT (a policy-level ELT could not be used to evaluate location-level financial structures). The financial structures are simply applied to the event losses for each year and the losses net of the financial structures evaluated and captured. The flexibility of Monte Carlo simulation often means that more complex structures can be

evaluated than can readily be done using either closed form integration or numerical convolution techniques.

Historically, insurance and reinsurance companies tend to take the net loss pre-cat ELT tables out of the model and use a simulation framework to calculate the losses for other perspectives. In many catastrophe models the convolution of loss distributions to different levels, and the application of financial structures are performed using either closed form integration or a discretized convolution technique. However, models are emerging that use Monte Carlo simulation as the basis for aggregating and applying financial structures from location-coverage level upwards. We anticipate this trend continuing as computing power increases.

1.10.5.10 Combining ELTs

It is often necessary to combine ELTs in order to calculate metrics at a more aggregated level. For example, ELTs for individual locations could be combined in order to calculate the metrics from a multi-location policy. Alternatively ELTs from separate policies may need to be combined to give a portfolio view. In order to combine ELTs, and preserve the correlation provided by the catastrophe model, the ELTs must be of the same version of the model (or at least a version with consistent event IDs). Although the model itself will contain a consistent set of events, each ELT may contain a different number of events, as events that do not generate any loss are sometimes excluded from the model output.

The first step in combining two ELTs is to compare the event IDs; events that are not in both ELTs can be added to the combined ELT. For events that are in both ELTs, the following steps are necessary:

1) The event ID and rate should be the same in both ELTs, so these can be bought across without change.
2) The mean loss, the correlated standard deviation, and the exposure impacted, for the same event ID, should be added.
3) The independent standard deviation for the same event in each ELT should be squared, summed and square-rooted.

The resultant combined ELT can then be used to calculate the metrics (e.g. AAL, SD) described above.

1.10.6 Year Loss Table (YLT)

YLTs can be divided into two different types:

- *Type 1*: A YLT where each loss entry loss represents the realization of an uncertainty distribution – NOT the mean loss. There is no uncertainty distribution provided in such a YLT – it is inherent within the loss figures.
- *Type 2*: A YLT where each loss entry represents the mean event loss, and a standard deviation, or distribution of uncertainty, is provided together with the mean loss.

A typical YLT structure is shown in Table 1.9. The standard deviation (SD_i) of each event, i, is only present for Type 2 YLTs.

1.10.6.1 Limitations and Benefits of YLTs

A disadvantage of YLTs is that making adjustments to the frequency is not as straightforward as for ELTs as the events have essentially already been drawn from the frequency distribution.

A second disadvantage is that the number of years in the YLT may not be the same number of years needed for the simulation framework that the insurance company is using. For example, perhaps an internal model framework is set up that uses 10,000 years and the YLT contains

Table 1.9 Sample year loss table.

Year (k)	Loss Number in Year (j)	Event ID (i)	Loss (X_i)	Exposure Impacted (E_i)	Standard Deviation (SD_i)
1	1	8	850,000	200,000,000	2,150,000
1	2	6	700,000	50,500,000	1,950,000
2	1	9	1,000,000	100,000,000	2,300,000
3	1	11	800,000	60,000,000	2,100,000
3	2	7	650,000	150,000,000	2,050,000

20,000 years. In this case, stratified sampling (e.g. Calder, Couper and Lo, 2012) can be used to reduce the number of years while minimizing the impact on the simulation error. If an increase in the number of years is required, then years can either be re-sampled from the YLT, however, the increased number of years will not reduce the simulation error for a type 1 YLT and will only partly reduce the simulation error for a type 2 YLT.

An advantage of YLTs is that any dependency between events in a year can be explicitly incorporated (e.g. clustering). There is no requirement that the distribution of events in a year fit a particular parametric distribution; for example, they could be the output of a global climate model, or the output of a complex clustering scheme where there is some dependency between the frequency and severity of events. A dependency between years and seasonality could also be incorporated in a YLT (although this is unusual).

A practical advantage of type 1 YLTs is that they can be used 'as is', without recourse to a simulation engine, to calculate metrics; including the impact of financial structures that depend on multiple events in a year such as reinstatements and aggregate event limits. A type 2 YLT still needs the uncertainty around the mean event losses to be simulated.

An advantage of a type 2 YLT is that the uncertainty is explicitly represented, and so increased sampling from the uncertainty distribution within a simulation framework can result in a more robust propagation of the uncertainty. Since the mean and the standard deviation of event size are known, a type 2 YLT can also be used to construct an ELT: each event having rate 1/N (assuming no repeated events). Although the event rate can be calculated, and the mean and standard deviation of loss populated for each event, the precise frequency distribution for the ELT may not be known, given the flexibility of frequency distributions that can be represented in a YLT.

1.10.6.2 Calculating Metrics from YLTs

For either kind of YLT, the AAL is calculated in the same way: by summing the losses (mean losses or realizations of loss depending on the type of YLT) across all years (k) and dividing by the number of years (N), i.e.

$$AAL = \frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{m_k} X_{j,k} = \frac{1}{N}\sum_{k=1}^{N} S_k = \frac{Z}{N} \tag{1.9}$$

where m_k = number of events in year k, $X_{j,k}$ = realization of loss or mean loss for the j^{th} event in year k, S_k = the sum of event losses within year k, and Z = the sum of event losses for the whole YLT.

The method of calculation of the standard deviation of a YLT depends upon the type of YLT. If the YLT is type 1 (i.e. represents the realization of each event loss), the uncertainty is inherent

within the YLT realization and so the variance must be calculated explicitly from the data using the formula for sample variance:

$$Var[S] = \frac{1}{N-1}\sum_{k=1}^{N}(S_k - AAL)^2 \tag{1.10}$$

where:

$$S_k = \sum_{j=1}^{m_k} X_{j,k} \tag{1.11}$$

$$SD = \sqrt{Var[S]} \tag{1.12}$$

If, however, the YLT is of the second type (i.e. the loss represents the mean and the SD of each event is provided), then using the formula above would understate the uncertainty as it would not take into account the SD in each event loss. Instead, the variance and standard deviation of this type of YLT can be calculated as (courtesy of Ye Liu):

$$Var[S] = \frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{m_k}(SD_{j,k})^2 + \frac{1}{N}\sum_{k=1}^{N}(S_k - AAL)^2 \tag{1.13}$$

$$SD = \sqrt{Var[S]} \tag{1.14}$$

As a practical note, in some cases, the number of years in a YLT may be less than N, perhaps because for some years there are no event losses generated for the portfolio. This might be particularly the case for low frequency perils (such as an earthquake in Austria). However, even if this is the case, the full value of N must still be used in calculating the AAL and SD; N must not be reduced even if there are years with no events.

The calculation of the OEP and AEP from a type 1 YLT is described in Section 1.10.5.9. YLTs can be combined in the same way as ELTs; the years and events are matched between YLTs and the losses and SDs combined in the same way as for ELTS. The correlation between two YLTs can be calculated in the same way as for two ELTs.

The calculation of the OEP and AEP from a type 2 YLT is, however, more complex. An approximate AEP and OEP can be calculated by treating the mean losses for each event as realizations of loss and following the process described in Section 1.10.5.9. However, this approach would only provide an approximation and would not use the uncertainty information provided in a type 2 YLT. A more comprehensive approach would involve simulating a number of type 1 YLTs by sampling the uncertainty distributions around each event in the type 2 YLT. For each generated type 1 YLT, an AEP curve can be calculated (as described in Section 1.10.5.9). At each probability level the distribution of losses can be averaged to calculate a final AEP. This, however, is considerably more complex and computationally intensive than the approximate approach. Research would be needed to establish under which conditions the approximation is valid, and when the more complete but complex approach should be used.

1.11 Statistical Basics for Catastrophe Modelling

This section introduces some of the basic distributions used within catastrophe modelling. This is a catastrophe modelling and management textbook, not an actuarial or statistical primer, and so many statistical topics will be beyond the scope of this book. The purpose of this section is to enable the reader to have some statistics for relevant distributions 'to hand' rather than provide a

complete primer in statistics for a full range of distributions. Readers are referred to Forbes *et al.* (2011) for more details of many distributions, and Chapters 9 and 10 of Sweeting (2011) for some basic statistics.

Throughout this section, $\boldsymbol{X}$ is a random variable with $\boldsymbol{x}$ a specific value of this random variable. For the distributions described below, the following statistics are provided.

- Probability function $\boldsymbol{Pr}(\boldsymbol{x})$ (for discrete distributions): the probability that each discrete value occurs.
- Probability density function (pdf) $\boldsymbol{f}(\boldsymbol{x})$ (for continuous distributions): a function that gives the relative chance of a value occurring, such that the integral over the entire function is equal to one, but the probability of each precise value occurring is zero.
- Cumulative distribution function $\boldsymbol{F}(\boldsymbol{x})$ (for continuous distributions): the probability that the random variable is less than or equal to a given value. It is equal to the integral of the pdf from the lowest possible value of the distribution to the given value.
- Mean or expectation of the random variable $\boldsymbol{E}[\boldsymbol{X}]$
- Variance of the random variable $\boldsymbol{Var}[\boldsymbol{X}]$.

The formulas for estimating the mean and variance from a sample of data of N observations (rather than the full but unobservable population) are as follows:

$$Sample\ Mean = \overline{E[X]} = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{1.15}$$

$$Sample\ \ Variance = \overline{Var[X]} = \frac{1}{N-1}\sum_{n=1}^{N} (X_n - \overline{E[X]})^2 \tag{1.16}$$

Also, note that the standard deviation (SD) is the square root of the variance, and the coefficient of variation (CV) is the standard deviation divided by the mean.

In some cases it can be useful to estimate parameters of the distributions from the observed (or sample) mean and variance. Where possible, these estimators are also given for the readers' convenience. Note that this 'method of moments' type approach is often less robust than other methods of parameter estimation; in particular, maximum likelihood estimation (e.g. Forbes *et al.*, 2011) is the preferred technique. However, for ease of application, very often the method of moments approach is still used.

The concept of a *compound distribution* is used in this chapter; the sum of a random number of independent and identically distributed random variables is a compound distribution. Namely, if $S = X_1 + X_2 + X_3 + \cdots + X_N$, where the X variables are independent and identically distributed random variables and N is also a random variable that is independent from the X variables, then S has a compound distribution. The expectation and variance of a compound distribution (e.g. Kaas *et al.*, 2009; Klugman *et al.*, 1988) are as follows:

$$E[S] = E[X]E[N] \tag{1.17}$$

$$Var[S] = E[N]Var[X] + Var[N](E[X])^2 \tag{1.18}$$

As an example of the application of this, consider the example of a property worth £1,000,000 with a 1% annual frequency of being destroyed:

$E[N] = 1\% = 0.01$
$Var[N] = 0.01 \times (1.00 - 0.01) = 0.0099$ (assuming a Bernoulli distribution which is appropriate for this scenario)
$E[X] =$ £1,000,000

$Var[X] = 0$ (there is no uncertainty in the size of loss as the property is assumed completely destroyed if an event happens)

$$E[S] = E[N]E[X] = 0.01 \times 1{,}000{,}000 = £10{,}000$$

$$Var[S] = 0.01 \times 0 + 0.0099 \times (1{,}000{,}000)^2 = £9{,}900{,}000000$$

Which gives a standard deviation of £99,499.

1.11.1 Discrete Distributions

Discrete distributions are often used in catastrophe modelling to model the frequency of events occurring within a particular year for a particular peril-region. The main distributions used are the Poisson and the negative binomial distributions, and these are described below.

1.11.1.1 Poisson

The Poisson distribution is commonly used as the distribution for event numbers in a catastrophe model. It gives the probability of a number of independent events occurring in a specified time. The assumption of independence and the fact that the mean equals the variance indicates that it may not be suitable as a frequency distribution for some perils; for example, for some European windstorm regions (see Chapter 3.3.1) or for North Atlantic hurricane numbers (see Chapter 3.2.1) where the variance in numbers is known to be greater than the mean (over-dispersion/clustering).

For $x = 0, 1, 2, \ldots$ and parameter $\lambda > 0$:

$$\Pr(x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{1.19}$$

$$E[X] = Var[X] = \lambda \tag{1.20}$$

A probability function for a Poisson with $\lambda = 2.0$ is shown in Figure 1.8.

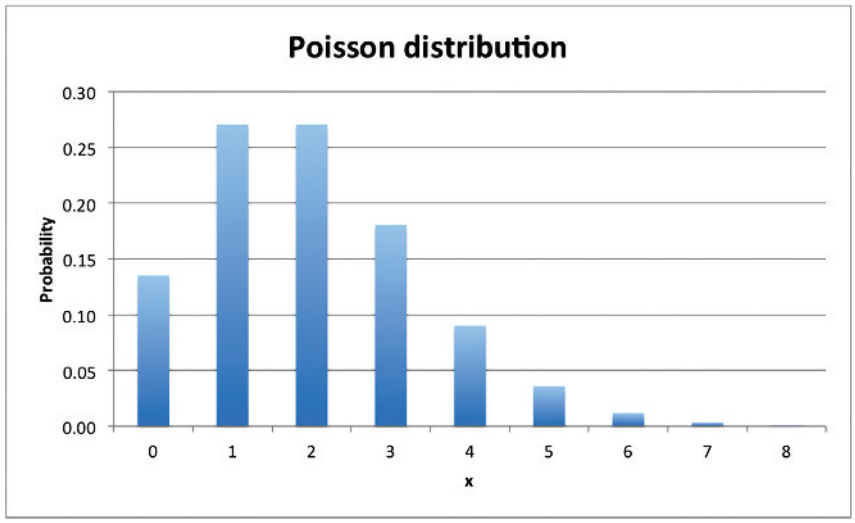


Figure 1.8 Poisson distribution with mean = 2.0

1.11.1.2 Negative Binomial

The negative binomial distribution is sometimes used within catastrophe modelling when a parametric frequency distribution is required, but there is some known dependence between events. The variance is greater than the mean for a negative binomial distribution. This distribution has been used to model numbers of European windstorms or North Atlantic hurricanes. There are many different ways of defining the negative binomial distribution. The definition below utilises the gamma function (Γ), which allows the parameter r to take non-integer values.

For $x = 0,1,2, \ldots$ and parameters $r \geq 1$, $m > 0$

$$\Pr(x) = \frac{\Gamma(x+r)}{x!\Gamma(r)}\left(\frac{m}{r+m}\right)^{x}\left(\frac{r}{r+m}\right)^{r} \tag{1.21}$$

$$E[X] = m \tag{1.22}$$

$$Var[X] = m + \frac{m^2}{r} \tag{1.23}$$

If the mean and variance are known then m and r can be estimated from:

$$m = E[X] \tag{1.24}$$

$$r = \frac{E[X]^2}{Var[X] - E[X]} \tag{1.25}$$

A negative binomial distribution for $r = 2$ and $m = 2.0$ (mean = 2.0) is shown in Figure 1.9.

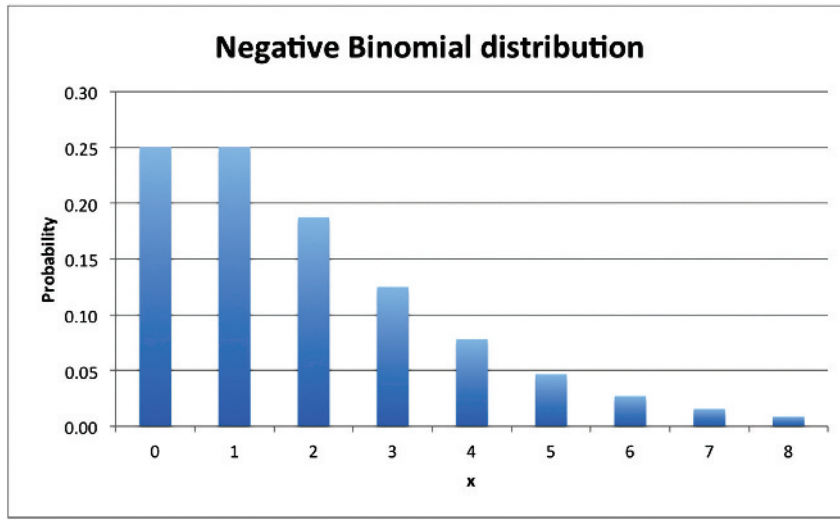


Figure 1.9 Negative binomial distribution for r = 2.0 and m = 2.0

1.11.2 Continuous Distributions

There are many parametric continuous distributions available. In this section we describe only two, the Beta distribution and the single parameter Pareto distribution, because of their common use in catastrophe model building and catastrophe pricing respectively. For information on other distributions, refer to Forbes *et al.* (2011).

1.11.2.1 Beta

The Beta distribution is commonly used within the vulnerability component of catastrophe modelling to describe the uncertainty around the mean damage ratio (the expected percentage of sum insured lost given a certain level of hazard). It is well suited to this, since it describes values in the range zero to one, which encompasses all possible values of damage ratio. This distribution is also used within the financial component of some catastrophe models to parameterize the aggregated uncertainty of event losses in order to quantify the impact of financial structures at that particular level of aggregation; in this usage the random variable X is the mean event loss divided by the exposure within the footprint of the event.

Before describing the Beta distribution, we must introduce the incomplete Beta function and the Beta function. The incomplete Beta function:

$$B(x;a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt \tag{1.26}$$

The Beta function is the incomplete Beta function where x = 1:

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt \tag{1.27}$$

The Beta distribution is then given as follows:

For $0 < x < 1$ and parameters $a > 0,\ b > 0$:

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \tag{1.28}$$

$$F(x) = \frac{B(x;a,b)}{B(a,b)} \tag{1.29}$$

$$E[X] = \frac{a}{a+b} \tag{1.30}$$

$$Var[X] = \frac{E[X](1-E[X])}{a+b+1} = \frac{ab}{(a+b)^2(a+b+1)} \tag{1.31}$$

If the mean and variance are known, the parameters a and b can be estimated as follows:

$$CV = \frac{SD}{E[X]} \tag{1.32}$$

$$a = \frac{E[X]^2(1-E[X])}{Var[X]} - E[X] = \frac{(1-E[X])}{CV^2} - E[X] \tag{1.33}$$

$$b = \frac{a(1-E[X])}{E[X]} \tag{1.34}$$

The probability density function for a Beta distribution with $a = 4.0\ and\ b = 3.0$ is shown in Figure 1.10.

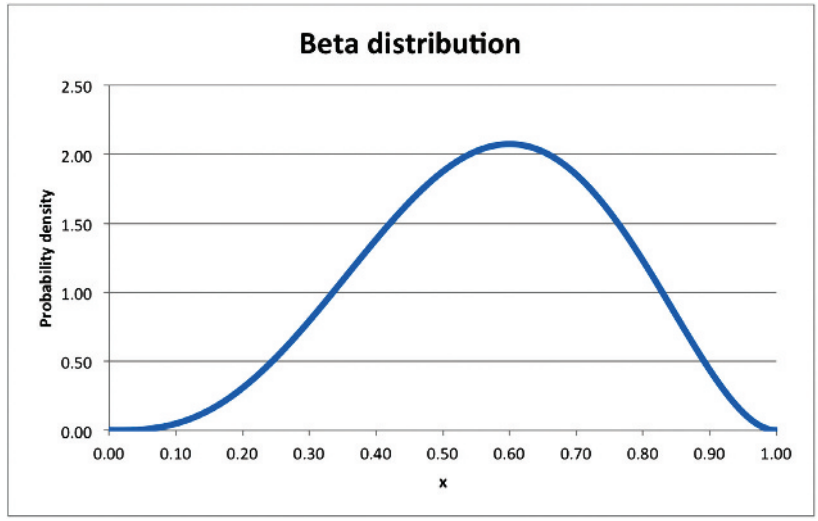


Figure 1.10 Beta distribution with a = 4.0 and b = 3.0

1.11.2.2 Pareto (One Parameter)

There are many different types of Pareto distribution. These distributions are 'heavy-tailed' distributions which means that there is an increased probability of extreme losses compared to many other distributions – which makes them appropriate for modelling catastrophe losses. The distribution described here is usually described as the one parameter Pareto distribution. Although it has two parameters, one is prescribed up-front by the threshold beyond which the loss distribution applies, so it is not a free parameter. This distribution is not normally used within the building of exposure-based catastrophe models, but is often used within catastrophe pricing to describe the severity of event losses. Particular values of the shape parameter, α, are commonly used as starting points to parameterize the distribution for different types of peril (see Table 2.3 in Chapter 2.6.4.2). It can therefore also be a useful distribution when trying to represent non-modelled risk using an actuarial approach (see Chapter 5.4.7).

Other types of Pareto distribution commonly used include the two parameter Pareto and the Generalized Pareto distributions. Description of these is beyond the scope of this book, but references can be found in standard actuarial and statistical texts (e.g. Klugman, Panjer and Willmot, 1988).

The one parameter Pareto distribution is given as follows:

For threshold $t > 0$, $x > t$ and parameter $\alpha > 0$:

$$f(x) = \frac{\alpha t^{\alpha}}{x^{\alpha+1}} \tag{1.35}$$

$$F(x) = 1 - \left(\frac{t}{x}\right)^{\alpha} \tag{1.36}$$

$$E[X] = \frac{\alpha t}{\alpha - 1} \tag{1.37}$$

$$Var[X] = \frac{\alpha t^2}{(\alpha - 1)^2(\alpha - 2)} \tag{1.38}$$

If the mean is known, then α can be calculated from:

$$\alpha = \frac{E[X]}{E[X] - t} \tag{1.39}$$

The probability density function of a Pareto distribution with $t = 10.0$ *and* $\alpha = 2.5$ is shown in Figure 1.11.

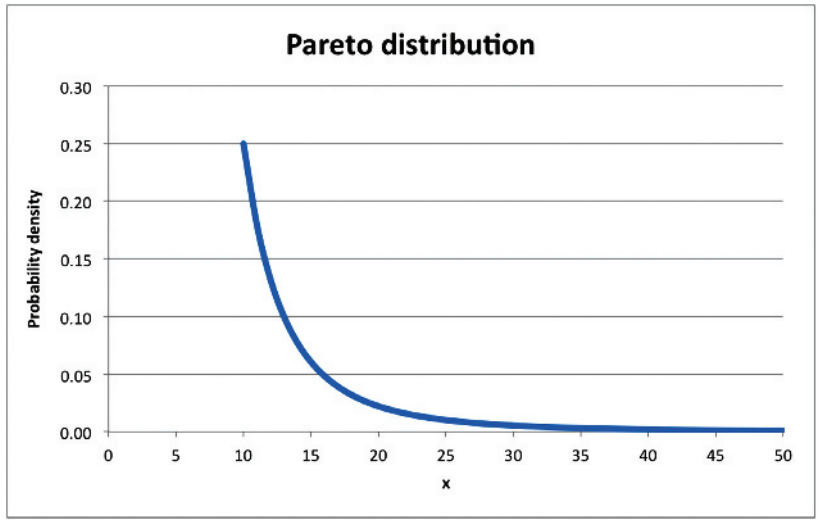


Figure 1.11 Pareto distribution with t = 10.0 and $\alpha = 2.5$

1.11.3 Coherent Risk Measures

This section describes coherent risk measures. Understanding of this term may useful to the catastrophe risk analyst when considering appropriate metrics to use for reporting and risk tolerance purposes (see Chapter 2.10.1). A coherent risk measure (e.g. Artzner *et al.*, 1999; Sweeting, 2011) is one which satisfies the criteria below, where $F(\ )$ is a function which implements a particular risk measure:

Sub-additivity: $F(X + Y) \leq F(X) + F(Y)$

Portfolio diversification should always lead to lower risk. In other words, combining two risks should not create any additional risk. The VaR metric does not always satisfy this criterion.

Positive Homogeneity: $F(cX) = cF(X)$, where c is a constant.

Multiplying the losses by a constant factor (e.g. inflation or a currency conversion – or even aggregating identical risks) should result in the risk measure changing by the same factor. Variance does not satisfy this criterion: $Var(cX) = c^2\ Var(X)$.

Monotonicity: $F(X) > F(Y)\ if\ X > Y$

If each entry in an Event Loss Table for portfolio X is greater than that for portfolio Y, the risk measure for X should be greater than the risk measure for Y. In other words, if the losses increase, the risk measure should increase.

Translation Invariance: $F(X + c) = F(X) + c$, where c is a constant.

If the amount of loss is increased by a fixed amount, then the risk measure increases by the same amount.

Notes

1. http://www.aon.com/reinsurance/analytics/remetrica.jsp
2. https://www.towerswatson.com/en/Services/Tools/igloo
3. http://www.oasislmf.org/

4. https://www.acord.org/webfiles/acord_knowledge/20130313_CatExp.pdf
5. See, for example, the US Department of Defense standard: http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html
6. https://www.cresta.org/
7. http://www.unicede.com/
8. http://www.ehso.com/siccodes.php
9. http://www.ehso.com/naics.php

References

ABI (2011) *Industry Good Practice for Catastrophe Modelling: A Guide to Managing Catastrophe Models as Part of an Internal Model under Solvency II*. Association of British Insurers, London.

Actuarial Standards Board (2000) *ASOP No. 38: Using Models Outside the Actuary's Area of Expertise (Property and Casualty)* Actuarial Standards Board, London.

Artzner, P., Delbaen, F., Eber, J-M. and Heath, D. (1999) Coherent measures of risk. *Mathematical Finance*, **9** (3), 203–228.

Calder, A., Couper, A. and Lo, J. (2012) *Catastrophe model blending: techniques and governance. In General Insurance Convention (GIRO)* UK Actuarial Profession, Brussels.

Embrechts, P. and Frei, M. (2009) Panjer recursion versus FFT for compound distributions. *Mathematical Methods of Operations Research*, **69** (3), 497–508.

Forbes, C., Evans, M., Hastings, N. and Peacock, B. (2011) *Statistical Distributions*. John Wiley & Sons, Inc., Hoboken, NJ.

Friedman, D. (1984) Natural hazard risk assessment for an insurance program. *Geneva Papers on Risk and Insurance*, **9** (30), 57–128.

Grossi, P. and Kunreuther, H. (2005) *Catastrophe Modelling: A New Approach to Managing Risk*. Springer, New York.

Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2009) *Modern Actuarial Risk Theory Using R*. Springer, Berlin.

Klugman, S., Panjer, H. and Willmot, G. (1988) *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc., New York.

Ley-Borrás, R. and Fox, B. (2015) Using probabilistic models to appraise and decide on sovereign disaster risk financing and insurance. *Financing and Insurance Policy Research Working Papers*. World Bank Group, Washington, DC.

LMA (2013) *Catastrophe Modelling: Guidance for Non-Catastrophe Modellers*. Lloyd's Market Association, London.

Mouroux, P. and Le Brun, B. (2006) Presentation of RISK-UE Project. *Bulletin of Earthquake Engineering*, **4** (4), 323–339.

Slingsby, A., Wood, J., Dykes, J., Clouston, D. and Foote, M. (2010) Visual analysis of sensitivity in CAT models: interactive visualisation for CAT model sensitivity analsis. Paper presented at Accuracy 2010 Symposium, Leicester.

Sweeting, P. (2011) *Financial Enterprise Risk Management*. Cambridge University Press, New York.

Swiss Re (2013) *A History of Insurance.* http://media.swissre.com/documents/150_history_of_insurance.pdf

Swiss Re (2015) Natural catastrophes and man-made disasters in 2014. *Sigma,* 2/2015.

Thoyts, R. (2010) *Insurance Theory and Practice*. Routledge, London.

Wald, D.J., Quitoriano, V., Heaton, T.H. and Kanamori, H. (1999) Relationship between peak ground acceleration, peak ground velocity, and modified Mercalli intensity in California. *Earthquake Spectra*, **15** (3), 557–564.

Woo, G. (2002) Natural catastrophe probable maximum loss. *British Actuarial Journal*, **8** (V), 943–959.