The Actuarial Profession
making financial sense of the future

13 August 2012

# Catastrophe Model Blending
## Techniques and Governance


***Alan Calder, Andrew Couper, Joseph Lo, Aspen*** *
***foreword by Stephen Postlewhite***


### Foreword

Catastrophe modelling has become a key underwriting and risk management tool over the past 15 years. However with the advent of increasingly complex stochastic catastrophe modelling tools, new risks and dangers appear. The proprietary nature of vendor models has led to opaqueness and a protectionist attitude, which itself leads to a lack of understanding and ability to properly challenge the assumptions made. Modelling firms are making strides to rectify this, making the models more transparent and highlighting uncertainty within the key assumptions. This is a trend which should be encouraged and embraced by the industry.

As modelling platforms open up it will be increasingly possible to critically review models and assumptions and thus develop in-house views of risk and analyse both pricing and accumulation across multiple model versions. These versions may take the form of adjusted vendor models or blends between models. Successfully implementing these nuanced views of risk requires collaboration between the fundamental research scientists and those with the ability to apply this scientific research to the business problems we are trying to solve.

Actuaries are well placed to play an important role in this area. This paper neatly summarises one such collaborative project at Aspen, which has been successfully implemented thanks to our research and development team across both catastrophe risk and actuarial.

*Stephen Postlewhite, MA,FIA*

*Acting Group Chief Risk Officer, Aspen*


### Acknowledgement

The authors are grateful for the interest from and the opportunities to discuss ideas with:

Ian Cook (Willis Re), Guillermo Franco, John Major and Imelda Powers (Guy Carpenter), and Ed Tredger (Beazley).

The authors would also like to thank their colleagues in Aspen for their assistance.



* Respectively, Group Head of Catastrophe Risk Management, Chief Actuary and Chief Risk Officer of Aspen Bermuda, and Head of Actuarial Research & Development. Corresponding author: Joseph Lo, London, U.K.; jo.lo@aspen.co

# 1 Introduction

## 1.1 The Purpose of the Paper

In this introductory section, we shall see the following pattern of activities when using multiple models in pricing:

- Obtain results from multiple models
- Critically compare and contrast these results against one another, as well as against one's own more qualitative assessments, so that strengths and limitations of each of the models would be highlighted
- Given this critical assessment, weights are placed on them for weighted averages to be taken (some models can and will have zero weights)

This is a common practice in our day-to-day actuarial work. A key contribution this paper will make is that these three steps can also be taken when accumulating catastrophe exposures. In Sections 4 and 5.1, given that there is typically more than one model available, we will argue that to understand catastrophe accumulations well, the first two steps are very important to perform. We expect these two steps to be common practice already for firms that have material exposures to catastrophe risk. However, for a variety of technical and operational reasons, the third step is not typically taken up, even in cases when blending would be very commonly performed in other areas of work such as pricing. The paper will suggest that:

- *Openness* to blending is important when using multiple models – even if one will not blend in every single catastrophe peril and zone. With this openness, the undertaking of the first two steps would have greater value.
- There are technical methods available for blending catastrophe accumulations. This will be discussed in general in Section 2 and a specific example technical solution is presented in Section 3. Through the use of a general discussion and a discussion surrounding a detailed example, technical and conceptual considerations will be reviewed in depth, allowing us to share with the reader some of the challenges we have faced in our experience of blending.
    - Among the many technical concepts considered in this paper, we shall consider the merits of two commonly employed ways of blending outputs of catastrophe models, indicating our preference for one over the other (see Section 2.1).
        - We shall also stress that there is a variety of ways to conduct "a straightforward 50-50 blend", urging practitioners to qualify such phrases with methodological comments (see Section 2.2.1)
    - We shall explore how blending can be related with uncertainty. In particular, we shall highlight that:
        - We cannot reduce all uncertainty by blending multiple models. In particular, common datasets and common scientific theories are two dominant reasons why we cannot do so.
        - We can reduce our vulnerability to drivers of uncertainty that are specific to the individual models, through using particular types of blending (see Section 2.1.2, and in particular 2.1.2.3).
        - We also begin to consider blending as a tool to reduce the negative *consequences* of uncertainty (see Section 2.1.1.2, and also 2.1.2.4), posing questions for future research.

- o The example technical solution will also indicate how one may enter different adjustments for different portfolios in an accumulation. This is extremely important where one would like to blend with, for example, outputs of experience rating models.

- Governance is a key operational part of the use of multiple catastrophe models. We shall explore this in Section 4. Referring the reader to existing literature on general governance concepts, we shall highlight model validation and governance activities that we have found particularly helpful. In relation to blending, we consider how specific ways of blending – for example, as in Section 3 – can also help the governance of the use of catastrophe models in general, by providing a faster quantitative feedback loop to channel expert opinions.

As the reader is (or will be) aware, existing good and thoughtful literature already exists on the areas of the use of multiple models and model governance. As much as possible, we shall make references to it and avoid repetition. We note that an overwhelming proportion of this literature is authored by brokers, catastrophe modelling specialists, and industry and regulatory bodies. It is, therefore our hope throughout its production, that this paper would make a unique contribution to these topics from the perspective of a (re)insurer, giving us an opportunity to begin a debate among (re)insurer practitioners on the various challenges and solutions facing us.

For ease of navigation and reference, we include a table of contents here. In summary:

- The remainder of this introductory section sets up the background and motivation for the discussion in the further sections.

- We shall explore existing blending methodology in Section 2, discussing the merits of different approaches, indicating a preferred approach.

- A detailed example blending solution is described in Section 3, using our preferred blending approach from the previous section. It will also explore a wide variety of issues in blending such as adjustments for individual portfolios.

- A discussion on governance of uses of catastrophe models can be found in Section 4: especially in relation to the blending demonstrated in the previous two sections.

- Further topics are discussed in Section 5, touching on Solvency II, non-modelled perils and the future of catastrophe model blending.

This paper assumes a basic understanding of the concepts of catastrophe models and their outputs, relying on the readers' familiarity with commonly used terminology. We expect the printed version of this paper will be in black and white. A full colour version can be downloaded from the Actuarial Profession website.

# Contents

## 1.2 Catastrophe Modelling

Catastrophe Modelling has been in use within the (re)insurance industry since the late nineteen eighties. The earliest models were largely deterministic models, allowing users to estimate “as-if” losses for either historical events, or “what if” type “worst case” set of scenarios. However, beginning in the early nineties, models began to be produced on a fully probabilistic basis, simulating tens to hundreds of thousands of possible catastrophe events in order to estimate a more complete range of potential losses for a given peril.

Catastrophe models are now adopted by the (re)insurance industry, in an attempt to provide improved estimates of exposures to extreme events. Entities may have only limited actual loss experience with which to quantify such exposures. This was the situation in Florida in 1992 when Hurricane Andrew hit to the South of Miami as a strong category four hurricane. A number of insurers failed principally as they relied on trending loss experience from the previous two decades: decades with few hurricane landfalls in the state and during which insured values increased, construction codes changed and the coverage detail also changed.

Well researched catastrophe models have the benefit of using longer records of observed hazard information of up to a hundred years to attempt to quantify the recurrence of extreme events, although often vulnerability components within models are inferred from much more limited observations.

Models are typically used as part of two main underwriting processes:

- in order to calculate expected losses, standard deviations of losses, and probabilities of losses to (re)insurance contracts for individual risks or portfolios for pricing;
- or to estimate the overall level of accumulation of several risks or contracts to a given peril or all perils, usually reported as an annual loss percentile such as 1% or “1 in 100”, or fed directly into stochastic capital models.

We shall discuss each of these processes in Sections 1.4 and 1.5 in more detail. For more details on catastrophe models themselves:

- The UK Actuarial Profession organises regular catastrophe modelling seminars – a recent one was held in March 2011, with speakers from major catastrophe modelling companies, as well as from actuarial practitioners. In particular, the present paper will regularly refer to the slides from the talk (Cook, 2011).
- At least two recent GIRO Working Parties considered natural catastrophes and reported back: one in 2002 (Sanders & others, 2002) and then in 2006 (Fulcher & others, 2006). The 2002 report deals with extreme events in general, with its Section 4 devoted to catastrophe modelling. The 2006 report discusses catastrophe modelling, with a focus on North Atlantic hurricanes.
- In the US, the Casualty Actuarial Society organises annual Ratemaking and Product Management Seminars. The March 2009 and March 2010 seminars both had catastrophe modelling workshops, the slides of which can be downloaded.
- (Grossi & Kunreuther, 2005) is a collection of essays that introduces catastrophe modelling, including consideration of how natural hazards are assessed and the topic of uncertainty. Insurance applications also feature in this book.
- (Woo, Calcualting Catastrophe, 2011) is a thoughtful – and thought-provoking – account of the subject. It considers catastrophes more generally, citing recent financial catastrophes as examples. It can be considered as a second edition to (Woo, The Mathematics of Natural Catastrophes, 1999).

## 1.3 Types of Uncertainty

A key concept with catastrophe modelling is that of epistemic and aleatory uncertainties. These are two scientific terms used to categorise different sources of uncertainty – we quote from Section 4.2 of (Grossi & Kunreuther, 2005):

- Epistemic uncertainty is the "uncertainty due to lack of information or knowledge of the hazard"
- Aleatory uncertainty is the "inherent randomness associated with natural hazard events, such as earthquake, windstorms and floods"

To this, it is helpful to add an additional uncertainty category: *implementation uncertainty*. There is a difference between: a scientific / mathematical or statistical model, and an *implementation* of such a model. Even a perfect scientific model is implemented by human beings, who are fallible: mistakes can be made even with very good governance and peer review regimes! With the constraint in computing power, simplifications of the scientific models – for example, through discretisations of continuous models or simulations of statistical models – lead to *approximations* of the models, not the precise models themselves. Implementation of models, then, introduces another layer of uncertainty.

The key characteristic of epistemic uncertainty is that it can be reduced through, for example, collection of more data, and development of scientific knowledge. Implementation uncertainty can also be reduced through, for example, better implementation techniques and computing power. However, aleatory uncertainty can never be reduced by our modelling or model implementation efforts. In light of this, when we consider uncertainty in relation to catastrophe models in this paper, where unqualified, we refer to epistemic and implementation uncertainty.

Epistemic uncertainty is especially dominant in catastrophe modelling, as it relies on models to deal with large extrapolations into extreme events – many of which we have never observed in history. Simple statistical bootstrap procedures can show wide uncertainty bounds at the higher return periods (such as 1 in 100 year). Surely to the benefit of the industry, and the consumers and investors the industry serves, Solvency II is already giving a big push into understanding such uncertainty: we expect this effort to continue.

We shall consider in this paper how the use of independently developed catastrophe models can be used to engage with epistemic and implementation uncertainty. In particular, in Section 2.1.2.3, we shall define a new type of uncertainty, which is a subset of epistemic and implementation uncertainty: our reliance on uncertainty from idiosyncratic features of particular model (implementations) can be reduced by the some methods of blending.

Before we move on, it would be useful to be clear on how the terms "model" and "model uncertainty" will be used in this paper.

The term "model uncertainty" is more familiar to actuaries. Model uncertainty is the uncertainty associated with the specification of the model: e.g. whether hurricane landfalls should be modelled by Poisson or Negative Binomial. This narrow type of model uncertainty is difficult to judge, and is usually assessed through sensitivity testing.

"Model uncertainty" can also mean the uncertainty associated with using *implementations* of models as with vendor "models". This is a handy interpretation when we have a few model outputs of the same risk to compare and contrast. It is at the same time narrower than the strict definition of model uncertainty and wider. It is narrower in the sense that by gauging only a handful of models, there is little chance that we would obtain the full space of model uncertainty. For example, there may be more than three competing scientific theories in modelling a particular physical process; or there is every chance that our current scientific or engineering knowledge in the future will include significant concepts and understanding of reality that we do not know of now. It is wider, in the sense that epistemic uncertainty would include a whole range of other "uncertainties", such as "black swans", parameter uncertainty, data uncertainty, uncertainty surrounding the interpretation of data, or modelling resolution uncertainty. These example uncertainties are linked with one another. If more relevant data is available, parameter uncertainty would reduce, while simultaneously, we would have a better chance of testing and refining model specifications, and hence reduce model uncertainty (in the stricter sense).

To avoid confusion, we shall refrain from using the term “model uncertainty” in this paper. However, in line with common practice, we shall abbreviate implementations of scientific models to “models”: and it is in this sense that we shall describe, for example, “vendor models” as models.

## 1.4 The Pricing Process

“Pricing” is a critical part of the overall underwriting process and must allow for all applicable terms and conditions that may impact the potential cash flows arising from a contract.

The process defined as “pricing” can include providing premium quotations to brokers and/or current and prospective policyholders and also deciding upon the share of a risk to accept at the offered premium. Often a distinction is made between the premium that a policyholder is charged, the internal technical premium and a walk-away or minimum premium.

The technical premium is an internal view of how much should be charged to cover all anticipated cash flows as well as provide the expected profit margin the company requires. Commercial considerations will drive the difference between the actual premium charged and the technical premium.

For a given class of business, the technical price will generally be calculated systematically and include:

- A contribution towards the expected cost of all claims that could be suffered by the company. This would include allowance for large and catastrophe claims that may not necessarily be seen within the observed claims experience. The relevant loads for large and catastrophe losses may be determined by exposure analysis, portfolio analysis and/or underwriting judgment.

- Loadings to cover internal expenses and external costs, including outwards reinsurance; and

- A profit margin consistent with the company’s long term target rate of return.

The expected loss cost for catastrophe perils will generally be determined using exposure analysis. In many cases catastrophe models will be used as the basis. However this should be supplemented where applicable by experience rating. The limitations of both experience rating and exposure rating have to be borne in mind by the pricing actuaries and underwriters.

The use of catastrophe models is fraught with difficulties. The validity and accuracy of the exposure data used as the input to the model is critical in determining the reasonableness of the output. The exposure data should be analysed carefully before use, ensuring the geocoding is as accurate as possible. Any changes expected in the portfolio between the data presented (which will usually be at a given point in time and provided by a broker) and an appropriate point during the period to be insured (usually the mid-point), should be incorporated.

Below are some further questions in the process:

- Have all perils been allowed for in the modelling? If not, how will they be incorporated?

- What view of catastrophe model output should be used for pricing?

    - With or without secondary uncertainty?
    - With or without demand surge?
    - With or without fire-following?

- What allowance could or should be made for uncertainty?

- The company may have an in-house research and development team. This team may also have a view of certain models, or parts of certain models. Will this be incorporated into the pricing process?

This paper will discuss blending tools to engage with the last two questions, although we shall touch on non-modelled perils in Section 5.2.

There is a wide variety of methods to calculate the profit margin. Capital requirements and allocations are typically key inputs into the calculations. A key common theme is that the size of the profit margin should in some way be linked with the risk associated with the contract to the company. Catastrophe model outputs are instrumental to assessing the catastrophe components of this risk.

When pricing, underwriters review output from the vendor models. For modelled perils, they may blend models together or they may make adjustments to a single model to take into account specific characteristics of the account being priced. Non-modelled perils will be allowed for separately. Once the account has been bound, the account will be added to the portfolio and the focus will then turn to accumulation, which we shall discuss in the next section.

Some companies also use a walk-away or minimum premium. This will follow the same process as the technical premium but with modified assumptions for some items. For example the expense assumptions may be modified so that only direct expenses are considered and no additional contribution to indirect expenses is required.

The references in Section 1.1 above should give the reader more detailed information on pricing catastrophe risks in insurance and reinsurance (for example, Section 4.3 of (Sanders & others, 2002) and Section 2.3 of (Fulcher & others, 2006)). A range of good papers are available on pricing in general: the UK profession's working party 2007 report on pricing (Anderson & others, 2007) is a good reference, and itself contains a bibliography in specific areas of pricing.

## 1.5 The Accumulation Process

For catastrophe perils, the accumulation process considers the aggregation of modelled losses from a given peril. The aggregation can be at many different levels including across classes of business, segments and also the entire portfolio of the company.

Catastrophe models are generally event-based systems. In a model with a consistent event set, the allowance for correlation between different contracts within a portfolio is then very straight forward. The expected losses across the portfolio can be summed for each event. Other statistics can then be evaluated for each event such as the standard deviation and also the maximum loss for the event given the underlying exposure (provided there are no unlimited policies within the portfolio).

The accumulation process will then consider the overall losses that could be expected for different segments of the portfolio in a given year. The catastrophe models tend to focus on two different views for modelled perils:

1. The maximum loss that occurs in a given year. The distribution curve for the maximum loss is known as the occurrence exceedance probability ("OEP") curve; and

2. The aggregate loss in a given year. The distribution curve for the aggregate loss is known as the aggregate exceedance probability ("AEP") curve.

The metrics can be evaluated across the entire portfolio, or broken down by segments such as a class of business. It could also be constructed for all perils or just based on a single peril. Some companies like to focus on a specific region for a given peril such as Florida Wind within all US Wind, or California earthquake within all US earthquake.

For each peril the company will generally have a standard view of how it should be modelled. This will either be a single vendor model or a blend of two or more models. Where the peril is material to the company, one expects this standard view to be arrived at and governed by appropriate *active* processes, after careful model evaluation and research activities.

The underwriters' view of each individual account could also be incorporated into a (separate) accumulation process. This relies on being able to make adjustments to single vendor models, or blend models together, on a consistent basis from account to account. If this can be done – and we shall illustrate an example approach in this paper – then accounts can be accumulated at portfolio level on a consistent basis. We can then accumulate on a number of bases:

- One or more vendor models;
- The standard agreed company blend; and
- The bottom-up account by account adjusted view.

The company can then decide which basis it would prefer to use for its internal and external reporting purposes.

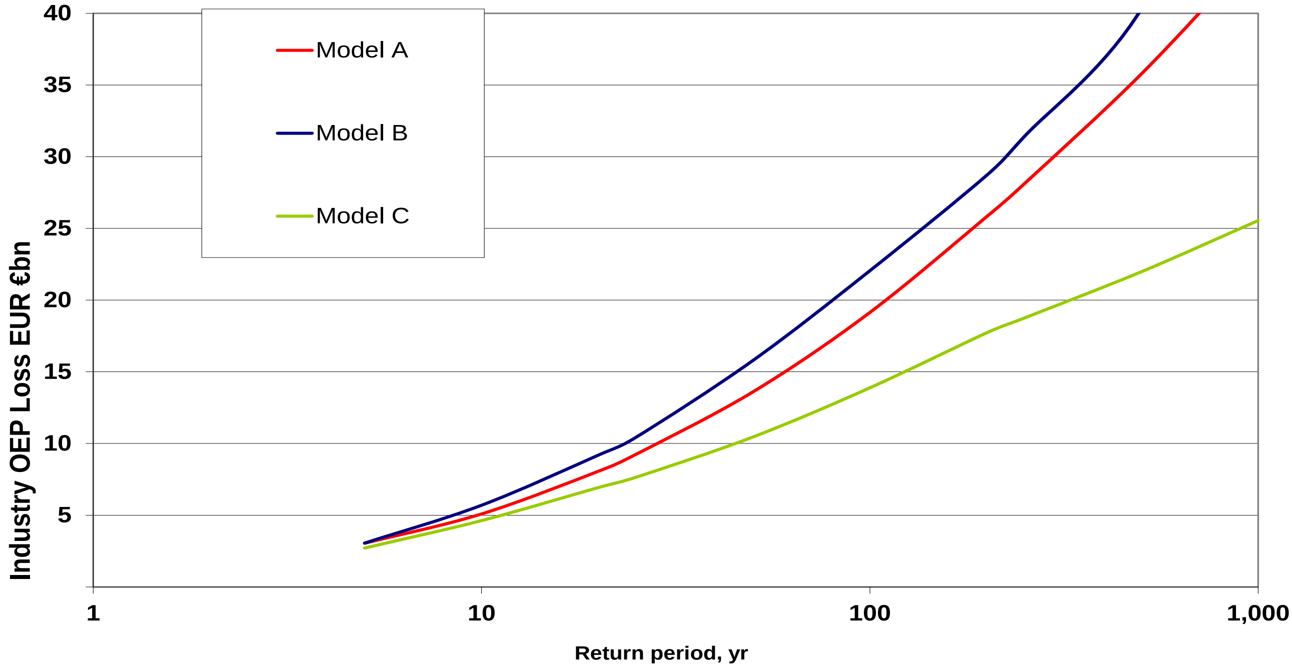
## 1.6 Pricing and Accumulation Example: UK and France Windstorms

Significant catastrophe losses for Europe in the UK and France illustrate two significant issues of using catastrophe models for both pricing and for portfolio accumulations. The most significant peril from an insured loss perspective is windstorm, which are also referred to as extratropical cyclones as they form outside of the tropics in the mid-latitudes.

Two such large losses from Extratropical Windstorms in Europe in the 1990s are (estimates are taken from Table 9 of (Swiss Re, 2012)):

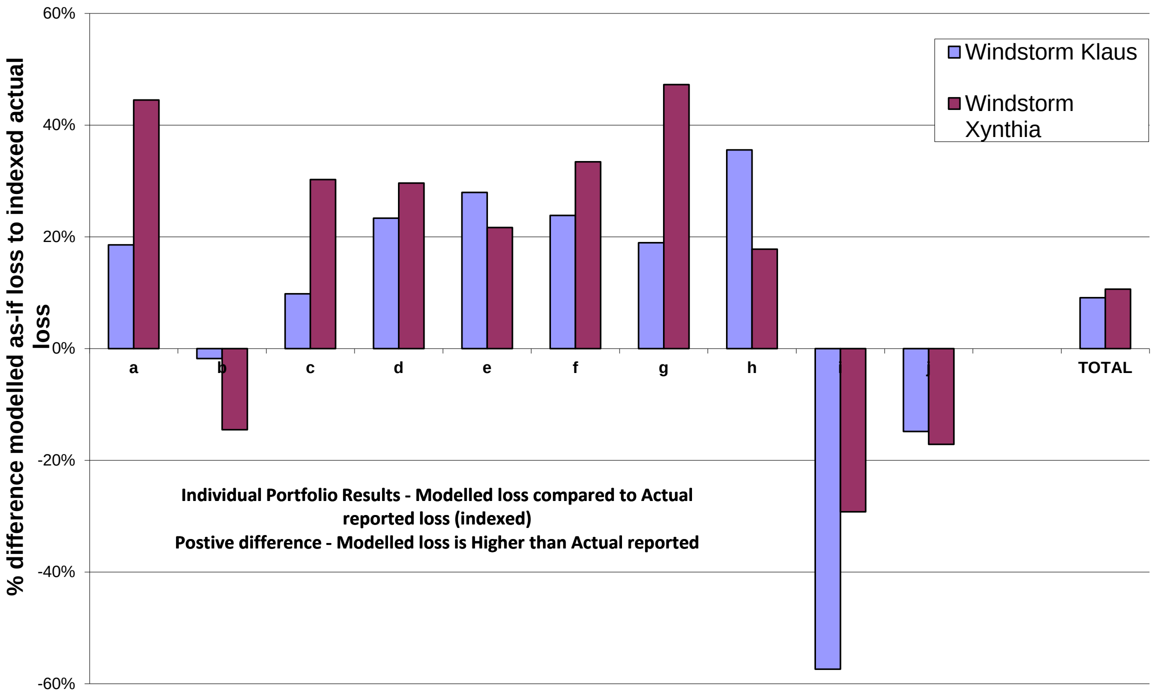
- Windstorm Daria in 1990 (LUNCO code: 90A) in the UK with a market loss for the UK of approximately US$8bn, from the UK and other parts of Europe, in 2011 values
- Windstorm Lothar in 1999 (Lloyds code: 99AD) approx. US$7.8bn in 2011 values for all of Europe including losses from, southern Germany and Switzerland

Such events spurred the development of models in Europe for windstorm. Several specialist model vendors and other market participants such as brokers developed cat models to simulate losses from such events. Due to the lack of standard met office data and accepted methodologies for simulating European windstorms, models for European windstorm have tended to produce a significant variation in outputs on broadly similar exposures. Despite significant revisions and enhancements, a significant degree of variation in results can be obtained between models even at an industry loss level, using consistent original exposure assumptions:




Source: Aspen model validation for European Windstorm model revisions, 2011

Secondly, further uncertainty in the use of such models can be demonstrated by benchmarking of various individual insurance portfolios in France, with recent loss events such as Windstorms Klaus in 2009 and Xynthia in 2010 where issues of trending actual losses to today's values are not as material as attempting to analyse more historical events. Although overall market losses are reasonably well simulated within 10% overall, however individual portfolios in this example reveal differences of up 40%-45%, in either direction:




Source: Aspen model validation for European Windstorm model revisions, 2011

This principally reflects two main facets of cat models. Models use "weighted average" assumptions for general vulnerability functions (which in Europe often tend to cover only broad risk types such as residential, commercial and agricultural), which may not always reflect the complete range of loss experience for specific types exposures written. Models typically use "weighted average" approaches when the user enters information without providing a high degree of granularity. This aspect of “averaging” can be reduced where the original data quality is improved through capturing more detailed attributes of the risks such as construction and year built, number of stories and roof types, etc. to reduce the levels of inference made by the model. However this requires the model developers to be able to discriminate by such characteristics: this may not always be the case, or can mainly be achieved from a significant degree of expert judgement rather than from actual loss experience. Even where such detail exists, a modelling company will derive a single mean damage relationship with a measure of uncertainty rather than produce a bespoke relationship for each and every exposure entered.

Particularly for smaller localised portfolios, such a portfolio may have significant biases in exposures on the ground relative to the geographical resolution of hazard within the model: essentially the resolution of the model may not be adequate enough to capture such effects. Finally claims management practices such as loss adjustment can also impact estimates significantly and again are “averaged” by current generation models. The above example demonstrates the variance of results possible within a single model – the use of multiple models may help to smooth individual biases, where the different models treat exposure and claims data differently.

Given the degree of uncertainty – at the overall exposure level and at the portfolio level – many brokers, (re)insurance companies have adopted a multiple model strategy in order to understand the significant variance in potential exposures. Indeed, it may be desirable to use a combination of actuarial models and cat models, particularly for specialised exposures or to assess high frequency exposures. In our experience of using current generation cat models, there can be a significant degree of overlap between what has been traditionally be considered as “attritional” losses (modelled using an actuarial experience based model) and what was regarded as “large losses” (modelled using catastrophe models). In reality, small or medium sized storms are included within cat models but results may not always align with trended experience resulting in further uncertainties for decision makers.

## 1.7 Use of a Single Unadjusted Model

In some ways the modelling process is easiest when only one of the vendor models is used, perhaps the model with the best global coverage or by selecting one individual vendor model for each region and peril. This would make the underwriting and the accumulation processes relatively straight forward. The documentation associated with the model is generally readily available. This is important as we enter a Solvency II world, provided the company is comfortable the use of a single model is a reasonable position to take. A certain level of due-diligence needs to be applied to satisfy both underwriters and senior management that the use of a single model is reasonable.

In reality this is often too simplistic and represents a potentially significant operational risk. Senior management will be expecting underwriters to be best in class and take more information into account in the decision making process, including their own judgement. The issue of model error associated with using one single model needs also to be addressed. For companies materially vulnerable to perils where there are wide legitimate disagreements between modelling assumptions – a potential example is that of European Windstorms as we discussed in Section 1.6 – it can be difficult to justify using a single model source for assessing the risk.

## 1.8 Use of Multiple Models

In reality when pricing an individual risk, underwriters benefit from having many different views of potential outcomes on an individual risk. Underwriters can compare and contrast results and seek an explanation if the models are producing materially different results. When considering catastrophe risk, they have three key model sources, on top of, and as input into, their own underwriting judgement.

- The use of *external catastrophe models* provides a packaged view of the risk. In return for the costs incurred from licensing the models and from managing and investigating them, the underwriters can tap – to the extent that the modelling companies are open and transparent – into the thinking of teams of professionals and experts to help quantifying risk. The use of *more than one* external catastrophe model provides different such views of the same risk.

- Underwriters will want to take into account *the company's internal view* of a given region and peril. This can be based on a blend of multiple models, or very occasionally from an in-house catastrophe model. The model blend approach to internal view can be considered as a distillation of the company's own assessments of vendor models as mentioned in the bullet point above. Where the internal view is from an in-house model, the underwriters would have much more transparent access to modelling assumptions – provided, of course, the in-house model is maintained and supported appropriately by experts.

- Also, depending on the risk being priced, experience rating may also come into play. The results of the experience rating can be compared with the results of the exposure models.

Many aspects of vendor models are essentially black boxes. It is often difficult to review all the underlying assumptions and decide whether they are reasonable. Even when underlying assumptions could be examined and appropriate modifications on them calibrated, some parts of the model are easier to modify than others. For some of the models it may be possible to overlay different event frequencies associated with each event in the event set. However, if you believe there should be a different distribution of physical events in the overall event set or a different distribution in specific regions, adding these events in and keeping the coherence of the models intact can be extremely difficult.

One of the ways to solve some of the issues with vulnerability is to use blended model output. If you believe model A understates severity of loss while Model B overstates, it may be possible to take an average or a weighted average of the modelled expected losses from the two models as a proxy.

If available, underwriters are likely to want to take into account multiple models and perhaps also overlay either experience, such as at the lower end of a curve, or expert judgement (from either the underwriters themselves or from specific research) at the upper end. This creates difficulties in the blending process. The major difficulty is not necessarily pricing an individual risk but in ensuring a consistent approach is retained for every risk. The deployment of a consistent approach would enable model output to be accumulated across the entire portfolio.

There are other practical issues that come into play when blending models. The methodology and parameters used has to be decided upon and agreed. There are a number of different methods e.g. frequency blending and severity blending (please see Section 2). One of the key practical considerations is the weight to give each model. The internal R&D or actuarial team may advise on the blend. However this will need to be explained to, and signed off by, management. Coming up with a robust methodology for deriving the weights is important but also fraught with difficulties.

## 1.9 Comments on the Use of Multiple Models

To provide some context, it is useful to reference the views of the rating agencies. For many insurers, rating agencies are a significant consideration in defining risk appetite and capital requirement. To inform their assessment, rating agencies require information regarding catastrophe exposures to be submitted to them. The agencies have made various statements regarding the use of multiple models.

a) Standard & Poors ("S&P")

In the wake of widespread disaster-related losses for insurers and reinsurers, ratings firm S&P has reiterated its call for the use of multiple catastrophe models. In a press release, S&P said it preferred use of models from at least two of the big three modelling firms when assessing the catastrophe risk for natural peril catastrophe bonds. S&P says its criteria still allows the use of a single model when assessing catastrophe risk, but contends that using multiple models would increase transparency in the market and lower the risk of "model shopping" where risk managers purposely select the model that gives them the most desirable results.

The risks being modelled are typically low frequency and high severity in nature. Loss estimates from an event, or the likelihood of an event, can differ significantly between modelling agencies, according to how the data is interpreted. S&P therefore consider that a multiple-model approach would give existing and potential investors a better perspective on the range of potential outcomes. While it would not eliminate uncertainty, it should provide a greater insight into the risk a deal presents, and to some extent, address the perceived issue of model shopping.[1]

This represents a positive view for using multiple models in assessing catastrophe risk. However, the following comment from another rating agency is more cautious in tone.

b) AM Best

When companies provide output from multiple catastrophe models, A.M. Best's baseline approach is to take the straight average. This, however, can be adjusted to a weighted average in cases where more refined information is available that supports greater reliance being placed on a given model. In either case, A.M. Best expects a company's management to be able to explain why it has utilized the output selected to best represent its catastrophe exposure.[2]

In our view, both the positive and more cautious views are important when using multiple models. Using multiple models and thoroughly researching the reasons for differences can give a (re)insurance company a clearer framework to evaluate and understand vendor models, leading to more informed usages of the models. It can also provide opportunities to reduce and deal with some (epistemic and implementation) uncertainty. The more cautious tone is relevant and points to the need for appropriate governance surrounding the use of multiple (or, indeed, *single*) catastrophe models for assessing catastrophe risk. It is important to appreciate the models we are blending with: blending two inadequate models or assumptions does not make an adequate model! The industry practitioner would appreciate the concept of model governance is a major part of the Solvency II regulatory regime.

These concepts are explored in various industry and actuarial papers (for example, see the ABI guide (Garnons-Williams & Zink, 2011)), and we aim to add to these topics further in this paper. Specifically, in Sections 1.10 and 4 we shall propose an approach to help monitor how catastrophe models are being used by underwriters, giving opportunities to have informed conversations about model usages and to prioritise in-house catastrophe modelling R&D and investigations. We shall also explore how different ways of blending (see Sections 2 and 3) can help deal with epistemic and implementation uncertainty.

Before we leave this section, it is worthwhile noting that, as we shall see in our discussions below – for example Section 2.2.1 – the phrase "straight average" can mean a variety of blending methods. These methods can yield very different results: *we recommend that such phrasing should be qualified with the blending methodology where appropriate*.

## 1.10 Quantitative Feedback Loops

The current processes of assessing cat risk can be based on a narrow perspective and lack critical elements. There is the potential to obtain a broader and potentially more dynamic perspective on risk. This has two dimensions – a quantitative and qualitative aspect. Whilst the use of catastrophe models provides an objective and independent perspective on exposures and is regarded as the best current means of assessing extreme loss potential, many limitations are recognised. These limitations are well known and related to data and assumptions. Major improvements in cat models are normally associated with either scientific or technical

[1] Source: Reuters

[2] Source: Best's Briefing, 10 March, 2011, "Catastrophe Models and the Rating Process FAQ"

progress (e.g. parallel computing advances) or lessons derived from analysis or observed data from significant actual events. However such analysis can often take several years to feed into updated models.

Alternative quantitative perspectives for cat risk can also be derived from more traditional actuarial experience based approaches. Particularly for high frequency perils such as tornado hail or smaller wind storm events, physical based models may not have sufficient resolution in data and parameters to accurately represent specific exposures. Thus a combination or fusion of actuarial based outputs and event based models may be useful particularly for portfolio management where event based outputs can be adjusted to reflect actuarial based results. Overlaying such adjustments to event based outputs then enables geographical correlation between accounts to be maintained for purposes of portfolio management e.g. aggregation of exposures.

A more qualitative dimension that can also be harnessed to address limitations relates to the underwriting process. An underwriter, particularly a highly experienced underwriter, will also have gained significant insight into hazard, exposures and trends. Elements of risk cover human dimensions that have variable response to hazards, with a high level of uncertainty. The variance in cat losses not only include elements such as the degree of hazard and level of vulnerability but also elements such insurance practice (claims settlement and handling), adherence to building codes, building stock preparedness (e.g. building maintenance levels may drop without frequent occurrence of a hazard). Good underwriting practice will attempt to weigh up such softer factors. Insights into recent losses, insurance practice etc. may lead underwriters to question the results of models and are worth capturing, particularly given relatively long-lead times for model update cycles. Capturing such opinion based perspectives can help identify and particularly prioritise regions, issues or types of business (e.g. specialised lines) where significant divergence exists. Once the extent of any divergence is revealed, this then allows risk managers to assess the validity of this view and to determine which elements can be modified to improve existing models via structured adjustments or if further research is required to be undertaken.

A more traditional route of discussion and review can achieve such feedback loops but lacks the full benefit of a “wisdom of crowds” approach as this can be selective, lacking a degree of significance than the actual “voting” of opinion can achieve. This is possible within the implementation as a separate underwriter view of risk, with underwriters given flexibility to adjust raw outputs in a number of dimensions beyond simple linear scaling of results. This aspect is discussed further in the example technical solution of Section 3.

# 2 Catastrophe Model Blending

The interest of Section 3 is to provide a detailed illustration of an example technical solution for blending catastrophe models. As we will see in Section 2.3, the existing literature already alludes to high level ideas. Our example technical solution will give us an opportunity to explore some of these ideas in more detail, from both the theoretical and practical points of view. Before we dive into the example technical solution, we discuss catastrophe model blending in general in this section.

A natural interpretation of catastrophe model blending is the blending of the model's mathematical formulation as well as the model's formulas. This interpretation sees model blending as adopting "the best bits" of each component model, to produce a brand new model. One may take the hazard section of model A, vulnerability section of model B and financial section of model C, for instance, to produce a new "blended" model M. This is what (Guy Carpenter, 2011) terms as *Model Fusion*, and we would imagine, if successfully implemented, this to have deep potential for superior modelling that has high transparency and a much better handle on (epistemic) uncertainty.

This is a big *if*. The broker's report cites licensing obstacles. Further, to be able to select in such granularity which model is *best* for which part, one must also invest heavily on data and expertise for validation and verification, as well as for the overall governance of the blending process. The costs can be prohibitive for most firms, although with technological progress and standardisation of vendor models, this may be achievable in the future (see Section 5.3).

This paper, therefore, interprets model blending as the blending of outputs, which are more readily available. This is not uncommon and can be helpful in some circumstances: for example, (Woolstenhulme & Major, 2011) found it useful to identify the model outputs with the model themselves in their discussion of model risk. Granted, maintaining or requesting outputs of runs from different models can be a challenge. However, with the necessary IT support to respond to this challenge, and until new generations of more flexible catastrophe modelling platforms become available, blending of model outputs is a much more practical approach to blending models, even if it falls short of the ideal of model fusion.

From now on in this paper, model blending refers to the blending of outputs of models.

## 2.1 Two Natural Ways of Blending

There are two natural ways to blend model outputs: the severity approach and the frequency approach. We shall now describe them and discuss their merits, concluding that the frequency approach is superior, as it has better ability to perform blending more consistently between pricing and accumulation, as well as between gross and net of excess of loss reinsurance.

### 2.1.1 Severity blending

The severity approach takes the weighted average of the component model outputs at each exceedance probability (EP). If model A gives $X for the 1% EP (i.e. 1 in 100 year), and model B gives $Y at the same EP, then the 50-50 severity blended 1% EP number is 50%x$X + 50%x$Y.

The calculation is simple to carry out and is intuitive for simple queries. If necessary, different blending weights could be easily implemented for different return periods.

#### 2.1.1.1 Difficulties with event sets

There is not an easy way to construct an additively coherent catastrophe model consistent with severity blending – this makes it difficult to have pricing and accumulations performed on the same bases. Recall that the rank-ordering of events is a key part of calculating the exceedance probabilities (see, for example, Slide 20 of (CAS, 2010)). Effectively, each result – be it gross or net, by risk, by class or classes in aggregate – requires a different ordering of model outputs, with apparently little to link between them.

In the same way, but on a larger framework, severity blending is difficult to implement in a Monte Carlo simulation driven stochastic models in the usual way. By usual here, we mean the simulation from a predefined event set, looking up the loss cost parameters from the various portfolios in the Event Loss Table conditional on the simulated event, and possibly simulating secondary uncertainty losses. One may declare model A's event set to be used in such a setup, with loss severity adjustment weights for each event in the set, calibrated by severity blending model A with B on a reference portfolio. However, there is no guarantee that the adjustment weights would produce the same severity blended results from portfolios that are different from the reference portfolio. Worse still, when selecting model A's event set, the aggregation of two severity blended catastrophe simulations would be more weighted towards model A, through the exclusive use of model A's view of clashes between the two portfolios. We shall say more about how event sets define dependencies between portfolios in Section 3.3.4.1.

### 2.1.1.2 Narratives and consequences

Yet another way of considering this is from a narrative point of view. Narratives are relatively difficult for severity blending. An example might be: "at the 1 in 100 year return period, we believe reality is a 50-50 average between models A and B". A more refined narrative would be that:

- given model A is higher than model B at the 100 year return period,
- and that A and B are both judged as "good" models through a process like that described in Section 4.1
- then we have good reasons to believe that
    - there is greater chance for the "true" 1 in 100 figure to be lying below A than lying above A
    - there is greater chance for the "true" 1 in 100 figure to be lying above B than lying below B
- so by taking a (weighted) average of the figures from A and B, we would have an answer that is "less wrong" than if we used pure model A or pure model B alone

Closer scrutiny suggests difficulties in explaining why the "1 in 100" event from model A should be matched with the "1 in 100" event from model B. This can be especially difficult when considering a large territory (e.g. US or European windstorms). The "1 in 100" event from model A could be coupled with several different "1 in 100" events from model B when we consider different portfolios. Moreover, it is plausible that after application of outwards reinsurance on each of the component model outputs, the coupling of events gross and net can be different.

It is also interesting that the above (more refined) narrative does not deal with *epistemic and implementation uncertainty* head on: this is in contrast to frequency blending – see in particular, Sections 2.1.2.3 and 3.2.7.5 – where the uncertainty in the choice between the two (or more) models are factored in, and some reliance on idiosyncratic drivers of uncertainty is reduced.

Rather, under this narrative, severity blending attempts to manage the consequences of getting the estimate wrong. This is not an unreasonable motive, and one intuitively feels that there would be ways to further consider and estimate extreme losses with this type of concept. We have not seen any studies on this topic in the context of catastrophe exposure estimation, but we would expect further research in this area to help answer questions such as:

- whether consequences should be symmetrically defined: intuitively, one would imagine the "pain" of (a) overestimating the extreme tail and therefore capital requirements, leading to the company being less competitive, would be different from the "pain" of (b) underestimating the extreme tail – and capital requirements, leading to taking on too much risk for the business and/or holding too little capital, increasing the chances of costly capital raising activities or potentially the failure of the company if capital cannot be raised after a significant event.

- what kind of consequence function does a "50-50" severity blend imply?

- how can we quantify "pain" to the concept of "consequence"?

- how does such consequence analyses help to improve on weighting of model?

- how does the highly unlikely crystallisation of the losses affect the impact of consequences on the estimation of return period losses?

#### 2.1.1.3 Possibility of catastrophe modelling without consistent event sets

It is possible to think of model outputs without the use of simulated events, and thus by-passing the above narrative difficulties. The frequency-severity approach to modelling non-catastrophe large losses in stochastic models rarely labels each loss with an event narrative. Practitioners sometimes impose copulas between aggregate large loss distributions to mimic large clashing events that impact multiple classes simultaneously. A similar framework can be possible when modelling severity blended catastrophe losses. The event sets (be it from A or from B) would no longer be consistently simulated in the stochastic model for the different classes. Instead, each simulated loss from the different portfolios would have a selected copula imposed on them.

We shall leave the exploration of this technique to another paper: indeed, Slides 22 and 24 of (Cook, 2011) has begun to do so. Such exploration will need to consider challenges faced by not having consistently simulated events in the same model, from model interpretations and communication of outputs, to technicalities such as defining the copulas and making sure such copulas are consistent in the various ways the portfolios are cut and diced.

### 2.1.2 Frequency blending

The frequency approach takes the weighted average of the same outputs at each loss severity. If model A gives \$100m at the 1.5% EP (66.7 years) and model B gives \$100m at the 0.5% EP (200 years), then the 50-50 frequency blended return period for \$100m would be 1.0% EP = 50% x (1.5%) + 50% x (0.5%).

#### 2.1.2.1 Difficulties with back-of-envelope calculations

Given modelled losses from component models on fixed return periods, frequency blending does not automatically give loss amounts for the same return periods. These back-of-envelope calculations available to severity blending are not available to frequency blending. This could cause an additional layer of complexities when performing analyses of change, for example, between successive accumulations. However, once a company decides to blend models A and B together, there is no reason why management itself should work with outputs from A or from B. One would imagine that they should work with the blended figures. Of course, they would still like to review the component model outputs when reviewing the blend (say, after a comparison exercise as discussed in Section 4.1.1).

Secondly, the use of exceedance probabilities can be unusual if a company is used to return periods and may pose slight communication hindrances between modellers. However, we note that with modern calculators and spreadsheets that have reciprocal functionalities, such changes should not be difficult. Once again, it is more the modellers and practitioners who are facing this issue: users of the outputs of the blended model are unlikely to need to be aware of this in day-to-day work.

However, there may be occasions when we focus on levels of losses – and it is the probabilities that we are trying to estimate. For example, management may want to know the likelihood of a catastrophe loss greater than a threshold – where the threshold may be a proportion of, say, the net asset values on the balance sheet. Some stress and scenario tests primarily focus on severity, and leave the probabilities as an “after thought” (e.g. if Windstorm Lothar were to hit again, what losses would the company endure – and where on the EP curve does this sit now given significant model revisions).

#### 2.1.2.2 A proper catastrophe model

The strength of frequency blending is the ability to coherently work with simulated events, through implementation recipes such as the one described in Slide 14 of (Cook, 2011). (We shall give a demonstration in Section 3.2.5.2 that this recipe does indeed give the frequency blend as defined above.) Events from the component models will not need to be coupled for averaging. Instead, frequency blending works naturally in a Monte Carlo simulation framework, with the correct proportions of events (or years) being simulated from the component models, according to the prescribed weights.

In this way, as suggested in (Cook, 2011), frequency blending gives us a proper catastrophe model, that has its own event set and associated portfolio losses. This gives immense implementation advantage over severity blending for (re)insurers (please compare with Section 2.1.1.1).

#### 2.1.2.3 Narratives and full conditional confidence

Narratives are useful for further discussions around blending, to gauge the aptness of the methodology. Like severity blending, frequency blending also allows narratives for communication: "we believe reality is reflected by A 50% of the time and by B the other 50% of the time" is such an example for the 50-50 frequency blend.

From this narrative, we see a feature of the frequency blending approach is that when a component model A is simulated, it assumes that A would model reality well enough. It is as if we have full confidence in model A. The flip side of this is that under this particular simulation, component model B has no voice – as if we have zero confidence in it.

In reality, both models A and B have deficiencies, primarily due to epistemic and implementation uncertainty – (Guy Carpenter, 2011) lists some of them. Some deficiencies are common to both models, and some of them are idiosyncratic to the individual models. The deficiencies distort results to various levels of materiality. As model evaluation processes may reveal (see Section 4.1.1), some of these deficiencies are known and some of them are only known to a limited extent. Some of the known ones might be allowed for already through adjusting the component model. Conceptually, frequency blending is helpful to reduce vulnerability to specific, material, idiosyncratic, deficiencies that – for whatever reason – have not been allowed for: *but* it does not get rid of them totally.

#### 2.1.2.4 Consistency with expected value blending

As discussed in Section 1.3, a key component in pricing a contract is the contract's expected loss cost. In catastrophe modelling, the contribution to the contract's expected loss cost from the modelled peril is represented by the "AAL" (average annual loss). There would be an AAL from each catastrophe model.

It is usual practice for weighted averages of the contract AALs to be taken during pricing. Frequency blending of the pre-contract losses is theoretically consistent with this approach, while severity blending gives different answers when challenged with even the most usual of nonlinear contractual terms such as excesses and deductibles. A demonstration of this is in Section 3.2.5.2, with a particular implementation of frequency blending. This consistency helps to define a system in which pricing and accumulations are performed on similar bases, giving frequency blending an advantage over severity blending.

In the spirit of Section 2.1.1.2, we can also consider frequency blending as a means of reducing the negative consequences of getting our results – here we are considering the AALs, not specific points on the OEP curves – wrong through epistemic and implementation uncertainty.

### 2.1.3 Comparisons between frequency and severity blending

On balance, we see frequency blending as a more desirable way forward for consistent implementation throughout an organisation. Since a key strength of severity blending is the ease of calculating blended figures at individual return periods, it is worthwhile considering how the two blends can differ on these figures – we shall conclude that there is no *a priori* reason why the two blends should give similar results, although one can discern some patterns that could help explain the differences.

A stylised OEP chart depicting the pure model outputs and the two blends can be seen below (Figure 2-1). The **blue** and **red** graphs give the two pure model outputs. The **purple** graph is from frequency blending and the **green** graph from severity blending. The pure model curves are produced by very simple assumptions to let us see how the blends behave. The blends are both 50-50.

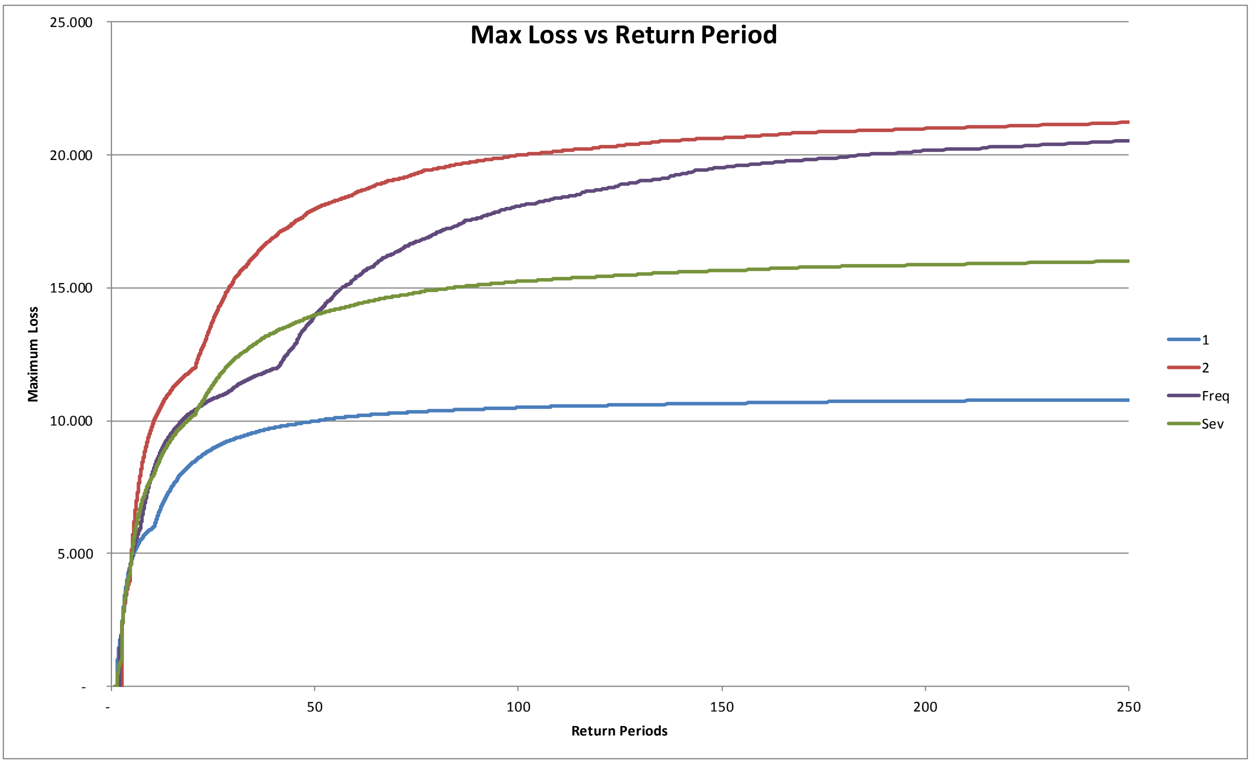



**Figure 2-1 Stylised OEP graphs demonstrating how frequency and severity blends may behave**

The severity blend is relatively easy to see. At each return period (i.e. at each vertical cross-section of the chart), the blend sits half-way between the two pure models, as suggested in Section 2.1.1.

However, the frequency blend is not so easily seen under this graph. The description in Section 2.1.2 demands a different graph. Specifically, we require the exceedance probabilities against the monetary amounts, as seen in Figure 2-2 below. The colour coding in Figure 2-2 is consistent with that in Figure 2-1. The straight lines are not typical – but this example is specifically constructed for demonstration purposes.

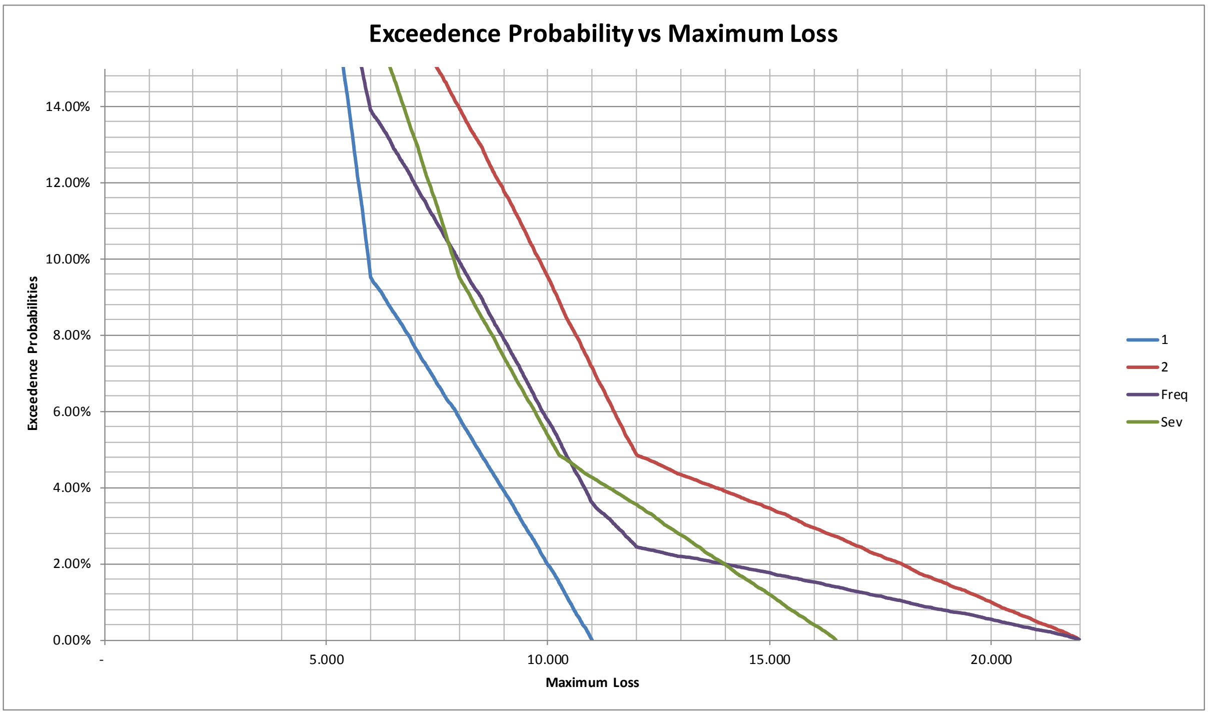


**Figure 2-2 Stylised OEP curves, as exceedance probabilities against losses**

A few observations can be made:

- The frequency blend curve is half way between the pure model curves for each vertical slice – this is in line with the discussions in Section 2.1.2.
- The less severe pure model curve is Model 1 (the blue curve). Frequency blending is influenced by this curve earlier in the exceedance probability dimension (and so, in the return period dimension). For example, there is a change in the model 1 curve at the 9.5% (10.5 years) point. This impacts the frequency blend curve as early as 14% (7.1 years).
- On the other hand, the more severe pure model curve is Model 2 (the red curve). Frequency blending is influenced by this curve later. For example, there is a change in the model 2 curve at the 4.8% point (20.8 years). Frequency blending is impacted at the 2.4% point (41.7 years).
- Of interest is that severity blend curve never reaches the extremes of the more severe pure model curve (Model 2, the red curve). The frequency blend curve (the purple one) attains this extreme: naturally, at half the likelihood versus the red pure model curve in our example.

The above is observed in many real situations. It means that we should not expect frequency and severity blends to give similar curves. The following is an example of the frequency blend graph changing direction sooner than the severity blend graph.

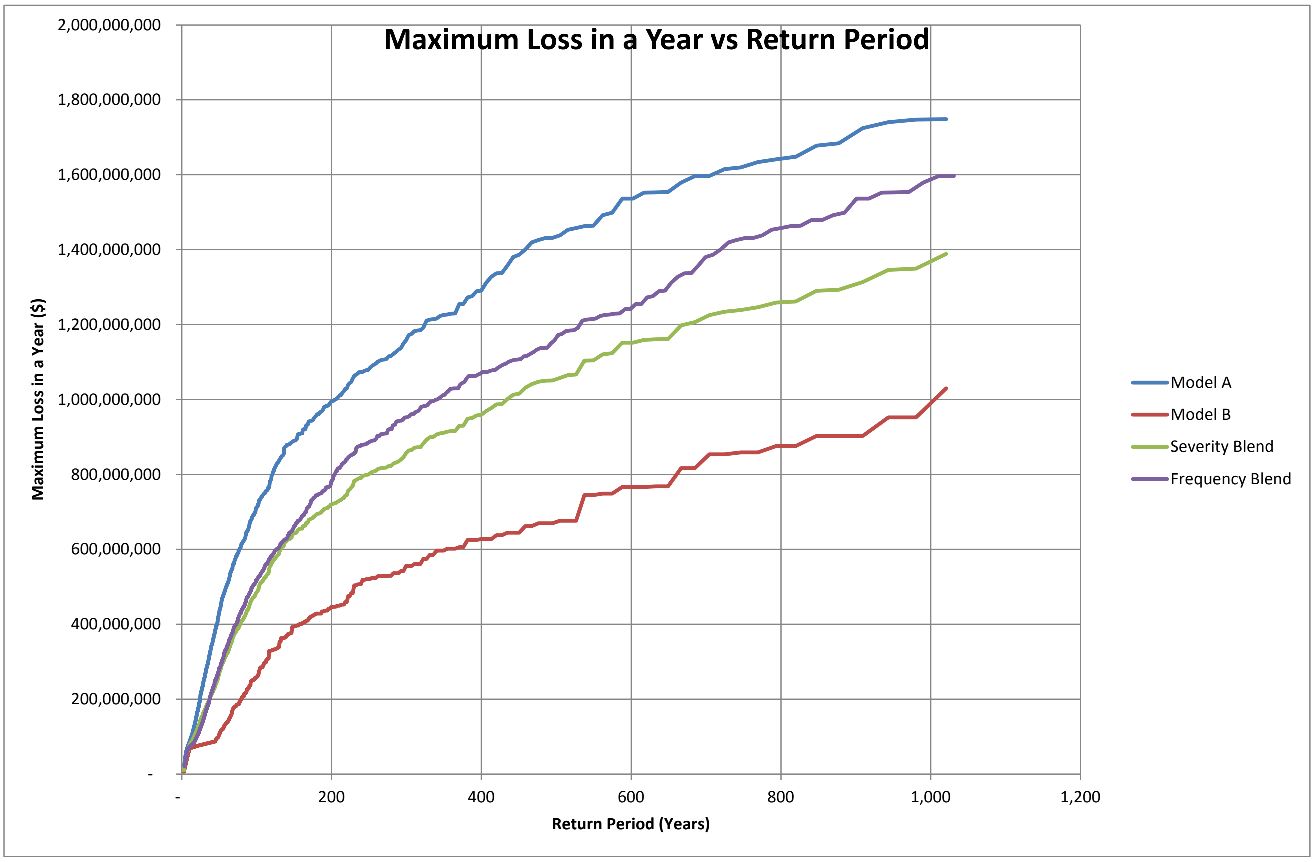



**Figure 2-3 Example OEP curves: frequency blending more severe due to sharp changes in direction in the less severe pure curve**

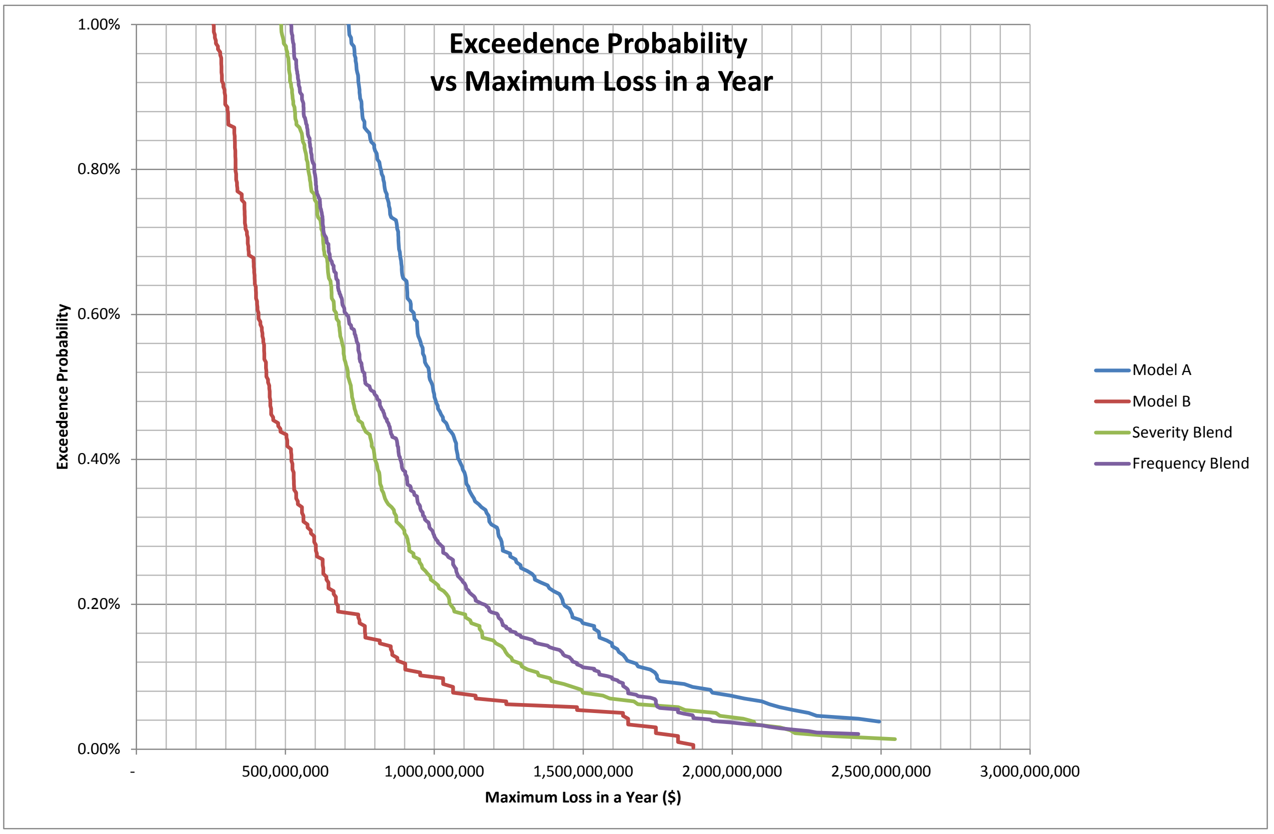



**Figure 2-4 The OEP curves as exceedance probabilities vs loss**

The following is relatively rare, where we have the frequency and severity blends being close to one another.

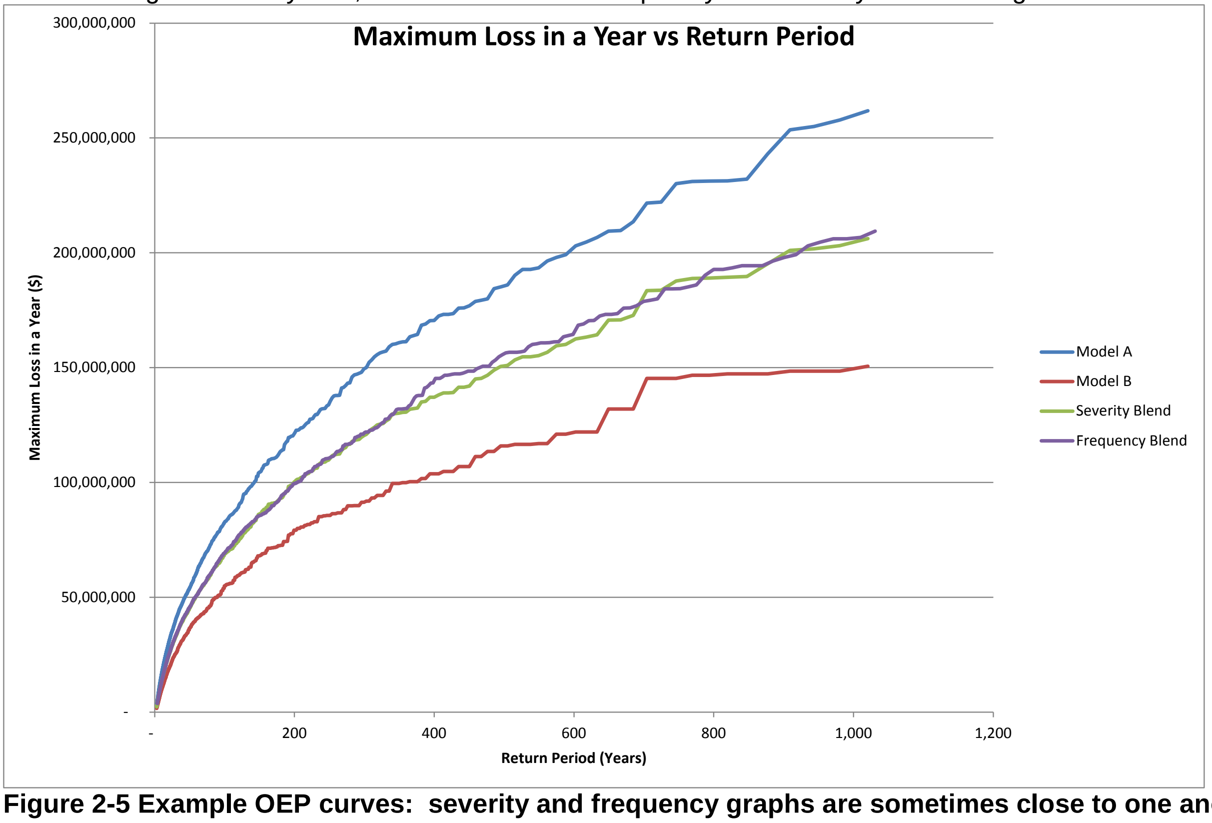



**Figure 2-5 Example OEP curves: severity and frequency graphs are sometimes close to one another**

When there are cross overs, the frequency blend graph is typically lower than the severity blend graph, due to the gradients of the pure model curves.

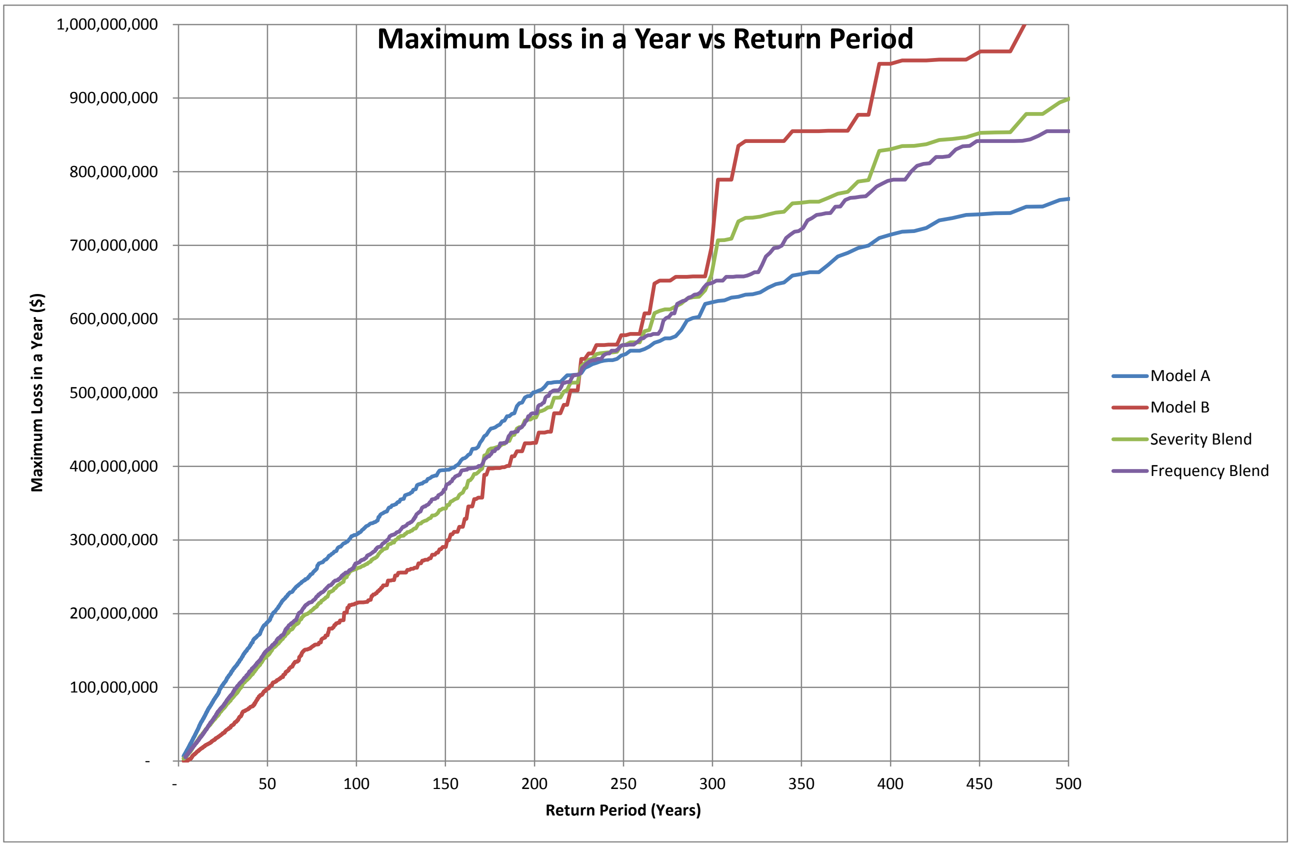



**Figure 2-6 Example OEP curves: the pure models cross**

## 2.2 Arithmetic and Geometric Averages

As the (Guy Carpenter, 2011) report alludes to, there are two algebraic ways to performing the averaging calculations: arithmetic averaging and geometric averaging. The paper (Major, 2011) gives more details. We followed (Cook, 2011) in the discussions above, focussing on arithmetic averaging.

Geometric averaging involves blending multiplicatively. In severity blending, we would have the 50-50 blended loss number as $\$X^{50\%} . \$Y^{50\%}$. In frequency blending, we would have the 50-50 blended EP number as $1.5\%^{50\%} . 0.5\%^{50\%}$. Geometric averaging can be thought of as arithmetic averaging in the log scale. Arithmetic averaging is the more easily explainable concept, although actuaries typically work with logarithms of loss amounts in statistical applications (e.g. the use of log link functions in performing GLMs).

### 2.2.1 What is a "50-50 blend"?

A message from this broad section – as seen above, and will be seen below – is that *the phrase "50-50 blend" can mean a wide variety of methods, with potentially widely differing results*. This can be important for users of blended outputs such as senior management, regulators and rating agencies to understand (see Section 1.9). Arithmetic and geometric averaging adds yet another complexity.

Due to a basic mathematical result (see for example (Wikipedia, 2012)), with both frequency and severity blending, geometric averaging gives lower results than arithmetic averaging. In severity blending, choosing geometric averaging could give a much less severe OEP curve. If severity blending with arithmetic averaging finds it difficult to reach the extremes of the more severe component model curves (see Section 2.1.3 above), then using geometric averaging would make this even harder – and could potentially be judged too optimistic for risk management purposes. This problem is much more acute with frequency blending: the (blind) use of geometric averaging on EPs could make the final blend only getting as far as the maximum simulated loss amount in the *least* severe component model curve. This is not desirable intuitively: a “50-50 blend” that never gets above the maximum of the least severe component curve has a lot to answer for! The following chart (Figure 2-7) adds the geometric blends to the stylised example of Figure 2-2 above, illustrating these points.

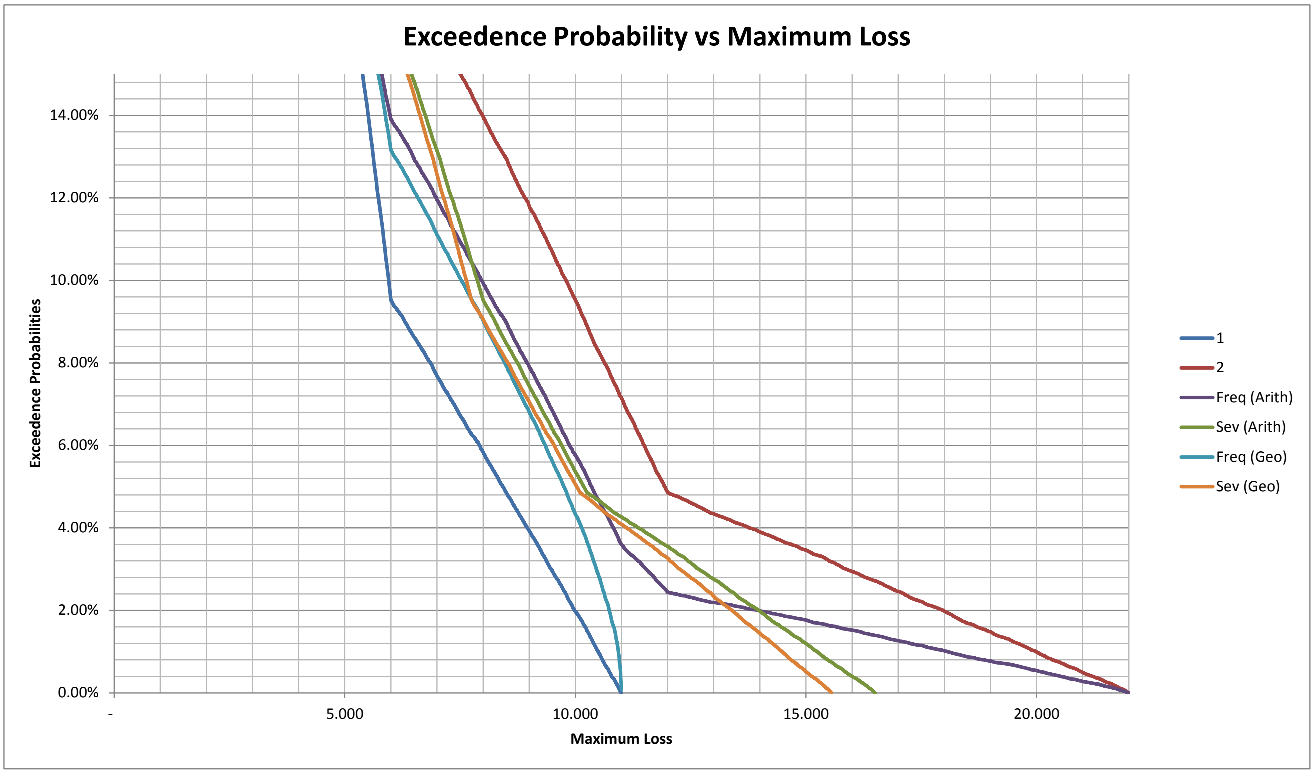



**Figure 2-7 Stylised Exceedance Probability vs Maximum Loss Example, comparing geometric and arithmetic blends**

## 2.2.2 Implementation difficulties

The availability of a coherent implementation recipe for frequency blending is a great strength. The recipes typically work on the assumption that we are performing arithmetic averaging – and no recipe seems to be available to work with geometric averaging. For simple purposes, it may possible to approximate the geometric averaging with a re-weighted arithmetic averaging: and the results can be very close. However, for applications with multiple risks and levels of aggregations, such approximations would hit a problem with regards to the selection of reference portfolios to calibrate the re-weights.

Moreover, in many applications, the blends are subject to mechanisms of insurance and reinsurance – outwards and inwards. These mechanisms, especially the presence of limits and deductibles, can make geometric averaging unstable. A blend on the gross numbers of 100 and 150 should net down additively after, say, a XL programme of effective 95 xs 5, become a blend on the net numbers of 5 and 55. Arithmetic averaging in the blends preserves the netting down relationship in many – although not all – occasions (here the gross 50-50 blend is 125, and the net 50-50 blend is 30, which is also intuitively 95 less than the gross blend). Geometric averaging will not do so (here the gross 50-50 blend is 122.5 and the net 50-50 blend is 16.5, which is 106 lower than the gross blend).

For these reasons, and for the sake of avoiding potentially optimistic blends, we will always assume arithmetic averages for the rest of the paper.

## 2.3 Literature on Model Blending

The blending of catastrophe model outputs is not a new phenomenon. A 2005 Florida Hurricane Catastrophe Fund (FHCF) report (Florida Hurricane Catastrophe Fund, 2005) cites the use of blending in its industry aggregate and more granular analyses. The report's Exhibit V (and Section E of Exhibit I) seems to suggest weighted averages of three or four model outputs were deployed to arrive at estimates of losses at different return periods – i.e. a severity blending approach. Section 4.5.1 of (Grossi & Kunreuther, 2005) illustrates severity blending, again in the context of understanding uncertainty arising from Florida hurricanes. The next section in the same book illustrates frequency blending when considering the uncertainty arising from South Carolina earthquakes.

The 2011 year received at least two public discussions of the use of multiple catastrophe models. In March 2011, Ian Cook presented the workshop (Cook, 2011) in London, and in December 2011, Guy Carpenter published (Guy Carpenter, 2011). Both consider the use of multiple models in the context of model uncertainty (we shall use the wider sense of the term "model uncertainty" in this section – see Section 1.3 for more detail). The blending of model outputs features in both as a way of using multiple models.

The (Cook, 2011) workshop has more practical implementation details such as illustrations of how the different blending methods work. Contrasting Severity Blending (the "Common Approach") and Frequency Blending (the "Alternative Approach"), it suggests that frequency blending has many advantages over severity blending: the listed advantages can probably be summed up by the availability of a coherent catastrophe model under the Frequency approach. It also suggests – not in as much detail – further blending possibilities, giving the example of decomposing the catastrophe models into components and blend separately. These ideas are further taken up by the ABI guide, (Garnons-Williams & Zink, 2011) – see, in particular, its Chapter 8 on using multiple models.

Cook's workshop goes on to discuss how one may practically choose model blending weights: suggesting a score-card approach to weigh up qualitative pros and cons of the models. The Solvency II section encourages the use of multiple models as a way to demonstrate understanding of external models. The section ends with the uncertainty surrounding whether catastrophe model changes should be deemed internal model changes.

Whereas (Cook, 2011) had an emphasis on methodology, (Guy Carpenter, 2011) had an emphasis on model uncertainty. The latter, having identified sources of model uncertainty, discusses how using multiple models could reduce the impacts from some such sources. Model blending would help deal with sources of uncertainty where the two models make different assumptions, or have different data sources. However, it identifies a key area where model blending would *not* help substantially with: this is the uncertainty arising from the two models making use of the same (limited) historic data. Data uncertainty associated with the common historic data would also be present in the blended model.

Sean Devlin's 2008 presentation (Devlin, 2008) joins (Cook, 2011) in pointing to the possibility of adjusting EP curves, by differing amounts at different return periods. Devlin's context is adjusting for seasonal forecasts. Cook's aim of "Shoehorning" is to replicate the results from one model under the event set of another model for accumulation purposes. The report (Guy Carpenter, 2011) suggests similar adjustments as part of "model morphing". Cook illustrates the concept most fully out of the three.

Finally, (Guy Carpenter, 2011) mentions model *fusions*, where a new model is constructed, made up of components from different vendor models. The literal implementation of model fusions is typically not possible due to licensing restrictions. However, the paper mentions that it may be possible, albeit very complicated, to perform multiple runs for every portfolio, which is spilt into different exposure classes.

Both (Cook, 2011) and (Guy Carpenter, 2011) very briefly mentioned the possibility of reducing the impact of major model changes when blending models.

# 3 A Technical Solution

## 3.1 Tables and Notations

Before moving on to discuss the example technical solution in detail, we define various data tables that we shall frequently reference. On the way of defining the tables, we also define notations that will help our technical discussion.

### 3.1.1 The Year Event Table

The Year Event Table (or YET) in its simplest form is a table with the following fields:

- SimulationYear (integers, duplicates allowed, running from 1 to *N*)
    - This labels the *n*th simulation
    - The number of simulations is *N*
- LossNumber (integers, duplicates allowed, running from 1 to a varying integer)
    - Each of the *N* simulations stand for a year's worth of catastrophe events. Each of the simulations can have a different number of catastrophe events from others.
    - When there is no catastrophe event for simulation *n*, then there will be no record available in the YET with SimulationYear = *n*.
    - When there are *k* catastrophe events for simulation *n*, then there will be *k* records available in the YET with SimulationYear = *n*. In this case, the LossNumber field for these records would run from 1 to *k*.
    - A combination of SimulationYear and LossNumber that appears in the YET can only appear once. Each record in the YET, then, stands for a simulated event.
- ComponentModel
    - For each simulated event, this field is used to identify which component model all the risk exposures should refer to for loss amounts.
    - In our example technical solution, we assume that each simulation year references the same component model. There are other possible ways – please see Section 3.2.7.1 below.
- EventID
    - For each simulated event, this field identifies the event in the event set of the model indicated by the ComponentModel field.
    - EventID and ComponentModel together define the physical event. Note that it is possible to have the same physical event appearing more than once in the whole YET.
- SUPercentile (real number between 0 and 1)
    - Upon simulation of an event, the losses are simulated through looking up a loss distribution conditional upon the event. The SUPercentile helps to look up the percentile of that distribution.
    - For example, a value of 0.75 for an event would always look up the $75^{th}$ percentile from the loss distributions associated with the event, for all the portfolios.

A key feature of the YET is what is *missing*. Unlike the other tables, the YET does not contain any loss amounts. It *does* contain enough information for us to go and look up the event sets of the component models and the losses of the portfolios in question through Event Loss Tables (ELTs).

However, as with Year Loss Tables (see Section 3.1.3), and *unlike* the ELTs (see Section 3.1.2), the YET has the ability to convey event occurrence dependencies. It can reflect clustering: the increased likelihood of occurrence of some events given the occurrence of a specific event. Even when the independence assumption is used, it is explicitly and positively given in the YET.

Some YETs have other technical fields such as a *Day* field, which gives the day in the year when the catastrophe event occurs. As many (re)insurance contracts are annual in duration, we leave the discussion of the use of this field outside the scope of this example technical solution.

### 3.1.2 Event Loss Tables

Event Loss Tables (ELTs) are possibly the best known tables to actuaries working with catastrophe model outputs (see for example, (Cook, 2011) and (Diers, 2008)). Each portfolio would have an ELT associated with it for each run of the component model. A basic requirement of catastrophe model management systems is to keep track of the different ELTs associated with different analyses and model runs for the portfolios of risk, and to flag the final and committed ELTs. Such systems would need to be more sophisticated when dealing with potentially twice (or more times) as many ELTs from runs of different component models.

We now very briefly describe ELTs in this section. Catastrophe model outputs attempt to produce a representative list of all possible events: each record in the ELT represents a possible event. At their most basic, an ELT would have three fields.

- EventID
  - The event IDs would be drawn from the overall event set for that peril from the component model. Each of them stands for a particular physical event.
  - Accumulations can be estimated by summing losses from different ELTs with the same Event ID (please see Section 3.1.2.1 below).
  - The Event IDs are the same as those in the YET (mentioned in Section 3.1.1) – to allow look up of losses to generate Year Loss Tables (we shall discuss this process in Section 3.1.3).
- Rate
  - For each event in the ELT, the rate is typically a very small figure and indicates the likelihood of the event occurring in a year.
  - The likelihood could be absolute or relative. Absolute likelihood refers to the actual numerical figure in this field being used direct as probabilities or parameters of distributions. Relative likelihood refers to when the rates indicate the likelihoods of events relative to one another: e.g. event *i* being twice as likely to occur versus event *j*.
  - If the events are assumed to independently occur in any given year, there are two ways to interpret the absolute rates: one as the probability in the Bernoulli distribution, and the other as the mean in a Poisson distribution. Since the rates are usually very small, the two interpretations should give similar occurrence results. However, theoretically, we prefer to use the Poisson assumption, as it better reflects the fact that the event sets contains *representatives* of all possible events. As representatives, there would be non zero probabilities (even if extremely small) of the same event ID occurring more than once in a year.
  - If the events are not independent, a negative binomial is the most likely distribution to use.

  - Please see Section 3.2.3.1 for more on examples of how rates are used.

- LossAmount

  - From the vulnerability and financial modules, the component model would output loss information for each event in the ELT.

  - The simplest model output would give a single loss amount.

  - It is often useful to know the currency the loss information is based on. This can either be as a separate field in the ELT or can be looked up from another table in the model management system.

Event losses have associated uncertainty, which can be represented by distributions. This uncertainty is known as *secondary uncertainty* (primary uncertainty arises from the chance of an event occurring). The more sophisticated ELTs would have parameters for such distributions. Popular ones are: mean, standard deviation, and maximum loss. For accumulation purposes, one may have further information with regards to how losses from different portfolios diversify against one another – please see Section 3.1.2.1 below. However, typically, one would not have the precise distribution definition in the ELTs: the losses can be driven by a wide variety of factors, making distribution definitions impractical in an ELT. Flexible families of distributions (e.g. the beta distribution to model the losses as proportions of the maximum losses) are often deployed.

Finally Event Loss Tables are silent on how the events are linked with one another. They do not indicate how the occurrence of one event might make the others (or even which others) more likely.

### 3.1.2.1 Accumulation of multiple ELTs

It is straightforward to accumulate the loss information from ELTs for individual risks or subportfolios to form an ELT for a larger portfolio of risks. The EventIDs from the different ELTs would be matched, and the following operations would be performed to obtain the fields of the aggregated ELT:

- Rates should be the same for all the individual ELTs against the same EventID. The accumulated ELT should take this common rate as its rate. If the rates are different, the reasons for this need to be investigated before rates could be assigned.

- LossAmounts / Means and MaximumLosses would simply be the sum of all the LossAmounts / Means and MaximumLosses, respectively, from the different ELTs at the same EventID.

- The StandardDeviations would depend upon how the risks or subportfolios diversify against one another. Some component model outputs have an associated recipe to help the users to do this. When no such recipe exists, one can add the StandardDeviations together, assuming that there is no diversification. This can be far too conservative, depending on the granularity of the risks and subportfolios in question. An alternative would be to reference outputs of other models to inform diversification assumptions.

### 3.1.2.2 Reduced ELTs

The *reduced ELTs* are ELTs, but without the Rate field. For some purposes, it is not necessary to have the full ELT: but just the ability for losses to be looked up from EventID values – and the rates are not relevant. We make this definition to help clarify concepts in this paper, rather than seeing these as separate physical tables in an implementation.

### 3.1.3 Year Loss Tables

Some component models give Year Loss Tables (YLTs) as standard: such tables are becoming more common, especially for perils that have clustering events. One of their strengths is their ability to reflect how events are linked with one another, in a way that ELTs cannot. Please see Section 3.1.2 for more on ELTs.

Similar to the ELTs, each model run for a portfolio of risks would have an associated YLT. The YLTs would also need to be managed properly by the internal catastrophe model management systems.

The YLTs are structured like the YET (see Section 3.1.1): they can be considered as a set of Monte Carlo simulations.

- SimulationYear
    - The concept here is the same as for the YET.
- LossNumber
    - The concept here is the same as for the YET.
- LossAmount
    - This is the portfolio loss associated with the event – similar to that of that for the ELTs.

While it is possible to have loss distributional information for each simulated event, as for the ELTs, it is not common to do so. Instead, the loss amounts in the YLTs would already be simulated from such distributions. Indeed, by joining on the EventID fields, the YET and a (reduced) ELT (see Section 3.1.2.2) would produce a YLT:

| Year Loss Table Fields | Source |
|---|---|
| SimulationYear | The SimulationYear field of the YET |
| LossNumber | The LossNumber field of the YET |
| LossAmount | Either “with Secondary Uncertainty”, which would be to take the SUPercentile to look up the loss distribution as implied by the loss parameters;<br><br>Or “without Secondary Uncertainty”, which would be to take the LossAmount field of the ELT (which would usually be – although not necessarily – the mean of the loss distribution) |

This mechanism is an important part of the example technical solution (see Section 3.2.5).

It is often useful to also have an EventID field for the YLT: each value in this field is unique to each combination of the SimulationYear and LossNumber fields. The map between EventID and SimulationYear-LossNumber combinations should be the same in each YLT to help accumulations and event lookups.

#### 3.1.3.1 Accumulation of multiple YLTs

As for accumulating ELTs (see Section 3.1.2.1), the YLTs of risks and subportfolios can be easily accumulated to form a YLT of a larger portfolio. Instead of matching EventIDs, we would be matching SimulationYear and LossNumber simultaneously. LossAmounts are simply summed together on the same SimulationYear and LossNumber.

#### 3.1.3.2 Reduced YLTs

The *reduced YLTs* are like YLTs, without the SimulationYear and LossNumber fields, but retaining the EventID field. This is a similar concept to the reduced ELTs (see Section 3.1.2.2): for some purposes, it is not necessary to have the full YLT: but just the ability for losses to be looked up from EventID values. Again, we make this definition to help clarify concepts in this paper, rather than seeing these as separate physical tables in an implementation.

## 3.2 Standard Agreed Blend

The first step of the example solution is to produce the *Standard Agreed Blend*. This can be considered as a top-down approach to blending, where the weights are pre-selected, and all risks in the accumulation are treated in the same manner and subject to these weights.

A frequency blending approach is employed in our example, based on simulations of years of catastrophe events. Following (Cook, 2011), pre-defined proportions of the simulated years will follow the component models, according to the pre-selected weights. All risks will follow these same pre-simulated years, so that accumulations could be done in a consistent basis.

The first piece of inputs is the selection of weights.

### 3.2.1 Selection of weights

The selection of weights is an art more than science. Both (Cook, 2011) and (Major, 2011) list possible criteria for this: (Cook, 2011) suggesting a scoring approach, qualitatively assessing the models; (Major, 2011) begins to place this problem on a more quantitative basis. We will not repeat the content of these sources in this paper, referring the reader to them. Together with Sections 4.1 and 4.2, we aim to complement these discussions by sharing our own experience in weight selection.

Selection of weights is difficult, and outcomes could be sensitive to this selection. Certain weight combinations might be easier to justify than others.

- Zero weights: there can be strong clear reasons for not using a component model in a blend. For example, putting an obviously “poor” model into the blend could make the blended model worse. Even if there are some good “bits” in the model, they may not be beneficial enough for the organisation to invest in the model and/or spend operational and management resources keeping track of an extra model for the peril.

- Equal weights: putting equal weights on the component models can be a good option (e.g. 50% on each of the two components) when: we believe the models are equally good; or we do not see any strong reasons as to why we should prefer one over the others.

Other weight combinations are a bit more difficult to discuss. There may be reasons to overweight on model A and underweight on model B (e.g. if A has been updated more recently, or if A has access to much better historical data, or if B is less transparent in their methodology, etc.), but currently, there does not seem to be a mechanical way of calculating the extent of the overweight.

Initially, one may be able to devise discrete levels of weights (e.g. constraining one’s decision on whether they should be 30-70, 50-50 or 70-30). An interesting feature of the example technical solution is that there is a potential to let the bottom-up adjustments (see Section 3.3) trigger further R&D work, leading possibly to further refinements of weights. We shall discuss this in more detail in Section 4.2.2.

Another angle of deciding weights may be to reference the narratives. The severity blending narrative in Section 2.1.1.1 suggests that we can put in different weights for different return periods. The frequency blending narrative in Section 2.1.2.3 suggests we should consider the relative chances of the material idiosyncratic deficiencies in the respective models crystallising seriously distorting results.

Finally, we shall propose as an area of further research the possibility of updating weights post events in Section 3.2.7.5.

### 3.2.2 Deciding on the number of simulations, *N*

We now decide on the overall number of simulations we will be producing (i.e. the value of *N* in Section 3.1.1). This is a relatively basic decision to make: the following considerations would help towards it:

- The higher the number of simulations, the less **simulation error** there would be in our blend. In its most basic form, simulation error from Monte Carlo simulations is inversely proportional to $\sqrt{N}$. Simple binomial procedures (see Section 3.2.2.1 below) can help inform this decision.

- The higher the number of simulations, the more **computing** infrastructure and power would be required to implement the blend. One might intuitively expect computing time and storage to be in the order of *N*, although there might be discrete cliffs with respect to *N* where significantly extra costs would be incurred. Advice from one's IT department could usefully be sought.

- **Perspective** is important. One could conceivably tolerate a relatively small level of simulation error when faced with much larger issues such as model uncertainty and issues to do with lack of historical data (see, for example, (Miller, 1999)).

- In some circumstances, **other internal parties** would be interested in the blended model outputs. A Monte Carlo stochastic internal capital model would be a prime example. The number *N* could be set with this in mind. For example, if *N* is set to be the same as the number of simulations in other models down the line, then the process of feeding outputs into these other models would be simpler.

- Similarly, *N* could be set with reference to the how the **component models** output their results. If the component models are themselves Monte Carlo simulation outputs (e.g. in the form of YLTs), then, again, for ease of processing, one may want to set *N* with this in mind. For example, if component model B outputs 10,000 simulation years as standard, and we want to 50-50 blend it with model A, then setting *N* to be 20,000 could be an option. Here, the 20,000 simulations would consist of 10,000 from model A and 10,000 from model B.

The last two points are not critical. As we will see in Section 3.2.3.1.5, there are natural ways to stratified sample large simulation sets down. It is also not impossible to cosmetically increase the number of simulation years (see Section 3.2.3.2.1).

#### 3.2.2.1 Simple binomial procedures for understanding simulation error

Simulation error could be assessed by the use of the binomial distributions. A key strength to Monte Carlo simulation is that tail probabilities from a distribution, $p = P(X>x)$, can be approximated by $\pi$ = #{simulations of $X$ that are greater than $x$} / $N$. Furthermore, good Monte Carlo simulation engines would produce draws from the distribution that look statistically independent – and a reasonable practical assumption to make is that the draws are independently made.

Under these assumptions, the number of simulations $N.\pi$ follows the binomial distribution with parameters $N$ and $p$. With these, confidence intervals can be obtained to give indications of simulation error (one can, for example, use the BINOM.INV function in Excel 2010).

Therefore, if we want to assess the simulation error around the 100 year return period, then we could set $p$ to be 1% and test for various levels of $N$. When $N$ = 10,000, the symmetric 95% confidence interval for $N.\pi$ would be (81, 120), from which we get the 95% confidence interval for $\pi$ to be (0.81%, 1.20%). This translates to (83.3, 123.5) in return periods in units of years. If $N$ is 50,000, then the confidence interval for the return period would be narrowed to (91.9, 109.4).

As the 200 year return period is particularly interesting from the Solvency II's point of view, we include the confidence intervals for various *N*'s here.

| p = 0.5% | | |
|---|---|---|
| N | 95% Confidence Intervals of Return Periods (Years) | |
| 5,000 | 142.9 | 312.5 |
| 10,000 | 156.3 | 270.3 |
| 20,000 | 166.7 | 246.9 |
| 50,000 | 177.9 | 227.3 |
| 100,000 | 183.8 | 218.8 |
| 500,000 | 192.5 | 208.1 |

**Table 3-1 Simulation error return period confidence intervals when *p* = 0.5%**

The procedure is easy to implement and can give us guidance of the amount of simulation error we are involved with. However, it does not give us simulation error of *means*, but just the simulation error of *probabilities* – and this is important when we are deriving expected loss costs or using the TVaR risk metric in stochastic models. While we have yet to come across analytic methods of deriving simulation error estimates for means, the procedure should give rough ball park figures of the simulation error. Heuristically, for appropriately selected *p*'s, one would expect simulation error of means should tend to be less than that indicated by this procedure: the mean is being taken over many simulations and the random absence or presence of the few simulations should not cause large disturbances.

### 3.2.3 Defining the Year Event Table

The Year Event Table (YET) will be made up of a certain number of simulation years from different component models. If

- the component models are A, B, C, …,
- the weights determined in Section 3.2.1 are w(A), w(B), w(C), …, and
- the number of simulations determined in Section 3.2.2 is *N*,

then the YET would have *N*.w(A) many simulations from model A; *N*.w(B) from B; *N*.w(C) from C; etc. We can therefore think of the YET as being made up of several *component YETs*, each containing simulations from one single component model.

The component YETs are generated in potentially two ways, depending on whether the component model outputs are in ELTs or in the YLT format.

#### 3.2.3.1 Deriving a Year Event Table from Event Loss Tables

From ELTs, the generation of the component YET is typically performed through Monte Carlo simulation. The use of ELTs in Monte Carlo simulations is commonplace: see for example (Diers, 2008). The contents of this section should pose little surprise for practitioners.

(The discussion of how to generate Monte Carlo simulations *in general* is outside the scope of this paper. We only mention that there are efficient and well-researched commercial software available for doing so: we expect many (re)insurance companies in the UK to already have such software in-house for satisfying Solvency II regulations with their internal models, which are typically Monte Carlo based.)

##### 3.2.3.1.1 Preparing the event set

For generating the component YET, we would not actually need all the fields from the ELTs. We shall only need the event IDs and the rates. However, we would need to make sure that any events that appear in any of the possible ELTs would be present in the YET generation process. The easiest way to make sure this happens is to take the entire event set from the component model for that peril. Conceivably, one could also take the events from all the portfolio ELTs – but this way may omit events that should be present in the YET generation process from future portfolios.

When preparing the event set, it is also important to take care of the events represented in the event set of the other component models. The event sets should represent events of similar characteristics. If model A's event set contains only category 3 hurricanes or above, and if model B attempts to have hurricanes of all categories in its event set, then model B's event set would need to be trimmed to only have the stronger hurricanes. The resulting blended model would only reflect the stronger hurricanes. (The weaker hurricanes may be modelled as a separate peril, 100% reliant on model B.) In this example, if model B's weaker hurricanes were not taken out, then the interpretation of the blended model would be extremely difficult and confusing. Section 3.2.3.3.1 gives a way of identifying wide discrepancies between modelled events.

### 3.2.3.1.2 Simulating the number of events

Following the notations in Section 3.1.1, we first simulate the number, *k*, of catastrophe events for each of the *N*.w(.) simulations. If events are generated independently of one another, then the Poisson distribution is a natural choice. When the occurrence of events depends upon common drivers (e.g. in the case of European windstorms with atmospheric persistence of overall climatic drivers such as the polar jetstream produces clusters of events temporally and spatially]), then a popular choice is the negative binomial distribution. There are other interesting interpretations of this distribution: we refer the reader to (for example) Sections 6.7 to 6.9 of (Panjer & Willmot, 1992) for further discussions. The providers of the component models may also recommend distributions and parameters.

If the distribution is chosen to be Poisson, then the Poisson parameter (which is equal to the Poisson mean) is typically equal to the sum of the rates from all the events. However, other parameter values are possible. This may be the case if for example we believe that there is significantly higher (or lower) risk of windstorm occurrence for the period we are modelling.

If the distribution is chosen to be negative binomial, then two parameters will be required. Typically, the mean would be calculated as in the Poisson case, summing up the rates from all the events. A second piece of information may be SD, variance or CoV of the frequency distribution. Note that the negative binomial distribution can only cope with variances that are greater than the mean. This may come from the component model providers, or may also be parameterised internally, as required. The method of moments can then be used to obtain the parameters of the negative binomial distribution.

When working with non-Poisson distributions, it can be easier to work with the entire event set from the component model, rather than just a subset of it. The component model provider may provide parameters for the distributions based on the full event set: and if only a subset is being used for generation, then *thinning* would be required. Thinning is discussed in, for example, Section 2.7.2 of (Sundt & Vernic, 2009) for the negative binomial distribution.

By the end of this stage, we would have enough information to fill in the SimulationYear and LossNumber fields of the component YET.

### 3.2.3.1.3 Sampling events

The next step is to assign events from the component model to each of the *k* losses in each simulation. Irrespective of the chosen frequency distribution, this is typically performed by weighted sampling, with weights being based on the rates as indicated in the ELTs.

Effectively, if event 1 has twice the rate as event 2, then we expect event 1 to appear twice as often as event 2 in the component YET. In the authors' experience, this step is surprisingly difficult to programme efficiently with Excel spreadsheets. We expect the readers' stochastic modelling software would be able to provide assistance in this area.

By the end of this stage, the EventID field of the component YET could also be filled in.

### 3.2.3.1.4 Secondary uncertainty percentiles

It is typically assumed that the secondary uncertainty percentiles are independent with all other variables in the YET. We recall that these percentiles are used to look up the portfolio loss distributions, conditional on the simulated event: the higher they are, the higher the event losses would be from the distribution. The secondary uncertainty – reflected by these loss distributions – expresses uncertainty of how the individual properties would respond to the particular natural event, and of how the losses then crystallise.

- Independence against *k*, the number of events in the simulation: there does not seem to be reasonable links or common drivers between the number of events – which is ultimately a physical quantity – and secondary uncertainty percentile – which is a variable associated with non natural objects and activities

- Independence against other events: the jury is still out on this. There is evidence that damages from earthquakes, for example, can affect the overall vulnerability of risks arising from later events in the year. However, it is not clear whether it is due to primary or secondary uncertainty. Please see Section 3.2.7.4 for alternative assumptions.

Under the independence assumptions, for each simulated event, we can generate a number between 0 and 1 from the uniform(0,1) distribution. It is a standard Monte Carlo technique that the inverse distribution function applied to a uniform(0,1) variable would give the distribution represented by the distribution function: this technique will be used in producing the YLTs from the ELTs (see Section 3.2.5).

With the completion of this step, we would have enough information to fill in the SUPercentile field also. Before we move on from the discussing how ELTs could be used to derive a YET, we consider a few technical tools that could help enhance this process.

## 3.2.3.1.5 Stratified sampling

We saw in Section 3.2.2 how we need to have a good balance between decreasing simulation error and putting a check on computing power requirements.

One compromise is to simulate a large simulation set (say, 100,000 simulation years of catastrophe events), but then reduce this simulation set in a uniform way to a smaller simulation set, which would feed into the YET. One way to do this is known as *stratified sampling*.

First, we note that this general method is potentially computationally more efficient. Computing power requirements stem mainly from the production of the Year Loss Tables (see Section 3.2.5 below) and from aggregations: as these are processes that would be repeated – for example, for whenever we bind a new portfolio, or whenever we perform a new aggregation. By contrast, the YET is produced once only: it only needs to be repeated whenever there is a new weight, or whenever the sampling needs to be updated.

For stratified sampling, we need to have a way of lining up the simulations in the large set. We could then divide the line-up into pots of equal size, where the number of pots would be equal to the number of required simulation years of the smaller set. Stratified sampling would then demand that one simulation year from each pot be selected to go into the smaller set. For example, if the larger set has 100,000 simulations, and we require 10,000 simulations, then:

- For each of the 100,000 simulation years, sum up the losses for the company from all the events in that simulation year. From this, we would have a set of 100,000 simulations of aggregate losses to the company from this peril.

- Decrease-order this 100,000 simulation years according to the aggregate losses

- Create 10,000 consecutive pots from the ordering, each with 10 simulation years: the first 10 simulation years would be in the first pot; the second 10 in the second pot; etc.

- Select one simulation year from each of the 10,000 pots, to create a set with smaller number of simulations.

- Create a new (smaller) YET with 10,000 simulation years, by looking up the simulated events associated with the selected simulation years.

This procedure would preserve the exceedance probabilities of the aggregate losses from the larger set into the smaller set, since the same proportions of the simulations would be above specific thresholds – except for the very extreme ones – in both sets.

We have seen practitioners select the simulation years from the middle of the pots, some from the largest of the pots, and some in a random manner. There are theoretical merits in each of these alternatives, but for sensible numbers of simulations, we do not expect the method of selecting simulations from the pots to be influential in calculating means or exceedance probabilities.

The ordering requires values to be associated with the simulation years. In the above example, we considered aggregate losses from all events in the simulation year. Effectively, this gives us the "AEP" curve. However, if the "OEP" curve is the more important quantity, then we would recommend using the *maximum* loss to the company of all the events in the simulation year (if there are no losses in the simulation, then this is set to zero).

We have used, as values, sums of the aggregate losses *to the company*. Instead of losses to the company, some practitioners use losses *to the market*.

- An advantage of using losses to the company is that the company's exposures may not be close to (a uniform share of) the market portfolio. Stratifying on market losses for measuring the company's own exposures would not be as effective in reducing simulation error as stratifying on the company's own losses.

- In the same way, stratifying on the company's own losses to measure exposures from individual classes would not benefit from the method as much as when measuring the company's aggregate exposures. However, conceptually, at least the losses from the classes on the smaller set would add up to a distribution at the company level with small simulation error ranges: the perceived simulation error at the class level would have some direction to it.

- A disadvantage of using losses to the company is that exposures could change dramatically – e.g. after large renewal dates, or for business that is being written dramatically. This would require recalibration of the YET post exposure change. However, we still maintain that this would be a more targeted way of stratify sampling than using market losses as the values.

- There can be circularity with using losses to the company. We would need the YET in order to produce the individual YLTs (as in Section 3.2.5) for accumulation. Therefore, when stratifying on losses to the company to produce the YET itself, one must be able to find another source. One way to do this is to sum up the constituent portfolio ELTs to give an ELT for the company (see Section3.1.2.1), and then create a company YLT from it. This should give a good proxy for the losses to the company. (They would have small differences from the losses calculated by adding up individual YLTs – see Section 3.2.5.1 for a discussion.)

### 3.2.3.2 Deriving a Year Event Table from Year Loss Tables

Some component models have YLTs as their standard outputs. Since the Monte Carlo step has already been done by the component model in producing the YLTs, deriving the YET from the YLTs is straightforward.

Similar to Section 3.2.3.1.1, a key thing is that we do not exclude relevant events – and this could accidentally happen if we make use of YLTs from only a few portfolios. The YLT of a "market" portfolio should have a good chance of containing all the events. The point mentioned in the same section about comparable event sets also holds.

The YLTs are not likely to have the required number of simulations. Stratified sampling (see Section 3.2.3.1.5) reduces the number of simulations in a way that does not increase simulation error for key distributions. We now discuss how one may be able to *increase* the number of simulations.

### 3.2.3.2.1 Repeated sampling

One easy way would be to allow repeat sampling. If my existing YLTs have 5,000 simulations, and I require 10,000, then I could make each of the 5,000 simulations appear twice in my 10,000 set.

Usually, higher numbers of simulations are associated with reduced simulation error. However, this method does *not* reduce simulation error; nor does it introduce more simulation error. The overall statistical outputs remain the same for both sets of simulations. This point may need to be communicated clearly to technical peers, lest a false sense of reduced simulation error is held.

What this *does* for us is that we would be able to continue with frequency blending to the required number of simulations.

### 3.2.3.2.2 Reshuffle of events

The repeated sampling method can be criticised as introducing a lot of redundancies for higher computing costs. If the recipe of event generation could be verified from the component model provider, then we could potentially reshuffle events. For instance, if the events are confirmed as modelled to be independent from one another, then, we could use the Poisson distribution and pick events from the YLTs (like in Sections 3.2.3.1.2 and 3.2.3.1.3) at random.

Again, we need to be cautious about simulation error. It does not give us as much reduction in simulation error as we would expect. We obtain reduction in simulation error in only two parts of the modelling: the overall event frequencies, and the relationships between events within a simulation. Simulation error related to the physical event set itself, the vulnerability and financial modules are not improved. The improvement, then, is seen more with aggregated losses (related to the AEP curve) rather than maximum losses (related to the OEP curve) in the year.

Note that if the component model simulates specific linkages between events, then this method would likely not produce comparable statistical distributions. An example of such a linkage is: increased likelihood of similar events (e.g. more intense hurricanes, hurricanes with similar tracks) given one such has occurred. However, one would imagine this method could be used again if the events could be categorised appropriately.

Finally, for audit trailing or debugging purposes, it may be useful for the YET to contain the simulations in the component model's YLTs. The reshuffling of events can be done for the required extra simulations.

## 3.2.3.3 Putting component YETs together

Once the component YETs are produced, they can be appended together into a single YET. The ComponentModel field would need to be appropriately filled in for the relevant component YETs. The SimulationYear field would need to be offset for the all component YETs after the first one. For instance, the first simulation from the second YET should *not* have a value of 1 in the YET, but $N$.w(A) + 1.

For simplicity, in this paper, we focus on blending the models from one single peril. When putting component YETs together, it is important to consider the corresponding blending for the other perils, lest we have correlations of *models* between different perils. For instance, if

- we are blending 50-50 between models from providers A and B for European Windstorms, and
- 50-50 between models from providers A and B for European Earthquakes,
- with $N$ = 10,000 for both perils, then

using the procedure in the first paragraph would always give provider A for *both* perils in the first 5,000 simulation years, and model provider B in the other 5,000 simulation years. It is likely that both A and B consider the modelling of the two perils relatively independently. A more appropriate allocation of simulation years would be to have: 25% of the simulation year has A supplying both perils, 25% with B supplying both, 25% with A supplying storms and B supplying quakes and the remainder with B supplying storms and A supplying quakes.

### 3.2.3.3.1 Mixed frequency distribution

After the component YETs are put together, the combined frequency distribution would form a *mixed distribution*. Except for the rare circumstance where the component models have Poisson frequencies *with equal means*, the combined YET would not have a simple Poisson or Negative Binomial distribution. This should not be regarded as a problem, but a natural feature of incorporating two different frequency assumptions in a way that keep each of them intact.

Indeed, mixed distributions are commonly used in a number of actuarial applications. The Mixed Exponential distributions are flexible and are used by ISO in the US (see, for example, page 19 of (Palmer) or (Botta & Harris, 1986)) for summarising claims data. The negative binomial distribution itself can be interpreted as a continuous mixture of Poisson distributions – where the Poisson rate is distributed with a gamma distribution (see, for example, Section 6 of (Panjer & Willmot, 1992) or Section 3.2 of (Sundt & Vernic, 2009)).

It is usual to consider the mixed exponential in the ISO case to be single distributions, as the component exponential distributions does not have direct physical interpretations. However, since the top-down standard agreed blend suggests full conditional confidence in individual component models (see Section 2.1.2.3) throughout a simulation year, a natural interpretation of the mixed distribution for us is through a simple Bayesian approach to uncertainty (see Section 3.2.7.5).

Finally, we would need to be aware of bimodality: if the blended distribution is bimodal with widely different modes, then this may be indicative of the component models considering very different event sets. Figure 3-1 is an example, where model A is based on a mean of 5 and model B on a mean of 15. Re-examination of the modelled events would be advisable in such a circumstance, lest we fall in the type of event mismatching problem discussed in Section 3.2.3.1.1.

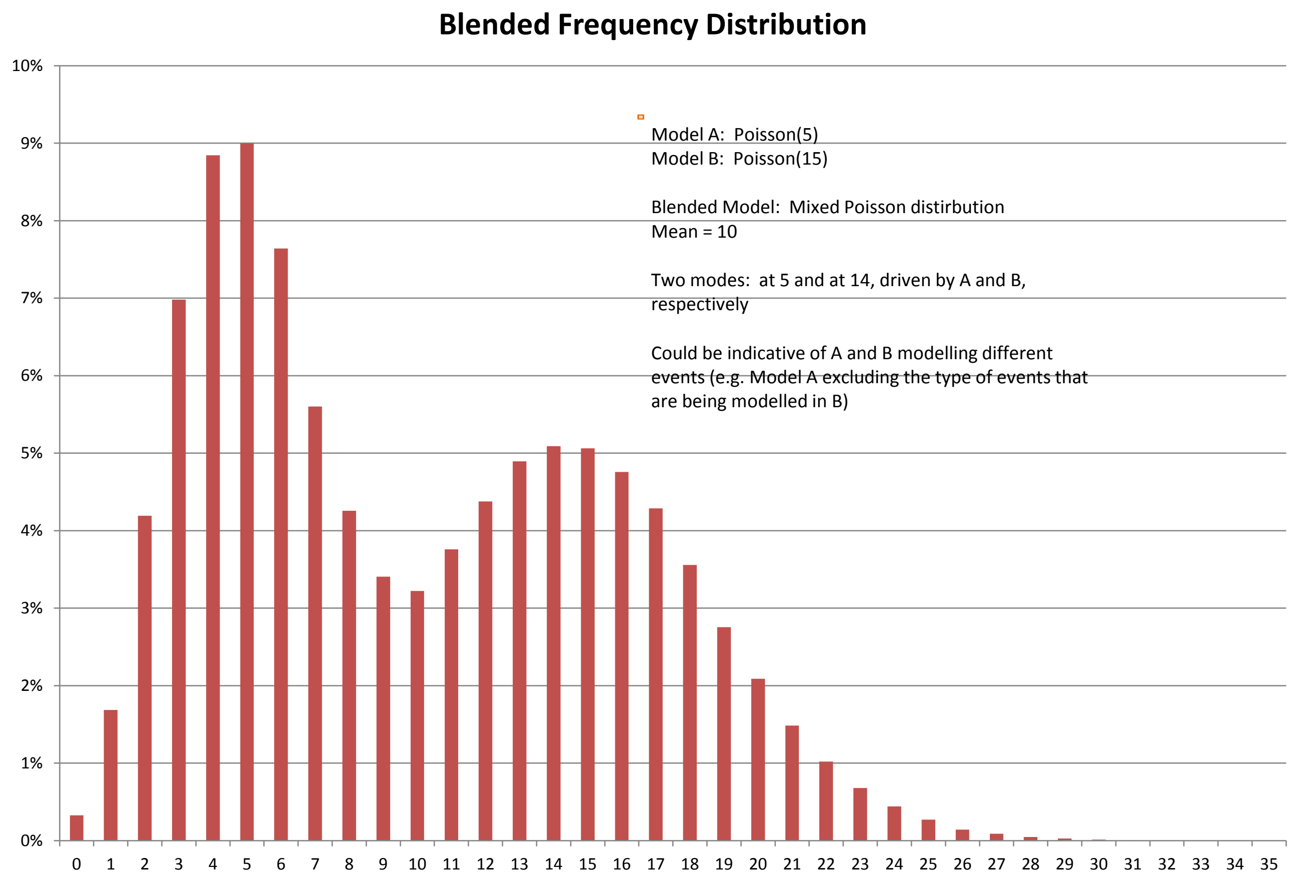



**Figure 3-1 Example frequency distribution after frequency blending: distance between two modes may point to large differences in modelled events**

### 3.2.4 Obtaining the portfolio ELTs and YLTs

For pricing a portfolio of risks, the ELTs or YLTs from all component models in relation to the portfolio would be required. For accumulation of all the portfolios, we would require ELTs or YLTs from all component models for all the portfolios.

This ideal situation may not materialise due to a variety of commercial reasons. The following are options to fill in for missing modelling outputs:

- Convert the portfolio inventory data from one model to another for modelling
- Market shares could be used on industry portfolio model outputs

Each of them has its own problems. Since the different models are developed differently, based on different assumptions and data, there is no reason why the inventory requirements and categories should match closely. The conversion of inventory data may therefore lose detail or the map between the different models may be open to uncertainty, leading to less reliable outputs. The market share approach is vulnerable to even more distortion: as it is unlikely that a portfolio would be a uniform share of the whole industry.

These options should be considered as "short-term fixes" and should only be deployed when the full modelling route is not practical. For these portfolios, appropriate adjustments are expected to be applied in the form of bottom-up adjustments – see Section 3.3 – for better reflection of the contributions made by the missing model outputs.

### 3.2.5 Producing the Year Loss Tables

The mechanism related in Section 3.1.3 is deployed in the example technical solution to produce YLTs for each portfolio of risks. The requirement would be a YET (as defined in Section 3.2.3) and the portfolio loss information in the form of the component model (reduced) ELT or YLT (as discussed in Sections 3.1.2.2 and 3.1.3).

#### 3.2.5.1 Secondary uncertainty pre contract

Where secondary uncertainty is simulated, it is better to produce the YLT to the client *before the contract*, and then apply the terms of the contract to produce another YLT which would be to the company.

An alternative is to calculate the ELT of losses from the contract to the company first, and then produce the YLT to the company. This can usually be done through analytically calculating the loss distribution after application of the contract. Numerical methods are then required to calculate the loss parameters.

The first approach is better because:

- It respects the usually non-linear nature of contracts – with individual limits and excesses – much more sensitively. Since the loss distribution parameters are usually the first two moments together with the maximum, it cannot easily take mass weights into account: mass probability weights are real features in excess of loss contracts, for example.

- It allows consistent simulations of exhaustions and utilisation of contracts in the same programme. In the first approach, a higher layer would only start produce losses to the company once the lower layers are exhausted. In the second approach, all layers in a programme would always have losses, which is unrealistic.

The following chart summarises the relative differences on the loss severities (measured on an OEP basis) of a reinsurance portfolio on an earthquake peril: applying the secondary uncertainty distribution before and after treaty terms. The differences are relatively muted in the nearer terms (+/- 2%), but becoming significant in the more remote return periods (here, from the 500 year point). Application of secondary uncertainty after treaty terms gives higher answers in the more remote return periods because it attempts to lump all risks in the portfolio together in the same distribution – here, specifically, in the second moment of the beta distribution. In reality, reflected better by applying treaty terms after secondary uncertainty, there would be points of discontinuity, where limits of specific reinsurance layers are attained.

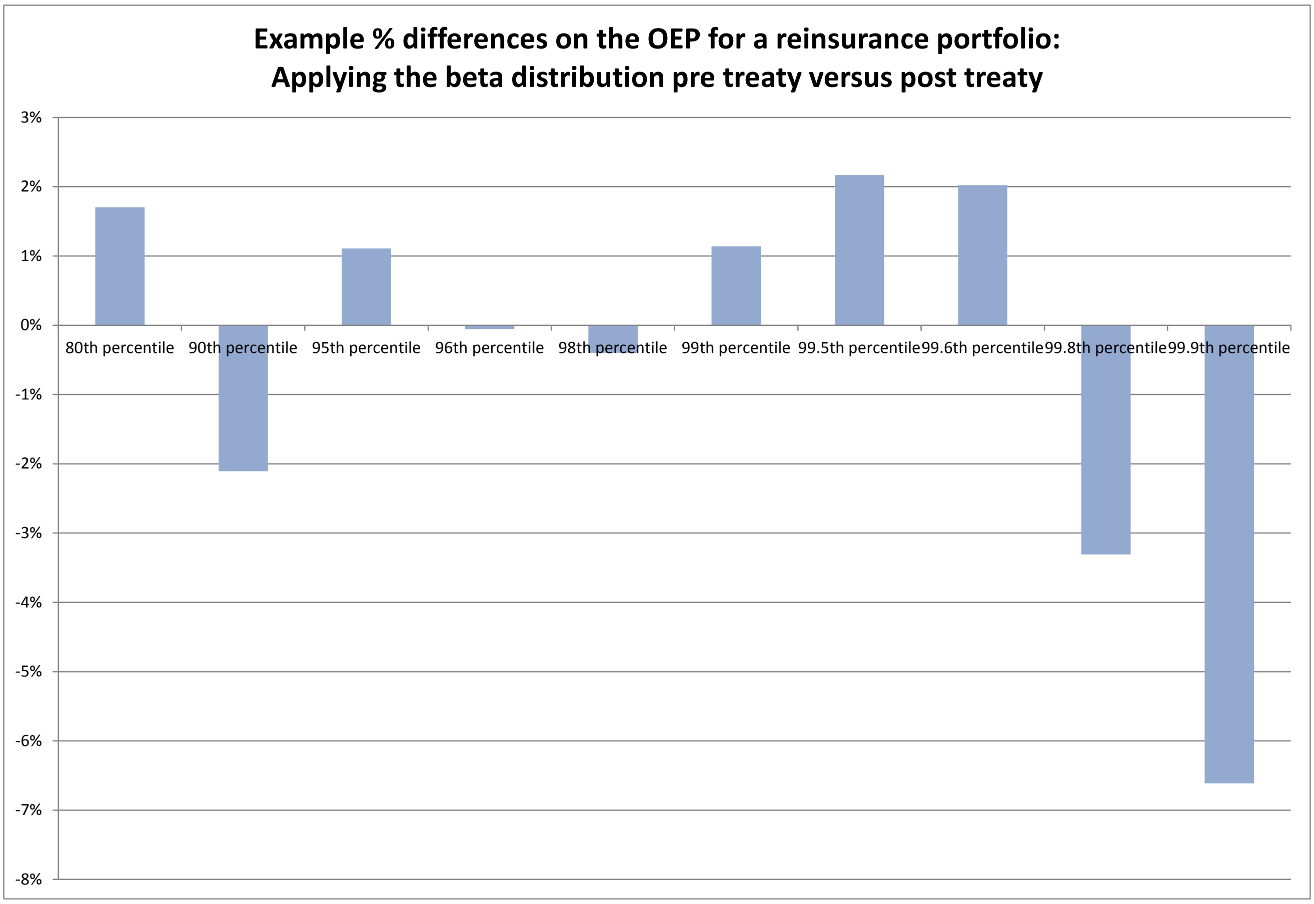



**Figure 3-2 The relative differences on the loss severities of an example reinsurance portfolio when applying secondary uncertainty before versus after treaty terms.**

### 3.2.5.2 Relationship with frequency blending

The top-down standard agreed blend part of the example technical solution, as described up to this point (from Section 3.2.3), results in the frequency blend as defined in Section 2.1.2. By a simple application of Bayes's Theorem, this can be easily seen for the case when two models, A and B, are blended with weights w(A) and w(B), where w(A) + w(B) = 1. Section 2.4 of (Parodi, 2012) has a more general exposition on the subject.

Let $\boldsymbol{X}$ be a random variable vector that contains catastrophe loss information. Typically, $\boldsymbol{X}$ would contain $K$, the random variable of the number of events, and $X_1$, $X_2$, … as the loss amounts from each event. Let $g(\boldsymbol{x})$ be a function that would map the instances, $\boldsymbol{x}$, of the random variable $\boldsymbol{X}$ to a real number. The function $g$ would help to determine the exceedance probability type. Let $y$ be a positive real number – this would stand for the loss threshold.

With regards to the models A and B, the example technical example is such that $N$.w(A) and $N$.w(B) many simulations are devoted to models A and B, respectively (see Section 3.2.3). Each simulation has equal likelihood of occurrence, meaning that **P**(A):= **P**(model A holds) = w(A), and similarly for B. Moreover, **P**(A) + **P**(B) = 1, with **P**(A and B) = 0.

Now application of Bayes's Theorem gives, $\mathbf{P}(g(\boldsymbol{X})>y) = \mathbf{P}(g(\boldsymbol{X})>y \mid A).w(A) + \mathbf{P}(g(\boldsymbol{X})>y \mid B).w(B)$, which is what we had in Section 2.1.2 for frequency blending.

The arguments above hold for the blending of more models. The function $g$ should be defined to give the maximum loss if we are considering OEPs, and to give the aggregate loss if we are considering the AEPs.

In a similar way, with $g$ giving the aggregate loss, it can be shown that the mean of the frequency blend is the weighted average of the means from the component models – which was stated in Section 2.1.2.4.

### 3.2.6 Secondary uncertainty dependencies

In theory, loss distributions from different risks and portfolios can offset one another to some extent. On the variety of modelled drivers common to all risks and portfolios, there is uncertainty with regards to how specific drivers apply to the different risks and portfolios, as well as further uncertainty arising from beyond the modelled drivers. If the drivers create dependencies between risks and portfolios, then these uncertainties create occasions of diversifications between risks and portfolios in their event loss distributions.

The example technical solution assumes conservatively that all the portfolios of risks are fully correlated with one another, for each event, through the use of one single SUPercentile field in the YET (see Sections 3.1.1 and 3.2.3.1.4). An alternative is to statistically reproduce the common drivers between portfolios and to define relationships between them and the losses. This route would require a large amount of research and assumptions: and may well be impossible if the component model providers would not divulge enough details to the company. We note that full correlation are not wildly unrealistic: for an unusual and extreme event – e.g. such as Windstorm Lothar in France, or an earthquake such as the recent Christchurch Earthquake in New Zealand – it is reasonable to expect a good amount of systemic tail risk, impacting on a large number of portfolios simultaneously.

Note that the example technical solution does not suggest *all the risks* are fully correlated with one another: it is suggesting that the component models be relied upon first for calculating the portfolio event loss distributions. Full correlation at the portfolio level has a couple of advantage over that at the risk level:

- It brings the conservatism to a more manageable level. Fully correlating all the millions of homes and office blocks across all the portfolios would be unrealistically conservative: we would be demanding that every individual property exposed to the event to incur similar levels of damages from the event.

- Catastrophe models are statistical models: model outputs are more reliable for a portfolio of many risks than for the individual risks themselves. Practitioners are usually more comfortable working with portfolio of risks. Imposing full correlation at the portfolio level would then avoid having to consider loss distributions of individual risks outside of the models.

For a company, there may already be natural definitions for portfolios and risks. A portfolio may be a business unit in an insurance context. It may be a cedant from a reinsurance perspective. In any case, the definition ought to take into account of practicality. For example, one may want a definition where the portfolio model outputs are produced as standard already; or one would want there to be a simple relationship between portfolios and contracts – whether inwards or outwards.

### 3.2.7 Other variations and further research

Before we go on to the bottom-up adjustment part of the example technical solution, we pause to list a few potential variations to the standard agreed blend part. Some of them are relatively simple ideas, and should be straightforward to implement. Some of them would require further research.

#### 3.2.7.1 Frequency blending at other levels

The example technical solution, following Slide 14 of (Cook, 2011), suggests frequency blending at the simulation level. Each simulation consists of model outputs purely from one component model. This level allows blending of event frequency outputs, including clustering.

However, one may already have good confidence in, say, the event frequencies and clustering – and hence would rather not blend models at this level. One obvious other level for blending, then, would be at the event level: where each event – irrespective of the model associated with other events in the same simulation – can be associated with randomly assigned component models. The event frequency simulation would be separately generated. This is somewhat similar to another way of frequency blending, as described in Slide 13 of (Cook, 2011). The focus here is on building up the event set, with the possible view of using the same frequency assumption for all events.

More so than in Section 3.2.3.1.1, the event sets from all the component models need to represent perils and events of similar characteristics. Indeed, they would also need to be consistent with what the event frequency simulation is representing. If we take this idea further, a possibility is to perform blending at lower levels, as we shall discuss in the following section.

### 3.2.7.2 Matching events

We now suggest a possible area of further research, an area that Slide 25 of (Cook, 2011) has touched on with its variable weighting blending. Typically, the component models attempt to generate a representative list of all possible events. For instance, if we were blending model A and model B, it may be possible to pair up an event from model A with an event from model B. We do not presently quite see how this could be done precisely, but the pairing up should respect essential physical characteristics that affect loss generation. For example, for a windstorm peril, we would expect each pair to have similar storm tracks, footprints and wind speeds. The multi-dimensional nature of matching may mean that exact pairing may not be possible. In this case, one may consider widening the definition of similarity and pairing up *groups* of events.

The work would rely on the component model providers to give clear guidance on detailed physical characteristics of each event in the models' event sets – we would imagine much more so than what some providers are currently divulging. We would also argue that this can be an immensely useful piece of information for the company to further understand the models that it relies on for assessing a major risk.

An event map between component models could help with more nuanced blending. For instance, event frequencies may be considered relatively straightforward to validate and understand, through the availability of historic data. It may be the vulnerability and financial modules that require blending, through the lack of claims data that are available to the company. Having events matched, the company could rely on a single model to generate frequencies of events as well as event selection, but then rely on the mapping to obtain the necessary event IDs to blend losses.

The blending could be performed through frequency or severity blending, depending on the nature of (conditional) confidence one gives to the models (see Section 2.1.2.3 for a further discussion).

### 3.2.7.3 Simulation error

Simulation error, as we have understood in this paper, is the error introduced through the use of the Monte Carlo simulation technique on the model output. As the example technical example relies on Monte Carlo simulations – and indeed, as is the case with many stochastic capital models – simulation error is relevant. We have already touched on this topic of in Sections 3.2.2 and 3.2.3.1.5.

Depending on the return period under consideration, simulation error is expected to be relatively small in comparison to some epistemic uncertainties (compare Table 3-1 in this paper with Figure 4 of (Miller, 1999)). The usual perspective is that simulation error is tolerable if we can tolerate other uncertainties. Indeed, it is usually argued, the existence of an uncertainty category that can be measured as in Section 3.2.2.1 is useful to keep the users on the alert that catastrophe model outputs should be interpreted with ranges.

We do not disagree with these statements: but we believe further effort should be made to control simulation error, where it is practical to do so:

- Simulation error is something that is *on top* of all the other uncertainties, meaning that its presence adds more uncertainty to model outputs.

- Moreover, if the return period under consideration is high, then given a fixed number of simulations, $N$, simulation error could become more significant with some of these uncertainties. One only needs to imagine a layer that attaches at the 1,000 year return period, when we only have, say, 2,000 simulations in our simulation set: the 95% confidence interval of simulation error would be (400, ∞) years – a very wide interval indeed!

The stratified sampling methodology mentioned in Section 3.2.3.1.5 is a practical method to control simulation error for a specific distribution. We suggest a couple of other methods.

- When pricing a very high attaching layer, it can be better to use analytical methods on the ELT to derive statistics such as mean and standard deviation to the layer.

- For the tail events, we can simulate a more detailed YET. For instance, if *N* is 5,000, then the top one percent of the simulations that give the most losses to the company would be represented by just 50 simulations. However, one could further define a YET, with, say, 1,000 simulations, that focuses on this part of the distribution. The overall YET would have 5,950 simulations (= 5,000 + 1,000 – 50). The interpretation of the resulting YLTs would be slightly more complex: each simulation in the original set would have a probability of 1 / 5,000 of occurrence. Each in the more detailed set would have a probability of 1 / (1,000 x 100) of occurrence.

In light of the points made thus far in this section, we have a couple of suggestions for the component model providers or for the market in general, in the area of output provisions.

- When placing very highly attaching layers, if YLTs are being provided, it would be helpful to provide the YLTs run on much higher number of simulations. The standard number of simulations for higher layer placements would not be anywhere near enough for assessing risk.

- Handling and transfers of large YLTs is unwieldy. We suggest the use of ELTs for communication of exposures. Through the methodology we have discussed in this paper (e.g. Sections 3.2.3 and 3.2.5), companies can easily produce YETs and YLTs, even ones with very high number of simulations: all they need are ELTs.

Finally, we end on a note of caution: that simulation error exists for all YLTs. It becomes explicit when we have ELTs to compare with – e.g. we can compare the AAL from the YLT and the AAL from the ELT. Some component models output YLTs as standard, without any accompanying ELTs. Since there are no easy comparisons, there is a risk that we mistakenly think there is no simulation error from the use of the output YLTs from such component models.

### 3.2.7.4 Secondary uncertainty correlations between events

In Section 3.2.3.1.4, the example technical solution suggests independence between loss distributions from different events. The second bullet point of event independence (in that section) could potentially be challenged along the following lines. Our understanding of how buildings respond to natural forces can be open to extreme systemic parameter uncertainty or even model uncertainty – that we have yet to see from historic data. Buildings within a region would have common features that may make them more or less vulnerable to similar perils. There may also be systemic data issue: the data for the region may not be granular or detailed enough for the component models to use the appropriate vulnerability or financial parameters, leading to potentially systemic over- or underestimation of losses. Moreover, the phenomenon of demand surge – where building costs increase due to high demand for reconstruction – would be exacerbated by multiple events from on the same region. If this is deemed necessary, the most conservative way would be to assume full correlation of the SUPercentiles from events from the same peril and region.

### 3.2.7.5 Uncertainty interpretations

In line with Section 2.4 of (Parodi, 2012), the simple mathematics in Section 3.2.5.2 points to another interpretation of the frequency blend: the Bayesian framework to uncertainty.

Referring back to Section 3.2.5.2, from this perspective, we are saying that model A is the “right” model for w(A) of the time, and B is the “right model” for w(B) of the time. More specifically, what we are saying is that the *outputs* of model A is reflective of reality w(A) of the time, and the same for B. This is the narrative in Section 2.1.2.3.

If we interpret the standard agreed blend from the Bayesian uncertainty perspective, then we may be able to extend the concept of this blend in two directions – we shall leave this for future research.

- The use of the Bayes's Theorem suggests Bayesian analysis, with w(A) and w(B) as forming the prior density. These weights could be updated through actual experience, after major catastrophe events. For instance, after five years of losses, if A is judged to be more reflective for all five years, then presumably we would increase w(A) and decrease w(B). The extent to which a year is judged to be an "A" or "B" year is likely to be qualitative in nature, but given the judgements, one would intuitively expect there to be an update formula for w(A) and w(B). The regular updates of component models may mean that stable long periods (or even just for five years!) of data may not be available, without having to engage in a large amount of as-if modelling.

- If we imagine there are different shades of "A" and "B", then we could define a series of intermediate models between A and B, and place weights on each of them. Doing so would likely be mathematically much more complex than the two model set up (such as what we discussed on event matching in Section 3.2.7.2), and hence, likely to be more complex to analyse the blended outputs and potentially less transparent. However, it would also allow a much less dualistic implementation of the blend, in which we have full confidence in A alone, or B alone.

## 3.3 Bottom-up adjustments

The standard agreed blend is a useful framework that produces a blended event set – or indeed it is a catastrophe model in its own right. Such an event set forms the back bone for accumulating exposures. It allows consistencies (up to a point) with pricing, where the pricing blends the AALs with the same weights, and where the AALs are purely derived from the component models.

It is conceivable that underwriting and further R&D in the company would point to certain portfolio of risks to be treated differently from the top-down assumptions (see, for example, Sections 1.3 and 1.8). A practical technical solution should allow for such deviations. In our example technical solution, we propose *bottom-up adjustments* as a set of tools for this.

These tools aim to adjust the severity of losses for each portfolio of risks – and do not alter the event set. This point is critical to the solution, since a key requirement for the solution is to be able to accumulate exposures. There are two key ways of doing so: scaling the severities uniformly and scaling the severities differently for different parts of the curve. Although the second way can be thought of as a generalisation of the first way, we separate their discussion as there can be quite different reasons for doing either.

Just as secondary uncertainty is performed on the *pre contract* losses (see Section 3.2.5.1), the adjustments in this section are also performed on them. More information is typically available from the original business's perspective, rather than from the contract's perspective. The market's observed stability of stable reinsurance programmes notwithstanding, in theory, the contract and programme structures can change: and it is more flexible to apply structures to the original business than to try and adjust the losses to the structures themselves.

Clearly, the adjustments apply more to reinsurance business than insurance business. However, it is possible that some of the techniques can also be applied to insurance at the portfolio level.

### 3.3.1 Uniform scaling

The most basic form of scaling is to multiplicatively scale (pre contract) event losses. The mathematics is straightforward: for each portfolio, define a scaling parameter $s$ so that each loss is multiplied by $(1+s)$. The key here is to record the *reasons* behind the scaling parameter selection, so that there are clear audit trails for peer reviews and future back testing (see Section 4.2 for more detail). Here are some example reasons for the scaling parameter (a discussion on the derivation of the adjustments is outside the scope of this paper):

- Adjustment for growth of the portfolio

- Adjustment for data quality (e.g. for completeness and accuracy of data)

- Adjustment for non-modelled or inadequately modelled loss / expenses (e.g. for loss adjustment expenses, additional living expenses, motor losses, unusual claim settlement practices)

- Adjustment for experience and other pricing analyses

The above examples could be applied either to the component model outputs themselves (i.e. pre standard agreed blend) or to the standard agreed blend, depending how equally adequately each of the component models engage with each item.

On top of the component models, underwriters and pricing actuaries may have other means of pricing catastrophe contracts. As well as possibly other exposure rating models, experience rating can also be helpful, especially for the lower layers. Closer scrutiny of the component models might reveal that a different set of weights might be relevant for a particular portfolio, which may be more focussed on particular regions or occupancy types. All of the above analyses would suggest a different contract AAL to that suggested by the standard agreed blend, even allowing for the adjustments as stated above.

We are then in the situation of the fourth bullet point in the list of reasons for adjustments. Here, one may execute simple search algorithms to find the value of $s$, so that the contract AAL after scaling is equal to the required AAL.

### 3.3.2 Variable adjustments

The second type of adjustment varies depending on the return period of the loss. This differs from what has gone before in the way that the scaling parameter $s$ is no longer constant, but dependent on the original return period of the modelled event loss pre contract.

Insurance companies are increasingly more sophisticated with regards to catastrophe modelling. Submissions for reinsurance renewals may now come with their in-house model outputs, as well as the standard component model outputs. Their in-house models will be unlikely to be a scaler multiple of the component models: and this is where some flexibility would be helpful.

Experience analyses may also reveal differing amounts of discrepancies below a return period. One may choose to rely more on experience below the return period, and on the top-down standard agreed blend for losses above the return period. An example of this approach would, arguably, be more robust for many current peril zones including European Windstorms as demonstrated by the example of recent French Windstorms in section 1.4 and for many Earthquake zones with high frequency, low magnitude events, with localised impacts.

As in uniform adjustments (see above), the system design needs to help with incorporation of comments for adjustments.

#### 3.3.2.1 Scaling on OEP

The scaling parameter in the example technical solution varies according to the return periods, as measured by the pre contract loss's OEP, after secondary uncertainty.

An alternative would be to base the return period on the severity distribution. The advantage of doing so is to enforce a stricter separation between what the individual portfolio could adjust – namely severities – and what they could not – namely frequencies. It would be difficult to maintain the frequency of earthquake occurrence should differ between portfolios!

However, unlike for other areas of actuarial work, the concept of the severity distribution as a separate entity is relatively foreign in catastrophe modelling. The most common statistics is the OEP. Since the derivation of adjustments would likely be reliant on wide conversations and research, and since the adjustments could require peer reviews by other catastrophe modelling professionals, it can be better to use the more widely understood OEPs.

In any case, for the more extreme return periods, the severity and OEP curves (and indeed, the AEP curves) are typically close to one another: the very extreme catastrophe years are broadly proportionally dominated by one single event. Whether we pick the OEP or severity distribution should be of fairly immaterial consequence.

### 3.3.2.2 Interpolation of the scaling parameter

Suppose we have identified *m* decreasing occurrence exceedance probabilities, $p_1$, $p_2$, …, $p_m$ to apply scaling parameters $s_1$, $s_2$, …, $s_m$. The example technical solution now defines a scaling function $s(p)$ that takes any exceedance probability to a real number. The idea is that we would look up each simulated event loss (with secondary uncertainty, before the contract), *x*, on the OEP curve to look up its exceedance probability $p(x)$. Then the scaling parameter to be applied to *x* would be $s(p(x))$.

Our example defines *s* to be a "join the dots" function between the successive points $(p_1, s_1)$, $(p_2, s_2)$, …, $(p_m, s_m)$. If *p* lies between $p_i$ and $p_{i+1}$, then $s(p) = s_i + (s_{i+1} – s_i) / (p_{i+1} – p_i) . (p – p_i)$.

As a simple example, if after experience analysis, we have evidence to suggest that losses are 10% lower than suggested by the standard agreed blend, below the return period of 20 years (i.e. OEP > 5%). One example of the scaling function might be: $s^1(p)$ =

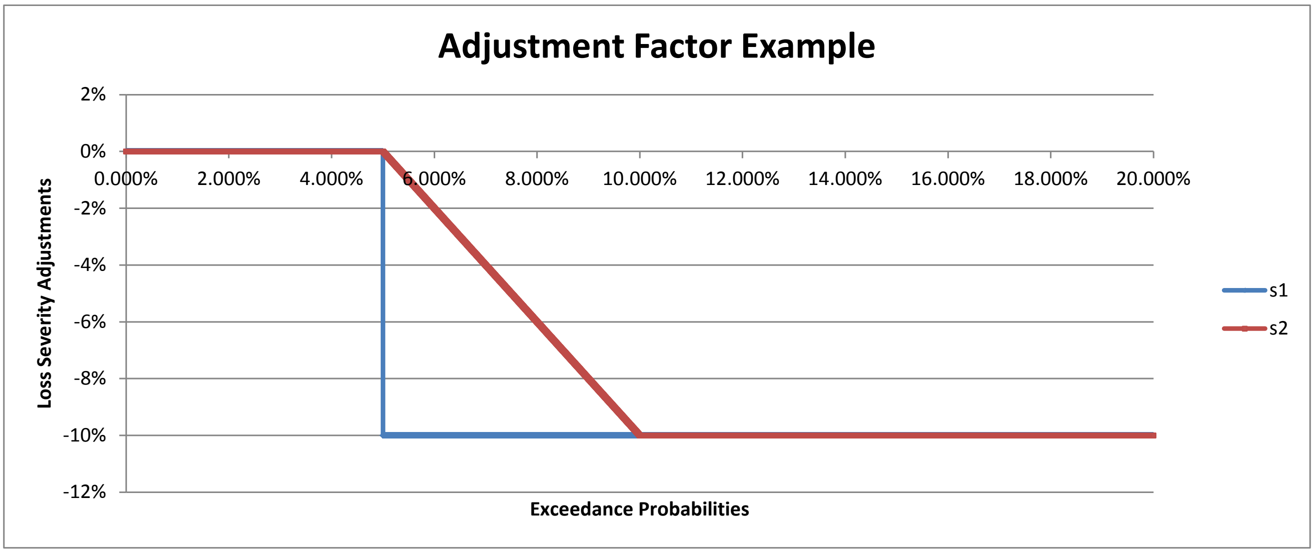
- -10% for $p$>5%
- and 0% otherwise.

This function would reduce all losses at OEP > 5% by 10%, and leave the remaining losses alone, unadjusted.

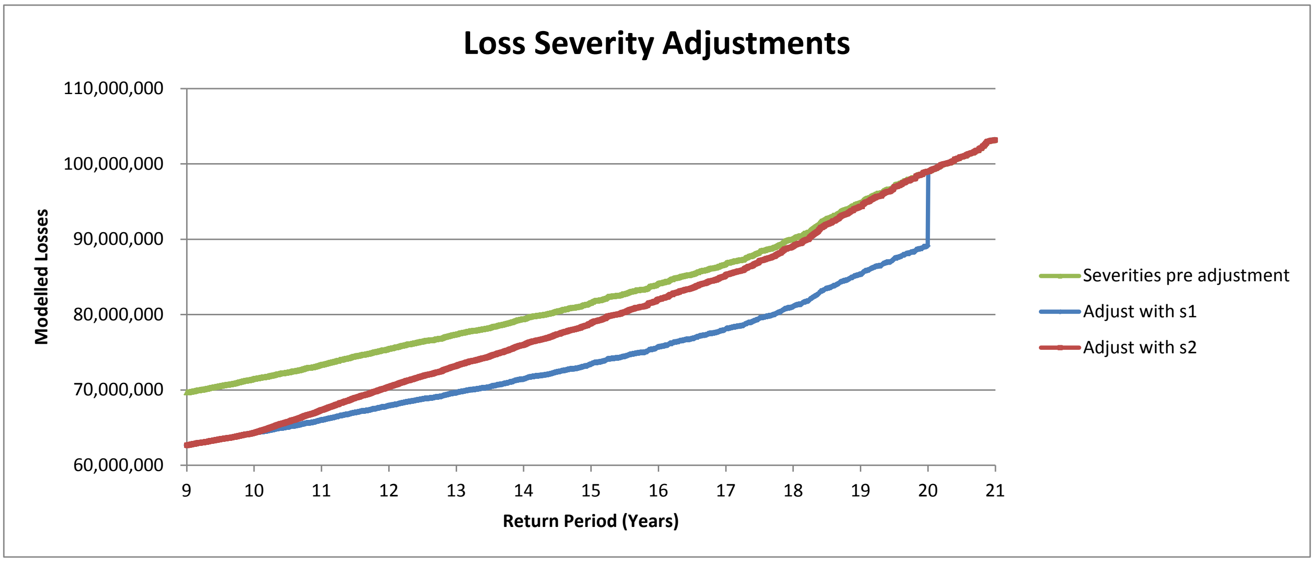
However, the resulting OEP curve would have a discontinuous step change – which might make this rather unrealistic. Indeed, we would imagine there to be a continuous transition from -10% to 0% in the scaling function itself. Another function might be: $s^2(p)$ =

- -10% for $p$>10%,
- 0% for $p$<5%,
- and -10% – 2 . ($p$ – 10%) for 5% ≤ $p$ ≤ 10%.

This function draws a line between $p$ = 10% and $p$ = 5%, linking the scaling parameters of -10% and 0%. The following chart illustrates $s^1$ and $s^2$.



The resulting OEP curve would be continuous, and hence more realistic than that from $s^1$. Heuristically, the link between the two scaling parameters can also be interpreted from a credibility angle: the credibility on our own experience analysis decreases as we move into further return periods. The chart below plots the loss curves pre and post the two adjustments.




We may note that the OEP curve from $s^2$ is not *smooth*, even if it is continuous (see the red line at return period = 10 above). This is because $s^2$ itself is not smooth. It is possible to further refine $s^2$ for it to be differentiable. However, given the adjustments are likely to have a range of uncertainty associated with it, we would expect further refinements would be spuriously accurate.

### 3.3.2.3 Monotonicity

Ordering of events is important in the calculation of the OEP curve. Intuitively, we would want the ordering of events to be maintained before and after adjustments. An event that gave a relatively high loss pre adjustment should continue to give a relatively high loss post adjustment. Mathematically, if we define:

- The function $x$ to map exceedance probabilities to the associated losses for the portfolio (pre contract, with secondary uncertainty)
- The *target OEP function* to be $p :\rightarrow (1+s(p)).x(p)$
- The *post adjusted OEP function* to be the OEP curve after the variable adjustments

then, we would want a decreasing target OEP function, so that the target OEP function and the post adjusted OEP function are equal. The two may differ because the post adjusted OEP function reorders the target OEP function, if the target OEP function is not already a decreasing function. Please see Section 3.3.4 for another reason – in relation to how the different portfolios accumulate with one another – why we would like the target OEP function to be decreasing.

When defining the points $(p_1, s_1)$, $(p_2, s_2)$, …, $(p_m, s_m)$, one validation rule that would be helpful to *reduce* the amount of reordering of events is to demand that the target OEP points, $(1+s_i).x(p_i)$, forms an increasing sequence with increasing $i$. The "join the dots" function, $s$, we obtain would not guarantee a decreasing target OEP function, $(1+s(p)).x(p)$, and hence, would not prevent reordering of the events. However, the target OEP function thus produced would be anchored at the defined points, which are increasing: this anchorage would help to reduce material decreases in the target OEP function – and hence reduce material reordering to produce the post adjusted OEP function.

### 3.3.3 Uncertainty of adjustments

In the uniform scaling case, *conditional on the level of adjustment*, we expect the event loss distributions in the component models would provide adequate reference volatility to cover the volatility associated with the scaling parameter, *s*. Effectively, the multiplication of (1+*s*) to the losses scales up the event loss distributions, keeping the coefficient of variation (CoV) constant. It is akin to saying that the extra losses imposed has the same volatility as the pre adjusted losses, and *are fully correlated to them*. If *s* has significantly different volatilities from what is already in the event loss distribution, then different scaling can be given for the mean loss and the SD of losses in the ELT. Such a situation, where we have clarity of how the volatilities differ, is expected to be rare in practice.

The procedure does not explicitly deal with the uncertainty associated with the level of adjustment itself. For example, while we may put in *s*% adjustment for growth of a portfolio, it is not *certain* that the growth would indeed be *s*%. The enthusiastic practitioners may allow the underwriters to put in ranges for *s*. Rough distributions (e.g. the normal distribution) may then be used to simulate the adjustment. These practitioners should then consider how the portfolio adjustment ranges should interact between one another: whether there should be positive associations because of (for example) systemic overestimation of data quality in the market, or whether there should be negative associations because of (for example) not all portfolios in the market can simultaneously grow beyond what is economically possible. Furthermore, we would also expect them to demand the underwriters to enter ranges *for every single portfolio* they price: since, in theory, even "stable" portfolios are open to uncertainty with regards to exposure changes.

We are also enthusiastic to take material uncertainty into account. However, we have not explicitly included a volatility factor for *s* here in the example technical solution, because it could make the management of the adjustments far too complex to manage, and the results could become difficult to interpret. This example technical solution recommends the underwriters to put in *prudent*, if deterministic, assumptions – perhaps something parallel to "the upper end of a reasonable range of best estimates" of reserving actuaries. In addition, sensitivity analyses and stress tests would be of help to assess uncertainty (e.g. what if the additional living expenses were systematically underestimated by x%).

For similar reasons, the example technical solution would also take a similar line to incorporate the uncertainty associated with defining target OEP points.

### 3.3.4 Dependency considerations

The example technical solution has two parts. We discussed the top-down standard agreed blend in Section 3. This part implements a frequency blend on two or more component models. In particular, it results in a blended event set. We then went on to consider bottom-up adjustments to the standard agreed blend in this section, where, for each portfolio of risks, underwriters would be encouraged to enter appropriate adjustments to the standard agreed blend and the rationale of the adjustments. Adjustments are aimed at scaling the loss severities of losses inside each portfolio: in particular, no explicit facilities *inside the example technical solution* are given for adjusting relationships between two or more portfolios. The dependencies between portfolios are entirely driven by the event sets from the top-down standard agreed blends.

#### 3.3.4.1 Event sets as copulas

Copulas are statistical constructs that are very useful in stochastic models. They are defined on the unit square (cube) $[0,1]^c$, where *c* are the number of variables that are being linked together, describing the dependencies between the variables. The Gaussian, *t*-copula and Gumbel copulas are in common use in insurance applications. Indeed, as soon as there is more than one variable in the stochastic model, "not using" copulas would in effect represent a positive choice of "using" the independence copula. Therefore, all but the simplest stochastic models deploy copulas – we cannot avoid them! We refer the reader to (Shaw & Spivak, 2009) for a more comprehensive guide in this area in general insurance economic capital models, and the classic (Joe, 1997) for a more general reference.

The example copulas we mentioned above are useful constructs: but their implicit nature makes interpretations of results tricky. Some practitioners therefore advocate more explicit ways of modelling dependencies by using drivers of risk (see, for example, (Kerley & Margetts, 2006)). One of the most famous drivers in insurance models is price and wage inflation rates: these variables to various extents affect claim costs and reserve developments. The end result (e.g. the dependencies between the claim distributions of two portfolios) can also be displayed as a copula: in graphical format, most convenient would be to use the "rank scatter" plots.

In catastrophe models, example risk drivers are (i) the event sets and (ii) the mechanisms of how the event characteristics translate to losses. Each event in the set has the ability to cause losses in many portfolios simultaneously: its physical characteristics are key determinants of the severities. As examples, the next two figures (Figure 3-3 and Figure 3-4) give the heat maps of the rank scatter plots from two component models, A and B. Typically, the heat maps would show a lot of independence between the two portfolios: this is because most of the time, losses from catastrophe events are very small. The example heat maps below only focus on events that give rise to large losses to *both* portfolios (more precisely, for events that give rise to more than a modelled 25 year loss).

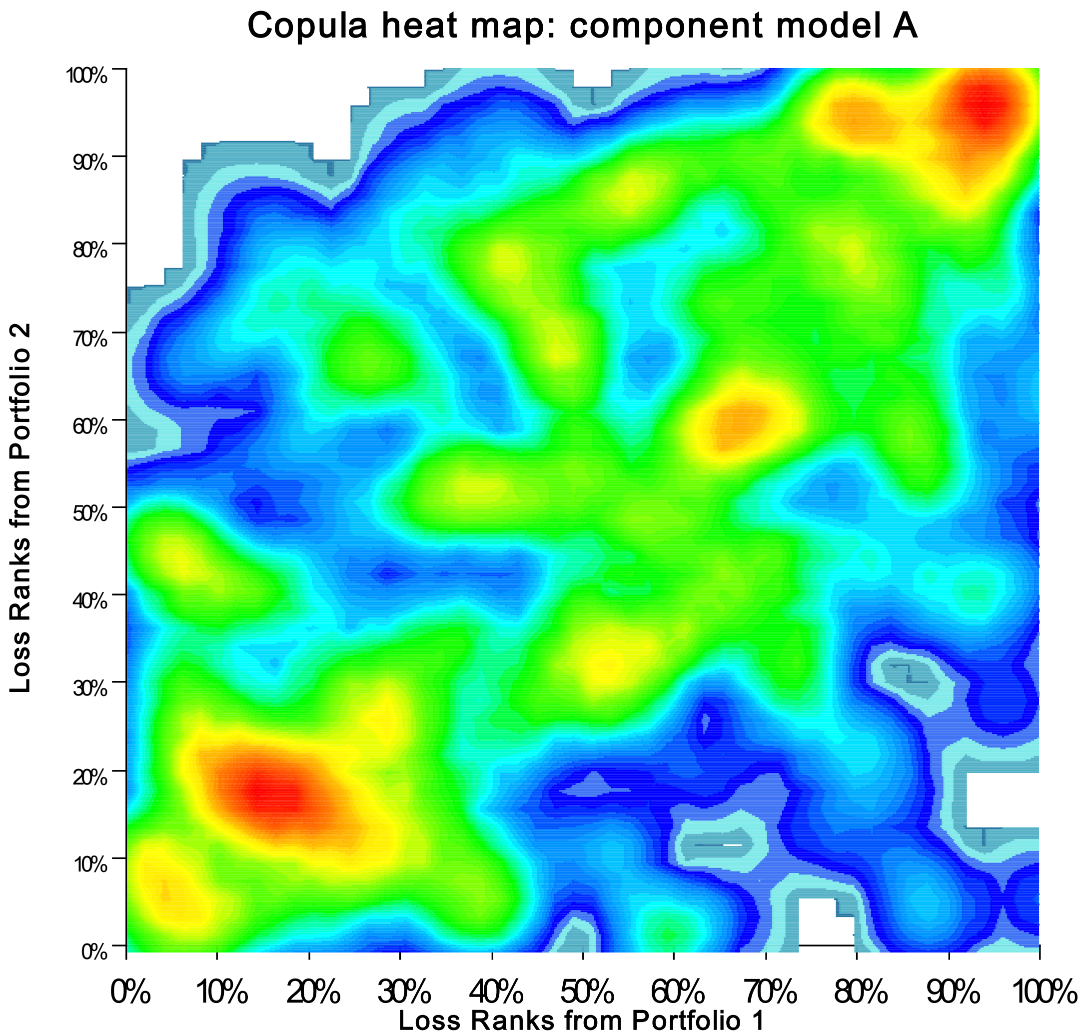



**Figure 3-3 Example copula heat map between two portfolios, for component model A; rank correlation coefficient is 54%**

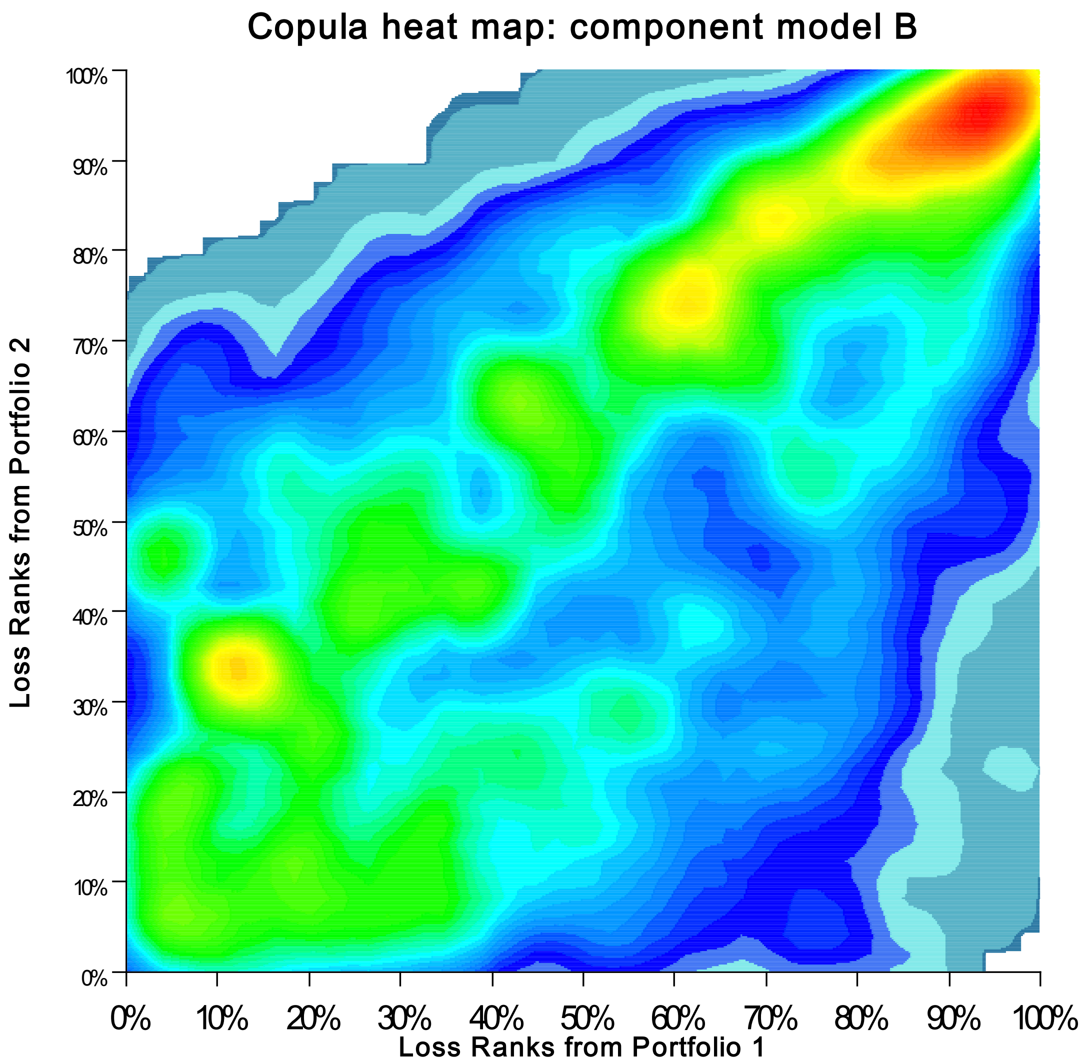



**Figure 3-4 Example copula heat map for losses between same two portfolios, for component model B; rank correlation coefficient is 64%**

The different component models give rise to similarities and differences in the implied copula between the two portfolios:

- Both A and B give tail dependencies in the first quadrant, as denoted by the red areas in the top right corners of the unit squares
- Both A and B have much less tail dependencies in the other quadrants: observe the much wider areas of deep blue and green towards the bottom left of the unit squares.
- Model A gives a more determined linear relationship in the first and third quadrants than model B: consider the shapes of the yellow and green areas for the two models.
- Model B's tail dependence reaches lower down the return periods than A in the first quadrant.
- Model A has more of a body than B, as we can see from the green areas in the middle of the unit square in Figure 3-3

These two examples suggest that the choice of component models can affect the way portfolios relate with one another. Below (in Figure 3-5), we show the effect of a 50-50 frequency blend between A and B. The resulting shape is more similar to B in the first quadrant, and to A in the third quadrant.

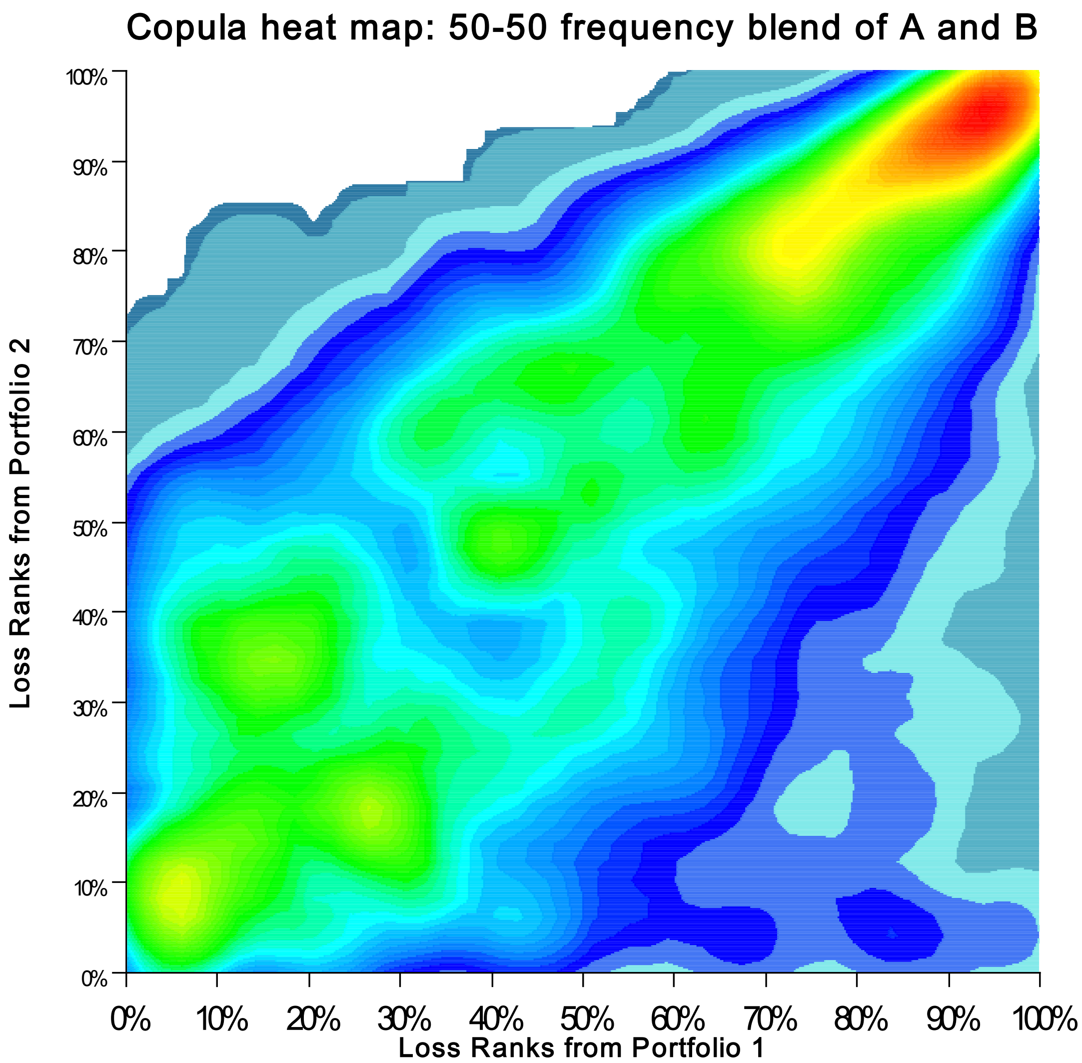



**Figure 3-5 Example copula heat map between same two portfolios, with 50-50 frequency blend; rank correlation coefficient is 70%**

It is well known that copula relationships do not change when the marginal distributions are monotonically transformed. If the adjustments to the severities are monotonic – as discussed in Section 3.3.2.3 –, then we can think of the top-down standard agreed blend as defining "the copula" in the stochastic model, and the bottom-up adjustments as defining the marginal distributions. This kind of thinking might be useful for discussing catastrophe exposure accumulations without the use of common event sets (see Section 2.1.1.3).

### 3.3.4.2 Common loads

The discussion in Section 3.3.4.1 notwithstanding, it is possible to incorporate relationships between portfolios in a deterministic fashion through common loads. The idea is similar to what we discussed in Section 3.3.3, where prudent deterministic loads are used to allow for uncertainty of the exposure adjustments. Here, through the use of *commonly agreed* prudent loads, we would allow for some relationships between portfolios through common drivers. For instance, we may agree that the top-down standard agreed blend (or indeed, one of the component models) would need a loading of *x*% for certain exposure types of occupancy.

This is clearly not the same as changing the copula (or changing the event sets, or the workings of the component models). There is a time and place for such dramatic change (see, for example, Section 4.2.2). But we need also be aware of simple and transparent methods to allow for common drivers.

# 4 Governance

In order to make informed risk management decisions that rely on the use catastrophe models a sufficiently robust governance framework needs to be established. This should consider all aspects of the use of models including devising policies for selection of models, processes for the validation of models and oversight and review of implementation decisions. Sufficient evidence and documentation of the process is considered essential particularly given the requirements of the Solvency II regulatory regime regarding internal capital models which, for entities with significant catastrophe risk, rely heavily on the outputs of catastrophe models.

Various forms of governance can be considered. One more extreme model would be a strictly centralised approach mandating only the use of one single catastrophe model for all exposure classes underwritten (in both pricing and accumulation)with all operational aspects undertaken by a centralised risk management team. This has the advantage of absolute consistency of approach and simplicity of operation. The disadvantage of this approach, as has been highlighted earlier in this paper (see, for example, Sections 1.7, 1.8 and 2.1.2.3), is the risk of systemic errors in relying on a single perspective of a highly uncertain risk. This approach would neglect the potential benefits of using the best available models for differing peril zones, for example, or differing risk types.

At the other extreme would be a heavily decentralised approach with a central committee delegating the choices of model selection and operation to each specific underwriting unit freedom of choice to select and validate models. The advantage of this approach is that the most appropriate model for each class of business could be used but at the expense of consistency of approach. This would be a significant limitation for classes exposed to the same correlated geographical perils. Determining the appropriate level of correlations between business units becomes a more complex blending challenge particularly if no centralised risk management process exists. Also model updates and revisions become more complex to validate if a large and diverse number of models are deployed.

Various other forms of governance between the two extreme cases set out above are conceivable with: for example, a centralised approach to the selection and validation of models; but with some operational autonomy provided to decentralised users to determine the appropriate settings and assumptions for specific regions or business lines.

The remainder of this section is devoted to how model blending might impact governance of catastrophe model use in the company. We shall also highlight governance activities that we found particularly helpful. In particular, with the advent of Solvency II, various advice and guidance (for example, the ABI recently published a guide (Garnons-Williams & Zink, 2011) in this area) are in the public domain: our aim does not include a repeat of their content.

Just as our example technical example is divided into two parts: the standard agreed blend and the bottom-up adjustments, we shall divide our discussion of governance into the governance of these two parts. However, the section should be read as a whole to gain a fuller feel of the discussion: as Section 1.10 anticipated for us, the process presented has a circularity element to it, in which the information gained from the bottom-up adjusted blend would help inform (next iterations of) the standard agreed blend.

We start with the governance of the standard agreed blend.

## 4.1 Governance of the standard agreed blend

### 4.1.1 Vendor model validation

The process of setting the top-down view of risk should ideally be driven by a comprehensive in depth model evaluation process. This is often also referred to as model verification or validation process. Catastrophe model validation is both an art and a science, especially for modelling long return periods. The straightforward, sledge-hammer, way of validating models through comparing with experience data direct is impractical. Not only do we not have hundreds of years of loss data, in particular, to validate 1 in 200 figures and beyond, even if such data existed, the fast-paced socioeconomic flux of the world renders such data irrelevant very quickly. However, there are activities – which we shall share here – that we have found particularly useful to engage in to help the validation process.

This process begins with a full review of the model(s) being evaluated, including the key model assumptions and limitations. Often extensive documentation is provided by the model developers or vendors which can be the basis for an evaluation. All key parameters and data sources for each key module – hazard, vulnerability and loss/financial calculations – would be reviewed and understood.

However, as well as reviewing documentation, a more robust level of validation can be achieved by undertaking a rigorous and systematic series of model tests and review.We shall split this discussion into two part:

- Validation of model outputs
- Validation of model assumptions

Expert practitioners carrying out this work can includ atmospheric scientists and earthquake scientists within one's in-house Catastrophe R&D team. In addition senior catastrophe modelling specialists perform many of the sensitivity and parameter tests and extractions to complement the fundamental R&D.

#### 4.1.1.1 Validation of model outputs

The validation of model outputs can be carried out with a series of reference exposures as a base. The use of industry portfolios is a common choice to provide a "weighted" outcome – the one that, we would imagine, model developers spend the most effort in. The use of significant portfolios, specific to the companies, is useful to gain helpful insights to judge the model's suitability for use within the organisation. Many levels of output could be usefully assessed including:

- overall industry losses
- portfolio specific losses
- historical as-if losses
- losses by zone / sub zone
- losses by major risk types
- construction and occupancy classes

For high return periods, these outputs can then be compared and contrasted against the outputs of other models, in the context of the individual models' assumptions and limitations. The rationale here is that independent efforts to model the same perils would help to highlight strengths and deficiencies of each other.

By "other models" here, we mean

- Current versions of available catastrophe models
- Previous versions of available catastrophe models (especially when we are validating a new version)
- Current version of the in-house standard agreed blend
- Current version of the bottom-up adjusted accumulations

A key feature here is that: just as blending would not remove all uncertainty for us due to, for example, the use of common historic hazard datasets (see Section 2.1.2.3), model assumption and output comparisons would not highlight all the limitations. The consideration of secondary events is helpful to supplement such comparisons (see below).

In addition, for lower return periods, if available to the company, the outputs can be compared against one's own historic loss data and experience. Indexed historic market losses can also be helpful. We expect this to be familiar to actuaries – especially those working in the pricing area, where experience and exposure rating are often used side by side. An extensive example can be seen in a study done on the US Tornado Hail peril, (Smosna, 2012).

We also refer the reader to Section 1.6 where we have included two charts from a recent internal model validation exercise: one comparing the model outputs at an industry level, and the other comparing historical losses against model outputs, at a portfolio level.

It goes without saying that the opinion of experienced underwriters is helpful on these outputs. With the example technical solution, their opinion now flows through quantitatively in the bottom-up adjusted accumulations (see Section 4.2.2 below), as well as via informal conversations and formal discussions in governance committees. Qualitative discussions can also assist this process: for example, Q&A with the model developers or other market participants such as broker evaluation teams.

We have found it useful to cover the modelling of secondary events following the modelled events in this assessment, particularly to determine the extent of the limitations of using a given model. Examples include:

- tsunamis following earthquakes
- fire following earthquakes
- storm surges from windstorms
- construction material and labour cost inflations due to sudden post event surge of their demand
- the pertaining insurance environment – for example, legal interpretations of Terms & Conditions

### 4.1.1.2 Validation of model assumptions

Where made available, individual key parameters can be assessed. We give two examples here: one on the hazard side, and the other on the vulnerability side.

An example for hazard parameters for earthquake would be the range of event magnitudes, considered in conjunction with their associated frequencies. Ideally hazard parameters should be benchmarked against independent scientific data. For example, rates of modelled event recurrence can be compared to recognised earthquake catalogues with published levels of completeness. Should a model noticeably diverge against accepted scientific data, further analysis maybe required to establish the cause. An example may be that a model developer chooses to add additional caution for extreme events. Where assumptions agree this also helps to demonstrate a level of confidence and understanding in a given model rather than simply relying on the model to have reasonable assumptions.

Similarly, independent sources of loss or vulnerability data could be used to review and evaluate models to give evidence on their fitness for purposes. For example, historical scenario results can be compared to observed losses to determine if any biases or limitations exist. This is often best done against a representative sample of portfolios. Alternatively if no direct claims experience is available in sufficient detail, engineering based data or damage surveys can be used to assess vulnerability curves or at least relative views of vulnerabilities, such as residential compared to commercial risk.

### 4.1.2 Governance committees and model adoption

Once a model evaluation is finalised, for further discussion and review, written findings could then be presented to any governance committees or processes, established with a specific remit to assess catastrophe risk. The written report should ideally contain, as well as thekey findings, recommendations for the adoption of the given model(s), possibly with adjustments. In situations where multiple models are considered, the strengths and weaknesses of each model in the key sub modules (hazard, vulnerability, loss calculations) are given. Should no clear "winner" emerge from the technical assessment and the models concerned have sufficient credibility; a case for blending models is considered an appropriate response, particularly for regions with significant exposures. Consideration of the weight of evidence will be undertaken to devise the appropriate weights for "blending" two or more models.

Thus any blend of models can comprise:

- 100% of unadjusted view of a single model if it is considered robust enough, or
- an adjusted view of a single model, or
- a blended view of models, also possibly with underlying adjustments

We recommend the make-up of governance committees determining any such blend should comprise relevant stakeholders with sufficient experience of catastrophe perils, with a careful balance of underwriting and risk management functions to ensure appropriate oversight. Decisions could be taken on a "materiality" basis as a guiding principle based on the undertaking's exposures. For example a more senior committee can be formed responsible for governance of the model selection and blending weights for major peak zone perils, with scope to form a more junior committee responsible for decisions regarding less material perils for the (re)insurance undertaking.

Finally, in agreement with Chapter 5 of (Garnons-Williams & Zink, 2011), we find documentation of model selection decisions and the rationale behind them to be helpful: not only does it provide audit trails, but such documentation encourages model selection in a disciplined manner.

## 4.2 Governance of the bottom-up adjusted blend

The bottom-up adjusted approach allows underwriters the freedom to adjust the default blends on an account by account basis for any of the modelled catastrophe perils and zones. As mentioned in Section 3.3, two main types of adjustment are enabled, both of which adjust the blended pre-treaty YLT. The treaty terms are then applied to the revised YLT. The first method allows underwriters to enter a revised expected loss to a layer by either reweighting the constituent modelled results, for example, shifting from a 50:50 split of given modelled outputs, to favour a specific model, such as to use a 60:40 split. Alternatively, the expected loss adjustment by layer maybe sourced from an actuarial based model providing a burning cost output. The second adjustment method is to adjust the exceedance curve itself by adjusting at different loss percentile points. For example losses can be adjusted at the 1 in 10, 25, 50 years, including non-linear type adjustments.

Once the adjustment is committed the underlying YLT is adjusted by the algorithm to derive a new loss to a layer and form a (treaty) YLT as a basis for accumulation reporting. Ideally the underwriter would be required to document the basis of the adjustment and could also choose one of a few predefined types of adjustment to provide clarity on the basis of adjustment.

Overall two distinct levels of governance deployed around the use of the "bottom-up" view:

- "Day to Day" via various levels of peer review for individual uses at account level
- Strategic monitoring and review of the overall use, impact and extent on the entire portfolio by comparison of the aggregated result to the standard blend

### 4.2.1 Account level governance

The governance of the adjustment process at an account level can be achieved well by three levels of review and sign off, calling on the expertise of senior underwriters, catastrophe risk modellers and actuaries. The first two are part of a broader peer review framework designed to consider all elements of the underwriting process.

- The first process is the underwriting peer review process. Here, a more senior underwriter is engaged to review and sign off all aspects of the underwriting including any pricing assumptions. This process is undertaken in conjunction with company's approved pricing policies and processes, and the catastrophe risk modelling process, which is itself subject to peer review for material accounts "pre-binding" of business.

- Secondly, all bound accounts would be subject to a Cat Risk Modelling team review undertaken by a modeller independent to the original modeller participating on the account. The governance of such reviews can be directed to ensure accounts with the most significant exposures are prioritised and reviewed by the most senior analysts. The review should challenge the basis of any adjustments considered to lack robustness.

- The third level of review is targeted purely at specific accounts: these could be carried out by a pricing actuary. The need for the review is triggered by the materiality of the adjustment (either positively or negative at certain threshold, which can be adjusted centrally), or if the exposure commitment exceeds a certain materiality threshold (again defined centrally and applying to all underwriting units). Furthermore, to cover the remaining contracts, a random sample approach could be introduced forcing a certain percentage of accounts to have an actuarial review.

Looking at publicly available literature, discussions on the subject of peer reviews are surprisingly limited, even though there is widespread agreement that peer reviews should be part of actuarial work in general. A comprehensive discussion is (Kucera & Sutter, 2007), and we also refer the reader to (Gibson, 2008) for an account in the context of reserving. The few detailed discussions on peer reviews usually centre on reserving, and we have yet to find one engaging in the subject of peer reviews on the use of catastrophe models. We now outline a few thoughts for future research.

The aims of peer reviews are likely to be similar between reserving and catastrophe modelling (e.g. ensuring processes are appropriate, spotting material errors, improving methodologies, helping with appropriate communication of results). However, peer reviews in the context of catastrophe modelling can require different approaches.

- The day-to-day multi-disciplinary nature of catastrophe modelling requires different communication techniques than an actuarial peer review of a reserving report produced by another actuary.

- Since we can be considering tail events, the data is likely to be much sparser or non-existent at these levels for validation and calibration.

- Probabilities ("1 in 100") are the more usual statistics than the mean ("Best Estimates"). Together with the previous point, similarities with peer reviewing capital model outputs are apparent.

In light of these observations, future research on peer reviews on the use of catastrophe models would benefit from considering latest practitioner insights from the compliance of Solvency II's validation test (see, for example, (Chhabra, Validating internal models: a practitioner's perspective, 2011)). Behavioural aspects of communication and eliciting expert judgement on remote return period events will also be important: see, for example (Chhabra & Parodi, Dealing with sparse data, 2010) and (Arbenz & Canestrano, 2012). The work of Daniel Kahneman and Amos Tversky (the former a 2002 Nobel prize winner in Economic Sciences) are likely to be helpful: Kahneman's recent book (Kahneman, 2011) has attracted good reviews.

### 4.2.2 Portfolio level governance

For major exposures, regular quarterly reporting of overall accumulations for the entire portfolio incorporates comparative analytics of the top-down standard agreed blend outputs and the bottom-up view. Significant variances between the two bases of reporting would be detected. Should this cross a materiality threshold, further analysis and report can be undertaken to investigate if any systematic pattern is applied across all accounts within a given accumulation zone, or if the adjustments are restricted to individual cases. Depending on the nature of the adjustments applied, and the overall impact, scope then exists to review in further depth within any committees mandated with monitoring and controlling overall catastrophe accumulations risk assessment. These committees can be identical to those alluded to in Section 4.1.2.

We expect any relevant governance committees comprising various stakeholders directly involved with catastrophe risk assessment but also those independent of the underwriting process in Risk Management functions can provide oversight. These committees would determine if the basis of the adjustments provides sufficient indication that further research projects should be commissioned. Such projects would aim to independently support adjustments with an eye to then “re-set” the top-down standard agreed view to correspond. However should the investigation instigated determine the basis for the adjustment to be insufficiently robust, the existing top-down view would then remain as the ultimate reference point for determining levels of catastrophe exposures.

# 5 Final Words

The final section provides closing commentary regarding the use of multiple models within Internal Models related to Solvency II (Section 5.1.1). We then focus our comments regarding the implications for the use of bottom up adjustments to a standard blending based approach and the level of documentation required (Section 5.1.2). The final sections discuss one of the major limitations of multiple models and blended approaches to capture the entirety of any undertaking's cat risk, namely that of non-modelled exposures (Section 5.2) and conclude by sign posting the future direction of cat risk modelling approaches related to blending (Section 5.3).

## 5.1 Solvency II

### 5.1.1 The use of multiple models in the Internal Models

Demonstrating compliance with Solvency II has the potential to be very onerous when using multiple models versus the same level of work performed on one model. As well as the actual validation and review work, compliance would likely also involve maintenance of documentation, minutes and correspondence (as discussed in (Garnons-Williams & Zink, 2011), especially in Chapters 3 and 7), with the aim of demonstrating the detailed and thorough review and understanding of the external models.

In demanding good understanding of external models, and in requiring demonstration of this good understanding, Solvency II encourages (1) the use of multiple models to promote understanding of the model and risk and (2) subsequently the openness to the possibility of blending models. The first point is echoed by CEIOPS's Level 2 advice (see quote in Section 8.2 of (Garnons-Williams & Zink, 2011)). Allowing for the principle of proportionality, if multiple models have not been considered, it would appear that a firm is not making use of existing available tools to understand catastrophe risk. The second point, we would argue, consolidate the first point: once such understanding has been attained, it would appear strange not to blend models when, for example, two models giving significantly different answers are judged to be appropriate for the company's risk profile.

As a consequence, when reviewing applications for model approval, we would expect regulators to consider the use of multiple models, where catastrophe risk is a material driver of risk.

### 5.1.2 Documentation and validation implications for the bottom-up adjustments

In Section 3.3, we have discussed a way to allow bottom-up adjustments to flow through into an accumulation, and therefore potentially into the company's internal model. Moreover, in Section 4.2.2, we have suggested how this accumulation could be deployed to govern the use of catastrophe models in the company, answering the concern voiced in Section 1.10, with regards to quantitative feedback loops (and the lack of them) in an organisation. There may be differences in assumptions internally between:

- the models used to underwrite each risk,
- the models used to accumulate risks across a portfolio, and
- the models used to consider overall risk capital requirements.

In an ideal world all approaches would be consistent or would even form a seamless whole. However, this is unlikely to be the case, at least simultaneously. If anything, there would be a time delay between the different points in the feedback loops: after bounding of risks, the bottom-up adjustments need time to be accumulated and reviewed against the standard agreed blend accumulations; R&D investigations would be prioritised and undertaken; model selection committees would take the R&D findings and take time to debate and decide on (possibly) new model blends; the new model blends would take time to be rolled out for the next iteration of pricing.

Moreover, under Solvency II, the Internal Model would require an appropriate level of documentation, to satisfy some of the requirements. It may be possible to document general principles when making adjustments, but it may be much harder to document every specific adjustment (and their validations) made to a single model or a standard blend of models. Therefore, it may never be the case that all the bottom-up adjustments would be flown through the Internal Model: it is likely then that an overall methodology will be used for the capital model – one that can be relatively easily documented and can still be validated to be fit for purpose.

Having said this, the bottom-up adjusted accumulation would be an invaluable source of information: the overall comparison of the bottom-up adjustment with the standard agreed blend can act as validation in itself. If the comparison between the two approaches yields similar results then the standard agreed blend can be documented and used. If material differences arise between the two approaches it may suggest that the standard agreed blend needs to be reviewed.

## 5.2 Non-Modelled Perils

A significant issue and limitation of cat risk assessment is the issue of "non-modelled perils". The definition of non-modelled perils is itself a huge question. Although cat modelling is now well established, the extent and depth of coverage by the models are not complete, particularly for entities with global exposures. There are two main aspects to these limitations particularly with respect to event or scenario based modelling to capture physical correlations:

- Peril zones with no available models such as riverine flooding or wildfires
- For modelled perils, non-modelled aspects particularly for secondary or interrelated perils such as the impact of a tsunami following an earthquake

In the latter case it is possible via research and development to incorporate factors to existing models to attempt to account for their impact. Such factors could be relatively crude "loadings" applied to base outputs based on loss observations or market observations. For example, for a major earthquake event a 10% of the total loss may be judged to have arisen from fire following the earthquake which is not incorporated within the models available for the particular earthquake zone. Thus a factor of 10% could be added to loss calculations. However, this is a relatively crude approach, and there may be significance variance in risk depending on the hazard and vulnerability in respect to areas of dense industrial activity, and more widespread commercial activity. Moreover, secondary effect may be localised and generated by only certain types of events. For instance, large scale tsunamis are most likely to be generated by megathrust subduction zone earthquakes and not necessarily by all offshore earthquakes (see, for example, (Fuako, 1979)).

A potentially more sophisticated approach is to define the range and extent of sources of secondary hazard and identify and determine the source events within given hazard models are most likely to produce such effects. The second stage then is to determine a means of assessing the intensity and extent of the impact to determine a factor to add. This is complex and requires careful assessment of exposures, insurance conditions (certain secondary perils may be for example sub-limited) and ultimately an assessment of damage caused by both the secondary peril but also the primary modelled loss to avoid double counting of risk. For example a tsunami may damage an already damaged building by the initial ground shaking and depending on the insurance terms the policy may already be a "total loss". On the other-hand, a tsunami may damage risks well beyond the extent of any earthquake shaking damage. Ideally one would seek to add factors derived from detailed research and development to specific events to adjust losses, the factors can be variable but allow for a consistent view to be taken across all exposures assessed. This can be achieved via the use of overall loading models that determine the valid events for each source model and any adjustments necessary.

For the situation of starting with no catastrophe model to determine exposures, more traditional actuarial based approaches to frequency and severity of loss using burning cost data and assumptions for frequency and severity can be integrated. A range of non-modelled peril zones can be incorporated into a solution and outputs from such pricing models are captured to provide a total technical price for a contract, combining the modelled zones and non-modelled zones. Care should be taken on the definition of these zones to ensure no correlation is assumed to the modelled zones.

In terms of managing the portfolio, management information (MI) and reporting of the in-force expected losses to non-modelled zones can be combined and reviewed. With such MI, it is possible to determine the most significant non-modelled exposures which can help determine research priorities or management mitigations by reporting the total limits exposed to non-modelled zones by contract types. Potential exists to define a correlation matrix between contracts to assess a “pml” type view within a zone recognising that in certain large zones (e.g. Australian flood), not all business particularly reinsurance business, will correlate 100%, even in the more extreme flood scenarios.

## 5.3 Future Direction for Blending Approaches

As discussed in, for example,Section 1.3 regarding uncertainty, we recognise that significant limitations of the blending approaches discussed within this paper remain, despite the emphasis of using multiple validated models to attempt to address variance in results and performance. Systemic aspects of error such as non-modelled elements discussed in the previous section are a key limitation but also the proprietary nature of current vendor models ensures a deep component level validation of all parameters is extremely difficult for users to undertaken independently of the vendors.

A component level approach to blending models (as briefly alluded to at the beginning of Section 2) to develop in-house views of risk by selecting the most appropriate hazard, vulnerability and exposure modules is in our view impossible to implement consistently and fully for a broad range of exposures within the current technological and scientific context of vendor catastrophe models.

However, we believe this situation is changing. Partly as a result of greater emphasis on validation of models brought forth by the discipline introduced by Solvency II and reactions of user’s experience to the application of vendor catastrophe models, a significant shift in approach is occurring. Greater transparency around modelling assumptions is becoming more apparent with emphasis on the provision of sensitivity options. Confidence intervals of key outputs such as vulnerability functions are being readily made available in the latest releases in 2012. Moreover, in future releases towards the middle of this decade, the ability to further deconstruct fully all the relevant modules may be possible with additional tools and approaches being developed by vendor firms and other market initiatives e.g. open source models.

Such initiatives could give rise to the opportunity to implement component level blending provided sufficient levels of standards and consistency are adopted by all vendors to enable interoperability. For example, the ability to use the hazard outputs of Vendor A with the vulnerability relationships of Vendor B. Whilst in theory this is readily conceivable to produce accurate results, a vital pre-requisite would be a consistent basis of output. For example, for an earthquake model component blend the shaking intensity of Model A would need to be defined or provided, such that impacts of soils were not also being reflected in the vulnerability functions of Model B. Provided such technical and organisational challenges can be overcome scope exists to improve the accuracy of blending and thus reduce the levels of uncertainty.

# 6 Works Cited

Anderson, D., & others, a. (2007). *General Insurance Premium Rating Issues Working Party.* UK Actuarial Profession.

Arbenz, P., & Canestrano, D. (2012, May). Estimating copulas for insurance from scarce observations, expert opinion and prior information: a Bayesian approach. *ASTIN Bulletin, 42/1*, 271-290.

Botta, R., & Harris, C. (1986). Approximation with generalized hyperexpoential distributions: Weak convergence results. *Queuing Systems, I*(2), 169-190.

Cairns, M., & others, a. (2008). Variable Capital Loads in Pricing. *GIRO.* UK Actuarial Profession.

CAS. (2010). Fundamentals of Catastrophe Modelling. *CAS Ratemaking & Product Management Seminar*, (p. 44).

Chhabra, A. (2011). Validating internal models: a practitioner's perspective. Liverpool: UK Actuarial Profession.

Chhabra, A., & Parodi, P. (2010). Dealing with sparse data. *GIRO.* Liverpool: UK Actuarial Profession.

Clark, D. R. (1996). *Basics of Reinsurance Pricing.* Retrieved July 2012, from Casualty Actuarial Society: http://www.casact.org/library/studynotes/clark6.pdf

Cook, I. (2011). Using Multiple Catastrophe Models. *The latest issues surrounding catastrophe modelling* (p. 38). London: The Actuarial Profession.

Devlin, S. (2008). Catastrophe Reinsurance Ratemaking. *Midwestern Actuarial Forum*, (p. 30).

Diers, D. (2008). Stochastic modelling of catastrophe risks in DFA models. *ASTIN Colloquium*, (p. 19). Manchester.

Florida Hurricane Catastrophe Fund. (2005). *2005 Ratemaking Formula Report to the State Board Administration of Florida.*

Fuako, Y. (1979). Tsunami Earthquakes and Subduction Processes Near Deep-Sea Trenches. *J. Geophysical Res., 84/B5*, 2303-2314.

Fulcher, G., & others, a. (2006). Report of the Catastrophe Modelling Working Party. *GIRO.* UK Actuarial Profession.

Garnons-Williams, T., & Zink, U. (. (2011). *Industry Good Practice for Catastrophe Modelling: a guide to managing catastrophe models as part of an internal model under Solvency II.* ABI.

Gibson, L. (2008). Defence against the dark arts: a practical handbook for reserving actuaries. *GIRO.* UK Actuarial Profession.

Grossi, P., & Kunreuther, H. (2005). *Catastrophe Modelling: A new approach to managing risk.* Springer.

Guy Carpenter. (2011). *Managing Catastrophe Model Uncertainty: Issues and Challenges.*

Hall, T. M., & Kleinn, J. (2011). Tropical cyclone return periods. *GIRO.* Liverpool: UK Actuarial Profession.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts.* Chapman & Hall.

Kahneman, D. (2011). *Thinking, fast and slow.* MacMillan.

Kerley, C., & Margetts, S. (2006). Top down / bottom up correlation. *GIRO.* Vienna: UK Actuarial Profession.

Kucera, J., & Sutter, R. (2007). Peer Review. *CAS Spring Meeting.* Orlando: CAS.

Major, J. (2011). *Catastrophe Model Blending: Overview.* Guy Carpenter.

Miller, D. (1999). *Uncertainty in Hurricane Risk Modeling and Implications for Securitization.* Retrieved June 06, 2012, from CAS Discussion Papaer Program: 1999 Discussion Papers on Securitization of Risk: www.casact.org/pubs/dpp/dpp99/99dpp273.pdf

Palmer, J. M. (n.d.). *Increased Limit Ratemaking for Liability Insurance.* Retrieved July 17, 2012, from Casualty Acutarial Society: http://www.casact.org/library/studynotes/palmer.pdf

Panjer, H. H., & Willmot, G. E. (1992). *Insurance Risk Models.* Society of Actuaries.

Parodi, P. (2012). Computational intelligence with applications to general insurance: a review. II. Dealing with uncertain knowledge. *Ann. Act. Science*.

Partner Re. (2009). *Catastrophe Portfolio Modeling: A Complete View of Risk.*

Sanders, D., & others, a. (2002). The management of losses arising from extreme events. *GIRO.* UK Actuarial Profession.

Shaw, R., & Spivak, G. (2009). Correlations and Dependencies in Economic Capital Models. *GIRO.* UK Actuarial Profession.

Smosna, H. (2012). Tornado / Hail: To model or not to model. *CARe.* Boston: CAS.

Sundt, B., & Vernic, R. (2009). *Recursion for Convolutions and Compound Distributions with Insurance Applications.* Springer.

Swiss Re. (2012). *Natural catastrophes and man-made disasters in 2011: historic losses surface from record earthquakes and floods.* Swiss Re.

Wikipedia. (2012). *Inequality of arithmetic and geometric means*. Retrieved June 12, 2012, from Wikipedia: http://en.wikipedia.org/wiki/Inequality_of_arithmetic_and_geometric_means

Woo, G. (1999). *The Mathematics of Natural Catastrophes.*

Woo, G. (2011). *Calcualting Catastrophe.* Imperial College Press.

Woolstenhulme, M., & Major, J. (2011). Model Risk in Financial Systems: The Need for Robust Decision Control. *CAGNY Fall 2011 Meeting*, (p. 50).