

## **CORRELATION, SIMULATION AND UNCERTAINTY IN CATASTROPHE MODELING**

Dag Lohmann  
Feng Yue

Risk Management Solutions  
7575 Gateway Boulevard  
Newark, CA 94560, USA

### **ABSTRACT**

The science of catastrophic risk modeling helps people to understand the physical and financial implications of natural catastrophes (hurricanes, flood, earthquakes, etc.), terrorism, and the risks associated with changes in life expectancy. As such it depends on simulation techniques that integrate multiple disciplines such as meteorology, hydrology, structural engineering, statistics, computer science, financial engineering, actuarial science, and more in virtually every field of technology. In this paper we will explain the techniques and underlying assumptions of building catastrophe models end to end. We especially will pay attention to correlation (spatial and temporal), simulation and uncertainty in each of the various components in the development process. Our Next Generation Simulation Platform will enable clients (insurance and re-insurance companies) to look at their catastrophe risks objectively while at the same time being able to make use of an open model architecture that opens up the infrastructure of risk modeling.

### **1 INTRODUCTION**

Insurance is a tool to protect individuals and companies from the small probability of a large unexpected loss. Engineering catastrophe models have been created to quantify this risk for a portfolio of subjects at risk (SaR, e.g. properties, cars, etc.). These models are also typically used to optimize a portfolio, to develop underwriting guidelines, for risk transfer, and for catastrophe bonds etc. A complete model typically consists out of the following four different modules (Dong 2001, see also Figure 1):

1. A stochastic model that randomly generates with various sizes, locations and frequencies
  - a. Tropical Cyclones: stochastic track model
  - b. Flood: precipitation, snowmelt, dam breaks
  - c. Earthquake: faults
2. A hazard model that generates local severity
  - a. Hurricane: peak wind gust
  - b. Flood: flood depth
  - c. Earthquake: attenuation, ground shaking
3. A vulnerability model that estimates the loss given the local severity
4. A financial model that calculates the flow of money between two parties, but can also compute industry and economic losses

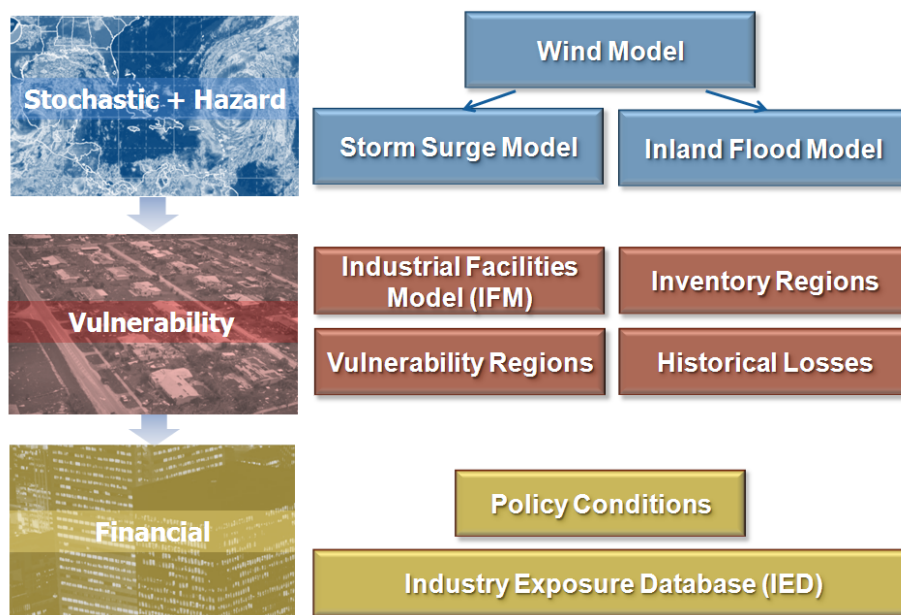


Figure 1: The building blocks of a catastrophe model (tropical cyclone + flood)

A typical question for an insurer is: What is the 100 year return period loss (defined here as the loss that gets exceeded on average once in 100 years) for a single location or a whole portfolio? To answer this question we need to build all the components of a catastrophe model with all the inherent features of nature. Important guiding principles are:

1. We need to model at least on the order of 10,000 scenarios for the next year in order to have locally stable results for a 100 year return period loss. Given that most computational models of the atmosphere and land surface (WRF 2006) are based on hydrodynamic equations, this poses a significant challenge to choose the correct models for the task. E.g. we need to run continental flood models on 100m grids for at least 10k years. This is a significantly larger computational task than currently undertaken in all ensemble prediction models run in weather agencies around the world.
2. Different perils (wind, flood, earthquake) have different spatial and temporal scales of the underlying physical phenomena. They can range from small scales (e.g. tornados, severe convective storms, flash floods) to large (tropical cyclones, earthquakes, floods from snow melt). All these driving variables need to be included to have a complete risk model. We describe the resulting correlation of these driving variables as primary correlation.
3. Some perils exhibit complex spatial and temporal correlations, such as persistence of large scale soil moisture anomalies from previous flooding, or large scale driving variables such as sea surface temperature. This imposes strict constraints on the temporal correlations, within a peril and in between perils.
4. Not all SaR in a portfolio will be affected by a catastrophic event the same way. Subjects at risk that are near to each other will have a much higher correlation in their damage ratios than those further apart (Dong 2001). Part of this primary correlation can be explained by the features of the underlying hazard model (e.g. steep gradients). However, given the same hazard two otherwise identical SaR might have different observed losses, since each SaR has a loss probability distribution around the mean loss. But the uncertainty around the mean loss can also be correlated or uncorrelated, depending on features such as construction type, distance, etc. The aggregation of risk for a portfolio must then take into account this correlation of uncertainty, sometimes referred to as correlation of secondary uncertainty.

A software system that wants to take all of the above features into account also needs follow design criteria, such as group-ability of losses across all possible portfolios and repeatability. In the following chapters we will present the basic building blocks of a catastrophe model in more detail. The rest of the paper has the following structure. Section 2 describes the basic building blocks of a continental flood risk model that is coupled to a tropical cyclone risk model. Section 3 describes some conceptual thoughts about correlation and uncertainty. Section 4 describes how a detailed simulation location level loss model can be build. Section 5 describes the resulting exceedance probability curve.

## **2 BUILDING A FLOOD CATASTROPHE MODEL**

### **2.1 Tropical Cyclones**

The stochastic driver for the example model of this paper is a tropical cyclone track model over China and Hong Kong. China is one of the largest economies in the world with its nominal GDP at 5.88 trillion U.S. dollars (2010 estimate). The country is exposed to typhoons along its eastern coast where some of China's most economically developed areas are located, for example, the Yangtze River delta Economic Zone (centered at Shanghai) and the Pearl River Delta Economic Zone (centered at Shenzhen and Guangzhou). Many major provinces with large populations are at risk from both wind and flood impacts from these typhoons.

Typhoons are a class of tropical cyclones, and all tropical cyclones are intense, cyclonic wind systems that develop over tropical waters. Tropical cyclones develop from tropical depressions, which are non-frontal, closed circulation regions of localized low-pressure in the presence of favorable environmental conditions, such as unstable atmosphere, high humidity, and warm sea-surface temperatures.

The structure of a typhoon is distinguished by the eye, which is surrounded by high walls of thick clouds and rotating winds. Typhoon winds normally increase in velocity toward the center of the storm. However, wind velocity is rarely symmetrical around the storm track. Generally, for the tropical cyclones of the Northern Hemisphere in which the circulation is rotating counter-clockwise, the right side of the storm (as viewed in the same direction as the typhoon's forward motion) experiences higher velocity winds than the left side. In addition to intense winds, typhoons often create heavy precipitation and typically range from 100 to 500 km in diameter. The farther north a typhoon moves, the more likely it is to encounter the mid-latitude jet-stream. When this occurs, the system becomes more asymmetric with strong winds largely on the right side of the system, covering a broader region than in a typhoon. These systems are undergoing extra-tropical transition (Wiki 2011) and are known as transitioning storms.

When typhoons make landfall, they can cause loss of life, extensive damage to buildings and their contents, and other infrastructure, as well as crop and forest destruction and water contamination.

The east coast of China is exposed to typhoon risk from April to December. The most active months for typhoons are July to September when the Northwest Pacific basin experiences its peak sea-surface temperatures, which fuel storm formation.

### **2.2 Stochastic Modeling**

We randomly generated over 24,500 unique stochastic events from historical data. The historical event data came from the China Meteorological Administration (CMA) catalogue of storms from 1945 to 2009. The resulting stochastic storms are distributed among 117 coastal gates and are modeled through their extra-tropical transition and inland decay. Figure 3 shows the stochastic tracks and an example wind profile. The size and shape of wind fields are generated using an analytical wind profile that is taken from a methodology originally developed at the Boundary Layer Wind Tunnel, University of Western Ontario, Canada (Georgiou 1985). The wind speed is calculated from the formula relating the site location relative to the storm track, the landfall location, and the physical parameters of the storm. At any given point in time and space, the 2-minute mean wind (equivalent over water) is entirely prescribed by the position

from the storm center and the following set of parameters: maximum wind ( $V_{max}$ ), radius of maximum wind ( $R_{max}$ ), four shape parameters giving the radial profile inside and outside the eyewall, the angle between the location of the maximum winds and the track, and four additional parameters that reduce the variance between observed and modeled wind fields.



Figure 2: China Typhoon Modeling Region

$R_{max}$  are simulated from a regression model with central pressure and latitude as predictors and autoregressive errors, following Vickery *et al.*(2000). The  $R_{max}$  model is fit to observations available in the Cooperative Institute for Research in the Atmosphere (CIRA) at Colorado State University (CSU 2010). For tropical storms, all other wind field parameters have been fit to the observations from 26 historical tropical cyclone events supplied by the CMA from 1975 onwards.

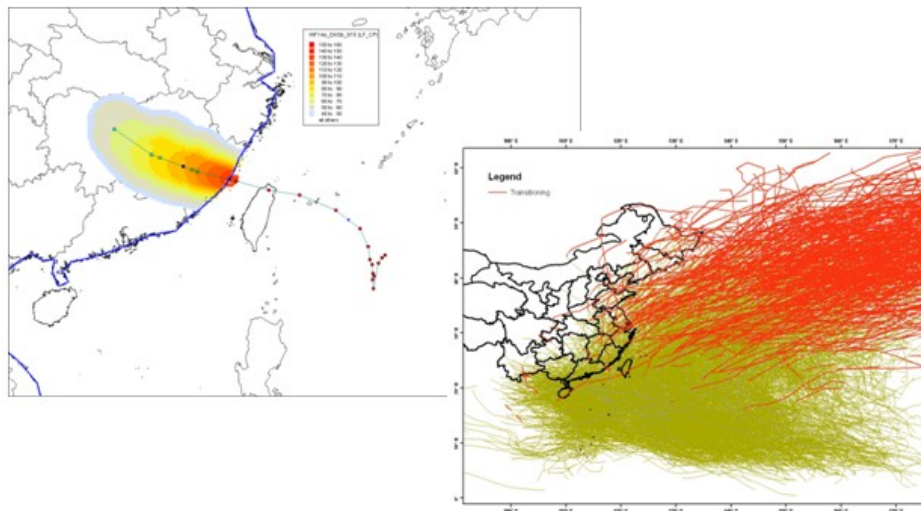


Figure 3: China Typhoon wind profile and track set. The track set forms the basis of the probabilistic model, since it contains already the important information about frequency and severity for the wind model. The flood model is driven by an added precipitation model.

### 2.3 Hazard Modeling

The core hydrological / hydraulic model features are highlighted in Figure 4. Essentially we have built a precipitation, hydrological and hydraulic model that is computational effective, but still state-of-the-art complex for the most important physical characteristics (Lohmann et al. 1999, Grieser and Jewson 2011).

#### Model Features

- China = 0.5 Billion 90m grid cells for distributed hydrological modeling
- Split up into ~13k catchments (each ~ 500km<sup>2</sup>)
- Atmospheric forcing with 3h time steps for rainfall in calibration period
- Calibration with Genetic Algorithm
- Stochastic model driven by track model coupled to RMS TC precipitation model (24k tracks)
- Snow model
- Runoff ratio based on 1k years run
- 1-d routing model with 5min time steps on 500m segments
- Direct input of 90m runoff 2-D flood plain model
- Model runs on 90m DTM, output will be nested dynamic VRG between 100m and 5km.

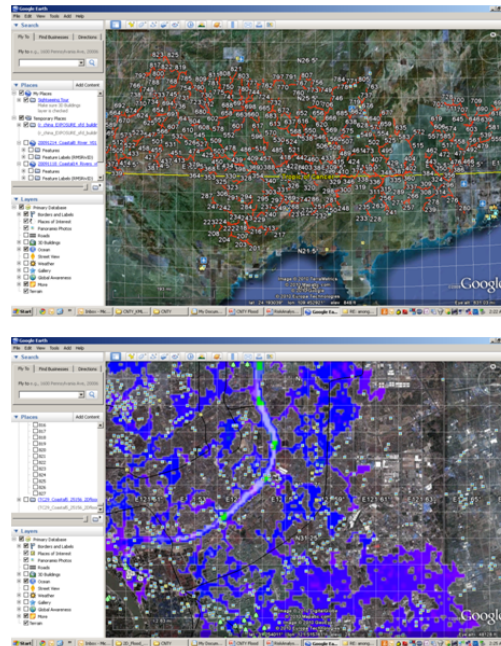


Figure 4: Features of a continental flood model driven by tropical cyclones (TC). Atmospheric forcing contains all the variables of the lower atmosphere (pressure, wind speed, radiation, precipitation, humidity). Output of this model is on a variable resolution grid (VRG).

The stochastic event module produces simulated inland flood events that are used in the probabilistic loss analysis. Each event is described by its frequency of occurrence derived from the wind module (not presented here). Due to the importance of antecedent conditions, we ran a continuous simulation of the rainfall and the rainfall-runoff models over 1,000 years as background events (without tropical cyclones) and used the resulting initial conditions for TC simulations. This ensures that realistic flood events are selected by the model; similar rainfall events can give very different flooding patterns, dependent on the soil moisture conditions and river levels prior to rainfall. Rainfall of a given intensity may or may not result in flooding, depending on the antecedent conditions. Figure 5 shows a historical reconstruction of a precipitation event.

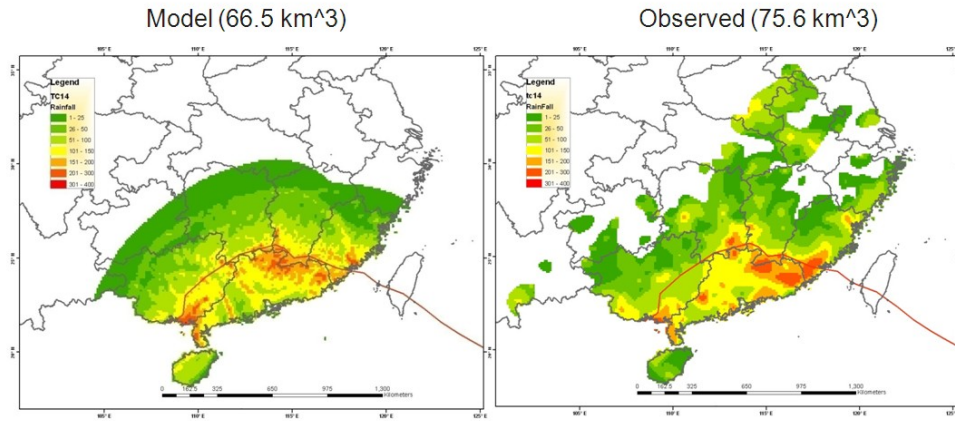


Figure 5: Modeled and observed precipitation patterns. The modeling methodology is described in detail in Grieser and Jewson (2011).

Important features of the hydrological model are a detailed calibration to historical data with a genetic algorithm. Each of the model parameters reflects the typical behavior of an individual component of the model, such as groundwater response, surface evapotranspiration, and snowmelt.

A state-of-the-art evolutionary multiobjective optimization algorithm, NSGAII (Deb et al. 2002), is adopted for calibration (Figure 6). Based on our study and objectives, the optimum parameter for NSGAII is as follows: population size = 20, mutation probability = 1/number of model parameter, crossover probability = 1.0, maximum number of runs = 1000, number of generation = 50.

For each parameter combination, the hydrological model is run, and the modeled runoff rates for the entire catchment are compared to observed rates. The performance behavior is determined in terms of two objective functions as defined by:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{obs,i} - Q_{mod,i})^2}$$

$$RMSE_{highflow} = \sqrt{\frac{1}{n_h} \sum_{i=1}^{n_h} (Q_{obs,i}^h - Q_{mod,i}^h)^2} \quad \text{with} \quad Q_{obs}^h : Q_{obs} \geq Q_{obs}^{80\%}$$

Analogous to Darwinian natural selection, NSGAII uses a population-based search in which high quality solutions (according to the two objective functions defined above) are evolved using selection, mating, and mutation. After 50 generations of evolution, the optimum population (best parameter set) are formed, which is called Pareto front (not shown here). For it is very important to capture the flood peak (or high flow), the parameter combination on the end of RMS\_highflow is selected as the final optimal paramters.

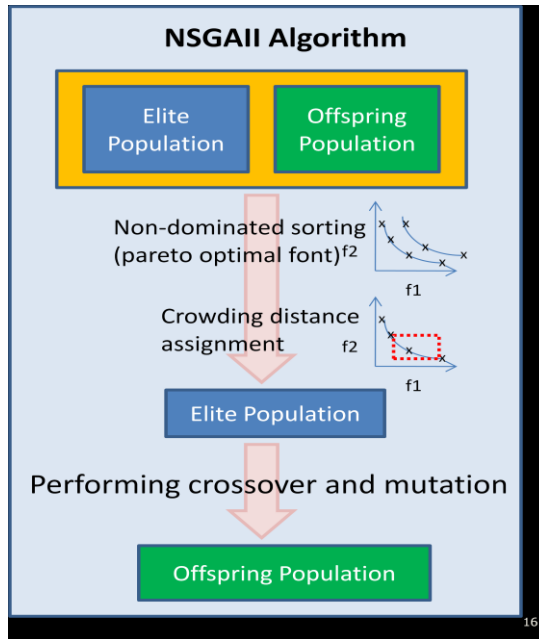


Figure 6: Simple illustration of NSGAI Algorithm

The calibration of the rainfall runoff model (RRM) is conducted for all basins separately using monthly data and published peak flow (whenever available). Typically, larger basins have a better model performance because runoff rates are, to a lesser degree, influenced by local phenomena, such as groundwater abstractions, and storage effects of weirs and lakes. Moreover, localized rainfall events, which are normally difficult to capture in the input data, do not determine the runoff regime of a large basin to such a degree as they do for small basins. The larger the basin is, the more these local effects tend to average out, which makes the calibration more stable and more reliable. Typically our calibration resulted in higher  $R^2$  values for larger rivers. One example is shown in Figure 7.

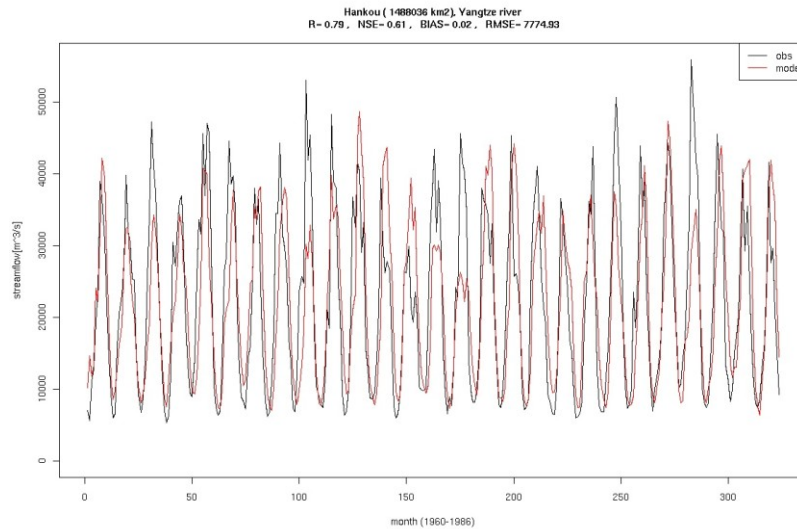


Figure 7: Example illustration of calibration results in Hankou along Yangtze River

The resulting computed water volumes are then used to flood large scale areas using hydraulic models (not shown). Figure 8 shows the resulting flood patterns.

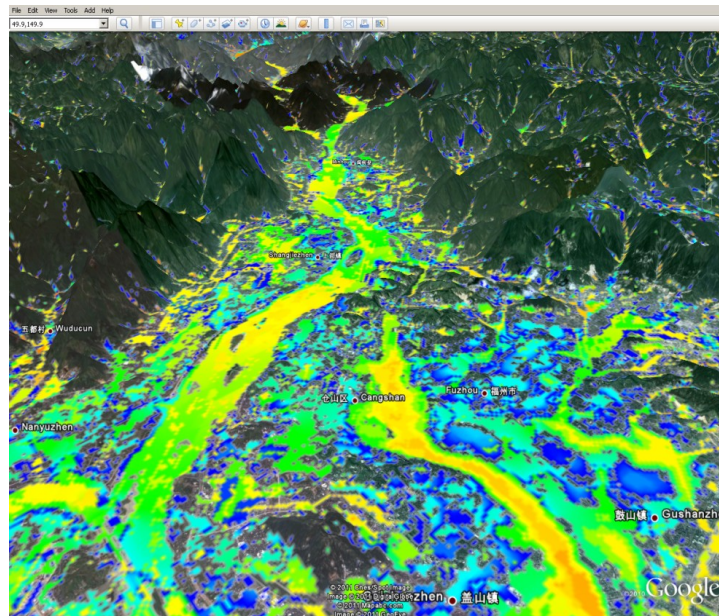


Figure 8: Flood pattern from one event. A complete event set would be comprised out of thousands of possible different outcomes given different precipitation events and different initial conditions of the soil. Each of the tracks from figure 3 produces one flood pattern that then can be used to create these flood footprints.

### 3 SOME CONCEPTUAL THOUGHTS ABOUT CORRELATION AND UNCERTAINTY

The concept of uncertainty and dependence or correlation in catastrophe models is normally not fully transparent to the end user. Typically the following types of uncertainty are considered:

- 1) *Rate or primary uncertainty* relates to whether a random hazard event will occur or not. Each event is typically assigned a location, size or magnitude and a rate. These events then have some primary correlation in the hazard fields (e.g. the tropical cyclone wind field) and in the resulting expected loss (the mean damage ratio (MDR)).
- 2) *Loss or secondary uncertainty* is the uncertainty of the loss given an event from 1). These uncertainties are often described as parametric distributions around the mean (Dong 2001). The correlation of secondary uncertainty then describes how losses are correlated around the MDR amongst different locations.
- 3) *Parameter uncertainty* can be quantified through e.g. calibration and sensitivity studies similar to the one shown in chapter 2. In principal parameter uncertainty can be reduced by increasing the knowledge about the system under investigation.
- 4) *Process and model uncertainty* can describe missing hazard sources, insufficient knowledge about the accuracy of the modeling process (unknown distributions, etc.). It therefore can have a significant impact on the correlation of the losses itself.

The first two categories describe how models typically describe uncertainty (Dong 2001). The uncertainties in 3) and 4) can be either grouped in aleatory or epistemic uncertainties, often dependent on philosophical arguments.



While some of these uncertainties can be quantified, others are impossible to know. In reality modelers select two or three alternative models and look at the different outcomes. This can be seen as some outer primary uncertainty loop over different model assumptions. Most of the resulting uncertainty is then expressed through secondary uncertainty in the vulnerability functions and the correlation of it. This can be seen as some inner secondary uncertainty loop around the mean damage ratio. It is this inner loop that we will perform our simulations of correlated losses in chapter 4 on. Figure 9 shows hurricane damage uncertainty for single-family wood frame structures as a function of the wind speed.

While a comprehensive uncertainty treatment would involve all part of a catastrophe model, this is sometimes not feasible given the computational demands of these models.

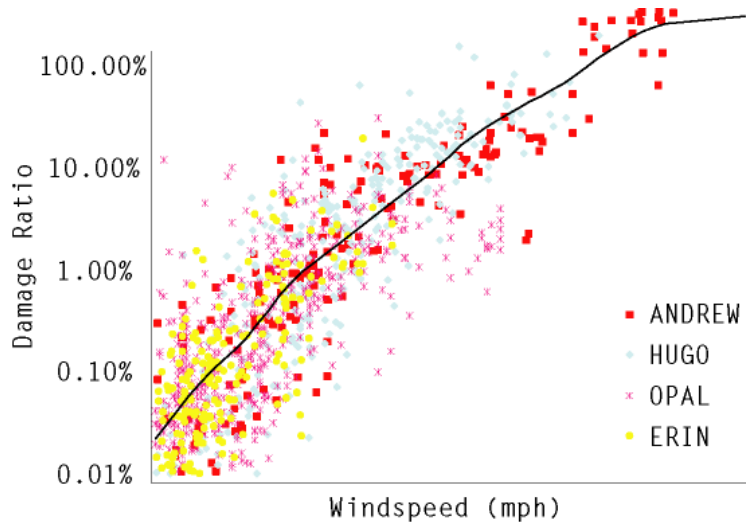


Figure 9: Hurricane Damage Uncertainty for Single-family Wood frame Structures. The expected loss (or mean damage ratio (MDR)) is the black line.

## 4 DETAILED LOSS MODELING

### 4.1 Overview

After an economic database of building values (not shown in this paper) and vulnerabilities (Figure 9) is built, we can do detailed loss modeling. Catastrophe models typically describe the outcome of the primary loss simulation in an event loss table (ELT). The ELT has for each event  $i$  and each location  $j$  an entry that lists the mean damage ratio ( $MDR_{ij}$ , here wind or flood cause these) and the distribution around it (the secondary uncertainty), often described parametrically by a beta distribution through the CV. We can then sample for each event multiple occurrences given that we can describe the correlation of secondary uncertainty with a simple model.

### 4.2 Severity Sampling

When a portfolio of insured SaR is modeled the sum of losses from all the properties as well as the correlation among these properties need to be considered. Obviously, no analytical method is suitable for this. Instead, the Monte-Carlo simulation method provides a straightforward solution.

In order to obtain loss samples  $X_j$  from the loss distribution of event  $i$  of the modeled portfolio with the use of Monte-Carlo simulation method, the correlation among the severities of properties  $j$  from the modeled portfolio needs to be modeled in the first place. In the real applications for client portfolios with hundreds of thousands insured properties the pair-wise correlations can make a very large correlation matrix that is difficult to decompose efficiently. For this reason, we created for illustration purposes a simple

one factor normal copula model (Hull and White 2005) to model a flat correlation matrix. The steps for simulation are the following:

- 1) Determine the linear or rank correlation of secondary uncertainty from observed losses (or engineering judgment)
- 2) Generate multivariate standard normal distributed random variables  $N_{ij}$  from Equation (1). In this equation,  $\rho_{global}$  determines the factor loading (the eigenvalues of the co-variance matrix or constant pair wise linear correlations between any two locations) for the one factor  $F_i$  in the correlation model and the residual factor  $\varepsilon_{ij}$ . If  $F_i$  and  $\varepsilon_{ij}$  are  $N(0,1)$  distributed, then  $N_{ij}$  is  $N(0,1)$  distributed.

$$N_{ij} = \sqrt{\rho_{global}}F_i + \sqrt{1 - \rho_{global}}\varepsilon_{ij} \quad (1)$$

- 3) Compute  $U_{ij} = \Phi^{-1}(N_{ij})$  where  $\Phi$  is the standard normal distribution function to get the loss quantiles
- 4) Compute  $X_{ij} = F^{-1}(U_{ij})$  using the pseudo inverse of the marginal distribution functions (e.g. beta)

The resulting loss distributions have the desired MDR and correlation structure. While these severity samples for a complete portfolio are all taken from the basic event set (just MDRs), they conceptually expand the risk surface for each SaR and through correlation of secondary uncertainty also for the entire portfolio (see chapter 5).

### 4.3 Frequency Modeling

Frequency modeling here refers to trying to preserving the rate of the original event set from which we take severity samples. These modeled event rates are typically on the order of 1/1000years to 1/1.000.000 years. In principal there are two different ways of how to do this.

- 1) one could specify an analytical distribution for the number of occurrences, e.g. Poisson or Negative-Binomial. This is a simple solution. Each severity sample would just have a rate of  $\lambda_i / n$ , where  $n$  is the number of samples. This solution has some fundamental shortcomings for the calculation of losses for an insurance portfolio, e.g. annual aggregate deductibles couldn't be computed.
- 2) one could produce a year event table (YET) in which one places the simulated severity events. E.g. we could choose to model 1 Million years. However, the rate of the original events would potentially not be preserved and therefore we might introduce significant errors in our simulation. However, this kind of simulation enables us to do much more than in 1). We can now easily specify correlations between different perils on different parts of the world. We can also compute any sort of annual deductible, etc.

### 4.4 Results

Risk models typically display their results in the form of exceedance probability (EP) curves. The insurance and reinsurance industry uses these curves to decide what cover to buy or to offer to achieve and acceptable level of risk and what premium to charge or pay (Dong 2001). These curves can also be used for case studies of varying deductibles and limits on single policies.

Based on the previous chapters, Figure 10 shows the economic EP loss curve for China due to wind plus flood losses after sampling and simulation.

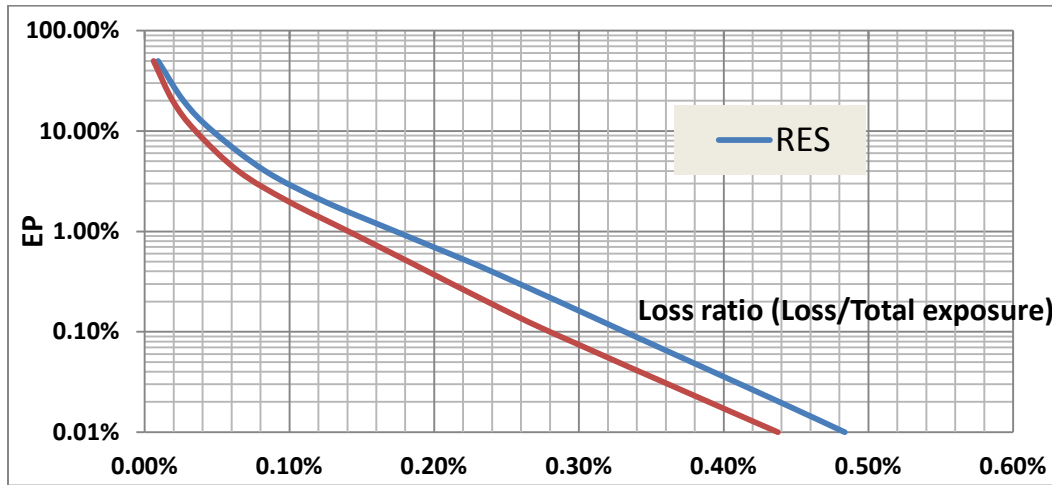


Figure 10: Loss results for China Typhoon are shown as exceedance probability as a function of the loss ratio (defined as the loss divided by total exposure). RES stands for residential exposure.

The effect of the correlation of secondary uncertainty is quite pronounced. Figure 11 shows simulated results from a hypothetical portfolio for 6 different levels of linear correlation ( $\rho_{global}$  in Equation 1). The results show that with higher return periods (lower EP) the portfolio results have larger losses. This results from wider portfolio loss distributions when the correlation is higher.

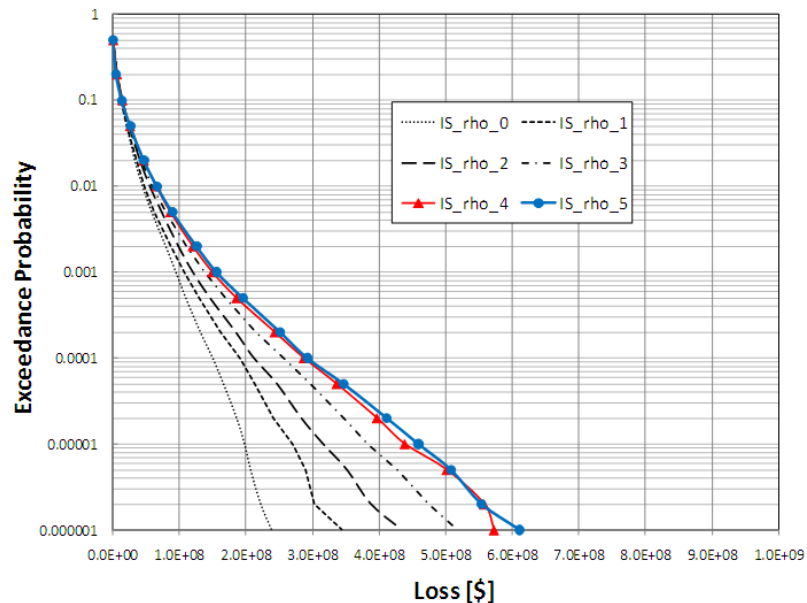


Figure 11: Impact of the correlation of secondary uncertainty on EP loss results. Larger correlation values lead to broader portfolio loss distributions and therefore to higher losses in the tail. This is compensated by lower losses at low return periods (not visible), since the average annual loss for the portfolio is independent of correlation.

## 4.5 Summary

In this paper, we presented a Monte-Carlo simulation method applied for detailed loss modeling for natural catastrophes that includes a simple model for secondary uncertainty correlation. We first built a physical model of the underlying natural phenomena (wind and flood). Building these models requires a detailed understanding of the underlying physical phenomena as well as computationally efficient methods to compute small scale hazard values for flood on a continental scale application. Afterwards we sampled loss distributions with the help of a one factor model to gauge the impact of correlated secondary uncertainty. The final result is a country wide EP loss curve that than can be used to make financial decisions.

## ACKNOWLEDGMENTS

The authors would like to thank their colleagues at RMS for their inspiration and collaboration.

## REFERENCES

- Dong, W. 2001. *Building a More Profitable Portfolio. Modern Portfolio Theory with Application to Catastrophe Insurance*. Reactions Publishing Group, London.
- WRF. 2006. Research-Community Priorities for WRF-System Development, WRAB, December 2006, <http://www.wrf-model.org/wrfadmin/publications.php>
- Wiki. 2011. Wikipedia, accessed July 2011: [http://en.wikipedia.org/wiki/Extratropical\\_cyclone](http://en.wikipedia.org/wiki/Extratropical_cyclone)
- Georgiou, P. N. 1985. "Design wind speeds in tropical cyclone-prone regions." PhD thesis, Fac. of Eng. Sci., University of Western Ontario, London, Ont., Canada.
- CSU. 2010. Colorado State University accessed July 2010: [http://rammb.cira.colostate.edu/products/tc\\_realtime/index.asp](http://rammb.cira.colostate.edu/products/tc_realtime/index.asp).
- Vickery, P. J., P. F. Skerj, and L. A. Twisdale, 2000. "Simulation of hurricane risk in the U.S. using empirical track model." *Journal of Structural Engineering*, **126**, 1222–1237.
- Lohmann, D., E. Raschke, B. Nijssen and D. P. Lettenmaier. 1998. Regional scale hydrology: 1 Formulation of the VIC-2L model coupled to a routing model. *Hydrological Sciences Journal*, vol. 43, no. 1, 1998
- Grieser, J., Jewson, S.. 2011. The RMS TC-Rain Model. *Meteorologische Zeitschrift*, in print. 2011.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. 2002. "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA. II." *IEEE Transactions on Evolutionary Computation*, **6** (2), APRIL 2002
- Hull, J. and White, A.. 2005. "The Perfect Copula." Rotman School of Management University of Toronto. First Draft: June 2005. Third Draft: July 2005.

## AUTHOR BIOGRAPHIES

**Dag Lohmann** is Vice President of Model Development at Risk Management Solutions in Newark, CA. He has been leading for 8 years the development of continental scale flood risk models in RMS, and recently the development of RMS' Next Generation risk modeling methodology. Before moving to RMS he worked for 6 years for the National Weather Service, NOAA/NCEP/EMC in Camp Springs, MD, where his main interest was data assimilation, forecasting and hydrological modeling. He received a Physics Diploma (Masters) from the Georg-August University in Goettingen (Germany) and a Ph.D. from Hamburg University (Germany) before working for 2 years as a postdoc at Princeton University. He received the 1999 Tison Award of the IAHS and has published numerous papers on hydrological modeling, model uncertainty, forecasting, data assimilation, and climate change. His email address is [Dag.Lohmann@rms.com](mailto:Dag.Lohmann@rms.com).

**Feng Yue** is a senior risk modeler in Model Development of Risk Management Solutions. Based in Newark, CA, Feng works on the development, implementation and application of RMS Next Generation de-

tailed loss model methodology and software. Prior to joining RMS, Feng worked as a hydraulic and hydrologic modeler for the Illinois State Water Survey on evaluating channel restoration measures and developing flood insurance maps. He has published a couple of scientific papers and reports in the fields of groundwater, hydrology and hydraulics. Feng has both MS and PhD in Civil Engineering from the University of Illinois at Urbana-Champaign and a BS in Water Resources from Zhongshan University, PR China. His email address is Feng.Yue@rms.com.