UNCERTAINTY MEASURES IN CATASTROPHE RISK MODELING

RMS® CCRA® Training Program

Certified Catastrophe Risk Analyst

 November 7, 2019

AGENDA

- **Unit 1: Types of Uncertainty Overview**
- Unit 2: Parameter Uncertainty
- Unit 3: Secondary Uncertainty
- Unit 4: Multiple Locations
- Unit 5: Including Secondary Uncertainty in the EP Curve
- Unit 6: Volatility Measures
- Unit 7: Correlation Between Portfolios

 November 7, 2019 2

This is the agenda for the Uncertainty Measures course. In addition there are three exercises to complete. Units 1 through 3 start with an overview of the types of uncertainty, including primary, secondary, and parameter uncertainty. Separate units then proceed to provide more details on parameter uncertainty and secondary uncertainty. The unit on secondary uncertainty includes discussion of the main sources of secondary uncertainty: hazard, vulnerability, and exposure data. We now start with an overview of the types and sources of modeled loss uncertainty.

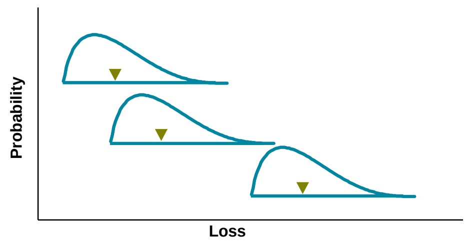
UNIT 1 – LEARNING OBJECTIVES

- Types of Uncertainty Overview
  - Define the types of uncertainty included in catastrophe risk analyses.
  - List the three main sources of uncertainty in catastrophe risk modeling.
  - Provide an example of how each type of uncertainty is characterized.

At the end of this unit you should have a concrete understanding of each of the three learning objectives listed on this slide.

TYPES OF UNCERTAINTY

- **Primary Uncertainty**: The uncertainty of which event, if any, will occur and the size of the event as measured by the mean loss.
- **Secondary Uncertainty**: The uncertainty in the amount of loss, given that a certain event has occurred.
- **Parameter Uncertainty**: The uncertainty in model parameters – some parameter uncertainty is included in secondary uncertainty.



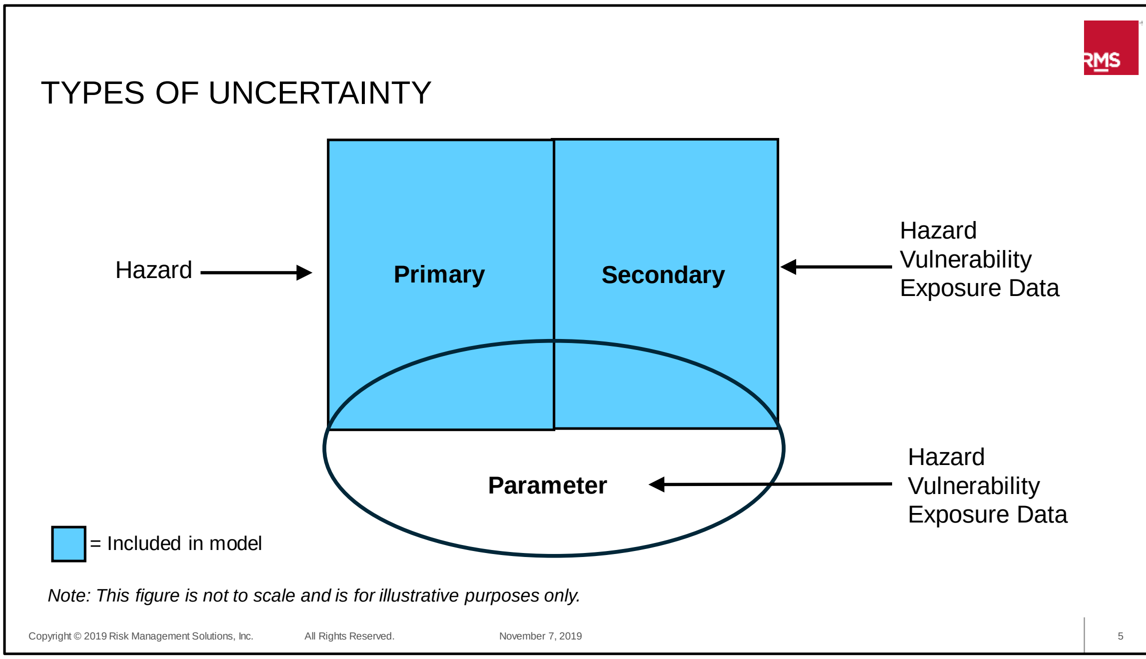
 November 7, 2019 4

There are three types of uncertainty in modeling.

Primary uncertainty is the uncertainty about which events will occur and their typical sizes. The measurements of primary uncertainty are the event rate and its mean loss. In the graph on the slide, the primary uncertainty is shown by the placement of the green triangles, which represent potential events, the probability of their occurrence, and the expected loss if one of them should occur.

Secondary uncertainty is the uncertainty in the amount of loss a given event can cause when it occurs. This is represented on the graph by the distribution around the green triangles for each of the events.

Parameter uncertainty is the uncertainty in the event parameters. Currently in RMS applications, a portion of the parameter uncertainty is modeled, and a portion is not modeled for both application performance and state of scientific knowledge reasons.



This graphic shows how the different types of uncertainty fit together. The blue rectangle represents the portion of the uncertainty that is modeled. We split this modeled uncertainty into primary and secondary uncertainty. Note that the oval, representing the parameter uncertainty, overlaps the blue rectangle, including both the primary and secondary portions. This means that part of the modeled primary uncertainty and part of the modeled secondary uncertainty is caused by parameter uncertainty. The majority of the parameter uncertainty is not modeled. The main sources of uncertainty are the unknown information about hazard, vulnerability, and exposure data.

VARIOUS UNCERTAINTY AND VOLATILITY MEASURES USED IN RISKLINK®

- Primary uncertainty
  - Event mean losses
  - Event rates
- Secondary uncertainty
  - Event standard deviations
- Overall volatility of the analysis
  - Standard deviation (σ) and CV (= σ/AAL) around the AAL in the RiskLink user interface
  - Incorporates both primary and secondary uncertainty

 November 7, 2019 6

RMS recognizes primary uncertainty in the respective country/peril models. For example, the primary uncertainty around Europe Windstorm is characterized by the event rates for each event in the Europe Windstorm model and the associated impact of the model on a book of business, as shown by each event's mean loss.

Secondary uncertainty is recognized through the event standard deviation. To get an understanding of the overall volatility of the analysis, which includes both primary and secondary uncertainty, RMS uses the standard deviation of the annual loss distribution. To compare the volatility between modeled results of portfolios, RMS uses the coefficient of variation (CV). The CV is defined as the standard deviation divided by the average annual loss (SD/AAL or σ/μ). Refer to Unit 4 of the Financial Modeling course for detail on the definition and calculation of uncertainty and volatility measures.

AGENDA

- Unit 1: Types of Uncertainty Overview
- **Unit 2: Parameter Uncertainty**
- Unit 3: Secondary Uncertainty
- Unit 4: Multiple Locations
- Unit 5: Including Secondary Uncertainty in the EP Curve
- Unit 6: Volatility Measures
- Unit 7: Correlation Between Portfolios

 November 7, 2019

This unit provides more detail on parameter uncertainty.

UNIT 2 – LEARNING OBJECTIVES

- Parameter Uncertainty
  - Define parameter uncertainty
  - Provide sources of parameter uncertainty
  - List examples of parameter uncertainty for different perils

 November 7, 2019

At the end of this unit you should have a concrete understanding of each of the three learning objectives listed on this slide.

SAMPLE PARAMETERS

- Earthquake hazard
  - Magnitude
  - Ground motion parameters (e.g. spectral acceleration)
- Hurricane hazard
  - Central pressure
  - Wind speed
- Vulnerability
  - Inter-story drift
- Exposure data
  - Construction type
  - Insurance to value

 November 7, 2019

Catastrophe risk analysis parameter uncertainty describes the uncertainty in the actual values that define specific catastrophe risk modeling parameters. Event parameter values in a catastrophe risk model can be characterized as single estimates, average values, or distributions around estimated or average values. This type of uncertainty is caused by measurement errors, analytical imprecision, and limited sample sizes during the collection and treatment of data. Parameter uncertainty is thus reducible through further study or experience. This is often termed as a combination of epistemic uncertainty, resulting from an inadequate understanding, and aleatory uncertainty, the intrinsic variability of nature which can only be measured over time.

Shown on this slide are examples of earthquake and hurricane parameters. These include event severity (earthquake magnitude or hurricane peak wind gust) and other event specific parameters such as earthquake ground motion or storm central pressure that help define the peril hazard. Parameter uncertainty also applies to other catastrophe risk modeling modules such as vulnerability. The example listed here is inter-story drift capacity. This is one parameter that influences the vulnerability curves applied to a building during analysis. Finally, parameters associated with exposure data, such as construction type and insurance to value also have their own associated parameter uncertainty.

As was mentioned previously, model and experience constraints require that only a subset of parameter uncertainties be modeled. An example where event parameter uncertainty is currently being modeled in RiskLink applications is for the New Madrid fault zone in the U.S. In this case, a scientifically well-studied region allows for the characterization of uncertainty around earthquake rupture scenarios and thus earthquake magnitudes.

SOURCES OF PARAMETER UNCERTAINTY

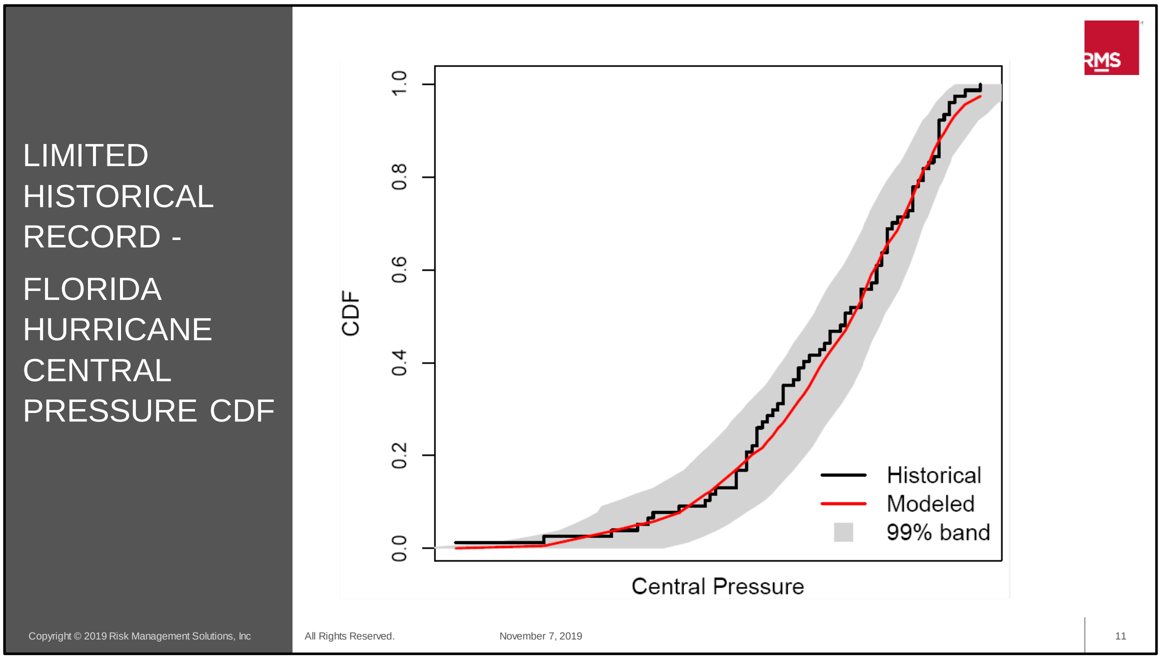
- Sampling error
  - Limited historical record
- Measurement error
  - Especially for older events
- Processes that fall outside the current methodology and could impact future losses
  - E.g. climate change

Three sources of parameter uncertainty are sampling error, measurement error, and processes that are beyond the current state of scientific knowledge.

Sampling error is due to the fact that the historical record is limited. For example, we have about 100 years of historical data for U.S. hurricane. This gives us a relatively small sample size from which to base a model that may undergo multi-decadal oscillations (or variations in the climate state). One strategy employed by RMS to combat the impact of a small sample size is to use supplementary scientific data, such as sea surface temperatures.

Measurement error is caused by the fact that the measurements of the parameters of the historical events might not be completely accurate. This is particularly problematic for older events. For example, 100 years ago, instrumentation was unavailable to measure the central pressure of a hurricane. Today's instrumentation allows for more accurate measurement of a hurricane. This difference in measurement must be recognized in the historical record as measurement uncertainty. Uncertainty is also an additional consideration for smaller events, because people tend to report larger events more than smaller ones. This is lack of data completeness for smaller events may lead to larger parameter uncertainty than for larger events.

A third source of parameter uncertainty is the fact that there may be changes to the assumptions underlying the model. For example, global warming may affect the frequency and severity of windstorms in the future, thus potentially reducing the validity of historical record averages as a predictor of future event parameters.



This slide, and the next four slides, are part of RMS' submission to the Florida Commission on Hurricane Loss Projection Methodology (FCHLPM).

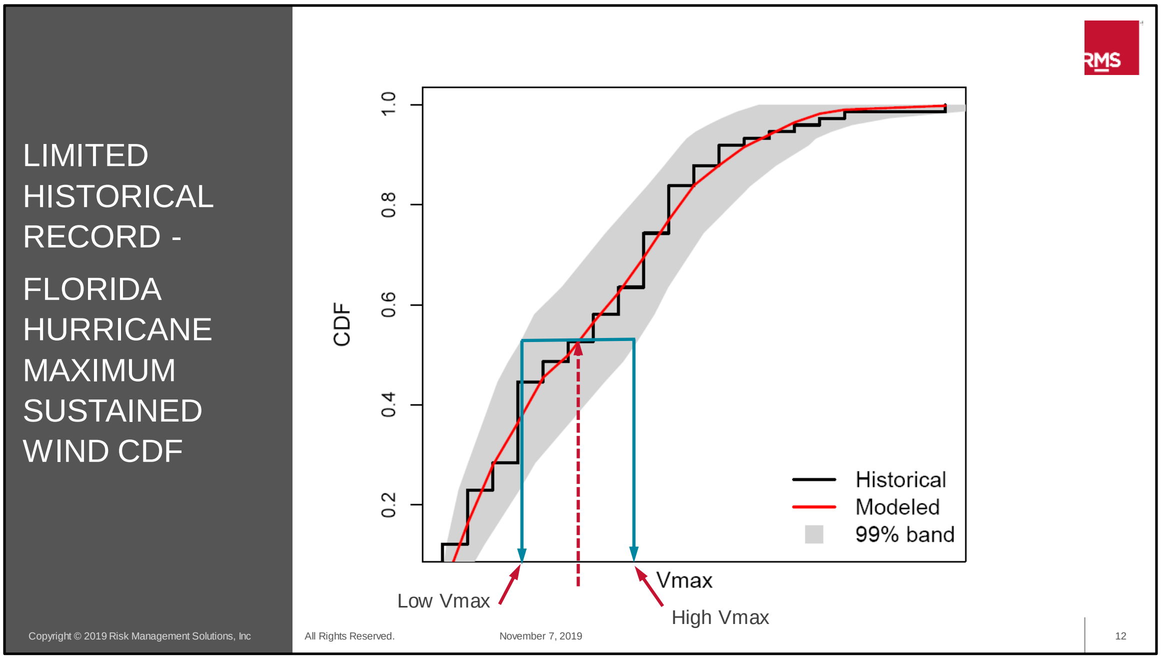
This is an example of parameter uncertainty due to a limited historical record.

One of the parameters in the hurricane model is the central pressure of the hurricane. The probability distribution is based on the historical record, which has a limited number of data points. Because of the limited amount of data, there is some uncertainty in what the "true" probability distribution is.

The black curve shows the empirical (historical) cumulative distribution function (CDF) of the central pressure of Florida hurricanes. The red curve shows the modeled CDF.

The gray shaded region, which was generated by a "bootstrap" analysis[1] of the historical data, defines the 99% confidence band of the modeled CDF. We can be 99% confident that the true CDF of the central pressure of Florida hurricanes lies within this region. Conversely, there is a 1% chance that the true CDF lies outside of this region.

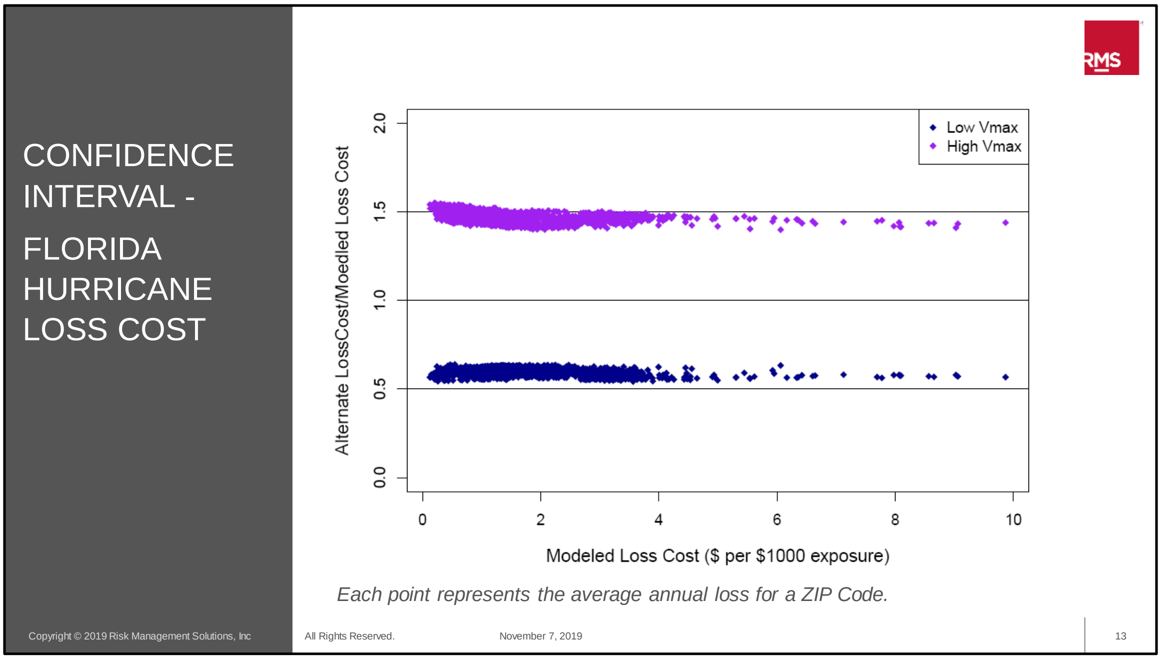
1. A bootstrap analysis is a Monte Carlo-based procedure used to quantify the uncertainty in a quantity that has been estimated from a limited set of data. In this case, the quantity of interest is the CDF of the central pressure, which was based on 73 historical storms. To estimate the uncertainty in the CDF using a bootstrap analysis, the 73 storms are repeatedly sampled <u>with replacement</u> to generate a large number (say 10,000 for the sake of clarity) of new samples, each consisting of 73 central pressures. Note that because we are sampling with replacement, most (if not all) of these new samples will be different from the original set of 73 data points used to estimate the CDF plotted in black. For a given central pressure, each of these 10,000 new samples will provide a different estimate for the CDF. If we sort these estimates from smallest to largest, the grey region in this plot corresponds to the area between the 50^{th} value (50/10,000 = .5%) and the $9,950^{th}$ value (9,950/10,000 = 99.5%).



Similar plots can be constructed for other parameters; for example, the Maximum Sustained Winds (Vmax).

To estimate the effect that the parameter uncertainty, due to the limited sample size, has on the modeled losses, we took each stochastic storm in the event set and looked at the upper and lower bounds of a 90% (99% is shown) confidence band around the stochastic storm's Vmax.

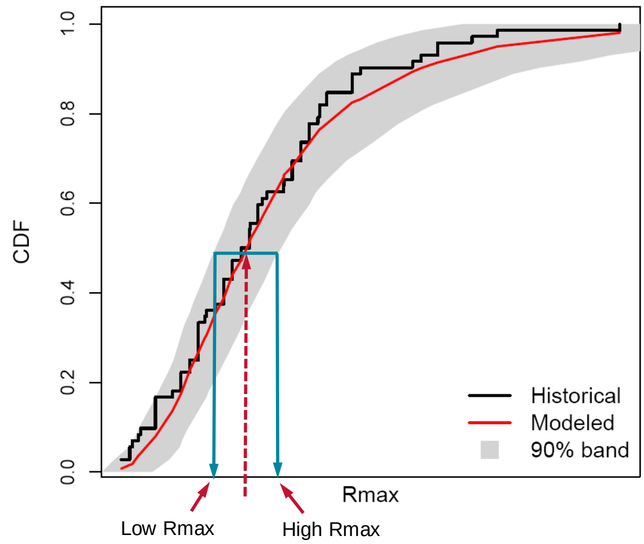
Take, for example, a stochastic storm with a Vmax indicated by vertical red dashed arrow – using the 90% confidence interval (5^{th} and 95^{th} percentile as low and high Vmax respectively) we can define the range of possible Vmax associated with this storm. This is one approach to quantifying uncertainty around Vmax.



Here are the results for two new stochastic storm sets. Each point represents the average annual loss for a ZIP Code. The blue points represent the stochastic storms with the Vmax at the lower bound. The purple points represent the storms with the Vmax at the upper bound. The exposure used in this slide is from the submission to the FCHLPM and represents about 90% of the residential Industry Exposure Database (IED) for Florida.

The vertical axis is the ratio of the new Loss Cost to the Loss Cost from the original modeled stochastic storms. Note that the Loss Cost can increase by more than 50% and decrease by more than 45% due to the limited sample size of Florida hurricanes.

LIMITED HISTORICAL RECORD - FLORIDA HURRICANE RADIUS TO MAXIUM WINDS CDF

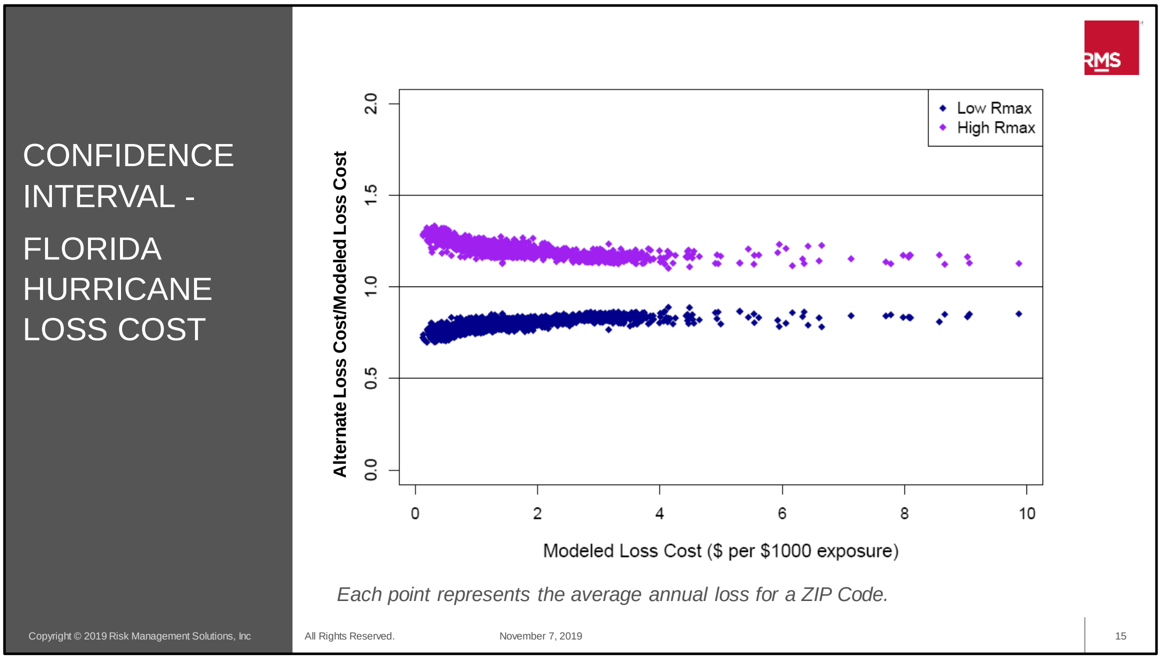


 November 7, 2019 14

We can also look at other parameters, such as the Radius to Maximum Winds (Rmax).

To estimate the effect that uncertainty in this parameter (due to the limited sample size) has on the modeled losses, we again took each stochastic storm in the event set and looked at the upper and lower bounds of a 90% confidence band around the stochastic storm's Rmax.

Let's look at a stochastic storm with an Rmax indicated by the vertical red dashed arrow. Using the 90% confidence interval, we can define the range of possible Rmax associated with this storm, thus quantifying uncertainty around Rmax.

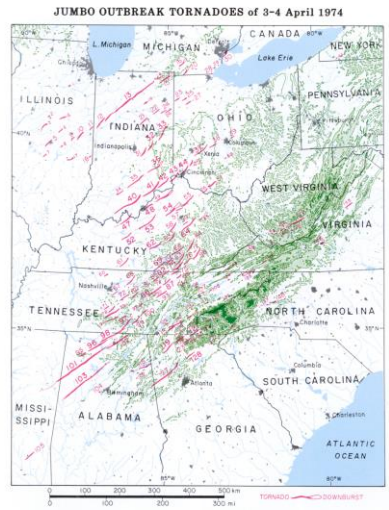


Similar to the previous graph showing the Vmax, here are the results for two new stochastic storm sets with each point representing the average annual loss for a ZIP Code. The blue points represent the stochastic storms with the Rmax at the lower bound, while the purple points represent the storms with the Rmax at the upper bound. The exposure used in this slide is also from the submission to the FCHLPM and represents about 90% of the residential Industry Exposure Database (IED) for Florida.

The vertical axis is the ratio of the new Loss Cost to the Loss Cost from the original modeled stochastic storms. Note that the Loss Cost can increase by more than 25% and decrease by more than 25% due to the limited sample size of Florida hurricanes.

PARAMETER UNCERTAINTY: TORNADO

- Our best understanding of how intensity varies within a tornado path comes from a very limited set of detailed research projects
  - 1974 Super Outbreak
  - Selected major urban tornado events (e.g. May 3, 1999, Oklahoma City)

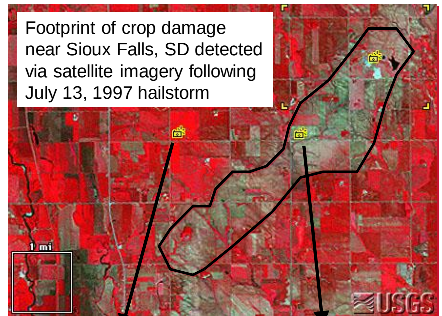


 November 7, 2019 16

A peril associated with severe thunderstorms is tornado, which is much more frequent than hurricane. However, most of what we understand about tornados comes from only the very largest events, which are few. So, even with a statistically large number of tornados in the past 100 years, meaningful understanding of intensity variability has been limited to a few data points thus adding to parameter uncertainty (e.g. the intensity of the storm).

PARAMETER UNCERTAINTY: HAIL

- Most hail research has focused on crop damage
  - Volume of hail (amount of ice per unit area) rather than hail size
  - Large hail is less of a threat to crops, but more of a threat to property



Undamaged soybean field

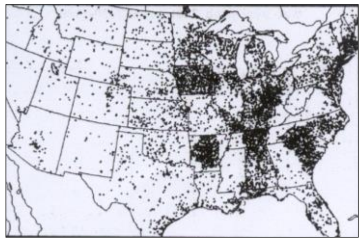
Damaged soybean field

 November 7, 2019 17

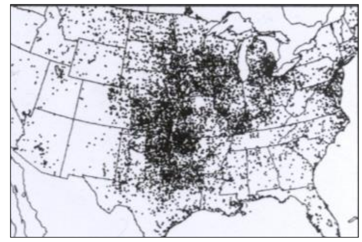
Another peril related to severe thunderstorms is hail. Hail is more frequent than tornados, but it still has parameter uncertainty associated with it. Most of the research around hail has focused on the damage that hail inflicts on crops, with the amount of hail in an area being the primary area of research. However, hailstone size is more important than the volume of hail when assessing hail damage to building structures. When examining current research for applicability to property damage, the inverse relationship about hail stone size must be accounted for – large hail stones are less damaging to crops but more damaging to structures. This highlights the parameter bias that may be introduced depending upon the focus of the reports, thus leading to event parameter uncertainty.

PARAMETER UNCERTAINTY:

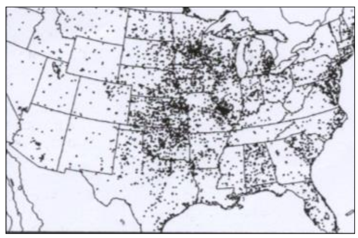
STRAIGHT-LINE WIND REPORTS



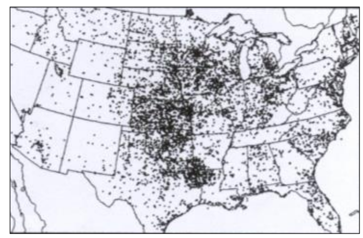
Plot of 58 mph wind gust reports 1970-1999



Plot of 60 mph wind gust reports 1970-1999



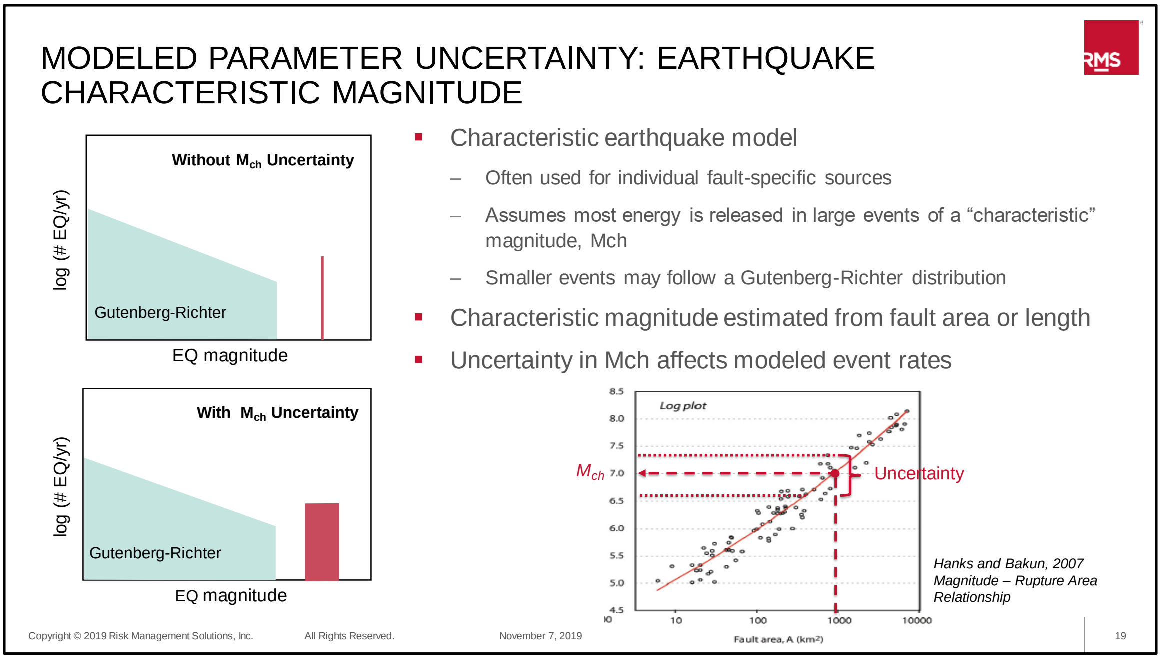
Plot of 65 mph wind gust reports 1970-1999



Plot of 70 mph wind gust reports 1970-1999

 November 7, 2019 18

For thunderstorms that produce neither hail nor tornados, parameter uncertainty comes from a different source – reporting patterns. Here is an example of how the measurement error is more pronounced for the smallest of these events. The map in the upper left is for the smaller wind reports. Note how the winds follow geo-political boundaries, with the border of Iowa being an example. This shows that the wind reports are more complete in Iowa than in neighboring states. Clearly, straight line winds do not stop at the Iowa-Illinois border, so assumptions must be made to complete the data which introduces parameter uncertainty. For the larger events, there is no evidence of this geo-political boundary, so additional assumptions are unnecessary, thereby minimizing parameter uncertainty.



Here is a simplified example of how a portion of the parameter uncertainty is included in the earthquake model.

The rate at which earthquakes occur (# EQ/year) is modeled through a recurrence relationship. For fault-specific sources, one theory holds that large, damaging events tend to be of a similar or "characteristic" size. This is because the stress needed to rupture or cause large-scale nearly instantaneous displacement along the faults builds to the same required level to create a similar size maximum event at similar intervals over time. Smaller earthquakes on or near the fault may be modeled with a Gutenberg-Richter (log-linear) recurrence, but these represent a much smaller proportion of the energy release and hazard for this source.

The characteristic magnitude is typically estimated from the source dimensions, either fault length or area, using relationships (like the one in the lower right) developed from past earthquakes. The size of the fault source is used to define the characteristic magnitude, which in turn is related to slip rate or estimated time between events to build the recurrence relationship.

One view is to assume the characteristic magnitude is a constant, which is shown in the top left diagram as a probability mass at M_{ch}.

A more comprehensive view is to recognize that there is variability in the magnitude for a given fault rupture, as well as the estimations of the fault itself, and model a distribution around this magnitude. The stochastic event set uses samples from this distribution as a means of including this parameter uncertainty in the model. This is recognized in the primary uncertainty.

[Note that the recurrence plots are displayed in terms of incremental rates, not cumulative; they represent the rate of a given magnitude, rather than that magnitude or larger. Also, the characteristic magnitude distribution does not have to be uniform, as illustrated for simplicity here.]

AGENDA

- Unit 1: Types of Uncertainty Overview
- Unit 2: Parameter Uncertainty
- **Unit 3: Secondary Uncertainty**
- Unit 4: Multiple Locations
- Unit 5: Including Secondary Uncertainty in the EP Curve
- Unit 6: Volatility Measures
- Unit 7: Correlation Between Portfolios

 November 7, 2019 20

This unit provides more detail on the sources of secondary uncertainty: hazard, vulnerability, and exposure data.

UNIT 3 – LEARNING OBJECTIVES

- Secondary Uncertainty
  - Define secondary uncertainty
  - Provide examples of hazard, vulnerability, and exposure data sources of secondary uncertainty
  - Explain how the level of uncertainty around mean damage ratios by peril is derived

 November 7, 2019

At the end of this unit you should have a concrete understanding of each of the three learning objectives listed on this slide.

SOURCES OF SECONDARY UNCERTAINTY

- Hazard
  - Earthquake: The uncertainty in the amount of ground shaking at a site
  - Windstorm: The uncertainty in the wind speed at a site
- Vulnerability
  - The uncertainty in the amount of damage, given an amount of ground shaking, wind speed, etc.
- Exposure data
  - The uncertainty in the location and construction details of the exposed building

 November 7, 2019 22

There are three main sources of secondary uncertainty: the hazard, the vulnerability, and the exposure data.

Hazard uncertainty deals with the variability around the loss-causing elements of the natural or man-made phenomenon that is being modeled. For instance, a loss-causing element of interest for earthquake is the intensity of the ground shaking and an element of interest for wind storms is the peak wind speed.

Vulnerability uncertainty deals with the variability of a structure's response to the hazard. As an example, you can have two identical structures experience the same amount of ground shaking, but the structures exhibit different damage patterns.

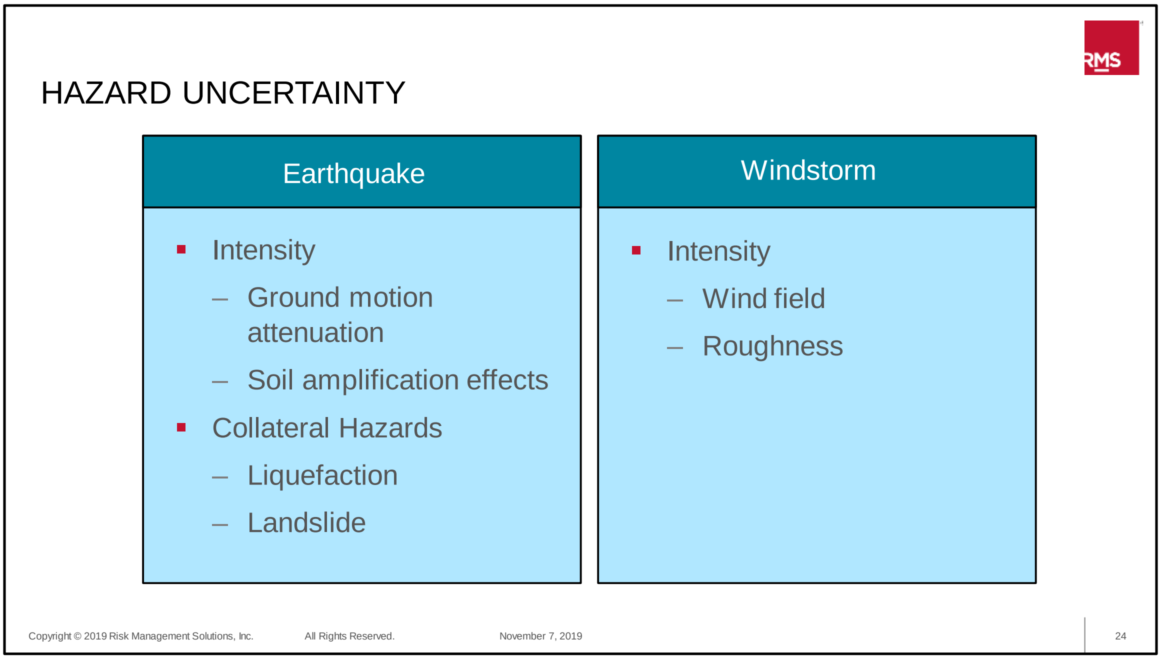
Uncertainty about the exposure refers to the difference between what is known about the structure in question and what is completely accurate about the structure. Even though a building looks like it has a specific roof nailing pattern, it may be inconsistent throughout the roof. However, when the exposure data correctly reflects the observed nailing pattern, that data may not completely reflect the inconsistent implementation of that nailing pattern.

SOURCES OF SECONDARY UNCERTAINTY

HAZARD

 November 7, 2019

The next three slides address the hazard component of secondary uncertainty in more detail.



Hazard uncertainty comes from various sources.

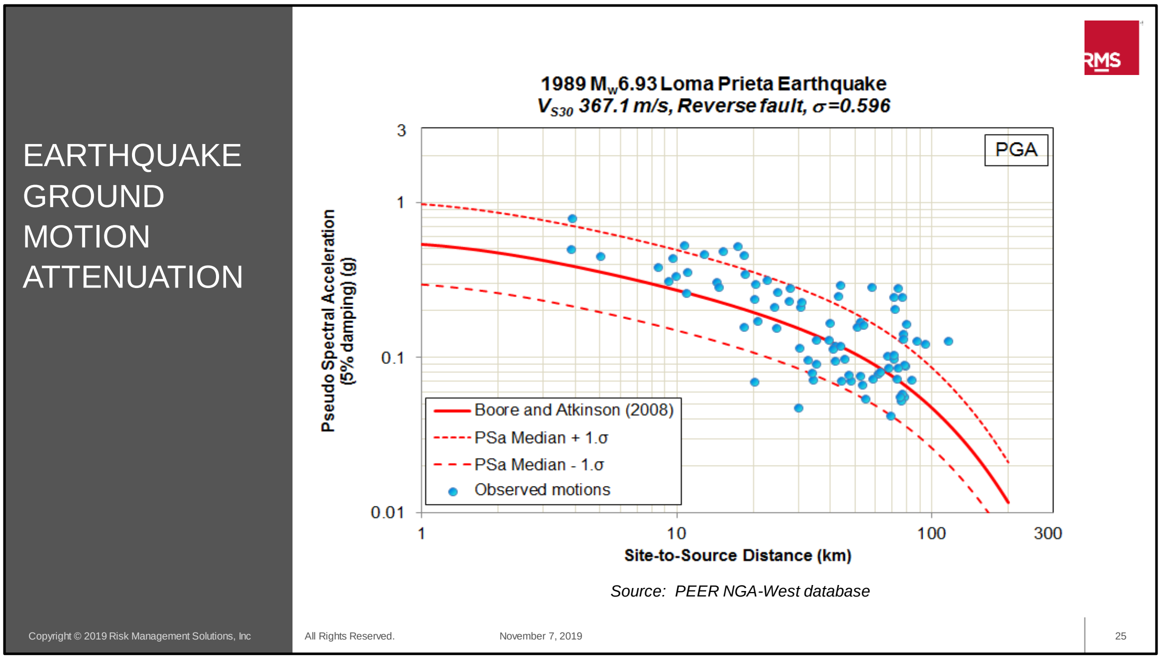
For earthquakes, the major source is the uncertainty in the ground motion attenuation. Attenuation refers to the dissipation of earthquake energy away from the fault rupture. This usually translates to a decrease in ground motion away from the source of the earthquake, however, there is uncertainty as to how much and where the ground motion decays.

Soil type has a significant impact on the amount of ground shaking at a particular location. For example, in general the ground shakes less over bedrock than over landfill. However, there is uncertainty as to how much less.

The collateral hazards of landslide and liquefaction also add to earthquake hazard uncertainty. An earthquake may initiate a landslide or cause the soil to behave almost as a liquid. However, there is uncertainty as to which locations will be impacted by landslide and soil liquefaction, and exactly how much damage the buildings will sustain.

For windstorm, the major source is the uncertainty in the wind field. Windstorms are by their nature very turbulent, complex, and asymmetrical. This creates a lot of variation over short distances, which creates uncertainty in the wind speed at any particular location.

Another source of windstorm hazard uncertainty is how the roughness of the surrounding landscape affects the wind speed. Roughness describes the type of terrain, land-use, and land-cover in an area. Tall natural and man-made features such as mountains, trees, and buildings tend to reduce the wind speed. However, there is uncertainty as to how much wind speeds decrease after passing over rough natural or man-made features.



An example from the 1989 Loma Prieta earthquake illustrates the uncertainty in ground motion attenuation. This graph plots the peak ground motion acceleration against the distance from the epicenter of the Loma Prieta earthquake. (The peak ground motion acceleration is measured in g's, a measure of ground motion used by seismologists that is measured as a fraction of the acceleration due to gravity's pull.)

The graph shows a general downward trend in the amount of ground shaking as distance from the fault rupture increases. However, the data points do not all fall exactly on the dark red curve, which is a predictive relationship developed using data from multiple events. Some points fall above the curve and some below the curve. This deviation from the mean estimate signifies that there is some uncertainty as to how much the ground shaking decays as the distance from the rupture increases. This uncertainty comes from both the individual characteristics of the event as well as the variability from site to site.

COLLATERAL HAZARD UNCERTAINTY

 November 7, 2019 26

Here is a rather dramatic example of liquefaction uncertainty.

This picture was taken after the 1964 Ni'igata earthquake in Japan and illustrates variations in liquefaction damage. Similar buildings on similar soil sustained markedly different levels of damage. Some of the buildings tilted to the side, while the building in the middle completely fell over. However, the buildings in the front were undamaged. Remarkably, many of the tilted buildings had little structural damage and were reused, once they were raised into upright positions.

Even though each of these buildings would be classified with the same susceptibility to liquefaction, the variability in the observed experience creates uncertainty in the modeled hazard.

SOURCES OF SECONDARY UNCERTAINTY

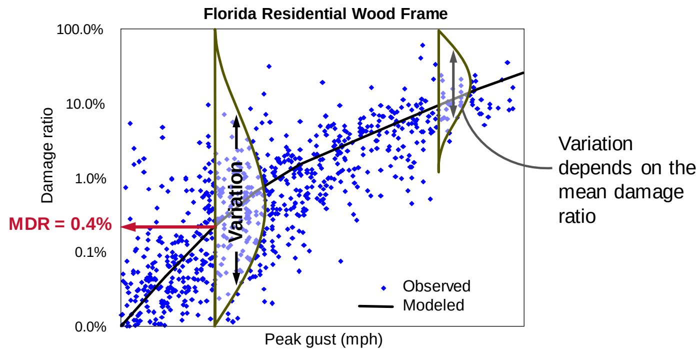
VULNERABILITY

 November 7, 2019

The next three slides address the vulnerability component of secondary uncertainty in more detail.

UNCERTAINTY IN VULNERABILITY: TROPICAL CYCLONE

- Vulnerability function models the mean damage ratio (MDR)
  – MDR = average loss / replacement value
- Variation around the mean is represented by the coefficient of variation (CV)



 November 7, 2019 28

This graph plots building damage ratios against wind speed for hurricanes in Florida. A damage ratio is the percent of the building that suffers a loss. Each data point represents the performance of wood frame single family dwellings for a ZIP Code in one of these hurricanes.

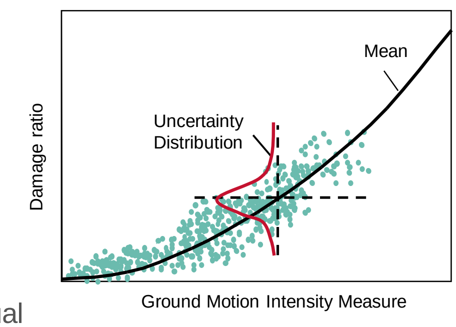
The graph shows a general upward trend in the damage ratio with increased wind speed, represented by successive line segments fit between mean values (the black line on the graph). This best-fit trend is termed a vulnerability curve. There is significant variation in the individual values observed around that curve. Some points are higher, some points are lower. This signifies that there is uncertainty as to how much damage occurs at a given wind speed, and how the damage increases as the wind speed increases.

Variability around the mean is noted by the coefficient of variation, or CV. The CV is the ratio of the standard deviation of the damage ratio to the mean damage ratio (MDR). *Note: Refer to Unit 4 of the Financial Modeling course for more information on the definition and calculation of the CV.* The leftmost green curve shows the distribution of loss around the MDR of 0.4%. Each of the data points that are crossed by the vertical green line are used to calculate the standard deviation, which is fairly large in this graph. The calculated standard deviation and the modeled MDR are then used to calculate the CV. Note that the standard deviation is dependent on the mean damage ratio.

UNCERTAINTY IN VULNERABILITY: EARTHQUAKE

- Shows relationship between ground motion (MMI, PGA, Sa) and damage (MDR)
- Damage represented as % of
  - Building replacement cost
  - Contents value
  - Time element coverage
- Vulnerability functions are region-specific and vary by building class, age, and height
- Vulnerability functions are calibrated with actual damage and loss data from past earthquakes

Vulnerability Function For Reinforced Concrete Building



 November 7, 2019 29

This graph looks at the performance of reinforced concrete structures when exposed to an earthquake.

The general trend is that the damage ratios increase as ground motion intensifies. As noted with hurricane, there is significant variation in the individual values observed around the black curve, which is the best-fit mean damage at each measure of ground motion. Some points are higher, some points are lower. This signifies that there is uncertainty as to how much damage occurs at a given level of ground motion for the same construction.
After reviewing the empirical and observed data, RMS breaks the data into meaningful groups. Specific vulnerability functions, such as the one depicted as the black curve in the graph, are developed for various region, building class, age, and height groups. The developed functions are then calibrated to and validated by claims data from past earthquakes. This same process is followed for all of RMS' available peril models.

UNCERTAINTY IN VULNERABILITY: FLOOD

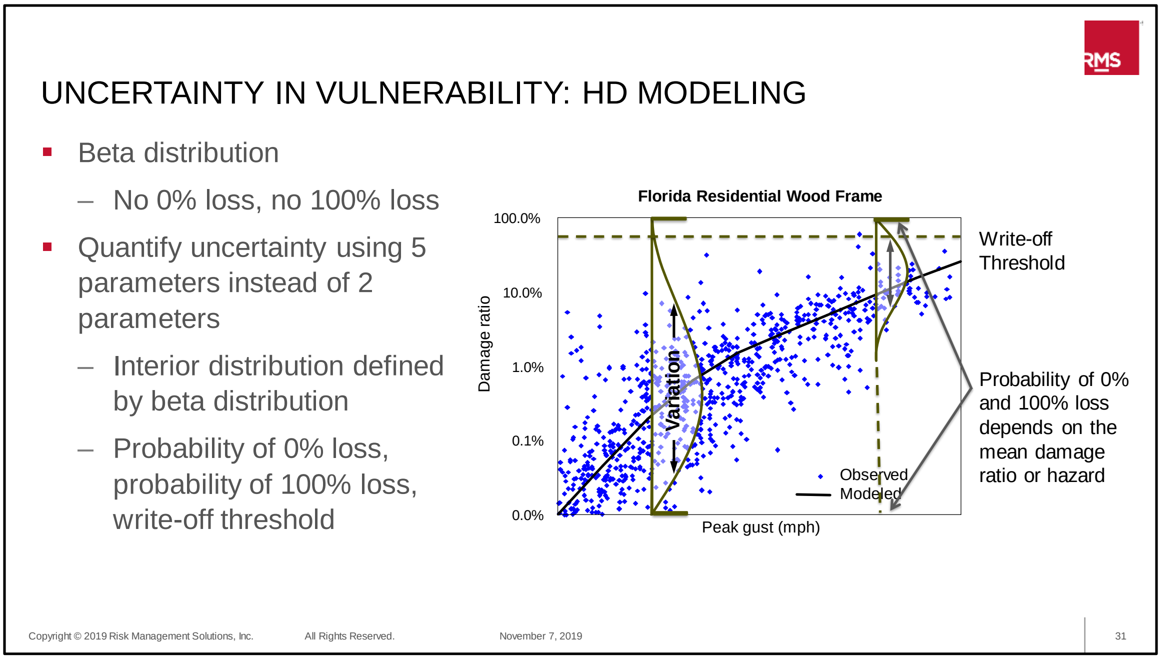
Building Collapse Due to Undermined Foundations
Prague, August 2002

Source: Radio Prague

 November 7, 2019 30

This graphic shows two pictures of a building that collapsed in a flood.

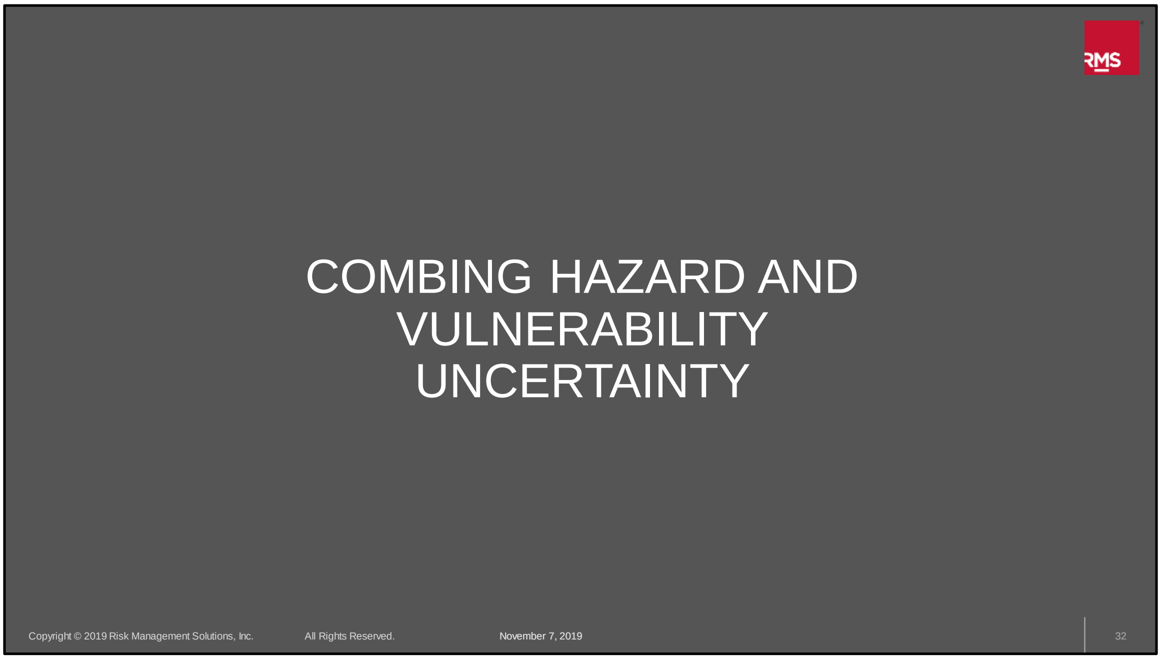
Assume that the primary construction characteristics (construction type and year built) are similar. Of interest is the fact that the building next to it did not collapse. Although this is similar to the earthquake liquefaction photograph from the 1964 Ni'igata earthquake in Japan, they highlight different sources of uncertainty. The Ni'igata pictures showed uncertainty in hazard (liquefaction), while this picture shows uncertainty in vulnerability because the hazard (depth of water) is identical. What is important to note in this picture is that similar foundational structures performed very differently when exposed to the same hazard.



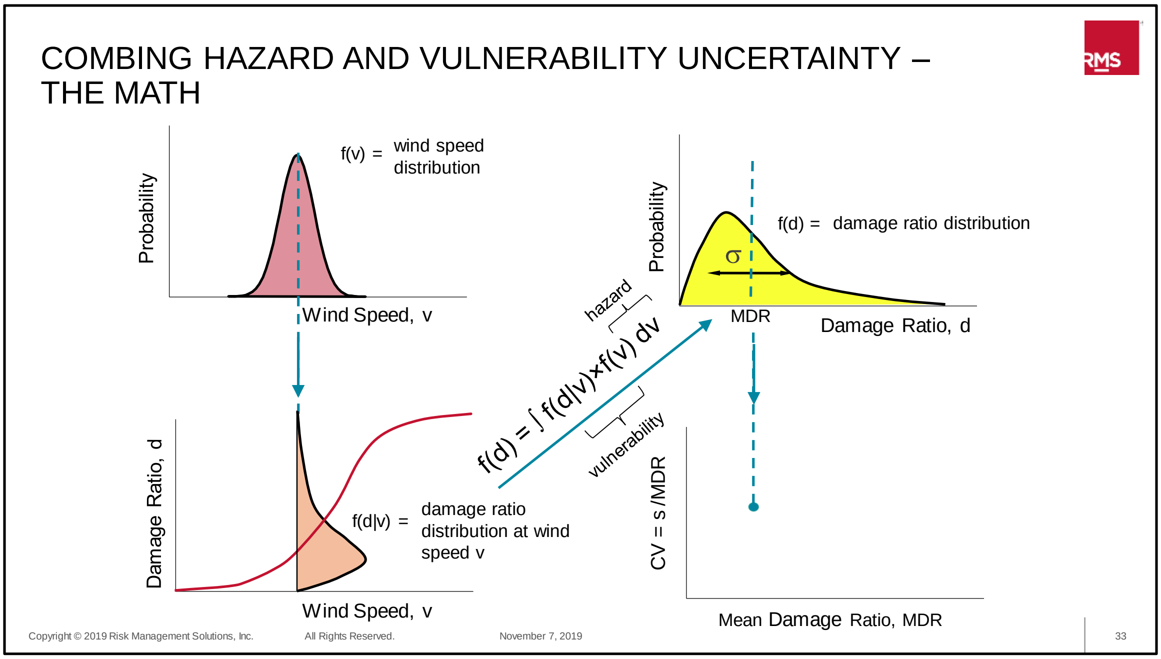
This graph is a replica of the earlier plot of building damage ratios against wind speed for hurricanes in Florida. As previously mentioned, the green curves show the distribution of loss around the MDR. This variability around the mean is noted by the coefficient of variation, or CV, and is modeled by a beta distribution. In RiskLink the probability of 0% or 100% loss is not explicitly defined within the context of the secondary uncertainty.

In HD modeling, the probability of 0% or 100% loss will be explicitly defined and the variation of damage ratios will continue to be modeled by a beta, or beta-like, distribution. The mean damage ratio will be a function of the probability of 0% loss, the probability of 100% loss and the mean of the variation of the damage ratios between 0% and 100%, or "internal distribution." Note the magnitude of 0% and 100% probability of loss in the figure are not from actual data, and are used for purely explanatory purposes.

Redefining the secondary uncertainty methodology also allows for consideration of a fifth parameter – a write-off threshold. Damage ratios above this threshold would be considered to be 100% loss. For example, if a location experiences a damage ratio of 97% a user may wish to consider this as a total loss. This parameter introduces an additional level of uncertainty, however, as the write-off threshold may vary between companies.



Once the hazard and vulnerability uncertainties have been accounted for, they are combined to arrive at a single relationship that describes uncertainty around the mean damage ratio (MDR), and ultimately the loss.



This slide illustrates how the hazard and vulnerability uncertainty can be combined mathematically.

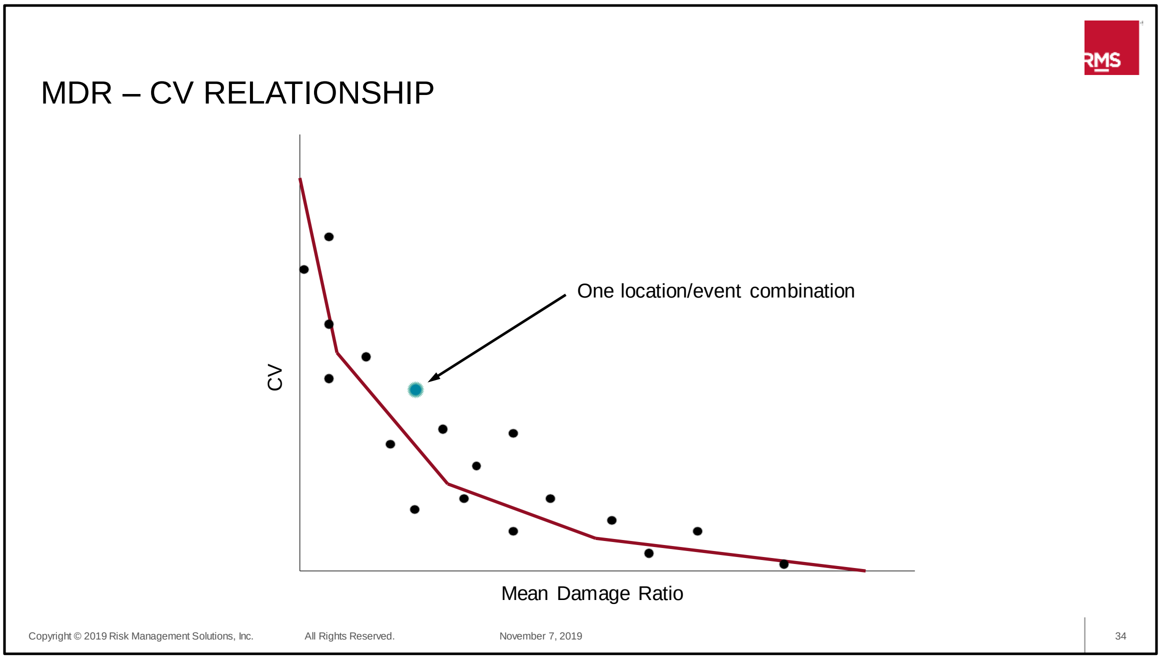
The graph in the top left shows the probability distribution of the wind speed at a given location for a given event. The probability distribution function of the wind speed, f(v), represents the hazard uncertainty.

The graph in the lower left shows a damage (or vulnerability) curve in blue, which relates the mean damage ratio to the wind speed. The probability distribution shown in the lower left image, f(d|v), represents the vulnerability uncertainty. That is, the uncertainty in the damage ratio, given the wind speed = v.

To combine these, we take a sample wind speed from the hazard distribution (shown as a dashed red line in the left-hand graphs). For this selected wind speed, we sample from the vulnerability distribution to get the sampled damage ratio. We do this many times to get the probability distribution of damage ratio for this event and location. This is mathematically equivalent to computing the integral shown in the middle of the slide.

The resulting damage ratio distribution, plotted in the upper right-hand graph, has a mean (MDR) and a standard deviation (σ) associated with it, as indicated. The coefficient of variation (CV) is the ratio of the standard deviation divided by the mean of this damage ratio distribution, and is plotted in the lower right-hand graph as a function of the mean damage ratio. This is one point on the MDR-CV relationship for this vulnerability function.

RiskLink DLM does not do this sampling procedure for every location for every event during an analysis run, since that would adversely affect performance. So, the results of the simulation are precompiled.



The precompiled results are then plotted as shown in this graph. Each point on this graph represents one location/event combination from the simulation to combine the hazard and vulnerability uncertainties. Note that the graph shown on this slide is the same one as the lower right graph on the previous slide. The red point on this slide represents the same red dot on the previous graph.

The horizontal axis shows the mean damage ratio, which is the mean of the distribution. The vertical axis shows the CV, which is the standard deviation divided by the mean and is a measure of the uncertainty in the mean damage ratio.

After repeating this simulation for a multiple locations and multiple events, a strong relationship between the mean damage ratio and the coefficient of variation becomes apparent and a piecewise linear curve is fit through the simulated MDR/CV pairs. When a RiskLink DLM analysis is run, RiskLink calculates the mean damage ratio, then uses this fitted curve to estimate the CV. This CV is then modified for the portfolio data uncertainty.

CALIBRATING THE MDR – CV RELATIONSHIP

Sort (on Mod. DR)

Loc	Obs. DR	Mod. DR	Obs. - Mod.
1	0.36%	0.28%	0.08%
2	0.00%	0.29%	-0.29%
…	…	…	…
99	2.64%	1.87%	0.77%
100	1.63%	1.88%	-0.25%
101	2.38%	1.89%	0.49%
102	1.63%	1.92%	-0.29%
…	…	…	…
199	1.88%	2.39%	-0.51%
200	4.62%	2.40%	2.22%
201	2.99%	2.43%	0.56%
…	…	…	…

$\}$ Locations 1 – 100
Modeled MDR = 1.66%
σ of Obs. - Mod. = 5.39%
CV = 5.39/1.66 = 3.25

$\}$ Locations 101 – 200
Modeled MDR = 2.11%
σ of Obs. - Mod. = 7.27%
CV = 7.27/2.11 = 3.45

 November 7, 2019 35

This table shows how we use claims data to calibrate the MDR-CV relationship.

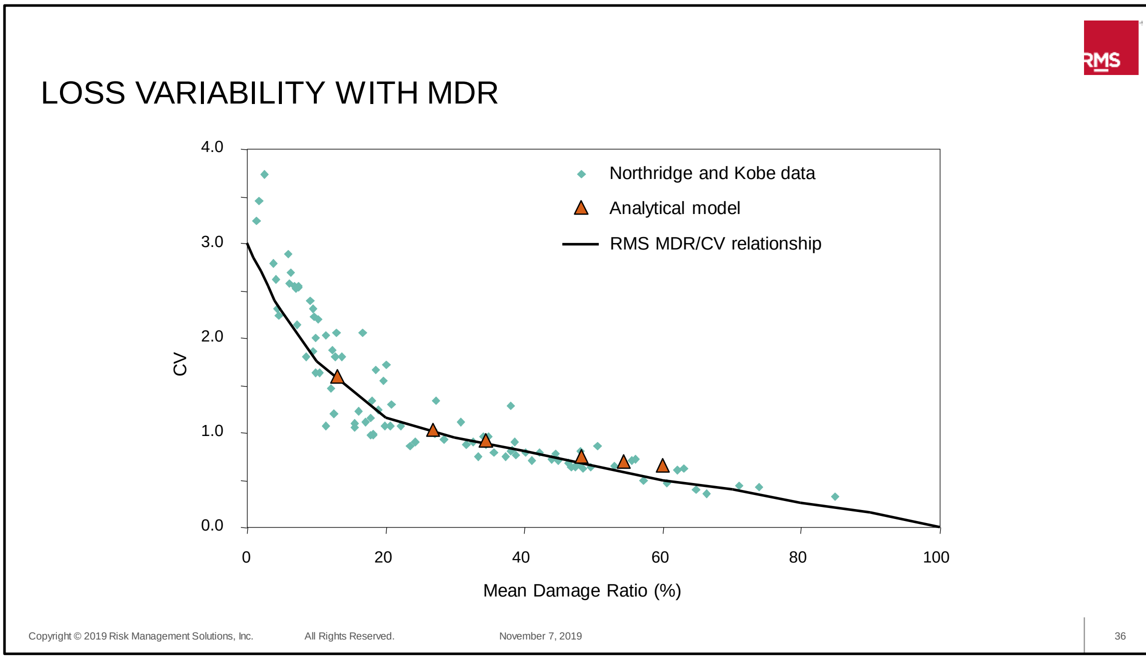
For each location (listed in column 1), we calculate the observed damage ratio (column 2), the modeled damage ratio (column 3), and the difference between the observed to modeled (column 4). We then sort on the modeled damage ratio (lowest to highest) and split the data into bands of 100 locations.

Thus, for location 99, the building damage based on field and claims data was 2.64% of the total value (Note: It is important to compare total values since the modeled MDR is defined as percent of damage relative to total value). The modeled MDR for location 99 is 1.87%, and the difference between the observed and modeled MDR is 2.64% - 1.87% = 0.77%.

To calculate the CV around the <u>modeled</u> MDR, the next steps are to:

1) Calculate the mean of the modeled damage ratio, and
2) Calculate the standard deviation (σ) of the values in the Obs. – Mod. column, then divide by the modeled MDR to arrive at the CV around the modeled MDR.

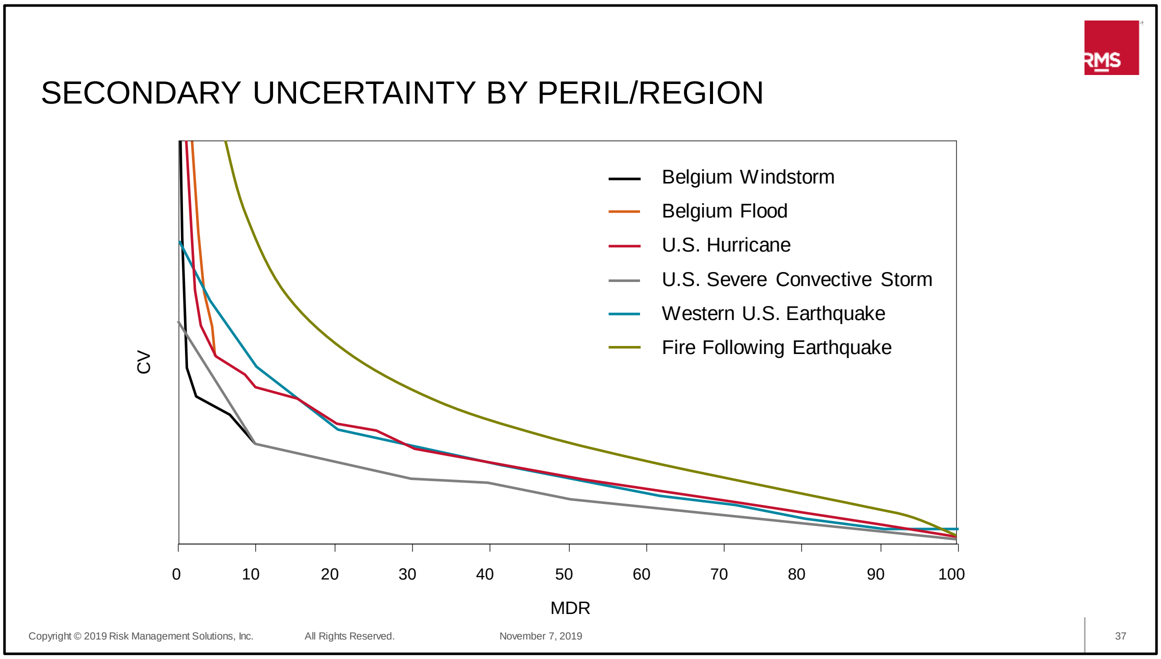
Thus, for the first band, the mean of the modeled damage ratios is 1.66%, and the standard deviation (σ) around this mean value is 5.39%. Dividing the standard deviation by the mean modeled damage ratio yields a CV of 3.25 for this band. This is represented as a point on a MDR-CV plot. In fact, each of these MDR/CV pairs for a band becomes a point on a chart similar to the red dots shown on the next slide.



Here are the results of our calibration for the earthquake model.

Each red point represents one MDR/CV pair from a band as explained on the previous slide. If the data is broken up into N bands, then there should be N dots.
The yellow triangles represent the results of some simulations that our vulnerability team did as part of the development of the spectral response model (the measure of structural response to earthquake ground motion).

The black curve represents the chosen MDR/CV relationship in the model. This is a piecewise linear best-fit to both populations of data.

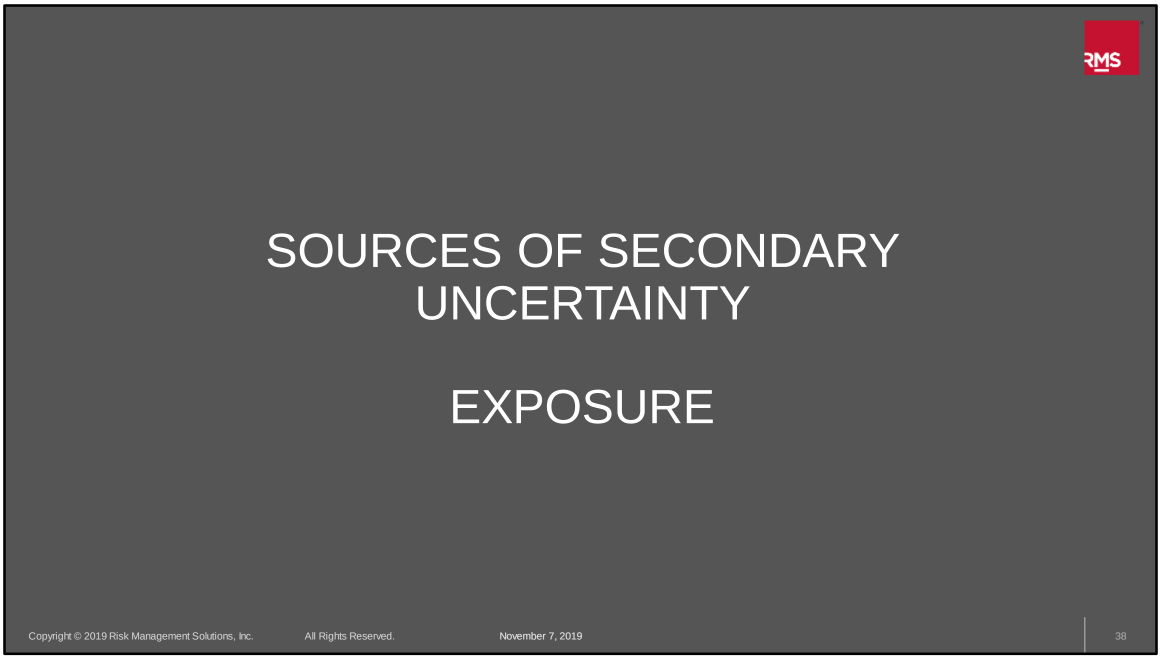


This slide shows the MDR-CV relationships used for a selection of RMS' peril/region models. Each is piecewise linear, with the exception of the fire following earthquake (FFEQ). The FFEQ MDR-CV relationship is unique in that it is formula driven, so it has a smooth curve.

In general:

- There is less variability at higher mean damage ratios for all perils.
- The high CV value for low MDR's is also a result of CV = σ/MDR. Thus the lower the MDR (assuming σ is not reduced greatly), the greater the CV.

You will notice the sharp drop in the low MDR portion of the Belgium Windstorm curve (black). It is important to note that the Europe Windstorm model has a unique MDR/CV relationship because observed damage is rarely greater than 10%, meaning there is much more information available from past events at lower mean damage ratios compared to other perils. You will note an inflection point around 2% MDR. This is due to two factors; the binning of observed damage by wind speed bands, and the fact that there is relatively more certainty around MDR at higher wind speeds as there are more building damage observations to use at higher wind speeds.



The next five slides address the exposure data component of secondary uncertainty in more detail.

EXPOSURE DATA UNCERTAINTY

- Geographic resolution
  – CRESTA, county, postal code, street address
- Construction characteristics
  – Known or unknown characteristics
    - Primary characteristics
    - Secondary characteristics

 November 7, 2019 39

Exposure data uncertainty addresses uncertainty in the model input data. The more accurate, complete, and highly resolved the data, the less uncertainty there will be in the modeled losses. This is similar to, but not the same as, parameter uncertainty, since the parameter uncertainty deals with the model parameters, not the actual data used as inputs to the model. For example, using portfolio data with location information at the street address level instead of at the ZIP Code level decreases the secondary uncertainty.

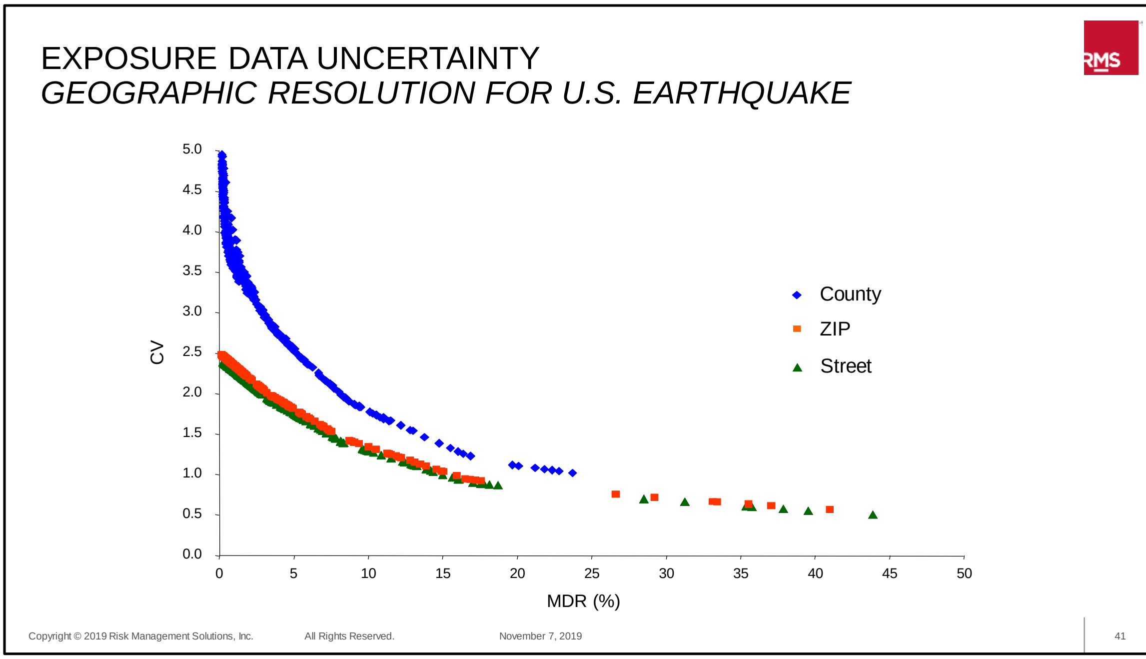
As another example, secondary uncertainty increases when building construction data is not known (e.g. the construction type, year built, number of stories). In similar fashion, the greater number of secondary characteristics that are known, such as the roof type (hip or gable), whether the building is bolted to the foundation, the presence of a soft story, etc., the less uncertainty there will be in the losses.

INCLUDING EXPOSURE DATA UNCERTAINTY

- The CV is modified based on the resolution of the exposure data.
- In general, the finer the resolution (the more characteristics are known), the lower the secondary uncertainty CV for a given MDR.
- When exposure information that decreases the MDR is added, the decreases in the secondary uncertainty CV due to the additional exposure information may be offset by the increase in the secondary uncertainty CV associated with the decrease in the MDR (see slide 36).
  - It is possible that the net effect is an increase in the secondary uncertainty CV.

 November 7, 2019

The MDR-CV relationships that are derived using the simulation method discussed earlier assumes a certain level of knowledge about the exposure data. In the baseline case, the CV derivations assume that the location is geocoded at the postal code resolution, the primary building characteristics are known, and the secondary building characteristics are unknown. To the extent that the exposure being modeled is different from this baseline assumption, the CV is modified upward or downward to reflect the amount of information known about the location. This is exemplified for a California earthquake location on the next slide.



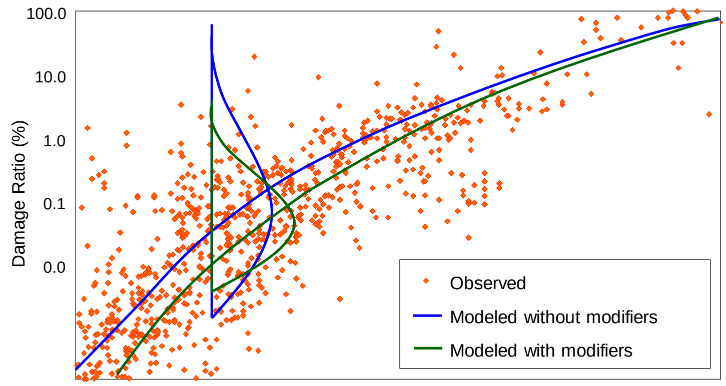
The chart on this slide shows the impact of geocoding resolution on the CV for a single family home in California. The same location was analyzed three times at varying data resolutions, street address, ZIP Code, and county.

Each point represents the mean damage ratio (MDR) and coefficient of variation (CV) for a stochastic earthquake event. Note that the CVs are much higher with county level geocoding. This is a reflection of the large size of the county, and thus the diversity in the mean hazard assigned to the county. As such, the uncertainty in losses is large if you do not know where in the county the location is physically located.

The ZIP Code level and street address CVs are not that different, because this particular ZIP Code is relatively small and homogeneous from an earthquake hazard perspective. The adjustments in the secondary uncertainty CV based on geocoding resolution are only applicable to RMS's earthquake models.

SECONDARY CONSTRUCTION CHARACTERISTICS
MEASURABLE IMPACT ON DAMAGE CURVES

- Vulnerability curve modification
- Secondary uncertainty measures

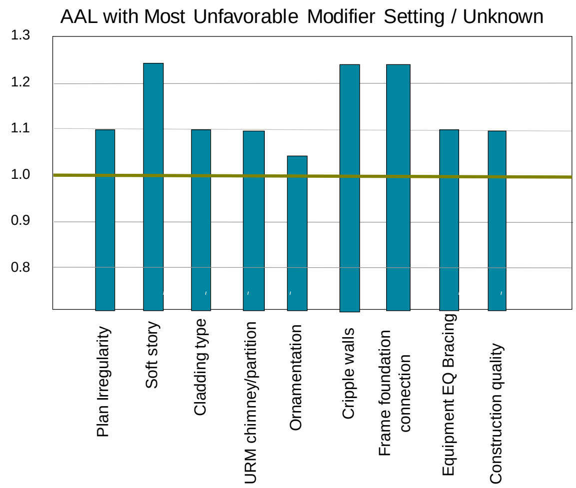


 November 7, 2019 42

Secondary construction modifiers impact both the MDR and the CV.
As has already been discussed, as you include an increasing number of secondary characteristics in your exposure input data, your exposure data uncertainty will decrease. In addition, it also has an impact on vulnerability uncertainty, as depicted on the graph above. This shows that secondary characteristics also change the spread of damage around the mean – by narrowing the uncertainty distribution from the blue to the green distribution. This graph also points out that this particular group of secondary characteristics lowered the mean damage curve as well. However, please note that secondary modifiers will not always decrease the building or coverage MDR, as is shown by the example in the next slide.

EFFECT OF VARIOUS MODIFIERS ON EQ AAL –
WUS SFD, 2-STORY, WOOD FRAME

- Cripple walls modifier unique to wood frame
- Frame foundation connection unique to wood frame and mobile homes



 November 7, 2019 43

In an effort to look at the impact of specific secondary characteristics (modifiers) on earthquake loss, we can examine this chart. The horizontal axis identifies secondary modifiers that are applicable to earthquake modeling. The vertical axis is the ratio between modeled loss from a location with an unknown secondary modifier setting and the modeled loss when the most unfavorable (most damageable) modifier setting is chosen. You will note that the impact on the MDR differs by modifier.

There is also an impact on the CV, although it is not shown here. In general, the greater the effect the modifier has on the losses, the greater the reduction in the CV if that modifier is known. So in this example, one of the modifiers that will cause the greatest reduction in the CV is knowing if the frame is bolted to the foundation or not. However, note that a setting of “unbolted” will increase the MDR.

It is worth noting some data quality issues in this example. Note that both cripple walls and bolting are unique to wood frame (in addition to mobile homes for bolting). If you profile a portfolio and find either of these identifiers being designated for other types of construction, this may be a flag for data accuracy issues in the portfolio. Regardless, in this case, neither of these modifiers will impact loss or loss uncertainty measures as RiskLink/RiskBrowser will not apply these modifiers to any vulnerability curves other than those describing wood frame or mobile home construction.

SUMMARY UNITS 1-3: TYPES AND SOURCES OF UNCERTAINTY

- Catastrophe models incorporate primary, secondary, and parameter uncertainty.
- Secondary uncertainty quantifies the uncertainty in hazard, vulnerability, and exposure data.
- Parameter uncertainty is the uncertainty in quantifying model parameters. A portion of this uncertainty is modeled, and a portion is not.

 November 7, 2019 44

Please review these summary points and key topics that were covered in Units 1- 3. If any of this information is unclear, we encourage you to go back to that unit to review the details prior to continuing with your study of this course.

AGENDA

- Unit 1: Types of Uncertainty Overview
- Unit 2: Parameter Uncertainty
- Unit 3: Secondary Uncertainty
- **Unit 4: Multiple Locations**
- Unit 5: Including Secondary Uncertainty in the EP Curve
- Unit 6: Volatility Measures
- Unit 7: Correlation Between Portfolios

 November 7, 2019 45

This unit provides more detail on the treatment of uncertainty for multi-location accounts and portfolios.

UNIT 4 – LEARNING OBJECTIVES

- Multiple Locations
  - Understand the differences between the independent and correlated location level uncertainty
  - Explain why the correlation of loss between locations varies by peril and regions

 November 7, 2019 46

At the end of this unit you should have a concrete understanding of each of the two learning objectives listed on this slide.

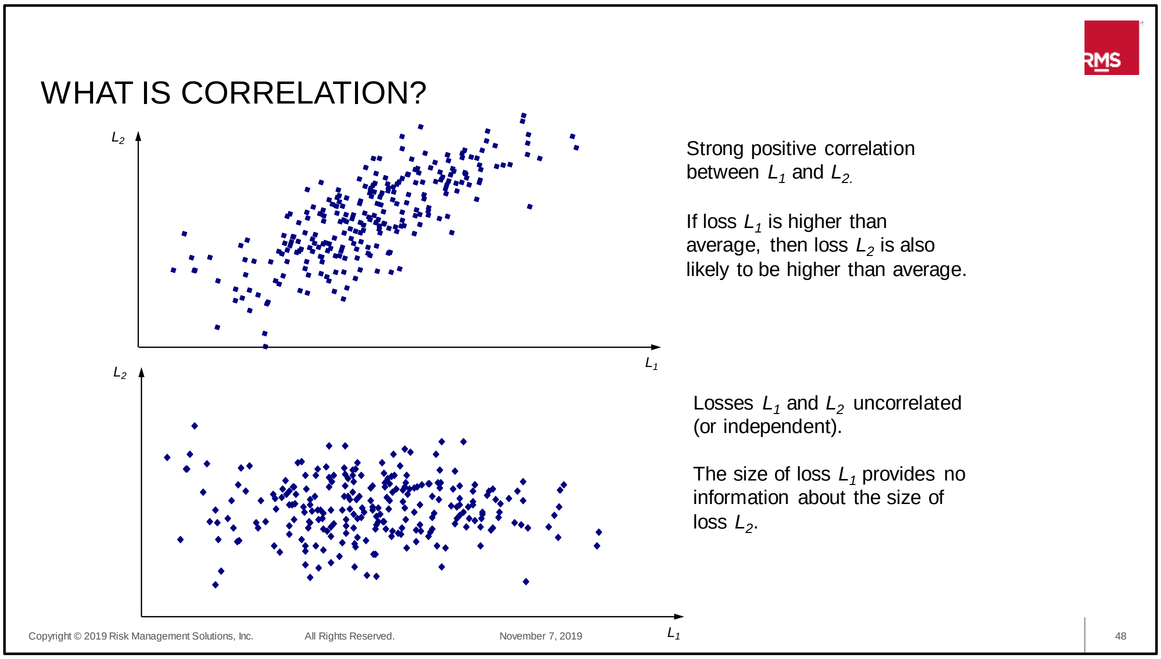
WHAT IS CORRELATION?

- The correlation between two random variables measures the degree to which the variables are linearly related.
- A strong positive correlation between the losses at two locations implies that if the losses are large for one correlation, they are likely to be large for the other location.
- When the losses are uncorrelated (or independent), then knowing the size of the loss at one location does not provide any information about the size of the loss at the other location.

The correlation between two random variables (such as the losses incurred at two different locations in a portfolio) is a measure of the degree to which the variables are linearly related.

If two random variables have a strong positive correlation, then when one of the random variables is large, the other is also likely to be large.
Conversely, if two random variables are uncorrelated – i.e., there is no discernible linear relationship between their values – knowing the value of one of the random variables does not provide any information about the value of the other random variable.

In RiskLink, we refer to the uncorrelated case as “independent” random variables. Strictly speaking, this terminology is misleading - while independent random variables are always uncorrelated, uncorrelated random variables are not necessarily independent. However, for all practical cases in the types of analyses performed by RiskLink, the lack of correlation between two random variables is due to their independence.



These plots illustrate a pair of random variables. The upper plot shows variables with a relatively large positive correlation, while the lower plot shows a pair of independent (uncorrelated) random variables.

When the losses are positively correlated, then information about loss L1 gives us some information about loss L2. For example, if loss L1 is higher than average, then L2 is also likely to be higher than average. On the other hand, when the losses are uncorrelated (or independent), then knowing that L1 is higher (or lower) than average does not provide any information about loss L2.

AGGREGATING FROM LOCATION TO PORTFOLIO

- The standard deviation of a portfolio loss is given by:

$$STDDEV = \sum_{loc} STDDEV_{loc}$$

when the location losses are perfectly correlated (*r* = 1)

and

$$STDDEV = \sqrt{\sum_{loc} STDDEV_{loc}^2}$$

when the location losses are independent (*r* = 0)

 November 7, 2019

Up until now, the discussion has centered around the secondary uncertainty for one location only. We now need to examine the aggregation of uncertainty up to the portfolio level.

First, consider the standard deviation of the portfolio loss for two extreme cases:

1. that all locations are totally correlated with each other, and
2. that all locations are completely independent of each other.

The first equation on this slide shows the calculation of the portfolio standard deviation under the assumption that all locations are completely correlated. The second equation shows the calculation of the portfolio standard deviation under the assumption that all locations are completely independent.

AGGREGATING FROM LOCATION TO PORTFOLIO

- In RiskLink, the portfolio standard deviation is computed as a weighted average of these two extreme cases

Correlated Piece
$$STDDEVC = w\sum_{loc} STDDEV_{loc}$$

Independent Piece
$$STDDEVI = (1-w)\sqrt{\sum_{loc} STDDEV_{loc}^{2}}$$

Portfolio Standard Deviation
$$STDDEVC + STDDEVI$$

 November 7, 2019 50

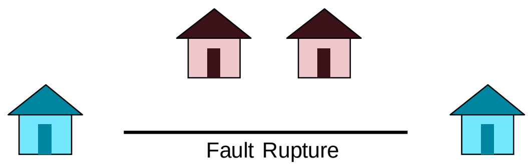
To minimize the runtime in RiskLink, the portfolio standard deviation is calculated as a weighted average of the two extreme cases shown on the previous slide.

If the weight applied to the perfectly correlated case is w and the weight applied to the independent case is (1-w), then the “correlated” and “independent” pieces of the portfolio standard deviation calculation are given by the first two equations shown on this slide. Adding these two pieces yields the total portfolio standard deviation, as shown by the last equation on the slide.

Note that when w = 1, the location losses are assumed to be perfectly correlated and when w = 0, the location losses are assumed to be independent. Consequently, the strength of the correlation between losses at different locations in a portfolio is quantified by the weight w in this formulation.

CORRELATION ASSUMPTIONS BY PERIL/REGION

- The correlation weight (w) is based on:
  - Exposure density: the closer the locations, the higher correlation



  - Dispersion of loss impacts: earthquake correlation > hurricane correlation
  - “Super Cats” haver higher correlation

 November 7, 2019 51

Correlation weights are estimated by simulating the losses by location for industry portfolios for the various regions and perils, and then calculating the loss-weighted average distance from the locations to the centroid of the event. The further apart the losses are from each other, the lower the weight given to the correlated piece of the standard deviation. The correlation weights used for a given region and peril are the same for all exposure data sets, and do not vary with the exposure density used in the analysis.

To illustrate how correlation weights are estimated for different regions, consider the picture above. All four houses are the same distance from the fault rupture. But because the blue houses are closer together than the red houses, they are more correlated than the red houses. This illustrates how exposure density contributes to the correlation weight; however, exposure density alone does not determine the weight.

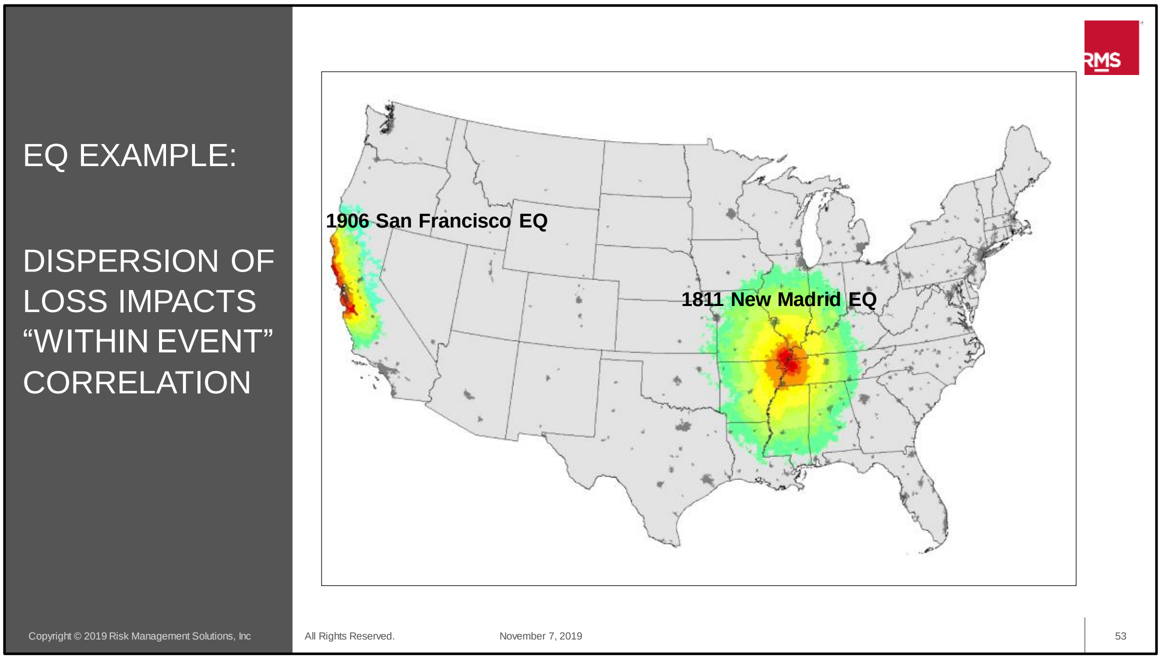
Regions and perils with loss impacts that are more focused and localized are modeled with a higher loss correlation weight than regions and perils with loss impacts that are scattered and more widespread, regardless of the density of exposures. Thus, the earthquake peril tends to have a higher loss correlation weight than the hurricane peril.

The next slide provides the empirically derived correlation weights by peril and region for current RMS peril models. These weights are termed regional "baseline" weights for catastrophe events. They are altered for super catastrophic (Super Cat) events whose large magnitude and association with other concurrent catastrophes (e.g. a category 5 hurricane concurrent with sea defense failure and subsequent catastrophic flooding) increase the spatial correlation of losses for highly impacted areas. Super Cat correlation weights are discussed in more detail shortly.

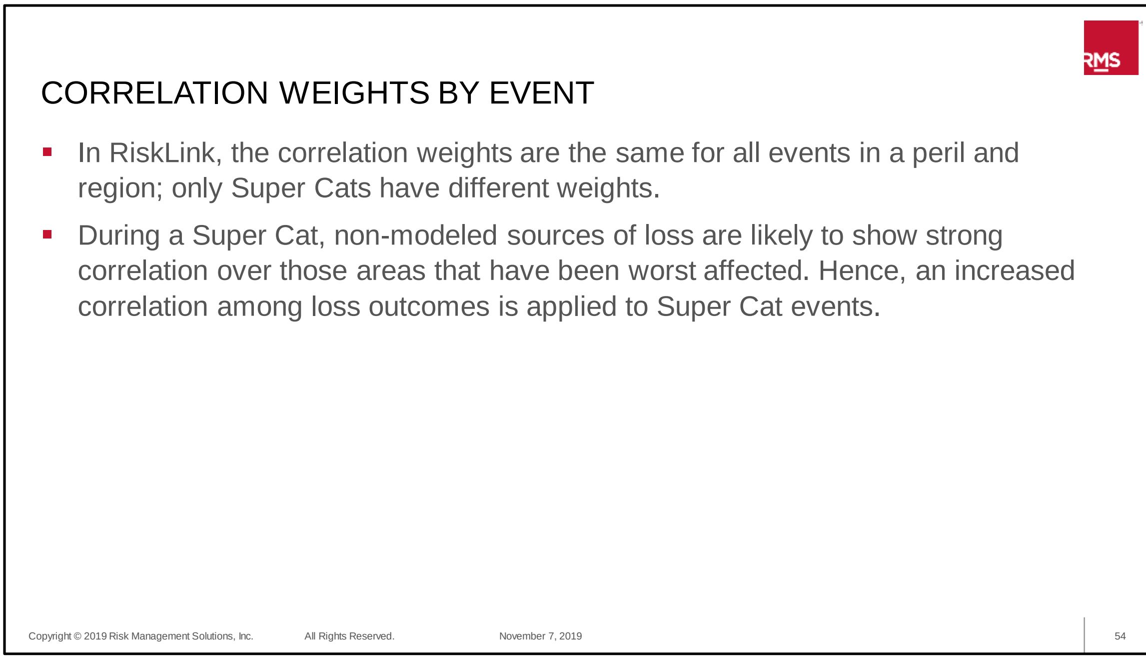
CORRELATION WEIGHTS BY PERIL / REGION

Peril / Region	Correlation Weight
California EQ	20%
Pacific N.W. EQ	20%
Japan EQ	20%
Eastern Canada EQ	17.5%
New Madrid EQ	17.5%
U.S. HU	15%
Severe Conv. Storm	15%
Japan TY	10%
European WS	8%

This table shows the weights for some of the perils and regions. The California earthquake region is assigned a 20% location correlation weight, whereas the New Madrid earthquake region is assigned a 17.5% location correlation weight. This correctly reflects the average geographic dispersion of the losses for events in those peril regions as discussed on the next slide.



To further illustrate the correlation weight variation by peril and region as shown on the previous slide, shown here are the extent of reported felt or damaging ground motions for two large earthquakes of nearly equal magnitude in the U.S. Note that the felt area for the New Madrid earthquake is much greater than that of the San Francisco earthquake. This is due to a slower attenuation of ground motion in the New Madrid region, resulting in losses extending to greater distances relative to California, which exhibits higher attenuation of ground motion. Assuming locations with similar geographic density, you will observe that the losses in the New Madrid region are more highly dispersed (covering a larger area) than for the California region. The farther apart the individual locations exposed to an event are from each other, the lower the loss correlation weight for that region.



The concept of Super Cat correlations was introduced several slides ago. Due to the extreme severity of the event hazard, a strong correlation exists between non-modeled loss (levee failure, evacuation cost, business closure, etc.) and modeled sources of loss for the highly impacted regions. Super Cats thus have their own component of loss escalation in addition to the demand surge and claims inflation common to other catastrophes. This Super Cat component of post-event loss amplification is accounted for in the increase of location correlation weights in densely populated urban regions.

The increase in correlation weights for Super Cats can be dramatic. For example, the maximum spatial correlation weight calculated for a California earthquake event is 32%, compared to 20% for non-Super Cat events. Similarly, the Super Cat correlation weight for New Madrid earthquake is 30%, compared to a baseline weight of 17.5%. Finally, for U.S. hurricane the maximum spatial correlation weight is 28%, compared to 15%.

CORRELATION: RISKLINK VS HD MODELING

- In RiskLink, the correlation is based on a simple and quick formula which is reasonable in most cases.
  - Occasionally results can be counterintuitive because correlations depend on the # of locations in the portfolio.
- In HD modeling, the correlation structures will be applied at building level, which avoids potential counterintuitive results (e.g. when adding locations).
- Applying correlation at the building level will allow future consideration of more complex correlation structures such as spatial (e.g. distance, region, country), construction type, year built, line of business, etc.

 November 7, 2019 55

The formulas for aggregating location level losses to portfolio loss have been discussed in this unit as a function of a correlation weight and the correlated and uncorrelated standard deviations. This methodology behaves reasonably for the majority of cases and minimizes the runtime in RiskLink. On occasion it can produce slightly counterintuitive results because the standard deviation is a function of the number of locations, which is dependent on the portfolio.

In HD modeling the correlation structure will be applied at the building level which will avoid these potential counterintuitive results when adding location loss to portfolio loss. More importantly it allows potential consideration of a number of more complex correlation structures to be applied spatially (locally or grouped), across building characteristics (construction type, year) and/or across occupancy types.

EXERCISE 1

 November 7, 2019

At this point in the course you should now complete Exercise 1, the objective of which is to calculate the portfolio level event standard deviation from location level information.

AGENDA

- Unit 1: Types of Uncertainty Overview
- Unit 2: Parameter Uncertainty
- Unit 3: Secondary Uncertainty
- Unit 4: Multiple Locations
- **Unit 5: Including Secondary Uncertainty in the EP Curve**
- Unit 6: Volatility Measures
- Unit 7: Correlation Between Portfolios

 November 7, 2019

Unit 5 steps through the process of including secondary uncertainty in exceedance probability curves.

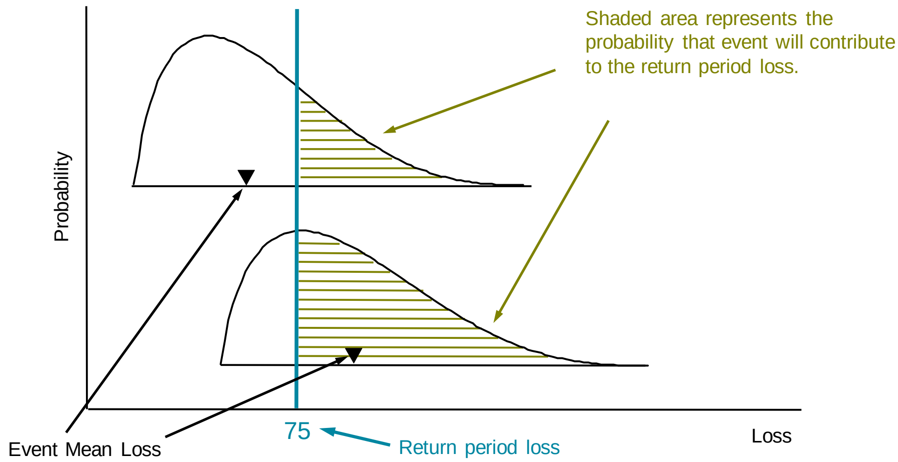
UNIT 5 – LEARNING OBJECTIVES

- Including Secondary Uncertainty in the EP Curve
  - Understand how uncertainty is incorporated into the OEP curve.
  - Assess which events are contributing to losses in excess of a specified return period or probability of exceedance

 November 7, 2019

At the end of this unit you should have a concrete understanding of each of the learning objectives listed on this slide.

OEP WITH SECONDARY UNCERTAINTY



 November 7, 2019 59

To examine the inclusion of secondary uncertainty in the EP curves, let us focus on two events from an Event Loss Table. Assume that we want to calculate the probability of exceeding a loss value of 75. Consider the situation depicted above of two events (black triangles) whose mean losses are nearly 75 - one event with a mean loss above this point (bottom event) and another with a mean loss below this point (top event). The mean losses for both events have a loss uncertainty distribution (black curves) that include the potential for the event loss to be greater or less than the modeled mean loss. Thus there is a positive probability that the top event will contribute to the loss value of 75 as shown by the area under the uncertainty distribution filled with green lines.

When an OEP curve <u>without</u> secondary uncertainty is constructed, all but the mean loss potential is ignored. Furthermore, all of the probability associated with the bigger event is included in the EP calculation, while all of the probability associated with the smaller event is excluded. When an OEP curve <u>with</u> secondary uncertainty is constructed, the potential for an event loss to fall below or exceed the mean loss is taken into account. Portions of each event curve, illustrated by the green shaded areas, are counted and contribute to the OEP curve.

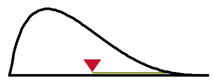
Often it is instructive to understand which events are contributing to losses in excess of a specified return period. This may be for comparison with legacy underwriting guidelines, with regulatory scenario analysis requirements, or to calculate the Excess Average Annual Loss (XSAAL). Excess Average Annual Loss is the contribution to the total average annual loss of events that exceed a certain threshold. The threshold is usually chosen as a point on the EP curve, such as the 100 or 250-year return period loss.

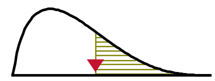
The inclusion of secondary uncertainty in the generation of the EP curve makes it difficult to determine which events are contributing to losses in excess of a specified return period or probability of exceedance. Given the uncertainty distribution around any event mean loss, there may be a positive probability that an event whose mean loss is below the specified loss level will still contribute to losses at or above this threshold – as seen in the first loss curve in the above graph.

OEP WITH SECONDARY UNCERTAINTY

Primary Uncertainty

Secondary Uncertainty





Event	Rate	Mean Loss	Std Dev	Exposure	α	β	$Pr(L_i>75)$	Exc. Rate
1	0.02	90	19	990	20.3070	203.0698	0.7783	0.0156
2	0.01	100	20	1,000	22.4000	201.6	0.9023	0.0090
3	0.04	80	17	970	20.2364	225.1303	0.5925	0.0237
4	0.09	20	12	920	2.6957	121.3043	0.0012	0.0001
5	0.03	70	18	980	13.9718	181.6332	0.3632	0.0109

Σ Exc. Rate = 0.0593

OEP(75) = $1 - e^{-0.0593}$
= 5.76%

 November 7, 2019 60

This slide is an illustration of how the contribution from each of the green-lined areas under the loss distributions on the previous slide are incorporated into the calculation of an OEP curve with secondary uncertainty. Note the numbers shown are not from actual events, and are used for purely explanatory purposes.

The light blue portion on the left of this table shows an event loss table similar to that output in RiskLink. For each event we have the annual rate of occurrence, the mean loss, the standard deviation of the loss, and the amount exposed to the event. The standard deviation is our measure of the secondary uncertainty for the event.

The medium-blue portion of the table shows numbers that are calculated from the ELT. α and β are the parameters of the beta distributions, and are calculated from the MDR and CV, as discussed in the Financial Modeling course.

Assume that we want to calculate the probability of exceeding 75. For each event, the beta distribution is applied to characterize uncertainty around the mean damage ratio. Thus, the probability that the loss is greater than 75 is calculated from the beta distribution. You can duplicate these numbers using the BETADIST function in MS Excel.

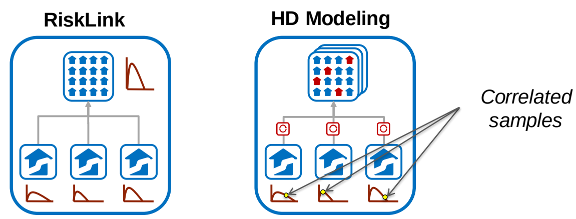
The last column shows the “exceedance rate” for each event. This is the annual event rate multiplied by the event probability of exceeding a loss of 75 and measures the contribution of each event to the total probability of exceeding 75.

Finally, sum of the exceedance rates is shown at the bottom of the two dark blue columns. Remember from Unit 6 of the Financial Modeling course that the Poisson distribution is used to characterize the frequency of events. Using the formula for the Poisson distribution, we can finally compute the probability of exceeding 75 = 5.76%.

One can get a sense of the importance of secondary uncertainty in the OEP by calculating the same value without incorporating it. In this case, the values in the Pr(L>75) column all become 1.0 for the first three events and 0.0 for the rest and the Exceedance Rate becomes the sum of their event rates, or 0.07. The resultant OEP(75) is 6.76%. If we were interested in the OEP for 105, however, neglecting uncertainty would give a result of 0.0% -- the largest possible mean loss is 100.

OEP WITH SECONDARY UNCERTAINTY: RISKLINK VS HD MODELING

- In RiskLink, the distribution of loss for each location in each event is propagated by tracking the mean and variance from location ground up to event gross loss
  - The method gives smooth results; however, it involves certain limitation and approximations.
- In HD modeling, calculations will be made as they occur in real contracts. Individual losses are sampled from a distribution and propagated resulting in many fewer approximations and limitations.
  - For aggregate data, significantly more calculations are required for smooth results. HD modeling results will be less *smooth* than the RL results.



This slide depicts the difference between the calculation of an event loss with secondary uncertainty in RiskLink and HD modeling. The illustration on the left shows how event losses are calculated in RiskLink. For each event the location level distribution of loss is propagated using the mean and variance from ground up to portfolio-level event gross loss using the correlation methods discussed in Unit 4. The analytical method produces smooth results, however, it involves certain limitations and approximations, such as the portfolio-level correlation weights (discussed previously) and the distribution of event-level losses.

The illustration on the right shows the same procedure in HD modeling. The location level losses are sampled from the corresponding distribution of loss (with correlation) and calculations can be made as they occur in real contracts. The losses are propagated resulting in fewer aggregate level approximations and limitations. Significantly more calculations/simulations are required for smooth results; consequently, the results of HD modeling may be less smooth than the analytical results produced by RiskLink.

The YLT unit of the Financial Model course discussed sampling of RiskLink results, a methodology used in the Simulation Platform (SP). The distinction here is that the sampling for SP takes place only for portfolio-level event losses (i.e. the top of the RiskLink graphic), whereas in HD modeling it starts at the location level and flows through the entire model.

AGENDA

- Unit 1: Types of Uncertainty Overview
- Unit 2: Parameter Uncertainty
- Unit 3: Secondary Uncertainty
- Unit 4: Multiple Locations
- Unit 5: Including Secondary Uncertainty in the EP Curve
- **Unit 6: Volatility Measures**
- Unit 7: Correlation Between Portfolios

Unit 6 provides detail on the definition, calculation, and appropriate use of modeled loss volatility measures.

UNIT 6 – LEARNING OBJECTIVES

- Volatility Measures
  - List the main drivers of volatility
  - Use the shape of the EP curve to compare volatility between two or more analyses
  - Explain why the coefficient of variation around the pure premium is higher for some peril regions than others
  - Identify the effect of volatility on excess layers

At the end of this unit you should have a concrete understanding of each of the four learning objectives listed on this slide.

SECONDARY UNCERTAINTY VS VOLATILITY

Secondary Uncertainty	Volatility
Uncertainty in the amount of loss, given that a particular event has occurred.	Uncertainty in the annual loss. Includes primary and secondary uncertainty.
Uncertainty around the mean event loss.	Uncertainty around the pure premium (not uncertainty in the pure premium).
Measured by event standard deviation, event CV	Measured by total standard deviation, CV (in user interface)
Beta distribution	OEP, AEP

 November 7, 2019

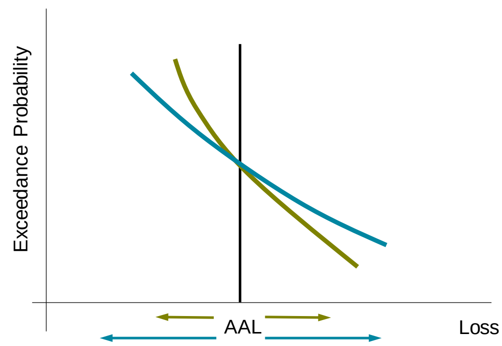
This slide provides comparative differences between how secondary uncertainty is defined and how volatility is defined. These are key definitions, and should be well understood before moving forward with the rest of the presentation.

Note that secondary uncertainty is applied to <u>single event</u> losses whereas volatility is applied to analysis output (e.g. portfolio AAL) from a <u>multi-event</u> analysis (e.g. an EP analysis).

Another important terminology that is important to clarify is the difference between uncertainty “around” a mean vs. uncertainty “in” a mean.
RiskLink reports the uncertainty around the AAL. This uncertainty is calculated from the probability distribution <u>around</u> the AAL. For example, based on 10,000 years of data, the AAL might be $100. However, in any given year, the loss will most likely be something other than $100. This loss should lie within the probability distribution around the AAL.

Uncertainty <u>in</u> the AAL implies lack of knowledge confidence in the AAL measure. For example, if the AAL is calculated from only three years of data, then the believability of the AAL is low and the uncertainty in the measure is high.

VOLATILITY AND THE EP CURVES



- No volatility implies a vertical EP curve at the AAL
- As the curve gets flatter, the range of possible losses increases, implying more volatility

 November 7, 2019 65

Different orientations of an EP curve represent different measures of volatility around the average annual loss. If there is no volatility, then the EP curve is a vertical line at the AAL and the standard deviation around the loss is zero (e.g. every year the same loss occurs). As the curve gets more horizontal, the range of possible losses increases, showing more volatility. In the case of the green curve, the standard deviation as a measure of loss volatility is smaller than the loss standard deviation for the blue curve.

VOLATILITY MEASURES – THE MATH

$$StDev = \sqrt{\sum_{i=1}^{N} \underbrace{L_i^2 \times r_i}_{\text{primary uncertainty}} \times \underbrace{\left(1 + cv_i^2\right)}_{\text{secondary uncertainty}}}$$

$$CV = \frac{StDev}{AAL}$$

$$CV = \frac{\sqrt{\sum_{i=1}^{N} L_i^2 \times r_i \times \left(1 + cv_i^2\right)}}{\sum_{i=1}^{N} L_i \times r_i}$$

L_i = mean loss for event i
r_i = annual rate for event i
cv_i = CV for event i
N = number of events

 November 7, 2019 66

As noted earlier, the volatility is measured by the standard deviation or coefficient of variation. Shown on this slide are the individual components in the calculation of standard deviation (StDev) and the CV. Note the difference in the designation between single event coefficient of variation (cv_i) and analysis loss output coefficient of variation (CV).

Use of these formulas allows you to develop an understanding of the drivers of modeled loss uncertainty. For instance:

- What happens to the CV if we double all L_i's? [CV remains unchanged]
- What happens to the CV if we double all r_i's? [CV decreases by a factor of sqrt(2)]
- What happens to the CV if we double all cv_i's? [CV increases. Amount of increase depends on size of cv_i's]

Exercise 2 will provide you with an opportunity to work with these parameter values in tables in the EDM and RDM shortly. We encourage you to complete this exercise when studying for the CCRA® exam.

Note: When we discuss the impact of doubling the event losses, it is under the assumption that everything else in the equation is being held constant, even though this is not possible due to the inverse relationship between the event loss and secondary uncertainty CV seen in slide 35.

MAIN DRIVERS OF VOLATILITY

- Frequency
  - The higher the frequency, the lower the volatility (CV)
- All other drivers

 November 7, 2019

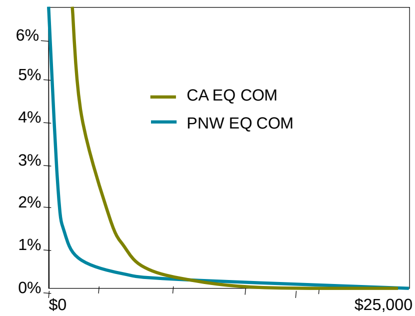
Based on the results of the variation in the parameter values in the CV formula, frequency (denoted by event rate r_i) has the direct impact [factor of sqrt(2)] on the volatility. Thus, frequency is the main driver of volatility.

VOLATILITY EXAMPLE - EARTHQUAKE

- The standard deviation is the square root of the variance of the aggregate loss distribution.

$$StDev = \sqrt{\sum_{i=1}^{N} L_i^2 \times r_i \times \left(1 + cv_i^2\right)}$$

- CV is the StDev divided by the mean pure premium.
- High CV for PNW exposure reflects the risk being dominated by infrequent large events.
- Low CV of CA exposure shows that there is less uncertainty in the losses due to more frequent events.



	CA EQ COM	PNW EQ COM
Pure Premium	273,216.73	130,817.79
Standard Deviation	1,250,861.14	2,902,756.19
CV	4.5783	22.1893

 November 7, 2019 68

This slide provides an excellent example of the comparison of portfolio level volatility. The blue OEP curve is from a Pacific NW earthquake analysis, and the green curve is from a California earthquake analysis. Note that the blue curve is less steep than the green curve, and therefore exhibits a greater loss volatility, as shown by the larger CV in the table. This reflects the nature of earthquake risk in the Pacific NW which is dominated high severity, low frequency events. The impacts of these events are less well understood than the higher frequency events in California.

EXERCISE 2

 November 7, 2019

Complete Exercise 2 and review the associated answers before continuing.

AGENDA

- Unit 1: Types of Uncertainty Overview
- Unit 2: Parameter Uncertainty
- Unit 3: Secondary Uncertainty
- Unit 4: Multiple Locations
- Unit 5: Including Secondary Uncertainty in the EP Curve
- Unit 6: Volatility Measures
- **Unit 7: Correlation Between Portfolios**

 November 7, 2019

Unit 7 explores the potential correlations between portfolios: grouping of portfolio results, quantification of portfolio correlation, and understanding the diversification impacts.

UNIT 7 – LEARNING OBJECTIVES

- Correlation Between Portfolios
  - Calculate the catastrophe risk correlation coefficient between portfolios
  - Understand the process of grouping portfolio modeled loss results
  - Define portfolio diversification

 November 7, 2019 73

At the end of this unit you should have a concrete understanding of each of the three learning objectives listed on this slide.

CORRELATION BETWEEN PORTFOLIOS

- Relationship between the annual losses of the portfolios
  - “Across event” correlation
- Measure of the amount of diversification benefit one will achieve by combining the portfolios
- One measure is the correlation coefficient, *r*
  - Recall that *r* quantifies the degree to which two random variables are linearly related
  - $-1 \le r \le 1$
  - A high correlation coefficient implies that if the losses are large for one portfolio, they are likely to be large for the other portfolio
  - Negative correlations are rare for catastrophe risks
  - Common events drive the correlation coefficient

 November 7, 2019 74

One way to measure the diversification benefit of writing multiple portfolios is the correlation coefficient. The correlation coefficient, ρ, has a domain of -1.0 to +1.0. Perfect positive correlation is represented by +1.0. Perfect negative correlation is represented by -1.0.

Common events for each portfolio drive the value of the correlation coefficient. So, perfect correlation (+1.0) means that the ELTs between two portfolios will be identical, except for the mean loss of each event. A correlation of 0.0 means that there are no common events between the two portfolios, maximizing the diversification benefit.

GROUPING PORTFOLIOS – COMBINING EVENT LOSS TABLES

- For dissimilar events: append tables
- For like events: combine results
  - Sum mean losses, exposures, and correlated piece of the standard deviations
  - Square root of the sum of the squares of the independent piece of the standard deviations

 November 7, 2019

To understand the correlation between portfolios, we have to examine each individual portfolio and the resultant portfolio group.

You should recall that each group creates a new Event Loss Table (ELT) comprising the union of ELTs from the portfolios to be grouped (underlying ELTs). First, the underlying ELTs in each of the portfolio results are examined to ensure that they were analyzed with the same event rate set (e.g., attempting to group a U.S. hurricane analysis that utilizes the 2011 stochastic event rate set with another U.S. hurricane analysis that uses the historical rates). Next, the individual events are examined. If the events have different event IDs, they are appended to the new ELT. If the events have identical event IDs, the mean losses, exposures, and standard deviation are combined. The new ELT is then sorted on event mean loss and the return period calculations and statistics are calculated.

CORRELATION BETWEEN PORTFOLIOS – THE MATH

$$\sigma_{a+b}^2 = \sigma_a^2 + \sigma_b^2 + 2\rho_{ab}\sigma_a\sigma_b$$

Solve for ρ_{ab} :

$$\rho_{ab} = \frac{\sigma_{a+b}^2 - \sigma_a^2 - \sigma_b^2}{2\sigma_a\sigma_b}$$

σ_{a+b} = standard deviation of group of portfolios *a* and *b*

σ_a = standard deviation of portfolio *a*

σ_b = standard deviation of portfolio *b*

ρ_{ab} = correlation coefficient of portfolios *a* and *b*

This is the formula to derive the correlation coefficient between two portfolios. Note that you need to know the standard deviations of each individual portfolio and the resultant group to complete the calculation.

Note: The upcoming Exercise 3 will provide you with an opportunity to work with these parameter values from EP analysis output. We encourage you to complete this exercise when studying for the CCRA exam.

GEOGRAPHIC CORRELATION OF RISK

EUROPE WIND EXAMPLE

Annual probability of an event causing greater than 100-year return period loss in both countries:

Belgium and Germany	0.255%
Belgium and France	0.100%
Belgium and the U.K.	0.117%
Belgium and the Netherlands	0.263%
Germany and France	0.064%
Germany and the U.K.	0.055%
Germany and the Netherlands	0.319%
France and the U.K.	0.073%
France and the Netherlands	0.028%
U.K. and the Netherlands	0.108%

 November 7, 2019 77

As we prepare for Exercise 3, let's look at wind risk in Europe. The chart above shows the annual probabilities of a single event causing a 100-year return period loss in both counties. Note that complete independence would be calculated as the annual probability of an event causing loss equal to the 100-year return period in both countries which is 1% x 1% = 0.01%.

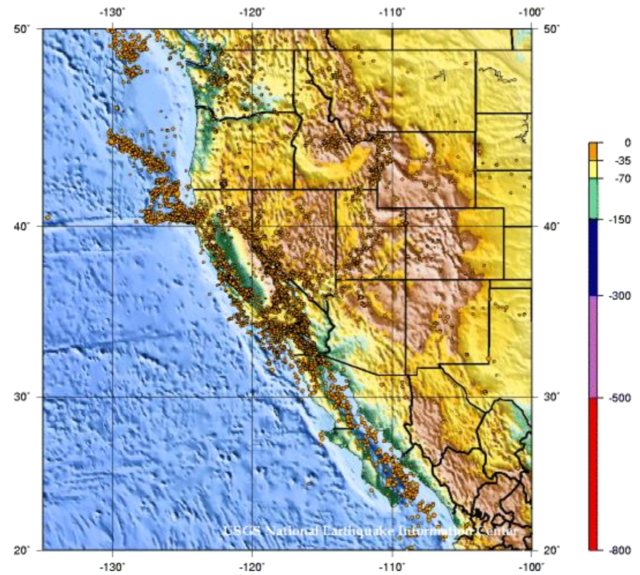
EUROPE MAP

  November 7, 2019 78

You may use this map of Europe as you investigate your findings in Exercise 3.

SEISMICITY OF THE WESTERN U.S.

1990-2000



November 7, 2019

You may use this map of California as you investigate your findings in Exercise 3.

EXERCISE 3

At this point in the course you should now complete Exercise 3, the objective of which is to calculate and provide an explanation of the correlation coefficient between Belgium and Germany portfolios as well as between Pacific NW and California portfolios.

GEOGRAPHIC CORRELATION OF RISK

EUROPE WIND EXAMPLE

Annual probability of an event causing greater than 100-year return period loss in both countries:

		Correlation Coefficient
Belgium and Germany	0.255%	0.62
Belgium and France	0.100%	0.35
Belgium and the U.K.	0.117%	0.37
Belgium and the Netherlands	0.263%	0.71
Germany and France	0.064%	0.23
Germany and the U.K.	0.055%	0.25
Germany and the Netherlands	0.319%	0.68
France and the U.K.	0.073%	0.25
France and the Netherlands	0.028%	0.18
U.K. and the Netherlands	0.108%	0.37

This chart shows some selected correlation coefficients along with the probability of 100-year return period events discussed earlier. There is a strong tracking between this probability and the correlation coefficients, meaning that as the correlation coefficients increase relative to other portfolio groups, so do the annual probabilities of a 100-year return period loss.

UNITS 4-7 SUMMARY: MEASURING UNCERTAINTY

- Loss correlation between locations must be taken into account when analyzing multi-location accounts or portfolios.
- Incorporating secondary uncertainty in the calculation of the EP curve allows for the potential loss contributions of several events to the return period losses.
- Risk and portfolio volatility are assessed by using measures of primary and secondary uncertainty.
- Correlation between portfolios can be measured by comparing individual portfolio volatility measures with grouped analysis results.