Portfolio Accumulation of Losses from Primary Insurance Policies

Loss accumulation has a special place in catastrophe modeling: It is the backbone of any enterprise risk modeling platform that includes support for detailed location-level modeling. Although loss accumulation sounds like a straightforward exercise, it is not. This process is both complex—requiring modern and sophisticated statistical methodologies—and computationally demanding.

Accurately modeling the dependencies in loss accumulation is a component of AIR's overall strategy of propagating and reporting all modeled uncertainty, due to the central role that loss accumulation plays in a catastrophe modeling platform. A robust catastrophe model must be able to do two things: roll up loss results from low-level granular analyses—losses by insurance coverage, by location and by event—to top-level insurance portfolios, including pre- and post-catastrophe insurance and reinsurance net loss; and propagate all modeled uncertainty to all financial and actuarial operations for insurance and reinsurance and for all reporting perspectives.

In a catastrophe modeling platform, there are a few general tiers of loss accumulation. The tiers themselves are defined by the generalized structure of the insurance portfolio and reflect market conditions, portfolio structuring, and risk management practices. These accumulation tiers also have an explicit and inherent element of geographical distance, and therefore include some dependencies based on their spatial proximity.

Propagating the uncertainty inherent in these loss calculations involves accumulating the loss distributions of multiple risks. There are a number of accumulations involved in this procedure and in the process of uncertainty propagation in general. These accumulation tiers can be generalized and thought of in the following order:

- Insurance coverages to an insured location
- Locations to sub-limits and other groups and campus structures
- Sub-limits to excess layers
- Layers to a contract and portfolio or book of business

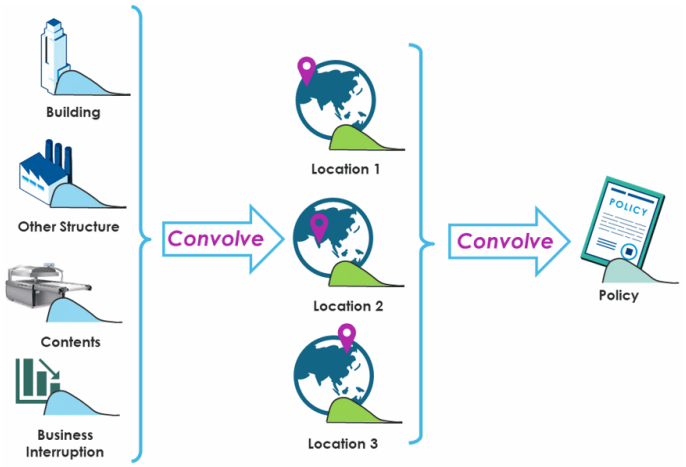


Figure 1. In AIR's current methodology each type of risk shown on the left is treated as independent; the loss distributions are convolved at both the risk and location level.

The accumulation of loss distributions itself also serves the purpose of capturing and modeling the dependencies between the underlying risks, which are the insured locations or groups of locations in a policy and are the contracts, layers, and other (re)insurance structures in a book of business. These dependencies vary by insurance coverage, geographical proximity, and insured peril. We have propagated these dependencies to all modeled perspectives and report them to our clients in the construction of all of the EP curves and all of the results in the year/event/loss tables in our databases.

Why Are We Introducing a New Generation of Loss Accumulation Methodologies Now?

How we resolve the business workflows and solve the industry's modeling challenges today goes back to the fundamental task we are striving for: to accurately propagate all modeled dependencies and uncertainty within our models and through our software. Currently, we perform all loss accumulation procedures with statistical convolution. As a loss accumulation methodology, statistical convolution does a good job rolling up the expected values while preserving the uncertainty of the distributions, from the insurance coverage to the book of business for both total event and annual losses. This

methodology also propagates the uncertainty from location to policy and portfolio because it builds up a probabilistic loss distribution for the policy. It is not well equipped, however, to model general dependencies during loss accumulation. Statistical convolution is agnostic to distances, geographical proximity, and to the spatial loss accumulation patterns, which vary based on the specific peril being analyzed. Instead, it applies the same statistical principles in all of these cases.

Newly Available Data Presents an Opportunity

As we have developed trusted relationships with our clients, we have had the opportunity over the years to gain access to a wealth of claims data from historical catastrophe events, which has allowed us to study this problem of loss accumulation in great detail. With vast amounts of new claims data at our disposal, we were able to calibrate our financial engine to provide the best possible methodologies for accumulating losses and develop new techniques for deriving and validating a new spatial correlation component for the entire Touchstone® platform. Such calculations are both sophisticated and computationally demanding. To handle the computational demands of such sophisticated calculations without adversely affecting performance, we have introduced new computational techniques optimized for memory management and high performance. These allow us to implement the new loss accumulation algorithms in a performant, stable, and scalable manner. These loss accumulation algorithms are specifically important for accurately representing multi-risk, multi-tiered, multi-peril contracts, where realistic loss accumulation is critical to achieving realistic results.

Accurately representing the accumulation of losses is critical: It has a central role in building up the loss distributions for each of these tiers and ensuring that the terms and conditions of the policies are applied correctly following the appropriate financial and actuarial principles. The new generation of probabilistic distributions, constructed using the methods described earlier and on which such multi-layered insurance and reinsurance contracts are applied, better reflect the loss and claims experience of firms in the industry whose claims data we have studied. With this new methodology in place, the application of these terms and the computation of these gross, reinsurance and net losses becomes significantly more accurate, realistic, and usable for underwriting, reserving, and risk management tasks. These new algorithms enable us to account for the spatial distances between risks, or their proximity to one another, and to realistically represent the appropriate degree of correlation of the losses incurred between locations, based on the peril(s) being modeled, just as one would expect when observing the impacts of a natural catastrophe.

Developing a Methodology to Capture Correlation

Accumulating various spatially dependent risks has been a challenge within the insurance industry, and statistical methods have been developed in academia and corporate research to address this challenge (see *Wójcik et al.,* 2019 and references therein in "References"). Technically speaking, we are solving the problem of computing an arbitrary sum of risks:

$$S = X_1 + ... + X_d$$

The word "arbitrary" in this context means "positively dependent" (i.e., positively correlated). Risks at locations are represented by the random variables $X_1, ..., X_d$, characterized by the loss distributions $F_1, ..., F_d$, and the covariance matrix $\sum$, which describes the geospatial dependencies between these risks. We assume that our arbitrary sum S is enclosed within two bounds: the independent sum $S^{\perp}$, where risks are assumed to be independent, and the comonotonic sum S^{+}, where risks are assumed to be maximally correlated. So, the sum S will be the riskiest or have the largest variance when risks are *comonotonic* (i.e., maximally correlated) and conversely, the sum will be the least risky if the risks are independent.

We assume that the distribution F_S of the arbitrary sum S is approximated by the weighted mixture of independent and comonotonic sums, referred to as the "mixture method":

$$F_S(s) = (1 - w)\, F_S^{\perp}(s) + w\, F_S^{+}(s) \text{ where } 0 \le w \le 1, \text{ for all } s$$

The weight w in the equation is dependent on how spatially correlated the pairs of risks are.

To provide insurance covering natural catastrophes, one needs to have a thorough understanding of both the physical attributes of the risk and of the peril itself. We have selected accurate statistical models to render these physical dependencies and translate them into business metrics. Our models realistically represent how the geospatial proximity of risks impacts the interdependencies in catastrophe-related damage. For example, if a hurricane makes landfall over a set of locations that are physically close to each other, these locations will likely have similar damage. This is because some aspects of spatial correlation in damage are implicitly accounted for by the model-predicted wind field, where locations near one another experience similar wind speeds.

Accounting for Spatial Correlation

Another type of correlation is correlation that is conditional on the catastrophe model loss estimates. Intuitively, if the predicted wind speed for an area is higher than the actual wind speeds, then the financial losses represented by insurance claims in that area are likely to be lower than estimated by the model. This means the actual losses will be correlated with each other, relative to the model estimate. Such conditional dependency is known as the *correlation of model errors*.

Insurance underwriting and risk management have developed strong financial and actuarial nuances for specific insured perils; our new methodology can successfully replicate these conditions. Correlation of model errors varies by peril and reflects the fact that damage patterns are shaped by the spatial scales characteristic of different perils. For example, the length of a typical damage path for a tornado is a few kilometers, while its width does not usually exceed 200 meters. The largest earthquakes, on the other hand, can have destructive and deadly effects up to 300-400 km away from the epicenter. To account for these spatial scales, our methodology identifies the size and hierarchy of nested grids that fully determine the correlation between model errors for a particular peril (see *Einarsson et al., 2016* for details).

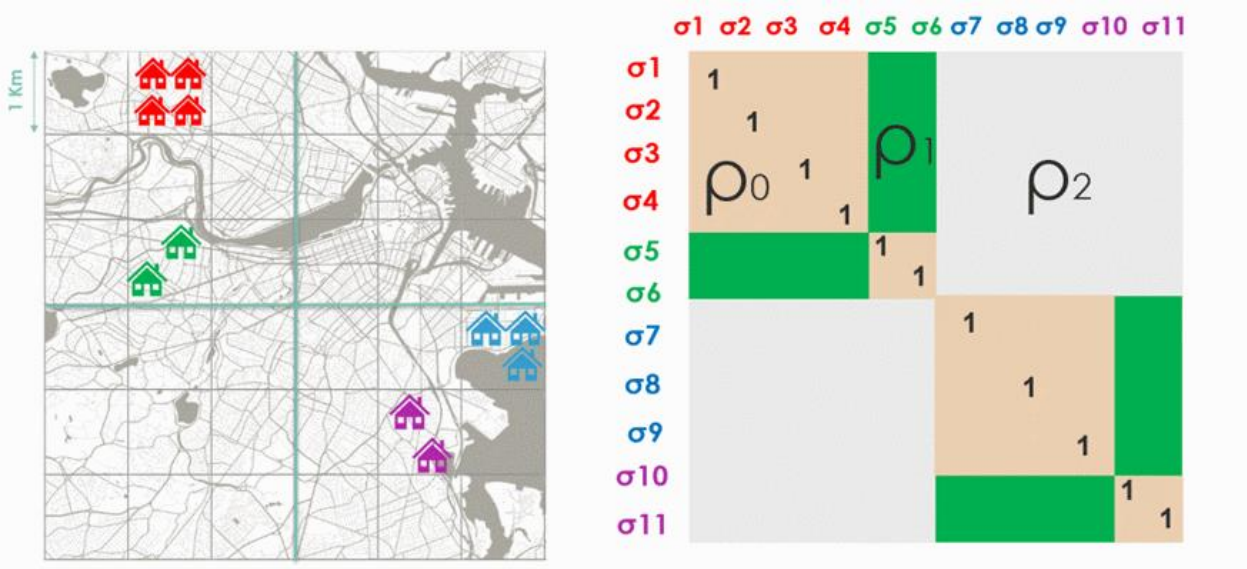


Figure 2. An example of a nested grid system superimposed on 12 locations (left); corresponding block-diagonal correlation matrix, which describes spatial dependence between these locations (right).

An example of such a nested 1- and 3-km grid system is shown in Figure 2. In this example, there are 11 locations that experienced damage from a flood. The block diagonal matrix on the right represents the correlation between these locations. For instance, the correlation between any two locations belonging to the same 1-km bin (for example, the group of red buildings) is referred to as ρ_0. Likewise, the correlation between any two locations belonging to the same 3-km bin, but different 1-km bins, is referred to as ρ_1. It is also possible for two locations to fall into distinct 3-km bins. Dependence between these locations is captured by the correlation coefficient ρ_2. Our nested block diagonal correlation matrix is chosen for computational efficiency. Organizing locations into grid blocks prevents the computationally expensive calculation and storage of a full correlation matrix. “Stationarity” is an underlying assumption that allows us to create this type of grid because it makes the estimates of spatial

correlation coefficients insensitive to shifts of the nested grid system. This assumption will be discussed further in the "Validation" section.

Accounting for the Order in Which Risks Are Accumulated

There is a complication our risk aggregation algorithm must account for in addition to the spatial correlation model we've discussed, namely the order in which the risks are being accumulated. In the business world, this complication is directly related to the structure of each policy and the order in which financial terms are applied. In the catastrophe modeling world, we need to incorporate this detail both to accurately reflect different policy structures and to ensure that the loss aggregation algorithm is as accurate and computationally efficient as possible.

For example, let us consider the problem of aggregating five dependent risks. Note that these five risks are not just numbers that can be added up, but rather dependent random variables; the way in which we sum these five dependent risks up determines the resulting loss characteristics of the total risk. The problem is non-trivial, as there are 236 ways in which these five risks and their partial sums can be combined. The process of summation can be visualized as an aggregation tree, in which every branching node represents the addition of two or more random variables. All possible aggregation trees for five dependent risks are displayed in Figure 3. For example, in the tree highlighted in blue, we first add X_4 and X_5 then we cointegrate X_3 and then add up this partial sum to X_1 and X_2 directly at the root node. In contrast, in the tree highlighted in orange, we first add X_1, X_4 and X_5 and then add X_3 before finally adding in X_2. There are three operations happening at every branching node of each aggregation tree:

1. Convolution, which assumes that random variables are independent.
2. Computation of the distribution of the comonotonic sum, which assumes that risks are maximally correlated.
3. Computation of the mixture distribution, which is a convex* combination of independent and comonotonic case; the weight in that mixture is a function of correlation between risks.

** Convex here means that the weights (i.e., w and 1-w) of each term are non-negative and sum to 1.*

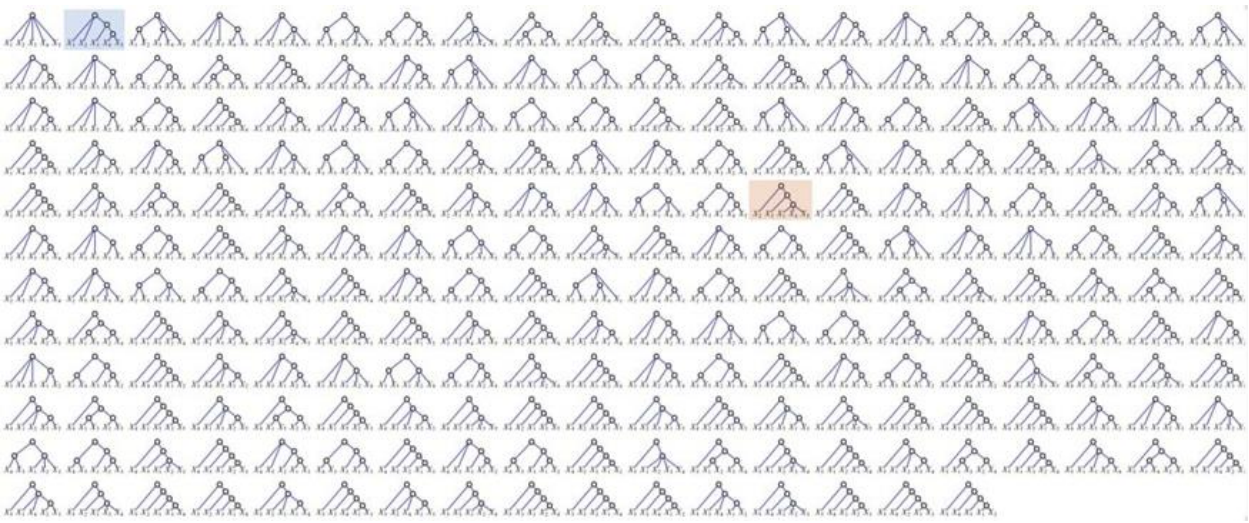


Figure 3. Aggregation trees represent a summation of five dependent risks.

In principle, both the order of loss accumulation and the grouping of risks need to be accounted for to effectively assess and manage risk and abide by accepted statistical principles. This has informed our choices, so in AIR's Next Generation Financial Module (NGFM), we use a combination of direct and hierarchical trees, as we see in Figure 4 for both ground-up and gross loss accumulation. Direct trees are computationally the most efficient because the weight in the mixture method only needs to be calculated once—at the root node. Whenever partial correlation between groups of risks and/or their partial sums is of interest (e.g., because of a particular portfolio structure), we use hierarchical sequential or hierarchical general trees.

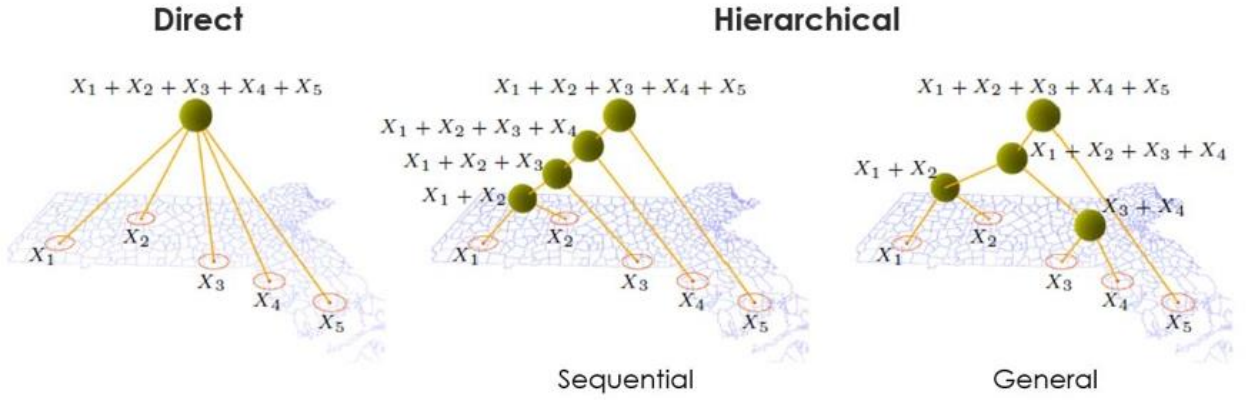


Figure 4. Two types of loss aggregation trees used in NGFM; the topology of a particular tree depends on (i) the type of results requested to be saved and (ii) the different tiers and types of financial terms.

Numerical Example of Loss Aggregation

To give you an example of how risk aggregation with spatial correlation works in practice, let us consider three risks characterized by marginal discrete density functions (fx_1, fx_2, and fx_3) of loss distributions

shown in black on the left in Figure 5. The distribution of the sum of these three risks, assuming independence, is shown at the top of this figure, in green. This is what Touchstone® 2020 computes using convolution. The distribution of the sum when the three risks are maximally correlated (i.e., comonotonic) is shown at the bottom of the figure, in red. Note that the shape of the red distribution is very different from the shape of the green one. Maximum correlation implies that the red distribution has the maximum variance attainable for the three marginal distributions shown on the left. In NGFM, we use a weighted mixture of the green and red distributions to represent the distribution of the sum of spatially correlated risks. Such mixtures are shown in the center, in blue, for correlation weights of 0.3, 0.5, 0.7. When the correlation between risks is small (e.g., 0.3) the shape is similar to what we would currently obtain from Touchstone. When the correlation is larger (e.g., 0.7), we approach the comonotonic case, and for intermediate correlation (e.g., 0.5) we have a balanced combination of the green and red distributions. The point to remember here is that spatial correlation has a direct impact on the shape of the distribution of the sum of risks and therefore a direct impact on both the ground-up and gross loss metrics.

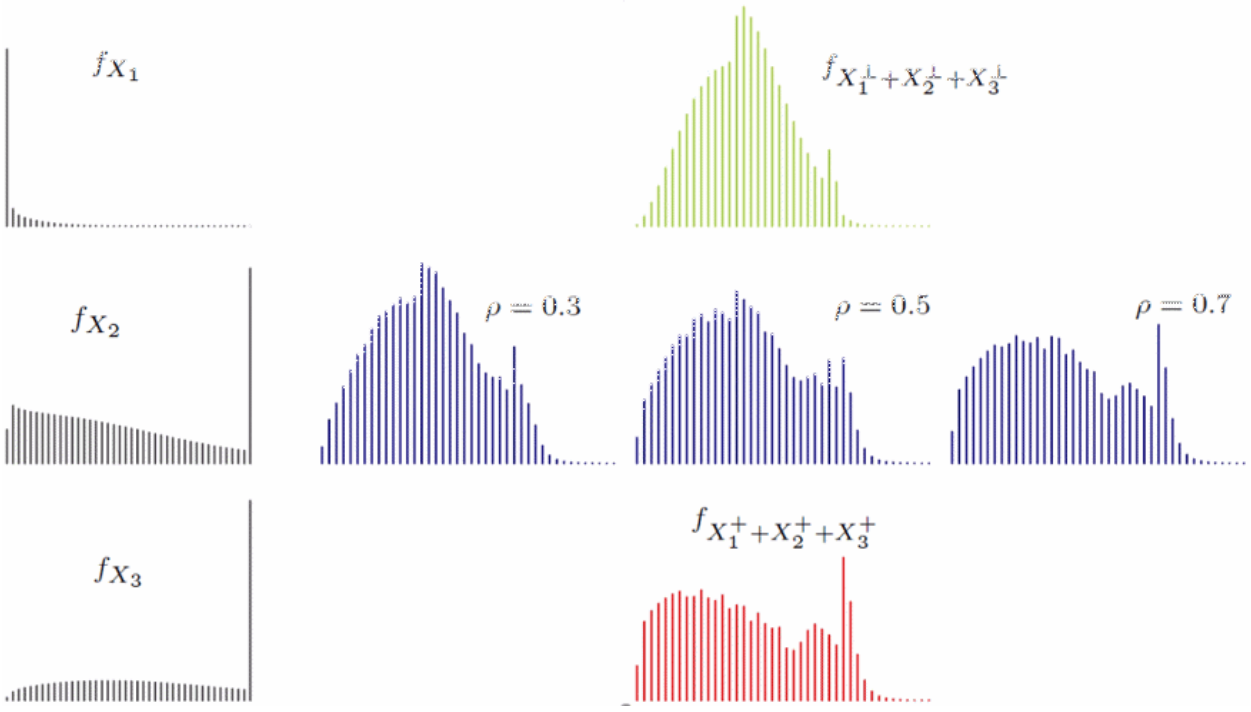


Figure 5. An example of sequential loss aggregation for three risks assuming different levels of positive dependence.

Validation of Our Methodology

We have conducted thorough and comprehensive model validation studies that demonstrate that our new methodology best reflects the physical reality of a natural catastrophe and closely approximates

the claims risk management workflows of insurers. These studies ensure that the methodology we've developed is modern, realistic, and addresses the needs of the industry today.

Predictability Analysis

One way to validate the accuracy of our modeled loss estimates for the aggregate risk of a portfolio including spatial correlation is to run a portfolio rollup for historical catastrophe events and compare the predicted distribution of the total loss with the sum of insurance claims for that event. An example for Hurricane Frances is shown in the upper left panel of Figure 6. The red curve is the distribution of the total loss predicted by our loss aggregation procedure and the green dashed line represents the sum of the claims.

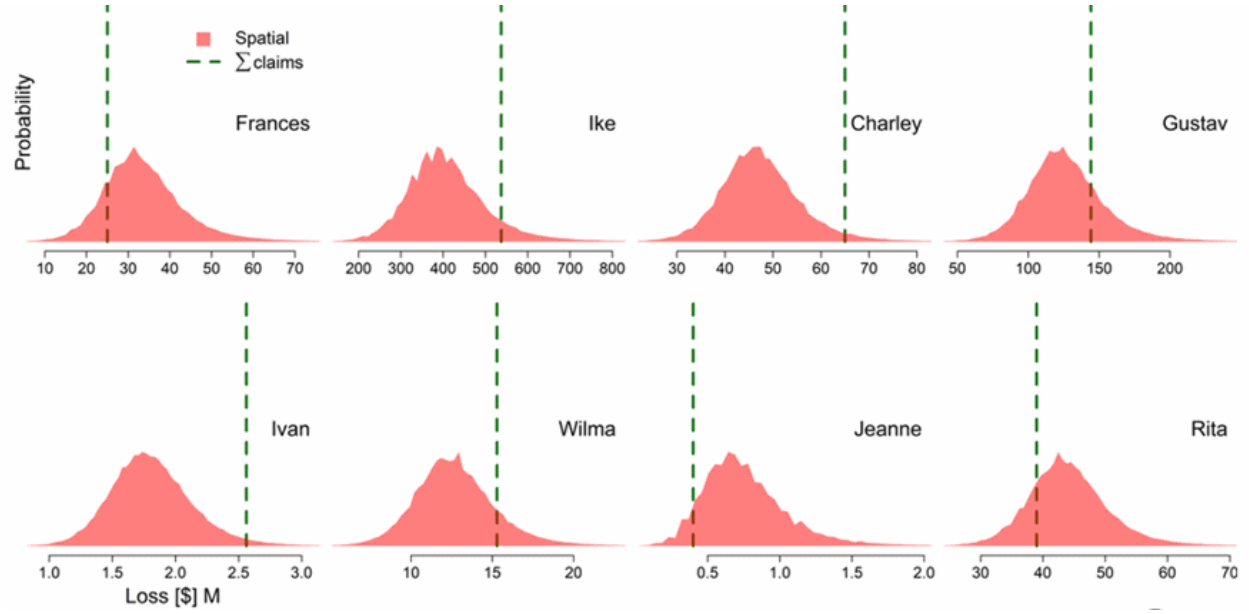


Figure 6. Validation of spatial correlation model in NGFM using historical claims data for U.S. hurricanes.

Good agreement can be seen in Figure 6, as the sum of claims falls within the potential support of loss distribution, and this sum has a relatively high probability within that distribution. We repeated this procedure for other historical hurricane events. Figure 6 shows that generally the sum of the claims falls within the predicted loss distribution for a particular event.

This kind of comparison is limited, however, by the fact that we are comparing a full probability distribution with only one number, i.e., one historical realization of the sum of claims. Ideally, we would like to have many such realizations for one historical event and would expect that the distribution of the sum of claims will closely reflect the total loss distribution estimated using our new aggregation scheme. One way to artificially make such a comparison feasible is to randomly draw subsets of available claims

data for a particular historical event and repeat the validation procedure for many random subsets. Preliminary results of this type of analysis indicate that the total loss distributions estimated using our mixture method compare favorably with empirical distributions of the sum of claims.

Non-Stationary Covariance Model

Recall from earlier in this article that one of the core assumptions underlying this model is stationarity. This assumption makes the estimates of spatial correlation coefficients insensitive to shifts in the nested grid system and allows us to use a simplified, yet computationally efficient, block diagonal correlation approach. To validate this methodology and see if this assumption is realistic, we compared loss estimates based on our stationary model to those obtained from a more complex and very computationally expensive non-stationary alternative (see *Higdon,* 2002 for details). Statistical properties of non-stationary phenomena vary in space. A good example is a vortex as shown in the left panel of Figure 7. This type of non-stationarity is common in hurricane modeling. The velocity field for a hurricane exhibits strong directional preference that varies in space as the air rotates around the eye of the storm. For comparison, in the right panel of Figure 7, we plotted the model errors for Hurricane Ike (2008). These errors exhibit locally stationary pattern in the areas of Texas and Louisiana, where the hurricane made landfall and caused the most damage. Globally, the model error field for Ike is non-stationary due to the distinct blue pattern caused by its move on an east-northeastward track.

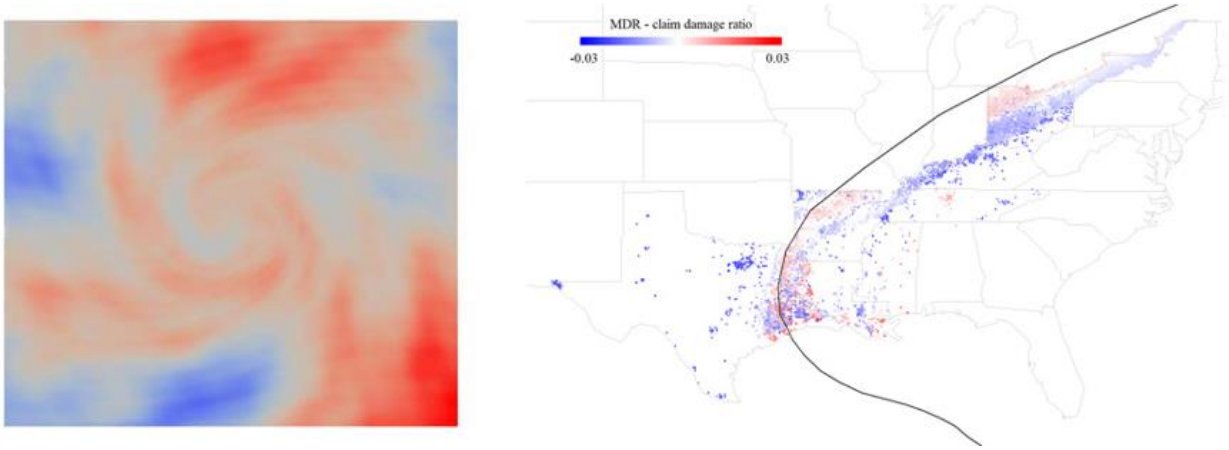


Figure 7. Non-stationary vortex simulation (source: Kleiber, William. (2016). High resolution simulation of nonstationary Gaussian random fields, Computational Statistics & Data Analysis. 101. 10.1016/j.csda.2016.03.005 (left); model errors for Hurricane Ike event in September 11, 2008. AIR posted estimated losses of between USD 8.2 billion and USD 12.2 billion on September 13, 2008 (right).

Figure 8 shows the validation results of our stationary spatial correlation model; we compared the total loss distributions (in red) with those obtained from a very computationally expensive non-stationary alternative (shown in blue) for a few historical hurricane events. The figure shows that while there are some differences in shape, those differences represent the trade-off between the computational speed and the accuracy of the results; but in general, these distributions are similar.

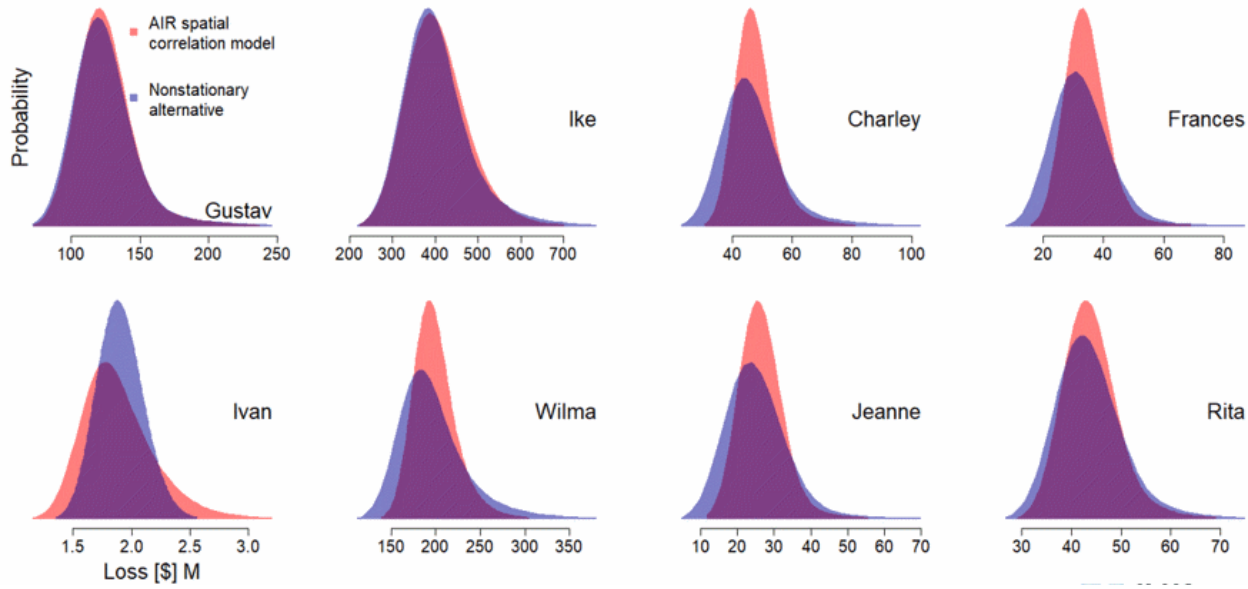


Figure 8. Validation of NGFM event loss distributions using non-stationary correlation model.

Many statistical loss accumulation algorithms are used in the industry, each of which solves various problems and has its own advantages and disadvantages. At AIR, we have developed and validated a methodology that optimally combines the solution to multiple challenges—modeling dependencies, proximity, and peril-specific correlation—while also reflecting actual industry practices in terms of the order in which policies are applied and claims are grouped together for accumulation and risk management. With these new features incorporated into our Next Generation Models, the underlying loss distributions—the foundation upon which insurance and reinsurance policies are applied—provide a more accurate and realistic view of risk.

References

Einarsson, B.; Wójcik R.; Guin J. Using intraclass correlation coefficients to quantify spatial variability of catastrophe model errors. Paper Presented at the 22nd International Conference on Computational Statistics (COMPSTAT 2016), Oviedo, Spain, August 23–26, 2016. Available online.

Wójcik, R.; Liu, C.W.; Guin, J. Direct and Hierarchical Models for Aggregating Spatially Dependent Catastrophe Risks. Risks 2019, 7, 54

Higdon, D. Space and space-time modeling using process convolutions. In Quantitative Methods for Current Environmental Issues. 2002, editors: Anderson, C., Barnett, V., Chatwin, P., and El-Shaarawi, A., 37-56. Springer London