# Model Behavior:
## Unreliable, yet indispensable

Written by Ivan Maddox

An explanation of why it is important to understand the limitations of risk models.

INTERMAP
Answers Now™

# Table of Contents

Introduction – 2

FEMA Flood Maps – 3

Cat Models – 4

Historical Claims Data – 6

Conclusion – 8

# Introduction

"All models are wrong, but some are useful."

With advances in computing capability and the incredible amount of raw data available to feed advanced computations, modeling has entered a golden age. Models are rapidly becoming more and more sophisticated, and new models are being built to help understand phenomena as diverse as the climate for the whole world, the birth and death of galaxies (not just stars, but galaxies), the spread of disease through individuals and populations, and even (most dauntingly) stock markets. As models increase in sophistication and application, it is becoming ever more important to remember the words of George Box, a pioneer in the statistical science that drives the type of modeling we have today. Here is the full quote, commonly abbreviated (as above):

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."
—George Box

In other words, it's more important to understand the limitations of models than it is to build robust models in the first place. Only by understanding the limits of a model can sound evaluations of modeled results be made. To illustrate this point, below is an exploration of the limits of three common types of models used in property insurance: FEMA's Flood Insurance Rate Maps (FIRMs), Cat Models, and actuarial models based on historical claims.

# 1 FEMA Flood Maps: An Incomplete Picture of the Risk Story

Perhaps the most commonly relied upon method of modeling flood risk in the United States is the Federal Emergency Management Agency (FEMA)'s Flood Insurance Rate Maps (FIRMs). That's because when it comes to flood risk in this country, FEMA is the authority — and rightly so. They have decades of engineering intelligence built into their FIRMs and they are refined regularly. In the end, though, they are models, and their limitations are as important as their strengths.

A recently published report that investigated how well the FIRMs predicted flooding from Superstorm Sandy in 2012 illustrates that while they performed well in some areas, there is also room for improvement. That's because there is no flood model, or model of any natural catastrophe, that will be 100% accurate. In fact, predicting 100% of a flood event is usually a sign of a weak flood model: Imagine a model where everything within 100 miles of water is labeled as High Risk. Flood models display their quality in the zone between, "You don't need a model to know that's a flood risk" and "that place will never flood." In statistical terms, the sweet spot is around 75% or 80% of flooding predicted. This is about where FEMA's FIRMs were on Sandy.

However, for insurance companies, understanding where 75% of their flood-risk lies is not good enough.

But after all the time and effort FEMA has spent perfecting the FIRMs, how is it possible to improve on what the FIRMs have to say? Luckily, there is more to understanding flood risk than statistics.

The first way is to compare multiple flood models. Where the models agree are the areas where the flood risk is probably well modeled. However, they will disagree in that zone between the obvious results, and looking closer at the areas of disagreement will be fruitful.

Another way to improve an understanding of flood risk is to use elevation data in tandem with flood risk maps. Regardless of how the modeling has been done, the process incorporated elevation information in some way. Using the actual elevation data lets someone visualize directly where water will flow and create the higher risk.

InsitePro™, Intermap's location-specific risk solution, now has features that allow users to do all this, and more. Through an intuitive profiling tool, InsitePro users have access to the world's best global elevation datasets. With both single and multi-location capability, it is now easy for InsitePro users to compare flood models and quickly understand risk from flood and other perils based on more than a single perspective.


*This location along the Monongahela River in Pittsburgh, PA, is outside of the flood risk area according to FEMA (on the left) even though this region has a history of heavy flooding.InsitePro (on the right) uses elevation data to show that this location is in a high-risk area.*

# 2 Cat Models: Incredible Technology Limited by Nature

Another popular tool used by property insurers for risk analysis—and by far the most complex—is the cat model. Short for catastrophe, cat modeling uses computer-assisted calculations to estimate the losses that could be sustained due to a catastrophic natural event. Perils analyzed include flood, hurricane (wind damage and storm surge), earthquake, tornado, hail, wildfire and winter storm. The principle function of cat models is to help insurers prove their financial solvency.

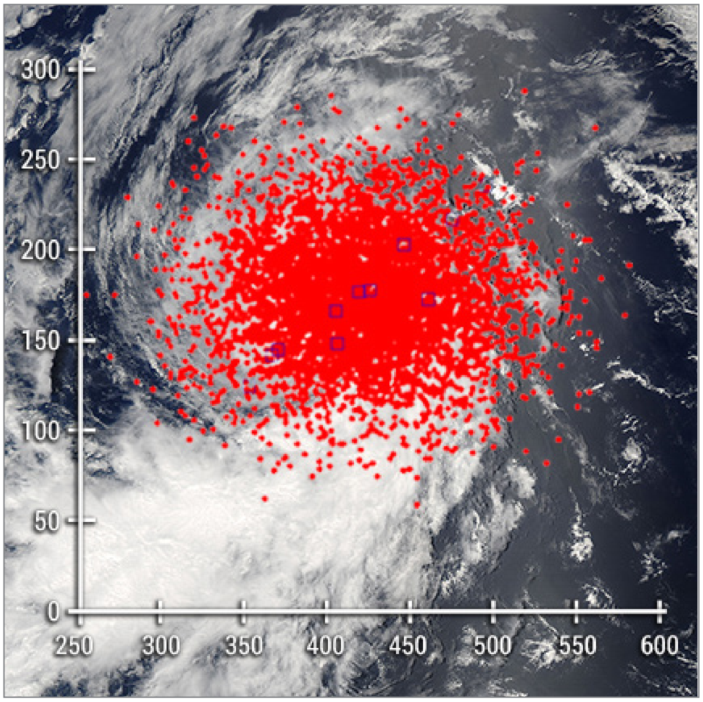
While most large companies, financial or otherwise, can prove and convey their solvency with accounting metrics, property insurers have a more difficult task. At renewal time, when they write the bulk of their policies, insurers are flush with cash as they collect premiums from their customers. An accounting snapshot would seem to show that they have enormous cash reserves and investments, but no obligations. In reality, their obligations are the claims they need to pay over the next year as the properties they are insuring are damaged. But the problem is that nobody knows how many claims they will need to pay.

Cat models solve this problem by calculating several statistical figures based on a carrier's portfolio of insured properties. The two most significant are:

- **Probable Maximum Loss (PML)**: An estimation of the most likely worst case for any given year. (Notice this is different the Maximum Possible Loss – that figure would be the entire portfolio being destroyed, which is very unlikely.)
- **Annual Expected Loss (AEL)**: An estimation of how much the insurer can expect to pay in the coming 12 months.

Cat models arrive at these calculations by utilizing three interconnected modules: hazard, exposure and financial.

- **The hazard module** estimates the risk of natural catastrophes affecting each of the locations within an insurer's portfolio. Which perils are evaluated depends on the location of the properties, with flood, earthquake, and wind/hurricane being the most common.
- **The exposure module** takes the hazard information and translates the chances of something happening into monetary values based on the physical attributes of the properties. Factors such as construction material, what the building is used for, and what is inside the building are all used to calculate the exposure.
- **The financial module** converts the exposure information into the output specifically needed to convey the solvency of the company; i.e. the balance between money available to pay claims and the amount of claims they might need to pay. These calculations are based on policy information (premiums, deductibles, and limits).



At the heart of all this is an estimation of what might happen in the coming 12 months. Monte Carlo statistical methods are used to create “event sets” — vast sets of different global scenarios that might happen in a given year based on the likelihood of each little aspect (including meteorological and climatic conditions) of a scenario happening. These statistical, or stochastic, methods allow the cat model to have a proxy for millennia of experience that is, of course, impossible to have based on reality.

Of course, we must always remember the words of George Box: all models are wrong, but some are useful. As sophisticated as they are, cat models are never accurate because it is impossible to estimate how a real natural catastrophe will impact property. Indeed, every event tends to expose a gap in the models.

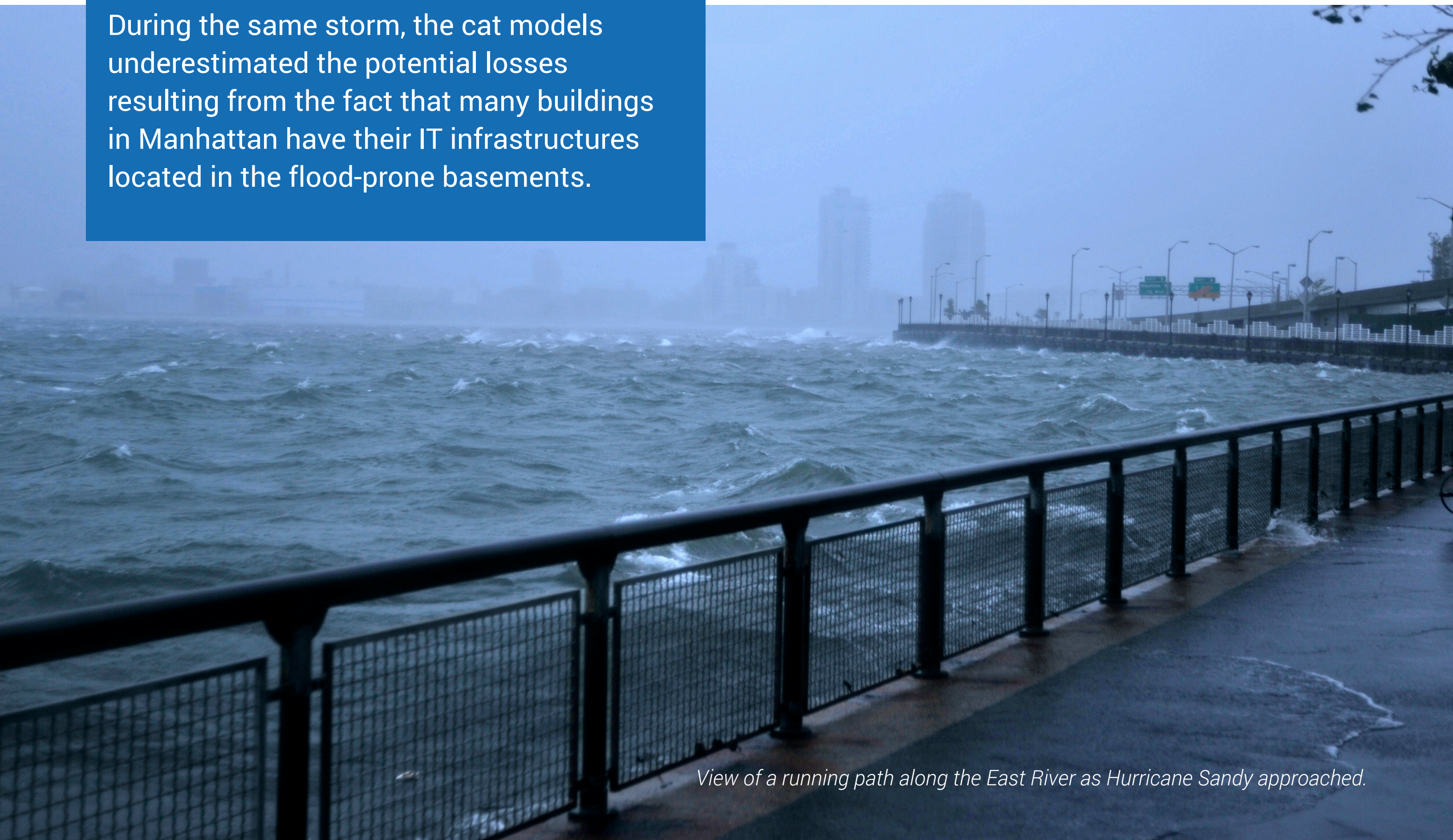
## Example

For super storm Sandy, the cat models expected the ports of New York to suffer flood losses only on the containers on the bottoms of the stacks on their docks. Unfortunately, all the containers were positioned on the docks to avoid the wind blowing the stacks over. Therefore that additional loss was unaccounted for.

During the same storm, the cat models underestimated the potential losses resulting from the fact that many buildings in Manhattan have their IT infrastructures located in the flood-prone basements.

Meanwhile, during the 2014 earthquake in Napa, California, some models did not account for the racks of wine barrels stacked high in the cellars coming tumbling down and shattering, which resulted in Business Interruption claims.

Nonetheless, the cat models perform amazingly well considering the raw complexity of what they do. All cat model vendors employ extremely bright minds to solve staggering statistical problems in a way that does the unthinkable – they predict the future with surprising good success.



*View of a running path along the East River as Hurricane Sandy approached.*

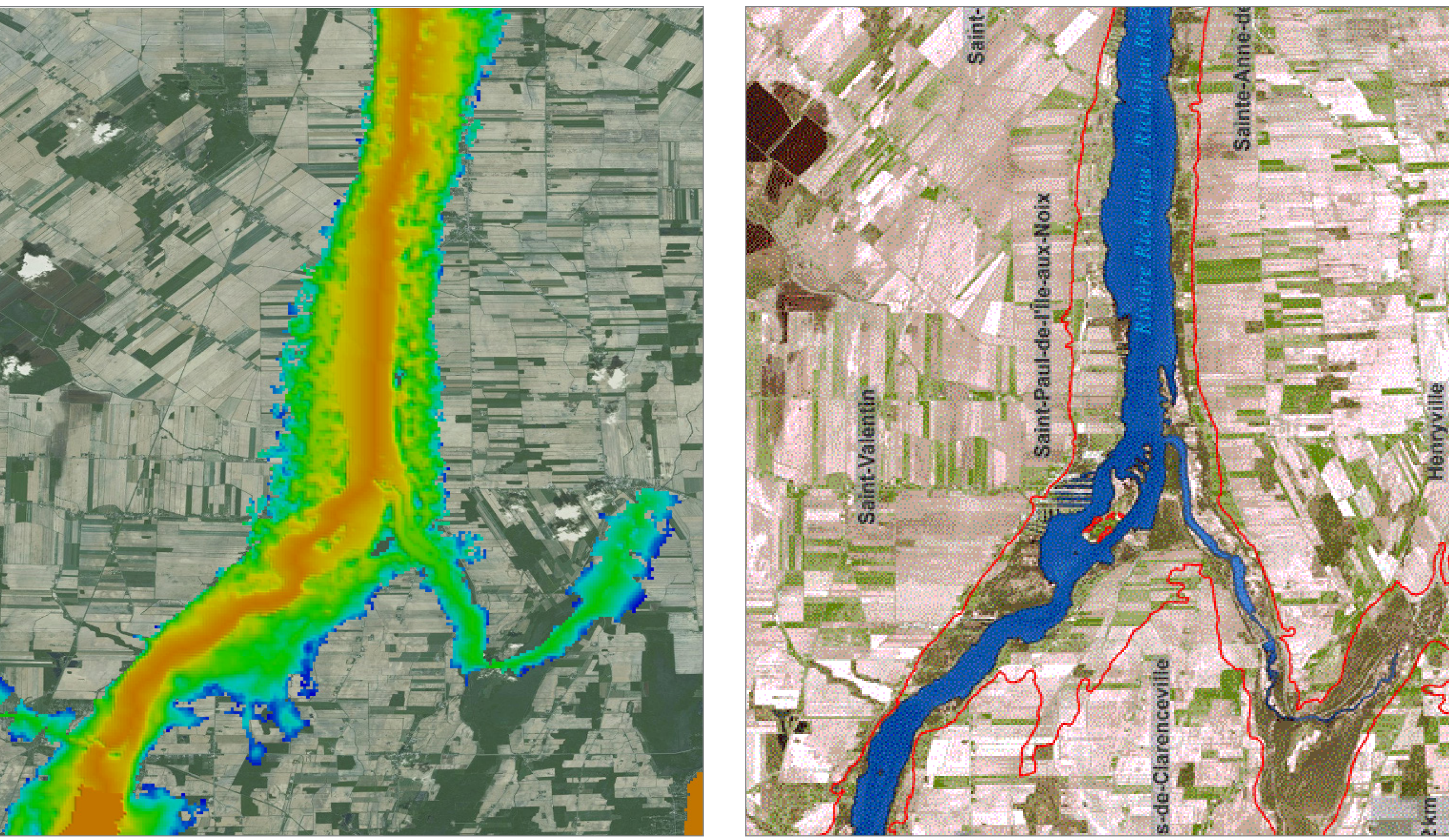
# 3 Historical Claims Data: Beware the Caveat

Lastly, property insurers and reinsurers frequently rely on a historical database of claims. While this information can be very useful,  it's important to remember that this dataset is simultaneously:

- their most expensive dataset;
- significantly overestimated in its value, and;
- terribly underutilized.

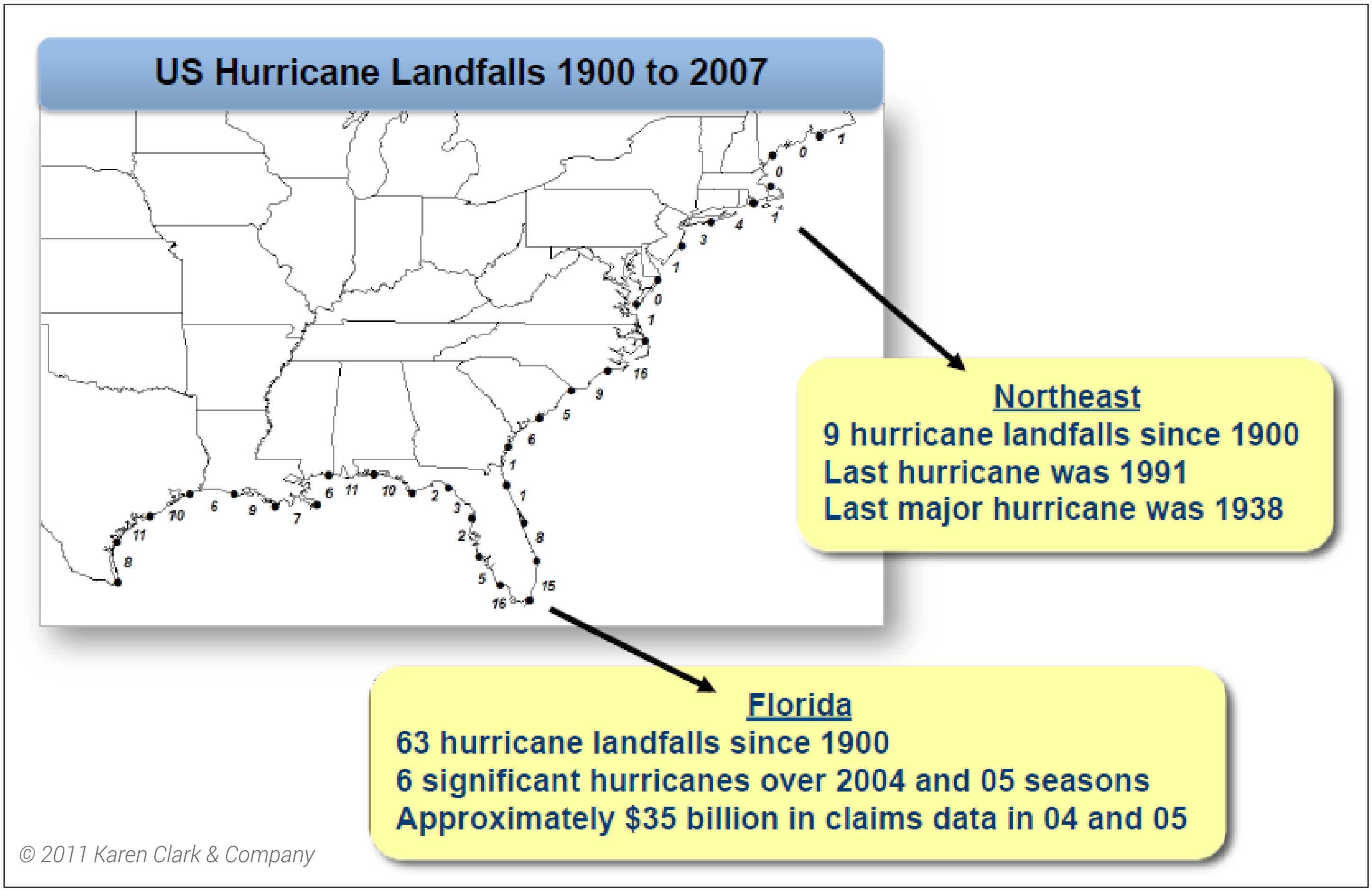
It's worth exploring how a historical database of claims can be all three things at once.

The history of claims paid by an underwriting organization is a fundamental collection of information, and from the very beginning of insurance in those smoky London coffeehouses in the 17th century they have been meticulously kept and guarded. They represent the sum-total of a company's experience. Such information does not come cheaply though – every single claim was paid for in straight cash, including some monumental losses. Such corporate memory must be kept away from competitors because it's frequently hard experience that gives insurers a competitive edge – twice shy is more profitable than twice bitten.




*Richelieu River near Henryville, Quebec, Canada: The left image showcases flood model outputs with a RTP of 100 years. The red polygon in the right image shows the extent of flooding in 2011 (CSA, 2011).*

Claims history, along with full policy history, are the core datasets used for designing underwriting guidelines and business rules. Actuaries spend a lot of time with these archives, to say the least. But in the case of property insurance, and more specifically, natural catastrophe coverage, history is an inadequate guide. The Toronto flooding from 2013 was caused by 90 mm of rain in one day, more than both the previous same-day rainfall record of 29.2 mm in 2008 and the roughly 70 mm monthly average for July. It was a completely unpredictable event based on historical records. Another example is the annual estimation of impending wind losses from Atlantic hurricanes, a wildly difficult prediction to make based on history, as shown by this graphic:




Natural events that wreak damage to property are, by their very nature, not bound by anything that has been witnessed in the past few hundred years, let alone insured by a single carrier in the past century.

However, there is information yet to be gained from these comprehensive catalogs of losses, especially for property losses. New techniques in geospatial analytics can convert a vast database into a map that can be incorporated into a geospatial analytic engine. Leveraging claims data in this way can lead to new insights into how losses can be modeled, by location, in a way that is immediate and intuitive, and can enable better portfolio management through better underwriting and better accumulation.


# Conclusion

Obviously, natural catastrophes are impossible to predict. This is both the *raison d'être* and principle challenge of the insurance industry: If these catastrophes were predictable, nobody would need insurance; but how do you base a business on the unpredictable? Insurers spend a lot of effort on inspection and mitigation, but in the end there is no defense against an act of nature, nor an inspection that will entirely capture all the variables.

Models are an indispensible part of any insurance business, and it is the intelligence of how those models are built and used that differentiates the insurers in the risk markets. Understanding the limitations of models is as important as using the best available models, because once the limitations are understood, the uncertainty can be identified and addressed. In insurance, more than in any other business, uncertainty is expensive, and anything that lowers uncertainty is a money saver.

# About The Author

Ivan Maddox is the InsitePro™ Product Manager at Intermap Technologies® and has over 20 years of experience in the Product Management and Operations fields, specializing in the delivery of SaaS solutions and platforms.

# About Intermap

**WWW.INTERMAP.COM**

Intermap Technologies is a global geospatial solutions leader focused on improving the ways in which people, businesses, and governments use location-based information. Through the creation of the Orion Platform®, the industry's first software-driven spatial data platform, Intermap provides products, services, and solutions that help business and governments solve their geospatial challenges.