Towards socially responsible (re)insurance underwriting practices: readily available 'big data' contributions to optimize catastrophe risk management

Zvezdov, Ivelin

26 September 2016

Towards socially responsible (re)insurance underwriting practices: readily available 'big data' contributions to optimize catastrophe risk management

Ivelin M. Zvezdov, M.Phil.

Senior product manager

AIR Worldwide, VERISK Analytics

izvezdov@air-worldwide.com

Ivelin.zvezdov@gmail.com

Mobile phone # 781.354.8661

Boston, MA

Dr. Sebastian Rath

Principal Insurance Risk Officer, NN Group

Advisor and Visiting Researcher, UNESCO-IHE

sebastian.rath@gmail.com

Mobile phone +31 6261 77668

Amsterdam, The Netherlands

Abstract

Today's advances in big data technologies readily allow for storing large inter-dependent data sets of historical and modeled natural hazard and financial data and unifying their granularity and accuracy with common geo-spatial and risk-type record identifiers. This is a significant component at both single insurance account, and even more so at the larger multi-policy portfolio scale for enabling optimal and socially responsible insurance underwriting practices. This supports insurance risk transfers by creating more accurate and all-uncertainty encompassing pricing techniques, and exposes these techniques and methodologies to all market players, including insurance policy holders via transparent statistical and actuarial principles.

Keywords

Big Data, (Re)Insurance Premium Pricing, Sustainable (Re)Insurance Principles

Introduction

Advances in big data methodologies with high degrees of granularity and transparency have made it possible to enhance the discussion on socially responsible (re)insurance underwriting practices. This article offers a definition of the microeconomic concept of socially responsible (re)insurance policy underwriting. This proposition draws on data components from natural peril and financial data modelling to bring it truly alive. In the view of the authors this proposition enhances transparency in risk metrics definitions, and hence improves the overall decision-making and underwriting process for an insurance policy or a reinsurance contract. Where consistently implemented and used this proposition builds and facilitates the basis for a more socially responsible (re)insurance policy underwriting, primarily but not limited to the context of (re)insurance of risks from natural catastrophes.

As such this paper aims to address various audiences: stakeholders at (re)insurers with responsibilities for fair risk pricing, fairness and customer interests in the underwriting process and specific IT interests and process in catastrophe risk modelling; stakeholders with responsibilities for ethical, social, responsible and fair risk transfers including investors, pension funds and NGOs; and stakeholders with responsibilities to further the use of available big-data at insurance companies, customer fairness and future (re)insurance products.

By addressing this audience this paper aims to contribute to discussions on the fair sharing of climate change risks, the evolving reinsurance market and regulatory requirements on fairness in (re)insurance risk transfers as part of evolving financial market regulations.

Optimal and socially responsible (re)insurance policy underwriting

The definition of such concepts cannot be purely an actuarial or modeling objective. The very idea of social responsibility carries an implied judgment of public good and optimal distribution of resources in the economy. Importantly, this also assumes a broad consensus among economic agents in these matters.

Without delving into the philosophy of public good and how the latter is enhanced by optimal actions from all micro-economic players, the authors propose the use of a simple equilibrium concept, which is sufficiently well recognized, and can be adopted for the purpose of this paper.

This equilibrium is between three groups of micro-economic players: [1] the insured persons or entities and their interests; [2] the insurance policy underwriters and their firm's interest and [3] the shareholders of these (re)insurance companies and their interests. The equilibrium status requires that (re)insurers charge a fair premium for risks, specifically in the instance of extreme catastrophe risks, such that the (re)insurers receive a fair and appropriate price for accepting and managing the financial impact of this risk, on behalf of the policy holders who transferred this risk. The equilibrium ensures that the firm with its employees [2] and its shareholders [3] are in fair balance with the other side of equilibrium, where insurance policy holders [1] originate risk transfers by accepting an insurance premium payment as a preferable economic option over alternatives, namely self-insurance through savings, or foregoing insurance altogether and carrying the (catastrophe) risk itself.

A role for ‘Big Data’

The attribute of ‘big data’ is derived from sheer volumes as well as from the complexity in data layering, both contributing to drive an accelerate speed implied in upgrading and updating of data vintages. Advances in software algorithmic and computational technologies and methods, as well as in hardware engineering components, are making possible the development of large data sets of geo-spatial physical hazard and financial (re)insurance historical and modeled variables and quantitative metrics. In this paper we examine selected data components and structures which together can introduce fair transparency in the policy underwriting process in big-data application and for the objectives of this paper.

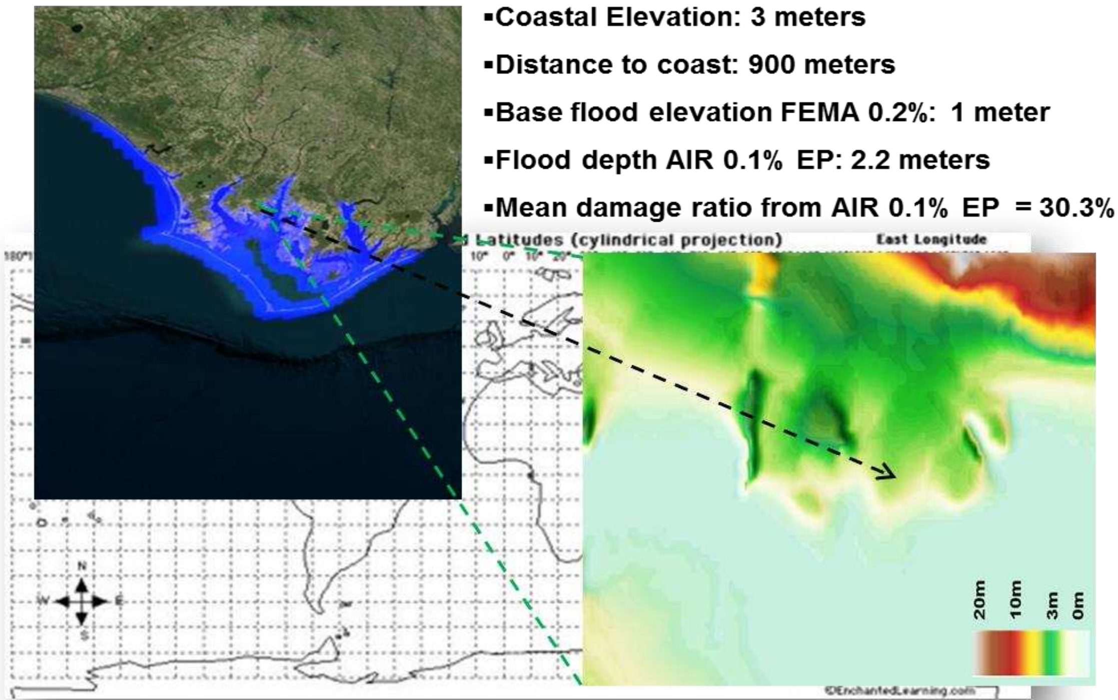
For the selected data components and structures this paper shows that new data-intense risk metrics and techniques have the potential to improve social responsibility in this business decision-making environment as their use enable policy-by-policy level granularity of risk metrics.

The techniques and their use proposed in this paper preserve granularity and support appropriate and fair high-level aggregate results in the process of computing, modeling and storing of sensitive risk-variables. Techniques discussed in this paper cover build-up of layers of risk metrics from various sources, including historical data, differentiating various modelled data sets and vintage for variables sets. For the purpose of (re)insurance data this paper addresses the use of unique geo-spatial records and unique record identifiers.

Data granularity and layering

While data volumes are a prerequisite and they are certainly present in natural hazard historical and modeled variable data-sets for insurance underwriting. The technology implementation of a historical and modeled data unique geo-spatial grid record and identifier, and big data algorithms for updating, storing and reusing such records in multiple parallel analysis runs allow for development of unique single risk geo-spatial hazard and (re)insurance risk metrics.

Fig.1: Unique geo-spatial record & identifier supports geo-physical and modeled hazard data-layers



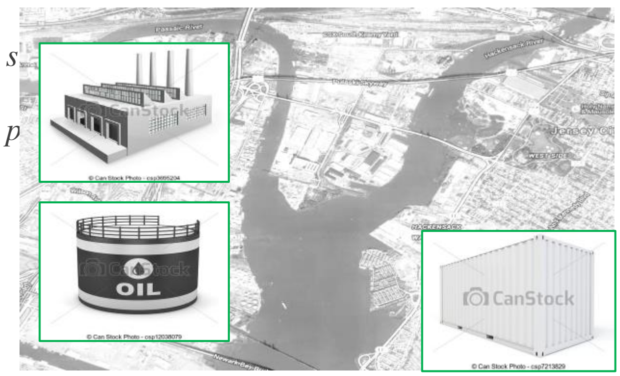
For the scope of this paper further components are distinguished. Typically, those pertain to a unique geo-spatial record. For the purpose of this paper such identifier enables to support multiple dependent data layers. Those dependent data layers, or components, may contain the

- insured risk attributes;
- historical data on exposure with related claims experience;
- data modelled from a natural hazard and financial models;
- resulting decision-making information layers (those may for example relate to risk analysis, risk pricing, risk mitigation or various kinds of strategic planning)

Where such frameworks are permanently maintained for data accumulation with a capacity to maintain these multiple components (with their layers of historical, natural hazard, geo-spatial and financial variables) this contributes to building-up an environment of physical and insurance risk transparency that can drive a fundamental change to the social requirements. This requires accountability for the users, be it local communities involved in risk assessment and mitigation, local, regional of international (re)insurers or any other stakeholder in a fair risk transfer.

Case-Study

Fig.2 Insurance policy for three industrial facilities, located in an area at high-risk for s… p…



This notional insurance underwriting case-study of an industrial facilities policy for three assets, located in geo-spatial proximity, in a highly vulnerable area to storm surge and river flood natural perils. Each unique geo-spatial record is enabled to contain modeled physical attributes of the perils such as distance to coast and coastal elevation, base flood depths and flood elevations. In parallel the same record hosts any existing historical data on previous catastrophe events at this geo-spatial location, and insurance policy claims, originated as a result, as well as known historical premiums. Lastly the same unique geo-record holds insurance loss risk metrics produced by a stochastic financial loss model. Thus structuring, permanently maintaining and using the four defined data layers (insured risk attributes; historical data; modeled data;

	derived uses for decision making) creates a qualitatively advanced risk management and insurance underwriting environment at multiple levels of decision making (e.g. for insurers in corporate underwriting, accumulation control, or risk management for single insured risk, accounts, line of businesses or portfolios).

The emergence of IT model architectures and supporting data that seamlessly facilitate merging of historical data components with actual modeled data environments - at transparent and fast availability - provides another technical prerequisite for socially responsible underwriting principles. Nowadays, those can for example be applied to residential home owner's insurance policies.

In the view of the authors today's engineering and IT technology enable the insurance market can adopt steps towards new, advanced industry-wide practice level. This level would draw on large data sets relevant to underwriting a (re)insurance policy contracts that are principally structured into three functional layers with common geo-spatial granularity and identification. Broadly the authors propose following definition of these four layers:

1. A combined exposure-and-history data layer. This layer stores attributes per unique geo-spatial record relating to the exposure and its physical location, its engineering attributes and vulnerabilities, as well as the insured risk records with their known insurance premiums and

claims history for known historic events informing contract data modelling implicitly. Amongst other user groups, this layer is important for insurance underwriters and claims handlers.

2. A hazard data layer. This layer contains the natural physical properties of the geography and terrain of this particular geo-spatial unique record or administrative unit, as well as stochastically modeled intensity and frequency of a natural catastrophe event. Amongst other user groups, this layer is important for catastrophe risk scientists, modelers, and model users who derive risk transfer solutions and decisions on model-basis.

3. A financial data layer. This layer is being informed by models and draws on the information in the prior two data layers. It contains modeled expected (re)insurance losses and a range of fully probabilistic uncertainty and risk metrics, and where all modeled financial quantities are dependent on the physical natural peril. Amongst other user groups, this layer is relevant as it informs risk carriers with their products and solutions for insured clients and customers.

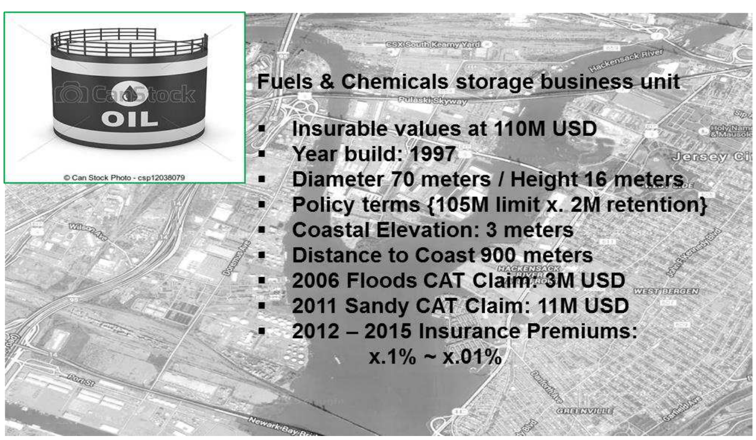
4. The use-and-decision-making data layer. This layer typically interprets data layers 1, 2 and 3. For the intended audience of this paper there may be a wide range of uses. Focusing on (re)insurer and stakeholders in financial risk transfers, the uses will relate in many instances to risk identification, risk pricing, risk selection, underwriting, and portfolio management. Those uses are inherently related to aspects that are strongly linked to fairness in the regulation of financial markets and are to be considered in risk management cycle at various stages. This layer determines key outcomes for all users and stakeholders involved in risk transfers. Amongst others it is relevant for the exposed insured, insurance brokers, modelers and financial risk carriers. Equally, this layer can inform the risk pooling of government initiatives and

intergovernmental organizations and other stakeholders supporting risk relief and mitigation efforts.

Historical, exposure, hazard and financial dependent layers of structured data enable for greater accuracy and flexibility in premium pricing by allowing creating dependencies and mapping functions across different data types and variables. Such an underwriting process ecosystem, underlined by systemic transparency for all data layers and for all unique variables, is a strong prerequisite for promoting socially responsible insurance policy underwriting practices by all involved market players.

Case-Study (continued): *Multiple inter-dependent data layers*

Fig.3 Historical and Exposure data layer is available at each unique geo-spatial data record and bounding box.



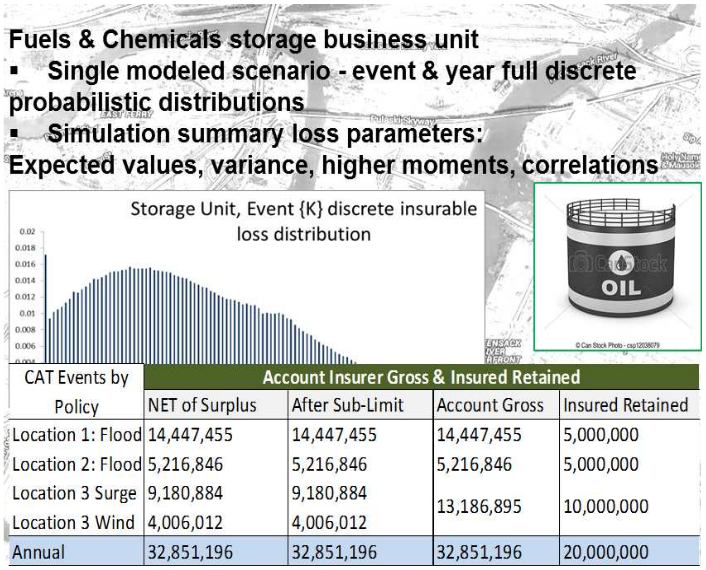
Historical & Exposure data layer contains available engineering information on the insurable asset, such as type of structure and construction materials, height and year build. It also contains the insured risk policy terms and any previous claims data from historical natural catastrophe events. Geo-physical properties of the insurable asset may also be become significant attributes of the historical data layer. This aggregation of data at a

	unique geo-spatial record allows the modeler or (re)insurance policy underwriter to build up statistical dependencies and correlation structures between physical parameters such as elevation and distance to body of water and historical claims data. Such correlation structures are effectively exploited in a geo-spatial surface –like statistical analysis for large sets of insured risks.

Big data software engineering algorithms enable fast writing of modeled data produced by simulation and analytical platforms, and then compressing such data sets to manageable proportions. Furthermore geo-spatial data manipulation and analytics algorithms enable the construction of statistical surface data sets, where statistical dependencies between historical and modeled physical and financial variables could be uncovered accurately and effectively.

Case-Study (continued): *Modeled financial and statistical insurance loss data layer*

Fig.4 Financial data layer contains statistical summary parameters for losses from single event – scenario and full stochastic simulations



CAT Events by Policy	Account Insurer Gross & Insured Retained			
	NET of Surplus	After Sub-Limit	Account Gross	Insured Retained
Location 1: Flood	14,447,455	14,447,455	14,447,455	5,000,000
Location 2: Flood	5,216,846	5,216,846	5,216,846	5,000,000
Location 3 Surge	9,180,884	9,180,884	13,186,895	10,000,000
Location 3 Wind	4,006,012	4,006,012		
Annual	32,851,196	32,851,196	32,851,196	20,000,000

Exact discrete probabilistic distributions from a single scenario, such a 1,000 year simulation, provide modelers and underwriters an opportunity to enhance product structuring and pricing, marginal impact and risk management tasks. As well as introduce transparency and accuracy in interaction within the insurance market place: between insurers, brokers, insureds and regulators. Statistical summary tables of expected value and variance of insured losses, as well as cumulative probability metrics, such as probability of breaching retentions and exhausting limits are direct inputs in (re)insurance policies and treaties pricing formulas and more complex simulation procedures. As with hazard data layers, constructing geo-spatial surfaces of insurance loss parameters presents an opportunity for studying and uncovering dependencies and correlations, which are not immediately evident in historical claims data.

Statistical and actuarial mechanics of socially responsible insurance policy underwriting

In the remainder of this paper the authors review a traditional and conceptual insurance underwriting model. The objective is to illustrate functionalities and capabilities of multi-layered, structured and dependent data in (re)insurance policy underwriting and risk management. The authors show how such data enables both, technical and actuarial transparency. In the view of the authors this contributes to transform the underwriting process from a closed-form-system to a socially-responsible and open ecosystem. The following section sets out where the authors see the potential for the insurance market to evolve form as a fair big-data user.

The authors propose the following conceptual insurance model, consisting of an annual data series $S\ \{1, \dots, k\}$ of (re)insurance occurrence claims X over a historical time period $T = 1, \dots, n$ years.

$$S_T = X_1 + X_2 + \cdots + X_k$$

With a traditional annual aggregate modeled loss distribution function $F(S)$

$$E(S) = E[X_1, \dots, X_k] = \int_{-\infty}^{+\infty} \int_{-\infty}^{s-x_1} \int_{-\infty}^{s-x_1-\cdots-x_{k-1}} f_X(X_1, \dots, X_k) dX_k, \dots, dX_1$$

This cumulative function is integrated over a stochastic simulation scenario of $n * p = N$ scenario years.

This paper represents the fully probabilistic expected annual aggregate insurance loss as

$$E[S_T] = \int_{-\infty}^{+\infty} S f_s(S) ds$$

and the annual standard deviation of insurance loss as traditional integral statistical quantities as

$$\sigma(S_T) = \sqrt{\int_{-\infty}^{+\infty} (S - E[S_T])^2 \, f_s(S) ds}$$

Both basic financial and actuarial measures of expected policy loss and its expected variation can be used to construct the classical insurance premium $P(r)$ formula with a known deterministic underwriting risk loading quantity, represented by a coefficient R as:

$$P(r) = E[S_T] + R * \sigma(S_T)$$

The audience of this paper is deemed to be familiar with a traditional underwriting environment, where these are the minimum required per-insured-risk quantitative measures to define the pure or technical natural catastrophe price.

In the view of the authors advances in big data storage and management techniques now readily support an evolution beyond this point, by using the full probabilistic function $E(S)$ for a single insured risk with modeled stochastic loss to be preserved and queried out of data storage for deriving accurate analytics and pricing metrics.

$$E(S) = \int_{-\infty}^{+\infty} \{F_S^{-1}(S_1), \{F_S^{-1}(S_2), \dots, \{F_S^{-1}(S_T)\} dS_T, \dots, dS_1$$

Where $S_T, \ T = 1, \dots, n$ is the modeled stochastic annual aggregate insured loss, and $F_S^{-1}(S_T)$ is its inverse cumulative densities function. The key benefit of using this full probabilistic single risk loss distribution $E(S)$ explicitly is that the insurance policy underwriter is able to easily compute any Value-at-Risk (VaR) and Tail-Value-at-Risk (TVaR) type of risk metric:

$$TVaR_{\propto}(E_S) = \int E(S)^{-1}(\propto)dP$$

Such tail risk metrics are considered beneficial and are used to enhance the classical underwriting formulas, to include premium pricing dependence on tail risk uncertainty, via an actuarial dependence and mapping function $\theta\{.\}$

$$P'(r) = EV(E_S) + \theta\{\sigma(E_S), \dots, TVaR_{\propto}(E_S)\}$$

More flexible pricing functions may include a more accurate risk loading factor R' for expected standard deviation of loss and dependence on tail risk via the actuarially developed, mapping function $\theta\{.\}$

$$P'(r) = EV(E_S) + R' * \sigma(E_S) + \theta\{\, TVaR_{\propto}(E_S)\}$$

The authors point out that a big insurance data user can readily utilize such capabilities today, for example with the most obvious candidate of integrating full probabilistic single-risk loss distributions in (re)insurance premium pricing. Analysis results for what has been defined as 3'rd inter-dependent data layer in this paper can typically be scaled-up as required, e.g. by a significant stochastic simulation such as of 100K scenario years; and by the size of an insurance account, line-of-business unit or an entire portfolio. Today, big-data storage, organization and manipulation tasks for the 4'th data layer which interprets big-data into decision-making metrics are quickly becoming an ever more engaging technological proposition.

This paper intends to address and include audiences that are not directly from an insurance background, and may therefore not be directly or deeply involved in today's IT capabilities of risk carriers, such as (re)insurers. Therefore it is worth noting that advances in technologies with introduction of massively parallel process (MPP) database platforms support high data-compression rates and node ‘cloning’ capabilities, enable the maintenance of the proposed four data layers. Therefore the authors consider that the proposed modelling and data-management tasks are quickly becoming ever more active enterprise goals.

Requirements for mapping across dependent data layers in addressing global climate change risks

The author's propose that making available structured and inter-dependent data layers in order to support fair and transparent insurance policy underwriting is an optimal (re)insurance industry micro-economic strategy. Such transparent underwriting ecosystem allows for all market players - insureds, brokers, and insurers in parallel to explore linkages between historical quantities of claims and hazard intensities, modeled hazard variables, and insurance loss and risk outcomes. Such underwriting practices are aligned with credible assessments of global challenges that were mutually acknowledged at COP21, the Paris based Climate Change Conference in December 2015, requiring transparent and fair tools for risk analytics in the global response to climate change risk mitigation and planning.

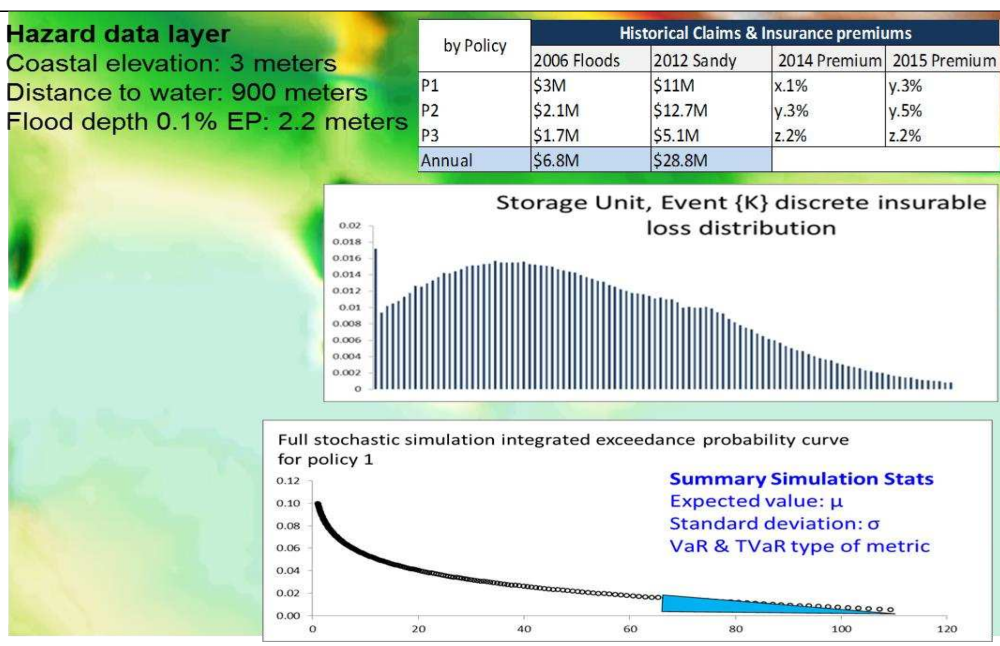
In this context there are three significant technological prerequisites, needed to develop meaningful inter-data layer operational capabilities. These are:

(1.) Common, unique, geo-spatial, identification record for variables and metrics across data layers, which have sufficient historical and modeled geo-spatial proximity;

(2.) geo-spatial grid scaling and transition algorithms – as we do not have all historical and modeled data realistically placed at a unified geo-spatial grid, it is necessary for the modelers to be able to quickly transition and rescale from various modeled and historical grid systems.

(3.) mapping, dependency and correlation mathematical functions, which allow sophisticated modelers and market practitioners to research and detail the linkages between data layers, and then utilize these in designing more sustainable (re)insurance pricing formulations and propositions.

Case-Study (continued): *Building of mapping and dependency functions across data layers*



Historical and Modeled data is of varying vintages and placed on varying geo-spatial grid systems. The first task of a modeler is to develop rescaling and interpolation algorithms such that analytics tools can move across data layers seamlessly and effectively by scaling up and down grid systems to equalize them at a common geospatial unit. Once this is accomplished statistical mapping functions across variable can be developed and integrated in (re)insurance pricing systems.

Loss and risk metrics from the financial data layer typically provide expected value of simulated loss, standard deviation of expected loss, and a VaR or tail VaR risk metric from the full simulated loss distribution. Dependencies across insured risks are also measured with correlation matrices.

$$P'(r) = EV(E_S) + R' * \sigma(E_S) + \theta\{\, TVaR_{\propto}(E_S)\}$$

Where the risk loading factor R' and the tail risk metric mapping function θ are derived and used to inform the underwriting process of the risk transfer. This expansion of pricing methodology in itself has an optimality and sustainability effect for both the (re)insurer and the (re)insured by capturing a more thorough view of risk and uncertainty.

When exploiting data layers 1, 2 and 3 as proposed in this paper, the 4'th data layer constructs summary data metrics for decision making. Focusing on transparent and fair underwriting processes, this paper explains foremost the most common underwriting approach.

The simplest and most intuitive underwriting function is established by exploiting the available linkages between historical, hazard and financial data layers while similarly accepting certain bounding intervals in which the established technical catastrophe insurance premium is considered as valid and statistically-technically fair. In this case an average historical claim and other historical statistics for this geo-spatial unit can be obtained from the historical data layer. Again, as previously introduced, the annual data series $S\,\{1, \dots, k\}$ of (re)insurance occurrence claims X over a time period $T = 1, \dots, n$ years is used.

$$S_T = X_1 + X_2 + \cdots + X_k$$

Using variables from the historical data layer permits to derive the classic actuarial historical statistics, comprising of the average claim $\mu(H)$, the standard deviation of historical claims $\sigma(H)$, and the largest historical claim $Max\{S_T\}$.

$$\mu(H) = \frac{1}{N}\sum_{i=1}^{N} S(T)_i$$

$$\sigma(H) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S.T_i - EV.AAL)^2}$$

$$Max\{S_T\} = \max\{S_1, \dots, S_T \mid T = 1, \dots, n\ years\}$$

The use of such information is common practice when engaging in financial risk transfers, challenging the policy seller understanding from a scientific, engineering or funding perspective, or when addressing future risk mitigation strategies. Where risk transfer via (re)insurance is concerned, users typically build bounding acceptability intervals. Those bounds draw on a mixture of historical and simulated statistical quantities, to explore the concept of historical and stochastic sustainability of the technical catastrophe premium $\boldsymbol{P}'(\boldsymbol{r})$

$$\mu(H) + \sigma(H) \leq \boldsymbol{P}'(\boldsymbol{r}) \leq VaR_{\propto=0.004} \leq Max\{S_T\} \leq VaR_{\propto=0.001} \leq Max\{S_T\} + \sigma(H)$$
$$\leq TVaR_{\propto=0.0001}$$

Of course such intervals are designed by each insurer on its pricing preferences, risk tolerance, market and client conditions. At this stage the objective of this paper is to remind reader of such basic principles for defining sustainable underwriting practices and demonstrate how big data principles and capabilities have made this possible.

Conclusions and work ahead

Today's advances in big data technologies readily allow for storing large inter-dependent data sets of historical and modeled natural hazard and financial data and unifying their granularity and accuracy with common geo-spatial and risk-type record identifiers. This is a significant component at both single insurance account, and even more so at the larger multi-policy portfolio scale for enabling optimal and socially responsible insurance underwriting practices. This supports insurance risk transfers by creating more accurate and all-uncertainty encompassing pricing techniques, and exposes these techniques and methodologies to all market players, including insurance policy holders via transparent statistical and actuarial principles. This paper introduced such modeling methods and principles conceptually via a case study, using historical and modeled hazard and financial data layers, which contain all of the individual risk factors entering a premium pricing equation, and informing a further decision making data layer. The authors advocate that the technological advances of 'big data' combined with a regime of transparency are creating the prerequisites for a sustainable and responsible insurance underwriting process, which itself should enhance credibility and trust among all stakeholders bounded in the (re)insurance market-place. The insurance risk underwriting and transfer market players have an incentive to create and pursue such sustainable and socially responsible practices, particularly those that reinforce their credibility as systemically stable institutions, demonstrating thorough understanding of risk profiles and skill in fairly managing insured assets. A regime of transparency and sustainability services the requirements of insureds and regulators acting on their behalf, as far as it guarantees close to optimal premium prices quoted and placed in stable market conditions without introducing unfair or adverse selection. Big data capabilities have the potential of leading and supporting a new level of utility in awareness analytics. Those may serve to detect significant gaps between insurance coverage, physical and financial asset values, as well as human preparedness and resources at risk in vulnerable geo-spatial areas with their specific supply-chain logistics. Continuously and consistently servicing the data requires a fair deal of standardization and transparency for efficiency reasons, while data security and data use requires appropriate levels of controls on their accessibility and usability. This facilitates to address vulnerabilities originating from exposure to natural perils and climate change risks, including financial liabilities and contingent business interruption risks that arise in such circumstances. Further developing integrated methodologies and models is a much needed and promising task for future research in order to map and correlate risk factors across physical, financial and demographic data layers.

References

Challa Aditya, Kolokoltsov, Vassili (2012), Insurance models and risk-function premium principle

Darkiewicz, Grzegorz Griselda, Deelstra Griselda (2007), Bounds for Right Tails of Deterministic and Stochastic Sums of Random Variables

Hürlimann, Werner (2005), On a Robust Parameter-Free Pricing Principle: Fair Value and Risk Adjusted Premium

Laeven Roger J.A., Goovaerts Marc J. (2012), Premium Calculation and Insurance Pricing