



Evaluating a Model's Appropriateness for Climate & Disaster Risk Finance Decision Making

ABOUT THE GRMA

Initiated by the Insurance Development Forum (IDF) and the Climate Vulnerable Forum and V20 Finance Ministers (CVF-V20), the Global Risk Modelling Alliance (GRMA) is a unique public-private partnership that provides countries with open-source models and data, as well as capacity development in risk analytics. The GRMA is hosted by the InsuResilience Solutions Fund (ISF), which is managed by Frankfurt School of Finance & Management, and is funded by KfW on behalf of the German Federal Ministry for Economic Cooperation and Development (BMZ).

This publication was prepared by Alastair Norris, a Climate & Disaster Risk Consultant with the Insurance Development Forum (IDF) as part of the Global Risk Modelling Alliance (GRMA). The author and GRMA would like to acknowledge Aon and Chaucer for their valuable contributions to the technical content of this note.

TABLE OF CONTENTS

Introduction	4
A framework for assessing model appropriateness	5
Step-by-step evaluation	7
1. Qualitative	
2. Exposure evaluation (optional)	
3. Event set characteristics (optional)	
4. Hazard parameter evaluation	
5. Damage function evaluation	
6. Correlation (optional)	
7. Historical loss - single-event analysis	
8. Historical loss - return period analysis	
9. Future climate risk evaluation (optional)	
Summary	16
Annex: Key questions to ask before trusting a model	17

ACRONYMS

AAL	Average Annual Loss
CDRF	Climate Disaster Risk Financing
EMDE	Emerging Market and Developing Economy
EP	Exceedance Probability
GCM	Global Climate Model
GED4GEM	Global Exposure Database for the Global Earthquake Model
GRMA	Global Risk Modelling Alliance
IPCC	Intergovernmental Panel on Climate Change
RCM	Regional Climate Model
SSP	Shared Socioeconomic Pathway
YLT	Year Loss Table

INTRODUCTION

Government officials tasked with ensuring fiscal resilience in the face of climate and disaster risk regularly make trade-offs: Protecting public finances from escalating disaster losses while ensuring that vulnerable populations receive adequate support before and after disasters occur.

Decisions on how to allocate resources are challenging given significant uncertainties. The insurance industry has long used catastrophe risk models to manage such uncertainties: these models quantify the potential impact of extreme events such as floods, tropical cyclones and earthquakes and thereby enable the pricing and prioritising of such risk. Governments are increasingly using these models to assess their own exposure to extreme events and inform financial decision-making.

Effectively applying climate and disaster risk models requires an understanding of how to interrogate the data and methodologies of individual models, and compare between models, in order to know which models are appropriate for a particular government objective. Government financial decisions based on inaccurate risk estimates can result in inefficient resource allocation or fiscal shocks following disasters, undermining the ability of governments to protect vulnerable populations and support timely recovery. Yet in contrast to insurance companies, government departments usually have limited access to the knowledge and resources to fully evaluate models, meaning that models used for public sector purposes are often not adequately evaluated.

This note outlines a framework to guide government officials in evaluating whether a particular model is fit-for purpose. The framework is designed to assess the adequacy of any climate or disaster model in support of disaster-risk financing decisions. It can also be used as a reference for governments to develop institutional capability in model evaluation, whether internally or through commissioning external expert support. The framework has been written for policy makers to understand what evaluation is required of a model before it can be trusted to inform decision making. The more detailed Annex is for technical experts or consultants more directly involved in the development and application of catastrophe risk models.

A FRAMEWORK FOR ASSESSING MODEL APPROPRIATENESS

Catastrophe risk models combine natural science, engineering, statistics and data to quantify the risk faced from natural hazards. Importantly their purpose is to extrapolate beyond the historical or observed record in a way that is considered realistic, whilst accounting for any observed trends in the underlying hazard such as climate change. These models can inform the design and selection of investments that mitigate or reduce risk (disaster risk reduction) and inform the design of financial instruments to finance the cost of responding to or recovering from these events (Climate & Disaster Risk Financing (CDRF)).

Model evaluation is an essential step for those that rely on the model's outputs to confirm that it has been robustly designed and developed, and provides a reliable representation of the risk, including credible loss estimates. Developing a risk model requires many assumptions on both the underlying science, and how it can be represented and communicated in a simplified way to support a specific decision. Many of these assumptions are subjective, made by the modeller or model supplier, and understanding them is crucial when selecting and using a model to inform public sector decision making.

Determining the appropriateness of a model requires the assessment of:

1. Whether the model captures the fundamentals of the climate system being modelled (e.g. how tropical cyclone windspeeds decay after landfall)
2. The model's ability to accurately assess the impact on exposure (e.g. for historical events, a comparison between modeled loss and recorded on-the-ground experience)

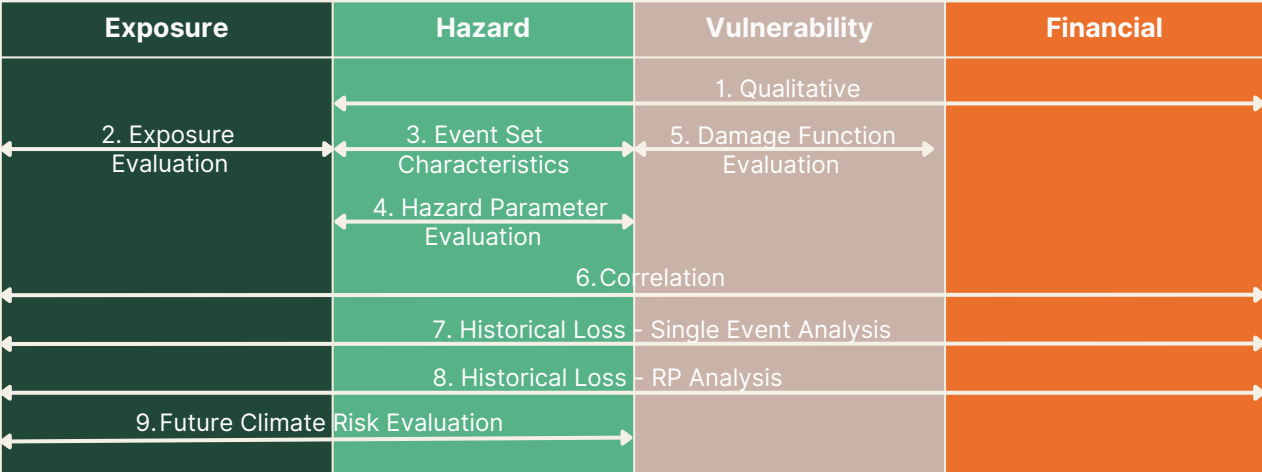
Figure 1 visualises a model evaluation framework and outlines the essential steps for fulfilling these two requirements. The figure highlights the four key components of a probabilistic catastrophe risk model (exposure, hazard, vulnerability and loss/financial) and allocates each step of the model evaluation to one or more of these components. The framework is specific to evaluating the appropriateness of a model for CDRF decision-making but can be adapted to apply to other risk-related use cases, for example Disaster Risk Management¹.

Before applying this framework, it is critical for there to be a full appreciation of the objective of the assessment as well as a clear description of the "risk question" that needs answering. Doing so directly informs the required depth of any assessment as well as any optional steps that need to be included. For example, is it the government's aim to evaluate:

1. whether an existing model or a model offered by a supplier is able to support the design of an insurance solution (e.g. does the Global Risk Modelling Alliance (GRMA) tropical cyclone model developed for Madagascar enable pricing of indemnity instruments in country), or
2. which one of multiple available models should be selected to inform policy decisions, such as the design of a risk financing strategy or for fiscal planning (e.g. which of the many commercially available earthquake and tropical cyclone models for the Philippines should be used to provide loss numbers for the risk financing strategy).

[1] If the focus were Disaster Risk Management rather than CDRF, the emphasis would shift toward understanding the spatial patterns of hazard and exposure (i.e. how the severity of hazard and exposure changes geographically) with more focus on scenario level estimates (e.g. a historical event) and less need for full probabilistic modelling. Within this evaluation framework it would likely remove the need for 3a, 3c, 4, 6b, 8.

Figure 1: Model evaluation framework (adapted from Aon)



STEP BY STEP EVALUATION

Catastrophe risk models generally consist of four components, exposure, hazard, vulnerability and loss. Together, these components provide an assessment of how likely an event is to occur, what exposure is at risk, the impact of events on the exposure (this could be quantified in terms of people or assets) and the financial implications of these impacts.

Any effective model evaluation depends on transparency. Model suppliers should provide comprehensive documentation that clearly describes the model methodology, data sources and inputs, key assumptions underpinning the model, and any known limitations that may affect its application or interpretation. In addition, all model evaluation tests performed by the provider should be made available along with clear explanation of why these tests were done and, importantly, what parts of the model could not be validated due to data limitations. This level of transparency is more achievable when models are open-source and developed based on open licences as the information is publicly available. The GRMA strongly supports open-modelling efforts and requires that all models and data developed as part of the programme are made openly available.

Steps 2 to 6 in the framework support component-level evaluation which compares specific components of the model to historical observations, whereas steps 7, 8 and 9 provide overall model loss evaluation which compares the model loss output with recorded historical experience. The steps marked as optional are only relevant if they are included as part of the model (e.g. exposure data is being provided by the model supplier) and if the government staff has the required analytical skill to evaluate that component.

1. Qualitative

Purpose

Confirm the intended purpose of the model, including what perils and causes of loss are covered, as well as ensuring the model runs as expected and there is sufficient data to support model evaluation.

1a. Check that there is sufficient access to the model, either directly or for a third party.

1b. Confirm which sources of loss are and are not covered by the model. This includes an assessment of which perils and sub-perils are covered by the model (e.g. for tropical cyclone is it solely wind or also storm surge and tropical cyclone induced precipitation), how losses are reported (e.g. are they reported separately for each sub-peril or combined), as well as what loss types are included (e.g. direct and indirect losses, and for the latter, which impact paths?).

1c. Check that the model runs without technical issues and as expected, producing output that includes all required risk metrics for the intended purpose (e.g. Typical key ones for risk finance include Average Annual Loss (AAL), Return Period losses, standard deviation (a measure of uncertainty), a Period Loss Table. Other use cases may require different metrics).

1d. Identify what data exists with which to evaluate component level output (e.g. for windspeed or rainfall this could be weather station data or satellite data) and model loss estimates (e.g. historical loss data from all available sources). In many parts of the world there is a significant lack of detailed up-to-date data to support a model evaluation, and judgement is required to balance the need for a robust, statistically worthwhile data set against biases in data collection, instrumentation differences, and whether old data represents current conditions. The availability of data can greatly impact the quality of an evaluation and may require data back-filling: researching alternative sources of data with which to back-fill existing gaps. Details on the amount of data required for an evaluation are included in section 1 of the Annex.

Expect: A clear understanding of what is being evaluated, against which sources of data as well as confirmation that sufficient model access exists for evaluation to continue.

2. Exposure evaluation (optional)

Purpose

Where exposure data is provided as part of the model, it should be evaluated to ensure completeness, accuracy and alignment with the evaluating organisation's objectives.

2a. Check that all asset types of interest are included in the exposure data to ensure that the model output provides a complete representation of risk.

2b. Assess the exposure data for completeness and accuracy. Key data attributes need to be captured (e.g. geographic location, construction type, occupancy type, number of storeys and value) and accurately reflect on-the-ground reality. This should include an assessment of the geographic coordinates of assets to ensure they are geocoded to the highest resolution possible and that all data is located within the modelled area (e.g. not in the sea if assessing a land-based flood model). Sensitivity testing of losses (sections 7 and 8) due to location uncertainty (e.g. accuracy of location information and respective geocoding) should be performed for high resolution perils (e.g. flood).

Expect: An understanding of the completeness and accuracy of the exposure data through a comparison with other sources of exposure data available either globally (e.g. GED4GEM, Overturn, Satellite data) or locally (e.g. government census data).

3. Event set² characteristics (optional)

Purpose

Directly assess the event set to ensure that the hazard has been appropriately modelled, aligns with scientific consensus, uses up-to-date datasets that include the presence of recent key events and can support the required outputs.

3a. Ensure the event set contains a sufficient number of independent events. The event set must contain enough events to support the required use cases (e.g. for small portfolios there needs to be a sufficiently large number of events to fully represent the risk³) and consist of independent events developed using an accepted methodology. For example, the event set should not consist solely of historical events with slightly perturbed tracks, but at a minimum also varies other characteristics (e.g. for tropical cyclone – perturbing the genesis location, windspeed, radius to maximum wind speed). A hybrid approach that combines statistical generation with simplified physical modelling would be the preferred methodology.

3b. Assess the appropriateness of the hazard metric as well as its relevance to the loss being modelled. For instance, for flood, it is possible to model flood velocity as well as depth. However, whether the former is a suitable metric depends on the feasibility to evaluate it (i.e. does data exist to evaluate the parameter) and whether there is a well understood and scientifically defensible link with damage.

Expect: An understanding for how the event set has been developed, along with its strengths and limitations and an assessment of its validity to provide the required loss output.

[2] An event set is a list of all plausible events that could occur of a particular hazard along with their frequency, severity and location or path.

[3] For smaller portfolios, a larger number of simulated events may be required to adequately capture the full range of potential losses. This reflects the Law of Large Numbers, whereby larger portfolios benefit from greater averaging across exposures, while smaller portfolios are more sensitive to event variability and the possibility that asset locations are not impacted by a limited event set.

Model Evaluation, Validation and Calibration

Evaluation, validation and calibration are sometimes used interchangeably which creates confusion for users and results in incorrect assumptions. These three terms are distinct in their meaning:

Evaluation – comparing the model against observed data that was not used as part of the development process to confirm it provides a reasonable reflection of reality.

Validation – testing of the model to ensure it accurately represents reality by checking the model against external data that was not used as part of the development process.

Calibration – this is the ‘fine tuning’ of key model parameters (e.g. windspeed) usually in accordance with observed data as part of the model development process. It can apply to hazard or vulnerability components as well as at the overall model level.

The key difference between model evaluation and model validation is that the former assesses the quality of the model to form an opinion on its appropriateness. This often involves subjective judgements using uncertain and incomplete data sets. Model validation is more objective and confirms the accuracy of the model, something which is difficult to achieve given the paucity of recorded data available.

4. Hazard parameter evaluation

Purpose

Ensure that modelled hazard intensity conforms to historical experience by comparing the interim hazard parameters (e.g. earthquake shaking intensity) with recorded observation data.

4a. Compare interim hazard parameters with historical observation data. Observation data can be obtained from global sources (e.g. global reanalysis data or aerial imagery), local instrumentation (e.g. in-country gauge or weather station data) or more qualitative sources (e.g. relief web, disaster loss databases, social media, etc.), although consideration should be made as to the hierarchy of data. Global reanalysis data will have the longest record but is likely to be less accurate locally and have a lower resolution. Ideally a minimum of 25 years of data would exist to support this analysis, to enable comparison of both frequent and less frequent events, but this is often not available in emerging markets and developing economies (EMDEs) and so techniques to blend different data sets and extend what observation data does exist might need to be used (e.g. extreme value distributions). Importantly, more data is being made available with each passing year, especially due to recent advancements in satellite observation.

4b. Compare hazard maps associated with the model against other available sources. An intermediate output of a catastrophe risk model is a hazard map developed using the interim hazard parameters often for key return period occurrences (e.g. 1-in-100-year hazard). The hazard maps produced from the model should be compared against other available sources of hazard (e.g. scientific datasets, peer-reviewed studies, government approved maps and alternative models) to provide additional validation of the spatial hazard distribution of the model and highlight any inconsistencies. Where material differences are identified, the drivers should be understood and documented, recognising that alternative sources may themselves differ due to methodology, data vintage or resolution, and so should be treated as benchmarks rather than ground truth.

Expect: Analytical outputs (e.g. Quantile-Quantile plots) that compare modelled data with historical observation data, both for specific events (i.e. well-known major events) as well as aggregated historical data, together with a comparison of the model’s hazard maps against alternative available sources. This comparison should include narrative around the fit of the modelled data to the recorded data, the consistency of the spatial hazard distribution with other available sources, and any limitations given the data available, such as what can’t be compared as recorded data does not exist to do so.

Catastrophe Risk Model or Climate Model

A catastrophe risk model combines science, engineering, statistics and data to quantify the potential impact that disasters such as floods, tropical cyclones or droughts have on a region. They provide an assessment of how likely an event is to occur and its potential impacts, which could be quantified in terms of people or assets (e.g. buildings) affected, or their financial implications (e.g. monetary loss).

In contrast, a Global Climate Model (GCM) or Regional Climate Model (RCM) is a mathematical model that aims to simulate the Earth's climate system and its respective connections. GCM's and RCMs are similar in design with RCMs having a higher resolution and therefore able to represent more local factors (i.e. to some extent tropical cyclones). These models are used to understand the Earth's climate system and how variables such as temperature and precipitation might change in the future in response to human activities.

GCMs/RCMs are crucial in helping to understand the potential long-term impacts of climate change and the outputs from these models can act as inputs to catastrophe risk models for determining future risk from natural hazards (see step 9).

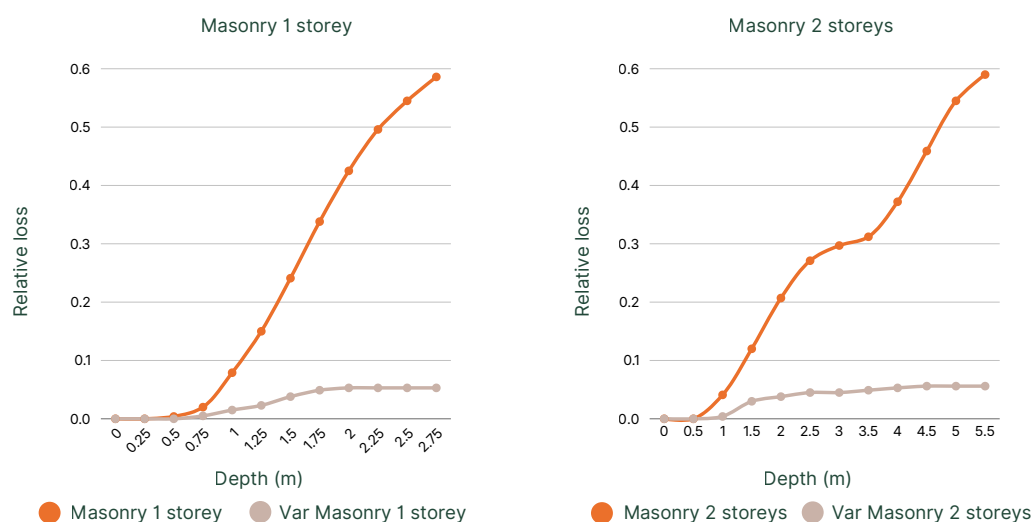
5. Damage function evaluation

Purpose

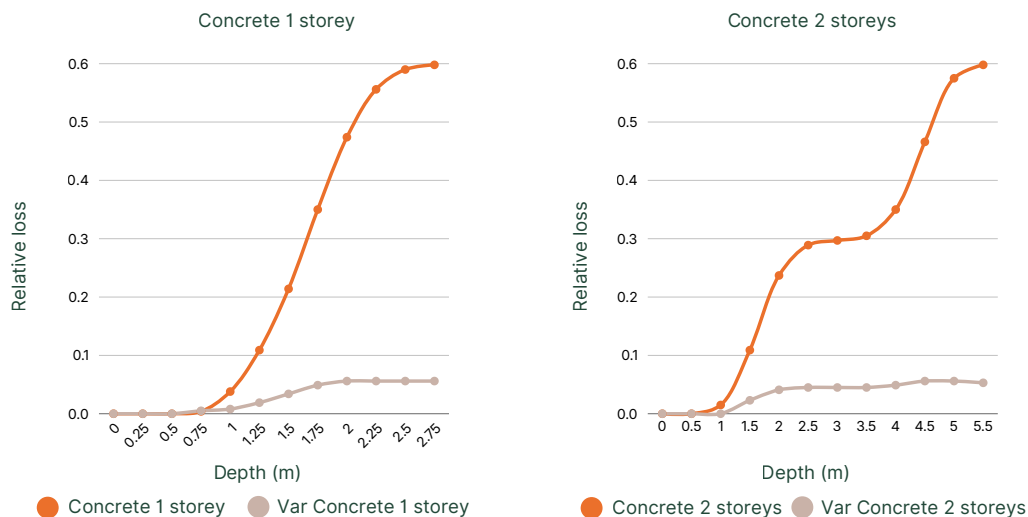
Assess whether vulnerability (damage) relationships are appropriate in the local context and that they are not overly sensitive to key modifiers.

5a. Understand the approach used to develop damage functions⁴ used in the model. Damage functions (Figure 2) can be developed from either an engineering-based approach (i.e. theoretical damage to building components) or historical loss/insurance claims data (e.g. post-event experience for different building types). Either approach is acceptable but should be clearly documented in the model methodology. In many EMDEs, detailed post-event data does not exist to develop damage functions using the claims data approach. Engineering-based approaches are more often relied upon.

Figure 2: Flood damage functions taken from GRMA Costa Rica report "Risk Assessment and Capacity Development in Costa Rica" (Ingeniar, July 2025)



[4] Damage functions translate hazard (e.g. water depth) into a damage ratio, or damage percentage, which in turn can be applied to either the insured value or number of people to obtain a loss estimate.



5b. Assess the appropriateness of damage functions for the modelled region and asset types.

Damage functions must have relevance to the local context and local building practices. This can be done either by developing new damage curves relevant to the local building stock and construction practices or by adjusting existing damage curves to make them align better with local level experience or practices. The relativities across different curves (e.g. construction types, geographic regions, etc.) should also be evaluated to ensure they are as envisaged.

5c. Evaluate damage ratio uncertainty and sensitivity. Damage curves should be evaluated by comparing them either to curves available in countries with similar building practices or to alternative curves developed by others (e.g. independent studies) using the same or an alternative approach. Understanding the sensitivity of results to key modifiers (e.g. construction type) is also required. Documentation should be available on what sensitivity testing has been carried out and the associated results (i.e. varying individual inputs to the model and observing associated changes in the outputs).

5d. Assess how secondary uncertainty correlation is represented in the model by examining the relationship between damage ratios across identical buildings in different locations and their consistency with observed loss data. This evaluation considers whether damage at one location is appropriately correlated with damage at nearby locations, ranging from negative correlation (where losses are independent across nearby sites) to positive correlation (where losses in close proximity have a positive relationship).

Expect: An understanding for how damage functions have been developed to account for the local context, with documentation evaluating the chosen damage functions against other sources of data and testing their sensitivity to key data inputs. Note that evaluation of secondary uncertainty correlation would usually be part of an advanced evaluation and should therefore be considered as optional.

6. Correlation (optional)

Purpose

Ensure that peril correlation, both spatially and temporally, is accounted for as this is critical to assessing tail risk (low frequency-high severity losses) and ensuring diversification benefits are accurately modelled.

6a. Assess how hazard spatial correlation is represented in the model by examining the consistency of hazard intensities across different locations within the same event and their alignment with observed hazard footprints and loss data. This evaluation considers whether the model appropriately captures the spatial dependence of hazard intensity, ranging from low correlation (where hazard impacts vary independently across nearby sites e.g. rainfall events where one area sees large amounts of rainfall when a nearby area does not) to positive correlation (where locations in close proximity experience similar hazard levels, such as coastal towns in close proximity experiencing similar wind speeds in a tropical cyclone event).

6b. Assess how temporal correlation is represented in the model. If the model includes correlation in the occurrence of events (e.g. monsoonal rainfall events cluster in time due to climate conditions being more positive for their formation) then this should be evaluated against historical data to ensure that the probability of events is being correctly represented. Temporal clustering increases annual aggregate losses.

Expect: An understanding of how spatial and temporal correlation have been applied within the model (often statistically) and confirmation that these can be evidenced from observational data, although for temporal correlation it is often difficult to accurately attribute losses to a single event. There should also be documentation explaining to what extent these correlations drive losses at the various parts of the return period curve.

7. Historical loss – single-event analysis

Purpose

Ensure that single-event modelled loss estimates align with on-the-ground impacts and can be relied upon for financial decision-making. This step complements the component level validation efforts and ensures that the model is not only scientifically defensible, but also provides robust loss estimates.

7a. Assess model loss estimates for deterministic events against on-the-ground experience (client-agnostic). Model suppliers will focus on assessing how well the model aligns with general historical experience in the region being modelled, typically using a broad economic exposure database rather than a client-specific portfolio. It is important to recognise that any such assessment must account for differences between the modelled scenario (i.e. using current exposure data) and the historical loss data recorded at the time of the event, as well as any differences in the types of losses covered by the model. In particular, exposure data must be normalised over time to ensure a like-for-like comparison, adjusting for changes such as inflation and other factors that affect the value and composition of the building stock between the historical event date and the present-day modelled exposure. This step should be run for key historical events and include a range of event severity (i.e. not only the most severe).

7b. Assess model loss estimates for deterministic events against on-the-ground experience (client-specific). Where possible, comparisons between modelled losses and historical losses should be conducted not only by the model supplier using broadly available datasets (often aggregated at a national level), but also using client-specific exposure information and, where available, the client's own historical loss experience. This is particularly important for government clients, where their risk may not be well represented by general picture. This step mirrors Step 7a, but is applied using client-specific exposure data and corresponding loss estimates.

7c. Assess model loss estimates for deterministic events against alternative analytical estimates. Where alternative modelled loss estimates are available for the region, whether publicly accessible or provided by the client, these should be used to benchmark results for individual events, with clear explanations provided for any observed differences. Comparative model evaluation is useful for understanding structural uncertainty and identifying the drivers of differences in hazard, vulnerability and loss estimates. The ability to interpret such differences may be constrained by the level of access to underlying model methodologies.

Expect: Comparisons between single-event modelled loss estimates and actual loss experience for multiple historical events with a range of different severities using both generic and client specific data. Included alongside these comparisons should be a narrative explaining potential sources of difference and any limitations of the analysis. If a model performs well on step 7a but not on step 7b then a comparison needs to be made between the two portfolios (client portfolio and client agnostic portfolio) to ensure they are comparable. If they are not, then the difference in performance could indicate certain strengths or limitations of the model and it may require adjusting to ensure greater alignment with the client in question. For example the client portfolio might contain richer exposure data and therefore provide a more accurate representation of the risk which offers valuable additional insight into model performance. For each of these steps, a big difference in losses should be a warning flag for further analysis. Large differences would be in the region of +/- 30%.

8. Historical loss – return period analysis

Purpose

Ensure that return period modelled loss for the full severity-frequency distribution provide estimates that support but are not constrained by recorded history.

8a. Assess the alignment of loss estimates between the model and historical experience (client agnostic). This step is similar to Step 7a, but uses the full event set to generate modelled losses and compares these against historical loss experience over the available record, collating data from all relevant data sources (e.g. EM-DAT, ReliefWeb, government datasets, and insurance industry data). Any differences between modelled and observed losses must be clearly accounted for, particularly given data limitations and the long-time span over which events occur. In many cases, sufficient historical data doesn't exist to compare against the full event set as less frequent-more severe events have not been experienced in recorded history. However, data should exist to support an assessment of the lower end of the Return Period curve. In addition to portfolio-level comparisons, where data permits, checks should also be undertaken at a sector-specific level.

8b. Assess the alignment of loss estimates between the model and historical experience (client specific). This step is similar to Step 7b, but applies the full event set to generate client-specific model losses, which are then compared against the client's observed loss experience where sufficiently detailed historical data is available (ideally greater than 20 years). This comparison is only feasible where the client maintains robust, granular loss records that allow for consistent and meaningful benchmarking against model outputs.

8c. Assess the alignment of modelled loss estimates with alternative analytical estimates. Where alternative modelled loss estimates are available for the region, whether publicly accessible or provided by the client, these should be used to benchmark results, with clear explanations provided to understand the drivers of any differences. Comparative model evaluation is useful for understanding structural uncertainty and identifying the drivers of differences in hazard, vulnerability and loss estimates. The ability to interpret such differences may be constrained by the level of access to underlying model methodologies. In cases where alternative models are not available, results may instead be benchmarked against parameterised historical loss data, such as extreme value distributions fitted to observed loss records.

Expect: Comparison of stochastic modelled loss estimates with observed loss data across the full frequency–severity distribution, as well as with any other available sources of loss information. Included alongside these comparisons should be a narrative explaining potential sources of difference and any limitations of the analysis, including what part of the return period curve can be well evaluated using this approach and importantly what can't be. e.g. insufficient historical record to validate more severe-less frequent events. If a model performs well on step 8a but not on step 8b then a comparison needs to be made between the two portfolios (client portfolio and client agnostic portfolio) to ensure they are comparable. If they are not, then the difference in performance could indicate certain strengths or limitations of the model and it may require adjusting to ensure greater alignment with the client in question. For each of these steps, a big difference in losses should be a warning flag for further analysis. Large differences would be in the region of +/- 30%.

9. Future climate risk evaluation (optional)

Purpose

Whilst a future climate risk component is not critical to supporting the development of DRF strategies or instruments, governments, regulators and multilateral development banks are requesting greater insight into projected physical risk (i.e. future damage from catastrophic events). This evaluation step provides confidence in the modelled output to be used for long-term financial decisions but does require a significant amount of expertise in relation to both climate modelling and catastrophe modelling.

9a. Ensure that projections of future exposure data are included in the modelling of climate risk given its relative importance in loss projections. Projections of how exposure will evolve over time are a key component in assessing the impact of future events and can be derived using sources such as Intergovernmental Panel on Climate Change (IPCC) reports, particularly Shared Socioeconomic Pathway (SSP) scenarios. These projections typically reflect changes in population, urban development, and migration patterns. Any provided projections should be validated against alternative datasets (e.g. FuturePop) and supported by a clear narrative explaining the underlying drivers of change.

9b. Evaluate how climate risk model outputs have been incorporated into the catastrophe risk model to support projected loss estimates for future time periods. The influence of climate change on the modelled peril is most likely obtained from either downscaled⁵ Global Climate Models (GCMs) or Regional Climate Models (RCMs). Three approaches to incorporating climate model outputs into the hazard component of catastrophe models are:

- **Direct:** The percentage change in the climate driver is applied directly to the hazard metric of interest (e.g. expected percentage changes in rainfall are applied directly to flood depth). Whilst this approach can theoretically provide future estimates it is very crude and the level of uncertainty is significant as complicated processes (e.g. how rainfall leads to flooding) are ignored.
- **Change factor:** A climate model or alternative scientific information is used to generate change factors that are applied to the original event set to alter the severity and frequency of events. For example, for tropical cyclones Knutson et al.⁶ published projected changes in tropical cyclone properties (frequency and severity).
- **Re-modelling:** Climate model data are directly used as the input data to the underlying event-based modelling used to develop a future climate event set (e.g. using climate model precipitation data as an input to the hydraulic and hydrologic modelling used to determine flood risk).

[5] Downscaling refers to the process of translating outputs from coarse-resolution global climate models (GCMs) into higher-resolution, regionally relevant climate information. This is achieved either through statistical relationships or dynamical modelling to better represent local-scale climate variability and extremes that are not resolved at the global model scale. After downscaling systematic errors need to be resolved through bias correction methods.

[6] Knutson, T., and Coauthors, 2020: Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. Bull. Amer. Meteor. Soc., 101, E303–E322, <https://doi.org/10.1175/BAMS-D-18-0194.1>.

9c. Assess the treatment and communication of uncertainty. As future risk is being assessed, no loss evaluation can be performed on this data so uncertainty levels are significant.

Expect: Estimates of projected future risk founded on scientifically accepted methodologies with consideration of changes to hazard and exposure. These estimates are expected to incorporate large levels of uncertainty given the assumptions involved and inability to evaluate any loss estimates.

CONCLUSION

Evaluating a catastrophe risk model is essential as governments increasingly rely on their outputs to support financial decision-making, where the credibility of those decisions depends on the extent to which the model accurately represents both underlying hazard processes and real-world impacts. Not all steps in the full evaluation framework are required for every use case: optional components such as exposure evaluation, event set characteristics, correlation, or future climate risk assessment are only necessary where these elements are included in the model or where sufficient capability exists to assess them. The framework provides governments with a step-by-step approach to evaluating risk models, setting out what to expect at each stage and supporting the reliability of key CDRF metrics (e.g. AAL, return period losses, standard deviation), which form the quantitative basis for disaster risk financing strategies, financial instrument design, and broader risk-informed financial decision-making.

ANNEX: KEY QUESTIONS TO ASK BEFORE TRUSTING A MODEL

This section is an addendum to the main content of the note. It provides an initial checklist of questions to ask of a model supplier to support in the assessment of the appropriateness of a model for CDRF decision-making. The checklist is designed for users who understand catastrophe risk models but without direct, hands-on modelling experience. This can include risk professionals in a government disaster risk management agency, staff within a Ministry of Finance risk unit, or external consultants providing risk advisory support to the government.

1. Qualitative

1. What perils, geographies, and sectors (e.g. buildings, infrastructure, agriculture) does the model cover, and what does it not cover?
Expect: A clear specification of what the model covers to clarify whether this meets requirements.
2. Who developed the model, how often is it updated, and when was it last updated?
Expect: A reputable model development company that has developed models for similar use cases previously and the model is kept up to date, incorporating new data and research as it becomes available (i.e. ideally updated within the last 5 years).
3. Does the model run consistently without technical errors?
Expect: Stable performance and a good user experience.
4. Is model documentation provided (e.g. model methodology), that explains data sources, assumptions, and validation methods in accessible language?
Expect: Transparent documentation publicly available or shared under a non-disclosure agreement detailing how the model has been developed, key data sources, any critical assumptions made throughout development, known limitations of the model and any validation checks performed by the model provider.
5. In addition to those used by the model provider are there sufficient data sources available to support a model evaluation?
Expect: A list of available data sources with which both the individual model components (e.g. weather station observations to evaluate the hazard component) and the full model output (e.g. recorded loss data) can be validated. The amount of data required depends on the use case and hazard being modelled. For example, to evaluate individual model components one would need 30+ years of wind data (to account for multi-decadal signals), 10+ years of up-to-date rainfall data (given changes in rainfall this must be close to the most recent 10 years) and 50-100+ years of seismic data (this is heavily dependent on source). To evaluate the full model, it is ideal to have 30+ years of detailed historical loss or claims data, although 20 years can sometimes be sufficient depending on use case.
6. What outputs does the model produce and are these sufficient for the intended use case (e.g. Exceedance Probability (EP) curves, year loss tables (YLTs), modelled hazard metrics)?
Expect: At a minimum the basic financial metrics required for CDRF instrument design are available (e.g. AAL, Return Period loss) along with associated model outputs (e.g. Exceedance Probability (EP) curves, YLTs).

7. Has the model been independently peer reviewed or validated by external experts?
Expect: For existing models ideally at least one independent review will have been performed, for newly commissioned modelling this likely won't be available.
8. Has the model been used by other government programmes in the region?
Expect: Evidence of previous operational use can provide additional confidence but should not be considered a prerequisite for model suitability.

2. Exposure evaluation (optional)

1. How has the model supplier validated the total insured values or population exposed, including by region and sector?
Expect: Comparison against national statistics and other available third-party data sets (e.g. Overture).
2. How does the model handle missing data, or unknown property characteristics?
Expect: Explanations for what assumptions are made where key attribute data is missing.
3. If exposure data is provided in aggregate form (i.e. number of people per province) how is it disaggregated to support model execution?
Expect: Catastrophe models depend on accurate geolocation of exposure data. Where exposures are provided in an aggregated format, they must be disaggregated to coordinate-level locations either before model execution or within the modelling process itself. The disaggregation should preserve the characteristics of the original dataset and be undertaken using a transparent, scientifically defensible methodology, such as the use of population, land-use, or building-density data. The relationship between the resulting exposure resolution and the hazard model resolution should also be clearly explained, as there is generally little benefit in disaggregating exposure data to a finer spatial scale than that supported by the underlying hazard model.
4. How sensitive are modelled losses to exposure resolution, geocoding accuracy or key property characteristics?
Expect: Evidence that higher resolution location data and inclusion of a greater number of key property characteristics reduces uncertainty and does not significantly alter loss metrics.

3. Event set characteristics (optional)

1. How many stochastic events are in the event set?
Expect: A sufficiently large number of events to capture the full range of possible events (both in terms of frequencies and spatial scale) and ensure statistical stability. What is sufficient will be defined by the use case, the peril and the size of the country/area being modelled.
2. How was the event set developed?
Expect: Detailed documentation outlining the methodology used, why it was chosen, what data sources were used in development, how data gaps were resolved and any limitations of the approach. Documentation should include reference to peer-reviewed studies/recognized methodologies.
3. What hazard metrics are provided by the model (e.g., flood depth, peak wind speed, peak ground acceleration)?
Expect: Scientifically accepted intensity measures that have a good relationship with damage and can be evaluated against recorded data.

4. Can historical events be individually identified within the model and evaluated against external sources?

Expect: Yes, model includes the ability to extract hazard footprints for specific, recent historical events to help with evaluation of the hazard model. This should include the comparison of visual interpretations of the event footprint as well as a statistical comparison of event level metrics.

5. Have sensitivity or convergence tests been carried out to show stability of the results?

Expect: Documentation of tests and assessment of how sensitive model results are to the number of simulations, including a minimum number of simulations for stability in results.

4. Hazard parameter evaluation

1. Has the hazard component (e.g., flood depth, wind speed) been calibrated to observed data (e.g. satellite data or station data)?

Expect: Quantitative comparisons between observed and modelled hazard data should be undertaken to assess the extent to which the hazard component aligns with historical observations. Any material discrepancies should be accompanied by a clear explanation, together with a transparent assessment of the range of hazard frequencies that could be evaluated. All data sources used in the assessment should be documented, including their vintage, purpose, and any modifications/extensions.

2. Are key model assumptions (e.g. decay rates) observable in both the modelled and recorded data?

Expect: Quantitative comparisons between modelled and observed data showing the model assumption and its correlation with recorded data.

5. Damage function evaluation

1. How were vulnerability (damage) curves developed?

Expect: An explanation for what data has been used to develop the damage curves (i.e. claims data, engineering studies, or both).

2. Are curves region/country-specific or imported from other regions where there could be different building practices or different material strengths?

Expect: Local derivation of vulnerability functions using local data is preferred; imported curves are acceptable but should be adjusted to ensure they are relevant for the local context.

3. How do the functions differ by construction type (e.g., wood, concrete) and use (e.g. residential, commercial)?

Expect: Logical sensitivity - stronger materials suffer less damage at the same hazard level.

4. Have validation checks been done using post-disaster claims data or independent damage surveys (where the necessary data is available)?

Expect: Comparison plots or summary results to be made available. Note that validation checks should be done using data that has not been used to develop the original curves.

5. When there is missing attribute data how are damage curves selected and what is the sensitivity of the approach (i.e. if missing data what happens)?

Expect: Clear documentation of the default damage curves applied where attribute data (e.g. construction type, occupancy, number of storeys) is unknown, and the rationale for those defaults. Sensitivity testing should also be available showing how modelled losses change when alternative curves are assumed, so that the materiality of missing attribute data can be understood.

6. Correlation (optional)

1. Are events independent or do they include clustering (e.g., aftershocks of earthquakes being related to the main earthquake event, or European windstorms occurring in clusters)?
Expect: Clear documentation outlining clustering methodology and comparison with observed records. This documentation should also identify how clustering relates to global climate drivers (e.g. ENSO, IOD, etc..).
2. Can correlation settings be modified in the model, and if so, how sensitive are losses to changes?
Expect: Sensitivity analysis explaining the impact or correlation on losses and why a certain correlation setting has been chosen.

7. Historical loss – single event analysis

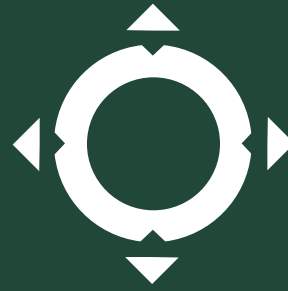
1. How was the appropriateness of the exposure database used for client agnostic model evaluation assessed?
Expect: Description of data sources used to develop the exposure data, their vintage (age) and completeness, and any benchmarks against which the exposure was compared.
2. Which historical events have been included in the model and do they compare well with recorded historical losses – for both client agnostic and client specific portfolios?
Expect: Several historical events to have been well documented with event reconstructions available either in the event set or as standalone (these should match historical data of the event to ensure minimal differences in hazard). Losses from these events for either client agnostic or client specific portfolios can then be compared with reported loss data for the event (E.g. post disaster damage assessments). Event comparisons should be provided along with detailed evaluation documentation that outlines how observed loss estimates were obtained, including what data sources were used and how they were adapted to overcome any challenges (e.g. differences in types of loss being covered or different vintage of data)
3. Are discrepancies explained (e.g., missing perils, data quality, exposure differences) sufficiently to enable confidence in the model?
Expect: A transparent discussion on the potential discrepancies, where these may originate from and where possible quantification of these differences.
4. How is uncertainty presented both within individual components and in the final output (e.g. confidence intervals, standard deviations, etc..)?
Expect: Numerical quantification of uncertainty in the analysis and impact on modelled loss outcomes, as well as for recorded data, if relevant.

8. Historical loss – return period analysis

1. How does the modelled AAL compare with long-term reported loss records for both client specific and client agnostic portfolios?
Expect: Modelled loss estimates should demonstrate a reasonable degree of correlation with observed losses and fall within a plausible range of historical averages. Any material discrepancies should be investigated and accompanied by a clear discussion of the potential drivers of the differences, including data limitations, modelling assumptions, and changes in exposure over time.
2. Do the exceedance probability (EP) curves align with historical large-loss experience?
Expect: Explanation of large historical events (i.e. those assessed in the previous step) and where they might sit in respect to the EP curve, with an assessment of any discrepancies in loss between the historical event and the EP curve.
3. How is uncertainty presented both within individual components and in the final output (e.g. confidence intervals, standard deviations, etc.)
Expect: Numerical quantification of uncertainty in the analysis and impact on modelled loss outcomes.

9. Future climate risk evaluation (optional)

1. What baseline climate period does the model represent, and how is this reference period defined?
Expect: A clear statement of the baseline climatology (e.g. present-day, 1981–2010, or a defined reference year) and confirmation that hazard, exposure, and vulnerability components are internally consistent with this period. All future projections should be explicitly referenced back to this baseline to ensure comparability.
2. What future climate scenarios, time horizons, and climate models are used to generate projections?
Expect: Availability of multiple recognised scenarios (e.g. SSP1-2.6, SSP2-4.5, SSP5-8.5) and time horizons (e.g. mid-century, end-century), supported by a transparent selection of GCMs and/or RCMs that represent a plausible range of future climate conditions relevant to the peril and region.
3. How are climate model outputs processed and incorporated into the catastrophe model to generate future hazard and event sets?
Expect: A clear description of the methodology used (e.g. direct scaling, change factors, or full re-modelling), including how downscaling and any bias correction are applied, and how these outputs are translated into changes in event frequency, severity, and spatial characteristics.
4. How are future exposure changes represented and validated within the modelling framework?
Expect: Transparent methodology for projecting exposure changes (e.g. population growth, urban expansion, migration, asset value change), using recognised sources such as SSP-based datasets, and validation against alternative datasets (e.g. FuturePop or national projections), with a clear narrative of underlying drivers.
5. How are projected changes in loss attributed and how robust are the results to key uncertainties?
Expect: Clear separation of loss drivers (hazard change, exposure growth, and vulnerability assumptions), sensitivity testing across scenarios, and explicit treatment of uncertainty arising from climate models, socioeconomic pathways, and modelling choices.



Global Risk Modelling Alliance

www.grma.global

Licensing and Reuse Disclaimer

This document is made available under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Under the conditions of this license, any individual or entity is permitted to copy, distribute, transmit, and adapt the content of this document, in whole or in part, and in any medium or format, for any purpose, including commercial use. Such use is permitted provided that appropriate credit is given to the original author(s) and source, a link to the license is included, and any modifications made to the original content are clearly indicated.

While reasonable efforts have been made to ensure the accuracy and reliability of the information presented in this document at the time of publication, the licensor does not warrant that the content is free from errors or omissions. The information is provided “as is” and may be subject to change without notice.

The licensor shall not be held liable for any direct or indirect loss or damage arising from the use of the material. Users are solely responsible for ensuring that their use of the content complies with the applicable license terms and does not infringe any rights of third parties, including but not limited to intellectual property rights.

The full legal text of the license is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>



KFW



Federal Ministry
for Economic Cooperation
and Development



Frankfurt School
of Finance &
Management