INVESTOR LEADERSHIP NETWORK

Sponsored By
Investor Leadership Network

WHY VENDORS **DISAGREE**

A Practical Guide To Evaluating Physical Climate Risk Data Vendors

Author: Nik C. Steinberg
June, 2026

INVESTOR LEADERSHIP NETWORK

EXECUTIVE SUMMARY

This study examines the significant methodological differences among climate risk data vendors, highlighting why their assessments of physical risk often produce widely varying results for the same location. The study presents a framework for institutional investors to assess various climate risk vendors using eight methodological categories and three dimensions of quality. Questionnaires and dummy portfolios were sent to seven vendors. Findings reveal that discrepancies stem from a multitude of choices, including how assets are identified to how estimates of physical hazard exposure are translated into financial loss estimates. We find that approaches vary considerably, underlying the vast differences in outputs presented in recent studies. Not all dispersion can be explained by these choices. Some disagreement stems from genuine, irreducible uncertainty in forecasting rare events. Though, if methods are sound and well-justified, the differences across vendors are valuable, adding richness and multiple views of an uncertain future. Conversely, methods lacking rigor and completeness could potentially lead to blind spots in screenings or even maladaptation if acted upon. The ILN hopes this report serves as a practical guide for investors navigating a marketplace lacking consumer guidance.

HOW TO USE THIS PAPER

Investors need to match their specific needs to the right vendor(s). The questions and evaluation criteria outlined below provide a flexible framework for engaging vendors and guiding selection choices. Evaluating vendor choices within these methodological dimensions could help investors understand vendor methods in a more systematic fashion, leveraging the evaluation criteria to understand whether those choices are justified or problematic.

A key takeaway from this exercise is that, although vendors differ widely in methods and findings, each brings a distinct view of how future climates will impact investable assets, and so long as methods are robust and well-justified, vendor differences add richness and diversity to our view of an uncertain future. As such, the vendor evaluation is intended to (1) discern the plausible reasons for dispersion across vendors, and (2) help investors make more informed decisions about vendor suitability based on their use case.

ABOUT THE INVESTOR LEADERSHIP NETWORK (ILN)

Launched at 2018 G7, the ILN champions initiatives and facilitates collaboration across leading global investors committed to accelerating the transition to a more inclusive and sustainable economy. The ILN's membership comprises 13 global institutional investors across six countries, with over US$10tn in assets under management. This platform encourages members to share resources, expertise and networks to develop, promote and deliver scalable initiatives and solutions on climate change, diversity and inclusion, and sustainable infrastructure.

The ILN established its Climate Change Advisory Committee (CCAC) to facilitate collaboration among global investors, build on existing guidance and best practices, and promote and operationalize practices for managing climate-related risks and opportunities . The ILN's CCAC provides investors and other industry stakeholders with resources and guidance to assess, manage and mitigate the impacts of climate change within their investment portfolios. The initiative's previously published reports have supported investors in integrating pertinent climate-related initiatives and practices, with an aim to share findings with the wider industry and advance the capacity of investors to understand and mitigate climate risks.

TABLE OF CONTENTS

INTRODUCTION

In the last decade, climate projections have found their way into the offices of investment institutions. Ushering in this data are dozens of physical climate risk data vendors ("vendors"), competing for market share. This creates pressure to achieve higher precision or offer more models, neither of which necessarily equate to higher accuracy. Meanwhile, the rapid increase in demand has coincided with an equally rapid increase in climate risks, regulatory nudges, and acknowledgement by leadership that climate risks pose a strategic and significant risk to investments. Physical climate risk data offers investors a glimpse at the level and character of physical climate risks facing the assets underlying their investments, from which decisions can be made to de-risk and formulate longer-term strategies better adapted to future climates.

Vendors offer investors risk data, which is often considered alongside other asset or company-level information, to inform investment and risk reduction decisions, but vendors – even when evaluating the same asset – disagree about which hazard presents the greatest potential risk (Section 2.1) and even disagree about the presence of acute risks in at least one instance examined in this study (Section 2.2). The 2025 study[1] by the Global Association of Risk Professionals (GARP) demonstrated significant dispersion across thirteen vendors and highlighted several modelling differences contributing to that spread. This paper presents additional areas of meaningful dispersion and builds on GARP (2025) by tracing how certain methodological choices throughout the modelling chain translate into differences in physical and financial risk estimates. Here, we take a step toward understanding those differences by examining methodological choices made in the long chain translating climate hazards to measures of physical and financial risk.

Applying climate models to financial risk is tricky business. Some would argue that climate models are not designed or capable of asset-level forecasting,[2] and climate models are heuristic, and their predictive accuracy will always be difficult to test.[3] Even when supplementing signals from climate models with more advanced approaches, several decisions must be made: which models to select, which risk metrics to generate, how best to integrate vulnerability factors, to name just a few. Complicating matters, there is no standard or well-worn path akin to risk measurement by insurers, measuring perils with natural catastrophe models. All of this leaves an enormous amount of responsibility with the research teams of vendors to make judgments. Teams make their best effort, undoubtedly, but that internal process and their ultimate decisions of how to proceed within the grey areas of methodological design vary so considerably that it creates material differences in their outputs.

Differences in climate risk assessments often stem not from raw climate model outputs, but from up- and downstream methodological choices. If methods are sound and well-justified, the differences across vendors are valuable, adding richness and multiple views of an uncertain future. Conversely, methods lacking rigor and completeness could potentially lead to maladaptation if investors fail to properly scrutinize. In both cases, a potentially sizable portion of the spread stems from compounding uncertainty, whereas an upstream decision - as nuanced as whether or not to include vegetation in wildfire hazard modeling - can significantly alter a vendor's ability to identify relevant hazards and size the financial risks.

[1]GARP (2025). A Risk Professional's Guide to Physical Risk Assessments A GARP Benchmarking Study of 13 Vendors. Available at https://www.fca.org.uk/publication/corporate/risk-professionals-guide-physical-risk-assessments-garp-benchmarking-study-13-vendors.pdf

[2]Fiedler, T., Pitman, A. J., Mackenzie, K., Wood, N., Jakob, C., & Perkins-Kirkpatrick, S. E. (2021). Business risk and the emergence of climate analytics. *Nature Climate Change*, 11(2), 87-94.

[3]Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263(5147), 641-646.

INVESTOR LEADERSHIP NETWORK

PURPOSE

The origins of this study can be traced to early 2025 under a G7 initiative, the ILN Climate Change Action Committee (CCAC). Members from 13 investment institutions, managing over USD$10 trillion collectively, shared their challenges in selecting and evaluating vendors. In recent years, the marketplace of physical risk providers has surged, and differences are not easily discernable from demo calls and vendor's own documentation. In response, the ILN commissioned a consultant to conduct an evaluation by asking vendors a series of questions about their approach and requested them to analyze the same dummy portfolio to better discern why some vendors might arrive at different viewpoints of risk for the same asset.

The purpose of this comparative study is to articulate specific methodological differences across vendors, and when possible, surface the reason for those differences, including important assumptions, as well as the gaps and strengths of the prevailing approaches. We hope this study reveals important insights into the reasons for those differences while providing investors with a practical guide for conducting their own vendor evaluation during the selection process.

This study presents findings from the engagements spanning June 2025 to April 2026, and is organized according to eight methodological areas (Figure 1) where differences are quantified and vendor methodologies are evaluated, including a brief diagnosis of the possible reasons for dispersion (Section 3).

01 METHODS

Vendors were presented with a dummy but theoretically plausible investment portfolio, containing three real assets, five listed equities, and two linear assets. Specific coordinates were provided to vendors for the real asset and linear assets, along with building type descriptions, to allow vendors to apply vulnerability factors. Vendors were provided with listed equity names and tickers and asked to return the number of affiliated corporate assets, as well as asset- and corporate-level risk scores by hazard, time period, and scenario. To enable comparison, some metrics were standardized (e.g., grouping “drought” and “water stress,” converting CVaR to average annual loss percentages). Revisions were made with vendors’ permission.

Vendors were then asked to provide responses to a questionnaire containing 24 questions (Section 3) to understand their methods across the eight methodological areas. Their written responses and published methodologies served as the primary input for evaluating the level of rigor, completeness, and transparency of their methods.

The eight methodological areas were chosen to cover a broad set of issues and to reflect the areas with the greatest subjectivity, where decisions can materially affect the quality of outputs. Rigor evaluates practices consistent with standards established in the literature; completeness ensures that critical components (e.g., exposure, hazard, vulnerability, or local datasets) are not omitted from approaches; and transparency is measured according to the extent of assumptions, rationale, and limitations provided by vendors to help end-users with interpretation and understanding uncertainty. Vendor-specific scores were then categorized and anonymized to facilitate vendor participation.

FIGURE 1. EVALUATION FRAMEWORK

Eight methodological areas covering meaningful aspects of physical risk modeling and assessment for financial stakeholders

- **Model Skill and Agreement**
- **Spatial Resolution**
- **Exposure and Materiality**
- **Scenarios and Probabilities**
- **Hazard Indicators**
- **Vulnerability Indicators**
- **Financial Metrics**
- **Physical Risk Reduction**

3 dimensions for evaluating vendor methodologies

Rigor | **Completeness** | **Transparency**

- 0 = Not suitable for climate risk analysis. Methods do not meet basic standards.
- 1 = Suitable for screening purposes only[4] and high-level hotspot identification based on strong directional change signals, using limited model testing, adequate spatial resolution, minimal asset validation, and simple indicators.
- 2 = Methods are reasonably sound and documented; appropriate for comparing relative risk across assets, identifying priority areas, and determining where deeper analysis is required.

[4]Screening is generally considered the first step in understanding broad climate risk exposure, and appropriate for directional insights but not for quantitative interpretation, prioritization, or site-based decision-making.

02 FINDINGS

Findings are divided into two-parts:

1. Quantification of the differences in results across vendors for a select number of metrics that proved comparable across most vendors (Sections 2.1 – 2.3).
2. An evaluation of the level of rigor, completeness, and transparency for each of the eight methodological areas (Section 3).

INVESTOR LEADERSHIP NETWORK

2.1 REAL ESTATE AND RANKING OF KEY HAZARDS

To explore agreement in risk identification, we asked vendors to provide a metric representing the combined risk (hazard intensity, vulnerability, and potential impact/ losses) for each of the three dummy real estate assets. Vendors provided scores in the form of dimensionless alphabetic (A – F) or numerical (0 - 100) ratings.

No two vendors agreed on which hazards pose the top two greatest potential risks to the three real estate assets provided (Figure 2). In only one instance did most vendors (6 of the 7 vendors) agree on the most material hazard: the Singapore Warehouse is highly exposed to extreme heat. Conversely, one vendor identified a key hazard when no others did, noting potential wildfire risk in Singapore. Curiously, the warehouse is situated within an almost entirely industrial area with little to no burnable vegetation, suggesting the vendor's wildfire model relies exclusively on fire weather conditions to measure hazard severity. and excludes important, non-climate factors such as land cover and vegetation dynamics. Another vendor noted some, albeit not very high, tropical cyclone risk in Singapore, even though Singapore lies outside the primary tropical cyclone tracks affecting the western North Pacific and is rarely directly impacted by tropical cyclones.

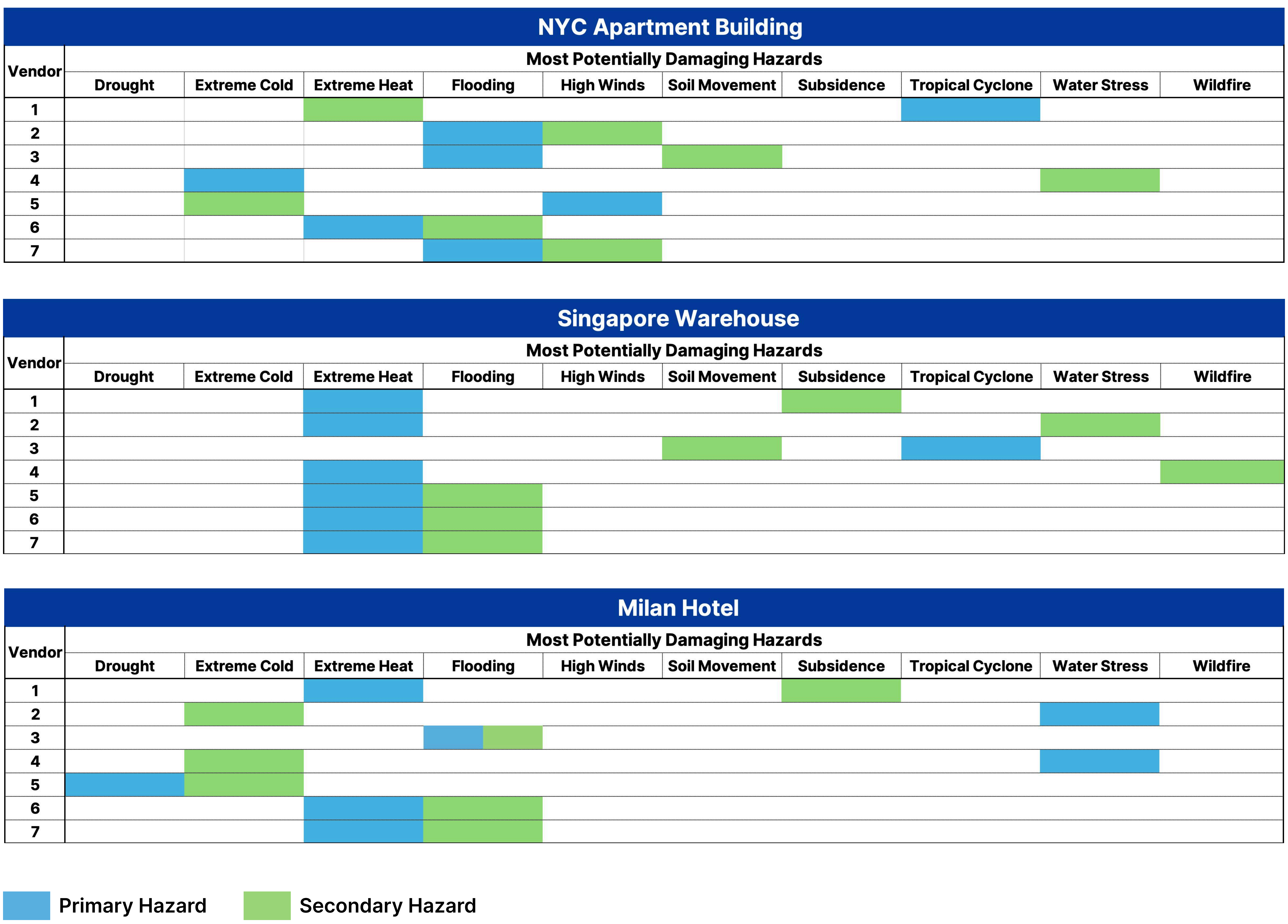

Figure 2. First and second most potentially damaging hazard facing three real assets

2.2 INFRASTRUCTURE AND FLOOD RISK

Identifying risk signals is perhaps one of the most important functions of climate risk assessments. Unlike some hazards where screening metrics can be interpreted more continuously, flood risk is highly threshold-dependent: small differences in elevation, drainage, defenses, or flood-model assumptions can determine whether a site is flood-exposed or not.

Due to the different measurements of flooding used by vendors, we assigned a binary function (True/ False) to vendor results across eleven locations along the A4 toll road that follows the Seine and Marne Rivers on the outskirts of Paris, France (Figure 3). A site was considered exposed to flooding if a vendor indicated a depth greater than zero millimeters - either riverine or rainfall-based flooding - during a modeled 1-in-100 year flood event by mid-century under SSP5-8.5 scenario.

In all but one instance, we find high disagreement among vendors. All vendors indicate at least some exposure at Location 3 (green dot), a hotspot due to its proximity to a confluence of the Seine (distance = 28 meters), elevation (roadway stands 5 meters above river banks) and its location just upstream from a flood-control structure where floodwaters are likely to accumulate during a flood event.

And yet the discord across the remaining ten sites is statistically indistinguishable from a coin flip.[5] When measuring the level of agreement across all eleven sites, there is virtually no inter-vendor agreement beyond random chance alone.[6] Put another way, across all but the most clear-cut cases of flood exposure along this stretch of roadway, the vendors are more likely to contradict each other than to agree. Only six vendors evaluated this roadway.

Figure 3-1. Eleven sites of the A4 toll road near Paris provided to vendors for flood analysis.

[5]Using a measure of disagreement per each site, we find a mean entropy of 0.78, whereas 0 is equivalent to full consensus, and 1, complete disagreement.

[6]Fleiss Kappa (κ) = 0.008. Fleiss Kappa gauges how consistently several independent raters assign binary (true/ false) outcomes to categories of measurement. It estimates how much of their agreement exceeds what random chance would produce, giving a measure of reliability among a group of raters randomly selected from a larger population. κ range: <0 = Less than chance agreement; 0.01 – 0.20 = Slight agreement; 0.21 – 0.40 = Fair agreement; 0.41 – 0.60 = Moderate agreement; 0.61 – 0.80 = Substantial agreement; 0.81 – 1.00 = Near-perfect agreement

Figure 3-2. Count of vendors that indicate flood risk (blue = no flood, red > 0 millimeters inundation) across 11 sites of the A4 toll road near Paris. Riverine and rainfall-based flooding were considered.

2.3 EQUITIES AND FACILITY COUNTS

Understanding what assets are exposed to hazards is a critical starting point for measuring the extent of climate risks facing equities. Yet establishing the physical operational footprint of publicly listed companies is inherently difficult. Vendors typically purchase or scrape data to identify corporate assets, but their underlying datasets vary widely in coverage and quality.

One of the most influential drivers of these differences is simply the number of assets each vendor identifies. We asked vendors to report the number of sites associated with five publicly listed companies of varying sizes and geographic footprints. For any company, reported asset counts differ substantially (Figure 3).

These discrepancies likely reflect a mix of internal capacity (e.g., staff or automated systems used to clean and build datasets), financial resources (e.g., how much third-party data a vendor can afford to acquire), choice of data providers, and methodological decisions, such as whether to include subsidiary assets or apply stricter criteria for linking sites to a parent company.

A larger asset inventory may indicate a more complete footprint, but data quality ultimately depends on how rigorously those assets are processed and validated (see Section 3.3). Because of these comparability issues, we could not compare risk scores across the five companies, though it is reasonable to expect that large differences in the number of identified assets could lead to material differences in vendor's evaluation of company-level physical climate risk. Only five vendors evaluated the equities.

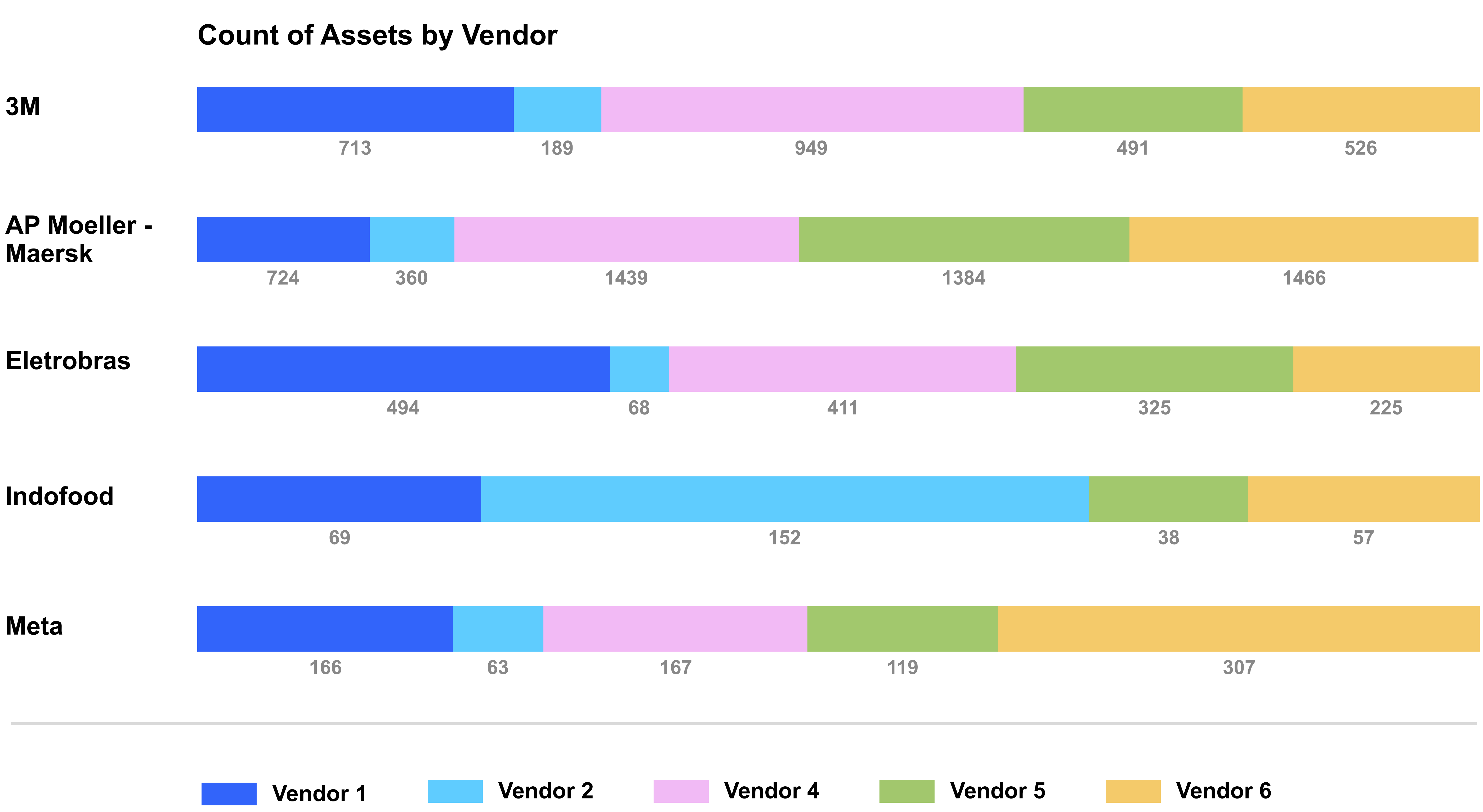

Figure 4. Number of corporate assets considered. Note, Vendors 3 and 7 do not provide analyses of corporate climate risk.

03

REVIEW OF VENDOR METHODOLOGIES

To surface some of the plausible reasons behind these differences, a synthesis of vendor methodologies is provided below. Vendors are assigned numerical ratings (0-2) based on varying levels of rigor, completeness, and transparency of vendor responses, and organized according to eight methodological areas. Identifiable information such as names and supplier information has been removed to preserve anonymity.

INVESTOR LEADERSHIP NETWORK

	Vendor 1	Vendor 2	Vendor 3	Vendor 4	Vendor 5	Vendor 6	Vendor 7
1.Model Skill and Agreement	5	4	4	3	5	5	4
2.Spatial Resolution	6	5	5	0	6	6	5
3.Exposure and Materiality	3	3	3	3	6	6	6
4.Scenarios and Probabilities	6	5	6	4	6	6	6
5.Hazard Indicators	5	2	5	3	6	6	4
6.Vulnerability Indicators	4	3	3	5	6	6	5
7.Financial Metrics	5	3	4	4	6	5	5
8.Physical Risk Reduction	4	0	4	2	5	5	5

Transparency — 6 — Rigor

Completeness

"6" is the highest possible score

Score = 0 Score = 1 Score = 2

Figure 4. Suitability (0-2) of approach assigned to each vendor for each methodological area and level of rigor, completeness, and transparency evidenced in vendor's responses and documentation.

INVESTOR LEADERSHIP NETWORK

3.1 MODEL SKILL AND AGREEMENT

Model agreement is a measure of the number or percentage of models among a larger ensemble agree across the same parameters. Some models may indicate more rain, while others less. Neither is necessarily incorrect. In fact, model variability adds richness to model ensembles, though vendors must first evaluate levels of agreement or disagreement in order to leverage techniques like clustering,[7] and ultimately, attach levels of confidence and uncertainty to model results.

Separately, model skill is a measure of how well General Circulation Models (GCM) and Regional Circulation Models (RCMs) (also referred to as Global Climate Models and Regional Climate Models, respectively) reproduce observed climate features: mean state, extremes, spatial patterns over a particular region, variable, or timescale. GCMs and RCMs operate at coarse spatial scales (~50 to 250 kilometers), which is in part why downscaling is so critical (see next section).

In applied, non-academic settings (i.e., risk assessments for investment portfolios), the Intergovernmental Panel on Climate Change (IPCC) assessment reports rely on large number of GCMs in modeling physical hazards (i.e., ensembles) due to the unique bias and approximations of each model. Over large areas, multi-model ensemble mean and median estimates generally outperform any individual model.[8] This ensures simulations span a range of physical process representations and emergent climate sensitivities, thereby providing a fuller spectrum of plausible climate futures. Notwithstanding this, in some cases it is justified to curate and select individual models. This is the case when depicting complex hazards that GCMs have historically struggled with,[9] such as precipitation extremes, tropical cyclones, wind extremes, humidity, fire weather, and complex rainfall dynamics. Using the 10-20% most climate sensitive models in an ensemble can also reflect worst case scenarios,[10] model skill testing is required to justify excluding models with known bias or weighing more skillful models based on how well they represent local, historical conditions and events.

The number of models considered is also vital considering the initial spread in model results. The CMIP experiment, CMIP6 (Coupled Model Intercomparison Project Phase 6), includes over 36 models, but most vendors use a subset of models due to computational and consistency considerations. As for disclosing model names and testing methods, vendors should be follow IPCC guidelines and peer-review publication standards, disclosing individual model names, performance results (e.g., Root Mean Square Error), and their methodological rationale for decisions. Investors should ask vendors why certain models were included or excluded as part of the model skill testing process.

Refinements to the ensemble group can be conducted through model skill tests, measuring a model's ability to reproduce past conditions at local levels. Testing model skill and agreement are critical first steps to

[7]Clustering is a technique that assessment teams can use to make sense of individual model results by quantifying the level of model agreement/disagreement across key variables. It indicates which and how many models project a certain direction and degree of change, enabling assessment teams to produce best- and worst-case scenarios and Representative Climate Futures (RCF) .Source: Whetton, P., Hennessy, K., Clarke, J. et al. (2012). Use of Representative Climate Futures in impact and adaptation assessment. *Climatic Change* 115, 433–442. https://doi.org/10.1007/s10584-012-0471-z

[8]Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of geophysical research: atmospheres*, 118(4), 1716-1733.

[9]HighResMIP and upcoming CMIP7 releases will offer higher native spatial resolution.

[10]Bevacqua, E., Fischer, E., Sillmann, J., & Zscheischler, J. (2026). Moderate global warming does not rule out extreme global climate outcomes. *Nature*, 651(8107), 946-953.

measuring and representing model and inter-model variability, the dominant sources of uncertainty out to mid-century,[11] and can help vendors decide which GCMs and RCMs to include/ exclude and weight as part of a model ensemble.

Rigor How are climate models tested for skill? How is model agreement determined? How are climate models selected?	**Completeness** Which GCM or RCM are used?	**Transparency** How do you test the skill of individual models? The agreement across models? How do you integrate findings from this exercise?
0/7 — 0 - No model testing	0/7 — 0 - Not enough models to run an ensemble	0/7 — 0 - No detail on models used or rationale
7/7 — 1 - Some model testing based on ability to reproduce past conditions using observational or reanalysis data	2/7 — 1 - A limited ensemble of GCMs or RCMs are considered	4/7 — 1 - Approach is detailed, but limited rationale for selected approach
0/7 — 2 - Models tested for skill across variables, time periods, and regions. Results integrated into model selection and/ or weighting	5/7 — 2 - All GCMs or RCMs are initially considered	3/7 — 2 - Fully documented approach, including critical assumptions and rationale

VENDOR RESPONSE

Four of the seven vendors test model skill, while three rely on testing from upstream public data sources. None of the vendors test model skill across all variables, regions, and time steps. Vendors conduct or rely on validation at the global scale and treat that as evidence of skill at the asset level when in fact models may perform poorly at the local scale, especially for variables driven by local processes (precipitation extremes, low-level winds, mesoscale phenomena). The four vendors that do test model skill rely on ERA5 reanalysis for bias correction, though test against global-mean quantities. Two vendors go further by conducting internal out-of-sample testing, simulating specific events to understand the ability of models to reproduce past extremes. One vendor takes a distinctly different approach, selecting the single model producing the greatest relative increase in severity per hazard (i.e., selecting the driest model for drought, wettest for flood). It is worth noting that a model that reproduces past climate conditions well can still diverge under non-stationary future conditions (shifts in circulation regimes, novel forcing levels, tipping dynamics). In the near-term, natural catastrophe models are more fit to capture tail-end extremes.

None of the vendors measure model agreement, or leverage measures of agreement for communicating inter-model variability and uncertainty. While model consensus may not equate to accuracy, when

[11]For an expanded discussion on the sources of uncertainty and their contributions to overall uncertainty through time, please see Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, 90(8), 1095-1108.

several models with distinct code lineage and parameterization agree on the direction and strength of the climate signal, it helps add confidence to models results. Yet, no vendors measure agreement, despite its widespread use in similar contexts.[12] As such, no vendors were assigned a “2” for rigor as investors need to understand levels of model agreement and disagreement when discerning the extent of model uncertainty shaping the vendor’s results.

Instead, most vendors rely on CMIP6 multi-model ensembles often using established statistically downscaled products such as NASA NEX-GDDP rather than conducting their own model curation. Model counts vary widely, from curated subsets of four to five models for specific hazards to more than thirty GCMs drawn from the full NEX-GDDP archive. One vendor instead uses a single GCM with a 100-member ensemble for tropical cyclones, prioritizing internal variability within a single model over inter-model variability.

Only a small number provide transparent documentation of how individual model skill assessments influence selection or adjustments. While model coverage is generally adequate, vendors were docked points for relying too heavily on only a few models, not performing any sort of model skill tests, or providing only limited visibility into how model uncertainty is managed and communicated to users.

3.2 SPATIAL RESOLUTION

The physical impacts of climate change are inherently local. Without the right level of detail, vendors may miss localized risks. Hazards that require additional spatial layers - such as floods, coastal storms, and wildfires - are especially localized, requiring finer spatial resolution. At the portfolio-level, low-fidelity data coverage - across hazards or assets - limits investors ability to compare hazard exposure or geographies and make even high-level inferences about risk concentrations or hot spots impossible.

In most cases, when vendors leverage ensembles of GCMs from downscaled datasets[13] additional downscaling (< ~25 x 25 kilometers) is not always necessary, especially for hazards like extreme heat or drought as these phenomena unfold at larger spatial scales.

However, hazards such as floods, coastal storms, and wildfires are especially localized, requiring a spatial resolution that approximately matches the resolution of the area of interest, which could range from building scale (e.g., ~30 x 30 meters), to neighborhoods (e.g., ~1 x 1 km), to natural landscapes (>1 km). While topographic overlays enable finer hazard modeling, finer resolution does not necessarily improve predictive skill.

Overprecision and hyper-local resolutions tend to produce errors, especially when topographic products carry sampling or inherent uncertainties. For example, digital elevation models (DEM) are often used to improve

[12] Whetton, P., Hennessy, K., Clarke, J. et al. (2012). Use of Representative Climate Futures in impact and adaptation assessment. *Climatic Change* 115, 433–442

[13] The most commonly used downscaled datasets included NASA’s statistically downscaled Earth Exchange Global Daily Downscaled Climate Projections derived from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (NEX-GDDP-CMIP6) and the World Climate Research Programme’s (WCRP) dynamically downscaled datasets for specific regions, called the Coordinated Regional Climate Downscaling Experiment (CORDEX).

the resolution of inland and coastal flood inundation models, but even some of the leading DEMs are known to carry ranges of uncertainty up to one meter, vertically. Vendors should carefully examine (and communicate) the types and dispersion of uncertainty across all products, including those used for bias-correction or to enhance topographic detail when determining the added value of depicting finer spatial resolution.

Rigor What is the level of geographic data (i.e., spatial resolution) in your model data?		**Completeness** How does the spatial resolution of your model data vary by geography? By hazard? By variable?		**Transparency** If model data has been processed to achieve a higher spatial resolution (i.e., downscaling), can you explain the methods used and potential uncertainties or limitations?	
1/7	0 - For localized risks, no downscaling is applied, and estimates reflect risk at the native resolution of the GCM/ RCM models	1/7	0 - Spatial resolution of model data is not detailed enough for the scale of assets or no downscaling was conducted from model's native resolution	0/7	0 - No detail provided on processing methods
0/7	1 - Downscaling is applied to better represent local conditions for the hazards described above, but is still much coarser than resolution of the asset(s) in question	0/7	1 - Spatial resolution is sufficient for some assets or hazards, but not all	1/7	1 - Some details provided on methods, but no rationale provided for chosen method (i.e., statistical vs. dynamical downscaling)
6/7	2 - Same as 1, with additional high-resolution layers applied (i.e., flood defense infrastructure, vegetation, land cover, etc.)	6/7	2 - Spatial resolution of model data is sufficient for all hazards and asset types	6/7	2 - Full documentation of spatial processing method and rationale, with references to peer-review work

VENDOR RESPONSE

With only one exception, vendors represent acute, highly localized hazards such as flooding and wildfire at finer scales. A majority deliver flood estimates at 90 x 90-meter grids, with a few offering meter-scale mapping in some regions due to the availability and integration of LiDAR or other high-fidelity elevation data. These vendors generally provide flood at the finest scales, wildfire and temperature at intermediate scales, and drought or water stress at coarser scales, reflecting the availability and quality of underlying data sources. Across the group of seven, there is shared recognition that higher resolution is not inherently better; as over-precision can introduce errors when downscaling or when applying geoprocessing methods that cannot fully account for local detail.

Most vendors offer at least partial descriptions of methods, acknowledging the downscaling approach taken but without explaining how or why the downscaling approach was conducted or chosen. Three apply more rigorous procedures, including quantile-delta mapping, out-of-sample skill testing, and a few dynamically downscaled RCMs to better represent regional conditions. At least one vendor reports

INVESTOR LEADERSHIP NETWORK

no downscaling at all. Those vendors working with high-resolution data report similar sources of uncertainty: dependence on DEM quality for flood modeling, smoothing of extremes during interpolation, uncertainties in boundary conditions, and reduced reliability in data-sparse regions. Only a few vendors explicitly describe validation steps, while others rely on documentation provided by downscaling products (e.g., NASA NEX-GDDP or CORDEX).

3.3 EXPOSURE AND MATERIALITY

Knowing what is in harm's way is critical, and location errors occur when vendors lack precision, validation, and multiple sources. For single assets, this could mean validating coordinates through simple desktop research; for linear assets and polygons, it could mean mapping the critical nodes and dependencies; and for corporates, it could mean identifying owned facilities, suppliers, and even supply sources. Ideally, vendors should retain enough spatial asset information to support modeling of upstream and downstream implications, with flexibility to define the scope and pathways of those cascading impacts. Many third-party spatial asset providers sell vendors large but messy data sets with incomplete or inaccurate location information, so vendors should have processes in place to check for accuracy. Vendors should also capture risks beyond the four walls of an asset, considering companies' entire value-chain or an asset's dependencies (e.g., water supply or power station). The completeness of their exposure asset data set can be evaluated by the number (or percent of known total), including the data sets they rely on to identify and map relevant and material assets.

However, it is difficult, if not impossible, to identify and map all the spatial locations that play a material role in the value of an organization or individual asset. As such, vendors should indicate the extent to that is known/ unknown through some expression of confidence.

Rigor How do you identify and validate the assets that are relevant?		**Completeness** What data sources are used to identify material assets?		**Transparency** Have all relevant material assets been considered? Can you describe what might be missing?	
1/7	0 - No validation of client-provided data (i.e., via satellite imagery, AI, or desktop checks)	0/7	0 - Location data is provided by client, no wider footprint or additional data sets are integrated	2/7	0 - Limited detail on validation process or data usage, no quantitative or qualitative expression of confidence is provided
3/7	1 – Some validation conducted for assets provided by client	6/7	1 - Additional data are incorporated to broaden the spatial footprint of an asset to capture, for example, flooded areas nearby that could limit access to asset	2/7	1 - Vendor has detailed their validation process and data sources, but has not provided an expression of confidence

INVESTOR LEADERSHIP NETWORK

3/7 — 2 - Validation of client-provided assets plus consideration of additional up/ downstream assets and dependencies from research or third-party data providers	1/7 — 2 - Multiple data sets are collated, capturing wider spatial footprint and reasonable share (e.g., beyond Tier 1 suppliers and/ or interdependencies) of the known value-chain	3/7 — 2 - Vendor has detailed their validation process, data source(s), and provided some expression of confidence

VENDOR RESPONSE

Across vendors, the rigor in validating an asset's location varies widely, and only a few vendors apply multi-step verification processes. One vendor provides no evidence of validation, having placed coordinates of a land-based asset in the ocean. All vendors must contend with the messy and incomplete nature of commercial asset databases; but only a minority have robust processes to detect errors, reconcile duplicates, or prevent misplacements such as coordinate drift or incorrect geocoding.

Completeness of exposure asset datasets also varies, with only a small subset integrating multiple sector-specific databases to capture relevant facilities (See Figure 4). Only one vendor attempts to integrate up or downstream dependencies broadening the scope of exposure analysis outside the four walls of a building. While some vendors achieve high precision for individual buildings through manual methods, only one vendor attempts to systematically map suppliers and dependencies linked to an asset.

Four of the seven vendors provide some measure of confidence in the extent or accuracy of their coverage, equivalent to "2" in the transparency category, describing known gaps, categorizing coverage quality ("low", "medium", "high"), or quantifying the share of assets captured at different levels of certainty. The remaining four vendors offer limited detail on methods, do not articulate uncertainties, and provide no indication of coverage confidence.

3.4 SCENARIOS AND PROBABILITIES

Scenarios play a pivotal role in expressing future uncertainty, especially over longer periods when policy-related decisions and decarbonization pathways play an outsized role in warming trajectories. In the nearer-term, warming is "locked-in" from past emissions and the outcomes and uncertainties of certain acute hazards (or in this context, perils) can be expressed through probabilistic analysis, leveraging catastrophe risk models and historical loss data from insurance claims.

At a minimum, vendors should provide multiple scenarios, distinct enough to represent a wide range of potential futures to capture low to high warming outcomes; leveraging scenarios designed by IPCC, Network of Central Banks and Supervisors for Greening the Financial System (NGFS), and other institutions. For discrete events, such as floods or high wind events, vendors are expected to provide discrete probabilistic estimates of how likely they are to occur.

Rigor What other futuring techniques are applied (e.g., narratives)? Has any probabilistic analysis been conducted for any hazards?		**Completeness** How many different scenarios are considered? For which hazards, have probabilities been applied?		**Transparency** Why have said scenarios been selected? How have other said approaches (e.g., narratives or probabilistic analysis) been constructed?	
0/7	0 - Vendor only uses scenarios to express future outcomes	0/7	0 – Vendor does not provide multiple scenarios	2/7	0 - Limited detail on approach, rationale, and critical assumptions
1/7	1 - Additional analysis is conducted, either via narratives (i.e., hot/ dry vs. cool/ wet futures) or probabilistic analysis	0/7	1 – Vendors provides multiple scenarios	2/7	1 - Detailed approach, but limited rationale for selected approach
6/7	2 - Vendor considers both climate scenarios and probabilistic events sets for relevant hazards	7/7	2 – Vendor provides multiple scenarios, and probabilistic estimates for all relevant hazards	3/7	2 - Fully documented approach, including critical assumptions and rationale

VENDOR RESPONSE

All vendors provide multiple SSP-RCP scenarios, most commonly SSP1-2.6, SSP2-4.5, and SSP5-8.5, and one vendor maps to NGFS scenarios given its work with banks. Probabilistic analysis is standard, with most vendors offering return periods (typically 5-year to 500- or 1,000-year) for acute hazards, and half of the vendors provide percentile distributions (10th, 50th, 90th). One provider offers an extreme Antarctic Ice Sheet collapse scenario for coastal flood and another vendor models 50,000+ synthetic years to estimate tail risks. Vendors take a broad definition of “probabilistic”, citing frequency analysis (Extreme Value Analysis), Bayesian or bootstrap methods, and Monte Carlo methods for generating return periods and confidence intervals. All vendors follow IPCC or NGFS guidance when selecting hazards, but only a few vendors describe how event catalogs are constructed, how uncertainty distributions are derived, or how tipping-point dynamics are incorporated. Fewer than half provide fully documented approaches that clearly articulate both assumptions and methodological reasoning.

3.5 HAZARD INDICATORS

Extracting climate signals requires robust measures of change hazards, especially in the tail-end of event distributions. Compared to changes in averages, changes in tail-end risks tend to generate larger, year-over-year losses. To do this, vendors should consider distribution of risks (i.e. percentiles, return periods) and incorporate thresholds with known or very likely material impacts (e.g., 30 mm of flood inundation triggers insurance claim or physiological threshold for reduced labor output is about 35°C Wet Bulb Globe Temperature).

Signals can be measured in relative (i.e., x more days, or x% more severe than baseline) or absolute terms (i.e., x meters of inundation, or x number of wildfire days). Ideally, vendors consider both relative and absolute changes, capturing tail-end risks, and measure as many characteristics of change as possible: change in timing, frequency, magnitude, duration, and spatial extent. Failure to depict these characteristics can lead to false negatives and limit the vendor's ability to identify relevant material transmission pathways, which could underestimate the sector-specific risk changes (e.g., the seasonal effects on labor stress in the construction sector). Along the long chain of translation, hazard exposure measures are a critical step, and vendors should provide a strong rationale for how they extract these metrics from the models and identify meaningful climate signals.

Rigor How are tail-end risks measured?	**Completeness** Do hazard metrics focus on absolute metrics, relative change, or both? What characteristics of change are measured?	**Transparency** Can you provide a list of indicators and metrics for measuring hazard exposure risk? How have they been processed and why have they been selected?
0/7 — 0 - No evidence of tail-end risk analysis	1/7 — 0 - May include measures of absolute or relative changes, but not both.	1/7 — 0 - Limited detail on approach, rationale, and critical assumptions
1/7 — 1 - Vendor has identified tail-end risks and constructed indicators that depict tail-end risks	3/7 — 1 - Considers absolute and relative change indicator for all hazards, but does not include all characteristics: severity, changing, timing, duration, spatial extent	2/7 — 1 - Detailed approach, but limited rationale for selected approach
6/7 — 2 - 1 + indicators representing thresholds evidenced to trigger material impacts	3/7 — 2 - Considers both absolute and relative measures of risk AND all relevant characteristics of change	4/7 — 2 - Fully documented approach, including critical assumptions and rationale

VENDOR RESPONSE

All vendors provide hazard-specific intensity measures, through flood depth, wind speed gusts, temperature threshold exceedances, wildfire probability or wildfire weather, and precipitation accumulation. Nearly all vendors quantify tail-end risks through return-period analysis, with most extending to 500- or 1,000-year events and applying methods such as Extreme Value Analysis and Monte Carlo simulation, generating large synthetic event catalogs. Only a minority go further by incorporating empirically grounded impact thresholds, such as flood depths known to trigger insurance claims (i.e., 30 mm) or heat-stress limits tied to human physiology (e.g., 35 WBGT for labor stress) or relative thresholds (i.e., number of additional days exceeding 95th percentile). Most vendors estimate changes in severity and frequency, but not spatial extent, duration, or timing, assumingly to align with insurers' approach to risk quantification, leaving gaps in understanding how risk evolves across operationally meaningful dimensions. Measuring relative changes are important for identifying emergent characteristics of change, such as the occurrence of an event in a place, or during a part of the year with no historical precedent, which is likely underinsured, and unacclimated.

All vendors report hazard-specific intensity metrics, but transparency in how these indicators are derived varies considerably. Most describe their processing steps, such as bias-correction, extreme value analysis, or back-testing against historical events, but provide limited rationale for why these processes or specific indicators and thresholds were selected. Over half of the vendors fully document their methodological choices, including the scientific or engineering basis for chosen metrics, the assumptions embedded in their processing pipelines, and the justification for thresholds tied to material impacts.

3.6 VULNERABILITY INDICATORS

How an asset responds to a hazard (i.e., vulnerability) is a key part to understanding overall risk. Robust methodologies will incorporate measures of vulnerability, which are specific to asset type, industry sector, geography, and ideally, some combination of all. More rigorous approaches will integrate vulnerability factors specific to an asset and its characteristics (i.e., data centers and the efficacy of their cooling technologies), while other, less robust approaches may only consider sector-specific factors, perhaps through overweighting of adjustments (e.g., water scarcity and food and beverage companies/ assets). Best-in-class approaches provide known, standardized hazard damage or business impact functions across key real asset classes, such as transmission and distribution (T&D), timberlands, and buildings, commonly held in global investor portfolios. For example, applying a consistent wind speed threshold or damage impact metric for electric transmission towers based on voltage rating.

It is widely acknowledged that assets respond differently according to asset characteristics, hazard type, and the severity of the hazard. A robust methodology will leverage vulnerability metrics for each of these dimensions. For example, the impacts of flooding on a commercial property with no drainage or pump equipment are worse than a building equipped with flood-adaptive designs. It can be difficult for providers to obtain asset-specific characteristics for all assets within a portfolio, and the use of certain proxies should be expected. As an example, these could be applied to assets (e.g., building age) or sectors (e.g., regionally-specific energy or water demand elasticity). Best-in-class approaches will provide flexibility to adjust default hazard thresholds when asset owners have more detailed or asset-specific information. For instance, while new buildings are built to withstand 165 km/h wind gusts, this can be calibrated up or down based on factors such as asset age, construction standards, or maintenance history.

If vendors are integrating vulnerability factors and applying them according to building types or by sector, then they are most likely leveraging peer-review data sets or proprietary loss/ claim information. Alternatively, more precise estimates may leverage asset-specific details provided directly by companies and asset owners. In either case, it is important to understand the data sources informing these vulnerability factors/ weightings, as they can play an outsized role in the total expression of vulnerability.

Rigor How is vulnerability of an asset factored in? What about sector-specific vulnerabilities?		Completeness How is vulnerability applied across different hazards and asset types?		Transparency What data sources were used to inform these views/estimates of vulnerabilities? What are some of their potential limitations and uncertainties?	
0/7	0 - No consideration for vulnerability	0/7	0 - No vulnerability factors are incorporated	0/7	0 - No detail provided
4/7	1 - Considers vulnerability at a more macro level, at the regional or sector-level only	4/7	1 - Vulnerability factors are applied to multiple asset types	2/7	1 - Some detail on methods, but no rationale provided for chosen data source(s) or method(s)
3/7	2 - Integrates quantitative measures of vulnerability at the asset-level, leveraging known relationships between hazard level and impact (i.e., via damage loss curves)	3/7	2 - Vulnerability factors consider sub-asset characteristics	5/7	2 - Full documentation provided with named data sources, peer-review citations, and rationale

VENDOR RESPONSE

All vendors integrate vulnerability factors, though granularity differs substantially. The most sophisticated approaches apply vulnerability factors at the building component level and assign damage curves according to individual physical building elements (i.e., roofing, foundations, electrical systems, HVAC, cladding). This enables building-specific vulnerability profiles that vary by construction type, age, occupancy, and hazard intensity. One vendor offers over half a million unique combinations: 390 asset types across 1,500 damage functions, with breakdowns per hazard (e.g., cooling costs for heat, physical damage for flood/wind/cyclone, water expenses for drought). It is worth noting that such granularity - across asset types and damage functions - may improve relevancy but not necessarily predictive skill. As is so often the case, increased complexity raises the risk of over-fitting.

Conversely, most vendors apply vulnerability factors based on broad archetypes, including three to five different building types (e.g., commercial office, industrial, residential) with standardized damage functions for each hazard. FEMA HAZUS is cited by multiple vendors as a foundational source for flood and wind vulnerability functions. Some vendors allow users to override default characteristics with asset-specific inputs when available, though the extent to which this is operationalized varies. In some cases, it is central to the workflow; in others it is an add-on available outside the vendor's platform. Archetype-based approaches offer scalability but may over- or under-estimate risk for assets deviating significantly from typical construction standards. FEMA does not advise the HAZUS loss estimates be applied to individual assets because values are calibrated at the portfolio-level.[14] A common limitation across the vendors is the reliance on loss information and damage functions from US or Western European data sources, which may affect accuracy when applied to assets in other regions with

[14]FEMA. (2023). Hazus Flood Model Technical Manual (Version 6.0). Federal Emergency Management Agency, U.S. Department of Homeland Security

different construction standards or regulatory environments. Several vendors supplement with proprietary data sources, including engineering firm partnerships, insurance claims databases, or integrate loss and damage information from laboratory tests. Most vendors acknowledge specific limitations, such as the US-centric nature of the HAZUS database, limited claims data for non-standard asset types, and over interpolating portfolio level loss estimates to individual assets. One vendor explicitly describes a bottom-up, first-principles engineering approach drawing on empirical disaster event data and engineering calculations to estimate damage function for individual components of a building, rather than relying on insurance claims data. Two vendors provide limited detail for their vulnerability estimates, which limits the ability to assess reliability or robustness. Other commonly acknowledged limitations include lack of lifetime degradation functions for chronic hazards, data availability gaps in emerging markets, potential bias from generic damage curves applied to specialized facilities and limited empirical validation data for building type.

3.7 FINANCIAL METRICS

Translating physical risk to financial risk requires pairing hazard, exposure and vulnerability characteristics with financial metrics at the asset-, firm-, or portfolio-level. For example, at the asset level, many vendors estimate expected average annual loss (AAL) as a way to price the financial risk or to size the cost of risk reduction investments. This can be further enhanced through a cost-of-failure analysis that incorporates repair costs and potential downtime impacts, both upstream and downstream. The way in which vendors arrive at these metrics is also important. Most commonly, vendors rely upon open-source datasets or proprietary datasets showing insurance claims, both of which contain significant variability for a single asset, due to building-to-building differences, limited data, and overfitting. Insurers and vendors typically select a median value for a large coverage region, failing to capture potential losses for specific building types based on their physical attributes, function, and location.

The transmission pathways from physical to financial risk are numerous, and robust methodologies will consider several financial measures at the asset- or firm- level, including, but not limited to direct capital impacts, damages, business disruptions, and changes to insurance premiums or coverage.

Few standardized approaches for translating physical risk to financial risk exist, and metric comparability at this stage is difficult as financial risk metrics are the result of a long chain of upstream assumptions. Nevertheless, vendors should articulate their methods, ideally through peer-reviewed research or via clear and defensible examples supported by strong rationale. For example, a vendor may provide comprehensive documentation on the derivation of damage functions. For many asset classes, such as electrical transmission and distribution infrastructure, standardized construction codes across countries can serve as a foundational reference for understanding damages to specific property types when confronted with hazards across multiple severity levels.

INVESTOR LEADERSHIP NETWORK

Rigor How is financial risk considered and calculated?	**Completeness** Which financial impacts are considered? What transmission pathways are modeled? How do these differ by hazard?	**Transparency** Can you share the methodological steps and key data inputs for translating physical risk to financial risk? What are the critical assumptions and limitations?
1/7 — 0 - No financial risk metrics are provided	0/7 — 0 - Financial metrics are not considered or provided	0/7 — 0 - Limited detail on approach, rationale, and critical assumptions
2/7 — 1 - Financial risk metrics are provided, but are derived from insurance claims, regional estimates, or other non-asset-specific extrapolations	3/7 — 1 – Multiple financial transmission pathways considered	3/7 — 1 - Detailed approach, but is not supported by other research or known practices
4/7 — 2 - Financial metrics are derived using a bottom-up approach, incorporating asset-specific information	4/7 — 2 - Consideration for relevant transmission pathways and financial impact metrics across multiple asset types and hazards	4/7 — 2 - Fully documented approach, including critical assumptions and rationale. The approach is well-established or supported, ideally through vendor's own peer-reviewed methodology.

VENDOR RESPONSE

Almost all vendors provide some form of a financial risk metric (though some financial institutions prefer to generate their own), most commonly an Expected Annual Loss (EAL) or Average Annual Loss (AAL) metric that aggregates expected damage across return periods. The most rigorous approaches consider specific components within buildings. One vendor explicitly quantifies likelihood of events and financial consequences through Monte Carlo simulations, producing repair costs (dollars) and downtime (days) with uncertainty bands based on component level replacement and upgrade costs. Most top-down approaches integrate some combination of company-level financial data, industry revenue benchmarks, and macro-level economic modelling, and produce metrics such as climate-adjusted asset values, credit probability of default adjustments, and portfolio-level Value-at-Risk (VaR). The distinction between direct asset damage (a CapEx loss) and business interruption or revenue loss (an operational expense) is made explicitly by most vendors, though the precision of these estimates - sometimes to the fourth decimal place - are incompatible with the uncertainty at this downstream stage of analysis.

Most vendors consider multiple financial transmission pathways: direct physical damage (CapEx), business interruption and revenue loss (OpEx and Revenue), productivity loss, insurance premium changes, and macro-level impacts on property values, and in some cases, expected GDP losses for the wider region. All vendors take a hazard-specific view of how hazards transmit to financial loss: flood translates to depth-damage repair costs plus downtime; wind to structural damage and associated business interruption; heat to productivity loss, HVAC costs, and employee health impacts; and drought to water cost or yields. At least one vendor explicitly limits its financial output to revenue impairment from direct asset damage without modelling a broader set of transmission channels. No vendors consider stranded assets or the inability to operate due to impacts in the wide operating environment, including energy or water use requirements that may conflict with local regulations or community needs. No vendors integrate indirect financial impacts, such as supply chain disruptions, utility outages, and access constraints.

Critical limitations noted include dependence on the quality of claims data, potential misalignment between generic vulnerability functions and the actual characteristics of specific assets, and the challenge of capturing compound or cascading financial impacts. Several vendors provide only a high-level description of the translation process without sufficient methodological detail to independently evaluate the approach.

3.8 PHYSICAL RISK REDUCTION

Identifying and quantifying risks alone is insufficient for decision-making. When climate risk assessment provides sufficient, decision-ready information, then users may decide to identify and size relevant risk reduction efforts.

However, assets are not islands and aspects of their resilience span far beyond their physical footprint, which means the completeness of risk reduction options can be judged by the diversity of options, spanning the wider operating environment. For example, options can span physical risk reduction (e.g., flood barriers or restoring natural drainage areas), risk transfer (i.e., insurance coverage and strategy), redundancy, planning, behavioral, and even community engagement efforts, to name just a few. Of all the topics evaluated so far, this is where the largest numbers of vendors falter. Risk reduction recommendations are almost entirely focused on asset hardening through engineering solutions.

Vendors should not only demonstrate how risk reduction measures were identified, but also how risk changes with different levels of intervention and investment, including the limitations and assumptions behind these calculations and recommendations. The level of rigor can best be judged by the scale at which judgements are made, the most sophisticated outputs are capable of generating quantitative metrics (e.g., ROI) tied to specific efforts.

Rigor How are risk reduction recommendations identified? How are associated costs (or costs of inaction) calculated?		Completeness What risk reduction areas are considered?		Transparency What are the limitations and critical assumptions of these metrics and recommendations?	
2/7	0 - No consideration given	2/7	0 - No consideration given	2/7	0 - Limited detail on approach, rationale, and critical assumptions
1/7	1 - Provides descriptive guidance only	5/7	1 - Options are limited, covering a small subset of plausible options	2/7	1 - Detailed approach, but limited rationale for selected approach
4/7	2 - Integrates quantitative measures of cost and benefit of different risk reduction options	0/7	2 - A wide range of risk reduction are considered, spanning physical risk reduction, transfer/ insurance, diversification, or community and municipal engagement strategies	3/7	2 - Fully documented approach, including critical assumptions and rationale

VENDOR RESPONSE

The range of capabilities across vendors for identifying and quantifying risk reduction opportunities is among the widest of any topic assessed. The most advanced approaches quantitatively match hazard-specific vulnerabilities to a curated library of measures and compute expected avoided loss, payback periods, and savings-on-investment ratios at the asset level. This enables direct cost-benefit comparison of different interventions and an explicit quantification of the cost of inaction (baseline minus adaptation option). One vendor leverages a component-based structural model, identifying the specific components driving losses and simulating alternative configurations, including both physical interventions and operational measures. Several vendors embed risk reduction measures within their impact functions, effectively quantifying the cost of inaction implicitly (i.e., no action = losses). At the time of writing, at least two vendors have “adaptation” capabilities actively in development but not yet in production. At least one vendor explicitly limits consideration of risk reduction to floods based on third-party data, with no other hazards addressed.

Risk reduction measures are predominantly engineering-based, covering physical interventions such as flood barriers, building elevation, roof upgrades, wildfire defensible space, wind-resistant construction standards, and water-recycling systems. This concentration reflects the relative maturity of quantitative methods for engineering interventions compared to governance, social, or environmental strategies. One vendor explicitly addresses operational measures (e.g., pre-retaining contractors for faster repairs, parametric insurance for faster access to financing) as a third category alongside physical hardening of the building. Several vendors note that their use of different Shared

Socioeconomic Pathways (SSP) - as part of their SSP-RCP scenarios - supports qualitative exploration of governance and social adaptation dimensions, though direct quantitative modelling of these areas is largely absent. Community-level, nature-based, and planning-based adaptation options, which may represent more cost-effective and durable interventions in some contexts, are not addressed by any vendor. Many vendors offer conflicting terminology, often conflating risk reduction (i.e., asset hardening or protection measures) with adaptation.

Common limitations cited include the use of generalized cost ranges that may not reflect site-specific conditions (local labor and materials costs, permitting requirements, design specifics), and assume that adaptation measures are effective long-term and properly implemented and maintained. Several vendors acknowledge that some risk reduction measures transfer rather than eliminate risk; for example, a flood barrier may protect against overtopping events but could increase residual depth if breached, and that interdependencies with external infrastructure can limit realized benefits of investments at the asset-level. At least one vendor notes that compounding climate risks (e.g., simultaneous riverine flooding and tropical cyclone wind) are not currently assessed, and interventions addressing cascading impacts are too complex and idiosyncratic to reasonably model.

04 LIMITATIONS

Differences in methodological approaches made metric-to-metric comparisons difficult, limiting the extent to which vendors' outputs could be evaluated side-by-side. As a result, only a brief quantitative analysis was feasible, focusing on the few metrics that were readily comparable across all vendors: ranking of key hazards (Section 2.1), presence of risk (Section 2.2), and corporate asset counts (Section 2.3). Some of the divergence observed in these comparisons may therefore reflect not only genuine methodological differences, but also the constraints of fitting heterogeneous vendor outputs into common metric and methodological categories.

The use of a standard questionnaire to evaluate methodologies introduces its own limitations. The eight methodological areas were selected by the author because they represent domains where ambiguity and disagreement are most common within the modeling community; however, future research may identify additional areas of meaningful divergence. Vendor materials were largely descriptive and did not include results from validation exercises (e.g., model performance against present-day hazards) or sensitivity analyses (e.g., the influence of specific assumptions on outputs) that vendors have presumably conducted internally, but did not share for this study. Consequently, a degree of subjectivity was required when ranking vendors. While Section 3 outlines standard practices, vendor scores were assigned on a relative basis. Vendors that provided insufficient methodological detail were penalized on transparency-related questions, and all scores reflect current capabilities rather than future roadmaps.

While several vendors were assigned highest possible score across methodological areas, it was not possible to discern if these vendors were also capable of informing high-stakes decisions at the site-level, which would require deeper analysis among a subset of those vendors using more advanced, bottom-up modeling approaches. Lastly, evaluations are not an endorsement of any vendor, and organizations should consider their own use case before using this guide to select vendors.

Finally, the seven vendors included in this study represent only a subset of the broader marketplace. There are likely twice as many physical climate risk data providers offering financial risk insights to investors.[13] The sample was limited to vendors that had previously engaged with ILN members, providing a constrained but hopefully representative view of the range of approaches currently in use.

[13] GARP (2025) analyzed 13 physical climate risk vendors

CONCLUSION

We can't directly validate physical climate risk outputs against future outcomes; the only thing we can evaluate today is whether the methods are robust, complete, and well-justified.

When evaluating the same asset or entity, we find that vendors present distinct views of exposure, the direction of risk, and the ranking of material hazards. Much of this disagreement stems from dozens of assumptions made along the long chain of decisions, translating hazard modeling to approximations of financial risk. Some dispersion will always exist due to irreducible uncertainty in modeling the future, but much of the dispersion stems from methodological choices.

Vendors can improve robustness of their outputs by communicating model spread and agreement, and highlighting what is uncertain or unknowable. Vendors can also demonstrate skill where validation is possible and trade precise, point projections for fuzzy explorations of the possible range of futures. This study also identified several blind spots in vendor methodologies and offerings. Vendors are narrowly focused on risks facing discrete assets, and mostly ignore the compounding and cascading nature of risks unfolding in the wider operating environment, communities, and supply chains surrounding an asset. Physical risk reduction efforts are similarly focused almost entirely on engineering solutions at an asset, with little attention paid to corporate governance, insurance strategies, workplace safety, and efforts that would benefit from closer collaboration with municipalities, utilities, and communities.

For investors and other downstream users, more robust decision-making is needed. Even with better vendor agreement and transparency, deep uncertainty will persist. Investors should not expect precision or perfect projections but instead consider solutions that improve resilience across a range of possible futures. Investors can also play a larger role at the outset, helping vendors understand what decisions are on the table. Most of the methodologies examined in this study are suitable for screening and regulatory disclosures by flagging and sizing relevant climate risks, but are not sufficient for site-specific capital allocation. High-stakes, capital allocation decisions at the asset-level require a deep understanding of an asset's unique vulnerabilities and dependencies, stress-testing to understand its performance around specific hazard thresholds, and supplemented by site-level inspections and independent, third-party reviews.

In the absence of standards, buyer education and vendor transparency are essential, but not sufficient. Formal standards may eventually emerge but are unlikely to keep pace with the rapid expansion of the vendor landscape. One immediate path is a voluntary vendor ensemble, modeled loosely on the CMIP experiments in climate science. By running diverse methodologies side-by-side under shared protocols, CMIP allows researchers to examine model spread, identify differences and their reasons, and ultimately turn dispersion into a richer understanding of plausible futures. A similar approach for physical risk vendors, anchored in a basic set of standards to ensure comparability, could help investors see where methods and findings converge, where and why they diverge, and identify places where deeper inquiry is warranted. For vendors, an ensemble would help expand their reach, strengthen credibility through comparability, and showcase the distinct strengths of their approaches without compromising proprietary information, and ultimately bring greater transparency to the market.

INVESTOR LEADERSHIP NETWORK

ACKNOWLEDGMENTS

In no particular order, we would like to thank the following vendors for their participation:
Class 3 Technologies
Climate X
First Street
Jupiter Intelligence
ICE
S&P
XDI (Cross Dependency Initiative)

Internal reviewers:
Caroline Kidnie (OMERS)
Katharine Preston (OMERS)
Daisy Streatfeild (NinetyOne)

Independent technical review by
Jo Paisley (GARP)
Maxine Nelson (GARP)

AUTHOR BIO

Nik Steinberg is a climate risk and adaptation specialist. He was managing director of climate research at Four Twenty Seven (now Moody's), senior advisor of climate risk for a USAID implementing partner, and now leads Vector Climate Group, supporting organizations as they navigate complex climate and nature related risks and opportunities.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

CONTACT

For questions about the ILN Climate Change Initiative Please contact:
secretariat@investorleadershipnetwork.org

www.investorleadershipnetwork.org