

<https://doi.org/10.1038/s44304-025-00090-7>

Quantifying the compounding effects of natural hazard events: a case study on wildfires and floods in California



Sam Dulin^{1,2}, Madison Smith^{1,2}, Beth Ellinport¹, Benjamin Trump¹, Jesse M. Keenan^{1,3} & Igor Linkov¹ ✉

Compounding weather events occur when two or more weather hazards combine to produce societal or environmental impacts. Despite the potentially catastrophic damage these hazards can cause to critical infrastructure, the quantification of such compounding hazards lags behind. This study presents a new method for analyzing how consecutive severe weather events interact and compound damage over time. The framework examines whether initial events amplify the impact of subsequent disasters, using flexible lag identification to determine optimal recovery periods. Applied to California wildfires and subsequent floods, it reveals regional variations in compounding effects and their temporal evolution. By quantifying these compounding hazards, this methodology provides crucial insights for policy makers and resilience managers to address climate change risks and the dynamic nature of compounding hazards. While developed for California, this approach can be applied globally to assess compounding hazards and inform disaster preparation strategies.

In recent years, the frequency and intensity of these extreme weather events have exhibited a rapid and concerning escalation attributed, in some part, to climate change and increased measures of physical exposure^{1–3}. This escalation has heightened the urgency to model and quantify such extreme events, particularly as they relate to the resilience performance of interconnected infrastructure systems and the risk of compound events^{4,5}. In ref. 6, the authors defined a typology of compounding hazards, and they defined compound hazards as ‘combinations of multiple drivers and/or hazards that contribute to societal or environmental risk’. The authors in ref. 5 narrowed this definition of compounding hazards to pertain specifically to infrastructure systems, wherein compounding hazards are defined as two or more hazard impacts that co-occur such that they concurrently affect interdependent critical infrastructure systems, thereby presenting multiplicative risks that may exceed the operating or functional capacity of systems and communities. In particular, compound events may lead to cascading effects that amplify the magnitude and complexity of disruptions⁶.

The emerging field of compound events research represents a critical paradigm shift in environmental risk assessment, moving beyond traditional single-hazard approaches to a more holistic understanding of interconnected environmental threats⁷. Compound events are characterized by multiple drivers that interact in complex ways, potentially amplifying their destructive potential^{8,9}. The typology defined by the authors in ref. 6

categorizes compound hazards as either preconditioned, multivariate, temporally compounding, and spatially compounding. Of particular interest to this paper will be preconditioned hazards, which are defined in ref. 6 as an instance where “one or more hazards can cause an impact, or lead to an amplified impact”. With this typology established, a discussion naturally followed about how compound driver play-out in terms of the disaster risk reduction (DRR) cycle and how to incorporate tools into risk assessments^{10–12}.

As the uncertainty around how to refer to compound events and how to incorporate them into existing disaster frameworks decreased, there still existed a lack of quantitative methods developed around real-world applications of these frameworks. It is now possible to see this trend shift on an international scale, especially in the case of flood-related compound events^{13,14}. As the body of research on these topics grow, there are many perspectives from which the quantitative methods can be applied. Whether the impact of a compounding event is measured via its damage to crops^{15,16}, its event complexity¹⁷, or its ability to disrupt energy systems¹⁸, there is a continued need to quantify the damage of these compound events in a manner that is useful to decision makers and emergency managers.

As compound events increase in their frequency and intensity, the overall risk exposure of assets and communities often increased without recognition. Of particular interest to policymakers and infrastructure operators are sequences of compound events and impacts that exceed the

¹Environmental Laboratory, US Army Engineer Research and Development Center, 696 Virginia Rd., Concord, MA, 01742, USA. ²Credere Associates, Westbrook, ME, 04092, USA. ³Center on Climate Change and Urbanism, Tulane Univ., 6823 St. Charles Ave., New Orleans, LA, 70118, USA.

✉ e-mail: Igor.Linkov@usace.army.mil

threshold for designed resilience. The concept of resilience was developed as a way to manage uncertainty, including from the prospects of compounding events. Resilience, in this context, refers to the capacity of a system—be it a community, ecosystem, or infrastructure network—to absorb and recover from disturbances, adapt to changes, and maintain an elasticity for its essential functions and structures^{19,20}. A system's resilience is one that can withstand and recover from extreme events while retaining its core identity and functions in a manner that may be measured by a wide variety of social, economic, environmental, and institutional factors. Systems that may have been considered resilient to individual hazards may find themselves overspecialized and overwhelmed when they face multiple, interacting shocks and stresses.

While extensive literature exists on multi-hazard events, quantifying these effects remains less explored. Many studies use mathematical modeling with fragility curves, damage simulations, graphical networks, and ensemble models to examine different relationships between events^{21–24}. However, these approaches often fall short of calculating the extent to which damages are amplified by preceding hazard events and impacts. This study aims to address these critical gaps by introducing a novel, quantitative framework for analyzing compounding events, risks, and impacts specifically in regard to the impact on infrastructure. This approach provides two key contributions to the field of environmental risk assessment and resilience planning. First, the study establishes an objective mathematical framing for defining and computing impacts of compounding threats. Furthermore, the proposed framework addresses missing and biased data issues²⁵ and over-dispersed infrastructure damages using Negative Binomial Regression to model the impacts of storm events^{26,27} and incorporate prior expert knowledge²⁸. An overview of the methodology is presented in Fig. 1.

This framework is evaluated through a case study of California, which frequently experiences wildfires followed by periods of heavy rainfall, leading to increased risks of post-fire flooding and debris flows. California was split into 6 regions based on the Koppen climate zones to examine regional variation of compounding hazards (Fig. 2). Post-fire flooding would fall under the preconditioned compounding hazard category in the typology developed in ref. 6. The weather and hazard data are from NOAA's National Center for Environmental Information (NCEI) Storm Events database²⁹ and meteo-stat³⁰. To deal with the high number of zero damages in NCEI, our framework uses a hurdle-model^{31,32}. Our framework is transferable to other regions. This framework contributes a standardized basis for identifying and classifying compounding threats across diverse geographic, climatic, and meteorological contexts, which offers the possibility for comparative analyses and prioritization among risk managers and policymakers.

Results

California serves as a case study to show the flexibility, comprehensiveness, and limitations of such a complex undertaking. California's geographic and climatic diversity makes it particularly susceptible to compounding natural hazards. Wildfires have increasingly become a significant hazard, particularly in California, where their frequency and severity have escalated in recent years³³. Furthermore, wildfires serve as a prime example where there is an existing body of literature on their effects^{34,35}, allowing one to incorporate prior knowledge to stabilize inferences. Wildfires profoundly impact soil properties by removing vegetation cover and organic matter and altering soil structure due to heat³⁶. This alteration increases the susceptibility of bare soil to erosion and runoff during precipitation events^{37,38}. Additionally, the heating from wildfires can bake the soil surface, reducing its capacity to absorb water and increasing surface runoff rates^{39–41}.

The relationship between wildfires and severe flooding can also be indirect. Both hazards may share climatic conditions that facilitate their occurrence. Extended periods of drought, marked by high temperatures and low precipitation, create ideal environments for wildfire ignition and spread. These drought conditions can also precondition soil by drying it out, reducing its moisture content, and increasing its susceptibility to hydrophobicity when exposed to high temperatures during a wildfire^{42,43}. Therefore, the compounding effect is not strictly causal but includes

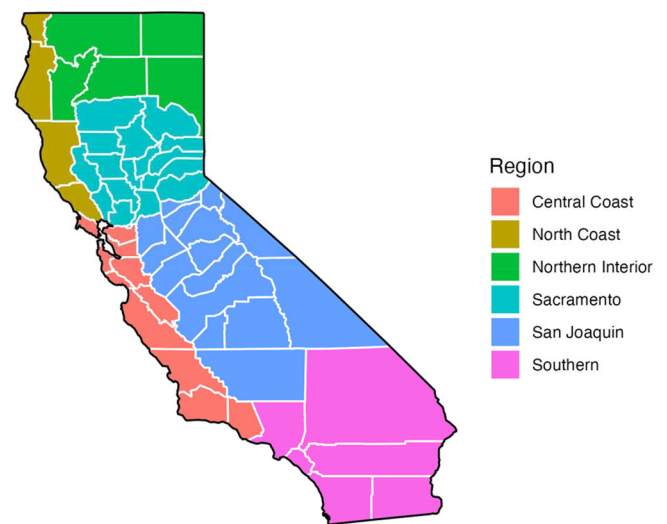


Fig. 2 | Counties mapped to Koppen Climate Zones.

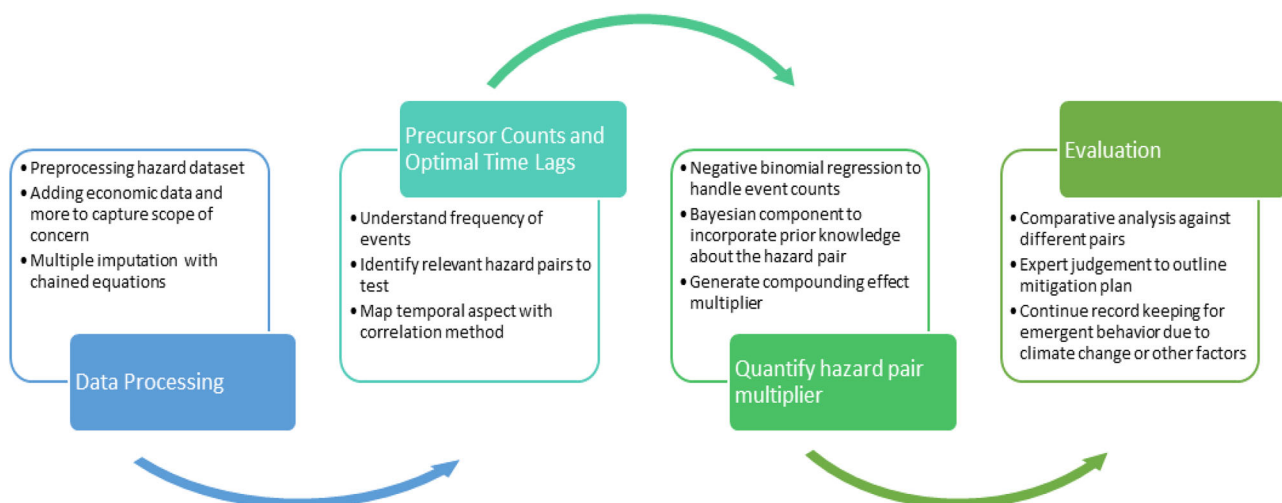


Fig. 1 | The framework from the beginning stages through to the end where policy, mitigation strategies, and decisions are formed.

Fig. 3 | Times lag between wildfires and flooding in California regions. Plots are of the month that maximized the cross correlation function between wildfires and flooding, along with an example of a cross correlation function for San Joaquin. **a** Time Lags for all six regions in CA. **b** An example of the optimal time lag calculation for San Joaquin California.

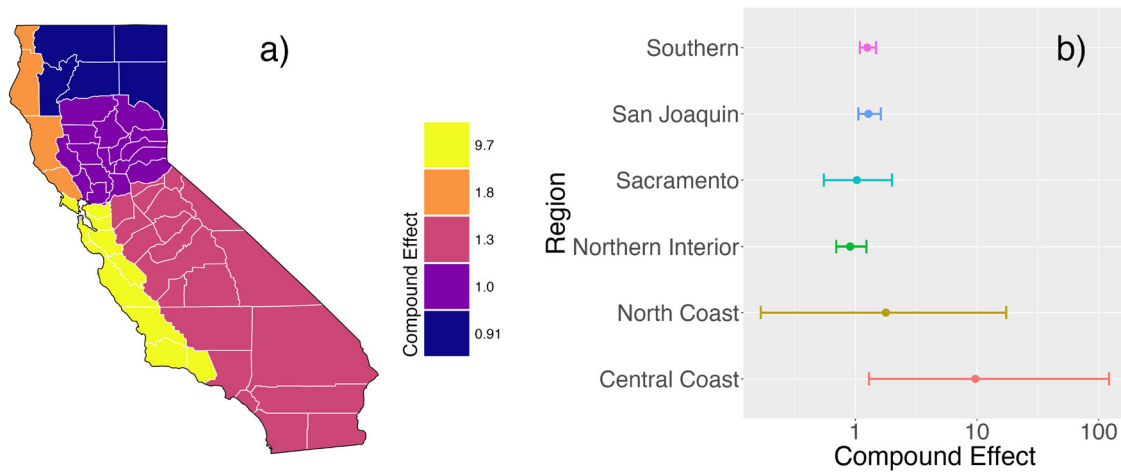
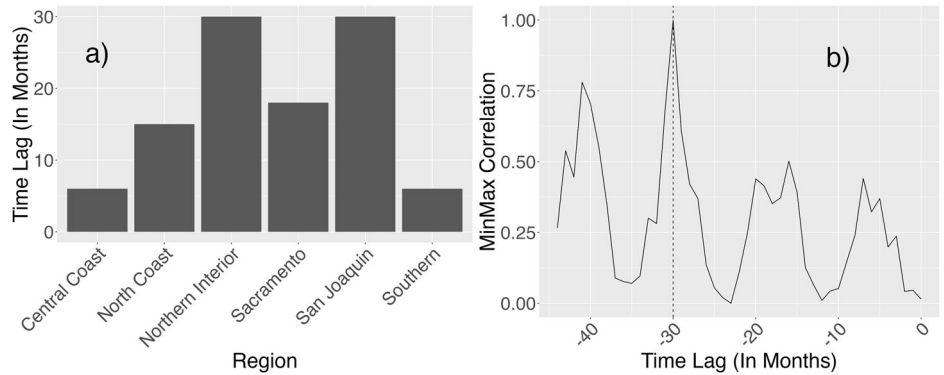


Fig. 4 | Compounding effect of wildfires followed by flooding in California. The map is of the point estimate of the compounding effect and the other displays the 95% credible intervals for the compounding effect. Values greater than 1 represent a compounding effect. **a** Regional differences in the compounding effect that wildfire

occurrence has on the property damage incurred by secondary floods.

b Compounding effect by region, including 95% credible intervals. For each region, we plot the mean and 95% credible interval that were obtained by sampling from the posterior distribution.

correlated possibilities. Our approach examines both potential causal relationships and correlated factors, recognizing the complex interactions between wildfires and flooding in our analysis. Because of these competing phenomena that lead to a single measured impact, it is important to note that the effects we can quantify are not necessarily causal. In other words, our methodology highlights a comparison between hazards that occur after a previous hazard and those that do not. This distinction is crucial for accurate interpretation and subsequent policy decisions.

The optimal time lags between wildfires and floods at a regional scale (Fig. 3a) show variability in the expected time between the two hazard events across the state. These time lags vary between 6 to 30 months, as determined by the maximum of the cross-correlation function methodology. As an example of the cross-correlation function can vary as a function of time lag, we show the results for the Central Coast in Fig. 3b. The correlations are normalized by the maximum correlation in the time series for clarity. The time lags found with our methodology align with estimates from previous studies on the effects of wildfires on flooding³⁴. While preexisting literature can be used to establish the time lags, this study uses a proposed cross-correlation method to demonstrate its applicability to any potential hazard pair whether there is preexisting notions on their temporal relationship.

Figure 4 shows the compounding effect of wildfires on floods in the six Koppen climate regions of California. The 95% credible intervals are included in Fig. 4b to guide the interpretation of these compounding effect values. A compounding effect greater than one suggests a hazard pair that has multiplicatively worse outcomes in terms of property

damage. In several cases, the credible intervals include a value of one, indicating an uncertainty whether the hazard pair truly is compounding. Possible reasons for this result include but are not limited to extremely damaging precursor fires that do not leave much property to be damaged by the ensuing flood events or previously enacted policies that have successfully reduced the impact of floods.

As represented in Fig. 4, there is significant geographical variability in the compounding effects of wildfires on flooding. We note that the Central Coast exhibits the most substantial effect on average. In particular, each additional preceding wildfire resulted in flood property damages increasing multiplicatively by a factor of 9.7 on average. The damages were highly variable for this region, resulting in a wide posterior interval. We note that the coastal regions have the widest posterior intervals, indicating that the compounding effects in this region were highly variable. In the central coast, for example, this variability may be arising because these regions are more prone to extreme flooding conditions, which can lead to a large spectrum of interactions between wildfires and floods. The fatter tails for these regions indicate that while mild and moderate impacts are possible, high-impact compounding effects are more likely in these regions.

The central coast compounding effect was followed by the San Joaquin region. In San Joaquin, each preceding wildfire resulted in a 30% increase to property damages, on average. We also note that San Joaquin region of California had a less variable posterior distribution, yet still a credible interval that excludes 1, indicating a more concentrated compounding effect in this region. So, while a compounding effect in this

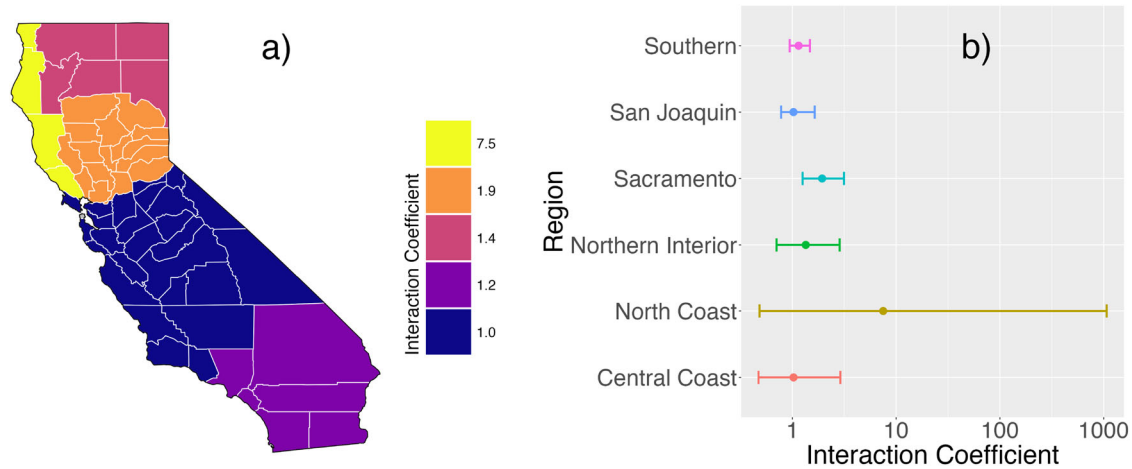


Fig. 5 | Changes over time in the compounding effect of wildfires on flooding in California, as measured by the interaction term coefficient. The map on the left is the point estimate, while the other is the 95% credible interval. Values greater than 1 indicate the compounding effect has increased over time. **a** Regional differences in

the year and precursor count interaction term. **b** Interaction coefficient by region, including 95% credible intervals. For each region, we plot the mean and 95% credible interval that were obtained by sampling from the posterior distribution.

Table 1 | Sample sizes of flood occurrence in the dataset

Region	Sample size
Southern	328
San Joaquin	439
Sacramento	300
Northern Interior	47
North Coast	85
Central Coast	272

region is highly likely, extremely high impact compounding effects are unlikely. These findings underscore the importance of regional analyses and caution against generalizing results across different geographical contexts.

The results for the year and precursor count interaction term (Fig. 5) revealed spatial patterns about how the compounding effects may be changing over time across California. The regression coefficients for the interaction term show that the Sacramento region of California exhibited a larger coefficient for this interaction term, suggesting a stronger temporal trend in the relationship between precursor wildfire events and flood damage in these areas. That being said, there is significant uncertainty in most of the estimates, in particular in the Northern Coast, and so nothing definitive can be said about these regions.

Notably, the Sacramento region of California showed the strongest evidence of a positive effect for the year and precursor wildfire counts interaction term when considering the 95% credible intervals (Fig. 5a). This finding indicates that in these regions, the impact of precursor wildfire events on flood damage has been increasing over time. In other words, for a given number of precursor wildfire events, the associated flood damage in recent years tends to be higher compared to earlier years in the study period. Further research is needed to understand the underlying drivers of these regional differences and temporal trends.

Aside from the distribution and time lag of precursor count variables, it is vital to ensure the model fits the data well. For example, for the Northern Coast the posterior predictive checks were less reliable compared to other regions, suggesting that the model may not adequately capture the complexities of the data here. This lack of fit could be due to the variability in the multiple imputed datasets or small sample size

(Table 1) for this region, also. Therefore, caution should be exercised in interpreting the coefficients for the North Coast. The Southern and San Joaquin regions showed moderate compounding effects of wildfires on floods. The results were more consistent compared to the Central Coast, indicating a more predictable relationship between these hazards in these areas. Additionally, the posterior predictive checks indicate that the model fits well in these regions.

Lastly, we demonstrate our proposed evaluation criteria to determine to what degree a hazard pair is compounding in Table 2. These criteria were developed to include information on the results of our framework and outside factors that may be critical for practitioners who aim to use this for decision making. For our case study of wildfires and floods, we observe that the hazard pair is most compounding, and thus poses the greatest risk, in the coastal regions. The conclusion is supported by the fact that these regions check nearly all the boxes in our evaluation framework, indicating a strong and consistent compounding effect. Decision-makers in these areas, therefore, should be particularly vigilant in preparing for this hazard pair.

Conversely, in a region like Sacramento, the pair does not appear to be compounding. That does not mean a wildfire and flood can never be compounding events, but merely that within our framework they are not a compounding hazard pair, and thus would not be as high of a risk for catastrophic damage as compared to the same hazard pair happening to the coastal regions. With the criteria outlined here, our framework indicates that the while wildfires and floods may pose great individual risks to Sacramento, their combined impact is less significant compared to coastal regions.

In summary, the occurrence of wildfires in certain regions can potentially increase the risk of severe flooding. Leveraging the Bayesian approach allows us to quantify this compounding effect in a rigorous way. Furthermore, our evaluation framework provided a way to assess if the hazard pair of wildfires and flooding was compounding, and to what degree, in a systematic way to assess regional-level risk, empowering stakeholders to make informed decisions and enhance preparedness for compounded hazard scenarios.

Discussion

The findings of our study, underscore the regional variation and uncertainty in nonlinear impacts of multiple, simultaneous environmental stressors. Our study introduces a flexible and threat agnostic framework, Bayesian Hurdle Negative Binomial Regression, to quantify these interactions, emphasizing the importance of how one hazard can amplify the effects of

Table 2 | Application of the evaluation criteria for the wildfire and flooding case study

Positive coefficient	Positive interval	Good model fit	Dense population	Social vulnerability	Region
.	Southern
.	San Joaquin
.	Sacramento
.	Northern Interior
.	North Coast
.	Central Coast

another over time. This approach, which systematically captures the temporal dynamics between sequential weather events through the use of precursor counts and defined time lags provides a critical advancement in risk assessment.

In the case study presented, we applied the proposed framework to the state of California to demonstrate its flexibility and comprehensiveness in analyzing the compounding effects of natural hazards. The results revealed optimal time lags between wildfires and floods across different regions of California, ranging from 6 to 30 months. This suggests that the timing of subsequent flooding in this case can be highly dependent on regional factors such as topography and climate. The cross-correlation method used not only aligns with the study at hand but demonstrates its applicability to other hazard pairs providing a robust tool for future risk assessments. Infrastructure managers across the state of California or broader US, have the ability to be proactive understanding the time until significant damage may occur.

Geographically, the compounding effects of wildfires on floods were most pronounced in the central coast, where the damage variability was high, indicating a broad range of potential impacts. Regions with significant coastal exposure may be more susceptible to extreme flood events following wildfires, likely due to the interplay between coastal hydrology and fire-altered landscape but understanding the potential compounded effect of coastal flooding due to wildfires enables risk managers to build local resilience to what typically may have been a usual flood. In contrast, California's southern region also exhibited notable effects, albeit with less variability suggesting this area might experience more predictable but still significant compounding impacts.

In regions which exhibited the most varied and extreme compounding effects of wildfires on floods, the findings suggest that extreme, catastrophic weather events, whether it was wildfires or floods, alter the distribution for the precursor variable such that there is a higher probability of observing extreme values far from the norm for the hazard that follows, such as a flood. For example, an extreme wildfire season with numerous fires can create conditions that significantly alter the probability and magnitude of subsequent flooding events. These so-called tail events are critical to consider in resilience planning because they can lead to disproportionately large impacts compared to more frequent less severe events. The variability seen through this framework, such as in the central coast's compounding effect, can be used to drive more local resilience policy planning noting the difference in effects across the state of California.

Temporal trends, addressed through the precursor and year interaction variable reveal that in certain regions, particularly Sacramento, the compounding effects on wildfires and floods have intensified over time. This trend could be attributed to factors such as climate changing patterns, urban growth and expansion, and alterations in land-use, which may amplify the risk of compounded hazard events. The increasing interaction between hazards over time points to growing need for adaptive risk management strategies that account for both short and long-term shifts in hazard dynamics. In most regions, there was too much uncertainty in the estimates to draw definitive conclusions.

While this study represents an advancement in the quantification of compounding effects in multi-hazard scenarios, it also highlights several areas for future research and development. First, there is a need for more in-

depth analyses of specific hazard pairs, leveraging more detailed, threat-specific datasets. This could involve integrating high-resolution climate models, remote sensing data, and on-the-ground observations to create more comprehensive and granular risk profiles. Such detailed analyses could reveal nuanced interactions between hazards that may not be apparent at broader scales, further refining our understanding of compounding risks.

Second, future work should explore the use of non-economic variables to measure the impact of environmental hazards. While property damage provides a quantifiable and comparable metric across different hazard types, it does not capture the full spectrum of impacts on communities and ecosystems. Incorporating metrics relating to public health, ecosystem services, social vulnerability, land-use, and post-event economic output could provide a more holistic view of hazard impacts and support more comprehensive resilience strategies. Furthermore, compounding effects on property damage could be compared to other measures such as social vulnerability, to better understand the specific impact these effects have on communities. This multi-dimensional approach to impact assessment aligns with broader sustainable development goals and could help identify synergies and trade-offs between different resilience-building measures.

Third, expanding the framework to include non-natural threats such as cyber, social, and political risks represent an important frontier for comprehensive risk assessment. In our increasingly interconnected world, the boundaries between natural and human-induced hazards are often blurred, and the potential for cascading failures across different systems is high. Developing methodologies that can integrate these diverse risk factors could provide invaluable insights for national security, critical infrastructure protection, and societal resilience planning. This expansion would require interdisciplinary collaboration, bringing together experts from fields such as environmental science, computer science, social sciences, and political science to develop truly integrated risk assessment frameworks.

The implications of this research for policy and disaster risk management are substantial, directly supporting the development of more effective adaptation strategies and improved resilience planning. Our framework provides decision-makers with a nuanced and accurate tool for assessing compound risks, enabling more informed and effective policy formulation. For instance, the ability to quantify the compounding effects of sequential hazards allows for precise risk assessments and the development of targeted mitigation strategies. This is especially crucial for infrastructure planning and urban development, where understanding the potential for cascading failures can inform more resilient design and investment decisions. Through a climate change perspective, with our framework, practitioners will be able to track these trends in an intuitive manner, as we continue to see an increase in the frequency and severity of extreme weather events.

The potential applications of our framework extend beyond traditional disaster risk management to inform a wide range of policy areas. For example, in urban planning, our approach could be used to develop more resilient zoning regulations and building codes that account for the potential compounding effects of multiple hazards. In agriculture, it could inform crop selection and farming practices that are robust to a range of potential climate scenarios, as well as optimal pricing strategies for risk transfer via insurance based on historical and future forecast environmental compounding threats. For public health systems, understanding the

compounding effects of environmental hazards could support more effective emergency response planning and resource allocation.

In conclusion, this study represents a significant step forward in our ability to understand, quantify, and prepare for compounding environmental threats. By providing a flexible, robust, and comprehensive framework for analyzing multi-hazard scenarios, we contribute to the development of more effective and adaptive risk management strategies. As we face the challenges of a changing climate and increasingly complex environmental risks, such innovative approaches will be crucial for building resilient communities and ecosystems.

Methods

Overview

Despite ongoing theoretical discussions exploring the ramifications of compounding threats, there remains a pressing need to quantify the associated risks and their implications for resilience design and management. This quantification is important for several reasons. First, it allows for more accurate and precise risk assessment, which may help communities and policymakers to allocate resources more effectively and to strategically prioritize investments in vulnerable infrastructure. Second, it provides a basis for investments that serve multiple adaptation co-benefits that can address multiple, interacting hazards simultaneously. Finally, it contributes to a more nuanced understanding of system vulnerabilities and thresholds, which reciprocally may inform resilience design and performance standards in infrastructure systems.

As outlined in Fig. 1, the research design is centered on four major steps: (i) data processing; (ii) precursor counts and optimal time lags; (iii) quantification of the hazard pair multiplier; and (iv) evaluation. The Methods section provides a more detailed outline of the following data sets and methodologies. In the data processing step, datasets from various sources were merged to create a set of variables that incorporates hazard event information as well as contextual information about the region of impact and additional weather drivers. NOAA's National Center for Environmental Information (NCEI) Storm Events database²⁹ was used to evaluate hazard impacts on property. This dataset includes an event's location, time, and impact in the form of property damage, as well as several other variables. This event data was used in conjunction with property value data gathered for the state of California to inform the capacity of regions to have damage inflicted upon them. Historical weather variables, such as the amount of precipitation experienced during the event, were included as well, using meteo-stat (v.1.6.8; ref. 30). Preprocessing for this data included aggregations, deduplication, and the handling of missing values. We employed Multiple Imputation by Chained Equations (MICE) to robustly estimate missing values.

In the second step, the study analyzed the frequency of natural hazard events in the dataset. This allowed for a regional understanding of which hazard pairs are co-occurring in time and space. At this step, the analysis incorporated the temporal aspect in the form of time lags, where one can identify the separation in time expected between two hazard events. Understanding where and when event pairs happen begins to address the risk of multi-hazard events but does not quantify the impact these multi-hazard events have on a critical infrastructure system. To transform these frequency and temporal analyses into actionable insights, this analysis utilized the concept 'precursor counts' to tie the spatial and temporal frequency of events into one value. This variable is critical in quantifying the compounding impact of the hazard pair co-occurring in the following step.

The third step in our compounding hazard framework was to calculate the multiplicative effect a precursor condition (first event type) has on a second event type. This calculation is done using the data and analytics calculated in steps 1 and 2 using a Bayesian Hurdle Negative Binomial Regression (BHNBR) model^{31,32}. The coefficient on the precursor counts variable is taken as the multiplicative effect of the impact (i.e., the compounding effect) that the hazard pair has on expected property damages relative to the second event happening in isolation. This study also incorporated an interaction term between the year and precursor counts to

investigate how the relationship between precursor events and a secondary event's damage might change over time, potentially due to factors such as climate change or land-use modifications. Weakly informative priors were incorporated in this model to balance prior beliefs with data influence. At this stage, posterior predictive checking is also used to evaluate model fits, ensuring reliability and validity. With this approach the analysis was also able to provide confidence intervals on the multiplicative compounding factor, giving decision-makers a range from which to expect these compounding effects. This interval is a crucial aspect of the process, in that it guarantees that even extreme events (e.g., "1-in-1000-year floods") are factored into the calculations.

The final step in the research design was a qualitative the evaluation of the previous steps. Key to this stage is the expert judgment needed to assess the model and develop mitigation strategies for hazard pairs that pose risks to the regions being assessed. Because steps 1 and 2 provide situational awareness for when, where, and how damaging a hazard pair can be, the user of this tool will be able to prioritize hazard pairs based on their own criteria. This design proposed an evaluation methodology with a set of criteria including model fit based on the posterior predictive checking, a positive precursor count regression coefficient, population density in the region, social vulnerability of the population, and a positive 95% credible interval. Understanding each of these criteria throughout the framework indicates significant compounding hazards, guiding risk prioritization and resource allocation before the subsequent hazard hits, thereby enhancing a community's resilience for the future events.

The challenges in quantifying compounding threats are manifold. They include the complexity of interactions between different environmental stressors, the spatial and temporal variability of extreme events, and the limitations of historical data in capturing emerging trends and unprecedented scenarios. Moreover, the cascading and nonlinear nature of impacts from compounding threats often defies simple additive models, necessitating more sophisticated analytical approaches. This framework aims to capture much of the complexity involved in the quantification of compounding threats, while leaving room for flexibility and expert input.

Data processing

The main source of data used in this study comes from NOAA's National Center for Environmental Information (NCEI) *Storm Events* database. Previous studies (e.g. refs. 44–46) have used this dataset as a basis for their research into specific natural hazards and their impacts. Although this database has its limitations and biases^{25,47}, it is one of the most comprehensive, open-source resources for historical natural hazards events, with national coverage over a substantial timescale (1996–2022). Once this data was downloaded from the source, we performed several processing steps to clean up the database and to add additional information about drivers and economic variables.

First, weather events were aggregated to the county-level, to ensure a sufficient number of data points and distinct boundaries for analysis. The original data, designated by the National Weather Service's (NWS) public forecast zones (NWSM 10-507), required mapping these zones to counties to achieve consistency in the representation of events. If an event occurred in a zone that intersects with two counties, the damages were allocated between the counties based on the proportion of the zone's area within each county. This approach differs from that taken by SHELDUS⁴⁸, where damages are split evenly between counties, as our method ensures a more accurate representation of the impact based on the affected area. After reallocating the damages to their respective counties, we removed any duplicate entries from the dataset to ensure that each row represented a unique weather event. This step was crucial for maintaining the integrity of our data. Subsequently, property and crop damages were adjusted for inflation, using 2022 as the reference year, to provide a consistent economic perspective across the timescale of the dataset.

Historical weather variables at the time and place of a natural hazard event were included using the meteo-stat library in Python³⁰. The event location was sometimes directly specified in the original data with latitude

and longitude values, which were used to retrieve historical weather data. When exact coordinates were unavailable, they were obtained via the Google Maps API (API, n.d.). If meteo-stat returned null values, the approximate locations from the Google Maps API were adjusted by 0.1 degrees in both latitude and longitude to find a location with available historical data, with the search radius not exceeding 0.5 degrees in either direction. Finally, for events with remaining missing weather values, the nearest weather stations were used as reference locations, determined through latitude and longitude comparisons.

Socioeconomic variables added to this expanding dataset include median housing price, number of houses, population density, and GDP for each county and year. These variables provided additional context for estimating property damage in a given county and year, the primary impact variable in this methodology. For instance, median housing price was used as a normalization factor in statistical modeling, which is crucial for comparing counties with varying damage potentials. Median housing prices for single-family homes in each California county were sourced from ref. 49. County GDP data was obtained from the U.S. Bureau of Economic Analysis⁵⁰, population estimates from California's Department of Finance⁵¹, and the number of housing units from the U.S. Census Bureau⁵². All monetary values were adjusted for inflation to 2022.

In order to ensure there is enough data for regression, the counties of California were grouped into six regions. Counties which are in the same Koppen climate region were grouped together, and the resulting regions are shown in Fig. 2. Sample sizes for flooding for each of these regions is provided in Table 1.

Multiple imputation by chained equations (MICE)

During the pre-processing phase, efforts were made to handle missing data systematically to ensure data integrity and completeness for analysis. Building upon this, multiple imputation by chained equations (MICE) offers a robust statistical approach to address remaining missing data. Introduced by Rubin⁵³, MICE generates multiple imputations to account for the uncertainty surrounding missing values⁵⁴. This method involves creating multiple complete datasets, analyzing each one separately, and then combining the results to produce estimates and confidence intervals that appropriately incorporate the uncertainty associated with missing data⁵⁵.

In particular, we are using the CART algorithm to perform the multiple imputations, due to its ability to capture complex and nonlinear interactions between multiple variables. CART constructs binary decision trees by identifying optimal predictors for tree splits. To impute missing values, we determine the terminal node to which the missing value belongs based on its predictors. Then, we identify observed values within the same terminal node as candidates for imputation. Imputed values are obtained by sampling from this candidate group. For a more detailed explanation, refer to ref. 56 and the citations therein. In this study, 20 multiply imputed datasets were aggregated to enhance the reliability of property damage estimates.

Precursor counts and optimal time lags

Having completed multiple imputation through MICE, addressing missing data concerns, we proceed to investigate and define whether two hazards are compounding. When examining a weather event Y that follows another event X, the *precursor count* refers to the number of occurrences of event X within a specified time period preceding event Y. The underlying assumption is that a higher precursor count—or frequency of event X—may be associated with a more severe manifestation of event Y. This, of course, may not be the case, and the purpose of this paper is to provide a systematic framework to be able to quantify this underlying assumption. In this framework, consider our interest in determining the number of wildfires that precede a specific flood event within a shared spatial area. To calculate the precursor count of wildfires before the flood event, it's essential to define a suitable time interval for identifying preceding wildfire occurrences.

This time interval varies significantly based on the characteristics of the hazards and the region in question. While domain expertise or existing research may offer insights into an appropriate time lag for specific

scenarios, determining this parameter analytically is often necessary, especially in novel or complex situations.

In such situations, we propose the use of cross-correlation functions (CCFs) to identify the optimal time interval for precursor counts. Cross-correlation is a widely used statistical technique that measures the similarity between two time series as a function of the lag between them^{57,58}. By analyzing the time-lagged cross-correlation between the time series of precursor event occurrences and the time series of subsequent event occurrences, we can identify the time lag at which the two series exhibit maximum correlation. This time lag is then used as the interval within which we can define precursor counts.

Let X_t and Y_t represent the counts of two different weather hazards at time t . These counts are recorded at regular intervals (daily, weekly, or monthly) depending on the specific weather hazards being studied.

The CCF $r_{XY}(k)$ between two time series $\{X_t\}$ and $\{Y_t\}$ is defined as:

$$r_{XY}(k) = \frac{\sum_t (X_t - \bar{X})(Y_{t+k} - \bar{Y})}{\sqrt{\sum_t (X_t - \bar{X})^2 \sum_t (Y_{t+k} - \bar{Y})^2}} \quad (1)$$

where k is the lag, \bar{X} is the mean of the X_t series, and \bar{Y} is the mean of the Y_t series. The cross-correlation function measures the similarity between X_t and Y_{t+k} as a function of the lag k .

To find the optimal time lag k^* where the two-time series are most correlated, we compute $r_{XY}(k)$ for a range of lag values and identify the lag that maximizes the cross-correlation function:

$$k^* = \arg \max_k r_{XY}(k) \quad (2)$$

the optimal time lag k^* indicates the timeshift for which the occurrence of weather hazard 1 (represented by X_t) and the occurrence of weather hazard 2 (represented by Y_t) are most highly correlated. This optimal time lag k^* provides valuable insights into the temporal dynamics between the two weather hazards, allowing us to quantify the delay with which the occurrence of one hazard may influence the occurrence of the other.

The optimal time lag identifies the starting point of the window used to calculate the precursor count for weather hazard 1 (e.g., wildfires) in relation to the occurrence of weather hazard 2 (e.g., floods). Specifically, the precursor count represents the number of occurrences of weather hazard 1 within a time window that begins at this optimal time lag and extends up to a defined end point (in our case study, three months prior to the flood event). This approach allows us to capture the most relevant period during which the occurrence of weather hazard 1 may be associated with the subsequent occurrence or intensity of weather hazard 2.

Acknowledging the potential biases highlighted in cross-correlation analysis of time series, it is crucial to clarify that our analysis does not assert the presence of direct correlation or causation between the weather hazards in study⁵⁹. Instead, the identification of the optimal lag allows us to discern potential temporal dependencies between these hazards. This lag serves as a methodological tool to capture and quantify precursor events, thereby enabling a more refined understanding of how the occurrence of one hazard may precede or influence another over time. This approach enhances our ability to later quantify compounding events associated with these hazards more effectively, recognizing that relationships can vary over time and space, so a static lag is not sufficient.

Bayesian hurdle negative binomial regression (BHNBR)

A Bayesian hurdle negative binomial regression (BHNBR) model is used to accurately quantify the compounding effects of sequential weather events on property damage. The negative binomial regression (NBR) model is particularly suitable for our analysis due to the over-dispersion present in the property damage data, a common characteristic of weather event counts and similar observational datasets^{60,61}. Traditional Poisson regression models often fall short in performance in the presence of over-dispersed counts data^{62,63}.

The probability mass function of the Negative Binomial distribution is given by:

$$P_{NB}(Y = y; \mu) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1)\Gamma(\alpha^{-1})} \left(\frac{r^{-1}}{r^{-1} + \mu} \right)^{r^{-1}} \left(\frac{\mu}{r^{-1} + \mu} \right)^y \quad (3)$$

where μ is the mean of the distribution, r is the dispersion parameter, and Y is the property damage count.

Furthermore, the dataset exhibits a significant prevalence of zero values. Therefore, a Hurdle NBR model is employed to effectively manage the challenges posed by this excess of zeros in the property damage data. This model approach is specifically designed for datasets where zero values are disproportionately represented⁶⁴. The Hurdle NBR model consists of two main components: the hurdle component and the positive component. The hurdle component addresses the likelihood of observing a zero versus a non-zero outcome, while the positive component models the non-zero counts using a truncated negative binomial regression approach. This method is akin to the standard negative binomial regression but focuses solely on positive counts.

Let Y_i denote the property damage count for event i . The hurdle model is structured as follows:

$$P(Y_i = y_i) = \begin{cases} p_i & y_i = 0 \\ (1 - p_i) \frac{P_{NB}(Y_i = y_i; \mu_i)}{P_{NB}(Y_i = 0; \mu_i)} & y_i > 0 \end{cases} \quad (4)$$

where p_i that the property damage is zero, μ_i is the mean damages, and P_{NB} is the negative binomial probability mass function defined in Eq. (3).

The primary objective is to associate the covariates of interest with the mean damages, denoted as μ_i . Additionally, we model the probability of observing zero damage, represented by p_i , based on the same covariates. The model is therefore specified as follows:

$$\log(\mu_i) = \beta_0 + \beta_1 \text{Precursor}_i + \beta_2 \text{HomeValue}_i + \beta_3 \text{Duration}_i + \beta_4 t_i + \beta_5 \text{PopDensity}_i + \beta_6 \text{prcp}_i + \beta_7 t_i \cdot \text{Precursor}_i \quad (5)$$

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 \text{Precursor}_i + \alpha_2 \text{HomeValue}_i + \alpha_3 \text{Duration}_i + \alpha_4 t_i + \alpha_5 \text{PopDensity}_i + \alpha_6 \text{prcp}_i + \alpha_7 t_i \cdot \text{Precursor}_i \quad (6)$$

where μ_i is the expected property damage from event i , Precursor_i is the precursor count, HomeValue_i is the total home value in the county event i occurred in, Duration_i is the length of event i , prcp_i is the precipitation that occurred on the day of event i , and PopDensity is the population density in the county event i occurred in.

The key parameter of interest in this analysis is β_1 , the coefficient of the precursor count variable. The exponential of this parameter captures the multiplicative effect of previous weather events on the property damages caused by the subsequent event. A positive β_1 indicates that prior events amplify the damage, while a negative β_1 suggests a mitigating effect. Another parameter of interest is β_7 , the interaction term between the year and precursor counts, which measures how the relationship between precursor events and on a secondary event type's damage might be changing over time. This interaction term is important in estimating how the compounding effects are expected to change over time.

Bayesian framework

In this study, the Bayesian approach extends the Hurdle Negative Binomial Regression model by integrating prior distributions for the parameters. This extension allows for the incorporation of existing knowledge and to quantify uncertainty in a rigorous manner⁶⁵. Bayesian methods are particularly suited for this task because they provide a

framework to systematically include prior information, which is crucial in scenarios where data are limited or noisy. This model was implemented using the brms package in R, leveraging its interface with Stan, a platform for Bayesian modeling. To maintain flexibility and stability in our estimates, we opted for weakly informative priors, as recommended by Gelman et al.⁶⁶. These priors strike a balance between incorporating prior beliefs and allowing the data to influence results significantly. Unlike non informative priors, which can lead to unstable estimates, especially with sparse data⁶⁷, weakly informative priors help mitigate the risk of overfitting and enhance the model's robustness.

Weakly informative priors are used for the covariate coefficients using the Student-t distribution instead of the normal distribution. This decision stems from the anticipated heavy-tailed nature of the effects being modeling. Weather hazards can exhibit highly variable impacts, which are less likely to be tightly concentrated around a mean, particularly when considering the influences of climate change⁶⁸. The Student-t distribution, characterized by its heavier tails, more accurately captures the potential for extreme values and offers a more realistic representation of the uncertainty in the model.

When a positive effect between two hazards is expected, but there is a lack of strong prior evidence, a weakly informative prior with a positive mean $\mu_j > 0$ is used, otherwise $\mu_j = 0$. For example, based on empirical observations, it is expected for heavy rainfall to approximately double the severity of flooding, in which $\mu_j = \ln(2)$ for the prcp_i covariate. In order to incorporate spatial uncertainty into the prior for the Precursor_i covariate, the degree of freedom parameter for the Student-t prior is determined as follows:

$$\beta_j \sim \text{Student-t}(7 - \lambda \cdot 4, \mu_j, 1) \quad (7)$$

where λ is calculated using min-max scaling of the region areas, ensuring that $\lambda = 0$ for the smallest region and $\lambda = 1$ for the largest region. In other words, λ is a measure of the spatial uncertainty around whether or not two hazards occurred in the same region. In total, there are six λ values corresponding to each of the six regions of California outlined in the "Data processing" section.

To compute the λ values, the following steps are used:

- Calculate the mean area of the counties in each region to obtain six average areas.
- Apply min-max scaling to these average areas to obtain λ values between 0 and 1.

The min-max scaling is performed using the following formula:

$$\lambda_i = \frac{\text{Area}_i - \min(\text{Area})}{\max(\text{Area}) - \min(\text{Area})} \quad (8)$$

where Area_i is the mean area of the counties in region i , and $\min(\text{Area})$ and $\max(\text{Area})$ are the minimum and maximum of the average areas across the six regions.

This scaling ensures that the degrees of freedom parameter of the Student-t distribution smoothly interpolates between 7 for the smallest region and 3 for the largest region, appropriately adjusting our prior based on the region size. The average areas for the regions are computed by taking the mean of the county areas within each region and then applying min-max scaling to the vector of these average areas, $\text{Areas} = (\text{Area}_1, \dots, \text{Area}_6)$.

Posterior inference is performed using Markov Chain Monte Carlo (MCMC) methods, with the Metropolis-Hastings algorithm, to sample from the posterior distribution. These samples allow for an estimation of the posterior means, credible intervals, and other summary statistics for the parameters. Since multiple imputation was used, the posteriors obtained from each imputed dataset are combined. In particular, the final posterior is the equally weighted mixture of the posteriors obtained for the imputed datasets. The posterior distributions obtained from the analysis can serve as

valuable priors for future studies. As more data becomes available and more studies are conducted, these posteriors will help refine our understanding of the compounding effects of weather hazards, improving the precision and reliability of future models.

Marginal effects

As previously discussed, the framework relies on the BHNBR coefficient for the non-zero damages, β_1 , to quantify the relationship between two arbitrary weather hazards. Equally important, however, are the marginal effects of the precursor counts. Marginal effects are a crucial concept in regression analysis, particularly in the context of nonlinear models like the Negative Binomial regression. This refers to the change in the predicted outcome (in this case, property damage) when that predictor variable (such as precursor counts) increases by one unit, while all other variables in the model are held constant.

Model checking

Posterior predictive checking (PPC) is an essential step in Bayesian analysis that allows researchers to evaluate the fit of a model by comparing the observed data to simulated data drawn from the model's posterior predictive distribution. This technique helps identify potential discrepancies between the model and the data, providing valuable insights for model refinement and validation⁶⁹.

PPC is based on the concept of generating new datasets, called posterior predictive replicates, from the model's posterior distribution. These replicates are then compared to the observed data to assess the model's performance. The process involves the following steps:

1. Draw posterior samples: Obtain samples from the posterior distribution of the model parameters.
2. Simulate replicates: Using these posterior samples, generate simulated datasets (replicates) from the posterior predictive distribution.
3. Compare to observed data: Compare the replicates to the observed data using various discrepancy measures or test statistics.

PPC offers several advantages for complex model validation^{70,71}. First, PPC provides a means to validate the model by assessing whether the observed data could plausibly have been generated by the model. PPC also aids in identifying areas where the model may not fit well, guiding improvements or refinements to the model. Finally, of the advantages of PPC that are relevant for this study, it also naturally incorporates the uncertainty in the parameter estimates.

By leveraging posterior predictive checking, researchers can gain valuable insights into the model's performance, identify potential areas of concern, and make informed decisions about model refinement or selection. This systematic approach to model evaluation is essential for ensuring the reliability and validity of Bayesian analyses and their subsequent conclusions.

Data availability

The datasets analyzed in this study were derived from: NOAA's National Center for Environmental Information (NCEI) storm events database, meteo-stat (v.1.6.8), California Association of Realtors housing data, U.S. Bureau of Economic Analysis GDP data, California Department of Finance population estimates, and U.S. Census Bureau housing units data. The processed dataset generated during this study are available from the corresponding author upon reasonable request.

Code availability

The code used in this study is available from the corresponding author upon reasonable request.

Received: 6 October 2024; Accepted: 2 April 2025;
Published online: 07 May 2025

References

1. Clarke, B., Otto, F., Stuart-Smith, R. & Harrington, L. Extreme weather impacts of climate change: an attribution perspective. *Environ. Res.: Clim.* **1**, 012001 (2022).
2. AghaKouchak, A. et al. Climate extremes and compound hazards in a warming world. *Annu. Rev. Earth Planet. Sci.* **48**, 519–548 (2020).
3. Zscheischler, J. et al. Future climate risk from compound events. *Nat. Clim. Change* **8**, 469–477 (2018).
4. Wied, M., Oehmen, J. & Welo, T. Conceptualizing resilience in engineering systems: an analysis of the literature. *Syst. Eng.* **23**, 3–13 (2020).
5. Wells, E. M., Boden, M., Tseytlin, I. & Linkov, I. Modeling critical infrastructure resilience under compounding threats: a systematic literature review. *Prog. Disaster Sci.* **15**, 100244 (2022).
6. Zscheischler, J. et al. A typology of compound weather and climate events. *Nat. Rev. Earth Environ.* **1**, 333–347 (2020).
7. Leonard, M. et al. A compound event framework for understanding extreme impacts. *Wiley Interdiscip. Rev.: Clim. Change* **5**, 113–128 (2014).
8. Pescaroli, G. & Alexander, D. Understanding compound, interconnected, interacting, and cascading risks: a holistic framework. *Risk Anal.* **38**, 2245–2257 (2018).
9. Zscheischler, J., Sillmann, J. & Alexander, L. Introduction to the special issue: compound weather and climate events. *Weather Clim. Extremes* **35**, 100381 (2022).
10. van den Hurk, B. J. et al. Consideration of compound drivers and impacts in the disaster risk reduction cycle. *Iscience* <https://doi.org/10.1016/j.isci.2023.106030> (2023).
11. Simpson, N. P. et al. Adaptation to compound climate risks: a systematic global stocktake. *Iscience* <https://doi.org/10.1016/j.isci.2023.105926> (2023).
12. Golzadeh, N. & Alizadeh, M. R. Flood, fires, and drought: do we have the tools to risk-assess for environmental management of extreme climatic events? *Integr. Environ. Assess. Manag.* **20**, 312–313 (2024).
13. Hsiao, S.-C. et al. Flood risk influenced by the compound effect of storm surge and rainfall under climate change for low-lying coastal areas. *Sci. Total Environ.* **764**, 144439 (2021).
14. Wang, Z. et al. Compound coastal flooding in San Francisco Bay under climate change. *npj Nat. Hazards* **2**, 3 (2025).
15. Hamed, R., Van Loon, A. F., Aerts, J. & Coumou, D. Impacts of compound hot-dry extremes on US soybean yields. *Earth Syst. Dyn.* **12**, 1371–1391 (2021).
16. Ben-Ari, T. et al. Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France. *Nat. Commun.* **9**, 1627 (2018).
17. Jiang, S., Tarasova, L., Yu, G. & Zscheischler, J. Compounding effects in flood drivers challenge estimates of extreme river floods. *Sci. Adv.* **10**, ead14005 (2024).
18. Ramenzoni, V. C., Comparini, L., Houlihan, A., Meyer, P. & Brouillard, P. Compounding disaster: perceptions of coastal risk, extreme events, and oil and gas energy production in the gulf of Mexico. *Energy Res. Soc. Sci.* **117**, 103735 (2024).
19. Keenan, J. M., Trump, B., Kytömaa, E., Adlakha-Hutcheon, G. & Linkov, I. The role of science in resilience planning for military-civilian domains in the US and NATO. *Defence Studies* <https://doi.org/10.1080/14702436.2024.2365218> (2024).
20. Linkov, I. & Trump, B. D. *The Science and Practice of Resilience* (Springer, 2019).
21. Fereshtehnejad, E. et al. Probabilistic risk assessment of coupled natural-physical-social systems: cascading impact of hurricane-induced damages to civil infrastructure in Galveston, Texas. *Nat. Hazards Rev.* **22**, 04021013 (2021).

22. Sanderson, D., Kameshwar, S., Rosenheim, N. & Cox, D. Deaggregation of multi-hazard damages, losses, risks, and connectivity: an application to the joint seismic-tsunami hazard at seaside, Oregon. *Nat. Hazards* **109**, 1821–1847 (2021).
23. Sedlmeier, K., Feldmann, H. & Schädler, G. Compound summer temperature and precipitation extremes over central Europe. *Theor. Appl. Climatol.* **131**, 1493–1501 (2018).
24. Bevacqua, E. et al. Advancing research on compound weather and climate events via large ensemble model simulations. *Nat. Commun.* **14**, 2145 (2023).
25. Gall, M., Borden, K. A. & Cutter, S. L. When do losses count? Six fallacies of natural hazards loss data. *Bull. Am. Meteorol. Soc.* **90**, 799–810 (2009).
26. Anastasopoulos, P. C. Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. *Anal. Methods Accid. Res.* **11**, 17–32 (2016).
27. Liu, H., Davidson, R. A., Rosowsky, D. V. & Stedinger, J. R. Negative binomial regression of electric power outages in hurricanes. *J. Infrastruct. Syst.* **11**, 258–267 (2005).
28. Choy, S. L., O’Leary, R. & Mengersen, K. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology* **90**, 265–277 (2009).
29. NOAA’s NCEI N. S. E. D. v3.1. <https://www.ncdc.noaa.gov/stormevents/> (2023).
30. Lamprecht, C. The Weather’s Record Keeper | Meteostat - meteostat.net. meteostat.net (2024).
31. Green, J. A. Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with poisson and negative binomial regression. *Health Psychol. Behav. Med.* **9**, 436–455 (2021).
32. Workie, M. S. & Azene, A. G. Bayesian zero-inflated regression model with application to under-five child mortality. *J. Big Data* **8**, 4 (2021).
33. Williams, A. P. et al. Observed impacts of anthropogenic climate change on wildfire in California. *Earth’s Future* **7**, 892–910 (2019).
34. Yu, G. et al. Process-based quantification of the role of wildfire in shaping flood frequency. *Water Resour. Res.* **59**, e2023WR035013 (2023).
35. Beyene, M. T., Leibowitz, S. G. & Pennino, M. J. Parsing weather variability and wildfire effects on the post-fire changes in daily stream flows: a quantile-based statistical approach and its application. *Water Resour. Res.* **57**, e2020WR028029 (2021).
36. Binkley, D. & Fisher, R. F. *Ecology and Management of Forest Soils* (John Wiley & Sons, 2019).
37. Benavides-Solorio, J. & MacDonald, L. H. Post-fire runoff and erosion from simulated rainfall on small plots, Colorado front range. *Hydrol. Process.* **15**, 2931–2952 (2001).
38. Stoof, C. R. et al. Hydrological response of a small catchment burned by experimental fire. *Hydrol. Earth Syst. Sci.* **16**, 267–285 (2012).
39. DeBano, L. F. The role of fire and soil heating on water repellency in wildland environments: a review. *J. Hydrol.* **231**, 195–206 (2000).
40. Certini, G. Effects of fire on properties of forest soils: a review. *Oecologia* **143**, 1–10 (2005).
41. Moody, J. A., Shakesby, R. A., Robichaud, P. R., Cannon, S. H. & Martin, D. A. Current research issues related to post-wildfire runoff and erosion processes. *Earth-Sci. Rev.* **122**, 10–37 (2013).
42. Richardson, D. et al. Global increase in wildfire potential from compound fire weather and drought. *NPJ Clim. Atmos. Sci.* **5**, 23 (2022).
43. Pausas, J. G. & Keeley, J. E. Wildfires and global change. *Front. Ecol. Environ.* **19**, 387–395 (2021).
44. Shimkus, C. E. et al. Winter storm intensity, hazards, and property losses in the New York tristate area. *Ann. N. Y. Acad. Sci.* **1400**, 65–80 (2017).
45. Young, A. M., Skelly, K. T. & Cordeira, J. M. High-impact hydrologic events and atmospheric rivers in California: an investigation using the NCEI storm events database. *Geophys. Res. Lett.* **44**, 3393–3401 (2017).
46. Ardon-Dryer, K., Gill, T. & Tong, D. When a dust storm is not a dust storm: reliability of dust records from the storm events database and implications for geohealth applications. *GeoHealth* **7**, e2022GH000699 (2023).
47. dos Santos, R. P. Some comments on the reliability of NOAA’s Storm Events Database. SSRN. <https://doi.org/10.2139/ssrn.2799273> (2016).
48. ASU Center for Emergency Management and Homeland Security. The Spatial Hazard Events and Losses Database for the United States, Version 21.0 [Online Database]. <https://cemhs.asu.edu/sheldus> (2023).
49. California Association of Realtors. Historical Housing Data. <https://www.car.org/marketdata/data/housingdata> (2024).
50. U.S. Bureau of Economic Analysis. GDP by County, Metro, and Other Areas. <https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas> (2024).
51. S. of California Department of Finance. Demographic estimates. <https://dof.ca.gov/forecasting/demographics/estimates/> (2025).
52. U.S. Census Bureau. Intercensal estimates of housing units: April 1, 2000 to July 1, 2010. <https://www.census.gov/data/tables/time-series/demo/popest/intercensal-2000-2010-housing-units.html> (2024).
53. Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys* (John Wiley & Sons, 1987).
54. Rubin, D. B. In *Flexible Imputation of Missing Data* 2nd edn. 29–62 (Chapman; Hall/CRC, 2018).
55. Buuren, Svan & Groothuis-Oudshoorn, K. MICE: multivariate imputation by chained equations in r. *J. Stat. Softw.* **45**, 1–67 (2010).
56. van Buuren, S. *Flexible Imputation of Missing Data* (CRC Press, 2011).
57. Podobnik, B., Wang, D., Horvatic, D., Grosse, I. & Stanley, H. E. Time-lag cross-correlations in collective phenomena. *Europhys. Lett.* **90**, 68001 (2010).
58. Shen, C. Analysis of detrended time-lagged cross-correlation between two nonstationary time series. *Phys. Lett. A* **379**, 680–687 (2015).
59. Olden, J. D. & Neff, B. D. Cross-correlation bias in lag analysis of aquatic time series. *Mar. Biol.* **138**, 1063–1070 (2001).
60. Hermans, E., Brijs, T., Stiers, T. & Offermans, C. The impact of weather conditions on road safety investigated on an hourly basis. In *Proceedings of the 85th Annual Meeting of the Transportation Research Board* (2006).
61. Lindén, A. & Mäntyniemi, S. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology* **92**, 1414–1421 (2011).
62. Gardner, W., Mulvey, E. P. & Shaw, E. C. Regression analyses of counts and rates: poisson, overdispersed poisson, and negative binomial models. *Psychol. Bull.* **118**, 392 (1995).
63. Ver Hoef, J. M. & Boveng, P. L. Quasi-poisson vs. Negative binomial regression: how should we model overdispersed count data? *Ecology* **88**, 2766–2772 (2007).
64. Hu, M.-C., Pavlicova, M. & Nunes, E. V. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *Am. J. drug alcohol Abus.* **37**, 367–375 (2011).
65. Mutiso, F. et al. Bayesian negative binomial regression with spatially varying dispersion: modeling COVID-19 incidence in Georgia. *Spat. Stat.* **52**, 100703 (2022).
66. Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y. S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2**, 1360–1383 (2008).
67. Lemoine, N. P. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* **128**, 912–928 (2019).

68. Rummukainen, M. Changes in climate and weather extremes in the 21st century. *Wiley Interdiscip. Rev.: Clim. Change* **3**, 115–129 (2012).
69. Gelman, A., Goegebeur, Y., Tuerlinckx, F. & Van Mechelen, I. Diagnostic checks for discrete data regression models using posterior predictive simulations. *J. R. Stat. Soc.: Ser. C. (Appl. Stat.)* **49**, 247–268 (2000).
70. Gelman, A., Meng, X. L. & Stern, H. *Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies*. 733–760 (Statistica Sinica, 1996).
71. Gelman, A. et al. *Bayesian Data Analysis* (Chapman; Hall/CRC, 2013).

Acknowledgements

This research project was supported by the U.S. Army Engineers Research and Development Center Funding Laboratory Enhancements Across (X) Four Categories (FLEX-4) Program and by the Strategic Environmental Research and Development Program: RC20-C1-1183, Networked Infrastructures under Compound Extremes (NICE). We thank Emily Wells for her substantial contribution to the early development of this project and valuable feedback during its initial stages. We also thank Mitchell Small for his assistance and expertise in shaping the project and revising the manuscript. Last, we thank the anonymous reviewers for their helpful feedback.

Author contributions

S.D. contributed to methodology, formal analysis, writing – original draft, writing – review and editing, and visualization. M.S. contributed to conceptualization, software, data curation, writing – original draft, and writing – review and editing. B.E. contributed to conceptualization, visualization, writing – original draft, writing – review and editing. I.L. contributed to conceptualization, supervision and funding acquisition. B.T. contributed to writing – original draft. J.K. contributed to supervision, writing – original draft, and writing – review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Igor Linkov.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2025