GAO

United States Government Accountability Office

Report to Congressional Addressees

December 2023

TECHNOLOGY ASSESSMENT

# Artificial Intelligence in Natural Hazard Modeling

## Severe Storms, Hurricanes, Floods, and Wildfires

GAO-24-106213

The cover image displays stylized examples of a severe storm, hurricane, flood, and wildfire overlayed by the type of grid used by traditional weather models and a computer microchip in the center (representing artificial intelligence).

Cover sources: Gavin(storm)/elroce and NASA(hurricane)/oobqoo(flood)/mbafai(wildfire)/stock.adobe.com; GAO (graphical elements). | GAO-24-106213

Highlights of GAO-24-106213, a report to congressional addressees

December 2023

## Why GAO did this study

Natural disasters cause on average hundreds of deaths and billions of dollars in damage in the U.S. each year. Forecasting natural disasters relies on computer modeling and is important for preparedness and response, which can in turn save lives and protect property. AI is a powerful tool that can automate processes, rapidly analyze massive data sets, enable modelers to gain new insights, and boost efficiency.

This report on the use of machine learning in natural hazard modeling discusses (1) the emerging and current use of machine learning for modeling severe storms, hurricanes, floods, and wildfires, and the potential benefits of this use; (2) challenges surrounding the use of machine learning; and (3) policy options to address challenges or enhance benefits of the use of machine learning.

GAO reviewed the use of machine learning to model severe storms, hurricanes, floods, and wildfires across development and operational stages; interviewed a range of stakeholder groups, including government, industry, academia, and professional organizations; convened a meeting of experts in conjunction with the National Academies; and reviewed key reports and scientific literature. GAO is identifying policy options in this report (see next page).

View GAO-24-106213. For more information, contact Brian Bothwell at (202) 512-6888 or BothwellB@gao.gov.

TECHNOLOGY ASSESSMENT

# Artificial Intelligence in Natural Hazard Modeling

## Severe Storms, Hurricanes, Floods, and Wildfires

## What GAO found

GAO found that machine learning, a type of artificial intelligence (AI) that uses algorithms to identify patterns in information, is being applied to forecasting models for natural hazards—such as severe storms, hurricanes, floods, and wildfires—that can lead to natural disasters. A few machine learning models are used operationally—in routine forecasting—such as one that may improve the warning time for severe storms. Some uses of machine learning are considered close to operational, while others require years of development and testing.

GAO identified potential benefits of applying machine learning to this field, including:

- Reducing the time required to make forecasts by replacing components of models that are slow and that increase the cost of modeling.
- Increasing model accuracy by more fully exploiting available data, using other data that traditional models cannot, and creating synthetic data to fill gaps.
- Reducing the uncertainty of model output by improving ensemble modeling—the processes of generating combined predictions from numerous models—and making better use of historical data.

**Forecasting natural disasters using machine learning**

Sources (left to right): Gavin(storm)/elroce and NASA(hurricane)/oobqoo(flood)/mbafai(wildfire)/stock.adobe.com. | GAO-24-106213

GAO also identified challenges to the use of machine learning. For example:

- Data limitations hamper the training of machine learning models and can reduce accuracy in some regions, such as rural areas where weather observations are sparse.
- A lack of trust and understanding of the algorithms as well as concerns about bias can make forecasters and other users hesitant to use machine learning models.
- Limited coordination and collaboration create challenges for fully developing some machine learning models. For example, some forecasters told us they lack opportunities to interact with researchers and convey their needs.
- Workforce and resource gaps also create challenges. For example, the upfront costs to develop and run machine learning models are high, and some companies working on these models do not fully understand the data and phenomena they are modeling, according to academic researchers.

GAO identified five policy options that could help address these challenges. These options are intended to inform policymakers, including Congress, federal and state agencies, academic and research institutions, and industry of potential policy implementations. The status quo option illustrates a scenario in which government policymakers take no additional actions beyond current ongoing efforts.

**Policy Options to Help Address Challenges to the Use of Machine Learning in Natural Hazard Modeling**

| Policy Option | Opportunities | Considerations |
|---|---|---|
| **Facilitate improved data collection, sharing, and use** (report p. 37).<br>Government policymakers could expand use of existing observational data and infrastructure to close gaps, expand access to certain data, and (in conjunction with other policymakers) establish guidelines for making data AI-ready. | • Efforts to address gaps in data sets can improve machine learning model performance.<br>• Expanded access to existing data would improve the ability of researchers and groups to develop and test machine learning technologies.<br>• Adopting standards for AI-ready data could reduce resources needed to curate data and facilitate more efficient modeling. | • Expanding observational infrastructure can be expensive and could divert limited resources from other efforts.<br>• Agencies need to weigh the benefits of greater data sharing against any increase in risks related to data security and privacy.<br>• Strict data standards may slow research and innovation if they burden or constrain machine learning researchers. |
| **Expand education and training** (report p. 38).<br>Government policymakers could update education requirements to include machine learning-related coursework and expand learning and support centers, while academic policymakers could adjust physical science curricula to include more machine learning coursework. | • Updating education requirements would better prepare students to use machine learning in government.<br>• More robust education can better prepare both researchers and end users in fields like meteorology and climatology to develop and use machine learning. | • Education and training reforms may need to be repeatedly adjusted, as technological change in this space can be rapid and unpredictable.<br>• Establishing and expanding professional development and training opportunities throughout government may require substantial investment. |
| **Address hiring and retention barriers and certain resource shortfalls** (report p. 39).<br>Government policymakers could address pay scale limitations for positions that include machine learning expertise and work with private sector policymakers to expand the use of public-private partnerships (PPPs). | • Providing workforce incentives to government employees for machine learning development could allow agencies to bolster workforce capacities.<br>• Expanding PPPs might help agencies overcome computational resource shortfalls and help industry draw on government expertise. | • Increasing salary limits for some employees would require agency budget increases or cuts to other budget items.<br>• Expanding the use of PPPs could magnify resource disparities between government and private industry.<br>• PPPs that entail hosting government data on collaborator systems may pose security risks that would need to be considered and addressed. |
| **Take steps to mitigate bias and foster trust in data and machine learning models** (report p. 40).<br>Policymakers could establish efforts to better understand and mitigate various forms of bias, support inclusion of diverse stakeholders for machine learning models, and develop guidelines or best practices for reporting methodological choices. | • Sustained efforts to address bias in data sets can reduce the likelihood of models negatively affecting certain communities.<br>• Acquiring diverse stakeholder perspectives throughout machine learning models' life cycles can help reduce certain types of bias in data and models.<br>• Fostering machine learning model transparency could improve end-user and decision-maker trust. | • Embedding efforts to address bias throughout the model life cycle may increase model costs and slow model development. |
| **Maintain status quo efforts** (report p. 41).<br>Government policymakers could maintain existing policy efforts and organizational structures, along with existing strategic plans and agency commitments. | • Some agency efforts are underway to address the challenges described. | • The extent to which agencies will meet their commitments under status quo efforts is unclear. Status quo efforts may not fully address challenges specific to the use of AI in natural hazard modeling. |

Source: GAO. | GAO-24-106213

# Table of Contents

## Tables



## Figures

# Abbreviations

| | |
|---|---|
| AI | artificial intelligence |
| AI-RI | Artificial Intelligence – Rapid Intensification |
| CPU | Central Processing Unit |
| CRADA | Cooperative Research and Development Agreements |
| FEMA | Federal Emergency Management Agency |
| FORTRAN | Formula Translation |
| GAN | Generative Adversarial Network |
| GSA | General Services Administration |
| GPU | Graphical Processing Unit |
| IIJA | Infrastructure Investment and Jobs Act |
| NASA | National Aeronautics and Space Administration |
| NIFC | National Interagency Fire Center |
| NIST | National Institute of Standards and Technology |
| NOAA | National Oceanic and Atmospheric Administration |
| NSF | National Science Foundation |
| NWS | National Weather Service |
| OPM | Office of Personnel Management |
| ORF | Operations, Research, and Facilities |
| PAC | Procurement, Acquisition, and Construction |
| QA/QC | quality assurance/quality control |
| SHIPS | Statistical Hurricane Intensity Prediction Scheme |
| USACE | U.S. Army Corps of Engineers |
| USFS | U.S. Forest Service |
| USGS | U.S. Geological Survey |

U.S. GOVERNMENT ACCOUNTABILITY OFFICE
441 G St. N.W.
Washington, DC 20548

December 14, 2023

Congressional Addressees

Since 1980, the U.S. has sustained 363 weather disasters that caused damages of $1 billion or more.[1] These disasters caused nearly 16,000 deaths and totaled an estimated $2.6 trillion in damages.[2] In 2022, there were 18 such "billion-dollar disasters" in the U.S. More than 80 percent of the damage between 1980 and 2023 was caused by four types of weather-related disasters: severe storms, hurricanes, floods, and wildfires.[3]

Forecasting natural disasters is important for preparedness and response, which can in turn save lives and protect property. The World Meteorological Organization estimates the global socioeconomic benefits of weather forecasting to be at least $158 billion per year.[4]

Computer modeling is a critical tool for natural disaster forecasting. Models are representations of real-world systems and can be used to approximate the real system's behavior. For severe weather systems, computer modeling allows forecasters to predict the system's behavior, such as the likely intensity of a hurricane or the location of flooding, and relay those predictions to the public and to emergency responders.

Artificial intelligence (AI) is a powerful tool that can automate processes, rapidly analyze massive data sets, enable modelers to gain new insights, and boost efficiency. It has already started to transform some sectors, such as finance and manufacturing, and has the potential to do so for many others. In natural disaster forecasting, modelers are beginning to use AI to efficiently harness more of the information that is available about the environment to significantly improve the timeliness, accuracy, and precision of natural disaster forecasting. They are doing this in a range of ways, from running AI models along with traditional models, to incorporating AI into traditional models, to fully replacing traditional models with AI models. However, these efforts are largely in the research and development phases and continued testing in operational settings will be needed to further mature the field. Proper safeguards will likewise be important

[1]Dollar amounts in this paragraph are in 2023 dollars.

[2]2023 damage statistics were partial due to the timing of this publication and are through August 8, 2023.

[3]The National Oceanic and Atmospheric Administration distinguishes hurricanes (tropical cyclones with wind speeds in excess of 74 miles per hour) from severe storms (winds in excess of 58 miles per hour or at least ¾-inch diameter hail). Multiple disasters may happen concurrently. For example, flooding may occur during a hurricane.

[4]Socioeconomic benefits include cost savings associated with disaster management, construction, water supply, transportation, agriculture, and energy. Systematic Observations Financing Facility, *Information Brief: The Value of Surface-Based Meteorological Observation Data: Costs and Benefits of the Global Basic Observing Network* (Geneva: World Meteorological Organization, 2020).

to minimize the risk of conveying inaccurate or otherwise harmful information, which can cost lives and property.

In light of broad congressional interest in forecasting natural disasters, we prepared this report under the authority of the Comptroller General to assist Congress with its oversight responsibilities. In the context of AI for modeling severe storms, hurricanes, floods, and wildfires, this report discusses (1) the emerging and current use of AI and potential benefits of this use, (2) challenges concerning the use of AI, and (3) policy options to address challenges or enhance benefits of the use of AI.

We collected evidence for this report from peer-reviewed articles and other reports, stakeholder interviews, and a meeting of experts we convened with the assistance of the National Academies of Sciences, Engineering, and Medicine. Our scope is limited to models that incorporate or use AI techniques (specifically machine learning) to enhance model input, output, or performance, or to predict the aforementioned hazards. We primarily focused on modeling for civilian use. We did not examine risk models, seasonal forecasting, long-term climate modeling, or hazard mitigation efforts. See appendix I for a full discussion of the objectives, scope, and methodology and appendix II for a list of experts who participated in our meeting.

We conducted our work from August 2022 through December 2023 in accordance with all sections of GAO's Quality Assurance Framework that are relevant to technology assessments. The framework requires that we plan and perform the engagement to obtain sufficient and appropriate evidence to meet our stated objectives and to discuss any limitations to our work. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

# 1 Background

## 1.1 Modeling for natural hazard forecasts

A *natural hazard* is an environmental phenomenon that poses a potential harm to society and the human environment. A *natural disaster* is the negative impact following an actual occurrence of a natural hazard in the event that it significantly harms a community. This report examines four hazards that can lead to billion-dollar disasters: severe storms, hurricanes, floods, and wildfires. Table 1 shows which federal agencies are primarily responsible for modeling these hazards for civilian applications.

**Table 1: Key federal agencies involved in modeling for natural hazard forecasts**

| Department/Agency | Description of role |
| --- | --- |
| Department of Commerce/National Oceanic and Atmospheric Administration (NOAA) | NOAA is the primary federal entity responsible for modeling and forecasting severe storms and hurricanes. NOAA provides forecasts for flooding driven by rain, as well as for flooding of rivers and coasts. NOAA supports wildfire modeling with fire weather forecasts and monitoring for potential wildfires. |
| Department of the Interior (DOI)/U.S. Geological Survey (USGS) | USGS's Natural Hazards Mission Area includes programs supporting emergency management, including flood and wildfire forecasting and response. |
| Department of Homeland Security/Federal Emergency Management Agency (FEMA) | FEMA leads coordination within the federal government to help communities prepare for and respond to disasters. Severe storm, hurricane, flood, and wildfire modeling is an essential component of FEMA's mission. FEMA relies on other federal, state, local, and tribal entities to lead some modeling efforts. |
| National Aeronautics and Space Administration (NASA) | NASA conducts monitoring and modeling of severe storms, hurricanes, floods, and wildfires. |
| Department of Defense (DOD)/U.S. Army Corps of Engineers (USACE) | The USACE uses various types of flood models to support efforts, such as emergency management, infrastructure development, and evacuation planning. They also employ a hurricane storm surge prediction model that uses forecast data from NOAA's National Hurricane Center to help forecast coastal flooding. |
| Department of Agriculture/U.S. Forest Service | The Forest Service manages all national wildland fires with the support of other federal, state, local, and tribal entities. For example, the National Predictive Services Oversight Group—an interagency group that includes the Forest Service, DOI, DOD, and others – provides fire weather and climate forecasts, combustible material (i.e., fuel) and fire danger models, and more. |

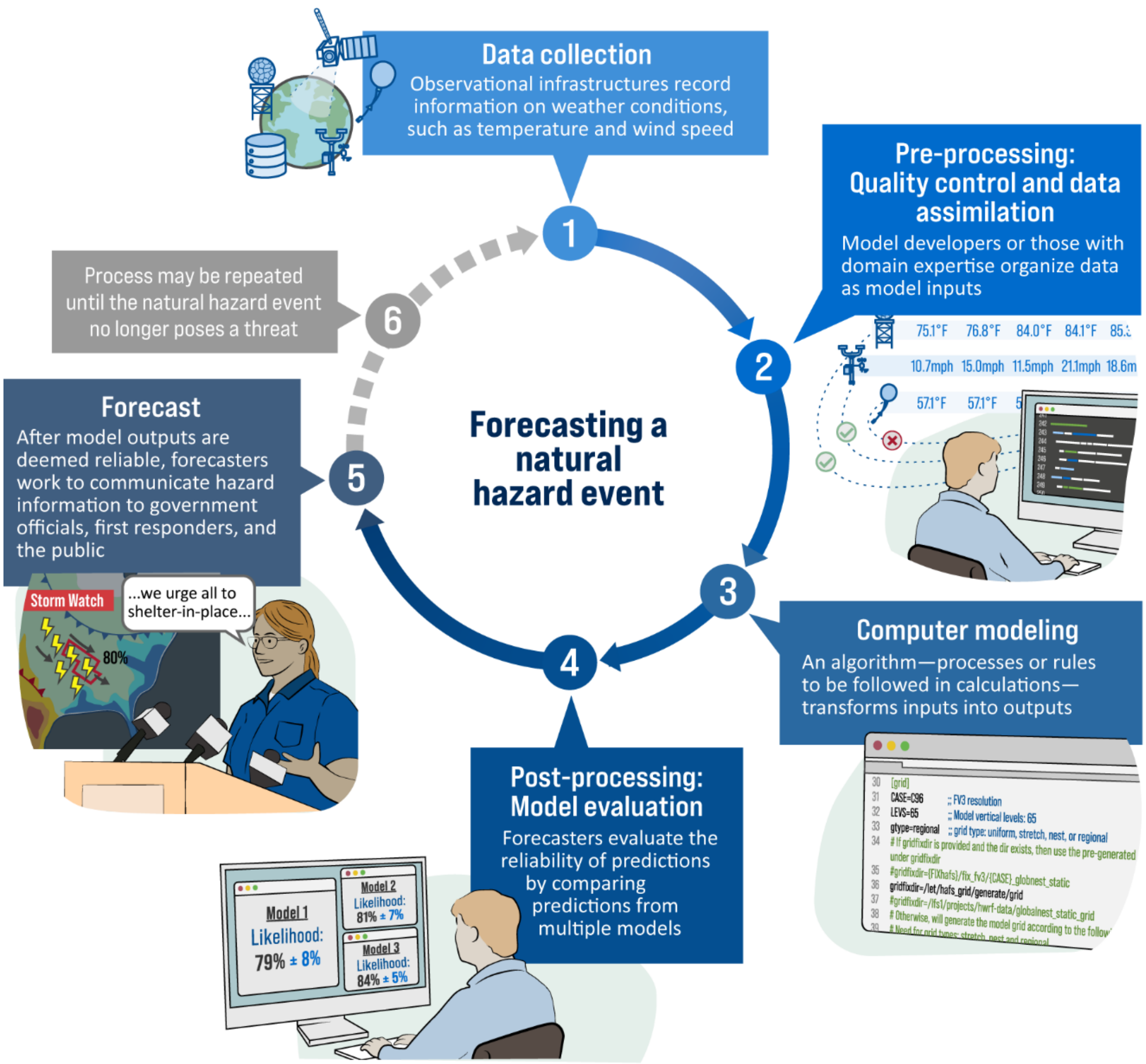
Source: GAO analysis of agency documents. | GAO-24-106213

Note: Includes key agencies responsible for modeling or forecasting severe storms, hurricanes, floods, and wildfires primarily for civilian applications.

Forecasters have used computer models for weather prediction since at least the 1960s. Models for flooding and wildfires have also been in use for decades. The basic function of a model is to use data from current conditions and generate useful predictions. For example, one inland flood model takes data inputs like precipitation, temperature, and relative humidity and predicts future water flow in a river or stream. A different flood model may use sea level rise as an input for predicting storm surge flooding. Traditionally, these models are based on scientists' understanding of the physical laws governing weather patterns and natural hazard event behaviors in the real world.

Forecasting is a multi-step process that includes modeling, which is represented in figure 1.

**Figure 1: Basic steps of forecasting a natural hazard event**




Source: GAO (analysis and illustrations). | GAO-24-106213

Note: Observations from step 2 to step 5 can be used directly to improve the interpretation of the forecast.

To start, data must be collected about real-world weather and environmental conditions. These data can come from satellites, field surveys, volunteer observers, thermometers, flood gauges, ocean buoys, weather towers, and other sources—collectively called *observational infrastructure.* Next, model developers, who may have domain expertise (e.g., atmospheric science, forestry, hydrology, and other fields), select the most relevant environmental data from past and current conditions. In doing so, they attempt to ensure a complete and representative data set to minimize *bias* (i.e., systematic errors due to a variety of causes). The completeness of the data set, among other factors, can help to increase the performance of the output of the model and decrease different biases that might influence the model's output.

The modelers then *assimilate* (combine) these data into specialized computer software and perform quality assurance and quality control (QA/QC) steps to verify the accuracy of data inputs. These and other preparatory steps are called *pre-processing.* The software then uses *algorithms*—quantitative rules programmed by a modeler, at least initially—to perform calculations. Modelers perform *post-processing* on the computer output, often corroborating model results by comparing the predictions of several models or of the same model run several times with different starting conditions (called *ensemble modeling*). The result is a useable output such as a flood map, the probability of a tornado, or the likely extent of wildfire spread. Finally, if the hazard is likely to pose a threat to life or property, forecasters communicate warnings based on these predictions to the public, government officials, and emergency responders to help inform on-the-ground actions. For example, a forecaster may use model results to issue a hurricane warning to a county, which can prompt residents to evacuate and government officials to prepare search-and-rescue personnel.

In the following vignettes, we provide background on the modeling of severe storms, hurricanes, floods, and wildfires.

# Severe Storms

Severe storms, which include severe winds, tornadoes, or hailstorms, cause an average of **48 deaths and $10.2 billion in damages** in the U.S. each year, according to NOAA.[5] The National Weather Service (NWS), part of NOAA, issues more than **15,000 severe storm and tornado watches and warnings** each year.

Severe storms can also disturb ecosystems, for example by knocking down large swaths of forest. Models and observations are essential to predicting storm location and intensity, as well as the likelihood of hail and strong winds.

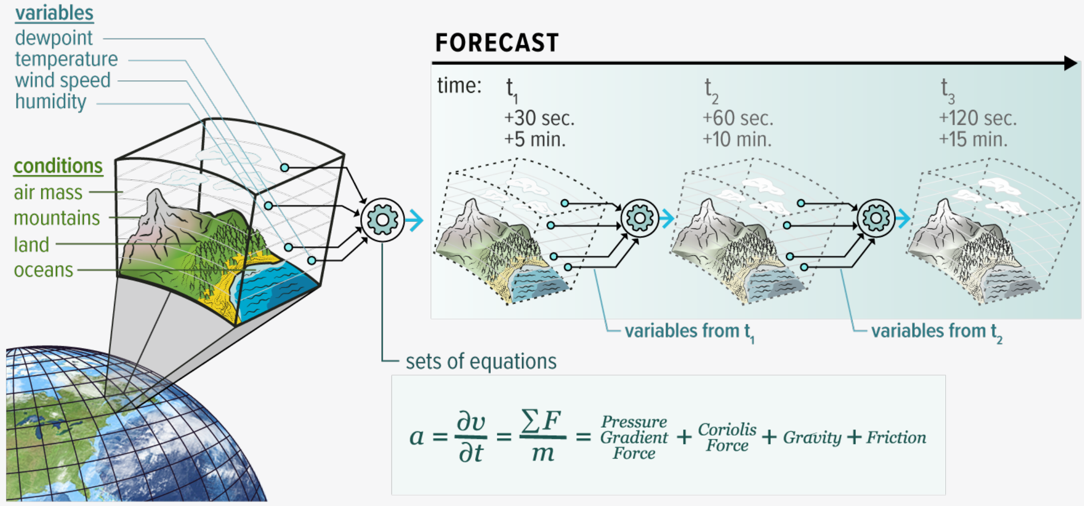
Source: Gavin/stock.adobe.com. | GAO-24-106213

## Modeling

The primary method for weather modeling is known as Numerical Weather Prediction (NWP). Modeling and forecasting of severe storms and weather in general is the responsibility of the NWS. The NWS uses several models that rely on NWP to predict the weather. These models range in their focus from short-term predictions for the U.S., to longer-term predictions for the entire globe.

NWP models the atmosphere by mapping current observations of temperature, moisture, and other variables onto a three-dimensional grid that covers the globe (see figure). It then uses a set of equations and estimates based on fundamental laws of physics to forecast how those variables will change in the next time step of the model (which can be anywhere from a few seconds to several minutes). By repeating these calculations many times over, the model generates a forecast that extends over hours or days and can be updated as new observations come in.

Example of how numerical weather prediction models the atmosphere and generates a forecast




Source: GAO (analysis and illustrations); chones/stock.adobe.com (earth). | GAO-24-106213

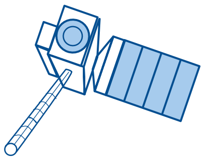
[5]These statistics represent billion-dollar events from 1980 through 2023 and are not representative of all severe storms that occur throughout the year across the U.S. See National Oceanic and Atmospheric Administration National Centers for Environmental Information, *U.S. Billion-Dollar Weather and Climate Disasters Summary Stats* (2023), https://www.ncei.noaa.gov/access/billions/summary-stats.

Source: Unless specified, all graphics on this page were created by GAO.

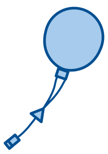
Source: Gavin/stock.adobe.com. | GAO-24-106213

## Data sources

The data underlying severe storm models come from an array of sensors and other sources, including:

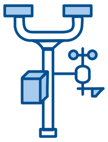


At least 17 different **satellites**, in polar orbits (which observe the same location twice a day) and geostationary orbits (which are fixed over one location and take many images per day), provide data on temperature and humidity variations, clouds, aerosols, winds, and other parameters.

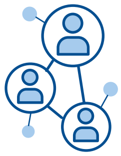


**Weather balloons** launched twice a day at almost 100 stations across North America, the Pacific, and the Caribbean are the primary source of data on upper atmosphere air pressure, temperature, relative humidity, wind speed and wind direction.

Over 150 surface **radar stations** provide data on precipitation, the rotation of thunderstorm clouds, airborne tornado debris, and wind strength and direction.



Thousands of surface **weather stations** provide data on precipitation, temperature, and wind up to 12 times an hour across the U.S.

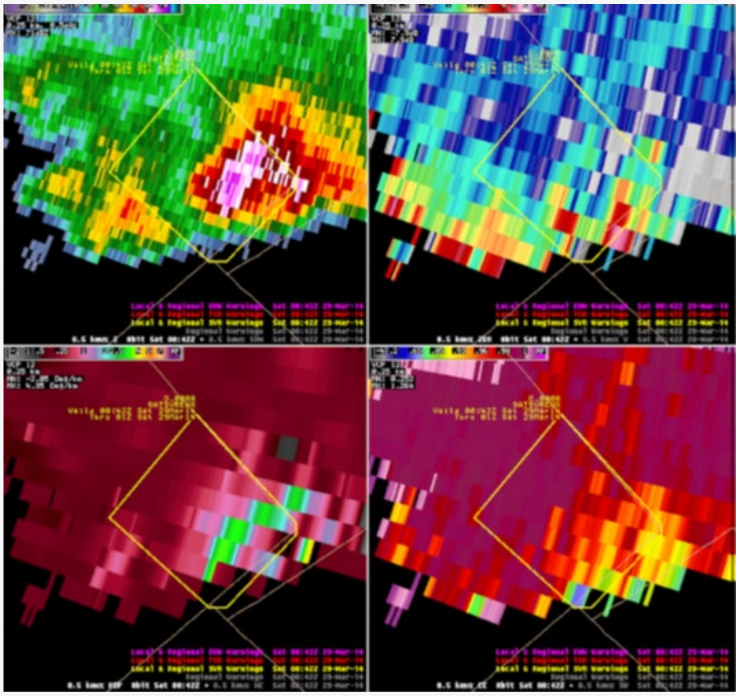


Networks of thousands of **volunteer observers** send firsthand observations that provide additional temperature, snowfall, and rainfall data.

Observations from **commercial aircraft** on temperature, wind, and humidity.

Source: Unless specified, all graphics on this page were created by GAO.

Example of a severe storm warning generated within an Advanced Weather Interactive Processing System



Source: NOAA. | GAO-24-106213

## Limitations

The accuracy of NWP short-term forecasts depends heavily on how well recent observations are assimilated into a model. New storms are often poorly represented because observations arrive too late or are not available to be fed into a model in a timely fashion. It is also difficult for models using traditional NWP to make predictions at a small scale for a large area, such as the behavior of an individual thunderstorm anywhere in the U.S.

# Hurricanes

A hurricane, also known as a tropical cyclone or typhoon, is a rotating cloud and thunderstorm system that forms over tropical waters. Hurricanes form from a combination of persistent storms, warm oceans, and moisture, and can produce violent winds, large waves, and torrential rains. Hurricanes **cost an average of $31.2 billion and cause 157 deaths** in the U.S. per year, according to NOAA.[6] They can also negatively affect ecosystems and seafood safety.

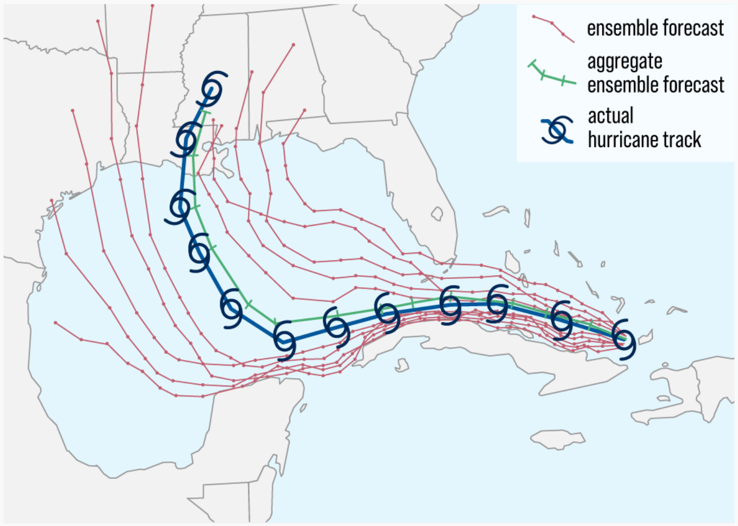
Hurricanes are projected by NOAA experts to become more intense due to changing climate patterns, and modeling these events is vital for mitigation strategies such as issuing evacuation alerts.

Source: elroce and NASA/stock.adobe.com. | GAO-24-106213

## Modeling

Hurricane forecast models use mathematical equations to predict the future behavior of a hurricane, including its path and intensity. Numerical weather prediction (NWP) is a widely used method for weather modeling. It analyzes current atmospheric observations (wind, pressure, temperature, and moisture) and, for hurricanes, uses physics to predict how the wind speed and direction in and around a hurricane will change over time.

Example of ensemble model forecasts of a hurricane




Source: GAO analysis of ensemble modeling for hurricanes; GAO (illustration). | GAO-24-106213

The primary method for estimating hurricane intensity is the Dvorak technique. This method uses automated processes to recognize cloud patterns in satellite data. Forecasters also use models that rely on historical relationships within hurricane data, such as behavior, location, and time of year, to provide a quick, statistical prediction. These models can be used alongside and integrated with NWP models to improve accuracy. Another technique is to use ensemble models, which either combine forecasts from several different models into a single forecast or use one single model several times with different parameters to make several forecasts. Ensemble model forecasts are generally more accurate because they combine several approaches, and the individual model errors get canceled out.

Forecasters use different models to predict a hurricane's impacts, such as storm surge, because hurricane forecast models are not suited to this purpose.

[6]These statistics represent billion-dollar events from 1980 through 2023 and are not representative of all hurricanes that occur throughout the year across the U.S. See NOAA National Centers for Environmental Information, *U.S. Billion-Dollar Weather and Climate Disasters* (2023), https://www.ncei.noaa.gov/access/billions/summary-stats.

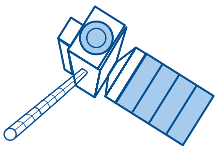
Source: Unless specified, all graphics on this page were created by GAO.

VIGNETTE
Traditional Modeling
# Hurricanes

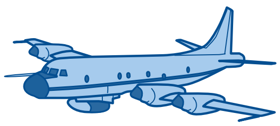
Source: elroce and NASA/stock.adobe.com. | GAO-24-106213

## Data sources

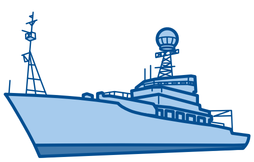
Hurricane specialists at NOAA use observational infrastructure to forecast, observe, and track hurricanes:



**Satellites**, such as NOAA's Geostationary Operational Environmental Satellites, are fixed over one part of the Earth to continuously observe hurricanes. NOAA also has low-orbiting satellites that orbit the Earth typically about every 90 minutes and can help determine a storm's structure.



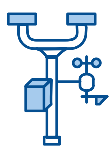
**Reconnaissance aircraft** operated by the U.S. Air Force Reserve and NOAA's Aircraft Operations Center collect data by flying through storms.



**Ships** routinely collect weather observations, such as air pressure, air temperature, sea surface temperature, and wind.

**Buoys** provide standard, hourly information about waves, winds, and the temperature of the air and sea surface.

**Radar** measure the concentration of precipitation in the air by transmitting pulses of energy and measuring the amount reflected.



**Other monitoring systems**, such as land-based monitoring stations, capture important data needed for hurricane forecasts.

Source: Unless specified, all graphics on this page were created by GAO.

Example of a ship used to collect weather observations, NOAA's Ronald H. Brown

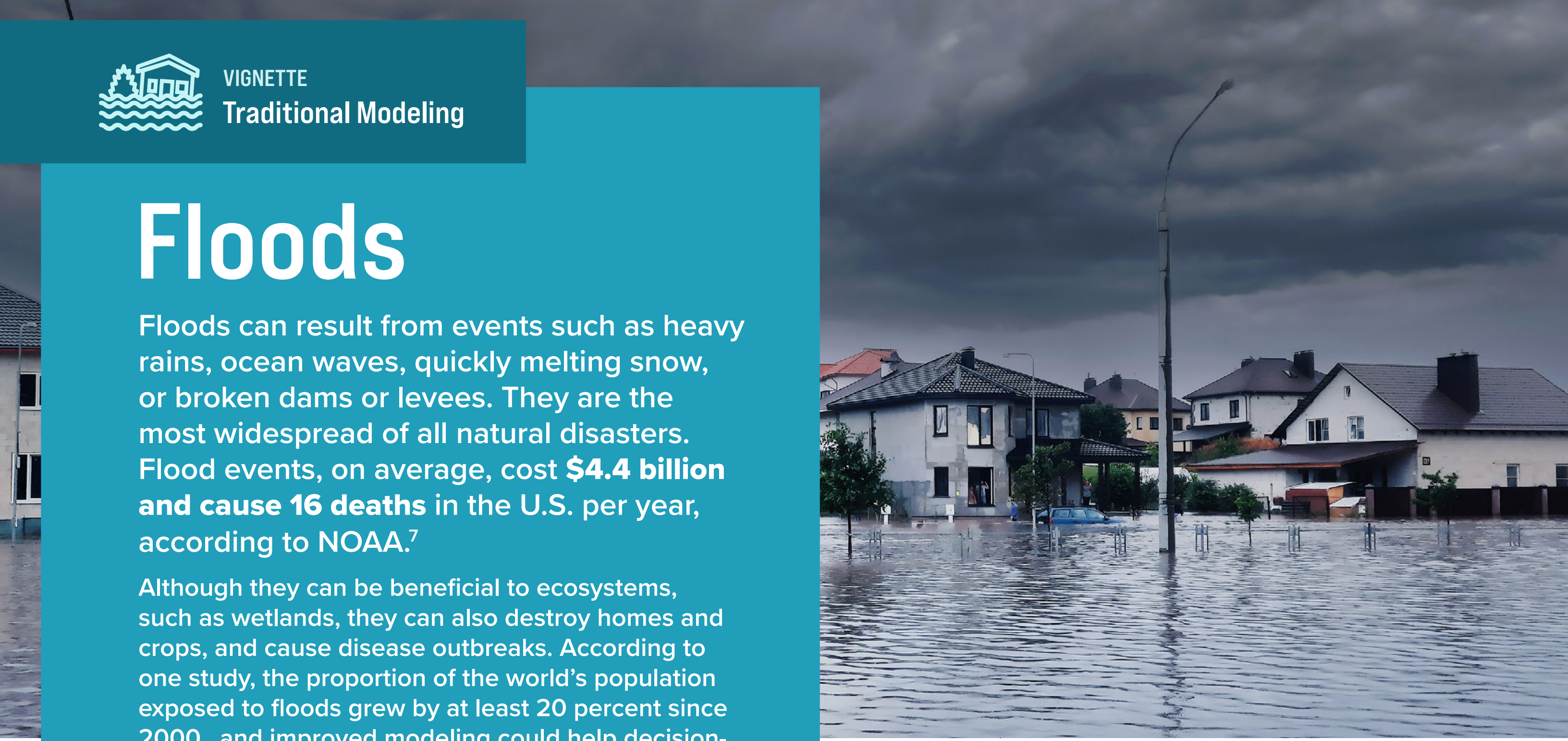

Source: Joseph Creamer/stock.adobe.com. | GAO-24-106213

## Limitations

There is uncertainty in hurricane predictions due in part to changing atmospheric conditions. Once a hurricane has formed, scientists can forecast its potential track, and the error in its prediction gets smaller as the hurricane is closer to making landfall. Predicting hurricane intensity, however, is more challenging. For example, a strong and destructive Atlantic hurricane in 2019, initially predicted to be a modest storm, rapidly intensified to a category 5 hurricane and left little time to prepare for its impacts.

Ensemble models for hurricanes also have limitations. For example, their forecasts typically have relatively low resolution and do not include the finer-scale details found in individual model forecasts.

VIGNETTE
Traditional Modeling



# Floods

**Floods can result from events such as heavy rains, ocean waves, quickly melting snow, or broken dams or levees. They are the most widespread of all natural disasters. Flood events, on average, cost $4.4 billion and cause 16 deaths in the U.S. per year, according to NOAA.[7]**

**Although they can be beneficial to ecosystems, such as wetlands, they can also destroy homes and crops, and cause disease outbreaks. According to one study, the proportion of the world's population exposed to floods grew by at least 20 percent since 2000, and improved modeling could help decision-makers better prepare and respond. Floods can occur within minutes or over a long period, and may last days, weeks, or longer.**

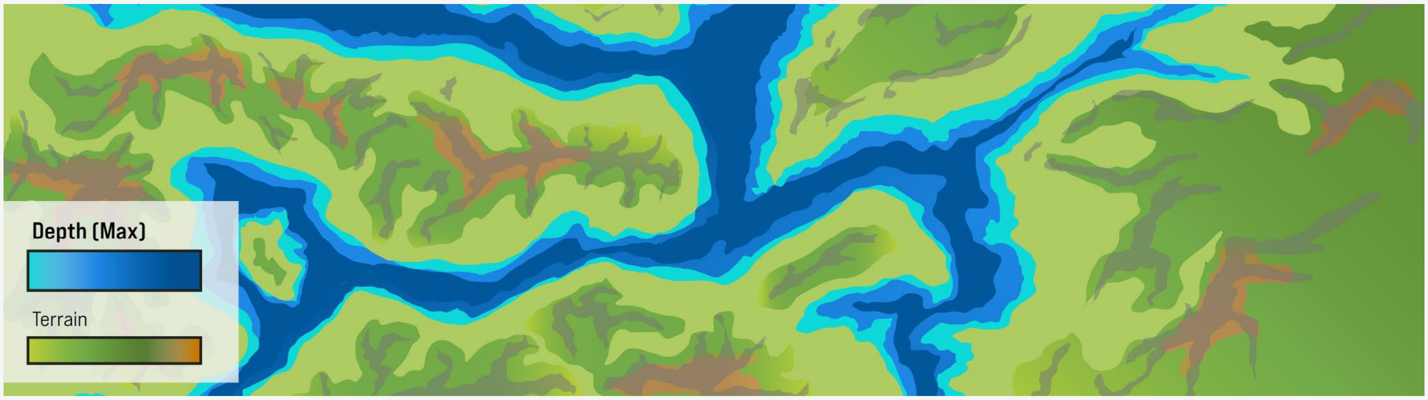
Source: oobqoo/stock.adobe.com. | GAO-24-106213

## Modeling

Federal, state, and local government agencies, as well as those in academia and industry, use flood models and tools for forecasting, disaster preparedness and response, risk management, insurance services, and other purposes. A widely used flood modeling software application is the U.S. Army Corps of Engineers' Hydraulic Engineering Center River Analysis System (HEC-RAS). It simulates floods using equations that represent fundamental principles such as conservation of mass and momentum to predict the flow characteristics of rivers, floodplains, and other areas. Users can build different kinds of models for their specific circumstances by inputting or creating data on terrain, soil, and more, and receiving detailed maps and graphs in return.

For example, once a model has run, the system allows the user to create visual representations of the model results, such as an inundation map. The HEC-RAS system includes a variety of one-dimensional and two-dimensional models, including shallow water, steady and unsteady flow, water quality analyses, and sediment transport.

Example of an inundation map




Source: GAO (illustration). | GAO-24-106213

[7]These statistics represent billion-dollar events from 1980 through 2023 and are not representative of all floods that occur throughout the year across the U.S. See NOAA National Centers for Environmental Information, *U.S. Billion-Dollar Weather and Climate Disasters* (2023), https://www.ncei.noaa.gov/access/billions/summary-stats.

[8]Beth Tellman, et al., "Satellite Imaging Reveals Increased Proportion of Population Exposed to Floods," *Nature*, vol. 596 (2021): 80–86. https://doi.org/10.1038/s41586-021-03695-w.

Source: Unless specified, all graphics on this page were created by GAO.

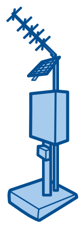
Source: oobqoo/stock.adobe.com. | GAO-24-106213

## Data sources

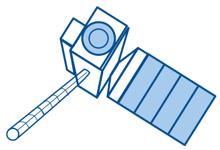
There are 4 main data sources used for flood detection and forecasting:

**Databases**, such as the gridded Soil Survey Geographic Database, which has detailed soil geographic data, provide various input data for flood models.

**Radar** can estimate the duration of rainfall, helping forecasters assess the threat of a flood.

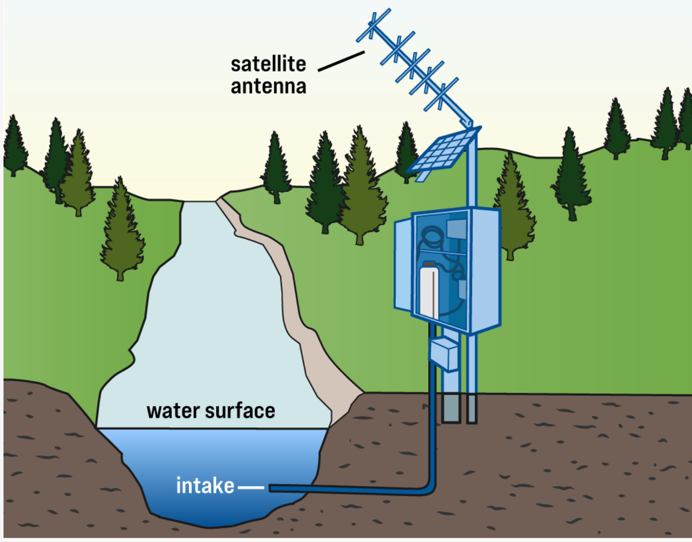


**Water gauges** measure the depth of water. For example, a stream gauge contains instruments that measure and record the amount of water flowing. Generally, these measurements occur automatically every 15 minutes, or more frequently during flooding. The USGS operates a network of more than 9,000 stream gauges nationwide.



**Satellites** produce less accurate estimates of rainfall than water gauges. They provide high-resolution coverage over oceans, mountains, and sparsely populated areas that may not have radar or gauge coverage.

Diagram of a typical stream gauge




Source: GAO diagram adapted from U.S. Geological Survey. | GAO-24-106213

## Limitations

HEC-RAS was introduced in 1995 and is now used by agencies, industry, universities, and others. However, some models within the system are not able to capture physical processes accurately. For example, some of the mathematical models included in HEC-RAS assume the input variables are stationery but, over long periods of time, parameters describing flood-prone areas can change. This limitation may create model output that is not entirely representative of actual flood characteristics, and in-depth knowledge and expertise of the model's parameters are often required to interpret the results. There are companion tools to HEC-RAS for use in flood forecasting that allow modelers to customize simulation time windows and refine model parameter to represent the current state of the system.

More generally, in some instances, flood models may be computationally demanding to run and require significant computer resources, which could impact who has access to the models and their ability to make short-term forecasts.

Source: Unless specified, all graphics on this page were created by GAO.

VIGNETTE
Traditional Modeling

# Wildfires

**A wildfire is an unplanned fire that occurs in a natural area, such as a forest or grassland. Wildfires cause an average of 12 deaths and cost $3.2 billion per year, according to NOAA.[9] USGS reports that wildfires burn about seven million acres each year in the U.S. They can disrupt power services and damage homes and infrastructure.**

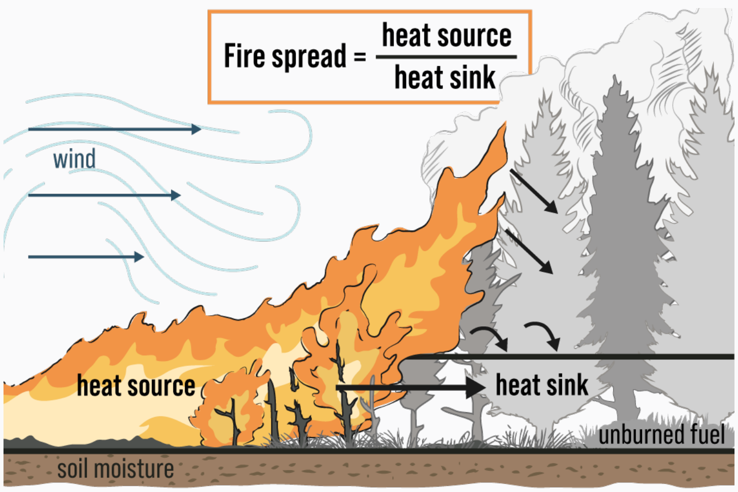
Wildfires play an important ecological role. For example, a wildfire can promote vegetation growth when it burns through dead leaves and other debris. However, wildfires can also cause soil erosion, increase the likelihood of landslides, damage vegetation, or otherwise harm ecosystems. Wildfires can be caused by natural phenomena (e.g., lightning) or human activity. The size and intensity of wildfires are expected to increase, making efforts to predict fire behavior an important step in mitigating their harmful impacts.

Source: mbafai/stock.adobe.com. | GAO-24-106213

## Modeling

The Rothermel surface fire spread model is a commonly used wildfire model in the U.S. It uses a variety of data, including data on wind, slope, the amount of dead and living vegetation (known as fuel), and the moisture content of the fuel. These variables, and others, are put into a mathematical model that uses the laws of physics to estimate how fast a wildfire will spread based on the relationship between fuel and the amount of heat required to ignite that fuel type. The model uses these variables to predict fire spread based on when the estimated heat generated exceeds the heat needed to ignite adjacent unburned fuel.

Simplified representation of the Rothermel fire spread model




Source: GAO analysis of Rothermel fire spread model; GAO (illustration). | GAO-24-106213
Note: This is an oversimplified representation of the Rothermel fire spread model and does not include all needed input variables or physical processes.

Fuel type is a key variable when modeling wildfires. For example, live trees generally require more heat to ignite than other fuels because they contain more moisture. To represent different fuel type characteristics, scientists have developed 53 fuel-specific models to use with the Rothermel model. Additional models are also available for different kinds of fires, such as those that spread across treetops. Another model can use information on slope, shading, elevation, and weather to more precisely estimate the flammability of dead fuels.

Analysts use additional support tools and systems when modeling wildfire behavior. For example, FlamMap is a fire mapping and analysis system developed, in part, by the U.S. Forest Service (USFS) to forecast wildfire growth and behavior with detailed sequences of weather conditions. FlamMap is an ensemble simulation system, combining multiple fire behavior models, such as the Rothermel model, with environmental input variables.

[9] These statistics represent billion-dollar events from 1980 through 2023 and are not representative of all wildfires that occur throughout the year across the U.S. See NOAA National Centers for Environmental Information, *U.S. Billion-Dollar Weather and Climate Disasters* (2023), https://www.ncei.noaa.gov/access/billions/summary-stats.

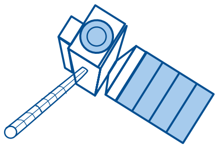
Source: Unless specified, all graphics on this page were created by GAO.



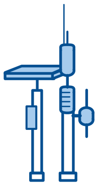
Source: mbafai/stock.adobe.com. | GAO-24-106213

## Data sources

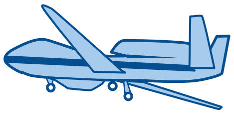
**Databases**, such as those accessed through Landscape Fire and Resource Management Planning Tools (LANDFIRE), a shared program between the wildland fire management programs of the USFS and U.S. Department of the Interior, provide data on vegetation, fuel, and other factors across the U.S.



**Satellites** operated by NASA and NOAA orbit the earth and provide information at various intervals to help monitor wildfires. For example, NASA has a satellite that filters light to help detect wildfires.

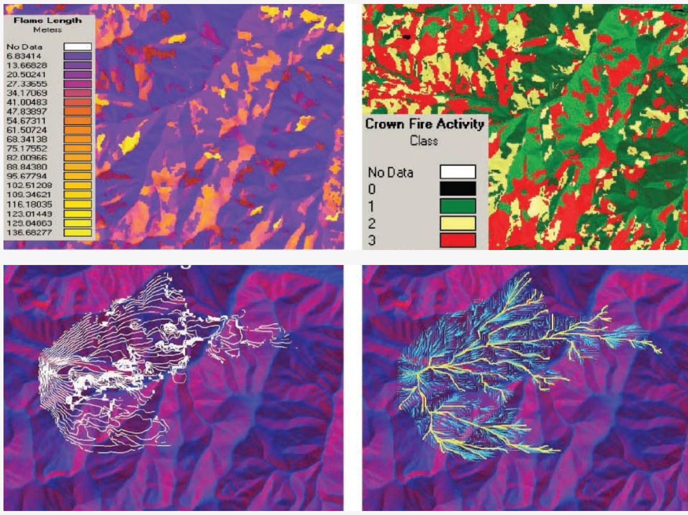


**Radar sensors** can estimate fire fuel and monitor long-term vegetation recovery after a fire, which helps scientists, researchers, and decision makers prepare for and respond to wildfire events.



**Airborne instruments** measure fire temperature and fuel conditions.

Examples of outputs from FlamMap




Source: U.S. Department of Agriculture, Forest Service. | GAO-24-106213

## Limitations

The Rothermel model has been used in fire and fuels management systems since 1972, but it does have limitations. For example, it assumes that fuel type is consistent in a given area and is evenly distributed. Real ecosystems often have multiple fuels distributed unevenly. The model also assumes that the fire is primarily spread by dead fuel and that moisture alone can stop it. In fact, other firefighting measures can be effective, including fuel breaks and chemical fire retardants. These limitations can cause the model to over- or under- predict the rate of spread of any given fire.

Wildfire analysis tools and systems also have limitations. For example, FlamMap will not simulate variations in fire behavior caused by weather fluctuations.

Source: Unless specified, all graphics on this page were created by GAO.

## 1.2 AI and machine learning

AI is a set of technologies able to perform tasks that normally require human intelligence. *Machine learning*, a type of AI, uses algorithms to identify patterns in information. In this report, we focus on machine learning since it is the predominant form of AI undergoing research and development for natural hazard modeling. In instances where we refer to AI, it is in reference to a certain requirement, capability, or use of AI other than machine learning.

The life cycle of an AI system (including a machine learning system) generally involves four phases: design, development, deployment, and continuous monitoring. These phases are often iterative and not necessarily sequential. This report assesses the use of machine learning in natural hazard modeling for technologies across this life cycle, from models that are undergoing research and development to *operational* models. The latter category denotes models routinely used and maintained to support important decisions, such as when to issue a storm warning or where to deploy firefighters.

## 1.3 Machine learning in natural hazard modeling

Machine learning can employ a fundamentally different approach to natural hazard modeling than physics-based traditional models. Many traditional models use algorithms that are designed and coded by developers, who sometimes have domain expertise, to mathematically represent the physical laws of nature. In contrast, machine learning is *trained* through many steps to detect patterns from large data sets (for example, many years of historical data on wind speed or temperature; see text box). In this way, a machine learning algorithm can generate outputs that are based on the patterns and statistical relationships within the training data, rather than on physical laws that are programmed into the model explicitly. At the same time, some of the patterns machine learning generates may reflect physical laws because the data inputs represent real-world natural hazard events, which are governed by physics. There are also hybrid models that employ both programmed physical laws and machine learning.

**Key concept: Machine learning models and training data**

Machine learning algorithms learn the mathematical relationships between inputs and outputs from a *training data set*. *Training data* are historical data and can be numbers, images, audio, or text. For modeling hazards, these data may include temperature measurements, wind speeds, and satellite images. Machine learning uses these data to create formulas based on mathematical patterns it identifies within the training data set. These formulas are what allow the machine learning algorithm to generate predictions when exposed to new data. The quality and quantity of a training data set are significant factors in a machine learning model's performance. A diverse, representative data set containing many examples of historical natural hazard events (e.g., hurricane landfalls) helps ensure that a machine learning model learns patterns that represent the range and complexity of real events. Machine learning models that are trained on data that are not diverse risk producing outputs that are inaccurate and not generally representative of real-world events. Some algorithms require training data that are extensively labeled and curated.

Source: GAO analysis of scientific literature. | GAO-24-106213

Machine learning can be applied to each step of the natural hazard forecasting process (fig. 1), and its use ranges from minor modifications to a traditional model (e.g., using machine learning to assimilate input data into a physics-based algorithm) to complete replacement of a traditional model using the machine learning algorithms described above. A variety of machine learning algorithms can be employed for natural hazard modeling, depending on needs, as shown in table 2.

**Table 2: Selected machine learning algorithms**

| Algorithm | Key features and use |
|---|---|
| Neural network | Groups of artificial "neurons" in layers of interconnected nodes. Typically process data by making a random guess regarding what the data mean, then correcting itself by adjusting its internal structure. Useful for identifying complex, nonlinear relationships in large data sets, which can help predict hazards that have nonlinear behavior, like flash floods. |
| Convolutional neural network | A specialized type of neural network with layers structured to detect patterns and features in images. Useful for translating visual hazard data into numerical data. |
| Generative adversarial network (GAN) | Uses two neural networks to create realistic data from a training set while also evaluating the accuracy of the generated data. GANs have been used to create better initial conditions for weather models. |
| Isotonic regression | Identifies a relationship between a predicted and observed value to minimize errors. Useful for calibrating forecast probabilities. |
| K-nearest neighbor | Predicts the classification of unlabeled data by looking at the nearest neighbors of those data. Useful when the data needed to model hazards have missing values. |
| Multilayer perceptron | A simple neural network with connected hidden layers that are used to learn complex, non-linear relationships in large data sets. |
| Random forest | Uses groups of "decision trees" to classify data into groups or measure relationships. Useful for identifying variables that can improve hazard prediction. |

Source: GAO analysis of scientific literature. | GAO-24-106213

## 1.4 Policy environment

AI has received significant attention from recent presidential administrations and Congresses. Existing federal laws and policies are meant to, among other things, "ensure continued United States leadership in artificial intelligence research and development; lead the world in the development and use of trustworthy artificial intelligence systems in the public and private sectors; prepare the present and future United States workforce for the integration of artificial intelligence systems across all sectors of the economy and society; and coordinate ongoing artificial intelligence research, development, and demonstration activities among the civilian agencies, the Department of Defense and the Intelligence Community to ensure that each informs the work of the others."[10] Most

[10]National Artificial Intelligence Initiative Act of 2020, Pub. L. No. 116-283, § 5101(a), 134 Stat. at 4525, to be codified at 15 U.S.C. § 9411(a), (referring to the National Artificial Intelligence Initiative). The National Artificial Intelligence Initiative Act of 2020 was enacted as Division E of the William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021. Although there have been numerous other recent laws and Executive Orders addressing the use of AI in the U.S., the purposes of this act made clear that its purpose through the Initiative was a comprehensive national strategy led by the

recently, in October 2023, President Biden issued an Executive Order on the *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, calling on federal agencies to lead both the advancement of AI development and efforts to mitigate risks related to its development and use. It also sets new policies and principles for the responsible development and use of AI.[11] In general, these efforts have not been specific to natural hazard modeling.[12]

Some agencies also have individual AI strategies. For example, in 2020, NOAA published two strategies. The *NOAA Data Strategy: Maximizing the Value of NOAA Data* seeks to ensure "readiness for artificial intelligence, machine learning, analytics, and other data science techniques," while the *NOAA Artificial Intelligence Strategy: Analytics for Next-Generation Earth Science* seeks to advance the national AI strategy by strengthening coordination, operational capabilities, workforce proficiency, and multisector partnerships within and beyond NOAA.[13] These initiatives funded items such as supercomputing for developing weather and climate models, wildfire research and infrastructure, and flood mapping and forecasting.[14]

National Artificial Intelligence Initiative Office. 15 U.S.C. §§ 9411(a) and 9412.

[11]Exec. Order No. 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (Oct. 30, 2023), 88 Fed. Reg. 75,191 (Nov. 1, 2023).

[12]One exception was in December of 2022, Congress enacted the Flood Level Observation, Operations, and Decision Support Act, or FLOODS Act. The FLOODS Act included a single provision that would support the integration of AI in natural hazard forecasting. It established a Hydrologic Research Fellowship Program within NOAA and prioritized the research fellows work which may include, among other priorities "apply[ing] artificial intelligence and machine learning capabilities to advance existing hydrologic modeling capabilities." The Flood Level Observation, Operations, and Decision Support Act, Pub. L. No. 117-316, § 14, 136 Stat. 4406, 4415 (2022), to be codified at 15 U.S.C. § 9709(b)(7)(B).

[13]These initiatives have been funded with appropriations from the Infrastructure Investment and Jobs Act, the Inflation Reduction Act of 2022, disaster relief supplemental appropriations, and annual appropriations laws. Infrastructure Investment and Jobs Act (IIJA), Pub. L. No. 117-58, div. J, tit. II, 135 Stat. 429, 1355-57 (2021), and see, *e.g*., Extending Government Funding and Delivering Emergency Assistance Act, Pub. L. No. 117-43, div. B, tit. II, 135 Stat. 344, 358 (2021), Consolidated Appropriations Act, 2023, Pub. L. No. 117-328, div. B, tit. I, 136 Stat. 4459, 4516-17 (2022).

[14]For instance, the Infrastructure Investment and Jobs Act (IIJA) appropriated $672 million for these items. The appropriation provisions in the IIJA supporting research supercomputing infrastructure and wildfire infrastructure are to remain available until Sept. 30, 2024. The provision supporting wildfire prediction, detection, observation, modeling, and forecasting was for fiscal year 2022 only and the provision supporting coastal and inland flood mapping and forecasting was appropriated funds to be expended in equal amounts in each of fiscal years 2022 through 2026 (5 years). IIJA, Pub L. No. 117-58, 135 Stat. at 1356-57, Procurement, Acquisition, and Construction (PAC) for supercomputing at (2); Operations, Research, and Facilities (ORF) for wildfire research at (5), PAC for wildfire infrastructure at (1); and ORF for flood mapping and forecasting at (3), respectively.

# 2 Machine Learning May Significantly Enhance Natural Hazard Modeling

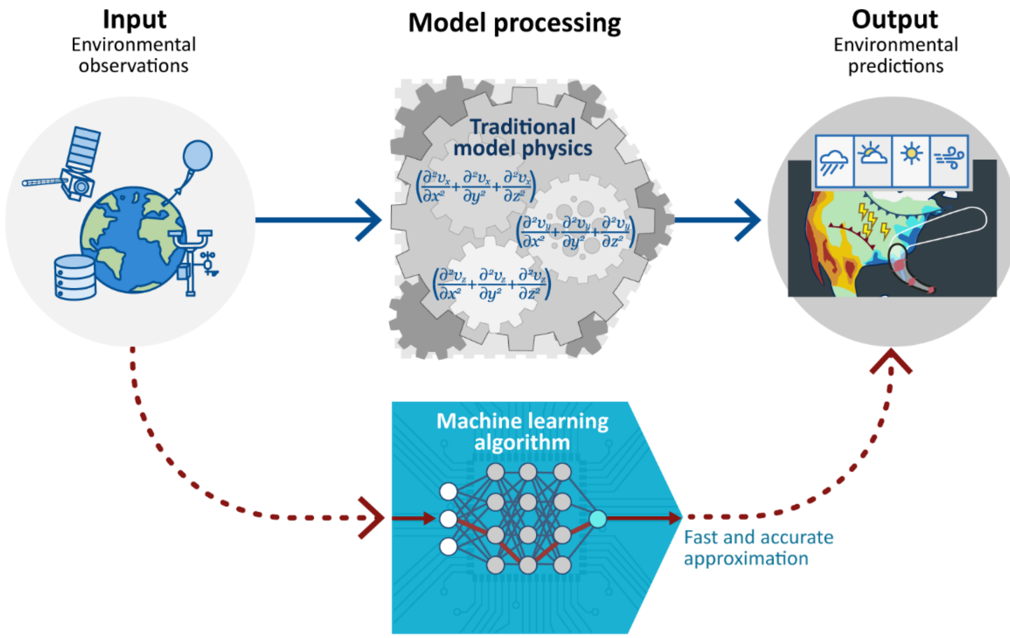
Machine learning has shown potential to significantly improve modeling capabilities for severe storms, hurricanes, floods, and wildfires. Machine learning is being used in at least one active, operational forecast model, for each of these natural hazards, according to agency officials and industry representatives. Other machine learning uses are getting close to operational demonstration, but most of the efforts to use machine learning for hazard models are still in research and development, according to agency officials. Machine learning can be used in many ways for modeling severe storms, hurricanes, floods, and wildfires and can lead to multiple benefits. In some research studies, machine learning has shown the potential to significantly speed up modeling results, leading to quicker forecasts and more timely responses that can save lives and property. Other research has suggested machine learning can increase the accuracy of models by more quickly incorporating and using data from existing sensors. We describe a range of these potential improvements, followed by a series of vignettes on how machine learning is being applied to each of the four hazards, from research efforts through operational forecast models.

## 2.1 Machine learning may speed up forecasting

Many traditional models repeatedly solve formulas based on the physics of the phenomenon being modeled (e.g., the flow of water or the spread of wildfire). Some of these calculations can take a long time to run and consume large amounts of energy. Two machine learning technologies—emulators and digital twins—could replace all or part of such models with components that require fewer calculations. These technologies have the potential to speed up predictions, reduce the cost of running the models, and reduce model uncertainty.

**Emulators** are designed to approximate physical processes represented in physics-based traditional models based on statistical patterns in the inputs and outputs of those models, as shown in figure 2. Once trained, they replace certain parts of the traditional models. Machine learning emulators vary in terms of their technical maturity. Emulators for natural hazard modeling are being widely researched but are generally still in development, although some private-sector companies are currently using emulators in modeling. One example is using an emulator to approximate rainfall for flood modeling.

**Figure 2: Representation of an emulator replacing part of a traditional model**




Source: GAO (analysis and illustrations). | GAO-24-106213

For severe storms, emulators can increase the speed and resolution of model predictions by replacing a component of the traditional model that slows down processing time. Traditional weather models can spend 30 to 80 percent of their computation time estimating how energy from sunlight moves through the atmosphere. An emulator that was trained on the existing observations of this process generated this estimate about 1,000 times faster than the traditional model. However, emulators initially require extensive training data, computational resources for that training data, and expertise to develop. Emulators also need to be tested under operational conditions to ensure their reliability. According to NOAA officials, these emulators are largely still in development.

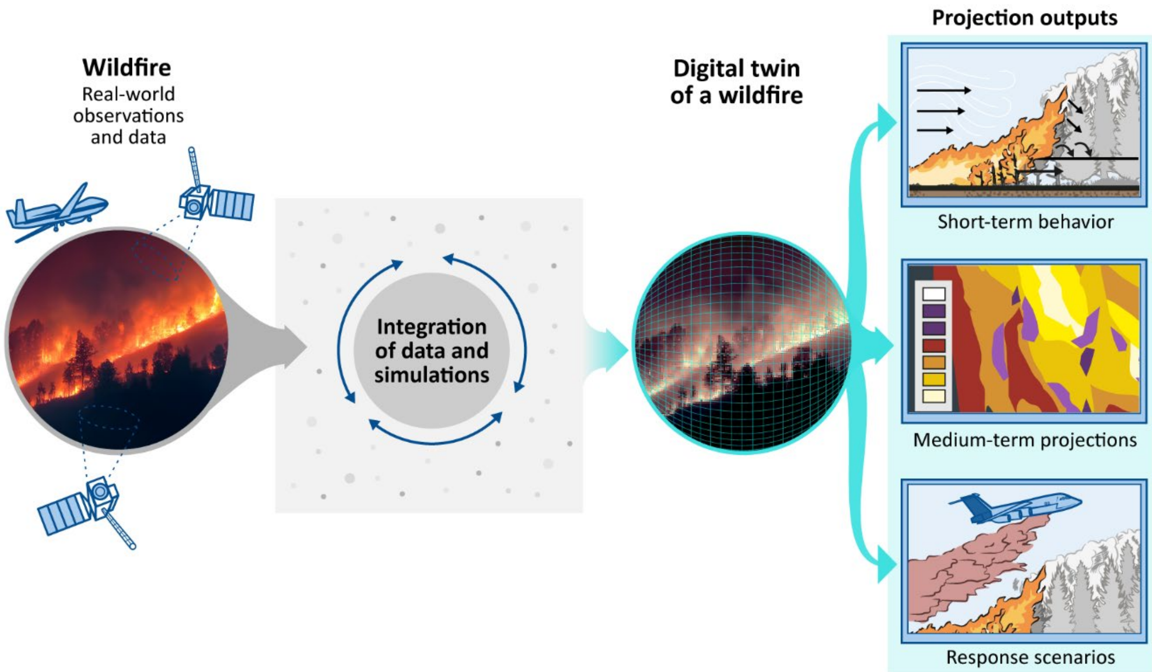
**Looking forward: Digital twins**

Digital twins are like emulators in that they are representations of physical processes that rely on large amounts of data from real-world observations, as shown in figure 3. However, as their name implies, they would represent multiple interrelated processes to create a (digital) "twin" of a natural hazard event. Digital twins for different applications vary in their level of maturity, but for the natural hazards we examined, they are still in an early research stage and require years of development. Digital twins for weather applications could use near-real-time observations to make much faster and higher-resolution predictions. This capability would allow decision-makers to see the impact of different scenarios faster, at different timescales, and with greater geographic specificity. For example, modelers are working to develop a wildfire digital twin to predict how a fire may respond to efforts to fight the fire.[a] Such a digital twin could help identify patterns of firefighting tanker plane deployment that best slow the spread of a fire. Digital twins need extensive training data to capture the range of potential hazard behavior. As a result, their predictions for rare but high-impact events could be inaccurate.

[a]Seong-Jin Yun, Jin-Woo Kwon, and Won-Tae Kim, "A Novel Digital Twin Architecture with Similarity-Based Hybrid Modeling for Supporting Dependable Disaster Management Systems," *Sensors,* vol. 22 (2022): 1-16.

Source: GAO analysis of scientific literature. | GAO-24-106213

**Figure 3: A digital twin recreating a wildfire**




Sources: GAO (analysis and illustrations); mbafai/stock.adobe.com (wildfire photo). | GAO-24-106213

## 2.2 Machine learning may improve the use of data

Machine learning offers multiple potential benefits related to better use of data. It could increase the speed of some modeling tasks, such as data assimilation and quality assurance. It can also quickly translate imagery data into usable numerical inputs for a model.[15] At the same time, preparing data for machine learning use can be a costly and labor-intensive process and machine learning can sometimes lead to inaccurate results when data sets are incomplete.[16]

**Improving data assimilation.** Machine learning can speed up data assimilation—the process of updating forecast models by repeatedly integrating the most current environmental observations to best represent the initial conditions for a prediction (e.g. the condition of the atmosphere for storms and hurricanes). Faster data assimilation could allow models to use more data, such as satellite data that is significantly underutilized.[17] This in turn would provide a more accurate picture of current conditions and should result in a more accurate forecast. Machine learning could also be used to produce model results more quickly. For example, in one study, a machine learning algorithm for assimilating satellite data used to model storms and hurricanes was able to process approximately 100 times as much data as the non-machine learning approach.[18] Several studies have shown benefits of using machine learning to enhance data assimilation, but it has yet to be implemented operationally, according to NOAA officials.

**Looking forward: Generative AI and Flood Modeling**

One expert we spoke with described how generative AI may be able to use textual reports of rainfall and flooding from across the globe and learn the relationship between the rainfall reports and the likelihood of neighborhood flooding. According to the expert, a user could ask a chatbot linked to such a model whether a recent storm in a city will lead to floods in a certain neighborhood. If enough rainfall reports are published online about that storm event, the expert said the answer a chatbot provides now is surprisingly accurate. Combining a generative AI chatbot that anyone could access via the internet with machine learning flood models could provide forecasts to people who may otherwise lack access to traditional real-time flood monitoring networks, such as in some lesser developed countries.

Source: Expert meeting and GAO analysis of scientific literature. | GAO-24-106213

**Increasing the speed of quality assurance review.** Machine learning could also improve the use of data by speeding up the quality assurance and quality control (QA/QC) processes used for natural hazard models. QA/QC is a pre-processing step to review incoming observations and ensure that inaccurate or unreliable data do not influence models. With NOAA hurricane data, pre-processing can delay transmission of new data to hurricane models by up to 30 minutes. Because the volume of data gathered is increasing, NOAA expects this delay to generally increase and that they will need to use machine learning to flag potential inaccuracies in incoming data for human review. By decreasing the amount of time required for QA/QC, NOAA officials said they

[15]Sarah M. Griffin, Anthony Wimmers, and Christopher S. Velden, “Predicting Rapid Intensification in North Atlantic and Eastern North Pacific Tropical Cyclones Using a Convolutional Neural Network,” *Weather and Forecasting*. vol. 37 (2022): 1333-1355. https://doi.org/10.1175/WAF-D-21-0194.1.

[16]Zihung Sun et al. “A Review of Earth Artificial Intelligence,” *Computers & Geosciences*, vol. 159 (2022): 1-16. https://doi.org/10.1016/j.cageo.2022.105034.

[17]For example, the traditional weather model that drives severe storm and hurricane prediction currently uses only 1 to 3 percent of relevant available satellite data due in part to its computational limits.

[18]Eric S. Maddy and S. A. Boukabara, “MIIDAPS-AI: An Explainable Machine-Learning Algorithm for Infrared and Microwave Remote Sensing and Data Assimilation Preprocessing - Application to LEO and GEO Sensors,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14 (2021): 8606-8616. https://doi.org/10.1109/JSTARS.2021.3104389.

hope machine learning will improve the use of data by their hurricane models.

**Expanding use of satellite imagery.** In a research setting, machine learning can be used to translate satellite imagery into numerical probability outputs that can be used to forecast hurricane intensity. For example, a machine learning model called AI-RI (RI is for "rapid intensification") can expand use of satellite imagery of a hurricane's structure to predict the probability of a hurricane rapidly gaining in strength within 12 to 72 hours. Rapid hurricane intensification is difficult for traditional models to forecast, but it is key for potentially life-saving decisions, such as when to issue an evacuation order. For wildfires, researchers have used machine learning to combine satellite imagery with existing ground surveys to estimate the amount of vegetation and other fuel in areas where data were sparse or outdated. In one study, machine learning use improved the accuracy of wildfire model estimates by 38 percent.[19] These machine learning applications are also still in the research stage, but some are close to being operational, including AI-RI, according to researchers.

**Drawbacks.** Although machine learning could help some aspects of data use in modeling, it also can require extensive up-front work to ensure the data are readily usable by machine learning. Machine learning can initially require large amounts of training data in a specific format, but there is currently a lack of "benchmarked" environmental data sets formatted and ready for machine learning training.[20] Pre-processing these data to make them "AI ready" requires input from researchers and modelers who are versed in both machine learning and the specific hazard being modeled and can be costly and time consuming. In addition, machine learning modeling based on data sets that aren't "AI ready" can sometimes lead to results worse than what a traditional modeling approach would yield (see text box). We discuss multiple data-related challenges later in this report.

**Key concept: Overfitting**

Overfitting can occur when there is insufficient training data to represent all conditions that could be modeled, the data contain too much irrelevant or "noisy" data, or a model is overly complex and learns too much from the irrelevant data. A model that is overfit does not produce reliable predictions because it describes a limited set of hazard conditions that cannot be generalized to hazard scenarios that differ from the training data set. If machine learning applications used for hazard modeling are built without sufficient training data and domain expertise, they are more likely than traditional models to "overfit" to a particular set of data.

Source: GAO analysis of scientific literature. | GAO-24-106213

## 2.3 Machine learning may improve predictions in the absence of data

Machine learning has the potential to reduce uncertainty in some situations where data don't exist or are insufficient by creating plausible *synthetic data* based on prior information. Synthetic data are new data generated by machine learning models after

[19]Amy L. DeCastro et al., "A Computationally Efficient Method for Updating Fuel Inputs for Wildfire Behavior Models Using Sentinel Imagery and Random Forest Classification," *Remote Sensing* vol. 14, no. 6 (2022): 1-12, https://doi.org/10.3390/rs14061447.

[20]M.G. Schultz et al., "Can Deep Learning Beat Numerical Weather Prediction?" *Philosophical Transactions A,* vol. 37 (2020): 1-22. https://doi.org/10.1098/rsta.2020.0097.

they have been trained on a data set of historical, real-world weather conditions. These data may be useful for modeling in locations that lack reported data, are difficult to track in real time (e.g., fuel moisture data), or are sparsely populated (e.g., data on the distribution of wildfire fuel).

For example, researchers have developed a machine learning algorithm that may be able to recreate radar observations of severe storms from satellite data for use in traditional models. These synthetic observations may be able to extend "radar" coverage to areas that currently lack it, such as mountainous areas in the western U.S.[21] However, this method still requires domain expertise of the simulated hazards, extensive training (and possible retraining) of the machine learning algorithm, and further research to confirm that the algorithm is generalizable to areas other than where it was trained.

## 2.4 Machine learning can help leverage multiple models to improve predictions

Machine learning has the potential to improve ensemble modeling. Ensemble modeling can reduce the uncertainty in model predictions by using multiple models (rather than one) and combining their predictions. This enables modelers to better represent the ways hazard may behave. This approach enables a larger range of potential outcomes to be forecast which helps to determine the most likely outcome. The result of ensemble models is generally a better forecast than that of a single model run. However, because of computational resource limits in traditional modeling, the number of model projections an ensemble model can incorporate is generally limited to 20 to 50 projections.[22] Studies indicate that because machine learning can use fewer computational resources than traditional models, machine learning may be able to create ensembles with hundreds of models, which could reduce uncertainty and produce more reliable forecasts. However, the machine learning techniques used to create these especially large ensemble models are more complex, harder to train, and potentially harder to understand and troubleshoot than other machine learning methods.

Another machine learning technique to improve ensemble forecasting, which is more efficient than traditional methods of ensemble modeling, uses machine learning to improve the post-processing and combination of ensemble model results. However, one study did not show large improvements in short-term predictions as a result of using

[21]In this study the researchers trained a machine learning algorithm on the occurrence of lightning observed in the tops of clouds by satellites with radar observations from multiple sources. Once the algorithm learned how lightning occurrence corresponded to radar observations, it could create "synthetic" radar observations for severe storms that lacked radar coverage by using the relationship between cloud-top lightning flashes and radar observations of severe storms. Kyle A. Hilburn, Imme Ebert-Uphoff, and Steven D Miller, "Development and Interpretation of a Neural-Network-Based Synthetic Radar Reflectivity Estimator Using GOES-R Satellite Observations," *Journal of Applied Meteorology and Climatology*, vol. 60 (2021): 3-21. https://doi.org/10.1175/JAMC-D-20-0084.1.

[22]Kirsten I. Tempest, George C. Craig, and Jonas R. Brehmer, "Convergence of Forecast Distributions in a 100,000-member Idealised Convective-scale Ensemble," *Quarterly Journal of the Royal Meteorological Society* vol. 149 (2023): 677-702. https://doi.org/10.1002/qj.4410.

machine learning for post-processing.[23] Machine learning ensemble applications are still in the research stage, according to NOAA officials.

**Looking forward: climate change**

Machine learning models depend on the availability of historical training data that are relevant to current environmental conditions. If climate change significantly alters the behavior of real-world natural hazards over time (e.g., the average hurricane becomes stronger, rainfall patterns across the U.S. change), then historical training data may not represent the hazard characteristics as accurately. It is not yet clear the extent to which this complication might affect the utility of machine learning models nor how best to mitigate it.

Source: GAO analysis of scientific literature and expert interview. | GAO-24-106213

[23]Vladimir M. Krasnopolsky and Y. Lin, "A Neural Network Nonlinear Multimodel Ensemble to Improve Precipitation Forecasts over Continental US," *Advances in Meteorology*, (2012): 1-11. https://doi.org/10.1155/2012/649450.

VIGNETTE
Machine Learning

## Machine Learning for Modeling

# Severe Storms

Source: Gavin/stock.adobe.com. | GAO-24-106213

Severe storms can be difficult to model because their features (e.g. tornadoes, hail, and strong wind) occur over a small area and can intensify within minutes. Traditional severe storm models are not able to predict at this small of a scale and typically only provide hourly forecasts.

Machine learning has the potential to improve forecasting by allowing modelers to incorporate larger data sets, discover new relationships in the data, correct bias, and produce faster and more frequent forecasts. One of the most common uses of machine learning to severe storm models is to refine the output of traditional models, also known as post-processing.

### Benefits of Machine Learning

Machine learning models have been able to address some key limitations of traditional severe storm modeling. For example, studies have shown that, once trained, machine learning models produce quicker predictions than traditional models. In addition, machine learning can uncover relationships within data, which could help scientists find new features that were not previously considered in a model.

The following are two of the studies we found demonstrating the potential benefits of machine learning for severe storm forecasting:

- To predict hail, scientists used isotonic regression, which identifies relationships between a predicted and observed value, and random forest, which classifies data into groups to measure relationships, based machine learning models trained on traditional model output and observational data. The resulting forecasts were more reliable and had lower bias than traditional methods.[24]
- Scientists used a random forest-based machine learning model trained on data from traditional models and storm reports to produce better forecasts of severe hail and wind than a traditional model. However, the traditional model performed better for tornadoes than the machine learning model.[25]

### Limitations of Machine Learning

Many types of machine learning models rely on traditional model output for training data, which could cause the machine learning models to have errors as a result of the existing errors within traditional models. Machine learning models can also be difficult to make generalizable which means their performance may vary based on location.

### Operational Use

Machine learning use for severe storm modeling is still largely in the research and development phase. One exception is ProbSevere (version 2), which is a machine learning model that has improved forecasters' ability to issue warnings of severe thunderstorms and tornadoes. Forecaster feedback suggests that ProbSevere allows them to warn of severe weather about 5 to 10 minutes earlier than they otherwise would have, providing more time to prepare for these hazards.[26] However, more research is needed to quantify exactly how ProbSevere affects lead time. Forecasters who evaluated the model said it increased their confidence in issuing warnings 95 percent of the time and increased lead time 76 percent of the time. Forecasters at the National Weather Service currently use ProbSevere as an additional tool when predicting severe weather.

[24] Amanda Burke et al., "Calibration of Machine Learning-Based Probabilistic Hail Predictions for Operational Forecasting," *Weather and Forecasting*, vol. 35, no. 1 (2020): 149-168. https://doi.org/10.1175/WAF-D-19-0105.1.

[25] Eric D. Loken, Adam J. Clark, and Christopher D. Karstens, "Generating Probabilistic Next-Day Severe Weather Forecasts from Convection-Allowing Ensembles Using Random Forests," *Weather and Forecasting*, vol. 35, no. 4 (2020): 1605-1631. https://doi.org/10.1175/WAF-D-19-0258.1.

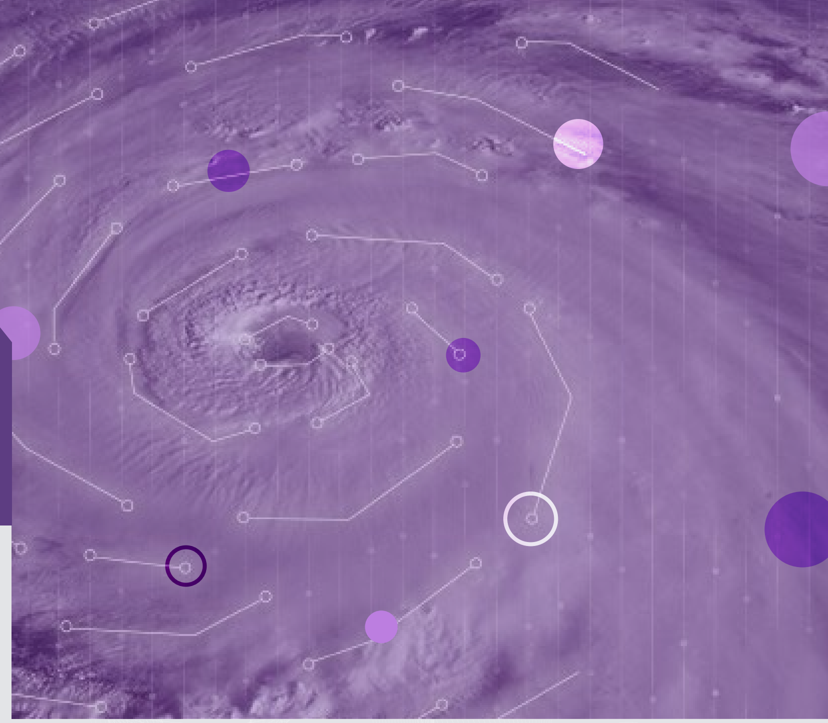
[26] John L. Cintineo et al., "NOAA ProbSevere v2.0 ProbHail, ProbWind, and ProbTor," *Weather and Forecasting*, vol. 35, no. 4 (2020): 1523-1543. https://doi.org/10.1175/WAF-D-19-0242.1.

Source: TechSolution/stock.adobe.com (mesh pattern). Unless specified, all other graphics on this page were created by GAO.

VIGNETTE
Machine Learning

Machine Learning for Modeling

# Hurricanes


Source: elroce and NASA/stock.adobe.com. | GAO-24-106213

Hurricanes are difficult to forecast due to their complex physical processes and interactions with the environment. Traditional hurricane models have improved in the past few decades, especially for forecasting the path of a hurricane. However, forecasting when a hurricane will form and how intense it will become remains a challenge.

## Benefits of Machine Learning

Machine learning has shown promise in helping to address some of the key limitations of traditional hurricane modeling. For example, studies have shown that machine learning models have provided more accurate hurricane intensity forecasts and reduced forecast lead time. Machine learning has also enhanced research and understanding of hurricanes by discovering new relationships in data. Machine learning models can also incorporate more data and decipher patterns that traditional techniques may miss.

The following are two of the studies we found that demonstrate these potential benefits:

- To predict hurricane intensity, scientists used a machine learning model based on a "multi-layer perceptron"—a simple neural network that learns complex, non-linear relationships in large datasets—and trained on traditional model output and observational data. It produced lower forecast error and correctly identified more rapid intensification events than a traditional model it was compared to.[27]
- To predict the locations of a hurricane's maximum wind speeds and wind radii, scientists used a convolutional neural network—which is structured to detect patterns and features in images—trained on satellite data and global track datasets. The model reduced forecast error of a traditional wind radii estimation method by 32 percent.[28]

## Limitations of Machine Learning

Machine learning shares some of the limitations of traditional numerical weather prediction models, such as a reliance on large amounts of data and high computational costs. Machine learning models also may perform worse than traditional models in forecasting high-intensity hurricanes because there are not enough data for them to train on.

## Operational Use

Despite its potential benefits, machine learning does not yet appear to be widely used for hurricane forecasting and is still largely in the research and development phase. One exception is the Statistical Hurricane Intensity Prediction Scheme (SHIPS), which uses a simple form of machine learning to more accurately predict rapid intensification. SHIPS has historically outperformed most operational hurricane intensity models through the mid-2010s and still provide skillful intensity guidance. In 2021, it was the best of 48 models evaluated by the National Hurricane Center at providing intensity guidance for hurricanes in the Pacific at longer lead times. The center currently uses SHIPS, along with other models, to help forecasters predict hurricane intensity.

[27] Wenwei Xu et al., "Deep Learning Experiments for Tropical Cyclone Intensity Forecasts," *Weather and Forecasting*, vol. 36 (2021): 1453–1470. https://doi.org/10.1175/WAF-D-20-0104.1.

[28] Jing-Yi Zhuo and Zhe-Min Tan, "Physics-Augmented Deep Learning to Improve Tropical Cyclone Intensity and Size Estimation from Satellite Imagery," *Monthly Weather Review*, vol. 149 (2021): 2097-2113. https://doi.org/10.1175/MWR-D-20-0333.1.

Source: TechSolution/stock.adobe.com (mesh pattern). Unless specified, all other graphics on this page were created by GAO.

# Machine Learning for Modeling Floods

Source: oobqoo/stock.adobe.com. | GAO-24-106213

Floods can be very difficult to forecast with traditional models, requiring complex data inputs and equations. These factors result in long run times for traditional flood models, which makes them less useful for sudden events, such as flash floods.

## Benefits of Machine Learning

Machine learning has the potential to model floods based only on historical data and current observations and without the complex equations of traditional models. It therefore shows promise in helping to address some of traditional modeling's key limitations. For example, machine learning models can be quicker to develop and, once trained, use far fewer computational resources. They can provide predictions faster than traditional models because they require fewer computing resources, making them more useful for flash floods.

In addition, machine learning may be able to analyze large data sets for insights into causes of flooding, which could lead to improved traditional models. One type of machine learning algorithm (artificial neural networks) can also deal with incomplete data sets, a major challenge for all environmental models. And machine learning models may be able to incorporate certain sources of data—such as satellite and road camera images—more efficiently than traditional models, according to researchers we interviewed.

The following describes three of the studies we found that demonstrate these potential benefits:

- A model using "multilayer perceptrons" has predicted water levels during a flash flood. In one study, the model was able to make real-time predictions of maximum water levels in an urban flooding scenario, which traditional models struggle to do.[29]
- A convolutional neural networks has been used to predict the depth and extent of flooding throughout a flood event. In one study, the network was able to predict the depth, size, and location of a flood from a relatively small amount of data, and more accurately than a commonly used type of traditional model.[30]
- A generative adversarial network has been used to predict maximum flood extent and depth based solely on rainfall images. Such models have been able to predict flood depth with reasonable accuracy, but they have struggled with shallower flooding.

## Limitations of Machine Learning

Machine learning is often unable to extrapolate from its training data to new situations, an issue known as the "generalization problem." For example, some machine learning algorithms may perform well for short-term but not long-term predictions, or for lower flood volumes but not higher volumes, or for locations without prior flood data.

## Operational Use

In the federal government, current flood models do not use machine learning, according to agency officials. In the private sector, one company used machine learning to reduce its computational costs for flood modeling by completely replacing a traditional flood model with a machine learning model. This change allowed the model to be applied on a larger scale than before. During the 2021 monsoon season in India and Bangladesh, the model outperformed the traditional model and enabled the company to help send 100 million alerts to people in flooded areas and to relevant authorities.[31]

[29]Simon Berkhahn, Lothar Fuchs, and Insa Neuweiler, "An Ensemble Neural Network Model for Real-Time Prediction of Urban Floods," *Journal of Hydrology*, vol. 575, (2019): 743-754. https://doi.org/10.1016/j.jhydrol.2019.05.066.

[30]Hossein Hosseiny, "A Deep Learning Model for Predicting River Flood Depth and Extent," *Environmental Modelling & Software*, vol. 145, (2021): 1-7. https://doi.org/10.1016/jenvsoft.2021.105186.

[31]Sella Nevo et al. "Flood Forecasting with Machine Learning Models in an Operational Framework," *Hydrology and Earth System Sciences*, vol. 36, (2022): 4013-4032. https://doi.org/10.5194/hess-26-4013-2022.

Source: TechSolution/stock.adobe.com (mesh pattern). Unless specified, all other graphics on this page were created by GAO.

# Machine Learning for Modeling Wildfires

Source: mbafai/stock.adobe.com. | GAO-24-106213

Traditional, physics-based models of wildfire use equations describing canopy biomass, heat transfer, and fluid mechanics to model fire behavior in space and time. These models demand detailed data sets on such factors as the location and dimensions of trees or fuel. These data are difficult to obtain for large areas, and the resulting models often require too much computing power to run in real-time. Some other traditional models are faster and can be adequate for large areas, but they are often less accurate because they leave out some of the complexity of wildfire spread

## Benefits of Machine Learning

According to wildfire researchers we interviewed, the majority of machine learning applications to wildfire modeling have been for model inputs. Applying machine learning to wildfire spread models is a promising research area still in the early stages of development.

Machine learning techniques may help address some of traditional modeling's key limitations. For example, researchers have used machine learning to better estimate fuel data based on multiple sources, thereby improving inputs to traditional models with more updated data. Others have replaced model components with a machine learning emulator, leading to faster predictions. And they have improved model outputs by overlaying historical information onto an ongoing fire, providing insight into where firefighting activities could be more successful based on past fires. Machine learning can also generally make more effective use of a wider variety of data.

The following describes three of the studies we found that demonstrate these potential benefits:

- One study trained random forest algorithms on radar data from satellites, data on land cover, and tree mortality surveys. It created more accurate estimates of wildfire fuel than previous estimates, although often did not correctly classify clusters of dead trees.[32]
- A type of K-nearest neighbor algorithm called "SMOTE" has been used to generate synthetic data to model large wildfires. Since these wildfires occur infrequently, this could address the challenge of a lack of training data for other machine learning models.
- Convolutional neural networks can learn spatial relationships and are used to predict how a forest fire will evolve. One study also showed they could be useful for detecting fires using satellite images.[33] However, they are harder to interpret than other models.

## Limitations of Machine Learning

Machine learning can be expensive and difficult to scale. It requires a large amount of training data, which are frequently unavailable for wildfires, and significant computing power during training. Machine learning accuracy also requires domain expertise in wildfire research to ensure realistic modeling.

## Operational Use

U.S. Forest Service officials told us three of their six models use machine learning. One of them, Potential Control Locations, was used in a 2022 wildfire in Colorado and discovered that the western side of the operating area offered a higher probability of success for controlling the fire. Responders used this information to plan their response and deploy resources to contain the fire.

[32] Amy L. DeCastro et al., "A Computationally Efficient Method for Updating Fuel Inputs for Wildfire Behavior Models Using Sentinel Imagery and Random Forest Classification," *Remote Sensing*, vol. 14 (2022): 1-12. https://doi.org/10.3390/rs14061447.

[33] Piyush Jain et al., "A Review of Machine Learning Applications in Wildfire Science and Management," *Environmental Review*, vol. 28 (2020): 478–505. https://doi.org/10.1139/er-2020-0019.

Source: TechSolution/stock.adobe.com (mesh pattern). Unless specified, all other graphics on this page were created by GAO.

# 3 Challenges to Applying Machine Learning in Natural Hazard Modeling

Using the information we gathered from agency officials, academic researchers, industry representatives, and the scientific literature, we identified several challenges affecting the development and use of machine learning for modeling severe storms, hurricanes, flood, and wildfires. Some of these challenges may slow the adoption of machine learning technologies, hinder their use, or lead to inequitable outcomes across society.

## 3.1 Data gaps, bias, and incompatibility

The development and performance of machine learning models depend on data. If the necessary data are missing or flawed, it can hinder the use of the technology for its intended purpose. We identified three broad categories of data challenges: gaps in observational data, incompatible data, and difficulty accessing data. Addressing these issues could improve forecasting of severe storms, hurricanes, floods, and wildfires.

### 3.1.1 Observational data gaps

Data gaps (insufficient data) have long posed challenges for natural hazard modeling. Availability of observational infrastructure, rarity of certain weather events, and limited access to certain data are common challenges for both traditional and machine learning modeling. However, data gaps may pose more challenges for machine learning models, in part because of the models' complexity. For example, using some machine learning technologies requires much more effort to understand why and when they produce unreliable or inaccurate results. With traditional models, forecasters and decision-makers may sufficiently understand how the models operate to interpret their outputs and recognize when they are usable, to what extent, and whether there may be flaws in the data the model relies on. For example, Forest Service officials told us that wildfire expertise can compensate for flaws in traditional models because those who understand the physical principles programmed into the models can recognize when the model makes a prediction at odds with those principles. In contrast, some machine learning models are based on patterns the algorithm finds in the data, and these patterns may not have a basis in previously understood physical processes.

The following are key gaps in observational data.

- **Geographic gaps.** Observational infrastructure (e.g., weather stations and stream gauges) tend to be located where extreme events are expected to happen, according to an expert. This limits model usability in areas where natural hazards may not have been historically frequent.

  For severe storms, there are more data for use in machine learning models in locations with greater population densities, according to an atmospheric research scientist familiar with machine learning applications. This could reduce the accuracy of such models in rural areas.

For hurricanes, historical records contain more data on the Atlantic basin than on the Pacific, according to hurricane researchers. As a result, machine learning models trained on these data will be less accurate for Pacific hurricanes.

For floods, some areas prone to flooding have less observational data available for modeling. For example, there are fewer ground-based stations that collect data at higher elevations, which can cause spatial "blind spots" and provide inadequate representation and increased model error within mountainous regions.

For wildfires, some machine learning models are accurate, but only in the local context on which they are trained, Forest Service officials told us. These models would be of little use in other locations. Traditional models based on wildfire physics have accuracy limitations but are usable across geographic locations with few modifications.

- **Data availability**. A variety of factors limit data availability. For example, some systems do not readily capture certain data, data are limited for rare natural hazard events, and some ground and satellite data are only obtainable intermittently. These issues limit the data available to train machine learning models for severe storms, hurricanes, floods, and wildfires.

  For severe storms, human-collected observational data are needed for AI models of tornadoes, according to NOAA officials at the National Severe Storms Laboratory. For example, because radar have difficulty detecting tornadoes, only about 45 percent of tornadoes are reported, which degrades the quality of the forecasts generated by AI, according to one expert. According to NOAA officials, better data would improve machine learning's ability to detect patterns in the physical processes leading to storm hazards.

  For hurricanes, due in part to intermittent satellite data, observations of the needed atmospheric variables for hurricane forecasting are not consistently available at all locations on the earth or at all locations in and around the hurricane. This limits data available for forecasting.

  For floods, flood events are generally infrequent, which limits the number of recorded events that can be used for machine learning model training, according to USGS officials. Additionally, a flood expert we spoke with told us that sensors used for data collection can get washed away during floods, limiting the availability of some important data.

  For wildfires, information about ground conditions—such as vegetation—is used to detect and predict wildfire behavior. However, according to academic researchers, the national database for fuels data—models created using field data, satellite imagery, and knowledge of related biological and physical properties—is only updated every 3 to 5 years, and therefore may be out of date.

- **Missing historical data.** Machine learning model training data need to be substantial and adequate for models to learn patterns and perform well. Having access to complete and accurate historical data that represent conditions appropriately is essential to model performance.

For severe storms, archives of historical data required to train machine learning models to perform well are not always available, NOAA officials at the National Severe Storms Laboratory told us.

For hurricanes, Category 5 hurricanes are a rare event and last for a short time, so the training data for that event type is limited, according to NOAA officials. The officials speculated that they only have 2- or 3-days' worth of historical data for category 5 hurricanes.

For floods, time series data are necessary for modeling. However, some machine learning algorithms consider events in a time series to be independent of each other. The Long Short Term Memory neural network model is popular for floods because it was designed to learn from long-term sequential data and can be used for time series forecasting. However, any gaps in the time series of the data (e.g., data missing for a certain day) will affect the model's forecast.

For wildfires, the National Interagency Fire Center (NIFC) compiles annual wildfire statistics for federal and state agencies. However, current reporting processes for official wildfire data did not begin until 1983. As a result, NIFC does not publicly share official data prior to 1983.

- **Private data**. Researchers said that some data that could be used for training or modeling may be unavailable or inaccessible due to privacy restrictions.

  For floods, image data from municipally owned or operated camera systems that could be used for flood mapping may be inaccessible due to privacy concerns.

  For wildfires, researchers told us that they are unable to obtain data about fuel characteristics on private land. According to these researchers, the USGS collects these data but cannot share them with the research community because of privacy restrictions, limiting the data researchers can use for models.

### 3.1.2 Data compatibility

In some cases, the necessary data are available for modeling, but they are not compatible with different priorities, programming languages, or organizational standards.

- **Observation network priorities.** Most national and individual observation networks in the U.S. are designed for use by agencies or organizations with a specific mission, and data from some of these sources can be unavailable to others. For example, a recent report from the President's Council of Advisors on Science and Technology on modernizing wildland firefighting recommended that the Department of Defense review and consider releasing classified space-based data containing incidental wildfire observations to the research community for use in wildfire modeling.[34] In addition, an expert we spoke with from a private company told us that footage from security and city-based cameras would fill major data gaps in their flood modeling efforts but that those data are not

[34] President's Council of Advisors on Science and Technology, *Report to the President: Modernizing Wildland Firefighting to Protect Our Firefighters* (February. 2023).

obtainable for use outside their original intended purposes.

- **Programming language.** Most machine learning algorithms are trained and used with Python code while some weather and climate models are written in FORTRAN.[35] Converting these programming languages is possible but requires additional labor and other resources.
- **"AI-ready" data.** Some agency officials and academic researchers we spoke with said that there are no unifying standards for how data are collected, labeled, or used (or made "AI ready"). For example, researchers we spoke with told us they are supervising a wildfire project that uses data from the Bureau of Land Management, NOAA, NASA, and the Forest Service, and none of these data are on the same scale, in the same format, or were compiled using the same standards. They told us that this compatibility issue is common and that it takes a lot of time to resolve. This issue exists for non-machine learning models as well, but some solutions have already been developed to help address it. NOAA officials at the Center for Artificial Intelligence—an AI knowledge sharing center that works across NOAA offices—told us that they are working to develop standards for AI-ready data. For example, they worked with Earth Science Information Partners, NASA, USGS, and others to develop a checklist for researchers to use when curating data for machine learning. The checklist includes items for data preparation, data quality, documentation, and data access. However, this effort is in its early stages and the impact the checklist has had on making data more usable, and the extent to which it is used, is unknown.

## 3.2 Trust in machine learning

Machine learning models are not always understandable by those who use them, which can reduce those users' trust in their output. For example, it can be difficult to determine whether they are producing an inaccurate result compared to non-machine learning models and, if so, why. This difficulty can contribute to researcher and end-user hesitance to adopt machine learning models in some cases. Moreover, the high complexity of some machine learning techniques, such as deep learning, can make determining how they make predictions nearly impossible. Models referred to as "black box" models are those for which the inner workings are not transparent and understandable to users. Trust in black box models is traditionally based on cross-validation of the model's performance.

According to researchers at the National Center for Atmospheric Research, knowing the inner workings of a model is important for validating the model for scientific use, tuning model settings to maximize accuracy, and trusting that the model is working as designed. There is no single standard across organizations or sectors for what makes

[35]Python is a programming language that is based on traditional languages but is better suited for current operating systems. FORTRAN is a high-level programming language developed in the 1950s for scientific and engineering applications.

machine learning trustworthy.[36] We identified three dimensions that help build trust in machine learning models, also discussed in the GAO AI Accountability Framework.[37] Though these terms are sometimes used interchangeably, they have distinct qualities:

- **Transparent AI** refers to making information about the data and decisions used to design develop, train, and operate the model, as well as information about the limitations of the model, accessible. According to the GAO AI Accountability Framework, transparency is important to detect errors or misuse and ensure equitable treatment of people affected by AI systems.
- **Interpretable AI** refers to building the model in a way that enables a user to simply look at the model and understand its development and output. The GAO AI Accountability Framework identifies interpretability as necessary for human supervision of the model to ensure accountability.
- **Explainable AI** refers to methods and techniques that produce accurate, explainable models of why and how an AI algorithm arrives at a specific decision. This process involves creating secondary methods or tools after the model has already been created in order to explain the original model. The GAO AI Accountability Framework states that a lack of explainability may limit confidence and trust in AI models.

Some machine learning models have proprietary restrictions and limited information about data sources, which can hinder transparency, interpretability, and explainability. For example, companies and researchers may restrict sharing information about the design and structure of their algorithms or the source of data used to train the models. Officials at the Forest Service and the Department of Homeland Security's Federal Emergency Management Agency told us that private companies they have worked with do not provide them with the details of their machine learning models, key descriptions of data variables used in the model, or data sources. According to NOAA officials, private companies are generally unwilling to share proprietary information, but this information is important for evaluating models' performance.

Some agencies have ongoing efforts to promote the development of machine learning systems that are trustworthy. For example, a NASA framework about the use of machine learning states that the basic elements of data and decisions must be available to tell a logical story of why a machine learning system is operating in a specific manner.[38] The framework refers to

[36]According to officials at the National Science Foundation (NSF), there is no standard definition for trustworthy AI. Multiple research communities are exploring this complex topic. In addition, the National Institute of Standards and Technology (NIST) is conducting research, engaging stakeholders, and producing reports on the characteristics of trustworthy AI, available at https://www.nist.gov/trustworthy-and-responsible-ai. NIST is partnering with other organizations to support initiatives on trustworthy AI. That includes a partnership NSF on an Institute for Trustworthy AI in Law & Society.

[37]GAO, *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities*, GAO-21-519SP (Washington, D.C.: June 30, 2021).

[38]NASA, *NASA Framework for the Ethical Use of Artificial Intelligence (AI)*, NASA/TM-20210012886 (Washington, D.C.: April 2021).

explainable AI as a crucial feature of machine learning systems but notes that it is difficult to develop and incorporate. Other users of machine learning models, such as forecasters or state foresters, might prefer a different level of trust in machine learning models than researchers do. For example, according to the National Association of State Foresters, many state foresters are open to using machine learning models for wildfire prediction and response support, but only if the model is proven in an operational context. They do not want to use machine learning models that are still in the R&D stage due to safety concerns, but they also do not need to necessarily understand the inner workings of the models. Similarly, forecasters interviewed by the National Science Foundation (NSF) AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography were more interested in the model input and evaluation than the inner workings of the models, researchers at the National Center for Atmospheric Research told us. National weather Service forecasters we interviewed similarly stated that they would be interested in using machine learning models if they had the opportunity to learn more about and test them.

## 3.3 Limited coordination and collaboration

Limited formal collaboration and partnership channels, along with competing priorities in certain sectors, also creates challenges for developing and adopting machine learning for natural hazard modeling, according to researchers and experts. For example:

- **Some end-users lack testing opportunities.** Atmospheric research scientists emphasize the importance of communication between the person applying machine learning and the machine learning model developer to, for example, determine what kind of features are important and whether machine learning is necessary to obtain the information required. However, some NOAA forecasters told us the opportunity had not arisen to discuss or test any machine learning models with researchers. They told us that if machine learning model outputs were ready in near real-time and easily accessible to them, they would look at the use of machine learning models alongside their current operational model outputs. Other NOAA officials we spoke with confirmed that user feedback is challenging to acquire.
- **Interdisciplinary disconnect.** Collaborating across disciplines can be challenging. For example, an expert we spoke with told us that the different review cycles for publication across disciplines hinder interdisciplinary collaborations because the timing is not aligned. Further, the type of journal researchers publish in has an impact on who sees their work. For example, a statistician might develop a new machine learning-based method for predicting wildfire spread and choose to publish it in a statistics-related journal. It is unlikely that researchers doing applied wildfire work would then read that journal and learn about the new method because it is outside of their field.
- **Competing objectives within organizations.** Competing objectives within an organization can hinder the use of available resources. For example, Forest Service officials we spoke with told

us that differences between wildfire management goals and research goals can sometimes create missed opportunities for data sharing and research development. However, when researchers and practitioners collaborate, data and processes that align across goals can be designed.

- **Funding and technology transfer.** According to NOAA officials, a lack of consistent funding and a historical lack of a formalized research-to-operations pipeline create barriers to collaboration. NOAA officials at the Global Systems Laboratory said that development of a model can stall because different stages of R&D are funded by different organizations or parts of institutions. These officials told us that advancing technology through this process can require understanding when funding will be available across multiple organizations. Further, according to these officials, it generally takes 10 years to transition a new tool through the research stages to operations, and the process is different across labs, even within NOAA. However, NOAA officials told us that the agency has put processes in place for developing transition plans that are roadmaps for the transition of R&D into operations and other uses. Transition plans aid in budget planning and help to accelerate research to operations. NOAA also has technology transfer specialists to facilitate the protection of intellectual property across all NOAA labs and programs.

## 3.4 Workforce and resource needs create barriers to uptake of machine learning

Knowledge gaps exist among machine learning experts and those with earth science domain expertise, which has hindered the development and applications of machine learning techniques to improve natural hazard forecasting. Agency officials, academic researchers, and industry representatives we spoke with told us that both machine learning expertise and domain expertise are needed for developing machine learning technologies for natural hazard modeling. However, the skills of various participants in the field are sometimes misaligned with professional needs. For example:

- **Developer knowledge of hazard modeling.** Academic researchers noted that companies developing machine learning models for natural hazards vary in their emphasis on knowledge of the hazard versus knowledge of machine learning modeling, and some companies may be developing machine learning models without domain expertise. For example, academic researchers told us that some companies create models that do not capture or predict certain key factors important for predicting fire behavior, such as wind. Similarly, researchers with the Cooperative Institute for Research in the Atmosphere told us they have seen private sector machine learning models in which several physical laws were not taken into consideration, such as the fact that hurricanes have a low-pressure center.

- **Educational requirements for government scientists.** NOAA officials said that educational requirements for government scientists are outdated and do not prepare a machine learning-ready workforce. For example, the U.S. Office of Personnel Management (OPM) General Schedule Qualification Standards for a meteorologist do not require any computer science or machine learning-specific coursework, and only suggest it as optional coursework. This gap amplifies challenges related to trusting machine learning model output. Industry representatives told us that universities and the government could do more to better prepare the workforce by training them with machine learning-related skills.
- **Few professional developmental opportunities in machine learning.** Some agency officials said that there are few professional developmental opportunities for training government scientists to integrate environmental data with machine learning workflows. For example, U.S. Army Corps of Engineers officials stated that they would like to use machine learning for flood modeling, but that extensive quality assurance and control evaluations, testing, and certification, as well as support and resources from management, are needed.
- **Federal hiring and retention.** Agency officials told us that the general federal pay scale limitations and workforce location requirements makes it difficult to compete with other sectors to hire and retain highly qualified staff with machine learning experience. For example, according to officials, the salary they can offer for certain positions is generally lower than the salary for the same position in the private sector. To address this challenge, agency officials focus on hiring those with domain expertise and then teaching them computer science and machine learning skills, as agency officials consider this easier than teaching natural hazard science to a computer scientist.

Resource needs also create challenges for federal agency use of machine learning. For example:

- **High costs for training machine learning models.** Agency officials said that some machine learning technologies, such as emulators, require extensive computational resources for training that generally require the use of GPUs (graphical processing unit) rather than CPUs (central processing units). GPUs are ideal for training certain advanced machine learning algorithms but can be approximately 10 times more expensive. Private sector companies can have better access to GPUs than some agencies due to resource disparities. Some government agencies have created partnerships with private companies in order to leverage such resources. However, these partnerships can be challenging and limited for both sides due to agency preferences for their own proven technologies and variance in the level of private companies' domain expertise.
- **Computational resources for running machine learning models.** Limited computational resources create barriers for government agencies interested in leveraging machine learning technologies. Agency officials said that using machine learning models incurs massive up-front computing costs that can be prohibitive or otherwise strain computing infrastructure. For example, NASA

officials told is that the agency has a variety of computing assets that machine learning models can use, but some models require more computing power than is readily available. Forest Service officials told us they had to completely abandon a project because they did not have the computational resources needed to run a machine learning model they were developing.

**Ethical considerations for machine learning in natural hazard modeling**

Private and public organizations have developed guidance on incorporating ethical principles such as fairness, accountability, transparency, and safety for machine learning use. For example, in June 2021, GAO published *Artificial Intelligence*: *An Accountability Framework for Federal Agencies and Other Entities*, identifying key practices for entities involved in the design, development, deployment, and monitoring of AI systems. These key practices, in part, intend to help entities mitigate negative risks that may arise from bias.

Bias is not unique to machine learning systems, and achieving zero bias or risk in an AI system is not possible. The use of machine learning has the potential to amplify existing biases and concerns related to civil liberties, ethics, and social disparities. Below are several examples of how different biases intersect with ethics when using machine learning for natural hazard modeling:

- **Technical bias** can develop in the algorithmic process of the AI system via system design choices and can reflect in the model output. For example, a machine learning algorithm designed to value or prioritize reducing economic loss from a flood event may signal emergency personnel to respond in areas that are more affluent or may prioritize mitigating damage to infrastructure. While the algorithm might perform well for its designed intent, it could lead to loss of human life if the flood event had a greater impact on less affluent areas. Responsible use of machine learning includes the prioritization of safety risks that have the potential of serious injury or death.
- **Interpretive bias** comes from the implicit cognitive biases of individuals or groups who develop or use AI systems. For example, NOAA officials told us it is difficult to intuitively understand the machine learning model's process, which can compromise perceptions of risk. Some forecasters use information from both traditional and machine learning models but may ignore a more accurate machine learning model output due to a lack of familiarity. Creating explainable and interpretable machine learning can provide insights into the functionality of the system and increase trust in its outputs, reducing the risk of harms related to interpretive bias.
- **Organizational bias** stems from the procedures and practices of institutions that develop and use machine learning systems. It affects how organizations or research teams are structured and the control or design of decision-making processes. Some federal agency officials we spoke with told us that there is no standardized or codified process for assessing machine learning models during R&D or for transitioning them to operations. Organizations that use machine learning without robust procedures that fully assess and manage potential risks of the AI system can cause harm by deploying systems that are inaccurate or unreliable.

GAO, Artificial Intelligence*: An Accountability Framework for Federal Agencies and Other Entities*, GAO-21-519SP (Washington, D.C.: June 2021).

National Institute of Standards and Technology (NIST), *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, Special Publication 1270 (March 2022).

NIST, *Artificial Intelligence Risk Management Framework* (AI RMF 1.0), NIST AI 100-1 (January 2023).

Source: GAO analysis of cited literature. | GAO-24-106213

# 4 Policy Options to Help Enhance Benefits or Address Challenges of Using Machine Learning Technologies for Natural Hazard Modeling

We developed five policy options that policymakers—legislative bodies, government agencies, academia, industry, and other groups—could consider taking to help address challenges related to the development, implementation, and use of machine learning for modeling severe storms, hurricanes, floods, and wildfires. This is not an exhaustive list of policy options. We intend for these options to provide policymakers with a broader base of information for decision-making.

## 4.1 Facilitate improved data collection, sharing, and use

**Challenges addressed**: Data gaps, bias, and incompatibility

Government, academic, and private sector policymakers could attempt to mitigate data limitations by expanding data collection efforts, improving existing data sets, and facilitating the sharing of available data.

According to GAO's AI Accountability Framework, data used to train, test, and validate AI systems should be of sufficient quality and appropriate for the intended purpose to ensure the system produces consistent and accurate results.[39]

**Potential implementation approaches**

- Government policymakers could expand the use of existing observational data and infrastructure to address data gaps and collect more, higher-resolution data.
- Government policymakers could explore opportunities and tradeoffs associated with the expansion of observational infrastructure, such as radar stations or flood gauges, where data gaps are most significant or where additional data would be most useful, such as in rural or mountainous areas.
- Government policymakers could expand researchers' access to archived satellite data from defense and other government sources, as well as from some private sources that are currently unavailable to researchers, where appropriate.
- Government, academic, and private sector policymakers could ensure there are guidelines to make data sets AI ready and maintain them for both public- and private-sector researchers. This effort could build, in whole or in part, on the existing federal partnership to develop an AI-ready checklist for data sets (see section 3.1.2).

**Opportunities**

- Efforts to address data gaps within data sets can improve machine learning model performance.
- Accessible data from a larger number of sources would improve the ability of researchers and groups to develop and test machine learning technologies.

[39] GAO-21-519SP.

- Adopting AI-ready data standards could reduce resources needed to curate data, which could allow researchers to spend more time modeling and less time preparing data.

**Considerations**

- Expanding observational infrastructure can be expensive and could divert limited resources away from other efforts.
- Agencies need to weigh the benefits of facilitating greater data sharing for natural hazard modeling purposes with ensuring protections for certain kinds of data. For example, making archived, defense-sensitive satellite data public could improve modeling outcomes but could also increase certain national security risks. Furthermore, use of some existing observational data, such as from traffic cameras, may be viewed as unethical or unlawful surveillance.
- Risks related to data security and data privacy could increase as more kinds of data are shared more easily and in greater volume.
- Data standards may impact research and innovation negatively if they are too strict because they could constrain machine learning researchers and developers to a few, potentially suboptimal, approaches for addressing their particular research questions.

## 4.2 Expand machine learning education and training

**Challenges addressed**: Workforce and resource needs

Policymakers could adjust government requirements for certain science occupations and expand training opportunities to foster machine learning expertise within academia and the federal government.

According to GAO's AI Accountability Framework, recruiting, developing, and retaining personnel with multidisciplinary skills and experiences in design, development, deployment, assessment, and monitoring of AI systems should be a key practice for AI system governance. Furthermore, the framework highlights the importance of staff having the necessary training, resources, and domain expertise to fulfill their role.

**Potential implementation approaches**

- Government policymakers could update OPM professional education requirements for some government science positions to include machine learning-related coursework.
- Academic policymakers could adjust certain physical science education curricula to allow greater flexibility to pursue data science, advanced statistics, and machine learning-related coursework.
- Government policymakers could expand the capacity or number of centralized or federated learning and support centers that provide hands-on training to researchers and end-users who interact with machine learning technologies, increasing workforce capacities and developing relevant skill sets for current and potential government employees. Some agencies are currently invested in providing training and workforce development, such as through NOAA's Cooperative Science Centers.

**Opportunities**

- Updating OPM professional education requirements for relevant government science positions to include machine learning-related coursework would encourage academic institutions to adjust educational curricula accordingly and better prepare students to use machine learning in professional positions within government.
- More robust machine learning-related education and training can better prepare both researchers and end-users in fields such as meteorology and climatology to develop and interact with these technologies.

**Considerations**

- Education and training reforms may need to be repeatedly adjusted, as technological change in this space can be rapid and unpredictable.
- Establishing and expanding professional development and training opportunities throughout government may require substantial investment and might conflict with existing agency priorities or commitments.

## 4.3 Address hiring and retention barriers and certain resource shortfalls

**Challenges addressed**: Workforce and resource needs; limited coordination and collaboration

Government policymakers could address staffing and resource issues by providing workforce incentives and investing in public-private partnerships.

According to GAO's AI Accountability Framework, recruiting, developing, and retaining personnel with experience in the design, development, deployment, assessment, and monitoring of AI systems should be a key practice for AI system governance. Furthermore, the framework highlights how AI systems require appropriate technologies to ensure intended goals and objectives are achieved.

**Potential implementation approaches**

- Government policymakers could address general schedule (GS) pay scale limitations. This could include adopting a special salary rate, for certain scientist positions that include machine learning expertise, such as data scientists, or implementing additional workforce flexibilities to attract and retain technical talent, as appropriate for different agencies.
- Government and private sector policymakers could expand the use of public-private partnerships through established legal mechanisms such as Cooperative Research and Development Agreements (CRADA) and authorities

similar to NASA's Space Act Agreements.[40]

**Opportunities**

- Providing workforce incentives to government employees for machine learning development could allow agencies to recruit new talent, reduce turnover, and enhance workforce capacities within the federal government.
- Expanding public-private partnerships might help agencies overcome computational resource shortfalls without the need for larger investments in computing infrastructure and hardware. For example, NOAA's Open Data Dissemination Program used CRADAs with major U.S. companies to migrate over 150 government data sets to collaborators' systems, dramatically increasing access to and use of such data, and storing the data next to cloud service providers' computational resources.[41]
- Expanding public-private partnerships helps industry draw upon government expertise.

**Considerations**

- Increasing salary limits for some GS employees would require increases to agency budgets or cuts to other budget items. According to agency officials, such increases allow agencies to hire at a higher salary but may lead to them being able to hire fewer people overall.
- Expanding the use of public-private partnerships, rather than investing directly in government capacities, could magnify resource disparities between government and private industry and increase government reliance on the private sector for computing-related services.
- Public-private partnerships that entail hosting government data on private sector storage systems may pose security risks for some sensitive information.

## 4.4 Take steps to mitigate bias and foster trust in data and machine learning models

**Challenges addressed**: Data gaps, bias, and incompatibility; ethical considerations for use of machine learning in natural hazard modeling; trust in machine learning

Government policymakers could address issues related to bias and transparency in

[40]NASA has specific "other transactions" authority under the National Aeronautics and Space Act, 51 U.S.C. § 20113(e), to enter into agreements with diverse groups of people and organizations, both in the private and public sector, in order to meet wide-ranging NASA mission and program requirements and objectives. See also National Aeronautics and Space Administration Transition Authorization Act of 2017, Pub. L. No. 115-10, § 841, 131 Stat. 72 (2017), 51 U.S.C. § 20113 note.

[41]NOAA officials stated that providing open cloud access to data alongside computational resources improves efficiency and scalability. According to NOAA, this data dissemination program became operational in 2020 with signed contracts with three cloud service providers. NOAA's Open Data Dissemination Program provides open access to hundreds of datasets from across NOAA, including atmospheric, oceanic, fisheries, weather, climate, surface observations, emergency response imagery, forecasting products, and near real-time data.

machine learning algorithm development through targeted project governance.

According to GAO's AI Accountability Framework, robust AI-specific risk management practices can ensure governance at the organizational level, such as identification of potential biases and societal concerns resulting from AI systems. Furthermore, one of the key governance practices identified by the framework is promoting transparency by enabling external stakeholders to access information related to AI systems. The framework also notes that, to demonstrate quality and reliability, agencies should document how training data is collected, prepared, and updated.

**Potential implementation approaches**

- Government, academic, and private sector policymakers could establish efforts, including studies and evaluation, to better understand and mitigate various forms of bias in recent and historic data sets.
- Government, academic, and private sector policymakers could support inclusion of a diverse array of stakeholders in data evaluation, model development, and testing.
- Government, academic, and private sector policymakers could develop guidelines or best practices for documenting and publicly reporting training data choices as well as data provenance[42], which can enable third-party assessment of machine learning models.

**Opportunities**

- Efforts to address bias within machine learning systems can reduce the likelihood of models negatively impacting certain communities, such as rural or poor communities. For example, observational data sets based on human reports of hail or tornadoes are inherently biased towards areas with higher populations and might create statistical bias in machine learning models to over-predict corresponding weather events in urban areas and under-predict them for rural communities.
- Acquiring diverse stakeholder perspectives throughout machine learning models' life cycles can help reduce certain types of bias in data and models.
- Fostering machine learning model transparency could improve end-user and decision-maker trust in these technologies, increasing acceptance for government and public use.

**Considerations**

- Embedding efforts to address bias throughout the model lifecycle may increase model costs and slow model development.

## 4.5 Maintain status quo efforts

Government policymakers could maintain existing policy efforts and organizational

[42] In this context, we use the term "data provenance" to refer to documentation of data, such as unique identifiers for data sets, prior use for machine learning, and contact information for appropriate subject matter experts for the data.

structures, along with existing strategic plans and agency commitments.

The current U.S. national strategy on AI is largely established through laws and executive orders, including, but not limited to:

- **Advancing American AI Act.** It was enacted most recently, and its intent is to encourage agency AI-related programs and initiatives; enhance the ability of the federal government to translate research advances into AI applications to modernize systems and assist agency leaders in fulfilling their missions; promote adoption of modernized business practices and advanced technologies across the federal government; and test and harness applied AI to enhance mission effectiveness.[43]

- **National Artificial Intelligence Initiative Act of 2020.** This law includes the initiative that directs the President and agency heads to sustain support for AI research and development, support AI education and workforce training programs, support interdisciplinary research and education programs, plan and coordinate federal interagency AI activities, conduct outreach to diverse stakeholders, support a network of AI research institute, and support opportunities for international cooperation with strategic allies on AI-related issues.[44]

- **AI in Government Act of 2020.** This law will codify the establishment of the U.S. General Services Administration's (GSA) AI Center of Excellence, directs the Office of Management and Budget to provide guidance for use of AI, and directs the OPM to update the occupational series for AI for federal employees. [45]

- **Executive Order No. 14110: *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*.** This Executive Order establishes new policies and principles for AI in several categories including: safety and security; advancing equity and civil rights; protecting privacy and civil liberties; innovation; and advancing federal government use of AI. In addition, it calls on several agencies to develop standards, tools, and tests to help ensure that AI systems are safe, secure, and trustworthy.[46]

- **Executive Order No. 13960: *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*.** This Executive Order, among other things, establishes a common policy for implementing principles related to lawfulness, performance, accuracy, reliability, safety, resilience,

[43]The Advancing American AI Act was enacted as part of the James M. Inhofe National Defense Authorization Act for Fiscal Year 2023, Pub. L. No. 117-263, § 7222, 136 Stat. 3668, Division G, tit. LXXII, subtit. B (2022), to be codified at 40 U.S.C. § 11301 note.

[44]The National Artificial Intelligence Initiative Act of 2020 was enacted as Division E of the William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021 (NDAA FY21), Pub. L. No. 116-283, 134 Stat. 3388 (2020) and has been codified at 15 U.S.C. §§ 9401-9415.

[45]The AI in Government Act of 2020 was enacted as Division U, tit. I of the Consolidated Appropriations Act, 2021, Pub. L. No. 116-260, 134 Stat. 1182, 2286 (2020), to be codified at 40 U.S.C. § 11301 note.

[46]Exec. Order No. 14110, Safe Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 30, 2023), 88 Fed. Reg. 75,191 (Nov. 1, 2023).

understandability, responsibility, transparency, accountability, and monitoring. In addition, it directs agencies to catalog their AI use cases, and calls on the GSA and OPM to expand AI expertise at agencies across government. [47]

- **Executive Order No. 13859: *Maintaining American Leadership in Artificial Intelligence.*** This Executive Order directs federal agencies to, among other actions, promote sustained investment in AI research and development in collaboration with non-federal entities, enhance access to federal data and computing resources, reduce barriers to the use of AI technologies, ensure that technical standards minimize vulnerabilities, train the next generation of American AI researchers, and develop action plans to protect American advantages in critical AI technology development. [48]

**Opportunities**

- Some agency efforts are already underway to address the specific challenges of using machine learning in natural hazard modeling. If these continue and agencies with natural hazard modeling responsibilities implement them, it could help address many of the challenges we describe and minimize potential negative outcomes of further policy interventions (as described in the considerations for other policy options above).

**Considerations**

- Although some status quo efforts direct agencies to take relevant actions that might address some challenges enumerated in this report, the extent to which agencies are meeting these commitments is largely unclear.
- Status quo efforts as defined through laws and Executive Orders relevant to the U.S. national strategy on AI are general and may not fully address the specific challenges identified in this report.

[47]Exec. Order No. 13960, *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government* (Dec. 3, 2020), 85 Fed. Reg. 78,939 (Dec. 8, 2020).

[48]Exec. Order No. 13859, *Maintaining American Leadership in Artificial Intelligence* (Feb. 11, 2019), 84 Fed. Reg. 3,967 (Feb. 14, 2019).

## 5 Agency and Expert Comments

We provided a draft of this report to the Department of Commerce, the Department of Defense, the Department of Energy, the Department of Homeland Security, the Department of the Interior, the National Aeronautics and Space Administration, the National Science Foundation, and the U.S. Department of Agriculture with a request for technical comments. We received technical comments from the Department of Commerce, the Department of Defense, and the Department of Homeland Security, which we incorporated as appropriate.

---

We are sending copies of this report to the appropriate congressional committees, the relevant federal agencies, and other interested parties. This report will be available at no charge on the GAO website at https://www.gao.gov.

If you or your staff members have any questions about this report, please contact Brian Bothwell at (202) 512-6888 or BothwellB@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in Appendix III.

Brian Bothwell, MS
Director,
Science, Technology Assessment, and Analytics

## *List of Addressees*

**The Honorable Debbie Stabenow**
Chairwoman
Committee on Agriculture, Nutrition, and Forestry
United States Senate

**The Honorable Ted Cruz**
Ranking Member
Committee on Commerce, Science, and Transportation
United States Senate

**The Honorable John Barrasso**
Ranking Member
Committee on Energy and Natural Resources
United States Senate

**The Honorable Gary C. Peters**
Chairman
Committee on Homeland Security and Governmental Affairs
United States Senate

**The Honorable Cathy McMorris Rodgers**
Chair
Committee on Energy and Commerce
House of Representatives

**The Honorable Jamie Raskin**
Ranking Member
Committee on Oversight and Accountability
House of Representatives

**The Honorable Frank D. Lucas**
Chairman
Committee on Science, Space, and Technology
House of Representatives

# Appendix I: Objectives, Scope, and Methodology

## Objectives, Scope, and Methodology

We describe our scope and methodology for addressing the four objectives outlined below:

1. How has the application of artificial intelligence (AI) affected modeling capabilities for severe storms, hurricanes, floods, and wildfires?

2. How could modeling capabilities for severe storms, hurricanes, floods, and wildfires be affected by future applications of AI?

3. What challenges exist with regard to the development and application of AI in modeling capabilities for severe storms, hurricanes, floods, and wildfires?

4. What policy options might help address the potential challenges related to the development and application of AI in modeling capabilities for severe storms, hurricanes, floods, and wildfires?

To address all four research objectives, we reviewed available and developing machine learning uses that government or the private sector could use to model severe storms, hurricanes, floods, and wildfires. To do so, we reviewed key reports and scientific literature describing current and developing machine learning uses; interviewed a variety of stakeholders, including agency officials, industry members, academic researchers, and professional associations; and conducted an expert meeting in conjunction with the National Academies of Sciences, Engineering, and Medicine.

### Limitations to scope

Our scope is limited to models that incorporate or use AI techniques (specifically, machine learning) to enhance model input, output, or performance, or to predict the aforementioned hazards. We focused on machine learning because, in recent years, environmental researchers and forecasters have focused on machine learning, as opposed to other types of AI, in recent years. We primarily focused on modeling for civilian use. We did not examine risk models, seasonal forecasting, long-term climate modeling, or hazard mitigation efforts. We focused our review on machine learning used to model the natural hazards in scope, including severe storms, hurricanes, floods, and wildfires. We excluded machine learning used for natural hazard modeling in other contexts, such global climate modeling, as well as catastrophe models used for insurance company risk management. Machine learning uses discussed are examples and not an exhaustive list of all machine learning uses for natural hazard modeling. We did not assess all available or developing machine learning uses. We selected narrative examples to demonstrate the breadth of machine learning uses for modeling the natural hazards in scope.

### Literature Search

In the course of our work we conducted one literature review. To establish background and identify machine learning models appropriate to our scope and their associated challenges, we reviewed key articles from the scientific literature. To support the four objectives we focused on the use of AI and machine learning in environmental modeling. We used search terms such as “artificial intelligence,” “machine learning,” “weather,” “meteorology,” and “forecasting,” and narrowed our search to articles published within the last 5 years. For this search, results could originate from scholarly or peer-reviewed material, government reports, conference papers, trade or industry articles, publications by associations, nonprofit organizations, or think tank, books, and legal materials, such as laws and Executive Orders. We selected the most relevant articles for further review based on our objectives.

### Interviews

We interviewed key experts and stakeholders in the field of machine learning for natural hazard modeling, including:

- Department of Commerce, National Oceanic and Atmospheric Administration, and National Weather Service; Department of Energy; Department of the Interior, U.S. Geological Survey; Department of Defense, Army Corps of Engineers; Department of Agriculture, United States Forest Service; Department of Homeland Security, Federal Emergency Management Agency; National Aeronautics and Space Administration; and the National Science Foundation;
- 10 academic researchers and research groups;
- two professional organizations;
- and one private firm.

Because this is a non-generalizable sample of the stakeholders involved in using machine learning for natural hazard modeling, the results of our interviews are illustrative and represent important perspectives, but are not generalizable.

### Expert meeting

We collaborated with the National Academies of Sciences, Engineering, and Medicine to convene a 3-day meeting of 17 experts on current and emerging machine learning tools for use in natural hazard modeling. We worked with staff from the National Academies of Sciences, Engineering, and Medicine to identify experts from a range of stakeholder groups including federal agencies, academia, and industry, with expertise covering all significant areas of our review, including individuals with research or operational expertise in using machine learning tools in the modeling of natural hazards.[49] We evaluated the experts for any conflicts of interest. A conflict of interest was any current financial or other interest (such as an organizational position) that might conflict with the service of an individual because it could (1) impair objectivity or (2) create an unfair competitive advantage for any person or organization. The 17 experts were

[49] This meeting of experts was planned and convened with the assistance of the National Academies of Sciences, Engineering, and Medicine to better ensure that a breadth of expertise was brought to bear in its preparation, however all final decisions regarding meeting substance and expert participation was the responsibility of GAO. Any conclusions and recommendations in GAO reports are solely those of the GAO.

determined to be free of reported conflicts of interest, and the group as a whole was determined to not have any inappropriate biases. (See app. II for a list of these experts and their affiliations.) The comments of these experts generally represented the views of the experts themselves and not the agency, university, or company with which they were affiliated, and are not generalizable to the views of others in the field.

We divided the 3-day meeting into seven moderated discussion sessions: (1) definitions, scope, and background for using machine learning in natural hazard modeling; (2) benefits, opportunities and drawbacks of using machine learning in natural hazard modeling; (3) future applications of using machine learning in natural hazard modeling; (4) technical challenges of using machine learning in natural hazard modeling; (5) non-technical challenges if using machine learning in natural hazard modeling; (6) an overview of the benefits and challenges discussed above and (7) policy options for using machine learning in natural hazard modeling (via breakout rooms and group discussion). Each session featured an open discussion among all meeting participants, most of which were based on key questions we provided. The meeting was transcribed to ensure that we accurately captured the experts' statements. After the meeting, we reviewed the transcripts to characterize their responses and to inform our understanding of all three researchable objectives. Following the meeting, we continued to seek the experts' advice to clarify and expand on what we had heard. Consistent with our quality assurance framework, we provided the experts with a draft of our report and solicited their feedback, which we incorporated as appropriate.

### Policy options

We intend policy options to provide policymakers with a broader base of information for decision-making.[50] The options are neither recommendations to federal agencies nor matters for congressional consideration. They are also not listed in any specific rank or order. We are not suggesting that they be done individually or combined in any particular fashion. Additionally, we did not conduct work to assess how effective the options may be, and we express no view regarding the extent to which legal changes would be needed to implement them.

We limited the policy options included in this report to those that met the policy objective and fell within the report scope. We present five policy options in response to the challenges identified during our work and discuss potential opportunities and considerations of each. While we present options to address the major challenges we identified, the options are not intended to be inclusive of all potential policy options.

To develop the policy options, we prepared a list of potential policy ideas (139 in total, including the status quo) based on our literature search, stakeholder interviews, and expert meeting. We removed ideas that were not likely to address the challenge or did not fit into the overall scope of our work. We grouped the remaining ideas based on

[50]Policymakers is a broad term including, for example, Congress, federal agencies, state and local governments, academic and research institutions, and industry.

themes (e.g., data collection). We combined those that (1) were duplicative, (2) could be subsumed into a higher-level policy option, or (3) were examples of how to implement a policy option rather than the option itself.

We conducted our work from August 2022 through December 2023 in accordance with all sections of GAO's Quality Assurance Framework that are relevant to technology assessments. The framework requires that we plan and perform the engagement to obtain sufficient and appropriate evidence to meet our stated objectives and to discuss any limitations to our work. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

## Appendix II: Expert Participation

We collaborated with the National Academies of Science, Engineering, and Medicine to convene a 3-day virtual meeting (May 2-4, 2023) of experts to inform our work on the use of machine learning for the prediction and forecast of severe storms, hurricanes, floods, and wildfires. The experts who participated in this meeting are listed below. Many of these experts provided additional assistance after the virtual meeting, such as by responding to follow-up questions via email and reviewing our draft report for accuracy and provided written comments.

**Ann Bostrom, PhD**

Weyerhaeuser Endowed Professor in Environmental Policy

Evans School of Public Policy & Governance

University of Washington, Seattle

**David Danks, PhD**

Professor of Data Science & Philosophy

University of California, San Diego

**Amy DeCastro**

Project Scientist

Research Applications Laboratory

National Center for Atmospheric Research

**Jenni Evans, PhD**

Director, Institute for Computational and Data Sciences

Professor, Meteorology and Atmospheric Science

Penn State University

**Sarvesh Garimella, PhD**

Chief Scientist

MyRadar / ACME AtronOmatic, LLC

**Michael Gollner, PhD**

Associate Professor and Deb Faculty Fellow

Department of Mechanical Engineering

University of California, Berkeley

**Kimberly Hoogewind, PhD**

Research Scientist

Cooperative Institute for Severe and High-Impact Weather Research and Operations

University of Oklahoma and NOAA National Severe Storms Laboratory

**Jonathan Koh, PhD**

Post-doctoral Researcher

Oeschger Centre for Climate Change Research

University of Bern

**Amy McGovern, PhD**

Lloyd G. and Joyce Austin Presidential Professor

School of Computer Science and School of Meteorology

University of Oklahoma

**Norberto Nadal-Caraballo, PhD**

Senior Research Engineer

Coastal Hazards Group

U.S. Army Engineer R&D Center

**Grey Nearing, PhD**

Senior Research Scientist

Google Research

**Christopher “Kit” O’Connor, PhD**
Research Ecologist
Wildfire Risk Management Science Team
Rocky Mountain Research Station
U.S. Forest Service

**Jaideep Pathak, PhD**
Research Scientist
NVIDIA Corporation

**Corey Potvin, PhD**
Research Meteorologist
National Severe Storms Laboratory
NOAA

**Vidya Samadi, PhD**
Assistant Professor, Water Resources Engineering
Director, Hydrosystem and Hydroinformatics Research Group
Clemson University

**Andrea Schumacher**
Project Scientist
Weather Risks and Decisions in Society
National Center for Atmospheric Research Program

**Katie Wilson, PhD**
Associate Policy Researcher
Homeland Security Operational Analysis Center
RAND Corporation

# Appendix III: GAO Contacts and Staff Acknowledgments

## GAO contact

**Brian Bothwell, MS**, Director, Science, Technology Assessment, and Analytics (STAA), at (202) 512-6888 or BothwellB@gao.gov

## Staff acknowledgments

In addition to the contact named above, the following staff made key contributions to this report:

**Laura Holliday, MS**, Assistant Director

**Jim Rice, PhD**, Analyst-in-Charge and Senior Physical Scientist

**Atiya Siddiqi, MS**, Senior Analyst

**Matthew Holly, PhD**, Analyst

**Holley Lawrence, MS**, Physical Scientist

**Nitin S. Panwar**, Computer Scientist

**Katherine R. Shlepr, PhD**, Biological Scientist

**Jehan Chase, JD,** Senior Attorney

**Ryan Han**, Visual Communications Analyst

**Ben Shouse, MS,** Communications Analyst

## These staff also contributed to this work:

Xiang Bi, PhD, Senior Economist

Jonathan Dent, Assistant Director

Andrew Stavisky, PhD, Assistant Director

Jessica Steele, Assistant Director

## GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

## Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's website (https://www.gao.gov). Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to https://www.gao.gov and select "E-mail Updates."

## Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, https://www.gao.gov/ordering.htm.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or
TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

## Connect with GAO

Connect with GAO on Facebook, Flickr, Twitter, and YouTube.
Subscribe to our RSS Feeds or E-mail Updates.
Listen to our Podcasts and read The Watchblog.
Visit GAO on the web at https://www.gao.gov.

## To Report Fraud, Waste, and Abuse in Federal Programs

Contact: Website: https://www.gao.gov/fraudnet/fraudnet.htm
Automated answering system: (800) 424-5454 or (202) 512-7470

## Congressional Relations

A. Nicole Clowers, Managing Director, ClowersA@gao.gov, (202) 512-4400,
U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548

## Public Affairs

Chuck Young, Managing Director, YoungC1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149, Washington, DC 20548

## Strategic Planning and External Liaison

Stephen Sanford, Managing Director, spel@gao.gov, (202) 512-9715
U.S. Government Accountability Office, 441 G Street NW, Room 7B37N, Washington, DC 20548