NOTE 3

Different Sources of Data and Where to Find Them

Climate and disaster risk financing (CDRF) is a data-hungry field. As should be clear from the notes in this series, data are critical to understanding risk, making effective decisions, and developing products that meet the needs of vulnerable populations, businesses, and governments. The data needs for any given activity often depend on the specific context and require extensive input from the government and relevant country experts. An in-depth guide to the data and expertise required for different DRF activities can be found in XXXXXX. This note focuses on common global disaster impact data and provides a summary of the data used in catastrophe risk modeling.

Data on Disaster Impact and Extent

Data on the impacts of disaster are frequently gathered and collated by a range of actors, from governments to scientists to international organizations. The methodologies for the development of these data vary, as do their consistency and coverage; but they can still be an excellent first source of information to inform or validate analytics.

Impact Data

Some of the main sources of impact data include the following:

- **Post Disaster Needs Assessment (PDNA):**
  A government-led process to determine the physical damage and associated losses from a disaster
- **Global Rapid Post-Disaster Damage Estimation (GRADE) report:**
  An initial estimation of the physical damage after an event that uses any and all available data and modeling
- **ReliefWeb:**
  Source of humanitarian information on disasters, including situation reports, evaluations, guidelines, assessments, and impact maps
- **Desinventar:**
  A disaster loss database compiled by the United Nations Office for Disaster Risk Reduction (UNDRR)
- **EM-DAT:**
  Disaster loss database with information from 1900 to the present day, compiled from a mixture of sources (UN agencies, nongovernmental organizations, reinsurance companies, and others)

FEWSNET (Famine Early Warning System Network), IPC (Integrated Food Security Phase Classification), and Cadre Harmonisé:

Entities that provide regular quantitative and qualitative assessments and forecasts of food insecurity conditions for at-risk countries across the world

GDACS (Global Disaster Alert and Coordination System) and WFP ADAM (World Food Programme Advanced Disaster Analysis and Mapping:

Data on occurrence and extent of disaster events, covering a range of perils

FAOSTAT (Food and Agriculture Organization Statistics):

Information on crop yields and production in countries across the world

WFP VAM (World Food Programme Vulnerability Analysis and Mapping:

A range of data to inform food insecurity assessments across the world

UNOCHA (UN Office for the Coordination of Humanitarian Affairs):

Source of high-level figures on human disaster impacts and humanitarian needs

Hazard Occurrence and Extent Data

Sources of data on hazard occurrence and extent, grouped by hazard type, include the following:

Flood

- ***Floodlist:*** Reports on all the major flood events from around the world
- ***Copernicus Global Flood Monitoring (GFM):*** Provides continuous monitoring of global flood events
- ***Dartmouth Flood Observatory:*** Provides global flood monitoring in near-real time

Drought

- ***World Bank NGDI (Next Generation Drought Index) Dashboard:*** Provides up-to-date drought data from commonly used remote sensing indexes; currently piloted in the Horn of Africa, being expanded across Africa
- ***FEWSNET:*** Provides remote sensing information on hydrometeorological conditions

Earthquake

- ***USGS (United States Geological Survey):*** Provides up-to-date reporting on the occurrence and (hazard) extent of earthquake events around the world

Tropical cyclone

- ***NOAA (National Oceanic and Atmospheric Administration) National Hurricane Center:*** Provides up-to-date reporting on the occurrence and (hazard) extent of tropical cyclone events around the world
- ***Regional Specialized Meteorological Centres and Tropical Cyclone Warning Centres:*** Tasked by the World Meteorological Organization with providing up-to-date meteorological information on tropical cyclones globally
- ***JTWC (Joint Typhoon Warning Center):*** Issues tropical cyclone warnings for parts of the Pacific Ocean and for the Indian Ocean to US military and government agencies

Catastrophe Risk Modeling Data

As explained in Note 2, a catastrophe risk modeling framework is made up of discrete modules that rely on specific data inputs. These data inputs are needed to develop and validate the calculations in each module and ensure that the end output is robust. To this end, each module should be developed using a combination of global and local data, and ideally each should provide high levels of transparency into how the data were incorporated.

Some relevant data sets for three of the modules and for model validation are outlined below. Note that these data aren't typically used in their raw format, but rather are used as inputs (in combination with other data sets and subject to various statistical and/or physical modeling methods) and/or as a means of calibrating and validating the models.

Hazard Module

For the hazard module, the data required depend on the peril being modeled (e.g., flood, tropical cyclone, etc.). Some initial data sets are outlined below for each peril, but the hazard module may require other information in addition to that listed.

Flood

- ***Precipitation data:*** Derived from either local weather stations or global gridded precipitation data sets derived from satellites
- ***Streamflow data:*** Derived from river gauges that record discharge, water depth, and other parameters
- ***Topographical data characterizing ground elevation:*** Used to determine how water flows through the landscape
- ***Data on the river network within the area of interest***
- ***Hydrogeomorphological parameters:*** Describe parameters of rivers (e.g., width, roughness) and floodplains, among others

Tropical cyclone

- ***Historical tropical cyclone tracks (such as those available from NOAA's International Best Track Archive for Climate Stewardship [IBTrACS]):*** Can be used to determine landfall frequency, windspeed frequency, and genesis location and frequency (i.e., number of events per month)
- ***A pressure-wind relationship***
- ***Meteorological data on characteristics of the atmosphere:*** Data on temperature, pressure, wind speed, rainfall, etc.
- ***Sea surface temperature data***
- ***Mean sea-level pressure***

Earthquake

- ***Historical earthquake catalog:*** Provides spatial and temporal data on earthquakes obtained from catalogs, seismological databases, and published sources
- ***Fault maps:*** Define fault locations and associated seismic source boundaries
- ***Strain rate of different faults***
- ***Soil maps:*** Classify the soil conditions across the area of interest to determine the amplification of the ground motion.

Drought

- ***Climate variables:*** Data on rainfall, surface/root-zone soil moisture, evapotranspiration, land surface temperature, etc.; derived from satellite observations, models, or weather stations
- ***Satellite-derived proxies for vegetation health:*** E.g., Normalized Vegetation Difference Index
- ***Hydrological data:*** Satellite-derived/modeled groundwater storage, streamflow data from river gauges, and reservoir levels (hydrological drought)
- ***Meteorological forecasts***

Exposure Module

As explained in Note 2, exposure data provide critical information on people and assets at risk (population, buildings, critical infrastructure, crops, etc.) and on attributes of people and assets that indicate potential vulnerability to disasters. Sources of data for the exposure module include the following:

Population data sets:
Various global population data sets provide geospatial data on population distributions and demographics at varying levels of granularity. Where available, these data sets are developed using census data for the lowest administrative level possible, and these data are then disaggregated to

increase granularity (e.g., to a 100 m x 100 m grid) using available geospatial data for the region/country (e.g., nighttime light data). Where census data are not available, then other survey data may be combined with other geospatial data sets to build a statistical model that estimates the population. It is important to be aware of differences between available population data sets when selecting a data set to use for risk modeling. Differences can include differences in underlying data sources, methods, and output resolutions.[1]

Examples of population data sets include WorldPop, Global Human Settlement (GHS-POP), and High Resolution Settlement Layer (HRSL).

Built environment data sets:
Data sets exist that are starting to provide detailed building footprint data globally. Building-level data can be supplemented by additional local data (e.g., census data) to improve their quality where necessary.

Depending on the specific peril and modeling approach, there will be different requirements for exposure data sets. A range of different attributes are collected in different data sets, including construction type, building height, etc. Please see the risk modeling notes in this series (Note 2, 2a, 2b, 2c and 2d) for more information about specific requirements.

An approach often used in developing countries is to develop aggregate built environment data. These are usually developed using a combination of census data and Earth observation data as an input to statistical analyses. They are then validated through field observations. These data sets provide a percentage of different building types, associated numbers, and building value within pre-determined grid cells (e.g., 100 m x 100 m).

Examples of built environment data sets include Google's Open Buildings, Overture Maps Foundation, OpenStreetMap, Ecopia AI, METEOR, and the Global Earthquake Model Foundation's Global Exposure Model.

Critical infrastructure:
Recently, partly in response to climate change, there has been much interest in efforts to increase the resilience of critical infrastructure globally. A range of global projects has enabled the accumulation of disparate data sets that identify different critical infrastructure globally and are a good source of information to complement local data.

Sources of data on critical infrastructure include (for example) the Global Infrastructure Resilience Index (GIRI), Coalition for Disaster Resilient Infrastructure (CDRI), and Resilient Planet Data Hub (RPDH). OpenStreetMap and Humanitarian Data Exchange are additional sources of relevant data.

Agriculture and crops:
Key data sets include land use and land cover, irrigation practices, and crop type classifications. In certain cases, information on soil types offers valuable additional insights. Agricultural census data provide additional information on crops, livestock, and farming practices, such as crop variety, livestock populations, farm sizes, and resource utilization. Together, these data sets help identify the exposure of agricultural systems to climate-related risks and inform vulnerability assessments.

Examples of agriculture and crop data include the European Space Agency's global land cover products, such as WorldCover; FAO's World Programme for the Census of Agriculture; and the Harmonized World Soil Database.

[1] UN Sustainable Development Solutions Network (SDSN), "Leaving No One Off the Map: A Guide for Gridded Population Data for Sustainable Development," 2020, https://files.unsdsn.org/Leaving%2Bno%2Bone%2Boff%2Bthe%2Bmap-4.pdf.

Vulnerability Module

Within disaster risk management, two types of vulnerability curves can be developed, empirical and analytical. (The two types can also be combined, considering all available data and models). Empirical curves are developed based on observations from past events, which can include both post-disaster damage assessment and claims experience (for insurance purposes). In contrast, analytical curves are developed through engineering analysis and judgment to estimate the damage a certain building may sustain from hazards of particular intensities. Empirical approaches rely on good-quality loss or damage observations along with accurate hazard intensity measurements. Analytical approaches can differ greatly from one another, and so care is needed to ensure these are adequately validated for the context in which they are applied.

For building vulnerabilities, there are globally available curves available, such as Joint Research Centre (JRC) flood curves and Emanual tropical cyclone wind curves[2]. However, these are not calibrated or validated for individual countries. Given that vulnerability can vary substantially between geographies (e.g., due to construction practices), curves will often need to be developed or adjusted for a given emerging market and developing economy (EMDE) setting. Depending on the vulnerability module setup, building vulnerability curves may also be adjusted using "vulnerability modifiers," which are factors that adjust the estimated vulnerability based on building characteristics such as construction type, age or construction date, number of stories, first-floor elevation, roof type, etc. The most important characteristics to consider vary between perils (e.g., roof design can be important for modeling wind-related damage, whereas first-floor elevation is particularly important for flood). The notes on catastrophe risk modelling in this series (Notes 2, 2a, 2b, 2c, and 2d) further discusses considerations relevant to vulnerability modules for individual perils.

[2] https://texmex.mit.edu/pub/emanuel/PAPERS/annrev.pdf;
https://twister.caps.ou.edu/METR3334/Papers/Emanuel-MeteorMono2018.pdf