Challenges of Integrating Physical Exposure and Human Impacts Data in Tropical Cyclone Studies – Reverse Outline 7/31/2022

**Introduction**

* Tropical cyclones frequently threaten coastal regions of the Eastern USA, resulting in damage to property, economic disruption, negative human health impacts, and loss of life, making them critical to study
* Tropical cyclones negatively impact human health (mental and physical) as well as create strains to local economies, so it is obvious that there are significant human impacts.
* Tropical cyclones represent a persistent threat in the future, and to build resilient communities that can last this threat, we require knowledge about their spatial and temporal impact of these storms.
* Data on the physical exposures of tropical cyclones and data on relevant human impacts come from different sources that are often at different spatial and temporal levels, making it difficult to integrate these datasets.
* This paper will explore why spatial and temporal misalignment occurs, what different scales are used, and the consequences of integrating human impacts and physical exposures datasets in tropical cyclone research.

**Physical Exposures**

* Physical exposure data often comes from vast monitoring networks across oceans and continents that is narrow in spatial and temporal resolution but large in geographic scope.

*Storm Tracks*

* The locations that the center of a storm moves through can be used to visualize the path the storm takes, which is known as the storm track.
* HURDAT2 is a data source of storm tracks.
* Distance from the storm track can be used to assign a county or zip code’s exposure to a tropical cyclone, and the distance threshold used can determine how much misclassification occurs.
* Another way to use the assign exposure using a storm track is to treat a county or zip code as exposed simply if the storm track moved through it.

*Wind Speed and Direction*

* Wind speed can be used to assign exposure to a tropical cyclone using measurements taken on the ground or above ground.
* Thresholds can be used to established a criteria for exposure (a zip code or county that experiences winds above a certain speed is considered exposed).
* Example of a study that used peak sustained wind greater than or equal to 34 knots when the cyclone was closest to the county as a way to assign exposure.

*Flooding and Storm Surges*

* Flooding due to intensive precipitation during a tropical cyclone on infrastructure, public safety, and economic strength of a community.
* One method to measure flooding is to use high water marks.
* Another method is to use stream gauges.
* Yet another method to assess exposure and damage due to tropical cyclones is to use satellite imagery, aerial photography, and remote sensing.

**Human Impacts of Tropical Cyclones**

* Human impacts data typically comes from census records, hospitalization records, vital statistics, and other sources that are typically aggregated at larger levels, usually based on geopolitical, cultural, and administrative boundaries.

*Human Health*

* Tropical cyclones have a number of human health impacts, and the data on these impacts come from various government and non-government public health agencies
* General health information can be obtained from the National Vital Statistics System, which takes data such as birth certificates from local health departments at the county and state level.
* Other sources of human health data come from the National Death Index, National Health Interview Survey, Nationally Notifiable Disease Surveillance System, Planned Parenthood Federation of America, Centers for Disease Control, and many others.

*Social and Economic Impacts*

* There are social and economic costs to tropical cyclones such as populations of displaced people and damage to local community infrastructure and businesses.
* US Census is a great source of information.
* Studying how employment and earnings change before and after a tropical cyclone is one method to assess damage.
* Tax returns allow researchers to observe economic changes with a point location since tax returns are typically linked to individuals with a known physical address.
* Insurance claims are a good source of information.

**Spatial and Temporal Scales and Misalignment**

* Human impacts data and physical exposures data come from different sources, often at different scales.
* The study question is an important factor in determining what spatial and/or temporal scales are used.
* The remainder of this section will focus on the different spatial and temporal scales commonly used in tropical cyclone impact studies.

***Spatial Scales***

* What we mean when we say “spatial scale” is the size of the geographic space we are collecting data from.
* It can be helpful to think of spatial data in terms of vectors or rasters. Vector data are discrete and well defined, often used for mapping human impacts.
* Raster data are gridded data that can display continuous exposure data, but also categorical human impacts data.
* This section will cover most common spatial scales, how physical exposure and human impact data are collected from these scales, and methods of integrating the physical exposure and human impacts data using these spatial scales.

*Vector Data*

* Vector data in tropical cyclone research is mostly point locations and polygons.

Point Location

* Point locations are the smallest spatial scale, and can be located using latitude and longitude lines.
* [@east2008monitoring] as example of using physical sensors at point locations
* [@yan2020tropical] and [@kinney2008autism] as examples of using storm tracks to assess physical exposure, and storm tracks are an example of a point location.
* [@lieberman2017self] as an example of using physical addresses for point locations.
* [@brunkard2008hurricane] as another example of using geocoded physical addresses as point locations.
* Point locations are high resolution spatial data, but there are reasons not to use them.
* There is not always enough lead time to set out temporary monitor networks as were used in [@east2008monitoring].
* Even if spatial data is available at the level of point location, it is sometimes best to aggregate this data for privacy concerns.
* The CDC suppresses identifiable data even at the county level when the number of cases in a year is below a certain number.

Polygons - Zip Code/County/Parish/State

* Polygons such as zip codes and counties are often used in tropical cyclones because a lot of administrative data from schools, hospitals, local government, etc. already have data at this level, plus it preserves privacy.
* [@huang2001long] used as example of using zip code level data to assess wind speed and damages after hurricanes.
* FEMA disaster declarations are an example of physical exposures being assessed at the county level. [@horney2021impact] is an example that uses counties experiencing one disaster declaration between 2003 and 2015 to look at the risk of disasters on suicide risk.
* [@lane2013health] used as example of using zip codes to analyze vulnerability of communities in New York City to the effects of hurricanes.
* [@kinney2008autism] used as example of the parish level being used to analyze effects of tropical cyclones exposure in utero and autism.
* [@parks2021tropical] used as example of county level data being used for analyzing hospitalization rates after exposure to a tropical cyclone.
* State spatial scale isn’t used much, but when it is it can be useful emergency preparedness and policy in different states/regions.
* [@willison2019quantifying] used as an example of using state level to compare response to hurricanes in Texas, Florida, and Puerto Rico.
* [@grech2015hurricane] used as an example of birth ratios and rainfall (in inches) taken at the state level from four Gulf States. Clarified in this paragraph that in utero exposure impacted whether male or female fetuses survived to be born months later after the storm (humans are not like frogs!).

*Rasterized Data*

* Rasterized data can show continuous data and variation across a geographic space.
* [@anderson2020assessing] was an example of using gridded precipitation data and using this to estimate county level precipitation data.

***Temporal Scales***

* There are large networks of sensors and monitoring stations that make it easy to have physical exposure data at a fine temporal scale, however most human impacts data is not at this scale, and is often not known until after a tropical cyclone.
* Time can be thought of as measurement snapshots (thermometer) or cumulative measurements (rain gauge).

*Snapshots of Time*

* Snapshots are values or measurements at a specific moment in time. Here temporal scales refer to frequency of snapshots.
* Real time (or at least very close to it) is available in minutes or hours because many sensors and monitors record storm tracks, flooding, precipitation, and storm surges at this level.
* [@east2008monitoring] used as example of sensors recording storm surge data at fine level (minute)
* [@grabich2016measuring] used as example of using storm tracks, where storm position is recorded every 6 hours.
* Measurements of human impacts are not usually at such short frequencies, but perhaps personal monitors and citizen science projects could change this.
* Measurements taken at daily or weekly frequencies are more common for human impacts, for example work absenteeism on a given day.
* Decade, yearly or semiannually measurements can measure things like population change, graduation rates, economic changes, etc.
* [[@jaycox2010children] used as example of taking measurements (in this case assessment of mental health outcomes) over longer periods, this study looked at 5 month intervals.

*Cumulative Measures of Time*

* Cumulative measurements measure everything up to a point which is why a rain gauge filling up with rain is a good measurement.
* On the smaller time scale is the amount of rain collected in an rain gauge over an hour, or storm cycle, or the total number of minutes that winds exceeded a certain speed. At longer time spans, days and weeks can record things like weekly hospitalization counts.
* [@parks2021tropical] used as example of hospitalization rates taken days after storm. The daily hospitalizations are the cumulative measurementCumulative measurements can sometimes be confused for snapshots, especially with things like daily counts, but important to remember that a count is accumulated over a specific period of time, not representative of a window in time.
* There are times where the delineation between snapshots and cumulative measurements up to a point are not as clear.

*Implications of not Improving this Integration*

* We will describe some of the ways to integrate datasets with different spatial and temporal scales: aggregation, matching, and interpolating.

Aggregating to Integrate Data at Different Scales

* Aggregating exposures is used by taking an average or a maximum of a physical exposure to assign it over a wider geographic or temporal scale.
* A threshold will often be chose to decide of a county or other area is exposed or unexposed.
* Saffir-Simpson scales is an example of using thresholds of maximum wind speed to classify severity of storms

Implications of Aggregating Data

* Aggregating data results in ecological bias, misclassification and measurement error, and

*Ecological Bias*

* Ecological bias occurs whenever the aggregate association between an exposure and an outcome does not properly reflect the association on the individual level

*Misclassification and Measurement Error in Aggregating Data*

* Misclassification error occurs when exposure and outcome variables are measured in categories and the wrong category is assigned to a particular case/observation - for example when a case that is exposed is incorrectly categorized as unexposed.
* Environmental epidemiology studies are often prone to misclassification error because the methods of assessing exposure are not always congruent with the way that researchers conduct human impact studies.

*When Data Have the Same Scale but are at Different Locations*

* Sometimes physical exposure data and human impacts data are both at the same scale (for example point location) but are not in the same place.
* One way to resolve this is to assign exposure to the individual based on exposure metrics taken from the nearest monitoring station.
* A disadvantage of matching exposure based on closest proximity to a monitoring station is that these can be sparse in certain areas (such as rural areas)
* Spatial interpolation is another method of resolving spatial misalignment by creating continuous surfaces that estimate exposure where there isn’t a monitor.
* Meteorological events can damage monitors.
* The more complicated a model becomes the harder it is to interpret. Some exposures like wind speed are relatively homogenous over large areas so different methods won’t yield wildly different results.

*Misclassification at Same Scale Different Locations*

* The more spatially heterogeneous an exposure is, the more room there is for exposure misclassification to occur.
* When assigning exposure to an individual point location based on the nearest monitoring site, the further this location is from the monitoring site, the more likely it is that the monitoring site won't reflect an exposure estimate accurately. Topography, complicated weather patterns, and other things could complicate this measurement.
* When interpolating, the environmental exposure of concern will partially determine the potential for misclassification. Using the examples from above of windspeed and rainfall, it is unlikely that much misclassification would occur over a spatial interface since they are homogenous over large areas.

**Conclusion/Discussion**

* This paper aims to show that in tropical cyclone studies physical exposure data and human impacts data come from different sources, have different scales, and that the integration of these datasets can lead to several sources of bias and error.
* Ecological bias and misclassification often arise from the integration methods we’ve mentioned.
* There is reason to be optimistic that we can harness the power of large datasets on physical, and increasing precision in model building, to understand and predict the effects of tropical cyclones on human health, economic viability, infrastructure, and social dynamics. This is of course contingent on our ability as researchers to make sure that the data on human impacts is of a similar quality and translatable resolution as the high quality data we already have on physical exposures.