**Challenges of Integrating Physical Exposure and Human Impacts Data in Tropical Cyclone Studies**

***Reverse Outline***

**Introduction**

* Tropical cyclones are very destructive and disrupt coastal communities, leading to health impacts, property damage, and costing billions.
* Tropical cyclones have a lot of impacts on physical health, mental health, and the economy, and here are some examples.
* Due to the destructive nature of tropical cyclones, it is important to study them to know how to build resilience against them.
* Tropical cyclone researchers come from various disciplines and so there is a challenge when integrating all of their datasets which are not often at the same spatial or temporal resolution.
* In this paper we will investigate why spatial and temporal misalignment exists, what spatial and temporal scales are used, and talk about some of the implications of integration.

**Physical Exposures from Tropical Cyclones**

* Let’s describe the meteorology of tropical cyclones and some of the measurements used to assess exposures to them.
* Define tropical cyclones.
* Atlantic Basin tropical cyclones originate in Africa and they develop across the Atlantic Ocean from summer to early fall.
* Wind is a main physical exposure of tropical cyclones that causes lots of damage and harm.
* Because tropical cyclones impact people through their physical properties, these have to be measured.

**Wind Speed and Direction**

* Wind speed is an important aspect of tropical cyclones, wind speeds above 74 mph is the requitement.
* Ground-based instruments used to measure wind include wind vanes, anenometers, and aerovanes.
* Above ground instruments used to measure wind include radar and doppler radar.

**Precipitation**

* Precipitation corresponds to magnitude and damage associated with a tropical cyclone.
* Ground-based instruments used to measure precipitation include rain gauges.
* Above ground, radar and doppler radar can be used to measure precipitation.

**Spatial and Temporal Misalignment: Origins of Integration Challenges**

* Tropical cyclone studies are multidisciplinary and this has led to spatial and temporal misalignment.
* Physical exposure data is comes from point locations in systems that cover vast geographic boundaries.
* Human impacts data are typically aggregated within geopolitical boundaries.
* Research question is part of what drives the spatial or temporal scale being used.
* Now that we have an idea of why spatial and temporal misalignment exists, it is important to know what scales are used when and where.

**Spatial Scales**

* The spatial scale is different for physical exposure data and human impacts data often.

**Point Location**

* Point locations are the smallest spatial scale and are often used in human impacts data to represent residential addresses, and here are some examples.
* There are certain advantages to point locations.
* Point locations of outcomes do not always line up with the exposure datapoints.

**Zip Code/County/Parish**

* Using zip codes/counties/parishes aggregates information at a wider spatial level.
* There are advantages to aggregating at the level of zip code/county/parish because it is a common method of assigning exposure.
* There are disadvantages to aggregating at this level because these areas are not always uniform in size, population distribution, etc. and can contribute to exposure misclassification.

**State/Metropolitan Region**

* Harville paper is example of looking at exposure in state of Louisiana at three different spatial levels.
* State level can show disparities in disaster preparedness, Willison paper looks at hurricane response in Texas, Florida, and Puerto Rico.

**Temporal Scales**

* Physical exposure data is often at very fine temporal scales (minutes, hours, etc.) and is often observed in real time, whereas human impacts data is not usually at this scale.

**Day**

* Physical exposure data is typically available at this scale, but human impacts data at this scale is mainly in the form of hospitalizations.
* Zahrah et al looked at casualty counts per day in United States.

**Week**

* Weeks are a common time unit when the human impact is related to birth outcomes or exposure in utero.

**Cumulative Measures of Time**

* Sometimes exposures are assessed over a broad period of time such as the duration of the storm, or a season, and in this case an aggregate measure of time is usually used.

**Implications of Not Improving this Integration**

* Several methods exist for integrating exposure and outcome data and this allows associations to be estimated between tropical cyclone exposures and human impacts
* When datasets are at different spatial and temporal scales, aggregating the data on the finer scale is a solution.
* Sometimes datapoints from physical exposures and human impacts are both at the same spatial level, usually a point location, but they aren’t in the same place.
* Integrating datasets introduces bias which will move the results towards or away from the null.
* In the last section of the paper we will explain the error and bias that can be introduced when we integrate the data through aggregation, interpolation, and matching.

**When Data Have Different Scales**

* When data are at different scales, you often aggregate.

**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.

**Categorizing Continuous Data**

* Researchers often aggregate a physical exposure measurement, often taking an average of maximum value.
* Thresholds of a physical exposure are chosen to decide of an area (such as a county or zip code) is exposed or unexposed.
* Saffir-Simpson scale is an example of categorizing continuous data (wind speed).
* Another scale used that categorizes wind speed is the Beaufort scale.
* Christopher2017effects looked at using outcomes as thresholds.
* Generally the drawback to categorizing or dichotomizing continuous data is that you lose information.
* Another problem with dichotomizing/categorizing is that cutoff points are arbitrary.
* Dichotomizing biases results (two measures that are close to each other but on the other side of a dividing line will be categorized as different).
* There are drawbacks to relying on a single exposure value, for example hurricanes may have different categories at different times and places during their trajectory.

**Misclassification and Measurement Error in Aggregating Data**

* Misclassification error results when individuals or aggregated groups are incorrectly categorized as exposed or unexposed.
* Misclassification error is common in environmental epidemiology, a good example is when storm tracks through a county make it exposed or not and population centers on the other side of the border but still close to the track are not classified correctly.
* Dasymmetric mapping is a potential remedy for misclassification.

**When Data Have the Same Scale, But are at Different Locations**

* Sometimes the point location for a human impact and a physical exposure related to the storm are not in the same place.
* Assigning exposure to a residential address based on the closest monitoring station/sensor is a typical way to resolve this, common to do in studies with wildfire plumes and smoke.
* Some problems that arise from the above method are that it excludes people who are too far away from a monitor to reasonably be included, and that in rural areas the in situ observations are more sparse.
* Weather itself can damage the sensors and monitors.
* Interpolation is another method of fixing the different locations problem.
* Kriging is one method of interpolation.
* Talk about modeling…[this is an idea that hasn’t been flushed out yet].

**Misclassification for Same Scale Different Locations**

* The more spatially heterogeneous that an environmental 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.
* 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.