
Multi-Source Data Fusion and Integration for Hazard Model Validation using Geospatial Information System

Edmund James Robbins, Florida Tech, Melbourne, FL; Florida Tech, Melbourne, FL; and E. Sziklay, J. P. Pinelli, and N. N. Kachouie

Abstract Text:

Multi-Source Data Fusion and Integration for Hazard Model Validation Using Geospatial Information Systems (GIS)

Due to the increasing effects of global climate change, the incidence and associated cost of disasters such as hurricanes are increasing. Hence, it is vital to develop integrated datasets and technologies to support and improve disaster response and predictions. The availability of large data sets from multiple organizations offers an opportunity to create standardized datasets or methodologies for the construction of such datasets that can serve as a basis for the creation, validation and comparison of a myriad of machine learning and physics-based models. The curse of the data driven approach is that many governmental, private sector, and academic modelers create and train models from unique datasets. The issue this causes is that models validated on one data set perform poorly when run on another compiled from different data. It then becomes difficult to make an honest evaluation of model performance.

To that end, the goal of this research is to develop and deploy an integrated database via the combination of reconnaissance, hazard, and exposure datasets. Through the application of geospatial data structures and techniques the various datasets will be combined into a single data structure in order to provide easy and intuitive access, appropriate data manipulation, computation, and interpretation. The implication of a united dataset is that modelers and machine learning professionals (Figure 1) will have the ability to access the entire spectrum of hazard related data in a single location in useable formats for machine learning or any other type of analysis. For example, the integrated database will enable the users to develop potential machine learning capabilities for damage estimation and insurance claim processing.

The current state of the database is best described as a case study focusing on the region of Florida affected. Hurricane Michael in 2018. More specifically Bay County, one of, if not the hardest hit locality in the state. Using Bay County as a jumping off point, an extensive data collection procedure was conducted to identify relevant and publicly available data from the local government. The data includes predominately information available from the local tax assessor and the GIS department and includes information on the building stock in the county. Building footprints derived from overhead imagery are included with this data, along with the delineation of the land parcels throughout the county. The collection effort was then expanded to hazard related information on the wind and storm surge levels. Using national level agencies like The National Oceanic and Atmospheric Administration (NOAA), The Federal Emergency Management Administration (FEMA), and the National Hurricane Center among several others, data located for the wind grid and the storm surge. The data was the best publicly available for the region for Hurricane Michael that did not represent a one-off analytical product.

Figure 1: Spatial Integration Process

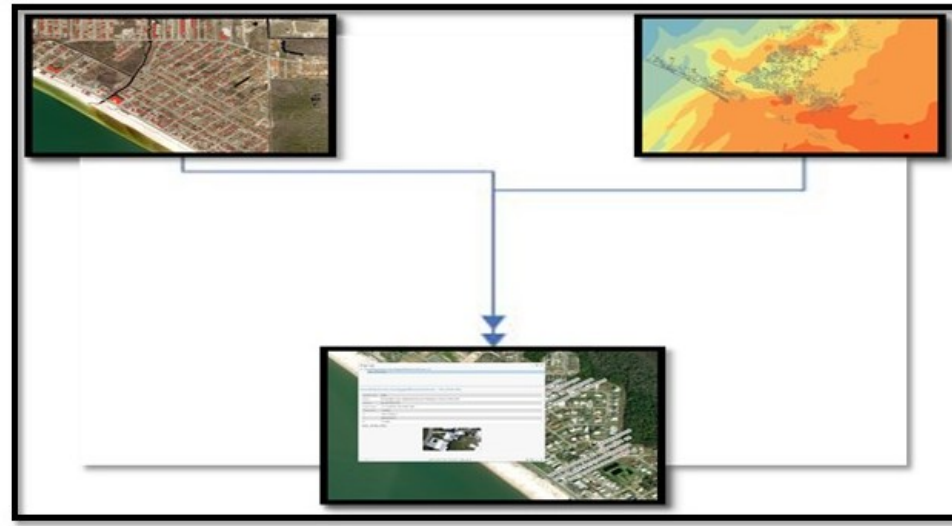


Figure 2: Spatial Database. Building shape file for Bay County, FL (Left); Building footprints for Bay County, FL (Right)



The final component to the data set was the reconnaissance data collected and stored on the DesignSafe platform. The data represented eld damage evaluations post Hurricane Michael collected by eld workers. This was either individual building evaluations or drone images which surveyed wide areas in the path of the storm and documented specific damage attributes.

Once the data was collected, GIS software was selected to be the primary data manipulation and analytical tool. Given that all of the data was collected in or could be joined with a spatial data structure and the large support infrastructure available for scaling it was the most appropriate choice. Using the GIS software, the data was imported, combined, and stored in a spatial lite database (i.e., geodatabase). The output of the integration and fusion is a shape le based on the satellite derived building footprints (Figure 2) which contains all of the information previously found spread across multiple locations and multiple agencies. This dataset could in the future be used for model training or validation and has the ability to continue to be expanded should additional information or products be created by another organization.