Implementing Extreme Value Analysis in a Geospatial Workflow for Assessing Storm Surge Hazards

Jason Catelli

The Pennsylvania State University

Master of Geographic Information Systems

Capstone Project

Advisor: Dr. Patrick Kennelly

May 6, 2015

Abstract

This project addresses the need to utilize extreme value theory in a geospatial environment to analyze coincident cells across multiple synthetic events. This work details a geospatial approach to move data from raster grids to SciPy's NumPy Array structure using the Python programming language. The data are then connected through a Python library to an outside statistical package to fit cell values to extreme value theory distributions and return values for specified recurrence intervals.

While this is not a new process, the value behind this work keeping the process in a single mainstream geospatial environment and being able to easily replicate this process for other natural hazard applications and extreme event modeling.

Introduction

Return periods or recurrence intervals can be defined as the percent chance of exceeding a certain threshold over 1 year. An example of this is a 100-year flood event, or event consisting of a 1 in 100 chance or 1% chance of occurring over 1 year. Insurance companies and public entities use return periods to understand the risk and prepare for catastrophic extremes. Using catastrophe models to simulate events, modelers can produce return periods based on historical and simulated results to understand these outlying events.

Gridded data of 100-year (1%) and 500-year (0.2%) storm surge flood elevations for the East Coast of the United States and Gulf of Mexico are critical to understanding this natural hazard. In this project, storm surge heights were calculated across the study area utilizing NOAA's National Weather Service SLOSH (Sea, Lake, and Overland Surges from Hurricanes) model data for thousands of synthetic hurricanes making landfall in the US. Based on the results derived from SLOSH, a series of interpolated surfaces were produced using spatial analysis in a geographic information system (GIS) at both the SLOSH basin and the synthetic event levels. The result was a single grid of maximum flood elevations for each synthetic event (Nong et al., 2010).

The former process relied on separate code to read the results and calculate the statistics before these results could be used in the GIS for visualization. This project fills this gap so that all work is contained in a single standard geospatial environment.

Method

The process begins with point data from the SLOSH model for one synthetic storm event.

The data points for each basin within the synthetic event are interpolated and a grid at approximately 2 km. is created (Figure 1). The grid is coincident to a larger grid at a similar resolution and is used to produce results for the entire geographic area (Figure 2). An additional series of interpolations based on Empirical Bayesian Kriging, a kriging method that automates a number of kriging parameters (O'Sullivan and Unwin, 2010 and ESRI, 2012), are then computed across the synthetic storm event to produce a grid at the same resolution for the storm event (Figure 3). This raster result contains storm surge heights across the spatial domain of the synthetic event. This process is then repeated through Python code for the remaining synthetic events.

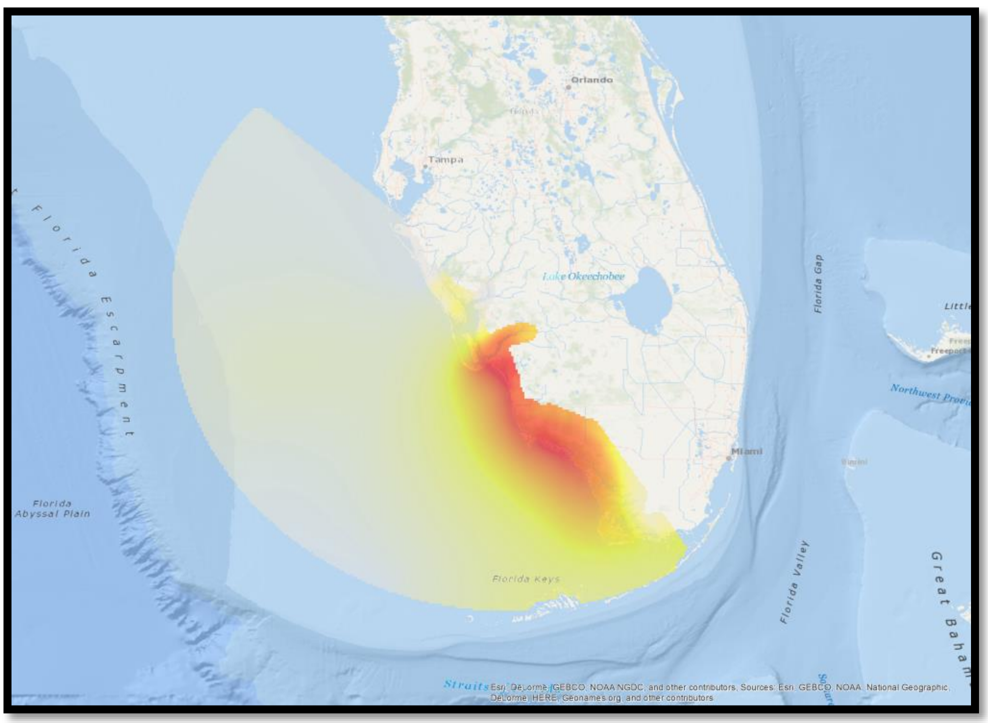

Figure 1. *Interpolation result across a single basin for one synthetic event.*

Figure 2. *Overall project domain*

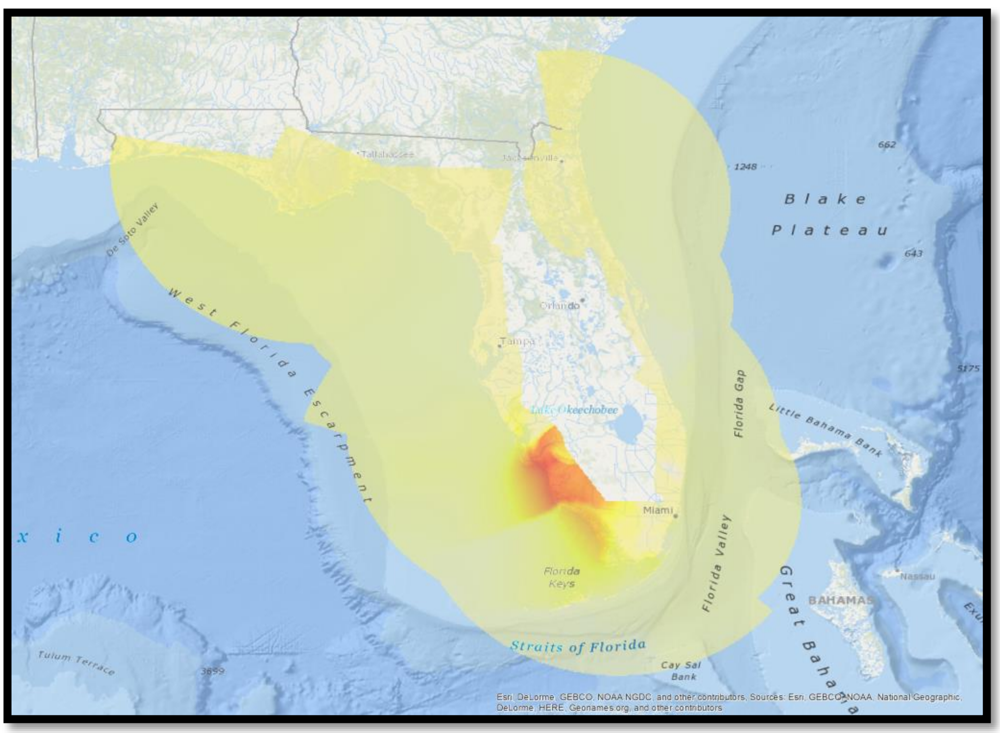

Figure 3. *Interpolation result across multiple basins for one event.*

Utilizing NumPy, the fundamental package for scientific computing with Python, each resulting grid showing storm surge for a synthetic event is converted to a NumPy array. A three-dimensional NumPy array is then created by stacking each two-dimensional event array vertically (Figure 4).

Through the use of Python indexing, the storm surge height values for each event at each grid cell can easily be retrieved. These values are then fit to an extreme value curve using the *genpareto* library within the SciPy statistical package (Coles, 2001). 100-year and 500-year values for each cell are returned based on the sample fitting as shown in Figure 5. Finally, the resulting 100-year and 500-year NumPy arrays are converted to raster grids.

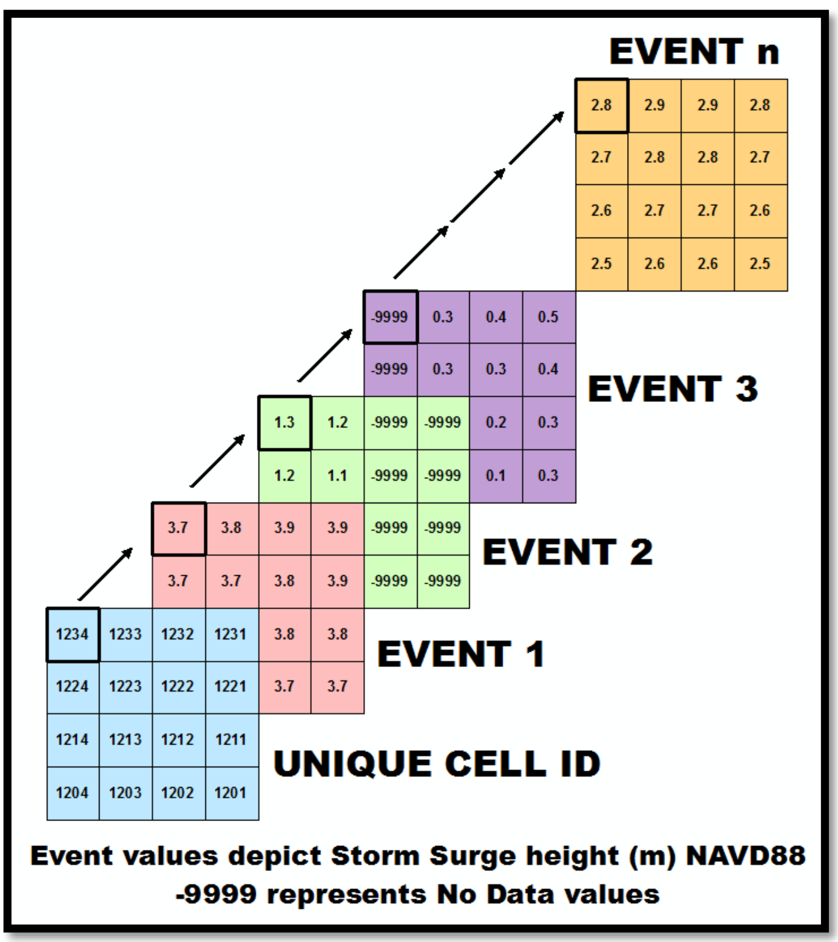

Figure 4. Example of three-dimensional NumPy array created by stacking each event array.

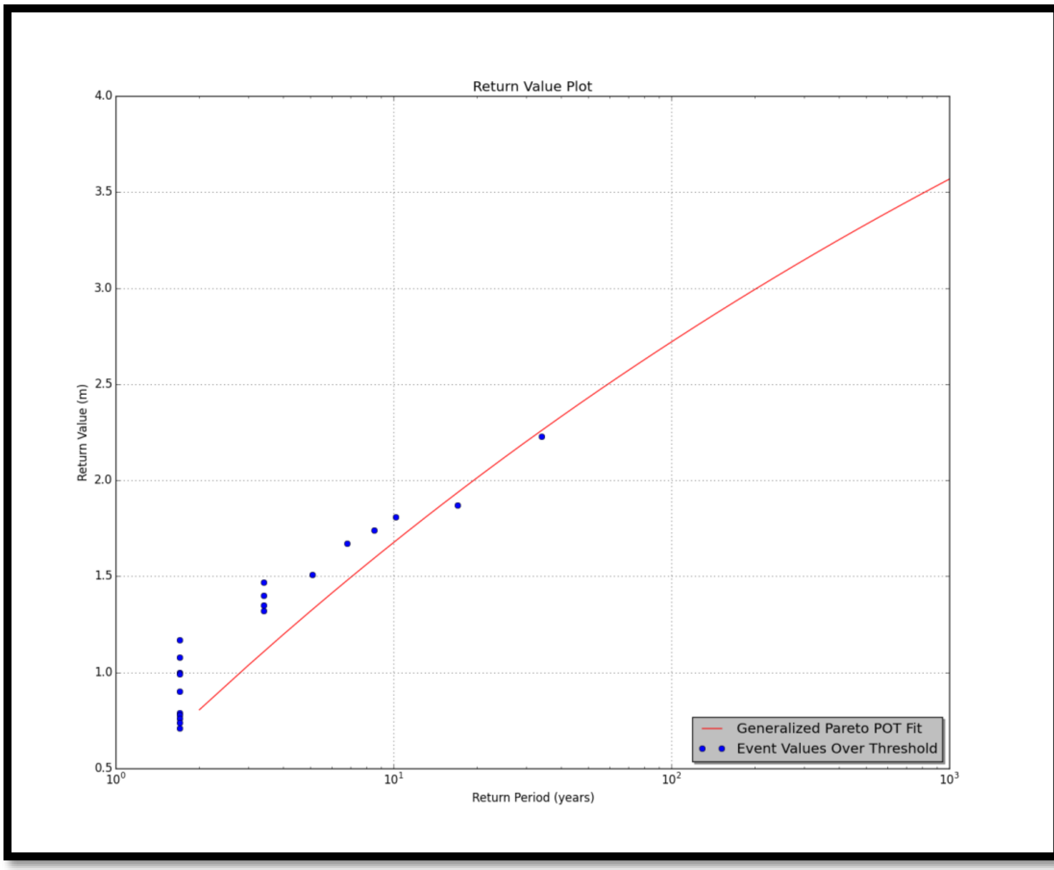

Figure 5. Return value plot for an individual cell using generalized Pareto distribution and peak-over-threshold approach.

Results

The process highlighted in Figure 6 details the workflow produced in this project. Raster data was successfully converted to NumPy array structures. Through Python code, storm surge height values for each synthetic event were returned at each individual cell. After importing the SciPy statistical package, the values collected at each cell were fit to a general extreme value distribution. 100-year and 500-year values at each grid cell were returned and converted back to raster grids for further geospatial analysis and visualization.

It is important to note that while the process was successful, memory errors surfaced due to the vast size of the datasets produced.

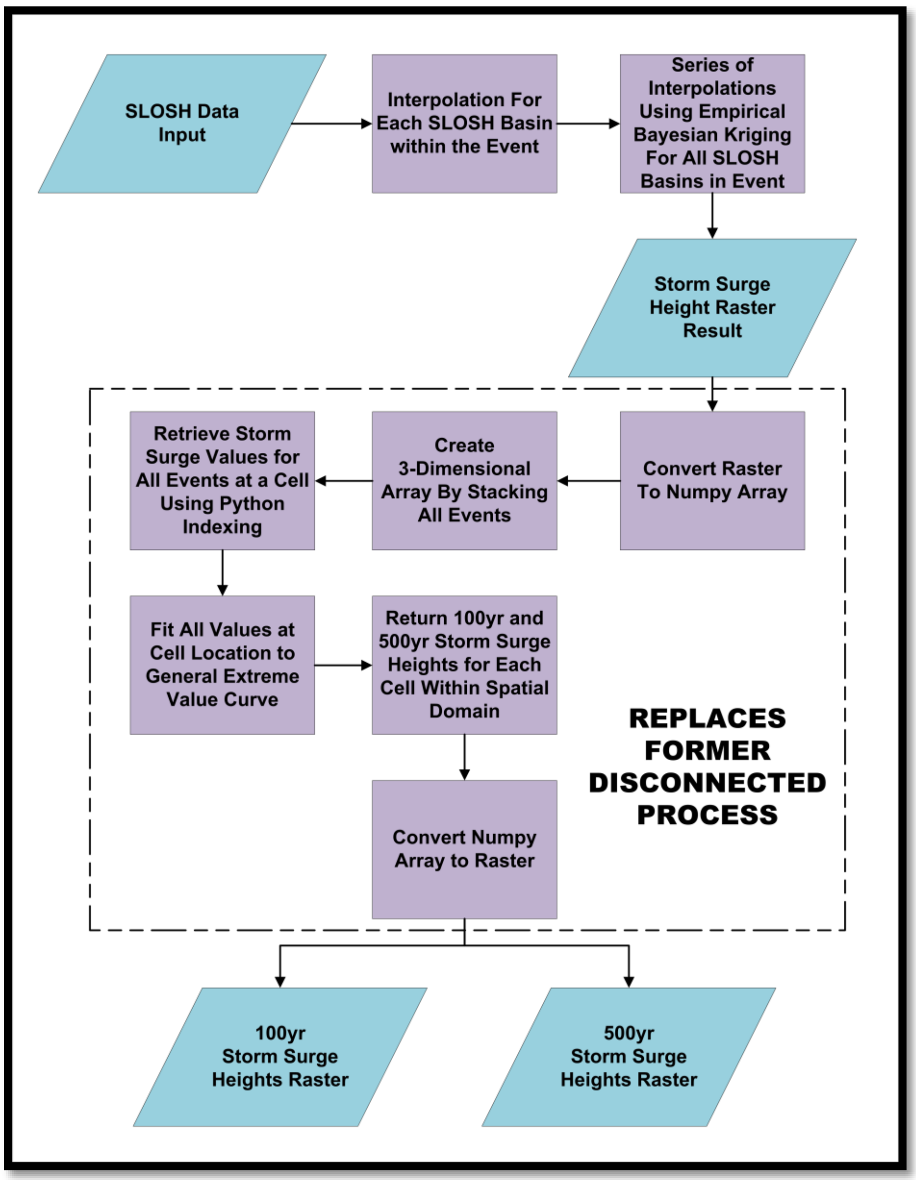

Figure 6. *Overall process.*

Summary

Utilizing the Python programming language along with NumPy and a statistics package like SciPy.stats, extreme value analysis can be performed in a single mainstream geospatial environment.

The process involves interpolating point data to produce coincident raster datasets representing multiple synthetic storm surge events. Converting the raster datasets to a three-dimensional NumPy array enables access to all event values at a single grid cell through Python indexing.

The event values at a single cell are fit to an extreme value curve using the SciPy.stats library within Python. Based on the fitting, 100-year and 500-year values are determined for the cell. The process is repeated for the remaining grid cells and the arrays are converted back to raster datasets representing 100-year and 500-year storm surge heights.

Acknowledgements

The author would like to thank his graduate advisor Dr. Patrick Kennelly for his continuous advice and guidance throughout this project. He would also like to thank his employer FM Global, his colleagues Dr. Shangyao Nong for his guidance throughout this project, and Dr. Sait Ahmet Binselam and Dr. Yasir Kaheil for their help with questions related to Python and statistics.

References

Albrecht, Jochen. *Key Concepts and Techniques in GIS.* London: SAGE Publications Ltd., 2007.

Chang, Kang-Tsung. *Introduction to Geographic Information Systems.* Texas: McGraw-Hill, 2003.

Coles, Stuart. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag, 2001.

Esri. 2012. "What is Empirical Bayesian Kriging?" Last modified November 2. http://resources.arcgis.com/en/help/main/10.1/index.html# //0031000000q9000000

Lin, N., K. A. Emauel, J. A. Smith, and E. Vanmarcke. "Risk Assessement of Hurricane Storm Surge for New York City." *Journal of Geophysical Research*. September 25, 2010.

Nong, Shangyao, Jeffrey McCollum, Liming Xu, Michael Scheffler, and Hosam Ali. "Probabilistic Storm Surge Heights for the US Using Full Stochastic Events." Paper presented at the AMS 29th Conference on Hurricanes and Tropical Meteorology, Tucson, Arizona, May 10-14, 2010.

O'Sullivan, David, and David J. Unwin. *Geographic Information Analysis.* New York:

Wiley, 2010.

Tonini, Francesco. “Drought Risk Assessment – A Customized Toolbox.” Presentation at the ESRI International User Conference, San Diego, California, July 8-12, 2013.

Walton, Todd L. “Distributions for Storm Surge Extremes.” *Ocean Engineering.* December 2000: 1279-1293.