**Question**

**What is the probability of having between x and y people affected?**

**Proposed Solution**

**Solution A**

* 1. Solution 1 (regressions)
     1. Plot different types of regressions of the impact values based on the dataset
     2. Look for high R^2 values as this is an indicator for how good a particular regression fits to the dataset
        + Ideally, identify one regression which provides best results for all datasets; this is an area for model development
     3. Solve the equation of the exponential regression for x (= population affected) with a given class limit y (e.g. 5, 15, 50,100).
     4. Calculate the class probabilities

**Attempt**

The problem is that the impact values (e.g. population at risk) differ in number (e.g. in some area we only have one value whilst in other there are up to six values). In those fortunate situations where we have six values we can certainly evaluate which regression fits better the data using R2.

What is unclear is what do we mean with class limits y and therefore how to calculate the class probabilities (points iii and iv).

|  |
| --- |
|  |
|  |

**Solution B**

* 1. Solution 2 (PDF)



1. Probability density functions can be used to determine the probability that a continuous random variable lies between two values, say *a* and *b.*
2. In this case *a* and *b* represent a population between *a* and *b* as the number of people comprehended in this hypothetical population and the formula evaluates what is the probability of *x* between those parameters.
3. In graphical terms the probability density function is the area under the graph of the density curve where the following conditions is true: 
4. Once the density function or  is known, the next step is to solve the Lebesgue integral and evaluate with the established *a* and *b* values and the probability of *x* of being between those values will be the result.

**Attempt**

Again the problem is that the impact values (e.g. population at risk) differ in number (e.g. in some area we only have one value whilst in other there are up to six values).

|  |  |
| --- | --- |
|  |  |

Calculating a Probability Density Function is doable but I wonder whether these makes any sense in our case where we have only a sample of six values over 975 years.

In other words, we have the values for six years (return periods or probabilities) namely 25, 50, 100, 200, 500 and 1000 but we ignore what is the number of people affected for all the other 975 years. Can we still assume that a six values set of discrete random variables (PMF) is representative of the whole time interval (continuous variables PDF)?

Assuming that the above problem is negligible, according to the rationale of solution b we could say that *x* represent a reliable probability of affected population between that *a* and *b* (see next figure).

|  |
| --- |
| x  b  a |
|  |

Why is the Lebesgue integral explicitly mentioned for evaluating the probability x for the *a* and *b* values. Does it add anything to the normal integration of the area under the curve (pdf)?

**Case Study**

(Best-case scenario six values of population affected for six annual probabilities)

|  |  |  |  |
| --- | --- | --- | --- |
| Return Period | Annual Probability | People Affected | Affected People  Cumulated |
| 25 | 4% | 1964 | 1964 |
| 50 | 2% | 1532 | 3496 |
| 100 | 1% | 9054 | 12549 |
| 200 | 0.5% | 16710 | 29260 |
| 500 | 0.2% | 1820 | 31080 |
| 1000 | 0.1% | 153 | 31233 |

Figure - GIS Calculated Values

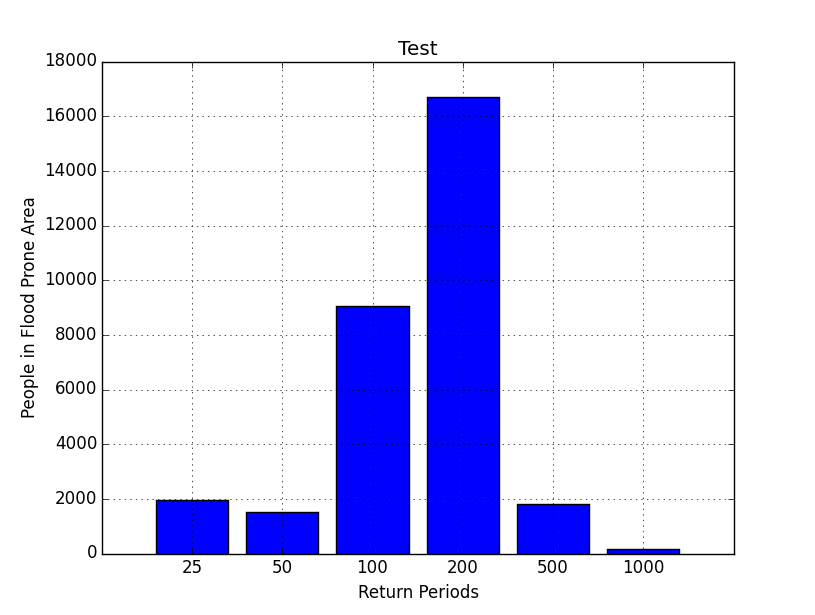
****

Figure - People Affected by Flood per Return Period

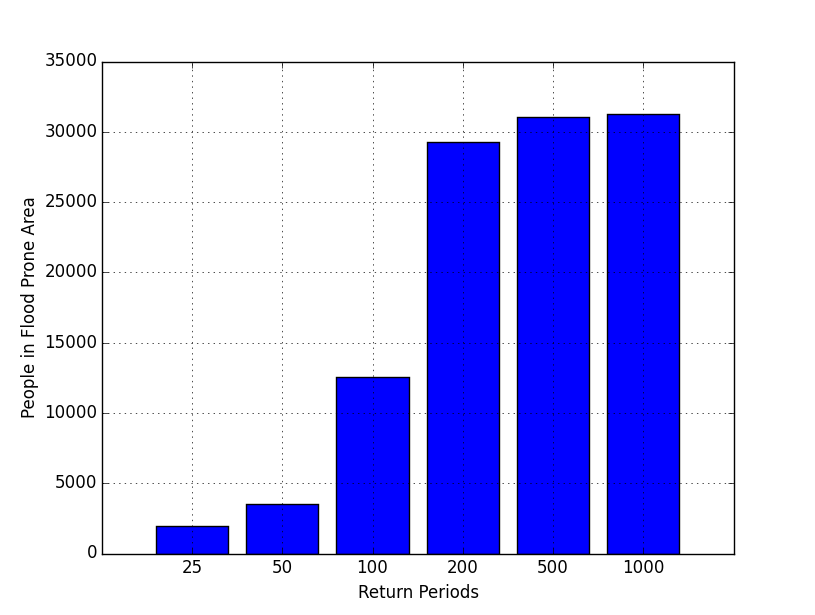


Figure - People Affected by Flood per Return Period (Cumulated Number of People)

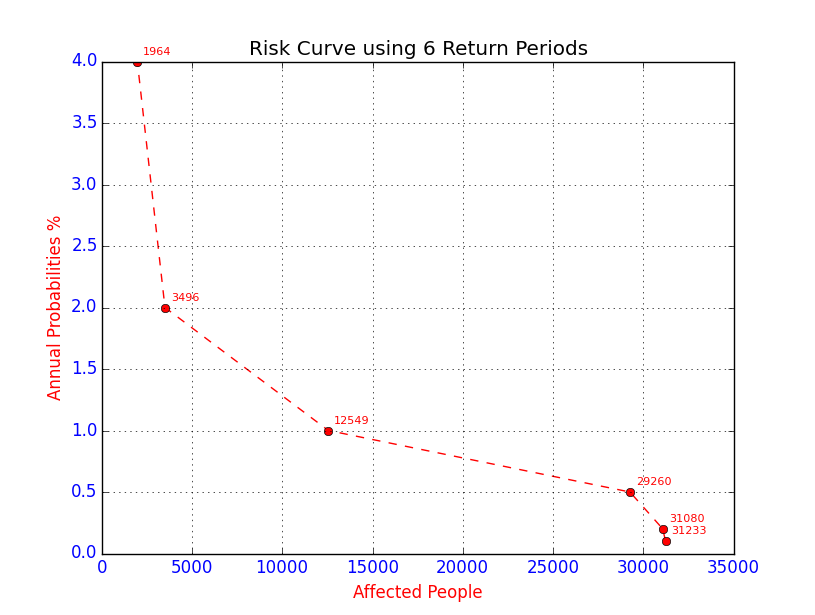
****

Figure - Risk Curve (annual probability of occurrence versus number of people affected)

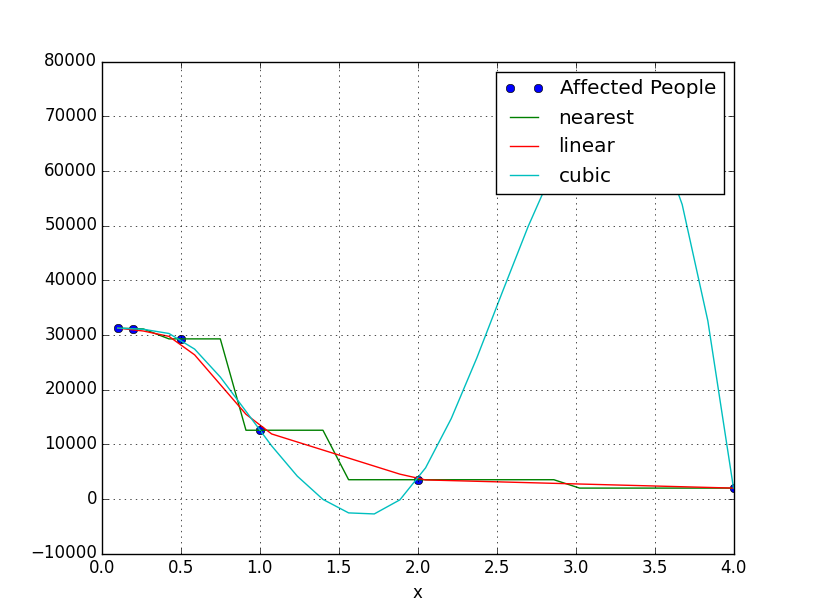
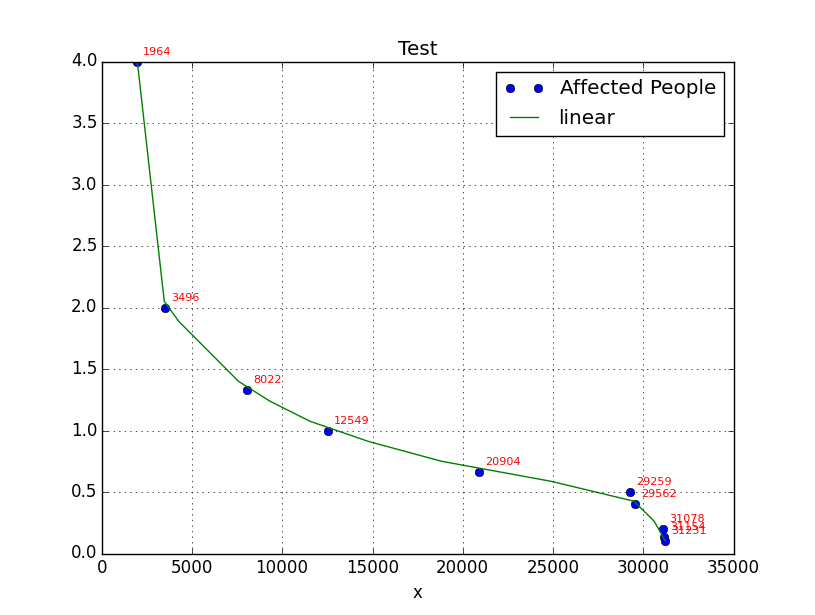
****

Figure - Various Interpolations evaluated using GIS Calculated Values

Figure - Interpolation chosen for Calculating Interim Values of Population (e.g. between x and y)