1 Confidence Factor

In this section, we describe our approach to assigning a confidence value to all the predictions made during the simulation. This confidence value of a prediction made for event e is based on two things: (1) the density of neighborhood around event e, and (2) the variance in the intensities of neighbors as well as their distances from event e. The confidence value of each prediction is inversely related to the variation in distances and intensities, and the confidence value is directly related to the neighborhood density.

There are different ways to calculate the neighborhood density. For our model, the neighborhood density of an event e, denoted by ρ_e, is the number of its neighbors divided by the maximal number of neighbors in the domain R. If the number of neighbors is greater than the threshold value, we set the neighborhood density to 1.0. For any event e, we then calculate a weighted intensity value that is a combined representation of each neighbor's intensity and its distance from the main event e. Suppose $[e'_1, ..., e'_n]$ is the list of neighbors of an event e, and that $[I'_1, ..., I'_n]$ is the list of intensities of the neighbors, On any day t, we assign weights to each neighbor denoted as $[w'_1, ..., w'_n]$. We then calculate the variance in the list of the weighted intensities $[w'_1.I'_1, ..., w'_n.I'_n]$ for each event. The weight is calculated based on each neighbor's distance to e, conforming to the Distance Principle of our model. The weight of the i^{th} neighbor is calculated as:

$$w'_i = \frac{R - d(e, e'_i)}{R} \tag{1}$$

To describe the variation in the weighted intensities, we use the coefficient of variation (C_V) [**?**], because we are interested in the relative spread of the intensity and distance of neighbors compared to each event rather than the numeric values of the variance. If, for a list of neighbors, the standard deviation and mean of their weighted intensities is denoted by S and M respectively, the coefficient of variation is computed as:

$$C_V = \frac{S}{M} \tag{2}$$

Note that C_V can take a value larger than unity, we therefore cap the value of C_V to 1.0 as it denotes maximum variability, this modified coefficient of variation is referred as C'_v and is calculated as:

$$C'_V = min(\frac{S}{M}, 1) \tag{3}$$

Since both of our variables, neighborhood density and variance are in the range [0,1], we calculate the confidence in a prediction made for event e is computed as:

$$confidence_{e,t} = wt_{var}.(1 - C'_V) + wt_\rho.\rho_e \tag{4}$$

In our validation methodologies, we examine the predictions made for an area rather than a single location. Therefore, we calculate the average confidence

($confidence_{A,t}$) in the prediction made for area A and day t as the average of individual confidence values calculated for each event found in that area A for all N number of days prior to day t in the simulation.