*Household Dislocation Algorithm 3: A Logistic Regression Approach**

Yi-Sz Lin
Walter Gillis Peacock
Jing-Chen Lu
Yang Zhang

January 2008

Hazard Reduction and Recovery Center
Texas A&M University

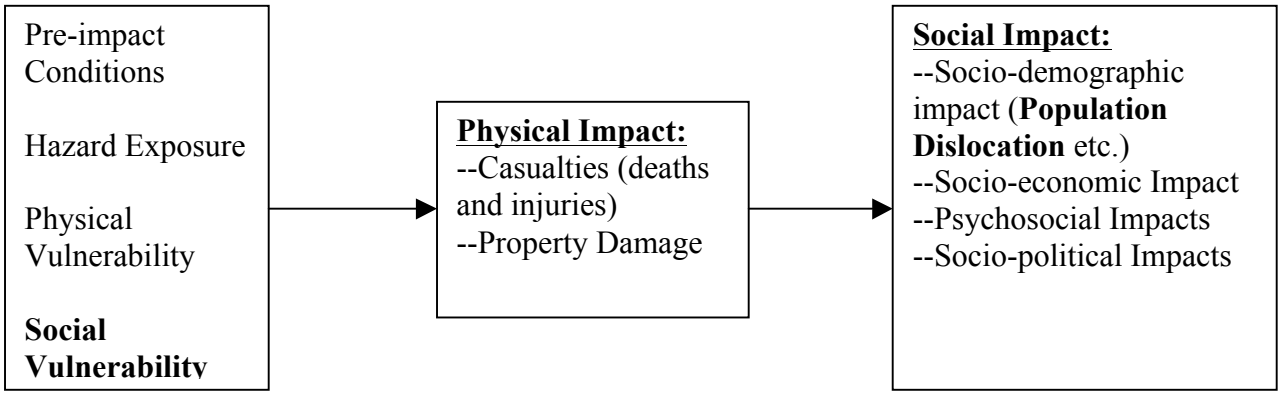
* This document discusses and provides detail instructions for the creation of the third household dislocation algorithm developed for and implemented into the Mid-American Earthquake Center's MAEViz program focusing on Shelby County, Tennessee. This work was supported by the Mid-American Earthquake Center with funding from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Mid-American Earthquake Center.

Suggested Reference: Lin, Y.S., W.G. Peacock, J.C. Lu, and Y. Zhang and (2008) *Household Dislocation Algorithm 3: A Logistic Regression Approach*. Hazard Reduction and Recovery Center, Texas A&M University. HRRC Reports: 08-05R. (hrrc.arch.tamu.edu/publications/research reports/08-05R Dislocation Algorithm 3.pdf)

Household Dislocation Algorithm 3: A Logistic Regression Approach

The disaster impact can be viewed as the result of interactions among three systems—the earth's physical system, the human social system, and the constructed system (Mileti, 1999). To understand the full nature of disaster impacts, we need to also understand better the nature of the social systems that are utilizing the built environment (buildings, transportation networks, infrastructure systems). Like the built environment, the social systems that produce and utilize that environment are far from homogeneous. They differ in their size, form and structure, in their access to scarce resources, such as wealth, power, social status and information, and in many other ways. These differences can shape and influence the nature of disaster impact—or again, the broader consequences of disasters. Estimates of population dislocation—one of the social impacts—will be inadequate if they are merely dependent on physical damages as demonstrated in the HAZUS model. Thus, this population dislocation algorithm is developed based upon not only the fragility of physical environment, but also the social factors and processes that generate vulnerability in terms of a person's or group's capacity to anticipate, cope with, resist and recover from the impact of a natural hazard (Blakie et al., 1994). These capacities are contingent upon a host of factors such as economic status and ethnicity that determine access to the scarce resources in the society.

Figure 1. The disaster impact model (Lindell & Prater, 2003)



Basic logic behind this approach:

The model is developed by employing the logistic regression method in which the natural logarithm of odds of event 1 is a linear function of a set of independent variables X_1, X_2, X_3 and X_4. This model is specified as the following equation.

$$\ln\left(\frac{\Pr(1)}{1-\Pr(1)}\right) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

In logistic regression the coefficients are estimated by the maximum-likelihood approach. It is a non-linear model that uses an iterative procedure to choose the estimates. The algorithm for predicting the probability that households living in a residential structure be dislocated is then derived from the logistic model, as shown in the following equation.

$$\Pr(1) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4)}}$$

In the equation the dependent variable $\Pr(1)$ is the probability that that households living in a certain residential structure be dislocated. The independent variables X_1 and X_2 represent the damage level and the housing type respectively for a certain residential structure while the other independent variables X_3 and X_4 represent the social characteristics for the block group in which the structure is located.

The Logistic Regression Dislocation Algorithm

I. Base data requirements.

1. **Census data, at block group level:** In the HAZUS package, the data at Census tract are used to estimate possible dislocation household. Block group data are used here to estimate more detailed information thereby facilitating planning within local communities and counties. The following are the data needed for the dislocation algorithm, and are similar to those used in the social vulnerability algorithms. They are provided in the excel spreadsheet for Shelby County (they originally were generated by Steve French's group).

   Variable name Variable definition
   - TOTPOP → Total Population
   - P_BLACK → Total Population -- Blacks Alone
   - P_HISP → Total Population -- Hispanics
   - TOT_HH → Total No. of Households
   - TOT_HU → Total Housing Units

   The above data are employed to calculate the following socioeconomic variables for a certain block group k in the impacted area.

   a. Percent of Black population in block group k:
   - $\%BLACK_{bg_k} = \frac{P\_BLACK}{TOTPOP} \times 100$

   b. Percent of Hispanics population in block group k:
   - $\%HISP_{bg_k} = \frac{P\_HISP}{TOTPOP} \times 100$

   c. Calculate average number of households per dwelling unit in block group k: By calculating the average number of households per dwelling units we get some notion of the number of households adjusting for occupancy rates. This adjusted mean will be used to estimate the number of dislocated households.
   - $AveHhDU_{bg_k} = \frac{TOT\_HH}{TOT\_HU}$

2. **From the inventory data:** The algorithm is based on damage to residential structures only. It is therefore critical that MAEViz be able to clearly identify residential structures and that residential structures be clustered into their respective census block-group areas. Two variables are calculated from the inventory data. The first one is a dummy variable

D_SF_{jk} for structure j in block group k recoded from the variable OCC_TYPE in the building inventory. In D_SF_{jk} the Single-family Residential (RES1 in OCC_TYPE) is coded as 1 while other types of residential structure are coded as 0. The second variable $\%VLOSS_{jk}$ is the percent value loss of residential structure j in block group k. It is estimated by using the SE-2 Fiscal Impact Analysis Algorithms from Steve French. The required inventory variables and calculations are shown as the following.

Variable name Variable definition

- OCC_TYPE → Structure occupancy type. The algorithm needs only single family structures (RES1) and multi-family structures (RES3)
- NO_DU → No. of dwelling units in the structure.
- BLDG_VAL → Total Value of the Building.

a. Create the dummy variable for residential structure j in block group k, D_SF_{ik}: The D_SF_{jk} is calculated with the following conditional operations.

- D_SF_{jk} = 1 if OCC_TYPE = RES1
- D_SF_{jk} = 0 if OCC_TYPE = RES3

b. Calculating the percent value loss of residential structure j in block group k, $\%VLOSS_{jk}$:

- $$\%VLOSS_{jk} = \frac{DED_{jk}}{Bldg\_Val_{jk}} \times 100 = \frac{\sum_{i=1}^{n} p(DS_i) \times DF_i \times Bldg\_Val_{jk}}{Bldg\_Val_{jk}} \times 100 = \sum_{i=1}^{n} p(DS_i) \times DF_i \times 100$$

Where,
DED_j = Direct Economic Damage to building j
$p(DS_i)$ = probability of the building being in Damage State i
DF_i = Damage Factor i from Hueste et al.
$Bldg\_Val_{jk}$ = Building Value for Building j in block group k from Inventory Database

II. Process for estimating dislocated households for block groups and the whole impacted area:

1. Calculate the dislocation probability for residential structure j in block group k be dislocated, $PrDis_{jk}$:

- $$\Pr Dis_{jk} = 1 \Big/ \left\{1 + e^{-\left[b_0 + b_1 \times (\%VLOSS_{jk}) + b_2 \times (D\_SF_{jk}) + b_3 \times (\%BLACK_{bgk}) + b_4 \times (\%HISP_{bgk})\right]}\right\}$$

The required coefficients are listed in the following **Table 1.**

Table 1. Default values of coefficients.

Coefficients	Default Values
b_0	-0.42523
b_1	0.02480
b_2	-0.50166
b_3	-0.01826
b_4	-0.01198

2. **Calculating dislocation factor for structure j in block group k using the dislocation probability ($PrDis_{jk}$) and probability cutoff value:** The dislocation factor ($DisF_{jk}$) is calculated by the following conditional operations. The default value for probability cutoff is 0.5. ***NOTE: We would like the ability to adjust this value as part of the algorithm.***

   - $DisF_{jk}$ = 1 if $\Pr Dis_{jk} >= 0.5$
   - $DisF_{jk}$ = 0 if $\Pr Dis_{jk} < 0.5$

   Here the probability cutoff value could be adjusted based on different circumstances.
   a. Hazard characteristics (e.g., lower cutoff value and thus higher dislocation estimates for earthquakes because of the aftershocks).
   b. Weather conditions (e.g., lower cutoff value and thus higher dislocation estimates for harsh weather conditions).
   c. Intensity of the event.
   d. Degree of damage to utilities.

3. **Calculating number of dislocated households for block group k:** After the dislocation factor for the residential structure ($DisF_{jk}$) is calculated, the number of dislocated households in block group k is calculated as the following.

   - $DisHh_{bg_k} = \sum_{j=1}^{m} \left(DisF_{jk}\right) \times \left(NO\_DU_{jk}\right) \times \left(AveHhDU_{bg_k}\right)$

   Where m= total number of residential structures in block group k.

4. **Calculating percent of dislocated households for block group k:** The Percent of block group households dislocated ($PDisHh_{bg_k}$) is then calculated as the following.

   - $PDisHh_{bg_k} = \frac{DisHh_{bg_k}}{TOT\_HH_{bg_k}} \times 100$

5. **Calculating total number of dislocated households for a jurisdiction covering p block groups:** The total number of dislocated households ($TotDh$) in a jurisdiction covering **p** block groups is calculated as the following.

- $TotDh = \sum_{k=1}^{p} DisHh_{bg_k}$

III. Expected output:

1. First there should be a report of dislocated household by block group and the total number of dislocated household at county level. See Appendix 3.
2. Second, there should also be a map of number of displaced household by block group (using $DisHh_{bg_k}$). See Appendix 4.
3. Third a map of percent of displaced household within the block group (using $PDisHh_{bg_k}$). See Appendix 5.

Appendix 1. Variable List

Variable Name	Description	Note
TOTPOP	Total Population	2000 Census (from Dr. French)
TOT_HH	Total No. of Households	2000 Census (from Dr. French)
TOT_HU	Total Housing Units	2000 Census (from Dr. French)
P_BLACK	Total Population -- Blacks Alone	2000 Census (from Dr. French)
P_HISP	Total Population -- Hispanics Alone	2000 Census (from Dr. French)
OCC_TYPE	Structure occupancy type	Building Inventory Data (from Dr. French)
NO_DU	Number of Dwelling Units in the Building	Building Inventory Data (from Dr. French)
BLDG_VAL	Total Value of the Building	Building Inventory Data (from Dr. French)
$AveHhDU_{bg_k}$	Average number of households per dwelling unit in block group k	$AveHhDU_{bg_k} = \frac{TOT\_HH}{TOT\_HU}$
$\Pr Dis_{jk}$	Dislocation probability for residential structure j in block group k be dislocated	$\Pr Dis_{jk} = 1 \Big/ \left\{1 + e^{-\left[b_0 + b_1 \times (\%VLOSS_{jk}) + b_2 \times (D\_SF_{jk}) + b_3 \times (\%BLACK_{bg_k}) + b_4 \times (\%HISP_{bg_k})\right]}\right\}$
$DisF_{jk}$	Dislocation factor for structure j in block group k	$DisF_{jk} = 1$ if $\Pr Dis_{jk} >= 0.5$ $DisF_{jk} = 0$ if $\Pr Dis_{jk} < 0.5$
$DisHh_{bg_k}$	Number of dislocated households for block group k	$DisHh_{bg_k} = \sum_{j=1}^{m} (DisF_{jk}) \times (NO\_DU_{jk}) \times (AveHhDU_{bg_k})$
$PDisHh_{bg_k}$	Percent of dislocated households for block group k	$(DisHh_{bg_k} / TOT\_HH_{bg_k}) \times 100$
$TotDh$	Number of dislocated households for an jurisdiction covering **p** block groups	$TotDh = \sum_{k=1}^{p} DisHh_{bg_k}$
The following 5 are defined for Estimation Procedure Section II. 1. only.		
$\%VLOSS_{jk}$	Percent of building value loss for structure j in block group k	See Section I. 2. b.
D_SF_{jk}	Dummy variable representing single family structure	See Section I. 2. a.
$\%BLACK_{bg_k}$	Percent of Black alone population in block group k	[(P_BLACK)/(TOTPOP)]*100
$\%HISP_{bg_k}$	Percent of Hispanics population in block group k	[(P_HISP)/(TOTPOP)]*100

Appendix 2. Example Calculations for population dislocation:

The following is the hypothetical damage and social characteristic for a certain block group:

Block Group	Structure	Direct Economic Damage to the Building	Pre-impact Building Value	$\%VLOSS_{jk}$	D_SF_{jk}	NO_DU	TOT_POP	TOT_HU	TOT_HH	$AveHhDU_{bg_k}$
47157 00130 xx	1	30,000	50,000	60	1	1	54	24	18	0.75
	2	350,000	500,000	70	0	20				
	3	30,000	120,000	33.3	1	1				
	4	30,000	50,000	60	1	1				
	5	12,000	40,000	30	1	1				

1. Calculate number of dislocated households for block group k:

$$\Pr Dis_{jk} = 1 \Big/ \left\{1 + e^{-\left[b_0 + b_1 \times (\%VLOSS_{jk}) + b_2 \times (D\_SF_{jk}) + b_3 \times (\%BLACK_{bg_k}) + b_4 \times (\%HISP_{bg_k})\right]}\right\}$$

$$DisHh_{bg_k} = \sum_{j=1}^{m} (DisF_{jk}) \times (NO\_DU_{jk}) \times (AveHhDU_{bg_k})$$

Block Group	Structure	$\%VLOSS_{jk}$	D_SF_{jk}	$\%BLACK_{bg_k}$	$\%HISP_{bg_k}$	$\Pr Dis_{jk}$	$DisF_{jk}$	$AveHhDU_{bg_k}$	$(DisF_{jk}) \times (NO\_DU_{jk}) \times (AveHhDU_{bg_k})$	$DisHh_{bg}$
47157001 30xx	1	60	1	54.7743	25.911	0.320934	0	0.75	0 * 1 * 0.75 = 0	15
	2	70	0			0.500044	1		**1 * 20 * 0.75 = 15**	
	3	33.3	1			0.195975	0		0 * 1 * 0.75 = 0	
	4	60	1			0.320934	0		0 * 1 * 0.75 = 0	
	5	30	1			0.1834	0		0 * 1 * 0.75 = 0	

2. Calculating percent of dislocated households for block group k

$$PDisHh_{bg_k} = \frac{DisHh_{bg_k}}{TOT\_HH_{bg_k}} \times 100 = 15 / 24 * 100 = 62.5\%$$

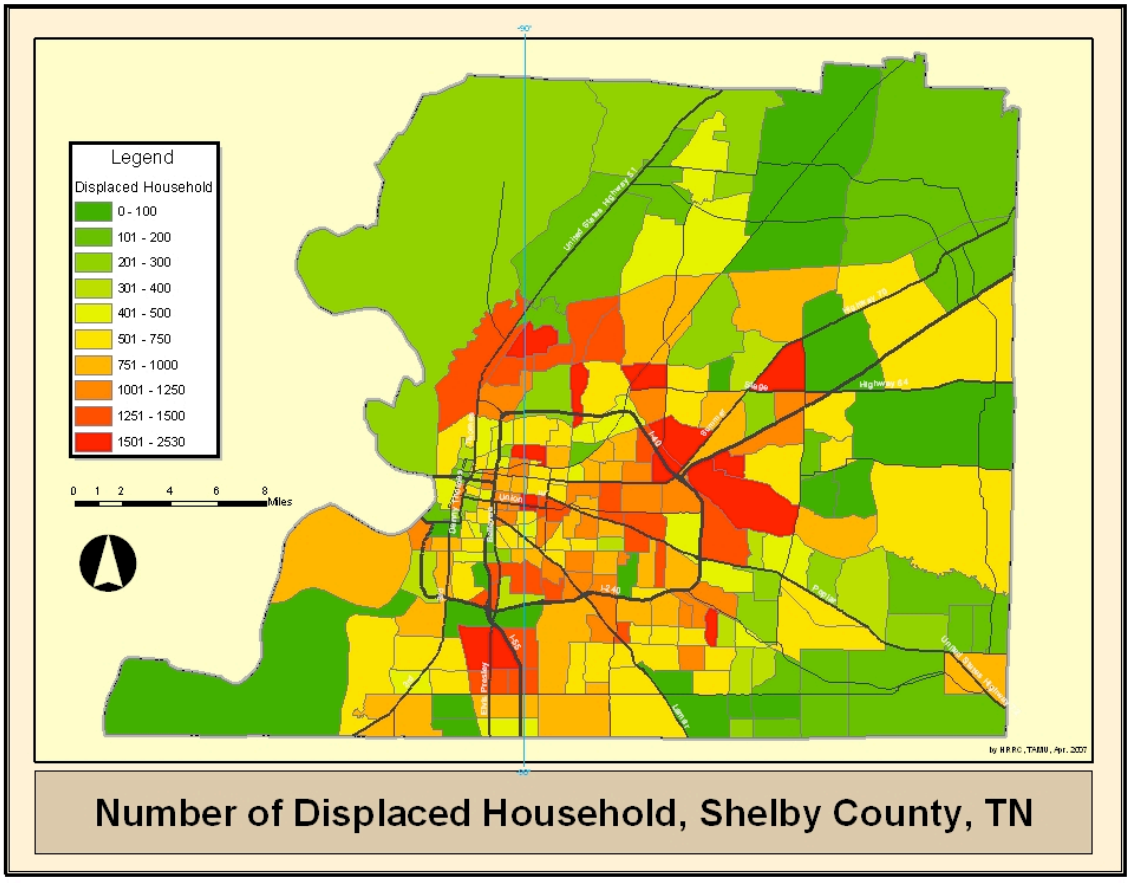
3. Calculating total number of dislocated households for a jurisdiction covering p block groups:

$$TotDh = \sum_{k=1}^{p} DisHh_{bg_k}$$

Appendix 3. Example of a fictions report of displaced household by jurisdiction (Shelby county) and by census block group.

	Number of Displaced Household	Percent of Displaced Household
Shelby County, TN	153232	46%
Block Group		
47157XXXXXXXX	383	73%
47157XXXXXXXX	453	68%
47157XXXXXXXX	494	59%
47157XXXXXXXX	231	71%
47157XXXXXXXX	673	58%
47157XXXXXXXX	1592	69%
……..		
47157XXXXXXXX	797	55%
47157XXXXXXXX	921	59%
47157XXXXXXXX	858	59%

Appendix 4 Example Number of Dislocated Household Map



Appendix 5 Example of Percent of Dislocated Household Map