

# **Technical Report**

## **Identifying Employment Subcenters in the LA Metro Area The Negative Exponential Procedure**

Jifei Ban

Department of Statistics, UC Riverside

[jban001@ucr.edu](mailto:jban001@ucr.edu)

July 19<sup>th</sup>, 2010

## Table of Contents

1. The negative exponential procedure
2. Maps Using 2003 TAZ Employment Data
3. Technical Appendix

### ***1. The Negative Exponential Procedure.***

The procedure we use is a revision of the GS procedure by Giuliano and Small (1991). The GS procedure defines a subcenter as a group of contiguous TAZ's, for which the employment density (number of employed people per acre) of each TAZ has to exceed  $d$  and the total employment of all the TAZ's in the group has to exceed  $D$ . In practice, one usually uses two pairs of cutoff values,  $d=20$ ,  $D=20000$ , and  $d=10$ ,  $D=10000$ . The problem with this procedure is: if the first pair of cutoffs is used, one will identify too few subcenters in the peripheral area; if the second pair of cutoffs is used, one will identify too many subcenters in LA county.

We propose a new procedure with varying cutoffs. We assume that the TAZ in LA county which has the highest employment density has cutoffs  $d_0 = 20$  and  $D_0 = 20000$ ; we call this TAZ *the employment center*. For example, in the 2003 SCAG employment data, its ID is "235132". As the distance from one TAZ to the employment center increases, the two cutoffs will decrease exponentially. If the centroid of one TAZ is  $r$  miles away from the centroid of the employment center, the cutoff  $d$  for this TAZ will be  $d = d_0 \exp(-k_1 * r)$ , where  $k_1$  is a positive number describing the decreasing rate. In the maps we generated, we choose  $k_1$  as  $\log(2)/40$ , which means the cutoff  $d$  halves when a TAZ is 40 miles away from the employment center. We call those TAZ's with employment density exceeding their corresponding cutoff  $d$ 's candidate TAZ's. A group of contiguous candidate TAZ's will be defined as a subcenter if the total employment exceeds  $D$ . We assume the same negative exponential model for  $D$  as  $d$ . We define the distance from a group of candidate TAZ's to the employment center as the employment weighted distance from these TAZ's to the employment center. For example, there are  $n$  TAZ's in a group of candidates.  $r_i, i = 1, \dots, n$ , is the distance from the  $i$ -th TAZ to the employment center;  $e_i, i = 1, \dots, n$ , is the employment in  $i$ -th TAZ. Then  $R$ , the distance

from the group to the employment center, is defined as  $R = \frac{\sum e_i r_i}{\sum e_i}$ . The cutoff  $D$  for a

group of candidate TAZ's will be  $D = D_0 \exp(-k_2 * R)$ . In the maps we generated, we choose  $k_2$  as  $\log(2)/40$ , which again means the cutoff  $D$  halves when a group of candidate TAZ's is 40 miles away from the employment center. We define those groups of candidate TAZ's with employment exceeding their corresponding cutoff  $D$ 's as subcenters.

Note: the coordinates of the centroid of each TAZ are contained in the shp file provided by SCAG.

## ***2. Maps Using 2003 SCAG Employment data***

Using the above procedure and the 2003 employment data of the LA Metro Area provided by SCAG, we generated the following maps with subcenters identified for the 6 counties of LA Metro Area. There are 51 subcenters in total, with 19 in LA, 11 in Riverside, 9 in Orange, 5 in San Bernardino, 4 in Imperial, 1 in Ventura, 1 on the boundary of LA and Orange, and 1 on the boundary of LA and San Bernardino.



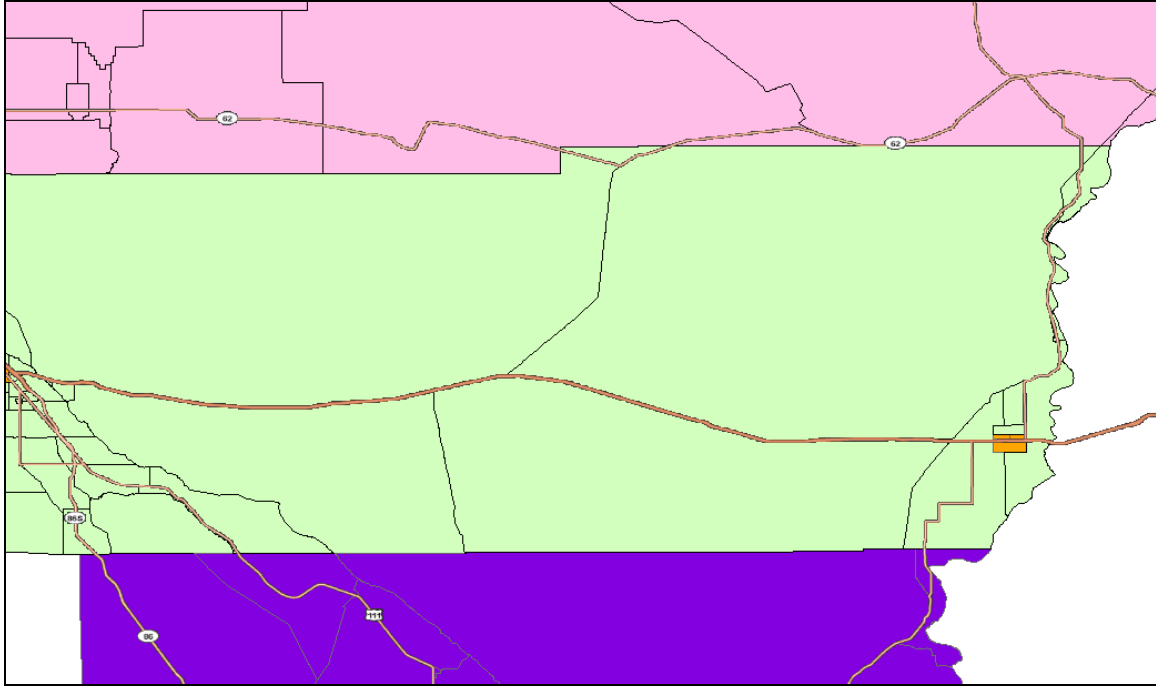


Figure 3: Riverside County East Portion, 2003

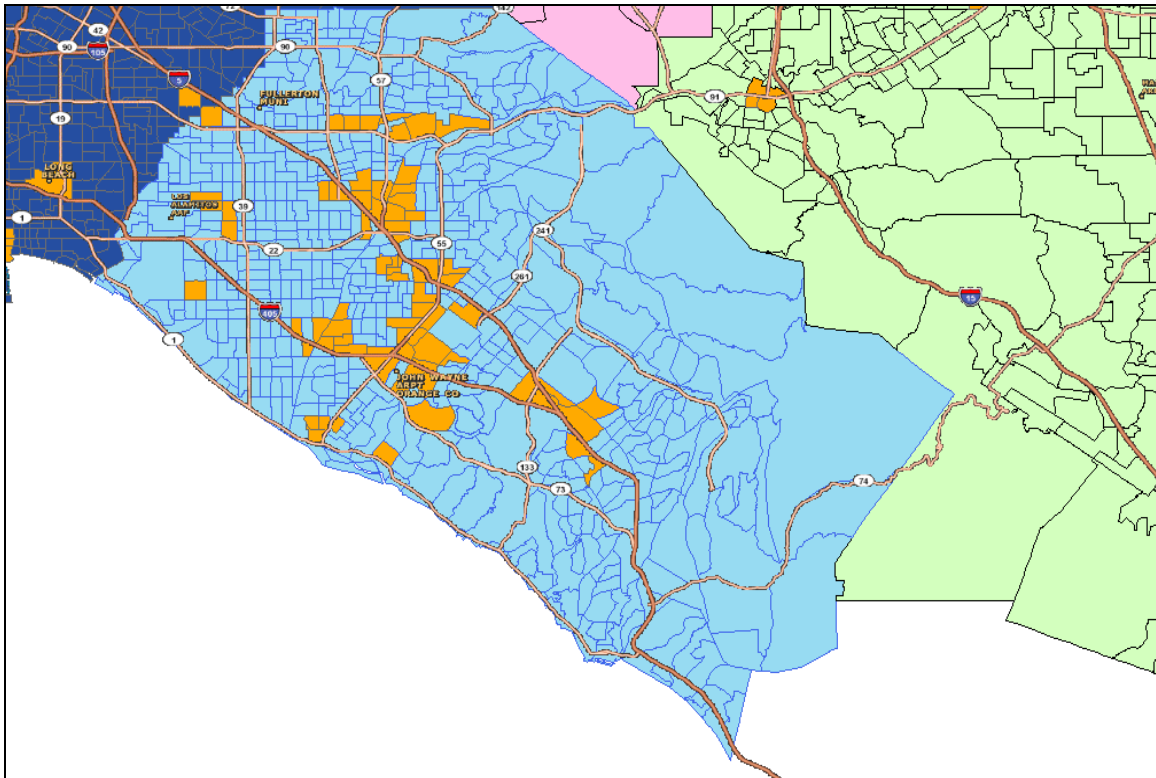


Figure 4: Orange County, 2003

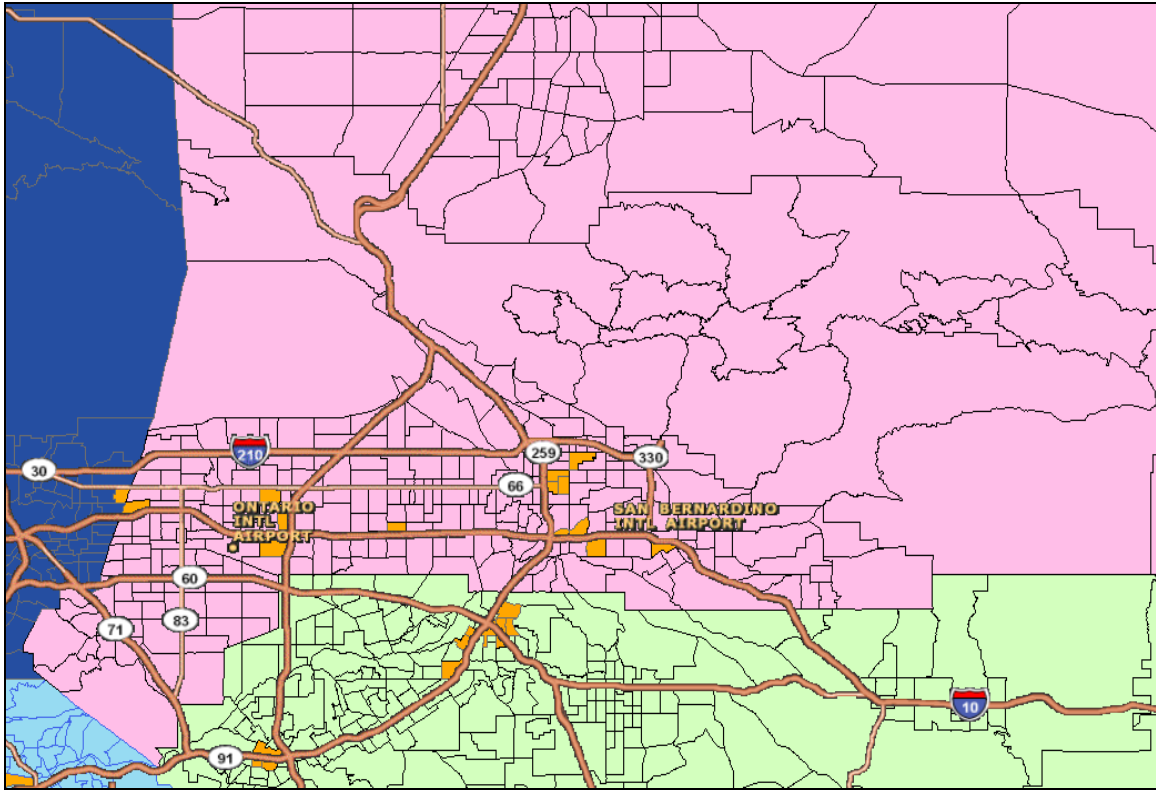


Figure 5: San Bernardino County, 2003

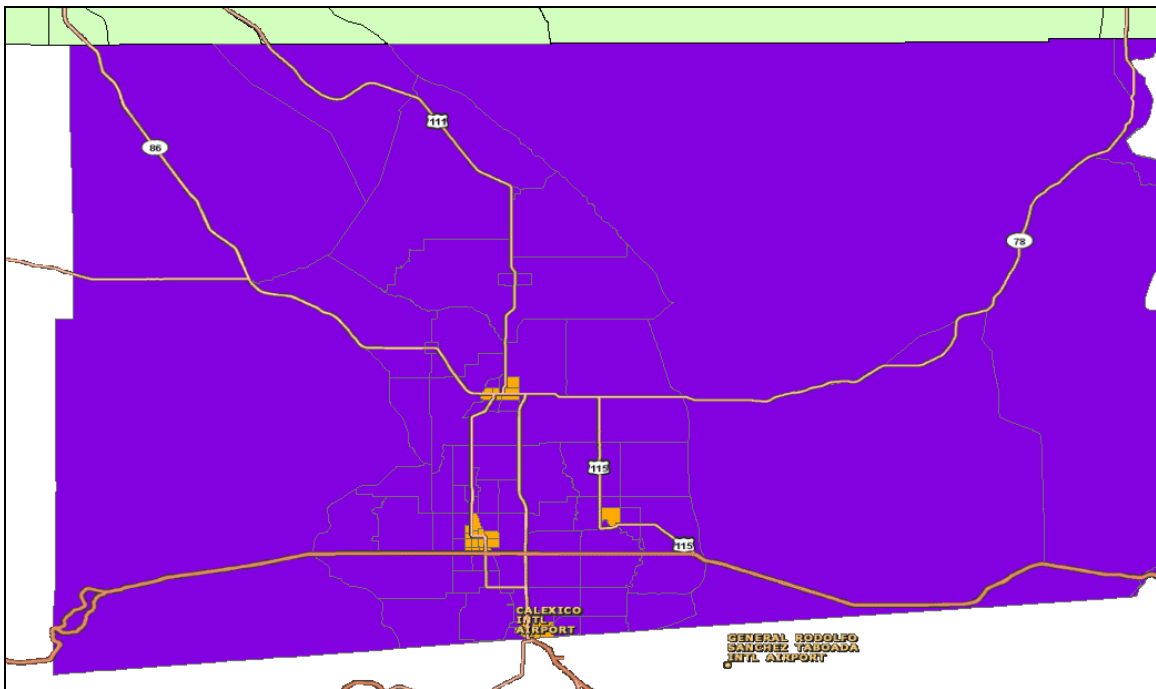


Figure 6: Imperial County, 2003

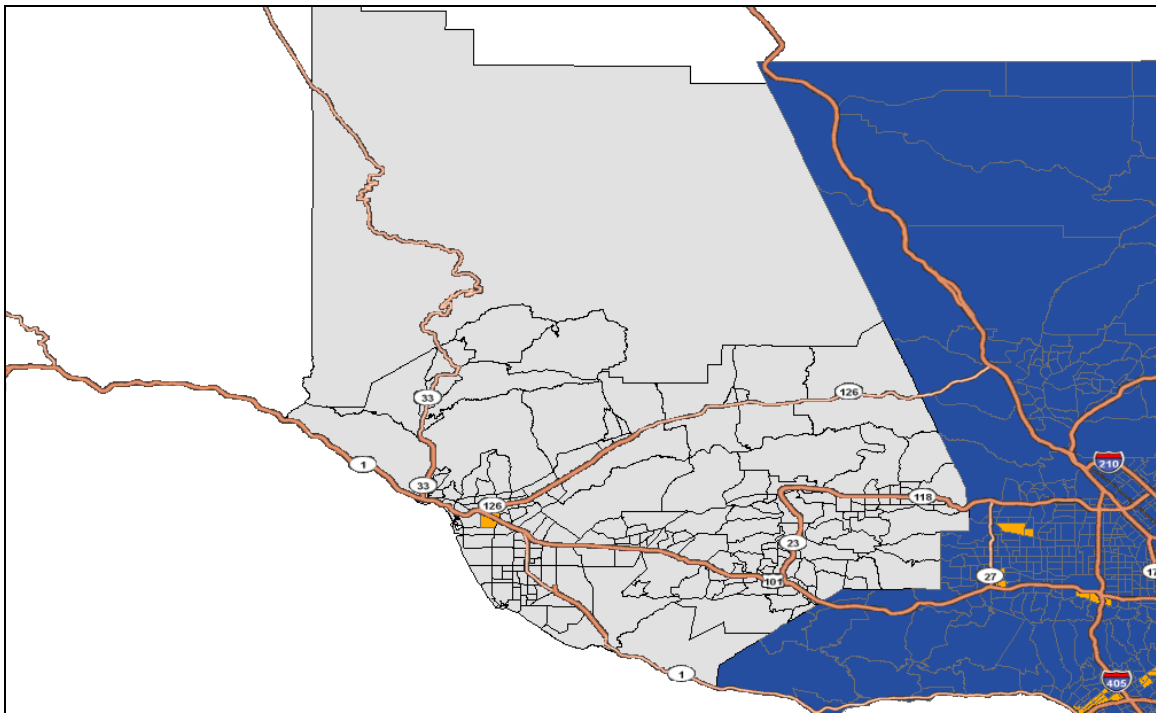


Figure7: Ventura County, 2003



### **3. Technical Appendix**

We have developed the function for the negative exponential procedure in the computer program R based on the code provided in Arnott and McMillen's first report. The function that will carry out the negative exponential procedure is "sub\_ne". Before calling that function, one needs to follow the direction in the technical appendix in the first report to set up the preliminaries before going into the command section. The commands that one needs to call in R by order are listed as follows:

#### **# Section 1 – read data files**

```
library(spdep)
shp_file<-readShapePoly("/scag/maps/la_taz_merged_shape.shp",ID="ID")
gal_file <- read.gal("/scag/maps/la_taz_merged_shape.gal",region.id=shp_file$ID)
```

#### **# Section 2 – finding geographic coordinates and re-labeling data**

```
lmat <- coordinates(shp_file)
longitude <- lmat[,1]
latitude <- lmat[,2]
employment <- shp_file$EMP03
area <- shp_file$ACRE
id <- shp_file$ID
taz2k <- shp_file$TAZ2K
```

#### **# Section 3 – Data manipulation and definition of contiguity matrix**

```
empdens <- employment/area
empdens <- ifelse(is.na(empdens),0,empdens)
summary(empdens)
lndens <- ifelse(empdens>0,log(empdens),log(.5/area))
wmat <- nb2mat(gal_file,style="B")
```

#### **# Section 4 – Negative exponential method**

```
source("/rfunctions/sub_ne.r")
nesub<-sub_ne(file=gal_file,coordinates=lmat,wmat=wmat,
dens=empdens,emp=employment,k1=log(2)/40,k2=log(2)/40,d0=20,D0=20000)
subobs_ne <- nesub$subobs
```

#### **# Section 5: Write results to file**

```
submat <- cbind(id,taz2k,longitude,latitude,subobs_ne)
write.csv(submat,"/scag/maps/subcenter.csv")
```

Discussion:

**Section 1-3.** The first three sections are exactly the same as those in the first report.

**Section 4.** This section identifies subcenters using the negative exponential procedure. The first line reads in our R function “sub\_ne.r”. Alternatively, the full program could be included directly in the main R program at this point. The first five arguments in the function simply feed the required data into the function. These arguments include the gal file (which was given the name gal\_file in the third line of section 1), the coordinates of the centroids of all the TAZ’s (“coordinates=lmats”), the contiguity matrix (“wmat=wmat”), the employment density variable (“dens=empdens”), and the variable representing total employment in a TAZ (“emp=employment”). The names on the right hand side of the equal signs can be varied if desired; the names on the left hand side are required by the function. The next two arguments, “k1” and “k2”, are the decreasing rates in the negative exponential models for the employment density cutoff and the total employment cutoff respectively. The last two arguments are the cutoff values for the employment center: “d0” is the minimum employment density and “D0” is the total employment cutoff. The variable nesub\_obs equals one if the TAZ observation is indicated to be part of a subcenter; it equals zero otherwise.

**Section 5.** This section is the same as the “Section 7” in the first report. It writes the matrix “submat” into a csv file. This file can be read into a GIS program to map the subcenters.

The code for the R function “sub\_ne.r” is provided below with some technical details explaining how the negative exponential procedure works.

### **The Negative Exponential Procedure**

```
sub_ne<-function(file,coordinates,wmat,dens,emp,k1,k2,d0,D0)
{ library(spdep)
  if (sum(wmat)==0) { wmat <- nb2mat(file,style="B") }
  dens <- ifelse(is.na(dens),0,dens)
  center_coor<-coordinates[dens==max(dens),]
  a<-57.2958
  dist_center<-3963.0*acos(sin(center_coor[2]/a)*sin(coordinates[,2]/a)+
  cos(center_coor[2]/a)*cos(coordinates[,2]/a)*cos(coordinates[,1]/a-center_coor[1]/a))
  cutoff<-d0*exp(-k1*dist_center)
```

```

obs <- seq(1:length(dens))
densobs <- obs[dens>cutoff]
wmat <- wmat[densobs,densobs]
n = nrow(wmat)
amat <- matrix(0,nrow=n,ncol=n)
amat[row(amat)==col(amat)] <- 1
bmat <- wmat
wmat1 <- wmat
newnum = sum(bmat)
cnt = 1
while (newnum>0) {
  amat <- amat+bmat
  wmat2 <- wmat1%*%wmat
  bmat <- ifelse(wmat2>0&amat==0,1,0)
  wmat1 <- wmat2
  newnum = sum(bmat)
  cnt = cnt+1
}
emat <- emp[densobs]
tmat <- amat%*%emat

empw<-t(emat*t(amat))
empw<-empw/rowSums(empw)
D_center<-empw%*%dist_center[densobs]

totemp<-D0*exp(-k2*D_center)

obsmat <- densobs[tmat>totemp]
subemp <- array(0,dim=length(dens))
subemp[obsmat] <- tmat[tmat>totemp]
subobs <- ifelse(subemp>0,1,0)

tab <- tabulate(factor(subemp))
numsub = sum(tab>0)-1

cat("Number of Subcenters = ",numsub,"\n")
cat("Total Employment and Number of Tracts in each Subcenter", "\n")
print(table(subemp))

```

```

out <- list(subemp,subobs)
names(out) <- c("subemp","subobs")
return(out)
}

```

This code is based on the code for function “sub\_gs.r” in the first report. Most of them are similar, except that: 1.”cutoff” here is a vector, which contains the minimal employment density for each TAZ, whereas, in the GS procedure, it is a single value represented by “mind”; 2. “totemp” here is also a vector, which contains the minimal total employment for each set of contiguous candidate TAZ’s, whereas, in the GS procedure, it represents a single value.

“center\_coor” represents the coordinates of the centroid of the employment center. “dist\_center” contains the distance from each TAZ to the employment center. It is obtained by calculating the distance between two pairs of latitude- longitude coordinates. “cutoff<-d0\*exp(-k1\*dist\_center)” gives the minimal employment density for each TAZ based on the negative exponential model. Those TAZ’s with employment densities exceeding their corresponding thresholds are defined as candidate TAZ’s. “empw<-empw/rowSums(empw)” weights each TAZ in a set of contiguous candidate TAZ’s by its employment. “D\_center” is a vector containing the weighted distance from each TAZ in a set of contiguous candidate TAZ’s to the employment center. ”totemp<-D0\*exp(-k2\*D\_center)” gives the minimal total employment for each set of contiguous candidate TAZ’s based on the negative exponential model. Any set which has a total employment exceeding its minimal total employment is defined as a subcenter.

The primary arguments for the procedure are:

- file: the gal file defined by GeoDa. If this file is specified without providing the argument wmat, the procedure will use spdep to generate the contiguity matrix.
- coordinates: the latitude-longitude coordinates of the centroid of each TAZ.
- wmat: the contiguity matrix.
- dens: the employment density variable.
- emp: the variable showing total employment on each TAZ.
- k1 and k2: the decreasing rates in the negative exponential models for the employment density cutoff and the total employment cutoff respectively.
- d0 and D0: the employment density cutoff and the total employment cutoff for the employment center.

## References

R. Arnott and D. McMillen, Final Report, Identifying Employment Subcenters in the LA Metro Area, A Comparison of Three Procedures (2009).

G. Giuliano and K. A. Small, Subcenters in the Los Angeles region, *Regional Science and Urban Economics*, 21, 163–182 (1991).