Random Trip Assignment Model

**Objective**

As part of the input files required by the simulation framework, a dataset containing trip information with nodes set as origin and destination needs to be prepared. This dataset can be generated based on an OD estimation matrix which provides the traffic flow data between different strategic zones. Thus this Random Trip Assignment Model aims to assign trips between zones to corresponding nodes within the zones, following some criteria that preserve some randomness as well as take some environmental factors into consideration.

**Setting**

Available files: Changi MTZ zones, Singapore nodes, Changi establishments data, Changi firm data, OD estimation matrix

Output file: OD data (node)

Scale: Changi area

**Methodologies**

To produce the required dataset, there are several steps of manipulation we need to implement on the current data we have:

Trip distribution from zones to nodes

Each zone may contain a multitude of nodes. To spread the total number of trips of an entire zone over all the nodes, we would like to be realistic, in terms of respecting the actual distribution of trips, as well as stochastic, so that such distribution is not deterministic. With the data we we have, there are a number of criteria that we can consider:

1. Number of establishments covered by fixed distance buffers

We could create fixed distance buffers around the nodes, and measure the number of establishments that are covered by each node’s buffer. Distribution of trips within the zone can be thus proportional to this number of all the nodes within the zone.

1. Number of jobs covered by fixed distance buffers

This method also makes use of buffers centred around the nodes, but relies on the number of jobs as criterion. The job number information is available from the Changi firm data.

1. Number of jobs of the nearest firm

Instead of capturing all the jobs that fall in the buffer region, we could also associate only one nearest firm to each node. Trip distribution can therefore be based on the number of jobs brought about the nearest firm to each node.

The problem with establishment data is that it is not a reliable gauge of the density of goods vehicle traffic flow. Establishments subsume residence area, which might not incur traffic of goods vehicles. The number of jobs, therefore, might give us more information about the likelihood of generating freight trips.

Moreover, after examining the distribution results of using buffers, we found out that it’s hard to fine-tune the size of the buffer to strike a balance. Larger buffers will create many double counting of the establishments/jobs because they could be densely distributed in some regions; on the other hand, buffers that are too small will leave a lot of nodes unable to capture any trip because there are simply no establishments/jobs within their buffer region. Thus, to avoid this complication we adopted the third method, which is to distribute trips in proportion to the number of jobs created by the nearest firm of each node.

However we do not follow this ratio deterministically. This distribution of jobs within the zone only serves as a probability benchmark to make our assignment. So node associated with the largest amount of jobs is more likely to assigned a trip starting from/ending at this particular zone. However there is no guarantee that the final distribution of trips will follow the distribution of jobs.

Cross-boundary trips distribution

Some OD trips could be across the Changi boundary, which means either the origin or destination is some region outside the predefined Changi area. To be able to still capture the traffic created by such cross-boundary trips, we appointed several boundary nodes to substitute external zones. Essentially this means any origin or destination that is outside Changi area will be assigned to one of the boundary nodes.

With this setting established, we need a criterion to decide which boundary node we should select for each cross-boundary trip. We compare between two options:

1. Road types

This method hinges on the road type of each boundary node. Here we assume some arbitrary values on each road type such that certain road types are more capable of attracting traffic than the others. For example, expressway might take up the largest proportion of trips, around 60%.

1. Proximity to Changi origin/destination

We could also make a more adaptive decision such that the assignment is determined by which Changi zone the trip starts from/ends in. The probability of assigning the boundary node to the trip is inversely proportional to the distance between this node to the Changi zone.

After some comparisons, the second method proved to be superior in two ways. First, it is able to differentiate each trip to a greater extent so that it’s far less crude than the first method. Second, it is a stochastic method that assign the trips based on some prior probabilities.

**Implementation Detail**

With some of the above key considerations, we were able to build this model with QGIS and R in conjunction. The basic workflow is demonstrated as follow (blue box implemented in QGIS and grey box implemented in R):

Extract associated number of jobs of each node based on “nearest” principle

Extract distance matrix of all the distances between boundary nodes and centroids of Changi zones

Calculate zonal probabilities probability based on job distribution

Calculate boundary node probabilities with respect to each zone based on distance matrix

Implement the random distribution based on the generated probabilities

Generation of dataset into accepted format

**Appendix:**

1. Implementation of number of establishments covered by fixed distance buffers + road types
2. Implementation of number of jobs covered by fixed distance buffers + road types
3. Implementation of number of jobs of the nearest firm + proximity to Changi origin/destination